Modeling and Forecasting Carbon Dioxide Emission Allowance Spot Price Volatility: Multifractal vs. GARCH-type Volatility Models*

Mawuli Segnon[†]

Department of Economics, University of Kiel,
Germany

Thomas Lux[‡]

Department of Economics, University of Kiel,
Kiel, Germany &
Bank of Spain Chair of Computational Economics
Department of Economics, University Jaume I
Castellón, Spain

and

Rangan Gupta[§]

Department of Economics, University of Pretoria,
South Africa

July 15, 2015

Abstract: This paper applies Markov-switching multifractal (MSM) processes to model and forecast carbon dioxide (CO_2) emission price volatility, and compares their forecasting performance to the standard GARCH, fractionally integrated GARCH (FIGARCH) and the two-state Markov-switching GARCH (MS-GARCH) models via three loss functions (the mean squared error, the mean absolute error and the value-at-risk). We evaluate the performance of these models via the superior predictive ability test. We find that the forecasts based on the MSM model cannot be outperformed by its competitors under the vast majority of criteria and forecast horizons, while MS-GARCH mostly comes out as the least successful model. Applying various VaR backtesting procedures, we do, however, not find significant differences in the performance of the candidate models under this particular criterion. We also find that we cannot reject the null hypothesis of MSM forecasts encompassing those of GARCH-type models. In line with this result, optimally combined forecasts do indeed hardly improve upon the best single models in our sample.

*Financial support by the European Union's 7th Framework Programme under grant agreement no. 612955 is gratefully acknowledged.

[†]Correspondence: Mawuli Segnon, Department of Economics, Leibnizstrasse 3, 24118 Kiel, Germany. E-mail: m.segnon@economics.uni-kiel.de.

[‡]Thomas Lux, Department of Economics, University of Kiel Olshausenstrasse 40, 24118 Kiel, Germany. E-mail: thomas.lux@bwl.uni-kiel.de.

[§]Rangan Gupta, Department of Economics, Pretoria, 002, South Africa. E-mail: rangan.gupta@up.ac.za.

Keywords *Carbon dioxide emission allowance prices, GARCH, Markov-switching GARCH, FIGARCH, Multifractal Processes, SPA test, encompassing test, Backtesting*

JEL classification Q47

1 Introduction

Since the market for European Union Allowances (EUAs) had been launched on January 1, 2005, a large number of studies have investigated the relationship between EUA spot prices and energy prices, such as oil, gas, and coal (cf. Hammoudeh et al., 2014a,b), and extreme weather events (cf. Mansanet-Bataller et al., 2007), economic activity (cf. Chevalier, 2009) and institutional changes such as the revelation of information on emissions caps (cf. Alberola et al., 2008). All these studies shed light on the different factors that may affect the short-term demand for EUAs. However, little attention has been paid to modeling and forecasting EUA spot price volatility, which with the development of derivative markets appears to be of particular importance for investors and energy companies. Daskalakis et al. (2009) utilize a geometric Brownian motion with an additional jump component to describe the random behavior of the carbon dioxide (CO_2) emission spot price. They find that the jump-diffusion model properly reproduces the non-stationarity and abrupt discontinuous shifts observed in CO_2 price levels. Benz and Trück (2009) use a Markov-switching model, and a standard GARCH(1,1) model to analyze the heteroskedastic behavior of carbon dioxide emission allowance return series, while Paolella and Taschini (2008) employ an AR(1)-GARCH(1,1) model with different innovations (Student's-t, symmetric and asymmetric stable, and the generalized asymmetric t-distributions). Benz and Trück (2009) evaluate and compare the one-day ahead forecasting performance of their models via the mean squared error (MSE), the mean absolute error (MAE) and the Kolmogorov-Smirnov and Kupiec tests. They find that both, the Markov-switching and the standard GARCH(1,1) models, perform well and dominate over models with constant variance. Paolella and Taschini (2008) also find that the GARCH model with different distributional assumptions for the innovation provides accurate one-day ahead out-of-sample value-at-risk forecasts. Recently, Benschop and Cabrera (2014) question the performance of the standard GARCH(1,1) due to the fact that structural breaks may occur in the market for European Union Allowances, and the standard GARCH(1,1) under such circumstances would not be able to properly capture the heteroskedastic behavior of the CO_2 emission allowance return series. These authors propose regime switching GARCH models that might be appropriate for covering infrequent structural changes in the data and compare their one-day ahead forecasting performance to those of the standard GARCH and simple AR models. Using MSE and MAE as criteria and Kolmogorov-Smirnov tests for performance evaluation they find that the Markov-switching GARCH models outperform both the simple Markov-switching and simple GARCH models.

In this paper, we propose a completely different approach for modeling and forecasting CO_2 price volatility. This new approach is based on multifractal processes that have first been developed in the theory of turbulent flows (e.g. Mandelbrot, 1974) and have been adapted to model financial volatility by Calvet and Fisher (2001, 2004). The robustness and the capacity of the process to dominate GARCH-type models in terms of forecasting volatility of various financial time series has been demonstrated in Calvet and Fisher (2004), Lux (2008), Lux and Morales-Arias (2010), Lux et al. (2014), and Lux et al. (2015). The attractiveness of the multifractal model stems from the fact that it provides a simple uniform framework for both long-term persistence in the volatility process and structural breaks through regime switching.

Previous studies of carbon dioxid emission allowances have concentrated on the comparison of one-day ahead forecasting performance of volatility models with constant and conditional variances. However, volatility forecasts for longer horizons are vital inputs for option pricing and risk management decisions. We, therefore, will compare the forecasting ability of the new Markov-switching multifractal (MSM) model with those of the standard generalized autoregressive conditional heteroskedasticity (GARCH), fractionally integrated GARCH (FIGARCH) and the two-state Markov-switching GARCH (MS-GARCH) at short and long horizons, thereby filling these gaps in the existing literature.

The rest of the paper is organized as follows. Section 2 presents the CO_2 prices and their descriptive statistics. The volatility models are described in Section 3. In Section 4 we provide the results of the empirical application and finally, Section 5 concludes.

2 Data

We use daily carbon dioxide (CO_2) allowance prices (in €) that are obtained from the European Energy Exchange (EEX) through Thomson Reuters DataStream. The data set covers the price observations from January 16, 2009 to January 20, 2015. During this time the carbon prices have fluctuated between 2.68 € and 16.84 € per tone of CO_2 (tCO_2). The starting point of our data set can be explained by the fact that the carbon price after October 2006 plunged to around € 0.01/tCO_2 as of March 11, 2008. A few potential reasons for this drop can be found in Yun and Baker (2009). Including this episode would presumably mean that we would have to cope with non-stationary dynamics within some subset of our time series and so we decided to confine ourselves to the post-2008 era over which no such collapse has been observed any more. The end-point of the sample is purely driven by data availability at the time of writing this paper. We compute the percent continuously compounded returns r_t as

$$r_t = 100 * [\ln(p_t) - \ln(p_{t-1})], \qquad (1)$$

where p_t denotes the carbon dioxide price at period t.

Fig. 1 depicts the time evolution of CO_2 allowance prices, and their log-returns and squared returns. The descriptive statistics of the log-returns, absolute and squared returns are reported in Table 1. Our data exhibit negative skewness and excess kurtosis. These results show that the computed log-returns do not follow a Normal distribution. This observation is confirmed by the Jarque-Bera test, which rejects the null hypothesis of Normally distributed log-returns at any level of significance. The augmented Dickey-Fuller (ADF) unit-root test of Dickey and Fuller (1979) indicates the stationarity of CO_2 log-returns. The Hurst indices reported in Table 1 are computed via Detrended Fluctuation Analysis (DFA) (cf. Weron, 2002). The Hurst indices are standard measures of long-term dependence. The values for log-returns are close to 0.5, implying absence of long memory features in CO_2 price returns. For absolute and squared returns the Hurst index values are significantly above 0.5, indicating the presence of long memory in CO_2 price volatility. We also compute the so-called Hill estimator for the tail index (cf. Hill, 1975) in order to quantify the decay of the unconditional distribution of CO_2 price returns in its extremal region. The estimate for the tail index (cf. Table 1) is in the vicinity of 3 and this result is in harmony with typical findings for financial assets, cf. Lux and Ausloos (2002). Again, this is a clear indication of non-Normality as the slow decay of the distribution of returns implies that large price changes occur with much higher probability than under a Gaussian shape.

Fig. 2 shows the autocorrelation functions for log-returns, absolute and squared log-returns. We observe that the absolute and squared log-returns are highly correlated, and this observation is in conformity with the Ljung-Box statistics, $Q(10)$ and $Q(20)$. However, the Ljung-Box tests also reject the null hypothesis of no serial correlation for raw log-returns at the 5% significance level. This indicates the presence of some serial dependence and predictability in the returns of CO_2 allowances. Overall, the carbon dioxide emission allowances, therefore, share the typical salient features of financial assets that are captured by the catchwords "fat tails" and "clustered volatility", but show some deviation from the "efficiency" of stock or foreign exchange markets (their complete lack of predictability). Hence, the data document a far reaching "financialisation" of the market for emission allowances, which, however, seems not deep and liquid enough to remove all predictable patterns like in archetypical stock markets of developed countries. Note that financialisation implies that large changes of the prices and waves of high and low fluctuations might be caused by the trading process and might not necessarily reflect exogenous sources of uncertainty.

Finally, we apply the modified iterated cumulative sum of squares (ICSS) algorithm [5] to test whether multiple breaks occur in the second moment at the 5% significance level. We obtain two break points that occur on March 24, 2009 and November 16, 2011, respectively. These break points motivate the use of Markov-switching models that can take into account such structural changes. In the multifractal model, such structural changes might be covered by its very principle of construction.

3 Model Framework

In this section we briefly present the volatility models used for forecasting CO_2 price volatility. In these models returns are formalized as

$$r_t = \mu_t + \sigma_t e_t, \tag{2}$$

where r_t is computed as in *eq.* (1), $\mu_t = \mathbb{E}_{t-1}[r_t]$ is the conditional mean of the return series, σ_t is the volatility process and e_t is standard Normally distributed. Defining $x_t = r_t - \mu_t$, the *centered* returns are given by

$$x_t = \sigma_t e_t. \tag{3}$$

Here we assume that μ_t follows an AR(1) process and consider the standard GARCH, the FIGARCH and Markov switching GARCH (MS-GARCH) models and the Markov switching multifractal (MSM) model for describing σ_t.

3.1 GARCH-type Models

Since its introduction by Engle (1982), the autoregressive conditional heteroskedasticity (ARCH) model and its subsequent generalized versions have become the major paradigm for modeling volatility due to their ability to capture the most important *stylized facts* (e.g. clustering effects, nonlinearity, long-memory and short-memory effects, asymmetric leverage effects) observed in all measures of volatility (e.g. absolute log-returns, squared log-returns, etc...). In the following we present three different GARCH models.

[5] We refer the reader to Sansó et al. (2004) for more details

3.1.1 The GARCH Model

The simplest GARCH(1,1) models the conditional variance as

$$\sigma_t^2 = \omega + \alpha x_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{4}$$

where $\omega > 0$, $\alpha > 0$, $\beta > 0$ and $\alpha + \beta < 1$. The nonnegativity constraints on ω, α and β guarantee the positivity of σ_t^2 (cf. Bollerslev, 1986).

h-step ahead forecasts from GARCH(1,1) are obtained recursively as

$$\begin{aligned}\hat{\sigma}_{t+h}^2 &= \omega + (\alpha + \beta)\,\hat{\sigma}_{t+h-1}^2 \\ &= \bar{\sigma}^2 + (\alpha + \beta)\left(\hat{\sigma}_{t+h-1}^2 - \bar{\sigma}^2\right) \\ &= \bar{\sigma}^2 + (\alpha + \beta)^{h-1}\left(\hat{\sigma}_{t+1}^2 - \bar{\sigma}^2\right).\end{aligned} \tag{5}$$

where $\bar{\sigma}^2 = \omega(1 - \alpha - \beta)^{-1}$ is the unconditional variance. As $h \to \infty$, it is clear that the volatility forecast in *eq.* (5) approaches the unconditional variance $\bar{\sigma}^2$ and $(\alpha + \beta)$ dictates the speed of the mean reversion.

3.1.2 The Fractionally Integrated GARCH Model

The FIGARCH model developed by Baillie et al. (1996) consists in introducing fractional differences in the GARCH process and thereby allows the model to reproduce the long-term dependence of financial returns volatility as documented in the high Hurst coefficients of absolute and squared returns. This means that these volatility measures are characterized by a slowly decaying autocorrelation function rather than an exponentially decaying one (as imposed, for instance, by the baseline GARCH approach). The conditional variance in the FIGARCH(1,d,1) model can be formalized as

$$\sigma_t^2 = \omega + \left[1 - \beta(L) - \phi(L)(1 - L)^d\right] x_t^2 + \beta \sigma_{t-1}^2, \tag{6}$$

where $\omega > 0$, $\phi < 1$, $\beta < 1$, $0 \le d \le 1$. L denotes the lag operator and d is the parameter of fractional differentiation. The parameters have to fulfill the following conditions:

$$\beta - d \le \phi \le \frac{(2 - d)}{3} \tag{7}$$

and

$$d\left[\phi - \frac{(1 - d)}{2}\right] \le \beta(d - \beta + \phi). \tag{8}$$

We can rewrite *eq.* (6) as follows

$$\begin{aligned}\sigma_t^2 &= \omega(1 - \beta)^{-1} + \left[1 - (1 - \beta)\phi(L)(1 - L)^d\right] x_t^2 \\ &= \omega(1 - \beta)^{-1} + \eta(L) x_t^2,\end{aligned} \tag{9}$$

where $\eta(L) = \eta_1 L + \eta_2 L^2 + \ldots$, $\eta_j \ge 0$ for $j = 1, 2, \ldots$.

$\eta(L)$ can be computed from the recursions:

$$\begin{cases} \eta_1 = & \hat{\phi} - \hat{\beta} + \hat{d}, \\ \vdots & \vdots \\ \eta_j = & \hat{\beta}\eta_{j-1} + \left[\left(j - 1 - \hat{d}\right) j^{-1} - \hat{\phi}\right] \pi_{j-1} \end{cases} \tag{10}$$

where $\pi_j \equiv \pi_{j-1}\left(j - 1 - \hat{d}\right) j^{-1}$ are the coefficients in the MacLaurin series expansion of the fractional differencing operator $(1 - L)^d$. The FIGARCH model reduces to the GARCH model when $d = 0$.

From *eq.* (9) one can easily derive the one-step ahead forecast of σ_t^2

$$\hat{\sigma}_{t+1}^2 = \omega(1-\beta)^{-1} + \eta_1 x_t^2 + \eta_2 x_{t-1}^2 + \ldots . \tag{11}$$

Using recursive substitution described above the h-step ahead forecasts of the FIGARCH(1,d,1) model are obtained as

$$\hat{\sigma}_{t+h}^2 = \omega\,(1-\beta)^{-1} + \sum_{i=1}^{h-1} \eta_i \hat{\sigma}_{t+h-i}^2 + \sum_{j=0}^{\infty} \eta_{h+j} x_{t-j}^2. \tag{12}$$

Long-term dependence shows up in the fact, that, in principle, all available past data should be used in the construction of forecasts of future volatility (while in GARCH, its short-term dependence makes it sufficient to use the filtered realization of the conditional variance, σ_t, at the forecast origin, time t).

3.1.3 The Markov-Switching GARCH Model

Following Klaassen (2002) the conditional variance in the two state Markov-switching GARCH(1,1) model can be formalized as

$$\sigma_{t-1}^2\{x_t|\tilde{\delta}_t\} = \omega_{\delta_t} + \alpha_{\delta_t} x_{t-1}^2 + \beta_{\delta_t}\mathbb{E}_{t-1}\left[\sigma_{t-1}^2\{x_{t-1}|\tilde{\delta}_{t-1}\}|\delta_t\right], \tag{13}$$

where $\tilde{\delta}_t$ is the regime path $(\delta_{t-1}, \delta_{t-2}, \ldots)$, $\delta_t \in \{1, 2\}$ is the variance regime at time t and follows a first-order Markov process, and $\omega_{\delta_t} > 0$, $\alpha_{\delta_t}, \beta_{\delta_t} \geq 0$. The expectation on the right-hand-side is across the regime path $\tilde{\delta}_t$ conditional on information $\Im_{t-1} = (x_{t-1}, x_{t-2}, \ldots)$ and δ_t.

Volatility forecasts are obtained as:

$$\hat{\sigma}_{t,t+h}^2 = \sum_{i=1}^{2} Pr\left(\delta_{t+h} = i|\Im_t\right)\hat{\sigma}_{t,t+h}^{2(i)}, \tag{14}$$

where $Pr\left(\delta_{t+h} = i|\Im_t\right)$ denotes the conditional probability of regime i at h periods ahead and $h-$step ahead volatility forecast in regime i made at time t can be computed recursively

$$\hat{\sigma}_{t,t+h}^{2(i)} = \omega^{(i)} + \left(\alpha^{(i)} + \beta^{(i)}\right)\mathbb{E}\left(\sigma_{t,t+h-1}^{2(i)}|\delta_{t+h}\right). \tag{15}$$

As with the standard GARCH(1,1) and FIGARCH(1,d,1), the MS-GARCH(1,1) is also estimated via the maximum likelihood approach as it is customary in applied finance.

3.2 The Markov-Switching Multifractal Model

Unlike the GARCH-type models, the recently developed Markov-switching multifractal (MSM) models are characterized by a multiplicative structure of the volatility process. In the MSM framework instantaneous volatility is modeled as a product of k volatility components or multipliers $M_t^1, M_t^2, \ldots, M_t^k$ and a positive scale factor σ^2 (cf. Calvet and Fisher, 2001, 2004; Lux, 2008). Formally, we have

$$\sigma_t^2 = \sigma^2 \prod_{i=1}^{k} M_t^{(i)}. \tag{16}$$

The multipliers or volatility components are assumed to be independent of each other at any time and satisfy $\mathbb{E}\left[M_t^i\right] = 1$. Each multiplier M_t^i is renewed at time t with probability γ_i depending on its rank within the hierarchy of multipliers and remains unchanged with probability $1-\gamma_i$. A formula for the transition probabilities, γ_i, that guarantee the convergence of the discrete-time MSM to a Poisson multifractal process in the continuous-time limit has been developed in Calvet and Fisher (2001). In this study we are not interested in the continuous-time limit, and therefore, we prefer to utilize the pre-specified transition probabilities proposed by Lux (2008) that are expressed as

$$\gamma_i = 2^{i-k}, \tag{17}$$

as this allows to reduce the number of parameters that need to be estimated. Note that this structure defines a Markov process with a large number of regimes that governs the volatility dynamics. The model avoids the curse of dimensionality of large Markov models via its systematic hierarchical arrangement of transition probabilities across states that in the specification of *eq.* (17) does not require any parameters to be estimated. The hierarchical structure with a spectrum of low and high frequencies of renewal of components by its very nature causes temporal dependence of the volatility process that indeed mimics a hyperbolic decay of the autocorrelation over a finite interval, i.e. up to time horizons $\approx 2^k$. With k large this becomes observationally undistinguishable from "true" long-term dependence. In this way, regime-switching and long-term dependence are intrinsically combined by the very construction of the multifractal model. To fully specify the MSM model we assume that the random multipliers follow a Lognormal[6] distribution with parameters λ and ν, i.e.,

$$M_t^i \sim LN(-\lambda, \nu). \tag{18}$$

Normalization of the distribution of the multipliers guarantees $\mathbb{E}\left[M_t^i\right] = 1$ which leads to

$$\exp\left(-\lambda + \frac{1}{2}\nu^2\right) = 1. \tag{19}$$

From *eq.* (19) it is obvious that the shape parameter ν can be expressed as: $\nu = \sqrt{2\lambda}$. With this restriction the Lognormal distribution of multipliers is fully defined by the scale parameter λ and the number of parameters to be estimated in the Lognormal MSM (LMSM) reduces to two, namely λ and σ. We estimate these for all specifications $k = 2, \ldots, 20$ using the GMM approach

[6] Other distributional assumptions such as Binomial, Gamma can be used as well, but have been found to make little difference in previous literature, cf. Liu et al. (2007), Lux (2008).

proposed by Lux (2008). We then choose the specification with the lowest GMM criterion as our preferred model for the subsequent forecasting exercise. Note that higher k increases the number of regimes (which is 2^k), and generates proximity to long memory over a longer number of lags, but comes at no additional computational cost in our approach. The pertinent moments used for the estimation can be found in Lux (2008). Note that maximum likelihood would be possible only for MSM models with a finite, discrete support of the multipliers, and computationally feasible only for a limited number of hierarchical components up to about 8.

Based on the LMSM model we adopt the standard approach for best linear forecasts outlined in Brockwell and Davis (1991) together with the generalized Levinson-Durbin algorithm proposed by Brockwell and Dahlhaus (2004) to conduct the recursive out-of-sample forecasting. This forecasting procedure can be summarized in two steps.

1. In the first step: We compute the following zero-mean time series

$$Z_t = x_t^2 - \mathbb{E}\left[x_t^2\right] = x_t^2 - \sigma^2, \tag{20}$$

   where $\hat{\sigma}$ is the estimate of the scale factor σ.

2. In the second step: Assuming that the CO_2 price volatility data follow the stationary process $\{Z_t\}$ defined in the first step, h-step best linear forecasts are given by

$$\hat{Z}_{n+h} = \sum_{i=1}^{n} \psi_{ni}^{(h)} Z_{n+1-i} = \Psi_n^{(h)} \mathbf{Z}_n, \tag{21}$$

   where the vectors of weights $\Psi_n^{(h)} = \left(\psi_{n1}^{(h)}, \psi_{n2}^{(h)}, \ldots, \psi_{nn}^{(h)}\right)'$ are solutions of

$$\Gamma \Psi_n^{(h)} = \gamma_n^h, \tag{22}$$

   with $\gamma_n^h = (\gamma(h), \gamma(h+1), \ldots, \gamma(n+h-1))'$ being the auto-covariances for the data generating process of Z_t at lags h and beyond, and $\Gamma_n = [\gamma(i-j)]_{i,j=1,\ldots,n}$ the pertinent variance-covariance matrix. The pertinent auto-covariances for the multifractal model are available in Lux (2008). Note that this approach provides only the best linear forecasts, but in contrast to the GARCH family the overall best nonlinear forecasts are not available to us as we estimate the MSM parameters via GMM.

4 Empirical Applications

4.1 Forecasting Methodology

To analyze the predictive ability of our models we adopt both the recursive and rolling forecasting schemes. We split our data set containing carbon price observations from January 16, 2009 to January 20, 2015 into two subsets. The first one covers the period from January 16, 2009 to November 16, 2012 and is used as in-sample data for model estimation. The second one contains carbon prices from November 19, 2012 to January 20, 2015 and serves as out-of-sample data that we use for forecast evaluation. Note that this division corresponds to the time before and after the second breakpoint indicated by the ICSS algorithm in Table 2. For the recursive scheme the estimation period is expanded forward by adding one observation day by day, so that the size of the data set used for the estimation increases over the out-of-sample period. This forecasting scheme should allow to more precisely estimate the parameters the more data become available. In the rolling scheme the estimation period is rolled forward by adding one observation and removing one day by day, so that the size of the estimation data sample remains fixed over the out-of-sample period. Forecasts are computed for horizons of various lengths: 1, 5, 10, 15, 20, 25, and 30 days. If the breakpoint identified by the ICSS would indicate a complete structural change of the dynamics, the forecasts computed for the out-of-sample period should at least initially not have much value beyond a naive forecast of future volatility. However, we do not observe much difference in the behavior of our forecasts so that the test might indeed rather have detected built-in regime changes like in the MS-GARCH or MSM models.

4.2 Forecast Evaluation Criteria

We evaluate the forecasting performance of our models via three loss functions, namely the mean squared error (MSE), the mean absolute error (MAE) and the value-at-risk (VaR)-based loss function proposed by González-Rivera et al. (2004) that represents the goodness-of-fit measure for the use of a volatility forecast in risk management, and the predictive ability test of Hansen (2005) to test the performance of each model under these criteria against all its competitors.

The three different loss functions are given by

$$\mathrm{MSE} = T^{-1} \sum_{i=1}^{T} \left(\sigma_{f,t}^2 - \sigma_{a,t}^2\right)^2, \tag{23}$$

$$\mathrm{MAE} = T^{-1} \sum_{i=1}^{T} \left|\sigma_{f,t}^2 - \sigma_{a,t}^2\right|, \tag{24}$$

$$\mathrm{MVaR}(\alpha) = T^{-1} \sum_{i=1}^{T} \left(\alpha - I_{t+1}^{\alpha}\right)\left(x_{t+1} - VaR_{t+1}^{\alpha}\right), \tag{25}$$

where $\sigma_{f,t}^2$ denotes the volatility forecast obtained using a GARCH-type model or MSM model, $\sigma_{a,t}^2$ is the daily actual volatility that is approximated by the daily squared returns, T denotes the number of out-of-sample observations, $I_{t+1}^{\alpha} = \mathbf{1}(x_{t+1} < VaR_{t+1}^{\alpha})$, $VaR_{t+1}^{\alpha} = F_{t+1}^{-1}(\alpha)\sigma_{f,t+1}$ is the conditional value-at-risk, and $F_{t+1}(.)$ is the forecast cumulative distribution of the standardized returns. As with MSE and MAE, a smaller value for the MVaR(α) hints at a good predictive performance. VaR forecasts provide an assessment of the loss that occurs in the α percent worst cases, and are used to determine the necessary level of underlying equity for risky assets to cover the risk of extreme market movement. The asymmetry of (25) covers the objective of avoiding type I-errors (failing to forecast a large negative change) to be more important than issuing an erroneous warning signal.

We note that the VaR-based loss function defined in *eq.* (25) is not differentiable due to the presence of the indicator function. The non-differentiability of MVaR(α) may cause a problem in the implementation of the superior predictive ability (SPA) test of Hansen (2005) that is based on the framework of White (2000). To obtain consistent results we also use a smooth approximation[7]

[7] As mentioned by Granger (1999) the issue associated with the non-differentiability may be just a technicality due to the fact that it should always be possible to find a smooth function which is arbitrarily close to the non-smooth one.

to MVaR(α), denoted as MSVaR(α), that is differentiable and given by

$$\text{MSVaR}(\alpha) = T^{-1}\sum_{i=1}^{T}\left[\alpha - g_\nu\left(x_{t+1}, VaR_{t+1}^{\alpha}\right)\right]\left(x_{t+1} - VaR_{t+1}^{\alpha}\right), \tag{26}$$

$g_\nu(y,z) = [1+\exp(\nu(y-z))]^{-1}$. The parameter, $\nu > 0$, governs the smoothness and for a higher value of ν MSVaR(α) gets closer to MVaR(α).

The three loss functions we use in this paper are well known in the literature and each of them can be used depending on the context and the objective of the users. To draw meaningful inferences about the relative forecasting performance of our volatility models, we apply the superior predictive ability (SPA) test of Hansen (2005). In the next Section we briefly describe the SPA test.

4.3 Superior Predictive Ability Test

The superior predictive ability (SPA) test proposed by Hansen (2005) allows to compare the relative performance of a particular model with its competitors via a pre-specified loss function. The test is a modification of the reality check test of White (2000). The null hypothesis that the benchmark model is not outperformed by any of the other competitive models is expressed as follows

$$H_0 : \max_{i=1,\ldots,K} \mathbb{E}\left[d_t\right] \leq 0, \tag{27}$$

where $d_t = (d_{i,t}, \ldots, d_{K,t})'$ is a vector of relative performances, $d_{i,t}$, that are computed as $d_{i,t} = L_{t,h}^{(0)} - L_{t,h}^{(i)}$. K is the number of the competitive models, h denotes the forecasting horizon and $L_{t,h}^{(0)}$ and $L_{t,h}^{(i)}$ are the loss functions at time t for a benchmark model M_0 and for its competitor models, $M_{i_{(i=1,\ldots,K)}}$, respectively.

The associated test statistic is given by

$$\text{SPA} = \max_{i=1,\ldots,K} \frac{\sqrt{T}\bar{d}_i}{\sqrt{\lim_{T\to\infty} \mathbb{V}ar(\sqrt{T}\bar{d}_i)}}, \tag{28}$$

where $\bar{d} = T^{-1}\sum d_t$. A stationary bootstrap procedure is used to obtain the p-values of the SPA. We refer the reader to Hansen (2005) for more details on technical issues. However, some remarks are in order on the framework of the SPA test concerning (i) the differentiability of the loss function used, (ii) the influence of the parameter estimation error, (iii) the impact of the relationships between models under comparison and (iv) the forecasting schemes used.

According to Theorem 2.3 in White (2000), the loss function under consideration must be differentiable, as in West (1996), Assumption 1, and either the same loss function has to be used for the estimation and prediction or $(T/R)\log\log R \rightarrow 0$ as $N \rightarrow \infty$ in order for the impact of parameter estimation to vanish.[8] As a result, the test cannot be immediately applied to non-smooth functions. However, as shown by White (2000) the asymptotic Normality of the

[8]Forecasts are to be produced for T periods, indexed from R through N (N is the total number of observations).

least absolute deviations estimator can be derived under conditions that allow to obtain results analogous to White's Theorem 2.3. Furthermore, it is important to note that non-differentiable loss functions can be used in the case they do not depend on parameter estimation. Sullivan and White (1999) used loss functions that depend on estimated parameters via an indicator function. The results of the stationary bootstrap reality check for their Monte Carlo experiments were in harmony with the desired limiting distribution and provide support to the conjecture that the test can be applied to non-differentiable functions as well.

Another restriction in the White's test framework is concerned with the nestedness of the models under comparison. In fact, the variance-covariance matrix of the limiting distribution must be positive semi-definite. This requirement is fulfilled if and only if at least one of the alternative models is nonnested with respect to the benchmark model. Our inclusion of the multifractal model guarantees that this condition is fulfilled as it is not nested in any GARCH-type model and vice versa.

We also note that the choice of the forecasting schemes has an impact on the limiting distribution. In fact, this has already been studied in detail in West and McCracken (1998) and McCracken (2000) who demonstrate how the parameter estimation uncertainty according to the choice of the forecasting schemes may have different impact on the variance-covariance of the limiting distribution. Hansen (2005) notes that a comparison of nested models that are estimated recursively may cause problems due to the violation of the assumption of stationarity. Hence, only rolling and fixed forecasting schemes are accommodated in his framework. In this paper we implement both recursive and rolling schemes because we want to know whether the violation of the stationarity condition may affect the SPA test results.

4.4 Estimating and Forecasting Results

The GARCH(1,1), the two-state Markov-switching GARCH(1,1) and the FIGARCH(1,d,1) are estimated via the maximum likelihood method and the results are reported in Table 3. The parameters β (that quantifies the effect of past volatility on current one) and α (that measures the effect of past squared innovations on current volatility) in the GARCH(1,1), MS-GARCH(1,1) and FIGARCH(1,d,1) models are well estimated and significant at the 1% level. The estimate of d in the FIGARCH(1,d,1) is also significant at the 1% level and indicates strong evidence of long memory in the conditional variance of CO_2 returns. The estimates of the transition probability of staying in regime 1 is 0.966, while that for regime 2 is 0.654. This indicates that a regime change from 2 to 1 happens more frequently than from 1 to 2, as can be seen in the time-varying probabilities of both states displayed in *Fig.* 4. Note that the major difference between the regimes is in the constants $\omega^{(1)}$ and $\omega^{(2)}$ while the coefficients for the volatility dynamics are relatively close to each other. Hence, the MS-GARCH model detects mainly differences between periods of generally high or low volatility while it implies almost the same structure of temporal dependence in both regimes. Finally, the estimates of the Lognormal parameter (λ) and the scale parameter in the MSM model are 1.171 and 3.328, respectively. While the second parameter just reflects the average rate of change, the parameter λ regulates the temporal dependency and is in line with typical findings for financial data.

Table 4 reports MSE, MAE, MVaR(5%) and MSVaR(5%) of volatility forecasts for our four volatility models used in this analysis and the superior predictive ability (SPA) test results. Note that MSE, MAE and MVaR(5%) for all four models are computed relative to the MSE, MAE and

MVaR(5%) of a constant forecast using historical volatility as estimated from the in-sample series and a value smaller than 1 would, thus, indicate that the model under consideration improves upon historical volatility under the respective criterion. Based on the MSE only the MSM model improves upon historical volatility at all forecast horizons. FIGARCH does so only at short horizons, and GARCH and MS-GARCH have average squared errors that are higher than those of historical volatility. The MSM is the best model and also improves on historical volatility at longer horizons (10 days and beyond) for the MSE criterion.

For MAE, all models have relative performance inferior to historical volatility, but MSM has the least deterioration. Under the MVaR(5%) criterion, and except for the MS-GARCH model at longer forecasting horizons (20 days and beyond), all volatility models dominate the historical volatility model with GARCH ahead of FIGARCH and MSM. Surprisingly weak is the forecasting performance of the MS-GARCH compared to the standard GARCH and FIGARCH models. The standard GARCH and FIGARCH models perform well and provide better average MVaR results than the MS-GARCH for longer horizons, 5 days and beyond. While the new MSM model has somewhat inferior statistics under the MVaR criterion than GARCH and FIGARCH, we will see below that the differences between GARCH, FIGARCH and MSM under the MVaR criterion appear insignificant for almost all time horizons under various backtesting procedures. As for the smooth MSVaR criterion, practically no differences are detected compared to the non-differentiable MVaR.

To see whether the recursive forecasting scheme affects the SPA results we also implement the rolling scheme that may also be more robust to structural changes in the data and is more appropriate to the test framework. The SPA results are reported in Table 5 (here we only display the probabilities of the SPA test statistics). The SPA test results are almost identical to those reported in Table 4 suggesting that the use of a recursive or rolling forecast scheme in this study does hardly affect any of our results.

For the SPA test we have used both the differentiable and non-differentiable VaR-based loss function. As can be seen in Table 6 the differences between both loss functions indeed are very small for all values of ν and completely disappear as the smoothness parameter ν increases. For $\nu = 5$ the difference is less than or equal to 0.001. However, to obtain consistent results for the SPA test we utilize the differentiable VaR-based loss function with ν set to 35. The results of the SPA test are practically identical with those obtained by using the non-differentiable loss function, cf. Table 4.

Overall, for MSE and MVaR, the performance differences are not that pronounced to indicate clear superiority of one of our models. Only MS-GARCH is clearly outperformed at usual levels of significance. However, for both MSE and MAE, MSM is the only one that cannot be outperformed at any forecast horizon (except for 1-day forecasts under MAE) at all traditional levels of confidence. Also, for MVaR and MSVaR it can only be outperformed at the 10-day horizon at a confidence level of 5%, and the 10 to 20 day horizons at confidence level of 10%. Across all criteria, it is, thus, the model with the highest rate of non-rejection of superior predictive ability test.

To access the performance of our volatility models in terms of forecasting value-at-risk at short and long horizons from another angle we additionally apply two backtesting procedures. The first one is the likelihood ratio test of Christoffersen (1998) that is based on the violations of the VaR forecasts and a Markov chain approach. This approach is well-known in the literature and the most used in practice. The idea of Christoffersen (1998) is that a model of VaR calculation is valid if (1) the expected frequency of observed VaR forecast violations is equal to the nominal coverage rate α, and (2) the VaR forecast violations are distributed independently. The second one is the duration-based backtesting approach proposed by Christoffersen and Pelletier (2004). Unlike the first approach the duration-based test is constructed by exploiting the statistical properties of the durations between consecutive violations. In fact, if VaR forecasts are valid, then, the duration between two consecutive violations must have a geometric distribution with a success probability that is equal to the coverage rate.

The test statistics and the p-values of both tests are reported in Tables 7 and 8 for the 1-day and 5-day horizons. The value of LR-UC is identical for all models. This is due to the fact that all our models lead to the same number of violations, even though they do not occur at the same times. The unconditional probability of violation is neither significantly higher nor smaller than 5%. This means that all our volatility models from this perspective perform equally well in estimating the level of market risk. We observe that the LR-IND and LR-CC tests of Christoffersen (1998) accept the null of independence and conditional coverage for almost all the volatility models used in this study except, in fact, for independence in the case of MSM at the 5-day horizons. These results have been confirmed for the standard GARCH, the MS-GARCH and the MSM but not for the FIGARCH model when applying the duration-based backtesting tests of Christoffersen and Pelletier (2004) to 1-period VaR forecasts. At the 5 days forecasting horizon all the models provide good performance. Surprisingly, we observe that the GARCH model also provides accurate VaR forecasts although there have been found structural breaks in the data. The structural similarity of all VaR forecasts can be contemplated in *Fig.* 5.

To gain insights as to whether forecasts from our competing volatility models may be combined to produce a new predictor that is more accurate than the individual forecasts we apply the forecast encompassing tests developed by Harvey et al. (1998) for non-nested models and its adjusted version proposed by Clark and West (2007) for nested ones. The results of both tests are reported in Table 9. In Table 9 the null hypothesis that forecasts from MSM do encompass those of the GARCH, FIGARCH and MS-GARCH models cannot be rejected at any standard confidence level for all three competing models with the only exception of FIGARCH at the 1-day horizon at a confidence level of 10 percent. This indicates that at the 1-day horizon the FIGARCH model should contribute useful information to the MSM forecasts while this is not obvious for the other models and forecast horizons.

The results of encompassing tests motivate us to explore the predictive capacity of combined forecasts from (FI)GARCH and MSM in a linear way in the hope of generating superior predictions (cf. Granger and Teräsvirta, 1999; Aiolfi and Timmermann, 2006). The new predictions $f_{n,t}$ are obtained by

$$f_{n,t} = (1 - \hat{\varsigma}) f_{1,t} + \hat{\varsigma} f_{2,t}, \tag{29}$$

where $f_{1,t}$ and $f_{2,t}$ are the single forecasts from the model 1 and model 2, respectively. $\hat{\varsigma}$ is the optimal weight of model 2 that is obtained from the following regression

$$\xi_{1,t} = \varsigma\,(\xi_{1,t} - \xi_{2,t}) + \epsilon_t, \tag{30}$$

where $\xi_{1,t}$ denotes the forecasting error from the model 1, that is MSM in this study and $\xi_{2,t}$ is that from (FI)GARCH, and ϵ_t an *iid* Normally distributed error term.

The SPA test results obtained by testing the new forecasts against all previous single models are presented in Table 4. The results show that for the MSE criterion, the null hypothesis that the combined forecast cannot be outperformed is never rejected at any confidence level. The same applies to the MAE criterion in almost all cases, but in both cases the forecast combinations with ex-post optimal weights do hardly improve upon the best single model, MSM. For the MVaR criterion, the linear combinations of both models provide average MVaRs that are in the most cases worse than those of their best ingredient, GARCH and MSM. However, the null hypothesis that the combined forecasts cannot be outperformed by the single models is again not rejected at the 5 % confidence level, except for the forecast combination from FIGARCH and MSM at the 10-day horizon. Note, however, that the optimal forecasts according to *eqs.* (29) and (30) have implicitly been calculated under an MSE criterion, so they need not be the optimal combinations for other criteria. Since we observe results for the MSE and MAE of the combined forecasts that are hardly different from the previous results for MSM as the single forecast model, this backtesting of optimal forecast combinations confirms the tendency of the test do not reject encompassing of the GARCH-type models by MSM. If MSM encompasses these alternative forecast, the forecast combination should not be superior which is indeed what we mostly observe.

5 Conclusion

This paper proposes a Markov-switching multifractal (MSM) model for modeling and forecasting carbon dioxide emission allowance spot price volatility. We have compared its forecasting ability with that of the standard GARCH, FIGARCH, and Markov-switching GARCH at shorter and longer forecasting horizons. The forecasting results indicate that the MSM model cannot be outperformed by its competitors across the vast majority of all performance criteria and forecast horizons. For MSE and MAE it provides the lowest average errors while for MVaR GARCH and FIGARCH have lower losses, but not to a significant extent. Various backtesting approaches for value-at-risk assessment also indicate that there are no significant differences between the candidate models in this respect. This result highlights the ability of the MSM processes to provide accurate volatility forecasts, especially at longer forecasting horizons. The MSM processes, therefore, seem to be at least as well suited as (FI)GARCH models for modeling of the spot price volatility of the European Union emission allowances market. Since GARCH and MSM are based on very different modeling concepts, they might capture different facets of the volatility process. Exploring the relationships between forecasts from different models, we find that MSM in the most cases encompasses GARCH and FIGARCH (except at 1-day horizon for FIGARCH). In line with this finding, combined forecasts using ex-post determined optimal weights do hardly improve upon forecasts based only on MSM.

References

Aiolfi, M. and A. Timmermann (2006). Persistence in forecasting performance and conditional combination strategies. *Journal of Econometrics 135*, 31–53.

Alberola, E., J. Chevallier, and B. Chèze (2008). Price drivers and structural breaks in European carbon prices 2005-2007. *Energy Policy 36*, 787–797.

Baillie, R. T., T. Bollerslev, and H. O. Mikkelsen (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 74*, 3–30.

Benschop, T. and B. L. Cabrera (2014). Volatility modeling of CO_2 emission allowance spot prices with regime-switching GARCH models. SFB 649 Discussion Paper 2014-050.

Benz, E. and S. Trück (2009). Modeling the price dynamics of CO_2 emission allowances. *Energy Economics 31*, 4–15.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*, 307–327.

Brockwell, P. and R. Dahlhaus (2004). Generalized Levinson-Durbin and Burg algorithms. *Journal of Econometrics 118*, 129–144.

Brockwell, P. and R. Davis (1991). *Time Series: Theory and Methods*. Berlin: Springer.

Calvet, L. and A. Fisher (2001). Forecasting multifractal volatility. *Journal of Econometrics 105*, 27–58.

Calvet, L. and A. Fisher (2004). Regime-switching and the estimation of multifractal processes. *Journal of Financial Econometrics 2*, 44–83.

Chevalier, J. (2009). Carbon futures and macroeconomic risk factors: A view from the eu ets. *Energy Economics 31*, 614–625.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review 39*, 841–862.

Christoffersen, P. F. and D. Pelletier (2004). Backtesting value-at-risk: A duration based approach. *Journal of Financial Econometrics 2*, 84–108.

Clark, T. E. and K. D. West (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics 138*, 291–311.

Daskalakis, G., D. Psychoyios, and R. N. Markellos (2009). Modeling CO_2 emission allowance prices and derivatives: Evidence from the european trading scheme. *Journal of Banking and Finance 33*, 1230–1241.

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

Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica 50*, 987–1007.

González-Rivera, G., T.-H. Lee, and S. Mishra (2004). Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. *International Journal of Forecasting 20*, 629–645.

Granger, C. W. (1999). Outline of forecast theory using generalized cost functions. *Spanish Economic Review 1*, 161–173.

Granger, C. W. and T. Teräsvirta (1999). A simple nonlinear time series model with missleading linear properties. *Economics Letters 62*, 161–165.

Hammoudeh, S., D. K. Nguyen, and R. M. Sousa (2014a). Energy prices and CO_2 emission allowance prices: A quantile regression approach. *Energy Policy 70*, 201–206.

Hammoudeh, S., D. K. Nguyen, and R. M. Sousa (2014b). What explains the short-term dynamics of the prices of CO_2 emissions? *Energy Economics 46*, 122–135.

Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business and Economic Statistics 23*, 365–380.

Harvey, D. I., S. J. Leybourne, and P. Newbold (1998). Tests for forecast encompassing. *Journal of Business and Economic Statistics 16*, 254–259.

Hill, B. M. (1975). A simple general approach to inference the tail of a distribution. *Annals of Statistics 3*, 1163–1174.

Klaassen, F. (2002). Improving GARCH volatility forecasts. *Empirical Economics 27*, 363–394.

Liu, R., T. di Matteo, and T. Lux (2007). True and apparent scaling: The proximity of the Markov-switching multifractal model to long-range dependence. *Physica A 383*, 35–42.

Lux, T. (2008). The Markov-switching multifractal model of asset returns: GMM estimation and linear forecasting of volatility. *Journal of Business and Economic Statistics 26*, 194–210.

Lux, T. and M. Ausloos (2002). Market fluctuations I: Scaling, multi-scaling and their possible origins. In A. Bunde, J. Kropp, and H.-J. Schellnhuber (Eds.), *Science of Disasters: Climate Disruptions, Heart Attacks and Market Crashes*, pp. 372–409. Springer, Berlin.

Lux, T. and L. Morales-Arias (2010). Forecasting volatility under fractality, regime-switching, long memory and Student-*t* innovations. *Computational Statistics and Data Analysis 54*, 2676–2692.

Lux, T., L. Morales-Arias, and C. Sattarhoff (2014). A Markov-switching multifractal approach to forecasting realized volatility. *Journal of Forecasting 33*, 532–541.

Lux, T., M. Segnon, and R. Gupta (2015). Modeling and forecasting crude oil price volatility: Evidence from historical and recent data. Working Paper, University of Kiel.

Mandelbrot, B. B. (1974). Intermittent turbulence in self similar cascades; divergence of high moments and dimension of the carrier. *Journal of Fluid Mechanics 62*, 331–358.

Mansanet-Bataller, M., A. Pardo, and E. Valor (2007). CO_2 prices, energy and weather. *The Energy Journal 28*, 73–92.

McCracken, M. W. (2000). Robust out-of-sample inference. *Journal of Econometrics 99*, 195–223.

Paolella, M. S. and L. Taschini (2008). An econometric analysis of emission allowance prices. *Journal of Banking and Finance 32*, 2022–2032.

Sansó, A., V. Arragó, and J. L. Carrion (2004). Testing for change in the unconditional variance of financial time series. *Revista de Economiá Financiera 4*, 32–53.

Sullivan, R. and H. White (1999). Finite sample properties of the bootstrap reality check for data-snooping: A monte carlo assessment. Technical report, QRDA, LLC Technical Report, San Diego.

Weron, R. (2002). Estimating long-range dependence: finite sample properties and confidence intervals. *Physica A: Statistical Mechanics and its Applications 312*, 285–299.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica 64*, 1067–1084.

West, K. D. and M. W. McCracken (1998). Regression-based comparison of some models of exchange rate volatility. *Journal of International Economics 35*, 23–45.

White, H. (2000). A reality check for data snooping. *Econometrica 68*, 1097–1126.

Yun, T. and R. D. Baker (2009). Analysis of a power plant investment opportunity under a carbon neutral world. *Investment Management and Financial Innovations 6*, 155–163.

Table 1: Descriptive statistics of the data

	Log-returns	Absolute returns	Squared returs
1567 observations (from January 16, 2009 to January 20, 2015)			
Minimum	-44.655	0	0
Maximum	21.060	44.655	1.994E+3
Mean	-0.035	2.134	11.111
Standard deviation	3.334	2.561	57.051
Skewness	-1.310	4.821	27.741
Kurtosis	27.299	57.837	938.163
Hurst index	0.468	0.903***	0.928***
Hill tail index at 5% tail	3.264 **[3.102 3.426]**		
Q(10)	63.992	636.600	50.968
Q(20)	84.744	1.180E+3	138.300
JB	3.900E+4		
ADF	-37.274		

Note: *** indicates 1% significance of Hurst coefficients based on the simulated boundary values of Weron (2002) for Wiener Brownian motion. For the tail index estimates, the brackets contain the 95% percent confidence intervals of the point estimate based upon the limiting distribution of the estimator.

Table 2: Structural breaks

Series	Break points	Period of time	Standard deviation
CO2 return	2	January 16, 2009 to March 24, 2009	4.844
		March 25, 2009 to November 16, 2011	2.019
		November 17, 2011 to January 20, 2015	4.057

Break points are identified by using the modified ICSS algorithm proposed by Sansó et al. (2004). The structural breaks happened on March 24, 2009 and November 16, 2011. The estimated break dates are indicated by the arrows in *Fig.* 1.

Table 3: Parameter estimates using the complete data

Parameters	GARCH	MS-GARCH		FIGARCH	MSM
Regimes		1	2		
ω	0.002*** (0.000)	0.032** (0.015)	0.170 (0.219)	0.003 (0.002)	
α	0.116*** (0.018)	0.046*** (0.009)	0.161*** (0.083)		
β	0.864*** (0.020)	0.930*** (0.013)	0.961*** (0.016)	0.318** (0.153)	
p_{ii}		0.966*** (0.013)	0.654*** (0.152)		
ϕ				0.030 (0.115)	
d				0.395*** (0.075)	
λ					1.171***
σ					3.328***
Diagnostic					
Log(L)	-3.764E+3	-3.675E+3		-3.757E+3	-
AIC	7.535E+3	7.365E+3		7.522E+3	-
BIC	7.551E+3	7.408E+3		7.543E+3	-

Note: The numbers in parentheses are standard errors of the estimations. Log(L) is the logarithm of the maximum likelihood function. AIC and BIC are the Akaike and Bayesian information criterion, respectively. *** indicate that the parameters are significant at the 1% level. The optimum objective function in the GMM procedures is obtained for $k = 20$ (volatility components in the MSM model).

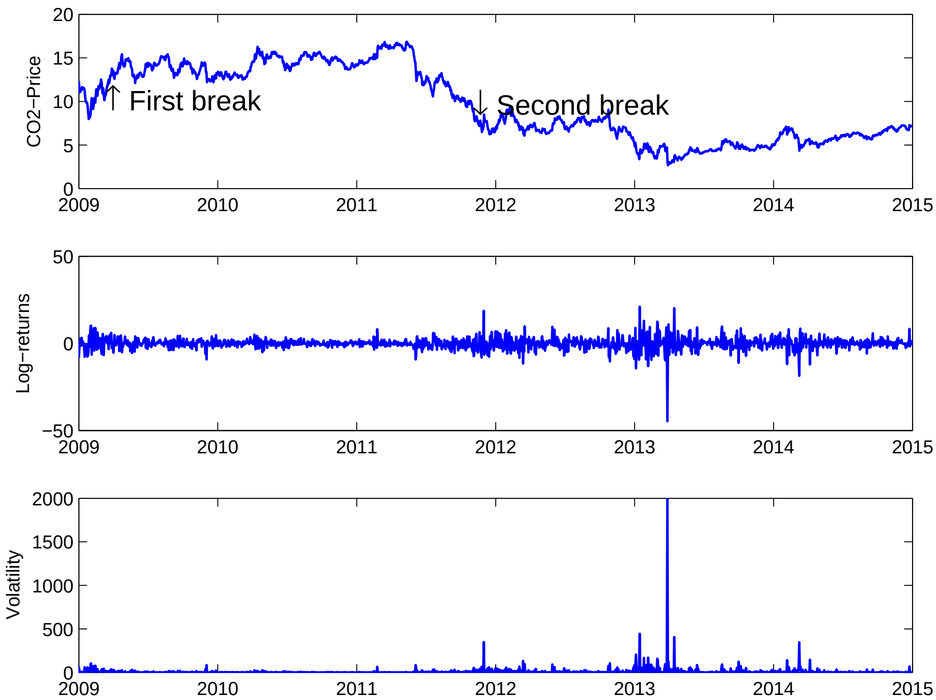

Figure 1: Plot of prices, returns and volatility.

Table 4: Recursive forecast scheme: MSE, MAE, MVaR(5%) and SPA test results using carbon dioxide prices from January 16, 2009 to November 16, 2012 as in-sample period and carbon prices from November 19, 2012 to January 20, 2015 as out-of-sample period.

	Forecast horizons						
	1	5	10	15	20	25	30
				MSE			
GARCH	1.021	1.022	1.010	1.027	1.047	1.049	1.029
	(0.063)	(0.097)	**(0.150)**	(0.075)	(0.045)	(0.050)	**(0.184)**
MS-GARCH	1.007	1.025	1.036	1.106	1.293	2.168	12.605
	(0.088)	(0.046)	(0.010)	(0.000)	(0.000)	(0.025)	(0.074)
FIGARCH	0.999	1.035	0.995	0.995	1.011	1.005	1.001
	(0.305)	(0.092)	**(0.352)**	**(0.331)**	**(0.104)**	**(0.123)**	**(0.571)**
MSM	0.985	0.987	0.982	0.987	0.992	0.991	0.991
	(0.827)	**(1.000)**	**(0.884)**	**(0.669)**	**(1.000)**	**(1.000)**	**(0.960)**
GARCH+MSM	0.984	0.987	0.982	0.987	0.992	0.991	0.990
	(0.960)	**(0.551)**	**(0.988)**	**(0.959)**	**(0.608)**	**(0.562)**	**(0.998)**
FIGARCH+MSM	0.979	0.983	0.982	0.986	0.991	0.991	0.991
	(0.816)	**(0.788)**	**(0.980)**	**(0.993)**	**(0.664)**	**(0.534)**	**(0.994)**
				MAE			
GARCH	1.166	1.209	1.198	1.240	1.259	1.295	1.284
	(0.029)	(0.026)	(0.039)	(0.034)	(0.024)	(0.011)	(0.015)
MS-GARCH	1.160	1.291	1.447	1.792	2.381	3.472	5.542
	(0.005)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
FIGARCH	0.993	1.178	1.138	1.170	1.180	1.179	1.182
	(1.000)	(0.083)	**(0.149)**	(0.082)	(0.046)	(0.044)	(0.041)
MSM	1.077	1.112	1.110	1.122	1.120	1.125	1.129
	(0.011)	**(1.000)**	**(0.851)**	**(1.000)**	**(1.000)**	**(1.000)**	**(1.000)**
GARCH+MSM	1.069	1.110	1.116	1.126	1.114	1.119	1.144
	(0.012)	**(0.839)**	**(0.294)**	**(0.110)**	**(1.000)**	**(1.000)**	(0.062)
FIGARCH+MSM	1.013	1.101	1.110	1.132	1.108	1.123	1.131
	(0.153)	**(0.795)**	**(0.810)**	**(0.104)**	**(1.000)**	**(1.000)**	(0.057)
				MVaR(5%)			
GARCH	0.926	0.924	0.910	0.914	0.922	0.965	0.946
	(0.908)	**(0.941)**	**(0.639)**	**(0.636)**	**(1.000)**	**(0.673)**	**(0.851)**
MS-GARCH	0.936	0.948	0.952	0.997	1.062	1.162	1.274
	(0.433)	(0.094)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
FIGARCH	0.980	0.940	0.914	0.921	0.959	0.969	0.964
	(0.089)	**(0.277)**	**(0.361)**	**(0.364)**	**(0.102)**	**(0.389)**	**(0.149)**
MSM	0.939	0.944	0.943	0.954	0.965	0.982	0.981
	(0.315)	**(0.241)**	(0.012)	(0.059)	(0.081)	**(0.150)**	**(0.106)**
GARCH+MSM	0.949	0.947	0.933	0.951	0.973	0.986	0.975
	(0.063)	**(0.107)**	(0.052)	(0.064)	(0.058)	**(0.160)**	**(0.176)**
FIGARCH+MSM	0.944	0.953	0.938	0.946	0.971	0.983	0.981
	(0.348)	**(0.191)**	(0.020)	(0.056)	(0.098)	**(0.186)**	**(0.179)**
				MSVaR(5%) with $\nu = 35$			
GARCH	0.926	0.924	0.910	0.914	0.922	0.965	0.946
	(0.909)	**(0.942)**	**(0.640)**	**(0.638)**	**(1.000)**	**(0.673)**	**(0.851)**
MS-GARCH	0.936	0.948	0.952	0.997	1.062	1.162	1.274
	(0.431)	(0.094)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
FIGARCH	0.980	0.940	0.914	0.921	0.959	0.969	0.964
	(0.089)	**(0.275)**	**(0.360)**	**(0.362)**	**(0.102)**	**(0.390)**	**(0.149)**
MSM	0.939	0.944	0.944	0.954	0.965	0.982	0.981
	(0.313)	**(0.239)**	(0.011)	(0.059)	(0.081)	**(0.150)**	**(0.106)**

Note: MSE, MAE, and MVaR(5%) and MSVaR(5%) for all four models are computed relative to the MSE, MAE, MVaR(5%) and MSVaR(5%) of a constant forecast using historical volatility as estimated from the in-sample series. The entries in parentheses are the p-values of the SPA test of Hansen (2005) for the pertinent model and criterion. The null hypothesis is that a benchmark model cannot be outperformed by other candidate models. The values in bold face represent the p-values that are greater than or equal to the 10% confidence level under a pre-specified loss function. The combined forecast are computed using the optimal ex-post weights defined in *eqs.* (29) and (30).

Table 5: Rolling forecast scheme: SPA test results using carbon dioxide prices from January 16, 2009 to November 16, 2012 as in-sample period and carbon prices from November 19, 2012 to January 20, 2015 as out-of-sample period.

	Forecast horizons						
	1	5	10	15	20	25	30
				MSE			
GARCH	0.053	0.078	**0.135**	0.094	0.035	0.047	**0.153**
MS-GARCH	0.093	0.016	0.000	0.002	0.010	0.020	0.036
FIGARCH	**0.294**	0.089	**0.345**	**0.248**	0.055	0.078	**0.295**
MSM	**0.840**	**1.000**	**0.887**	**0.752**	**1.000**	**1.000**	**0.890**
				MAE			
GARCH	0.033	0.021	0.028	0.024	0.014	0.004	0.008
MS-GARCH	0.011	0.000	0.000	0.000	0.000	0.000	0.002
FIGARCH	**1.000**	0.081	**0.112**	0.051	0.020	0.011	0.009
MSM	0.066	**1.000**	**0.888**	**1.000**	**1.000**	**1.000**	**1.000**
				MVaR(5%)			
GARCH	**0.734**	**0.841**	**0.747**	**0.736**	**0.814**	**0.677**	**0.808**
MS-GARCH	**0.445**	0.000	0.000	0.000	0.000	0.000	0.000
FIGARCH	0.022	**0.170**	**0.338**	**0.365**	**0.128**	**0.686**	**0.327**
MSM	**0.584**	**0.393**	0.098	0.085	**0.303**	**0.387**	**0.329**

Note: The entries are the p-values of the SPA test of Hansen (2005) for the pertinent model and criteria. The null hypothesis is that a benchmark model cannot be outperformed by other candidate models. The values in bold face represent the p-values that are greater than or equal to the 10% confidence level under a pre-specified loss function.

Table 6: MSVaR(5%) using a smooth function

	Recursive forecasting scheme						
	Forecast horizons						
	1	5	10	15	20	25	30
				Smoothing coefficient $\nu = 5$			
GARCH	0.926	0.923	0.908	0.914	0.921	0.964	0.946
MS-GARCH	0.936	0.947	0.951	0.997	1.063	1.163	1.276
FIGARCH	0.980	0.939	0.913	0.920	0.960	0.969	0.964
MSM	0.940	0.945	0.943	0.955	0.966	0.982	0.982
				Smoothing coefficient $\nu = 10$			
GARCH	0.927	0.924	0.909	0.914	0.922	0.965	0.946
MS-GARCH	0.936	0.948	0.951	0.997	1.062	1.162	1.274
FIGARCH	0.980	0.940	0.914	0.921	0.960	0.969	0.964
MSM	0.939	0.945	0.943	0.954	0.965	0.982	0.982
				Smoothing coefficient $\nu = 35$			
GARCH	0.926	0.924	0.910	0.914	0.922	0.965	0.946
MS-GARCH	0.936	0.948	0.952	0.997	1.062	1.162	1.274
FIGARCH	0.980	0.940	0.914	0.921	0.959	0.969	0.964
MSM	0.939	0.944	0.944	0.954	0.965	0.982	0.981
				Smoothing coefficient $\nu = 95$			
GARCH	0.926	0.924	0.910	0.914	0.922	0.965	0.946
MS-GARCH	0.936	0.948	0.952	0.997	1.062	1.16	1.274
FIGARCH	0.980	0.940	0.914	0.921	0.959	0.969	0.964
MSM	0.939	0.944	0.944	0.954	0.965	0.982	0.981

Note: ν is the smooth parameter. For $\nu \geq 10$ the differences between the non-differentiable and differentiable functions are vanishing.

Table 7: Backtesting tests of 5% VaR forecasts for CO_2 returns at the forecasting horizon, $h = 1$.

	GARCH	MS-GARCH	FIGARCH	MSM
EFV	0.048	0.048	0.048	0.048
LR tests of Christoffersen (1998)				
LR-UC	0.029	0.029	0.029	0.029
	(0.866)	**(0.866)**	**(0.866)**	**(0.866)**
LR-IND	0.416	0.416	1.985	0.416
	(0.519)	**(0.519)**	**(0.159)**	**(0.519)**
LR-CC	0.444	0.444	2.014	0.444
	(0.801)	**(0.801)**	**(0.365)**	**(0.801)**
LR tests of Christoffersen and Pelletier (2004)				
LR-CC	1.670	0.131	10.461	1.462
	(0.434)	**(0.937)**	(0.005)	**(0.482)**
LR-IND	1.540	<0.001	10.330	1.331
	(0.215)	**(0.993)**	(0.001)	**(0.249)**

Note: EFV denotes the ratio of VaR violations to the sample size (T=538) observed for the CO_2 returns. LR-UC, LR-IND and LR-CC denote the unconditional coverage, independence and conditional coverage test statistics, respectively. While the LR test in Christoffersen (1998) is based on the violation process and a Markov chain approach, the test developed by Christoffersen and Pelletier (2004) is constructed based on the durations between violations. The values in bold face are the p-values that are greater than or equal to the 5% confidence level.

Table 8: Backtesting tests of 5% VaR forecasts for CO_2 returns at the forecasting horizon, $h = 5$.

	GARCH	MS-GARCH	FIGARCH	MSM
EFV	0.048	0.048	0.048	0.048
LR tests of Christoffersen (1998)				
LR-UC	0.029	0.029	0.029	0.029
	(0.866)	**(0.866)**	**(0.866)**	**(0.866)**
LR-IND	1.985	1.985	0.416	4.413
	(0.159)	**(0.159)**	**(0.519)**	(0.036)
LR-CC	2.014	2.014	0.444	4.442
	(0.365)	**(0.365)**	**(0.801)**	**(0.109)**
LR tests of Christoffersen and Pelletier (2004)				
LR-CC	0.205	1.704	0.331	3.492
	(0.903)	**(0.427)**	**(0.848)**	**(0.175)**
LR-IND	0.075	1.573	0.200	3.361
	(0.785)	**(0.210)**	**(0.655)**	**(0.067)**

Note: EFV denotes the ratio of VaR violations to the sample size (T=538) observed for the CO_2 returns. LR-UC, LR-IND and LR-CC denote the unconditional coverage, independence and conditional coverage test statistics, respectively. While the LR test in Christoffersen (1998) is based on the violation process and a Markov chain approach, the test developed by Christoffersen and Pelletier (2004) is constructed based on the durations between violations. The values in bold face are the p-values that are greater than or equal to the 5% confidence level.

Table 9: Encompassing tests for non-nested models

	Model 1 vs. Model 2						
	MSM vs. GARCH						
ENC-T	-0.880	-0.236	0.777	0.109	-0.223	-0.194	0.325
	(0.810)	**(0.593)**	**(0.219)**	**(0.457)**	**(0.588)**	**(0.577)**	**(0.373)**
$\hat{\varsigma}$	-0.203	-0.059	0.149	0.045	-0.071	-0.065	0.145
	[0.269]	[0.242]	[0.217]	[0.204]	[0.196]	[0.189]	[0.186]
	MSM vs. MS-GARCH						
ENC-T	-1.172	-0.975	-0.795	-0.710	-0.985	-1.252	-1.450
	(0.879)	**(0.835)**	**(0.787)**	**(0.761)**	**(0.837)**	**(0.894)**	**(0.926)**
$\hat{\varsigma}$	-0.496	-0.503	-0.217	-0.180	-0.128	-0.054	-0.009
	[0.405]	[0.313]	[0.220]	[0.144]	[0.088]	[0.042]	[0.012]
	MSM vs. FIGARCH						
ENC-T	1.323	-0.989	1.181	0.338	-0.409	-0.153	0.045
	(0.093)	**(0.838)**	**(0.119)**	**(0.368)**	**(0.659)**	**(0.561)**	**(0.482)**
$\hat{\varsigma}$	0.366	-0.360	0.160	0.222	-0.258	-0.084	0.036
	[0.195]	[0.256]	[0.308]	[0.347]	[0.378]	[0.403]	[0.425]

Note: we test the null hypothesis that forecasts from model 1 encompass those of model 2 and ENC-T denotes the associated test statistics. The values in parentheses are the p-values of the tests. $\hat{\varsigma}$s are the estimates of the slope parameter ς in the forecast encompassing regression *eq.* (30). The values in square brackets are the standard errors of the estimation.

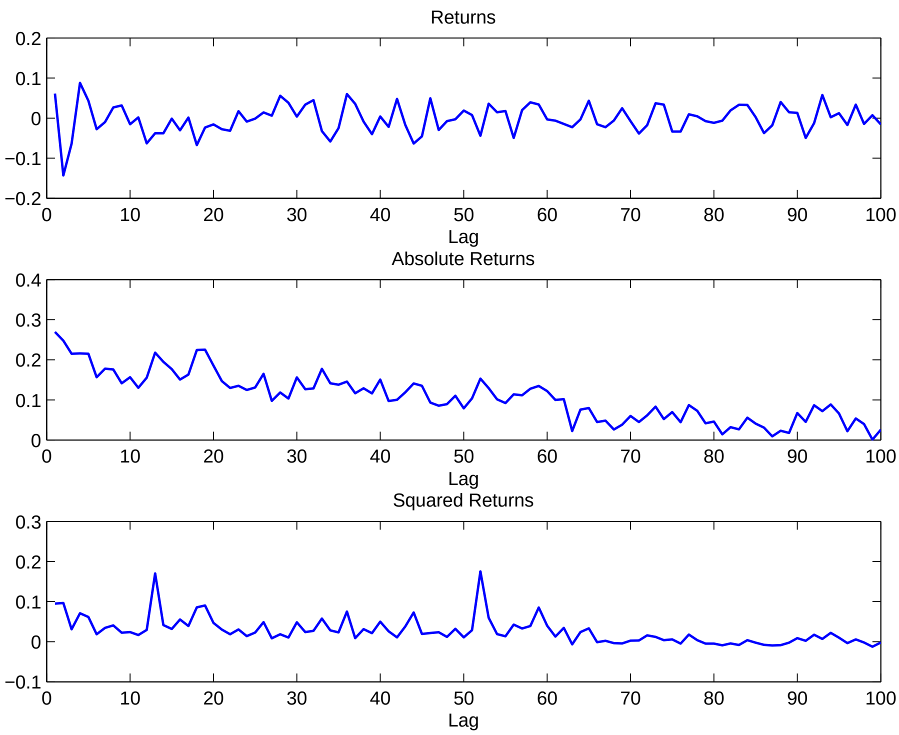

Figure 2: Plot of autocorrelation functions

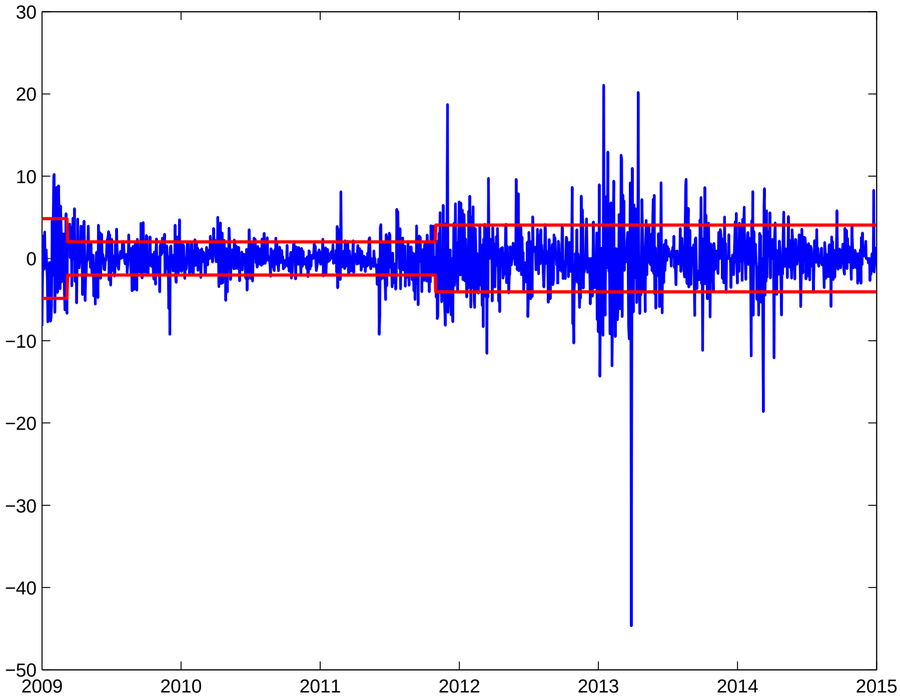

Figure 3: CO_2 returns and three-standard-deviation bands for the regimes defined by the structural breaks identified by the modified ICSS algorithm.

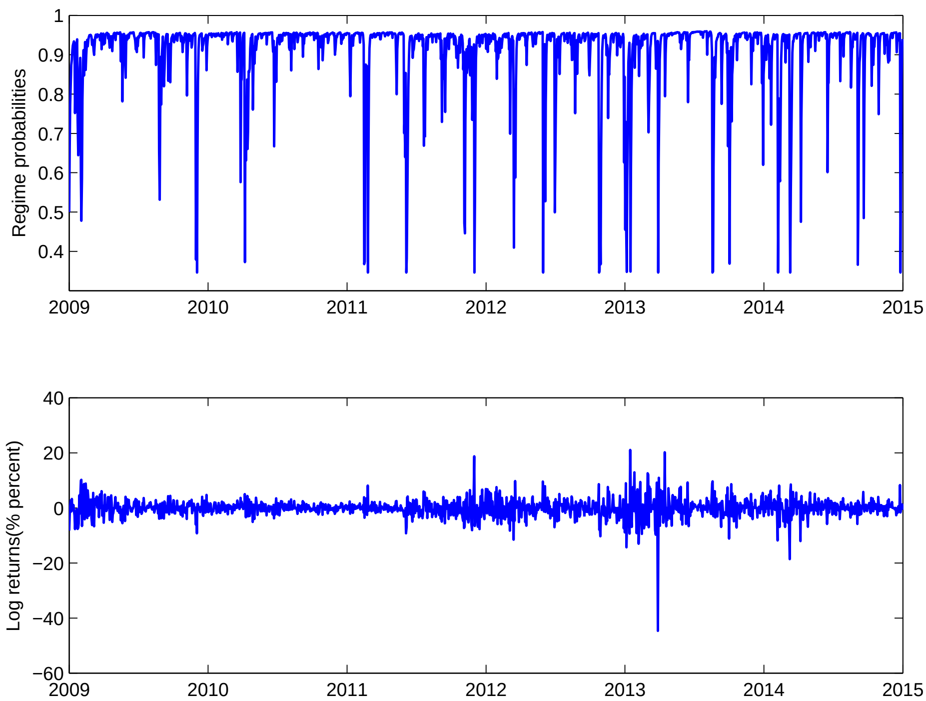

Figure 4: Estimated probabilities of MS-GARCH(1,1) of being in regime 1 (upper panel) and log returns (lower panel)

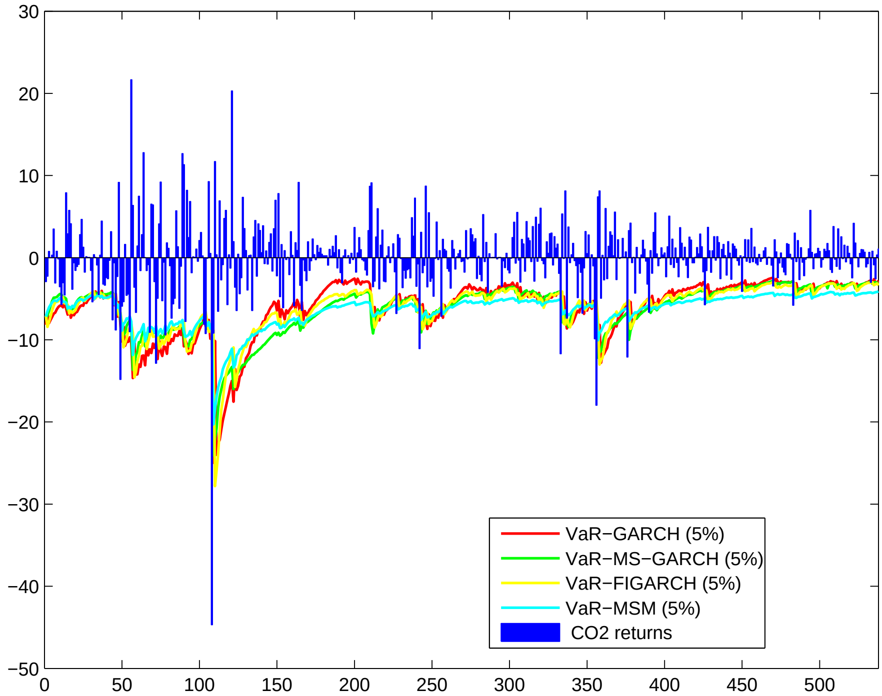

Figure 5: CO_2 returns with 5% VaR forecasts from GARCH, MS-GARCH, FIGARCH and MSM at the forecasting horizon, h=5