



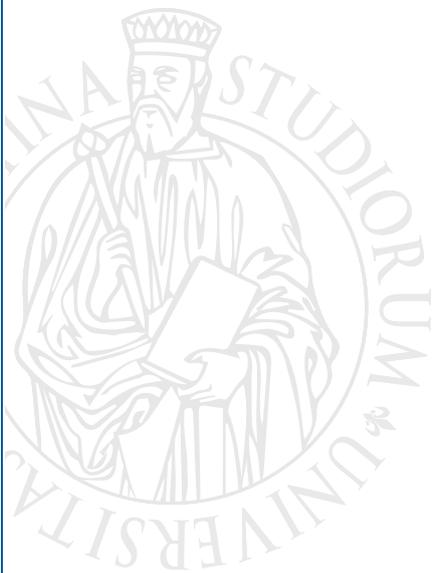
UNIVERSITÀ
DEGLI STUDI
FIRENZE

DISIA

DIPARTIMENTO DI STATISTICA,
INFORMATICA, APPLICAZIONI
"GIUSEPPE PARENTI"

Forecasting Realized Volatility with Changes of Regimes

Giampiero M. Gallo, Edoardo Otranto



**DISIA WORKING PAPER
2014/03**

© Copyright is held by the author(s).

Forecasting Realized Volatility with Changes of Regimes*

Giampiero M. Gallo[†]

Dipartimento di Statistica, Informatica, Applicazioni “G. Parenti”, Università di Firenze
Viale G. B. Morgagni, 59 - 50134 Firenze, Italy (*gallog@disia.unifi.it*)

Edoardo Otranto

Dipartimento di Scienze Cognitive, della Formazione e degli Studi Culturali and CRENoS,
Università di Messina, Via Concezione, 6 - 98121 Messina, Italy (*eotranto@unime.it*)

Abstract

Realized volatility of financial time series generally shows a slow-moving average level from the early 2000s to recent times, with alternating periods of turmoil and quiet. Modeling such a pattern has been variously tackled in the literature with solutions spanning from long-memory, Markov switching and spline interpolation. In this paper, we explore the extension of Multiplicative Error Models to include a Markovian dynamics (MS-MEM). Such a model is able to capture some sudden changes in volatility following an abrupt crisis and to accommodate different dynamic responses within each regime. The model is applied to the realized volatility of the S&P500 index: next to an interesting interpretation of the regimes in terms of market events, the MS-MEM has better in-sample fitting capability and achieves good out-of-sample forecasting performances relative to alternative specifications.

Keywords: MEM, regime switching, realized volatility, volatility persistence, volatility forecasting.

JEL codes: C22, C24, C58

*A previous version of the paper was circulated under the title Realized Volatility and Change of Regimes. Without implicating, we are grateful to conference and seminar participants (ERCIM 2010, London; SCO 2011, Padua; CREATES, EUI, U. of Salerno 2013).

[†]corresponding author

1 Introduction

A consolidated literature in financial econometrics is devoted to measuring asset volatility, exploiting the information contained in asset price data collected at a very high frequency. The volatility estimators, known as realized volatility measures, have allowed for deeper insights into the dynamics of volatility, traditionally analyzed in a modeling and forecasting framework within the GARCH paradigm as conditional variances of returns (Engle, 1982; Bollerslev, 1986; and further extensions – for a review, see Teräsvirta, 2009; for nonlinear models, see Teräsvirta, 2011). Starting from the plain realized volatility, largely studied in Andersen et al. (2000, 2003), other measures have been introduced to take into consideration the presence of jumps and other market microstructure issues (for a review, see Andersen et al., 2010). The most recent addition in the family of volatility estimators is the realized kernel volatility (developed by Barndorff-Nielsen et al., 2008), designed to possess robustness to market microstructure noise.

While volatility measurement from an end-of-day perspective has reached a mature stage, the question of volatility forecasting is still open, since there is still a wide debate as of which model is appropriate to characterize volatility dynamics, based on the current information available (cf. among others, Brownlees and Gallo, 2010; Shephard and Sheppard, 2010; Hansen et al., 2012; Cipollini et al., 2012). The modeling effort which we adopt in this paper makes two strong choices: the first is to mainly concentrate on Multiplicative Error Models – MEM – developed by Engle (2002), Engle and Gallo (2006); the second is to take observed oscillations around time-varying averages of kernel realized volatility as consistent with a regime-switching representation (Hamilton, 1989, 1990), rather than expression of long memory (cf. the discussion in the next section below). Overall, we keep model residual diagnostics and forecasting performance as a guideline for gauging a suitable specification next to the interpretability of the results.

In a MEM, the volatility series is modeled as the product of a time-varying scale factor evolving autoregressively and a random disturbance with a proper distribution. As it is applied to non-negative values, a MEM captures dynamics without resorting to logs (as in other models), thus producing forecasts of volatility (and not of log-volatility). We favor modeling volatility levels with multiplicative (rather than additive) disturbances, because of the Quasi Maximum Likelihood interpretation shown in Engle and Gallo (2006) which ensures consistency of MEM parameter estimators. For series such as the one at hand, the original MEM with an asymmetric response of volatility to the sign of lagged returns still suffers from residual autocorrelation. The consideration of time dependent regimes is not new to the volatility literature (see below) and our extended class of Markov Switching (MS) MEMs captures the presence of regimes with possibly different dynamics. Our approach is different than the mixture MEM (De Luca and Gallo, 2004 and Lanne, 2006, which suffers from the shortcoming of time independent regime indicator variables, cf. Bauwens et al., 2012; and De Luca and Gallo, 2009, who make the mixing weights dependent on a lagged observable variable) and the P-Spline MEM

used by Brownlees and Gallo (2010) for realized volatility with a time varying average level (which is a statistical fit to a supposed smooth underlying trend).

After deriving the theoretical properties of the model, we use the S&P500 series as a guideline for the discussion, with attention devoted to the issue of how many regimes characterize the data, and how to handle the episode of the short-lived burst of volatility in October 2008. We show that three regimes are appropriate for the evolution of the volatility in financial markets in the period taken into consideration (spanning from 2000 to 2012). A fourth regime (which would in principle accommodate the turmoil after the fall of Lehman Brothers) does not receive empirical support from a nonparametric Bayesian test for the number of regimes (Otranto and Gallo, 2002) and actually makes the attribution of periods to regimes less sharp. We find that in order to make the general inference less sensitive to the 2008 outlier, the insertion of a dummy variable to change the intercept of the model is a more successful strategy. Among several specifications, evidence supports a parsimonious common dynamics across some regimes. Moreover the MS specification seems to fit and interpret the changes in volatility better than alternative parameterizations, such as the Smooth Transition model (à la Teräsvirta, 1994) and the Heterogeneous Autoregressive (HAR) model (à la Corsi, 2009).

The illustration is carried over to the forecast comparison, both in- and out-of-sample: a better prediction fit than the base Asymmetric MEM is present as one would expect from the richer parameterization of the MS models; but there is also a good performance in one- and two-steps ahead forecasts in the out-of-sample period with the Diebold-Mariano (1995) test and the Model Confidence Set sequential tests (Hansen *et al.*, 2011) using absolute and squared losses. In view of the results obtained by Hansen (2010) on the theoretical inverse relationship between in- and out-of-sample performance, we interpret our results (better in-sample fit and equivalent, if not slightly better performance out-of-sample) as a support for our Markov Switching MEMs.

The paper is organized as follows: in Section 2 we summarize some relevant literature to modeling realized volatility. In Section 3 we introduce the Markov Switching extensions within the MEM framework and in Section 4 we briefly describe the alternative parameterizations used. After discussing the data of the S&P500 volatility, Section 4 is articulated along a discussion of the estimation results, the inference on the regimes, the regime-specific average volatility, and a comparison of the models based on confidence intervals and the forecast performance. Concluding remarks follow.

2 High Persistence in Volatility

There is a large debate in the literature about the nature of the high persistence in realized volatility and whether it may be the result of some nonlinearity in the process. The HAR model (Corsi, 2009) although formally not a long-memory model, can reproduce the observed hyperbolic-type decay of the autocorrelation function of log-volatility by specifying a sum of

volatility components over different horizons. The intuition by Baillie and Kapetanios (2007) about the existence of both non linear and long memory components in many economics and financial time series is developed by McAleer and Medeiros (2008) who introduce a multiple regime smooth transition extension of the HAR: their model is also able to capture the presence of sign and size asymmetries. Bordignon and Raggi (2012) propose an elegant solution to combine in the same model the non linearity effects, through a Markov switching process, and high persistence, through a fractionally integrated dynamics, capable of improving the accuracy of in-sample and out-of-sample forecasts. Alternatively, concentrating on long memory explanations, Andersen et al. (2003) suggest a fractionally integrated autoregressive moving average (ARFIMA) model for log-realized volatility; in exchange rate dynamics, Ohanissian et al. (2008) find evidence of long memory as well. However, as noted by Lanne (2006), the ARFIMA model may not be optimal for several reasons: a) a simple short-memory ARMA model can be as good in forecasting the realized volatility of stock returns as a long-memory ARFIMA model (Pong et al., 2004); b) the parameters of the FI part and ARMA part can capture similar characteristics (Bos et al., 2002); c) the ARFIMA model is an approximation to the “true” model anyway, as it involves an infinite-order lag polynomial that must be truncated in practical applications. Corsi et al. (2008) underline how the empirical distributions of ARFIMA and HAR residuals, derived from realized volatility series, tend to exhibit yet unmodeled volatility clustering. In this respect, the presence of regimes (mixture distribution) is able to capture the slowly decaying autocorrelation function of the observed realized volatility series. Maheu and McCurdy (2002) find strong statistical evidence of regime changes in both the conditional mean and conditional variance of realized volatility,¹ using a Markov Switching ARMAX representation, where the transition probabilities and the conditional mean of volatility are both functions of the duration of the state. Scharth and Medeiros (2009) extend a regression tree model to accommodate smooth splits in regimes controlled by significant past cumulated returns which account for long-range dependence in volatility. Other models using realized volatility are Maheu and McCurdy (2007, 2011).

A comparison across this rich set of models (volatility vs. log-volatility; long memory, smooth transition, spline approach and Markov Switching) is beyond the scope of this paper, also recalling the discussion by Diebold and Inoue (2001), underlining the observational equivalence between long memory and nonlinear effects in volatility. It may well be that for specific series and sample periods, one model may significantly outperform others (a common occur-

¹In a GARCH framework, previous contributions had addressed regime switching, cf. the SWARCH model (Hamilton and Susmel, 1994), the MS GARCH model (Dueker, 1997, Klaassen, 2002), and the recent multivariate extensions (Edwards and Susmel, 2003, Higgs and Worthington, 2004, Gallo and Otranto, 2007, 2008). An alternative way to consider changes in regime is given by smooth transition GARCH models (Teräsvirta, 2009) or other nonlinear models (Teräsvirta, 2011). Several other authors indicate the presence of level shifts in GARCH (Perron and Qu, 2010) or breaks in unknown points also in GARCH (He and Maheu, 2010) as the cause of an apparent high persistence. The issue of time varying underlying level of volatility is addressed also by Engle and Rangel (2008), who adapt a spline function in GARCH to capture a low frequency component of volatility (which they connect to macroeconomic factors).

rence for nonlinear models). What we favor in the sequel is the comparability within the sole class of MEMs with regimes (considering abrupt and smooth changes) or without, concentrating on the economic interpretability of the regime identification and their capability to capture certain episodes characterizing the evolution of volatility, especially at times of distress.

3 The Asymmetric MEM with Markov Switching

The basic MEM idea is introduced in Engle (2002) and successively developed in Engle and Gallo (2006). Cipollini et al. (2012) suggest an extension to the multivariate case. The volatility x_t of a certain financial time series is modeled as the product of a time varying scale factor μ_t , representing the conditional mean of x_t , which follows a GARCH-type dynamics, and a positive valued error ε_t :

$$\begin{aligned}
 x_t &= \mu_t \varepsilon_t, & \varepsilon_t &\sim \text{Gamma}(a, 1/a) \text{ for each } t \\
 \mu_t &= \omega + \alpha x_{t-1} + \beta \mu_{t-1} + \gamma D_{t-1} x_{t-1} \\
 D_t &= \begin{cases} 1 & \text{if } r_t < 0 \\ 0 & \text{if } r_t \geq 0 \end{cases}
 \end{aligned} \tag{3.1}$$

This base specification takes into account the presence of asymmetric responses of volatility to the sign of the previous returns (Engle and Gallo, 2006), similarly to the GJR–GARCH model (Glosten *et al.*, 1993) and the coefficient γ captures a stronger reaction in correspondence to past negative returns. We call this model *Asymmetric MEM* (AMEM); setting γ to zero gives us the simple MEM. Constraints can be imposed to ensure the positiveness of μ_t (e.g., $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$, $\gamma \geq 0$) and the stationarity of the process (persistence $(\alpha + \beta + \gamma/2)$ less than 1 under zero median of the returns). The Gamma distribution of the innovation depends only on a single parameter a , providing a unit mean and a variance equal to $1/a$. Larger values of a correspond to a symmetric density similar to the Normal distribution; Engle and Gallo, 2006, provide a graphical summary of various cases), whereas smaller values provide asymmetry with a larger variance.² Correspondingly, conditional on the information Ψ_{t-1} available at time $t - 1$, the mean and variance of x_t are μ_t and μ_t^2/a respectively. Further lags can be added to the specification of μ_t .

For the specification provided in equation (3.1), the unconditional mean of the volatility across the entire period is given by:

$$\mu = \frac{\omega}{1 - \alpha - \beta - \gamma/2} \tag{3.2}$$

²Of course less restrictive (but less parsimonious) specifications could be adopted, such as the Generalized Gamma.

As noted in the introduction, this feature may be too restrictive for series such as our leading example, especially if we want to allow for sudden and persistent changes in the level of the series. In order to extend the capabilities of the model to capture extreme events which may have an impact on market dynamics, we introduce switching parameters that follow a discrete Markov chain. We define the Markov–Switching AMEM (MS–AMEM) as:

$$x_t = \mu_{t,s_t} \varepsilon_t, \quad \varepsilon_t | s_t \sim \text{Gamma}(a_{s_t}, 1/a_{s_t}) \text{ for each } t$$

$$\mu_{t,s_t} = \omega + \sum_{i=1}^n k_i I_{s_t} + \alpha_{s_t} (x_{t-1} - \mu_{t-1,s_{t-1}}) + \beta_{s_t}^* \mu_{t-1,s_{t-1}} + \gamma_{s_t} D_{t-1} (x_{t-1} - \mu_{t-1,s_{t-1}}) \quad (3.3)$$

where s_t is a discrete latent variable which ranges in $[1, \dots, n]$, representing the regime at time t . I_{s_t} is an indicator equal to 1 when $s_t \leq i$ and 0 otherwise; $k_i \geq 0$ and $k_1 = 0$. Thus, the constant in regime j is given by $(\omega + \sum_{i=1}^j k_i)$ and reflects non decreasing levels of volatility passing to higher regimes. The changes in regime are driven by a Markov chain, such that:

$$Pr(s_t = j | s_{t-1} = i, s_{t-2}, \dots) = Pr(s_t = j | s_{t-1} = i) = p_{ij} \quad (3.4)$$

The positiveness and stationary constraints given for (3.1) hold within each regime in (3.3) as well. The unconditional expected value within state j , $j = 1, \dots, n$, is equal to:

$$\mu_j = \frac{\omega + \sum_{i=1}^j k_i}{1 - \alpha_j - \beta_j - \gamma_j/2} \quad (3.5)$$

where $\beta_j = \beta_j^* - \alpha_j - \gamma_j/2$. In view of this reparameterization, the same unconditional expected value is shared by the model:

$$\mu_{t,s_t} = \omega + \sum_{i=1}^n k_i I_{s_t} + \alpha_{s_t} x_{t-1} + \beta_{s_t} \mu_{t-1,s_{t-1}} + \gamma_{s_t} D_{t-1} x_{t-1},$$

the parameters of which are reported, for ease of interpretation, in the estimation tables below, keeping separate the contributions of the most recent data and of the past conditional expected volatility.

Extensions of the model to make the transition probabilities (3.4) dependent on past observable variables are possible (cf. Maheu and McCurdy, 2002, who make the transition probabilities dependent on the duration in the state); we will propose a particular specification in the next section.

We can estimate the model (3.3) adopting Hamilton’s filter and smoother (Hamilton, 1994, ch.22). There is a path dependence issue which causes a well–known computational problem, due to the unobservability of μ_{t,s_t} and the dependence on past values of s_t : evaluating the likelihood recursively, we would need to keep track of all possible paths taken by the regime between $t = 1$ and $t = T$, making the model non tractable. The solution adopted in this case

is the one proposed by Kim (1994) for a state-space MS model (cf. also the many examples in Kim and Nelson, 1999). After each step of the Hamilton filter, at time t we collapse the n^2 possible values of μ_t into n values, by an average over the probabilities at time $t - 1$:

$$\hat{\mu}_{t,s_t} = \frac{\sum_{i=1}^n Pr[s_{t-1} = i, s_t = j | \Psi_t] \hat{\mu}_{t,s_{t-1},s_t}}{Pr[s_t = j | \Psi_t]} \quad (3.6)$$

where a hat indicates the estimate of the unknown variable and the probabilities in (3.6) are obtained by the Hamilton filter.³

Tests based on the likelihood function cannot be used to compare the MEM and AMEM with respect to the corresponding MS models because of the presence of nuisance parameters present only under the alternative hypothesis; in this case, with the proper caution, a classical BIC and AIC could provide some information (see Psaradakis and Spagnolo, 2003, with the AIC giving better results when the parameter changes are not too small and the hidden Markov chain is fairly persistent).

An alternative which proved to behave well in previous applications is the nonparametric Bayesian approach of Otranto and Gallo (2002). The procedure identifies the number of regimes in Markov switching models, based on the detection of the empirical posterior distribution of the number of regimes, via Gibbs sampling, using the nonparametric Bayesian techniques derived from the Dirichlet process theory.

4 Alternative Specifications

From a theoretical point of view, the AMEM approach can be extended in different directions following the suggestions discussed in Section 2. In this section we explore three alternative approaches which aim at capturing possible features missed by a Markov Switching approach, namely:

1. a time-varying transition probability approach to explore the possibility that the passage from one regime to another depends on market conditions represented by, for example, lagged values of realized volatility or by an independent measure of volatility, such as the VIX (an average of implied volatilities on at-the-money options on the Standard & Poor

³A better approximation can be obtained by considering n^3 possible values of μ_t (considering $t, t - 1$ and $t - 2$) and collapsing them into n^2 , with a substantial increase in the size of the Markov chain and of the computational burden.

500 index). This model is called TVTP–MS(k)–AMEM and is specified as

$$x_t = \mu_{t,s_t} \varepsilon_t, \quad \varepsilon_t | s_t \sim \text{Gamma}(a_{s_t}, 1/a_{s_t}) \text{ for each } t$$

$$\mu_{t,s_t} = \omega + \sum_{i=1}^n k_i I_{s_t} + \alpha_{s_t} (x_{t-1} - \mu_{t-1,s_{t-1}}) + \beta_{s_t}^* \mu_{t-1,s_{t-1}} + \gamma_{s_t} D_{t-1} (x_{t-1} - \mu_{t-1,s_{t-1}})$$

$$D_t = \begin{cases} 1 & \text{if } r_t < 0 \\ 0 & \text{if } r_t \geq 0 \end{cases}$$

$$p_{ij,t} = \frac{\exp(\theta_{ij} + \phi_{i,j} v_{t-1})}{1 + \exp(\theta_{ij} + \phi_{i,j} v_{t-1})} \quad (4.1)$$

where v_t is the volatility index.

2. a smooth transition approach to explore the possibility that the abrupt shifts from one regime to another in the MS model should be replaced by a more progressive passage across average volatilities, similarly to the Smooth Transition model of Teräsvirta (1994). Still maintaining the lagged volatility as the forcing variable in the transition function, we have the following specification (ST–AMEM):

$$x_t = \mu_t \varepsilon_t, \quad \varepsilon_t \sim \text{Gamma}(a, 1/a) \text{ for each } t$$

$$\mu_t = \omega + (\alpha_0 + \alpha_1 F(v_{t-1})) x_{t-1} + (\beta_0 - \beta_1 F(v_{t-1})) \mu_{t-1} + \gamma D_{t-1} x_{t-1} \quad (4.2)$$

$$F(v_t) = (1 + \exp(-g(v_t - c)))^{-1}.$$

3. Finally, to investigate the possibility that (quasi) long–memory may be able to capture the slow moving underlying level of volatility, we adopt a specification similar to Corsi’s (2007) HAR where the conditional volatility is made dependent on past volatilities aggregated at different frequencies (HMEM; D(aily), 5 is for the W(eek), 22 is for the M(onth)):

$$x_t = \mu_t \varepsilon_t, \quad \varepsilon_t \sim \text{Gamma}(a, 1/a) \text{ for each } t \quad (4.3)$$

$$\mu_t = \omega + \alpha_D x_{t-1} + \alpha_W \bar{x}_{t-1}^{(5)} + \alpha_M \bar{x}_{t-1}^{(22)}$$

with the possible introduction of regimes (MS(k)–HMEM):

$$x_t = \mu_{t,s_t} \varepsilon_t, \quad \varepsilon_t | s_t \sim \text{Gamma}(a_{s_t}, 1/a_{s_t}) \text{ for each } t \quad (4.4)$$

$$\mu_{t,s_t} = \omega + \sum_{i=1}^n k_i I_{s_t} + \alpha_{D,s_t} x_{t-1} + \alpha_{W,s_t} \bar{x}_{t-1}^{(5)} + \alpha_{M,s_t} \bar{x}_{t-1}^{(22)}$$

We will also keep track of the performance of the original HAR where the innovation term is added to the conditional expectation, rather than multiplied by it.

5 Regimes in the Volatility of the S&P500 Index

Let us now turn to the empirical analysis, choosing the realized kernel volatility of the S&P500 index to represent some typical features of realized volatility of a stock index with a very liquid base which may be common to other stock indices. Our data are taken from the *Oxford-Man Institute's Realised Library* version 0.1 (Heber et al., 2009), and are expressed as percentage annualized volatility (i.e. the square root of the realized kernel variance multiplied by $\sqrt{252} * 100$), as shown in Figure 1 between January 3, 2000 and October 26, 2012.⁴ The time series shows the high degree of persistence discussed above. The presence of changing levels of the prevailing average volatility by subperiods or other types of nonlinearity can be conjectured: the series shows in fact alternating regimes which visually involve a shift in the average level but may be accompanied by changes in the dynamics of the series. This is particularly clear in the sample chosen, with the turbulence leading to the burst of the tech bubble, the 2001 recession, the low level of volatility in mid decade and then the explosion of uncertainty following the subprime mortgage crisis.

The figure may be complemented by some descriptive statistics reported in Table 1 for the overall period between January 3, 2000 to October 26, 2012: we also report a split analysis between in-sample and out-of-sample with July 1, 2011 as the end of the estimation period (marked by a vertical bar in the figure). It is clear that the first period has more turbulent features than the second with a very large range with a thick right tail (high kurtosis) determined by the bursts of volatility particularly after September 2008, and reaching a maximum above 150% relative to an overall mean of about 13%). The slow decay in time dependence is reflected by the autocorrelation function, reported at lag 1 (one day), 5 (one week), 22 (one month), confirming the features found in the literature. The more recent period is definitely less turbulent with more moderate peaks and with skewness and kurtosis reflecting a substantially less thick right tail.

5.1 Model Estimation and Testing

The MEM and AMEM are estimated first as benchmarks to be used in what follows. We are aiming first at establishing the possible presence of misspecification through model diagnostics; we gauge switching models accordingly, as being capable of solving possible residual autocorrelation and improving model fit. We then turn to the evaluation of relative forecasting

⁴The Realized Library, continuously updated, provides a number of indicators among which the realized kernel variances for many financial indices. The estimator uses a weighted combination of squared returns and cross products of returns h lags apart ($h=1, \dots, H$) with a Parzen weight function and the bandwidth H optimally chosen as in section 2.1. of Barndorff-Nielsen *et al.* (2009) who also note the similarity of the realized kernel to a HAC type estimator.

performance.

The estimation results are reported in Table 2. The asymmetry coefficient γ is statistically significant, in line with previous evidence, as well as the information criteria, favoring the AMEM. The estimated persistence, measured as $(\alpha + \beta)$ for the MEM, and $(\alpha + \beta + \gamma/2)$ for the AMEM, is very high (around 0.93 for both models). Model diagnostics shows the presence of serially correlated residuals at all relevant lags (cf. the Ljung–Box test statistics reported in the related columns of Table 5 and the associated p-values) which is not corrected by increasing the order of the AMEM model (results not shown here).

In exploring the possibility of the presence of regimes, we start by applying the nonparametric Bayesian procedure suggested by Otranto and Gallo (2002) described at the end of the previous section. Table 3 reports the results for two choices of the hyperparameter A regulating the prior probabilities on the number of regimes (a lower A favors a lower number of regimes; the presence of autocorrelation in the raw data suggests the choice of low values of A , as suggested in Otranto and Gallo, 2002). There is quite a strong indication of the presence of regimes (absence of regimes get zero posterior probability) and that three is the favored number of regimes, with some evidence that a fourth regime could be accommodated.

As with other MS models, it is crucial to avoid over–parameterization in order to achieve a parsimonious representation with reliable parameter estimates: we estimate both a 3–state and a 4–state MS–AMEM using the most general specifications and then proceed to some coefficient testing to select a more parsimonious model and improve the precision of parameter estimates. In particular, with an eye to a simplification of the specification, we can calculate the Wald test statistics for the joint hypotheses in the MS–AMEM:

$$\alpha_i = \alpha_j, \quad \beta_i = \beta_j, \quad \gamma_i = \gamma_j$$

for each $i, j = 1, 2, 3(, 4)$ and $i \neq j$. The null is always rejected in the 4–state case, whereas in the 3–state case the corresponding p-values are 0.00 for $(i, j) = (1, 2)$ and 0.97 for $(i, j) = (1, 3)$ and 0.28 for $(i, j) = (2, 3)$, due to large standard errors of the coefficients in regime 3. We suggest a restricted model to have the same AMEM dynamics in regimes 1 and 3: its estimation confirms the significant difference of the dynamic coefficients in state 2.⁵ On the other hand, the MS(4)–AMEM presents some computational problem in estimation with some coefficients easily stuck at zero: as a matter of fact, the inversion of the Hessian is possible only when such coefficients are constrained. Even doing this, the sandwich matrix is not invertible, and the standard errors must be calculated in a non–robust way by using the inverse of the Hessian. In Table 4 we show the estimation results for the selected MS–AMEM models with 3 and 4 regimes (first two models of each panel) where the estimated coefficients are indicated with the relative standard errors unless they are constrained to some value (equality across coefficients

⁵The restricted model with same dynamics in state 2 and 3 does not get a similar empirical support and induces residual autocorrelation.

or zero restrictions). The MS(4)–AMEM shows some puzzling features such as the zero values for the shift in the intercepts for states 3 and 4 and the coefficient α_4 .

The shapes of the Gamma densities based on the estimated parameters a in the AMEM case (no regime) and for each regime of the MS(3)–AMEM and MS(4)–AMEM (Figure 2) give an idea of the interaction between conditional means and the innovations in each regime. For the MS(3)–AMEM, we see an approximate symmetry in the state of medium volatility (23.52, the largest value) and a fairly pronounced asymmetry for the state of high volatility (smallest a equal to 10.63), as one would expect. The larger variance of the regime representing the highest volatility (regime 3 for MS(3)–AMEM and 4 for MS(4)–AMEM) is evident, having peaks of different magnitude belonging to this state; moreover, in the MS(4)–AMEM, innovations of regime 1 and 3 have a similar distribution. In the 3-state model the regime 1 of low volatility has a larger variance than the regime 2, which could be considered a sort of *normal* volatility state. Our findings seems in line with the well-documented finding that the volatility of volatility is high if the level of volatility is also high, also due to a heteroskedastic measurement error (Barndorff-Nielsen and Shephard, 2005). Similarly Corsi et al. (2008), using several models and different frameworks, show that realized volatility is subject to heteroskedastic errors due to the time-varying volatility of the realized volatility estimator.

Finally the log-likelihood value (and the AIC and BIC), do not show particular gains for the 4-state models, which could be overparameterized (but we recall that these measures are not directly comparable for the problem of nuisance parameters present only under the null hypothesis).

One of the motivations to adopt an MS model to analyze volatility is the presence of auto-correlated residuals in the MEMs with constant parameters. In Table 5 we show the Ljung-Box statistics (Q) and the corresponding p-values (p) also for the two MS–AMEMs to check how uncorrelated the residuals are up to lag 20.⁶ What we observe is that the model with three regimes is able to capture the strong residual dependence structure still present in the MEM and the AMEM; the model with four regimes fails, possibly for an incorrect inference on the regimes.

In order to shed more light on the nature of the dynamics in realized volatility we keep the alternative specifications in the background; their estimation (details available upon request) shows that autocorrelation plagues all models (cf. Table 6), with the notable exception of the ST–AMEM whose residuals reveal acceptable patterns up to and including lag 10.

5.2 Inference on Regimes

We can assign each day to a regime based on the mode of the smoothed probabilities $Pr(s_t|\Psi_T)$ of being in one of the states for each t . These results are shown in Figure 3. In the upper

⁶For MS–AMEMs we have used the generalized residuals, introduced by Gouriéroux et al. (1987) for latent variable models, defined as $E(\hat{\epsilon}_t|\Psi_{t-1}) = \sum_{i=1}^3 \hat{\epsilon}_{s_t,t} Pr(s_t = i|\Psi_{t-1})$, where $\hat{\epsilon}_{s_t,t}$ are the residuals at time t derived from the parameters of the model in state s_t .

Panel A, we report the realized volatility series together with the S&P500 index (log-scale, to appreciate the local trends) and the regimes attributed according to the smoothed probabilities from the MS(3)–AMEM: fortunately, these probabilities are generally near zero or one, favoring a clear-cut association. There are many oscillations between state 2 and 3 until the second half of 2003, a feature which is consistent with the uncertainty following the burst of the dot-com bubble (which had its peak in March 2000), the 2001 recession, 9/11, the start of the Afghan war in October 2001 and of the Iraqi war in March 2003. There is a long quiet period between October 2003 and July 2007, characterized by a long lasting bullish rally in the stock market, with a sudden passage to the regime of high volatility on July 19, 2007 which lasts for about a month, marking the beginning of the sub-prime mortgage crisis and the first measures by the FED to contain the credit crunch. The second part of 2007 and the first part of 2008 are basically characterized by mid-volatility with occasional passages to the higher regime until the beginning of Sep. 2008 when the switch to regime 3 is more persistent, gaining momentum with the collapse of Lehman Brothers (with an outburst in the series on Oct. 20, 2008), and lasting until May 2009. The second half of 2009 is spent mostly in regime 2 with a passage to regime 1 at the end of the year. The following months see changes between regimes 1 and 2, although the month of May 2010 is marked by a sudden switch to regime 3 (following the flash crash of May 6).

The characterization of the regime switching is confirmed by the lower half of Table 4, where we report the estimated transition probabilities. The diagonal elements can be translated into the duration in a certain regime i as $\frac{1}{1-p_{ii}}$. Making reference to the MS(3)–AMEM, the estimates imply, on average, a 87 days permanence in the state of low volatility, which decreases to 28 days for the intermediate volatility state, and to 13 days for the high volatility state. This result is in line with the empirical evidence that the turmoil periods have a lower duration than quieter spells.

The off-diagonal elements of the transition probability matrix point to a strong interaction between regimes 2 and 3 while the period of low volatility is a sort of self standing regime. As a matter of fact, once in regime 1 there is a very low probability to switch to either of the other two regimes. From the regime of intermediate volatility, there is a higher probability to move to the high volatility regime than to revert to a low volatility regime. By the same token, note that the downward transition from the high volatility states occurs preferably with a move to the intermediate state.

In Panel B of the same figure, we show the inference on the regime using MS(4)–AMEM. As far as period classification is concerned, we tend to get the same indications as before, although this model generates frequent changes in regime (cf. the many isolated dots in the figure); this time around, the largest smoothed probability is not always close to one as with the MS(3)–AMEM. Regime 4 is associated with the most severe downturns in the market and the highest spikes in the volatility series, with the longest period in the regime around October 2008. It looks like Regime 3 and 4 in the MS(4)-AMEM capture most of what is identified as

Regime 3 in the MS(3)–AMEM, with somewhat of a higher interaction between Regimes 1 and 2 than before. In the MS(4)–AMEM case the four regimes have an expected duration equal to 77, 18, 15 and 5 days respectively.

In Figure 4 we show how the observations are assigned to the regimes using the two models. For regime 1 we observe a certain consistency in the two classifications. The observations belonging to regimes 2 and 3 for MS(3)–AMEM are often classified differently within MS(4)–AMEM. In particular many observations belonging to regime 2 in MS(3) are assigned to regime 1 in MS(4); similarly regime 3 in MS(4) contains many observations of regime 2 in MS(3). We can argue that this classification of MS(4) is the main cause of the presence of autocorrelation in its residuals and may be seen as a sign of erroneous classification.

5.3 Unconditional volatility

Each state corresponds to a different average level of volatility, expressed by (3.5). From our estimates of the MS(3)–AMEM, we derive the corresponding levels of volatility by regime (cf. Table 7) which are 9.21% in regime 1, 14.39% in regime 2 and 28.80% in regime 3. Considering the MS(4)–AMEM we notice similar levels for the first three regimes, with a large increase in the level of regime 4 (56.97%). Bursts of volatilities, as well as sudden reductions in their values, correspond to a discrete change in the average value around which volatility follows its dynamics. Moreover, the estimated persistence is clearly different from the one in the MEM and AMEM (recall that it is around 0.93) and also within each regime: we estimate it to be equal to 0.83 during the medium volatility, and to a smaller 0.79 in the low and high volatility regimes in MS(3)–AMEM. These results are consistent with the intuitive notion that ordinary dynamics (with higher persistence) is captured by the second regime and very quiet and turbulent periods show a lower persistence. A different interpretation is obtained with the MS(4)–AMEM, where the persistence increases with the regime (0.80 in regime 1, 0.81 in regime 2, 0.91 in regime 3 and 0.96 in regime 4). The comparison with the alternative specifications shows that unconditional volatility levels and ranges are fairly similar: TVTP–MS(3)–AMEM comprises an average level of volatility in regime 3 which is more similar to the one in Regime 4 of the MS(4)–AMEM; MS(3)–HMEM has lower values than the others for regimes 1 and 2, while it has a much higher average level for regime 3. For the ST–AMEM, being a continuous function of time, unconditional volatility is comprised between a minimum of 7.20 and a maximum of 27.40 (sensibly smaller than the observed bursts). HMEM and HAR have levels in line with the baseline specifications.

Further insights on the value added in considering our Markov Switching AMEM can be gained by dividing the original data by the regime specific average volatility as done in Figure 5. It is apparent that the models manage to remove the underlying slow–moving trend in the evolution of volatility: now the series oscillates around 1, with the exception, in the MS(3)–AMEM, of the episode started in September 2008. The figure in Panel A shows a fairly regular

behavior of the transformed series with some signs of autocorrelation (captured by the multiplicative structure of the model) with much less pronounced peaks relative to the original series. The fact that the observed series is well above the average level in the last portion of the sample may be an indication that the height of the crisis may be marked by an even higher average level of volatility, and a fourth (very short lived) regime would in principle accommodate that. Panel B shows that this is indeed the case, but it implies a worsening of the inference in the other regimes and a bad performance in terms of Ljung-Box test. A good compromise would be to opt not to fit an extra regime, but a single change in level, given the exceptional nature of the market volatility dynamics following the demise of Lehman Brothers. We call this model the MS(3)–AMEM(d) (d stands for dummy) and the estimation results are shown in Table 8 to be compared with the other MS models in Table 4. We can notice that the coefficient relative to the dummy is significant and that the rest of estimation results is very similar to the previous 3-state case (except, obviously, the lower value of k_3 , the state of highest volatility which includes the peak represented by the dummy coefficient). The Ljung–Box autocorrelation statistics of Table 5 show again a good behavior, whereas Panel C of Figure 5 shows that the volatility level divided by the regime-specific average falls in a small interval around 1 for the full period considered. This results seems to provide a further strong evidence in favor of the 3-state case, fairly reassuring as of the little impact that the burst of volatility at the end of 2008 has on the overall picture, excluding the presence of a fourth regime.

Figure 6 shows that the problems addressed with the use of the dummy would not be solved with either of the alternative specifications: as a matter of fact, the TVTP–MS(3)–AMEM shows quite an increased range of fluctuations in all periods, notably for the periods of low and medium volatility and no improvement in the high turbulence period. SP–AMEM is certainly better, but fails to capture the 2008 episode. Finally, we needed to resort to an increased y–axis scale to accommodate the results of the MS(3)–HMEM, a model providing very unsatisfactory results in terms of capturing changes in regimes and low–frequency movements in the series.

5.4 Estimation accuracy

In order to evaluate the properties of our volatility predictions, we resort to the calculation of 95% confidence intervals for the expected volatility values, using our models. For the MEM, AMEM, ST–AMEM and HMEM we compute the lower and upper values of the intervals, respectively, as the 2.5 and the 97.5 percentiles of the Gamma distribution with parameters \hat{a} and $\hat{\mu}_t/\hat{a}$. For the MS models, instead, we derive the bounds numerically, using a grid search, from a mixture of distributions $Gamma(\hat{a}_{s_t}, \hat{\mu}_{s_t}/\hat{a}_{s_t})$ by regime, with weights equal to the corresponding smoothed probabilities. Finally the bounds for the HAR are derived by resorting to estimation residual moving–block bootstrapping (Künsch, 1989, to accommodate serial correlation).

In Table 9 we show the mean width of the intervals (together with their standard deviation)

as values computed across the T intervals and then, for the MS models, the corresponding values when the regime at time t is identified by model specific smoothed probabilities. The mean widths (and their standard deviations) on the whole period are by and large comparable with one another (with the MS–AMEMs being the smallest), with the exception of the HAR’s which is much larger. The analysis by regime shows that, again, the three MS–AMEM have comparable widths with the understandable exception of the MS(4)–AMEM where Regime 4 accommodates some of the uncertainty elsewhere captured in Regime 3. Somewhat surprisingly, the TVTP–MS(3)–AMEM has uniformly larger mean widths and standard deviations.

Note that conditional expected volatilities obtained by the MS–AMEMs explicitly consider a different underlying regime–specific average volatility. Relative to other models, this should imply fewer observed volatilities falling outside the interval bounds. For a nominal confidence interval of 95% and for each model, we show the percentage of out–of–bound observations in the left part of Table 10, distinguishing conditional overprediction (i.e. the lower bound of the interval is greater than the observed data) from underprediction (the upper bound of the interval is lower than the observed value). The relative incidence of such occurrences is lower for TVTP–MS(3)–AMEM and MS(3)–HMEM (less than 2.9% against more than 4.0% for the other models) but it is not equally spread between overprediction and underprediction (the first case is more frequent), differently from the MS(3)–AMEMs and HAR (which has constant intervals). By the same token, the MEM, the AMEM, the MS(4)–AMEM, the ST–AMEM and the HMEM underpredict more frequently. We find it instructive to further decompose the incidence of overprediction, respectively, underprediction calculating the corresponding fraction in each regime (central and right part of the Table 10). It is clear that the MS models, in general, have a decreasing pattern of overprediction as the volatility regime increases, but there is also a large percentage of errors for MS(4)–AMEM in regime 4, which contains a small number of episodes (see Panel B of Figure 3). Both MS(3)–AMEMs have higher shares of underprediction within regimes 1 and 3, whereas for the MS(4)–AMEM and the MS–HMEM it typically shows up in regime 1 and for TVTP–MS(3)–AMEM in regime 3. Summarizing, the MS(3)–AMEM and MS(3)–AMEM(d) have patterns better balanced between over– and underprediction than the other models.

To give a better idea of what is implied by these models in prediction, we expose in Figure 7 some episodes (March 2005 – panel A, October 2008 – Panel B, and March 2011 – Panel C) characterized by different volatility behavior, comparing the observed realized kernel with the 95% prediction intervals derived from the MS(3)–AMEM(d), ST–AMEM and MS(3)–HMEM (we select one model for each kind of regime change or dynamics). In the first period there is a prevalence of low volatility: all the prediction intervals contain the observed volatility. The second period contains the highest peak in the sample: it is interesting to notice that the MS(3)–AMEM(d) is able to detect the jump on October 10, while the ST–AMEM and MS(3)–HMEM react to the shock with a lag and the ST–AMEM underpredicts its size. Given the cost in inserting a fourth regime both computationally and in terms of misprediction in other periods,

we conclude that a few sizable peaks in the series are better not attributed to an extra regime. Admittedly, the insertion of a dummy can be made only after the fact, but the benefit is definitely higher in the overall performance of the corresponding 3–state model. This is confirmed by the inspection of the third period where several regimes are present. It is possible to notice that the MS models are able to follow these dynamics, while the ST–AMEM is at a loss, widening and tightening the bands at the “wrong” times. Relative to the 3–state models, the wider prediction interval for the MS(3)–HMEM is also apparent.

5.5 In– and Out–of–sample Performances

In this section we evaluate the forecasting performance of the Markov Switching models relative to the base model and the extensions presented in Section 4. There is a large debate about the usefulness of adopting loss functions such as the Mean Squared Error (MSE) and Mean Absolute Error (MAE) both in– and out–of–sample. They can be used descriptively, such as in Table 11 for the ten estimated models pointing to the better performance of the TVTP–MS(3)–AMEM in–sample, of the ST–AMEM for the MSE loss function (1– and 2–steps ahead) and of TVTP–MS(3)–AMEM and MS(3)–AMEM(d) for the MAE loss function (1– and 2– steps ahead respectively). From an inferential point of view, the popular Diebold–Mariano (1995) test can be used to verify the null hypothesis of no difference in the accuracy of two competing forecasts: it boils down to verifying whether the mean of the differences of the squared forecast errors and absolute forecast errors of each pair of models is zero. Traditionally, Diebold–Mariano tests (more generally, any predictive ability test) are applied out-of-sample and do not take into account parameter uncertainty; Inoue and Kilian (2005) show, both analytically and via Monte Carlo simulations, that out-of-sample tests have lower power than their in-sample counterparts for linear model with the same number of parameters (in our nonlinear context the reference to Inoue and Kilian’s results should be meant as a heuristic extension). Diebold (2012) stresses that the Diebold–Mariano test is intended for comparing out–of–sample forecasts, not for model selection, while the in-sample test can give some comparative indications about model properties in terms of fitting capabilities. In general, an interesting result by Hansen (2010) shows that there is a strongly negative relationship between in– and out–of–sample performances, in the sense that in–sample criteria would favor over–parameterized models (an extreme case is a model with T dummy variables), while the contrary tends to be true in an out–of–sample framework. An important implication of the theoretical result of Hansen is that an over–parameterized model always tends to do worse than a more parsimonious model in the out–of–sample evaluation. Therefore, the fact that a more complex model has a better in–sample performance, but at the same time it is not worse out–of–sample than a simpler model, can be interpreted as a strong evidence in favor of the more complex model.

Having this in mind, let us consider Table 12, where we show the results of the Diebold–Mariano statistics with the correction proposed by Harvey et al. (1997); for the out–of–sample

case we use a span between Jul. 5, 2011 and Oct. 26, 2012 (334 observations), but we do not re-estimate coefficient values. We synthesize results, using a symbol to indicate that the model by column performs significantly better (at a 5% significance level) than the model by row in terms of in-sample (presence of a star), or out-of-sample (1-step with a diamond and 2-steps-ahead with a black circle). Notice that the TVTP-MS(3)-AMEM fits the data (in-sample performance) significantly better than the other models in MAE terms while the MS(3)-AMEM(d) retrieves substantial ground when squared errors are involved, since individual episodes connected to peaks are better captured by it. The nonlinear models have a significantly better performance relative to the linear models (excluding AMEM) in the 1-step out-of-sample comparison. We notice a similar better performance of the models containing the asymmetric effect of returns (excluding MS(4)-AMEM) relative to the models without this effect. It is interesting to notice that the difference in the estimated level of the volatility in regime 3 between MS(3)-AMEM and MS(3)-AMEM(d) amounts to a better out-of-sample performance of the latter (when absolute errors are involved). The similar performance of the MS(3)-AMEMs when compared to the AMEM has to be interpreted in the light of Hansen's (2010) comments.

In Table 13 we perform a simultaneous comparison of the models showing the results of the Model Confidence Set (MCS) approach of Hansen et al. (2011);⁷ in this case the symbols are inserted in correspondence of the models belonging to the set with best forecasting performance. The results show that the TVTP-MS(3)-AMEM is better than other models in-sample; in the 1-step ahead forecasting the models provide a similar performance, apart from HMEM and HAR, and in the 2-step ahead forecasting, AMEM, all the MS(3)-AMEMs and ST-AMEM are the best ones when absolute errors are involved.

6 Concluding Remarks

In this paper we have focused on the identification and interpretation of regime switching in volatility, using the realized kernel volatility of the S&P500 as our representative application. We have suggested several extensions of the univariate Asymmetric Multiplicative Error Model (AMEM) to accommodate sudden changes in the underlying average level of volatility. Such a feature is shared by most volatility series measured from the year 2000 until at least the end of 2012, since financial markets underwent an alternation between persistent periods of turmoil and of quiet. This phenomenon goes beyond the type of volatility clustering which can be accommodated by a standard AMEM (as shown by its severe residual autocorrelation). Recalling the discussion in Sections 1 and 2 and the references therein, in fact, there are several studies trying to model it either by the insertion of a spline function, or recurring to Markov Switching or resorting to some long memory model.

⁷This by-passes a limitation of the Diebold and Mariano test which involves two models at the time with possible inconsistencies. To perform the MCS approach we have used the Quasi-Likelihood loss function with the semi-quadratic statistics; see Clements et al. (2009) for details.

Modeling the series at hand, the presence of a sudden and short-lived burst in volatility in October 2008 highlights important points about regime switching models: the fact that regimes may be sample-specific, posing some challenges for out-of-sample behavior and the possible presence of outliers may cloud the picture of how many regimes are to be considered. We tackle this problem by running a nonparametric Bayesian procedure prior to estimation which points to the presence of three regimes and, therefore, give way to the interpretation of the October 2008 as an isolated episode. The estimation results show that the 3-state MS-AMEM has indeed better properties in capturing the behavior of the series: the 4-state version suffers from numerical difficulties in convergence; while it classifies the episode as extreme, it fails to capture the size of the jump, and ends up signaling as extreme other periods with different characteristics making the overall inference on the regimes less stable. We manage to further improve the fit and general properties by allowing a dummy variable to change the value of the constant in the conditional expectation expression just for the day of the jump. Relative to the AMEM and the 4-state model, the better performance of the 3-state model is also confirmed by the analysis of the residual autocorrelation, prediction confidence intervals (especially for specific meaningful episodes) and MAE and MSE results in- and out-of-sample. Another extension, the TVTP-MS(3)-AMEM, seems to have good properties in terms of and out-of-sample performance but it fails in terms of residual autocorrelation. We are reassured that we are not incurring in the curse of overparameterization by the discussion in Hansen (2010): our 3-state models outperform the AMEM in-sample and fare equally well out-of-sample in formal forecast comparison tests with absolute and squared loss functions.

A comparison with other approaches goes beyond the scope of this paper: one main difference is that we model volatility directly with a proper distribution assumption on the innovations rather than using regression-like models on log-volatility. From a forecasting point of view, it is well possible that other models may reveal a better out-of-sample performance, but we consider the discussion on the economic interpretability of the inference on the regimes to have provided interesting insights.

This good performance can be replicated also for other time series, as done in Gallo and Otranto (2013) where other American stock indices, differing by degrees of capitalization were analyzed. The lesson we learn is that the increasing flexibility of the distribution hypothesized for the error terms favors the possibility to respect the statistical hypothesis of the model; in practice the correct hypothesis about the error distribution seems to be a crucial task to increase the quality of the model estimation.

The extensions that may be envisaged see Markov Switching models with time-varying transition probabilities depending on alternative forcing variables: a source of inspiration is the linear MS model for volatility adopted by Maheu and McCurdy (2002) with duration in the state as the driving variable. Other choices related to market activity are certainly possible and may shed further light on the volatility dynamics, especially in the analysis of the short-lived burst of volatility which have characterized several recent financial crises.

References

- [1] Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys. 2000. Great Realizations. *Risk* 13: 105–108.
- [2] Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys. 2003. Modeling and Forecasting Realized Volatility. *Econometrica*. 71: 579–625.
- [3] Andersen, T.G., T. Bollerslev, and F.X. Diebold. 2010. Parametric and Nonparametric Volatility Measurement. In Y. Aït-Sahalia and L.P. Hansen (eds.), *Handbook of Financial Econometrics*, 67–138: Amsterdam: North-Holland.
- [4] Antoniak, C.E.. 1974. Mixtures of Dirichlet Processes with Applications to Bayesian Non-parametric Problems. *The Annals of Statistics*. 2: 1152–1174.
- [5] Baillie, R., and G. Kapetanios. 2007. Testing for Neglected Nonlinearity in Long Memory Models. *Journal of Business and Economic Statistics* 25: 447–461.
- [6] Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde, and N. Shephard. 2008. Designing Realised Kernels to Measure the Ex-post Variation of Equity Prices in the Presence of Noise. *Econometrica* 76: 1481–1536.
- [7] Barndorff-Nielsen, O.E., P.R. Hansen, A. Lunde, and N. Shephard. 2009. Realised Kernels in Practice: Trades and Quotes. *Econometrics Journal* 12: 1–32.
- [8] Barndorff-Nielsen, O.E., and N. Shephard. 2005. How Accurate Is the Asymptotic Approximation to the Distribution of Realised Variance? In D. Andrews, J. Powell, P. Ruud and J. Stock (eds.), *Identification and Inference for Econometric Models. A Festschrift for Tom Rothenberg* 306–331. Cambridge: Cambridge University Press.
- [9] Bauwens, L., C.M. Hafner, and S. Laurent. 2012. Volatility Models. In L. Bauwens, C.M. Hafner and S. Laurent (eds.), *Handbook of Volatility Models and their Applications*, 1–45. Hoboken, New Jersey: J. Wiley & Sons.
- [10] Bollerslev, T. 1986. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 31, 307–321.
- [11] Bordignon, S., and D. Raggi. 2012. Long Memory and nonlinearities in realized volatility: a Markov switching approach. *Computational Statistics and Data Analysis*, 56, 3730–3742.
- [12] Bos, C.S., P.H. Franses, and M. Ooms. 2002. Inflation Forecast Intervals and Long Memory Regression Models. *International Journal of Forecasting* 18, 243–264.

- [13] Brownlees, C.T., and G.M. Gallo. 2010. Comparison of Volatility Measures: a Risk Management Perspective. *Journal of Financial Econometrics* 8, 29–56.
- [14] Cipollini, F., R.F. Engle, and G.M. Gallo. 2013. Semiparametric Vector MEM. *Journal of Applied Econometrics* 28, 1067–1086.
- [15] Clements, A., M. Doolan, S. Hurn, and R. Becker. 2009. On the Efficacy of Techniques for Evaluating Multivariate Volatility Forecasts. Working Paper Series 41, Financial Economic Program, NCER.
- [16] Corsi, F. 2009. A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics* 7, 174–196.
- [17] Corsi, F., S. Mittnik, C. Pigorsch, and U. Pigorsch. 2008. The Volatility of Realized Volatility. *Econometric Reviews* 27, 46–78.
- [18] De Luca, G. and G.M. Gallo. 2004. Mixture Processes for Financial Intradaily Durations. *Studies in Nonlinear Dynamics and Econometrics*, 8. DOI: 10.2202/1558-3708.1223.
- [19] De Luca, G., and G.M. Gallo. 2009. Time-varying Mixing Weights in Mixture Autoregressive Conditional Duration Models. *Econometric Reviews* 28, 102–120.
- [20] Diebold, F.X. 2012. Comparing Predictive Accuracy, Twenty Years Later: a Personal Perspective on the Use and Abuse of Diebold-Mariano Tests. Working Paper 18391, National Bureau of Economic Research.
- [21] Diebold, F.X., and R.S. Mariano. 1995. Comparing Predictive Accuracy. *Journal of Business and Economic Statistics* 13, 253–263.
- [22] Diebold, F. X., and A. Inoue. 2001. Long Memory and Regime Switching. *Journal of Econometrics* 105, 131–159.
- [23] Dueker, M.J. 1997. Markov Switching in GARCH Processes and Mean-reverting Stock-market Volatility. *Journal of Business and Economic Statistics* 15, 26–34.
- [24] Edwards, S., and R. Susmel. 2003. Interest-rate Volatility in Emerging Markets. *The Review of Economics and Statistics* 85, 328–348.
- [25] Engle, R.F. 1982. Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation, *Econometrica* 50, 987–1008.
- [26] Engle, R.F. 2002. New Frontiers for ARCH Models. *Journal of Applied Econometrics* 17, 425–446.
- [27] Engle, R.F., and G.M. Gallo. 2006. A Multiple Indicators Model for Volatility Using Intradaily Data, *Journal of Econometrics* 131, 3–27.

- [28] Engle, R.F., and J.G. Rangel. 2008. The Spline-GARCH Model for Low-frequency Volatility and its Global Macroeconomic Causes. *Review of Financial Studies* 21, 1187–1222.
- [29] Escobar, M.D., and M. West. 1995. Bayesian Density Estimation and Inference Using Mixtures. *Journal of the American Statistical Association*. 90: 577–588.
- [30] Ferguson, T.S. 1973. A Bayesian Analysis of Some Nonparametric Problems. *The Annals of Statistics*. 1: 209–230.
- [31] Gallo, G.M., and E. Otranto. 2007. Volatility Transmission across Markets: a Multichain Markov Switching Model. *Applied Financial Economics* 17, 659–670.
- [32] Gallo, G.M., and E. Otranto. 2008. Volatility Spillovers, Interdependence and Comovements: a Markov Switching Approach. *Computational Statistics and Data Analysis* 52, 3011–3026.
- [33] Gallo, G.M., and E. Otranto. 2013. Volatility Swings in the US Financial Markets. In M. Grigoletto, F. Lisi and S. Petrone (eds.), *Complex Models and Computational Methods in Statistics* 137–148. Milan: Springer-Verlag.
- [34] Garcia, R. 1995. Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models. *International Economic Review*. 39: 763–88.
- [35] Glosten, L., R. Jagannathan, and D. Runkle. 1993. On the Relation between Expected Value and the Volatility of the Nominal Excess Return on Stocks *Journal of Finance* 48, 1779–1801.
- [36] Gouriéroux, C., A. Monfort, and E.R.A. Trognon. 1987. Generalized Residuals. *Journal of Econometrics* 34, 5–32.
- [37] Hamilton, J.D. 1989. A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica* 57, 357–384.
- [38] Hamilton, J.D. 1990. Analysis of Time Series Subject to Changes in Regime. *Journal of Econometrics* 45, 39–70.
- [39] Hamilton, J.D. 1994. *Time series Analysis*. Princeton: Princeton University Press.
- [40] Hamilton, J.D., and R. Susmel. 1994. Autoregressive Conditional Heteroskedasticity and Changes in Regime. *Journal of Econometrics* 64, 307–333.
- [41] Hansen, B.E. 1992. The Likelihood Ratio Test under Nonstandard Conditions: Testing the Markov Switching Model of GNP. *Journal of Applied Econometrics*. 7: S61–S82.

- [42] Hansen, P.R. 2010. A Winner's Curse for Econometric Models: on the Joint Distribution of In-sample Fit and Out-of-sample Fit and its Implications for Model Selection. Mimeo, Stanford University.
- [43] Hansen, P.R., A. Lunde, and J.M. Nason. 2011. The Model Confidence Set. *Econometrica*, 79, 453–497.
- [44] Hansen, P.R., Z. Huang, and H.H. Shek. 2012. Realized GARCH: a Joint Model of Returns and Realized Measures of Volatility. *Journal of Applied Econometrics* 27, 877–906.
- [45] Harvey, D., S. Leybourne, and P. Newbold. 1997. Testing the Equality of Prediction Mean Squared Errors. *International Journal of Forecasting* 13, 281–291.
- [46] He, Z., and J.M. Maheu. 2010. Real Time Detection of Structural Breaks in GARCH Models. *Computational Statistics and Data Analysis* 54, 2628–2640.
- [47] Heber, G., A. Lunde, N. Shephard, and K. Sheppard. 2009. OMI's realised library, version 0.1. Oxford-Man Institute, University of Oxford.
- [48] Higgs, H., and A.C. Worthington. 2004. Transmission of Returns and Volatility in Art Markets: a Multivariate GARCH Analysis. *Applied Economics Letters* 11, 217–222.
- [49] Inoue, A., and L. Kilian. 2005. In-sample or Out-of-sample Tests of Predictability? Which One Should We Use? *Econometric Reviews* 23, 371–402.
- [50] Kim, C.J. 1994. Dynamic Linear Models with Markov-switching. *Journal of Econometrics* 60, 1–22.
- [51] Kim, C.J., and C.R. Nelson. 1999. *State-Space Models with Regime Switching*. Cambridge, MA: MIT Press.
- [52] Klaassen, F. 2002. Improving GARCH Volatility Forecasts with Regime-Switching GARCH. *Empirical Economics* 27, 363–394.
- [53] Künsch, H.R. 1989. The Jackknife and the Bootstrap for General Stationary Observations. *Annals of Statistics* 17, 1217–1241.
- [54] Lanne, M. 2006. A Mixture Multiplicative Error Model for Realized Volatility. *Journal of Financial Econometrics* 4, 594–616
- [55] Maheu, J.M., and T.H. McCurdy. 2002. Non-linear Features of Realized FX Volatility. *Review of Economics and Statistics* 84, 668–681.
- [56] Maheu, J.M., and T.H. McCurdy. 2007. Components of market risk and return. *Journal of Financial Econometrics* 5, 560–590.

- [57] Maheu, J.M., and T.H. McCurdy, 2011. Do high-frequency measures of volatility improve forecasts of return distributions? *Journal of Econometrics* 160, 69–76.
- [58] McAleer, M. and M.C. Medeiros. 2008. A Multiple Regime Smooth Transition Heterogeneous Autoregressive Model for Long Memory and Asymmetries. *Journal of Econometrics* 147, 104–119.
- [59] Ohanissian, A., J. Russell, and T. Tsay. 2008. True or Spurious Long Memory? A New Test. *Journal of Business and Economic Statistics* 26, 161–175.
- [60] Otranto, E., and G.M. Gallo. 2002. A Nonparametric Bayesian Approach to Detect the Number of Regimes in Markov Switching Models. *Econometric Reviews* 21, 477–496.
- [61] Perron, B., and Z. Qu. 2010. Long-memory and Level Shifts in the Volatility of Stock Market Return Indices. *Journal of Business and Economic Statistics* 28, 275–290
- [62] Pong, S., M.B. Shackleton, S.J. Taylor, and X. Xu. 2004. Forecasting Currency Volatility: a Comparison of Implied Volatilities and AR(FI)MA Models. *Journal of Banking and Finance* 28, 2541–2563.
- [63] Psaradakis, Z., and F. Spagnolo. 2003. On the Determination of the Number of Regimes in Markov-Switching Autoregressive Models. *Journal of Time Series Analysis* 24, 237–252.
- [64] Scharth, M. and M. Medeiros. 2009. Asymmetric Effects and Long Memory in the Volatility of DJIA Stocks. *International Journal of Forecasting* 25, 304–327.
- [65] Shephard, N., and K. Sheppard. 2010. Realising the Future: Forecasting with High-frequency-based Volatility (HEAVY) Models. *Journal of Applied Econometrics* 25, 197–231.
- [66] Teräsvirta, T. 1994. Specification, estimation and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association* 89, 208–219.
- [67] Teräsvirta, T. 2009. An Introduction to Univariate GARCH Models. In T.G. Andersen, R.A. Davis, J.-P. Kreiß and T. Mikosch (eds.), *Handbook of Financial Time Series*, 17–42, Berlin-Heidelberg: Springer.
- [68] Teräsvirta, T. 2011. Nonlinear Models for Autoregressive Conditional Heteroskedasticity. CREATES Research Papers 2011-02, School of Economics and Management, University of Aarhus.

Notes

¹In a GARCH framework, previous contributions had addressed regime switching, cf. the SWARCH model (Hamilton and Susmel, 1994), the MS GARCH model (Dueker, 1997, Klaassen, 2002), and the recent multivariate extensions (Edwards and Susmel, 2003, Higgs and Worthington, 2004, Gallo and Otranto, 2007, 2008). An alternative way to consider changes in regime is given by smooth transition GARCH models (Teräsvirta, 2009) or other nonlinear models (Teräsvirta, 2011). Several other authors indicate the presence of level shifts in GARCH (Perron and Qu, 2010) or breaks in unknown points also in GARCH (He and Maheu, 2010) as the cause of an apparent high persistence. The issue of time varying underlying level of volatility is addressed also by Engle and Rangel (2008), who adapt a spline function in GARCH to capture a low frequency component of volatility (which they connect to macroeconomic factors).

²Of course less restrictive (but less parsimonious) specifications could be adopted, such as the Generalized Gamma.

³A better approximation can be obtained by considering n^3 possible values of μ_t (considering $t, t-1$ and $t-2$) and collapsing them into n^2 , with a substantial increase in the size of the Markov chain and of the computational burden.

⁴The Realized Library, continuously updated, provides a number of indicators among which the realized kernel variances for many financial indices. The estimator uses a weighted combination of squared returns and cross products of returns h lags apart ($h=1, \dots, H$) with a Parzen weight function and the bandwidth H optimally chosen as in section 2.1. of Barndorff-Nielsen *et al.* (2009) who also note the similarity of the realized kernel to a HAC type estimator.

⁵The restricted model with same dynamics in state 2 and 3 does not get a similar empirical support and induces residual autocorrelation.

⁶For MS-AMEMs we have used the generalized residuals, introduced by Gouriéroux *et al.* (1987) for latent variable models, defined as $E(\hat{\varepsilon}_t | \Psi_{t-1}) = \sum_{i=1}^3 \hat{\varepsilon}_{s_t, t} Pr(s_t = i | \Psi_{t-1})$, where $\hat{\varepsilon}_{s_t, t}$ are the residuals at time t derived from the parameters of the model in state s_t .

⁷This by-passes a limitation of the Diebold and Mariano test which involves two models at the time with possible inconsistencies. To perform the MCS approach we have used the Quasi-Likelihood loss function with the semi-quadratic statistics; see Clements *et al.* (2009) for details.

TABLES

Table 1: Descriptive statistics for the S&P500 realized kernel volatility (in annualized percentage terms). Full span: January 3, 2000 to October 26, 2012; out-of-sample period starts on July 5, 2011.

	Full	In-sample	Out-of-sample
Mean	15.54	15.60	15.00
Median	13.10	13.10	12.95
Min	3.49	3.49	4.50
Max	153.19	153.19	62.43
St.Dev.	10.02	10.16	8.76
Skewness	3.26	3.34	2.06
Kurtosis	20.67	21.43	5.73
$\rho(1)$	0.82	0.82	0.74
$\rho(5)$	0.70	0.71	0.57
$\rho(22)$	0.53	0.55	0.41

$\rho(i)$ indicates the autocorrelation at lag i .

Table 2: S&P500 realized kernel volatility (in annualized percentage terms). Coefficient estimates for the two benchmark models, MEM and AMEM (standard errors in parentheses), together with likelihood-based criteria. Sample: Jan. 3, 2000 to Jul. 1, 2011.

	ω	α	β	γ	a	log-lik	AIC	BIC
MEM	1.05 (0.07)	0.40 (0.03)	0.53 (0.03)		13.44 (0.48)	-7834.23	5.47	5.48
AMEM	0.94 (0.07)	0.27 (0.01)	0.61 (0.01)	0.11 (0.01)	14.21 (0.51)	-7752.06	5.42	5.43

Table 3: Nonparametric Bayesian approach to detect the number of regimes. Empirical posterior distribution of the number of regimes as a function of the hyperparameter A .

A	number of regimes			
	1	2	3	4
0.1	0.00	0.00	0.93	0.07
0.3	0.00	0.00	0.91	0.09

The nonparametric Bayesian approach of Otranto and Gallo (2002) is a procedure to identify the number of regimes in Markov switching models, based on the detection of the empirical posterior distribution of the number of regimes, via Gibbs sampling, using the nonparametric Bayesian techniques derived from the Dirichlet processes theory. The hyperparameter A regulates the prior probabilities on the number of regimes. The other prior distributions adopted are the same as in Otranto and Gallo (2002).

Table 4: Coefficient estimates for Markov Switching AMEM specifications with 3 and 4 regimes and Gamma innovations (standard errors in parentheses; when the s.e. is omitted, the corresponding coefficient is constrained to take on the reported value). Likelihood-based criteria are reported in the lower portion. Sample: Jan. 3, 2000 to Jul. 1, 2011.

Mean Coefficients											
ω	k_2	k_3	k_4								
MS(3)–AMEM											
1.92 (0.13)	0.44 (0.19)	3.64 (0.36)									
MS(4)–AMEM											
1.92 (0.18)	0.66 (0.42)	0.00	0.00								
α_1	α_2	α_3	α_4	β_1	β_2	β_3	β_4	γ_1	γ_2	γ_3	γ_4
MS(3)–AMEM											
0.14 (0.01)	0.12 (0.01)	0.14		0.58 (0.03)	0.67 (0.03)	0.58		0.13 (0.01)	0.08 (0.01)	0.13	
MS(4)–AMEM											
0.14 (0.02)	0.15 (0.03)	0.05 (0.02)	0.00	0.59 (0.04)	0.62 (0.05)	0.81 (0.03)	0.91 (0.02)	0.13 (0.01)	0.08 (0.01)	0.09 (0.02)	0.09
Gamma parameters								Likelihood-based criteria			
a_1	a_2	a_3	a_4								
MS(3)–AMEM											
17.72 (0.70)	23.52 (1.25)	10.63 (0.91)						-7603.06	5.32	5.36	
MS(4)–AMEM											
17.70 (0.89)	26.02 (1.97)	20.28 (2.27)	7.42 (0.97)					-7559.93	5.30	5.37	
Transition Probabilities											
MS(3)–AMEM											
p_{11}	p_{12}	p_{13}	p_{14}	p_{21}	p_{22}	p_{23}	p_{24}				
0.99 (0.00)	0.01 (0.00)	0.00		0.01 (0.00)	0.96 (0.01)	0.03					
MS(4)–AMEM											
0.99 (0.01)	0.00	0.00	0.01	0.00	0.94 (0.01)	0.02 (0.01)	0.04				
MS(3)–AMEM											
p_{31}	p_{32}	p_{33}	p_{34}	p_{41}	p_{42}	p_{43}	p_{44}				
0.01 (0.00)	0.06 (0.01)	0.93									
MS(4)–AMEM											
0.00	0.07	0.93 (0.02)	0.00	0.06 (0.03)	0.11 (0.04)	0.01 (0.01)	0.82				

Table 5: Residual autocorrelation of AMEMs and MS–AMEMs: Ljung-Box Q statistics and corresponding p-values. Sample: Jan. 3, 2000 to Jul. 1, 2011.

lag	MEM		AMEM		MS(3)–AMEM		MS(4)–AMEM		MS(3)–AMEM(d)	
	Q	p-value	Q	p-value	Q	p-value	Q	p-value	Q	p-value
1	5.44	0.02	5.12	0.02	0.84	0.36	15.93	0.00	1.17	0.28
2	7.09	0.03	7.38	0.02	1.81	0.41	22.19	0.00	1.91	0.38
3	7.74	0.05	7.77	0.05	2.34	0.51	22.22	0.00	2.36	0.50
4	10.96	0.03	9.72	0.05	2.63	0.62	25.15	0.00	2.53	0.64
5	14.39	0.01	12.31	0.03	3.58	0.61	31.52	0.00	3.36	0.64
6	15.36	0.02	13.15	0.04	3.58	0.73	37.58	0.00	3.39	0.76
7	17.01	0.02	14.22	0.05	3.78	0.81	41.57	0.00	3.55	0.83
8	19.81	0.01	16.81	0.03	3.83	0.87	48.17	0.00	3.59	0.89
9	26.36	0.00	21.71	0.01	3.97	0.91	55.30	0.00	3.60	0.94
10	35.71	0.00	30.56	0.00	6.09	0.81	67.34	0.00	5.17	0.88
11	50.33	0.00	40.33	0.00	7.71	0.74	75.33	0.00	6.80	0.82
12	58.10	0.00	47.61	0.00	8.93	0.71	84.53	0.00	8.37	0.76
13	71.43	0.00	58.53	0.00	10.31	0.67	94.71	0.00	10.04	0.69
14	82.12	0.00	69.81	0.00	13.11	0.52	107.47	0.00	14.02	0.45
15	87.92	0.00	75.01	0.00	13.97	0.53	114.85	0.00	15.26	0.43
16	93.04	0.00	79.41	0.00	14.25	0.58	120.31	0.00	15.71	0.47
17	93.82	0.00	79.77	0.00	14.79	0.61	121.78	0.00	16.21	0.51
18	95.84	0.00	80.97	0.00	15.01	0.66	124.04	0.00	16.52	0.56
19	102.04	0.00	87.11	0.00	15.54	0.69	133.35	0.00	16.82	0.60
20	108.26	0.00	93.77	0.00	16.97	0.65	145.47	0.00	18.30	0.57

Table 6: Residual autocorrelation of the alternative models: Ljung-Box Q statistics and corresponding p-values. Sample: Jan. 3, 2000 to Jul. 1, 2011.

lag	TVTP-MS(3)-AMEM		ST-AMEM		HMEM		MS(3)-HMEM		HAR	
	Q	p-value	Q	p-value	Q	p-value	Q	p-value	Q	p-value
1	52.05	0.00	0.04	0.84	1.59	0.21	3.23	0.07	6.12	0.01
2	76.10	0.00	0.90	0.64	38.25	0.00	54.87	0.00	48.49	0.00
3	99.15	0.00	1.76	0.62	40.74	0.00	71.89	0.00	59.25	0.00
4	132.53	0.00	3.29	0.51	46.21	0.00	92.96	0.00	60.21	0.00
5	177.49	0.00	5.45	0.36	46.24	0.00	98.49	0.00	66.01	0.00
6	207.78	0.00	6.02	0.42	47.98	0.00	108.47	0.00	68.82	0.00
7	234.29	0.00	6.39	0.50	52.32	0.00	119.01	0.00	69.16	0.00
8	264.40	0.00	7.72	0.46	56.87	0.00	127.54	0.00	69.18	0.00
9	293.85	0.00	10.39	0.32	64.57	0.00	136.43	0.00	86.96	0.00
10	332.39	0.00	15.80	0.11	74.82	0.00	148.65	0.00	88.92	0.00
11	363.26	0.00	20.36	0.04	89.91	0.00	162.64	0.00	88.93	0.00
12	394.20	0.00	22.55	0.03	97.59	0.00	171.92	0.00	88.93	0.00
13	418.79	0.00	27.12	0.01	111.65	0.00	183.08	0.00	95.94	0.00
14	446.49	0.00	30.76	0.01	121.71	0.00	194.03	0.00	95.98	0.00
15	472.29	0.00	31.33	0.01	127.59	0.00	201.47	0.00	99.40	0.00
16	497.53	0.00	31.70	0.01	133.41	0.00	206.70	0.00	99.97	0.00
17	509.40	0.00	32.48	0.01	134.01	0.00	206.91	0.00	100.66	0.00
18	521.13	0.00	32.75	0.02	135.46	0.00	108.14	0.00	100.85	0.00
19	545.15	0.00	34.05	0.02	140.66	0.00	211.16	0.00	100.32	0.00
20	565.59	0.00	35.21	0.02	145.23	0.00	212.98	0.00	101.32	0.00

Table 7: Estimated unconditional volatility according to the different models. Sample: Jan. 3, 2000 to Jul. 1, 2011.

	No regime	Regime 1	Regime 2	Regime 3	Regime 4
MEM	14.25				
AMEM	14.39				
MS(3)-AMEM		9.21	14.39	28.80	
MS(4)-AMEM		9.28	13.74	26.76	56.97
MS(3)-AMEM(d)		9.20	14.35	28.37	
TVTP-MS(3)-AMEM		8.69	14.37	55.99	
ST-AMEM		7.20	27.40		
HMEM	14.06				
MS(3)-HMEM		9.13	7.30	67.33	
HAR	13.25				

In the ST-AMEM case the two values are referred to the minimum and maximum unconditional volatility in the span considered.

Table 8: Coefficient estimates for the MS(3)–AMEM(d) specification (standard errors in parentheses; when the s.e. is omitted, the corresponding coefficient is constrained to take on the reported value). Likelihood-based criteria are reported in the lower portion. Sample: Jan. 3, 2000 to Jul. 1, 2011.

Mean Coefficients								
ω	k_2	k_3	δ					
1.91	0.47	3.51	71.75					
(0.17)	(0.31)	(0.65)	(14.20)					
α_1	α_2	α_3	β_1	β_2	β_3	γ_1	γ_2	γ_3
0.14	0.13	0.14	0.58	0.67	0.58	0.13	0.08	0.13
(0.02)	(0.02)		(0.04)	(0.04)		(0.01)	(0.01)	
Gamma coefficients						Likelihood-based criteria		
a_1	a_2	a_3				Log-Lik	AIC	BIC
17.71	23.62	11.0				-7594.89	5.32	5.36
(0.94)	(2.11)	(0.92)						
Transition probabilities								
p_{11}	p_{12}	p_{13}	p_{21}	p_{22}	p_{23}	p_{31}	p_{32}	p_{33}
0.99	0.01	0.00	0.01	0.96	0.03	0.01	0.06	0.93
(0.00)	(0.00)		(0.00)	(0.01)		(0.00)	(0.01)	

Table 9: Average widths and standard deviations of the 95% confidence intervals of the conditional expected volatility for the whole period and within each regime. Sample: Jan. 3, 2000 to Jul. 1, 2011.

	Whole		Regime 1		Regime 2		Regime 3		Regime 4	
	Mean	st.dev.	Mean	st.dev.	Mean	st.dev.	Mean	st.dev.	Mean	st.dev.
MEM	16.34	8.40								
AMEM	15.93	8.30								
MS(3)–AMEM	15.23	10.06	8.75	2.22	13.65	3.97	32.47	9.91		
MS(4)–AMEM	14.69	10.89	9.27	2.55	13.71	5.88	22.83	5.50	41.00	19.99
MS(3)–AMEM(d)	15.15	10.21	8.72	1.93	13.42	3.92	32.07	11.16		
TVTP–MS(3)–AMEM	15.63	10.43	11.33	5.04	15.38	6.75	33.99	13.70		
ST–AMEM	15.48	8.61								
HMEM	16.48	8.48								
MS(3)–HMEM	16.01	10.58	12.40	5.31	23.27	14.09	32.97	14.07		
HAR	19.30	0.00								

Table 10: Percentage of predictions falling above (overprediction) and below (underprediction) the 95% confidence bands with share composition by regime. Sample: Jan. 3, 2000 to Jul. 1, 2011.

	Overall			Overprediction % by Regime				Underprediction % by Regime			
	total	over	under	1	2	3	4	1	2	3	4
MEM	4.6	1.6	3.0								
AMEM	4.5	1.4	3.1								
MS(3)–AMEM	4.0	1.9	2.1	48.1	38.9	13.0		48.3	20.0	31.7	
MS(4)–AMEM	4.3	1.6	2.7	58.7	17.4	8.7	15.2	76.3	4.0	10.5	9.2
MS(3)–AMEM(d)	4.1	1.9	2.2	48.1	38.9	13.0		47.6	23.8	28.6	
TVTP–MS(3)–AMEM	2.5	1.9	0.6	70.9	20.0	9.1		5.6	22.2	72.2	
ST-AMEM	4.6	1.5	3.1								
HMEM	4.4	1.6	2.9								
MS(3)–HMEM	2.9	2.1	0.8	74.6	23.7	1.7		79.2	0.0	20.8	
HAR	5.0	2.4	2.6								

Table 11: In-sample (Jan. 3, 2000 to Jul. 1, 2011) and out-of-sample (Jul. 5, 2011 to Oct. 26, 2012; 1- and 2-steps ahead) performance in terms of Mean Square Error and Mean Absolute Error of the models estimated.

	In-sample		Out-of-sample			
	MSE	MAE	1-step-ahead		2-steps-ahead	
	MSE	MAE	MSE	MAE	MSE	MAE
MEM	29.10	3.25	29.61	3.72	38.27	3.98
AMEM	27.06	3.16	27.65	3.62	35.16	3.82
MS(3)–AMEM	25.57	2.91	27.00	3.62	31.05	3.70
MS(4)–AMEM	22.54	2.81	28.95	3.63	35.58	3.97
MS(3)–AMEM(d)	23.01	2.89	26.96	3.61	30.92	3.68
TVTP–MS(3)–AMEM	19.90	2.60	27.00	3.59	31.63	3.71
ST-AMEM	26.09	3.13	26.92	3.61	30.84	3.77
HMEM	29.38	3.28	29.91	3.76	36.16	3.93
MS(3)–HMEM	23.37	2.95	29.26	3.75	36.20	4.10
HAR	28.72	3.28	29.70	3.79	35.12	3.94

Table 12: Significance of the Diebold–Mariano statistics for comparison of in-sample predictions (Jan. 3, 2000 to Jul. 1, 2011) and out-of-sample (Jul. 5, 2011 to Oct. 26, 2012) 1- and 2-steps ahead forecasts.

	Absolute Errors										
	MEM	AMEM	MS(3)–AMEM	MS(4)–AMEM	MS(3)–AMEM(d)	MS(3)–AMEM	TVTP–MS(3)–AMEM	ST–AMEM	HMEM	MS(3)–HMEM	HAR
MEM	★	◆	●	★	★	●	★	◆	●	★	
AMEM			★	★	★		★			★	
MS(3)–AMEM				★	◆	●					★
MS(4)–AMEM			★								
MS(3)–AMEM(d)			●		●			●			
TVTP–MS(3)–AMEM				★			★				
ST–AMEM				★							★
HMEM	★	◆	◆	★	◆	●	★	◆			
MS(3)–HMEM			●	★	●	●	★	◆	●		
HAR	★	◆	◆	★	◆	●	★	◆			★
		Squared Errors									
MEM	AMEM	MS(3)–AMEM	MS(4)–AMEM	MS(3)–AMEM(d)	MS(3)–AMEM	TVTP–MS(3)–AMEM	ST–AMEM	HMEM	MS(3)–HMEM	HAR	
MEM	★	◆	★	★	◆	★	★	◆	★		
AMEM		★	★	★		★			★		
MS(3)–AMEM			★		★					★	
MS(4)–AMEM				★							
MS(3)–AMEM(d)					●			●			
TVTP–MS(3)–AMEM				★		★					
ST–AMEM			★							★	
HMEM	★	◆	★	★	◆	★	◆				
MS(3)–HMEM			★	★	●	★	◆	●			
HAR	◆	★	★	★	◆	★	◆			★	

The model by column is significantly better in forecasting performance than the model by row (at a 5% significance level) if a symbol is present: in-sample (★), 1-step (◆) and 2-steps (●) ahead out-of-sample.

Table 13: Models belonging to the best forecasting set based on MCS approach for individual in-sample predictions (Jan. 3, 2000 to Jul. 1, 2011) and out-of-sample (Jul. 5, 2011 to Oct. 26, 2012) 1- and 2-steps ahead forecasts.

Model	Absolute errors	Squared errors
MEM	◇	◇
AMEM	◇ ●	◇ ●
MS(3)–AMEM	◇ ●	◇ ●
MS(4)–AMEM	◇	◇ ●
MS(3)–AMEM(d)	◇ ●	◇ ●
TVTP–MS(3)–AMEM	★ ◇ ●	★ ◇ ●
ST–AMEM	◇ ●	◇ ●
HMEM		
MS(3)–HMEM	◇	◇
HAR		●

The presence of the symbols indicates that the model belongs (at a 95% confidence level) to the best set in terms of in-sample (★), 1-step (◇) and 2-step (●) ahead out-of-sample forecasting performance.

FIGURES

Figure 1: Realized kernel volatility of S&P500 index (Jan. 3, 2000 to Oct. 26, 2012); the vertical gray line separates the in- from the out-of-sample periods (the latter starting on Jul. 5, 2011).

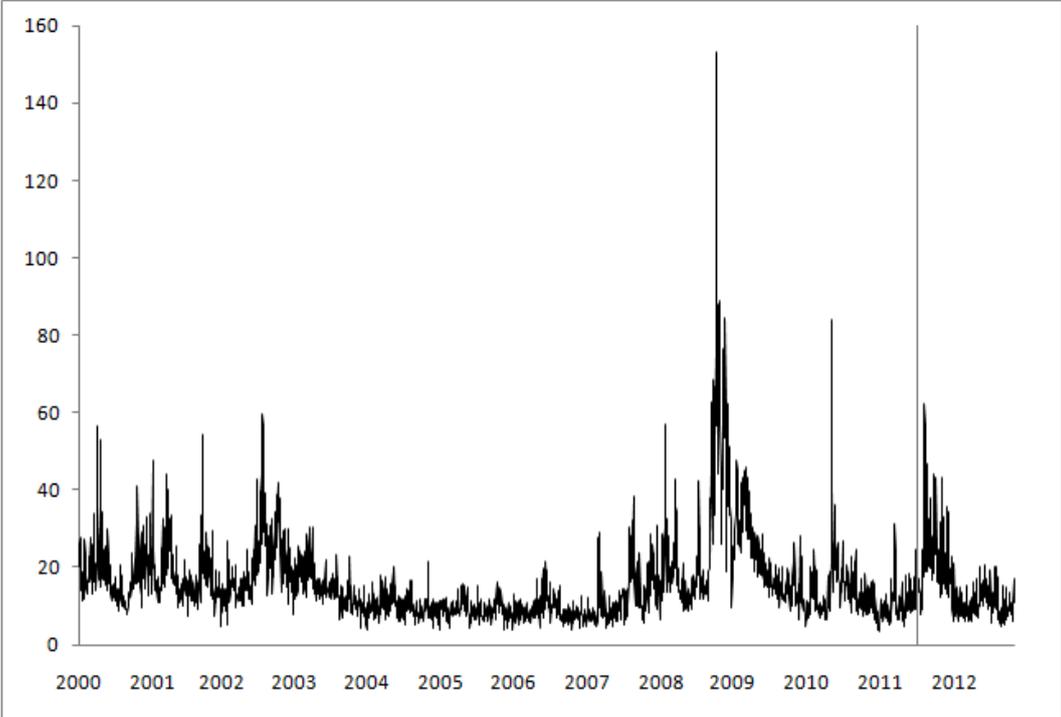
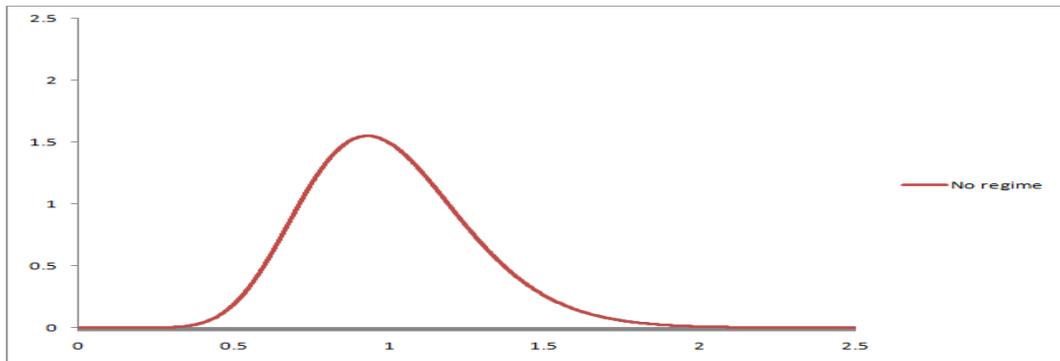
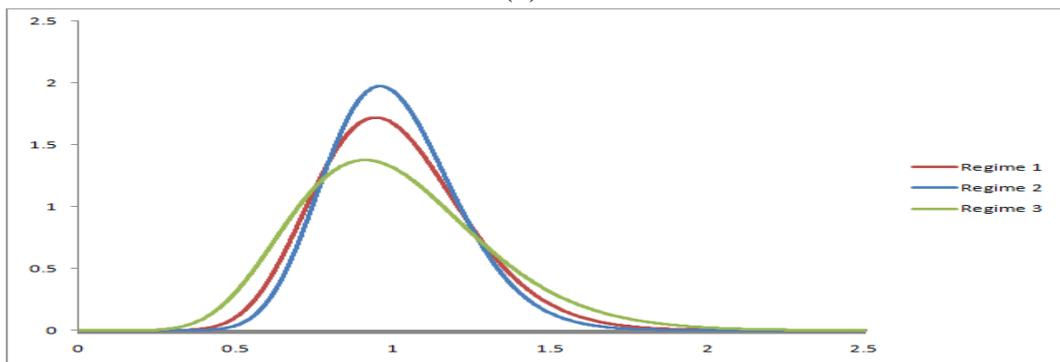


Figure 2: Gamma densities for the AMEM and by regime for the MS(3)-AMEM and MS(4)-AMEM.

A. AMEM



B. MS(3)-AMEM



C. MS(4)-AMEM

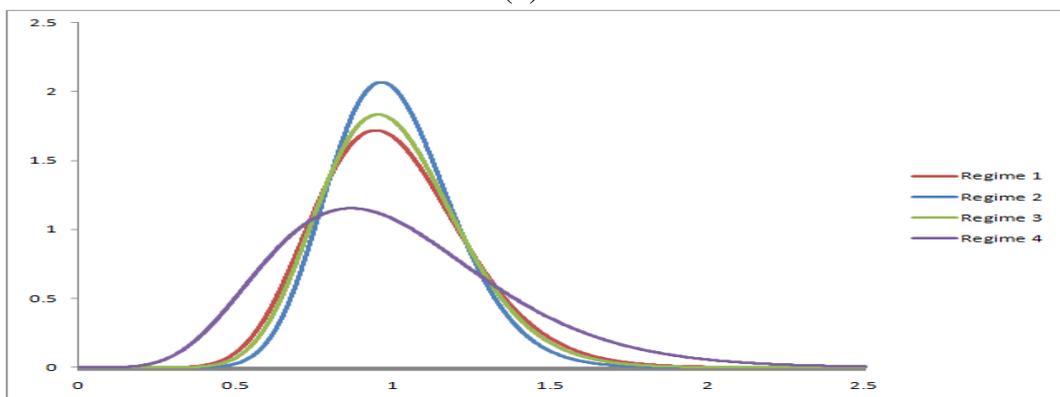
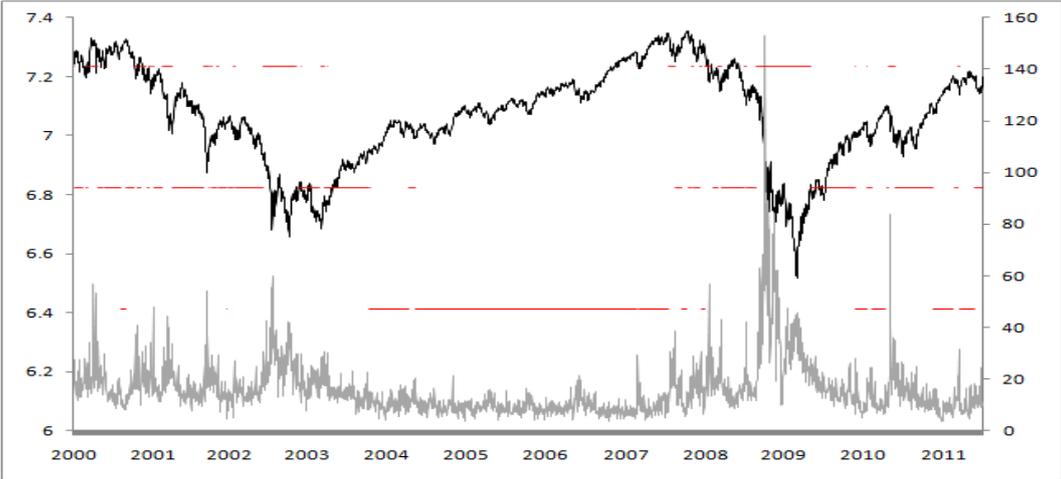


Figure 3: Inference on the regimes (horizontal dots), obtained as the mode of the smoothed regime probabilities using the MS(3)–AMEM (panel A) and MS(4)–AMEM (panel B). For reference, logs of the S&P500 index (black line) are reported against the left axis and its realized kernel volatility (gray line) against the right axis.

A. MS(3)–AMEM



B. MS(4)–AMEM

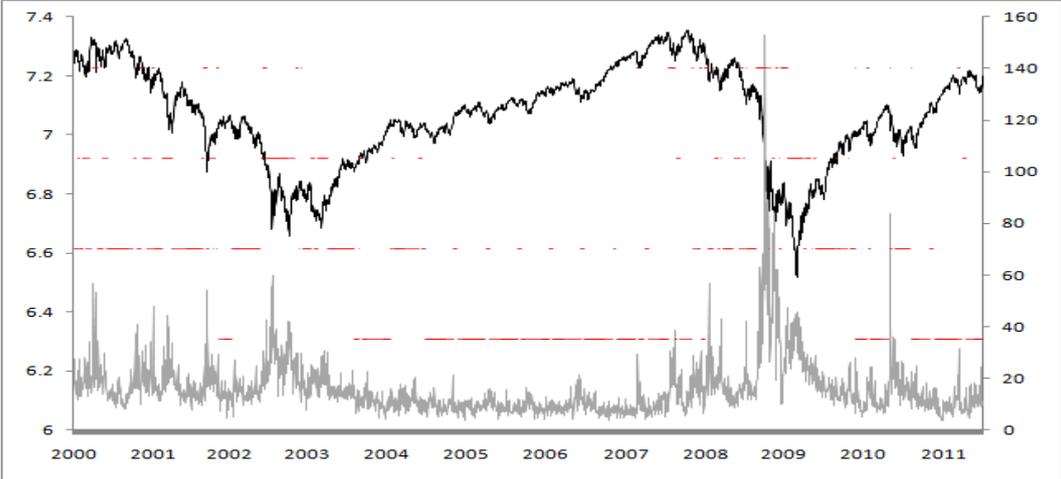


Figure 4: Inference on the regimes obtained as the mode of the smoothed regime probabilities using the MS(3)–AMEM and MS(4)–AMEM. The bars show as the observations assigned to the regime on the vertical axes for the 3-states case are assigned to the regimes in the 4-states case.

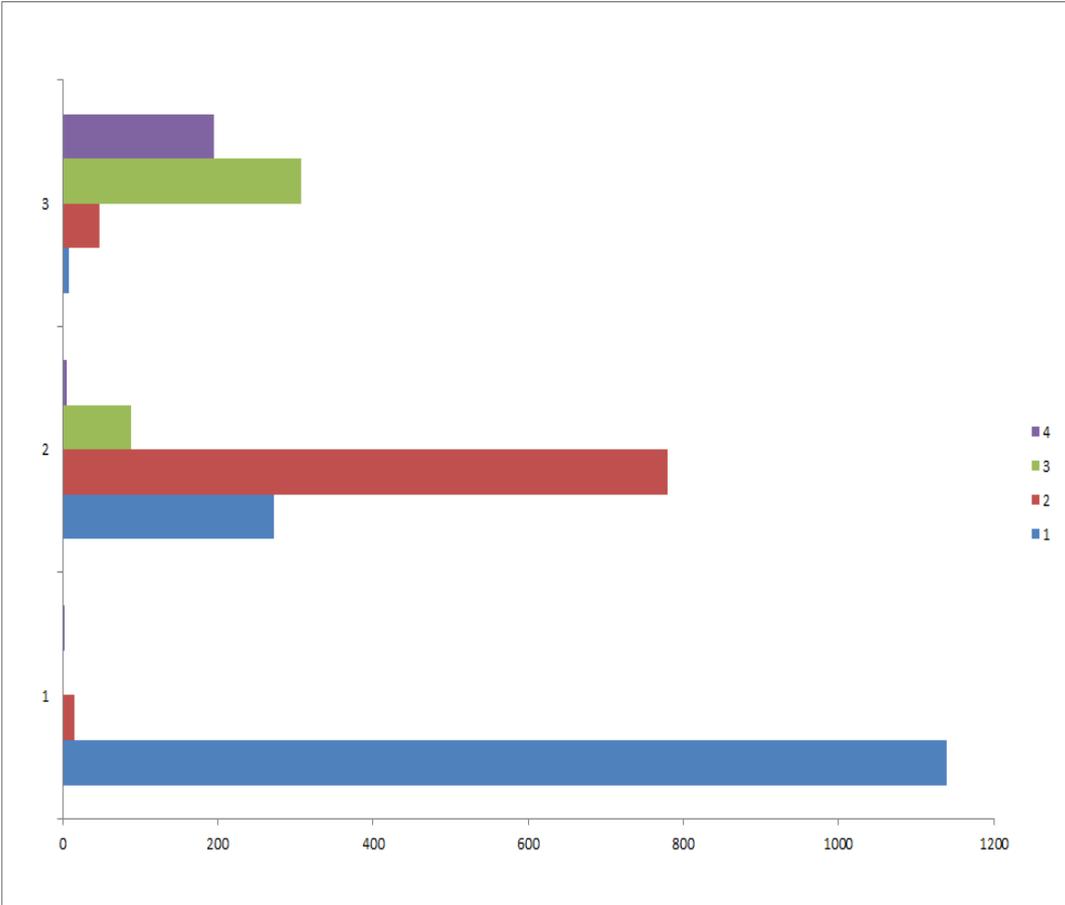
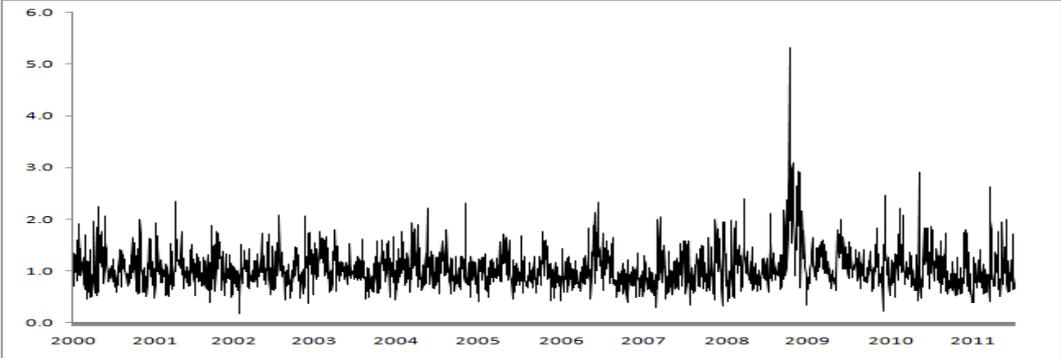
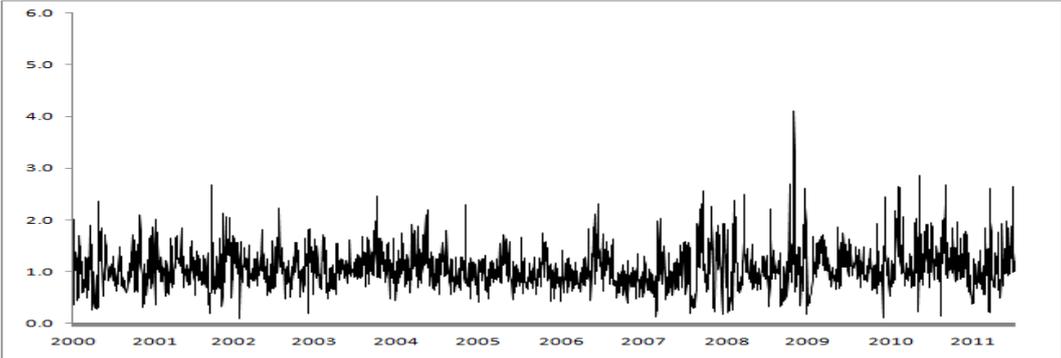


Figure 5: S&P500 realized kernel volatility series divided by the regime-specific average volatilities obtained from the MS(3)-AMEM, MS(4)-AMEM and MS(3)-AMEM(d).

A. MS(3)-AMEM



B. MS(4)-AMEM



C. MS(3)-AMEM(d)

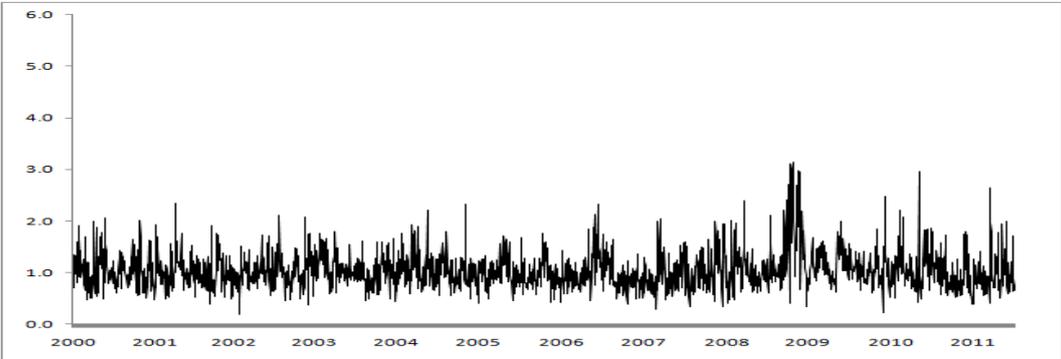
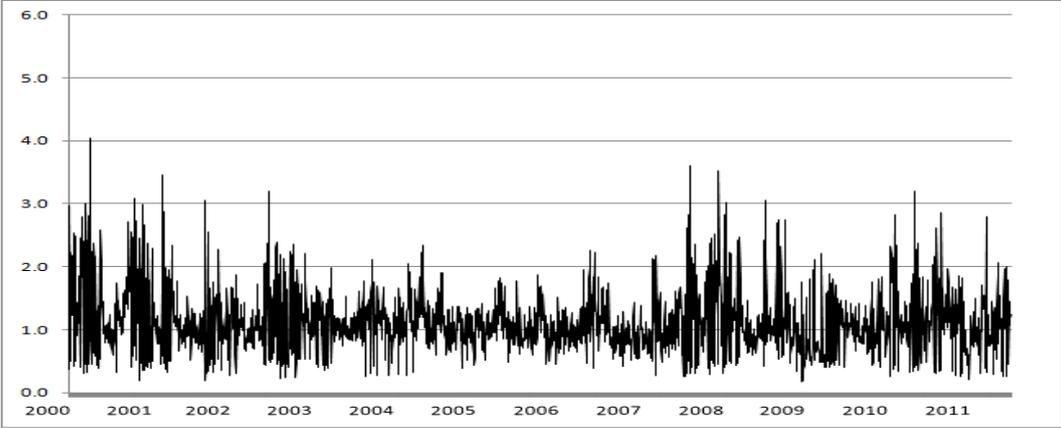
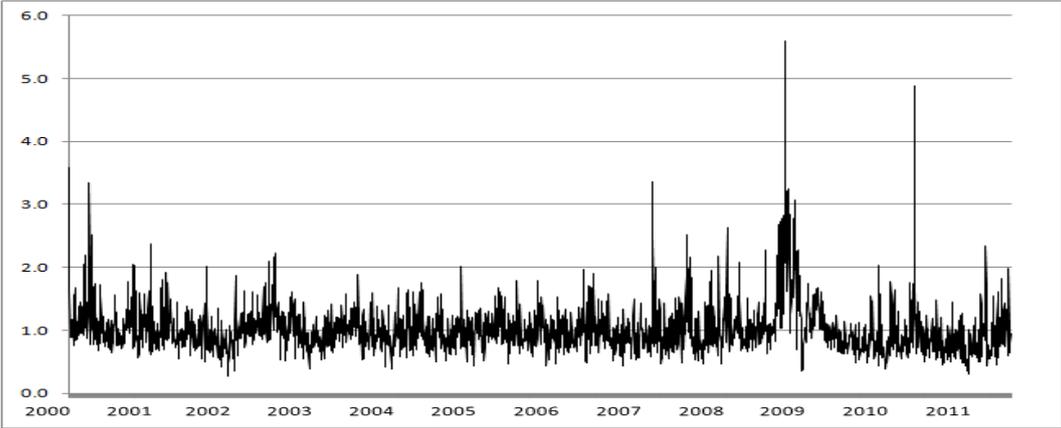


Figure 6: S&P500 realized kernel volatility series divided by the regime-specific average volatilities obtained from the TVTP-MS(3)-AMEM, SP-AMEM and MS(3)-HMEM.

A. TVTP-MS(3)-AMEM



B. SP-AMEM



C. MS(3)-HMEM

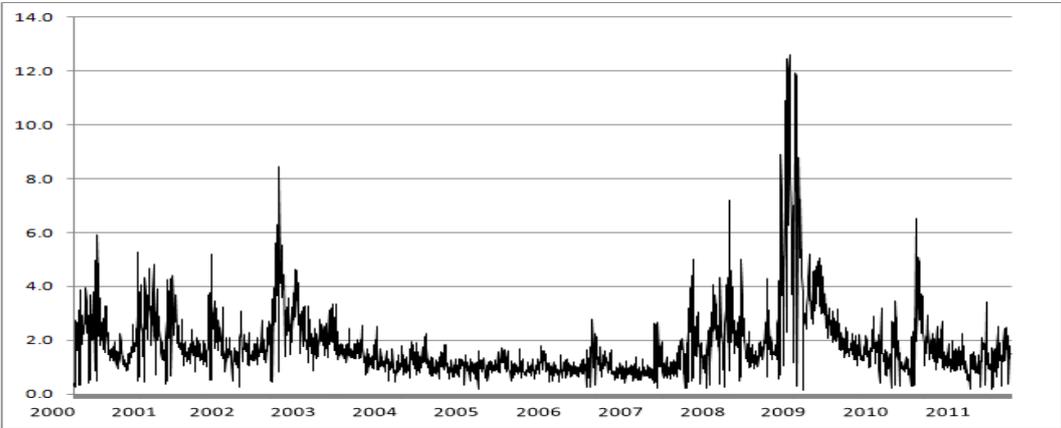
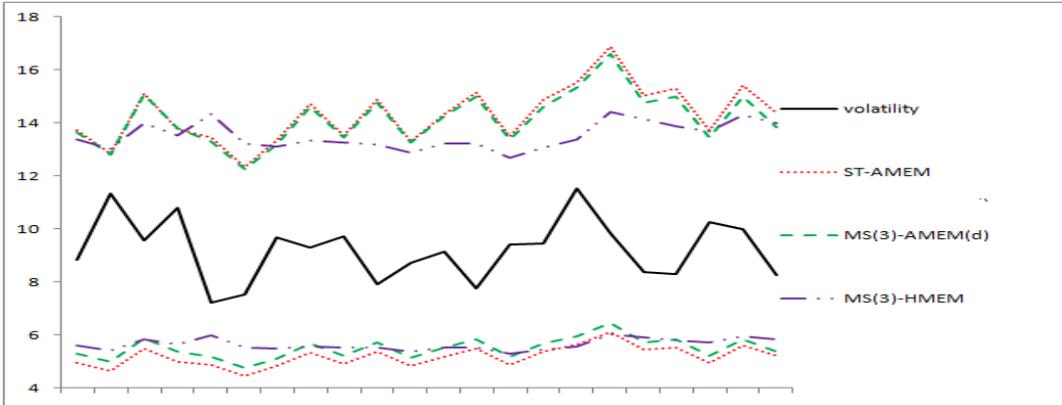
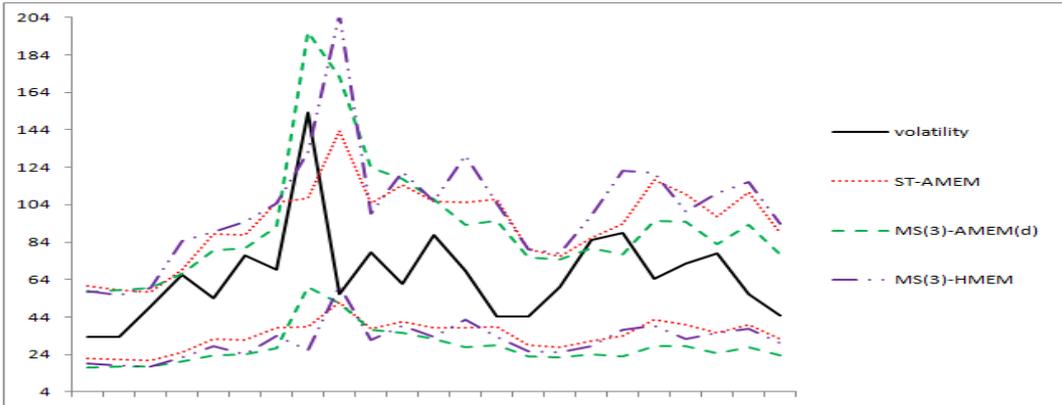


Figure 7: S&P500 realized kernel volatility series. 95% Prediction intervals derived from the AMEM, MS(3)-AMEM, MS(4)-AMEM and MS(3)-AMEM(d) for three representative market volatility spells.

A. March 2005



B. October 2008



C. March 2011

