UNIVERSITÀ DEGLI STUDI FIRENZE
DISIA
DIPARTIMENTO DI STATISTICA, INFORMATICA, APPLICAZIONI "GIUSEPPE PARENTI"

Combining Sharp and Smooth Transitions in Volatility Dynamics: a Fuzzy Regime Approach

Giampiero M. Gallo, Edoardo Otranto

DISIA WORKING PAPER
2017/05

Combining Sharp and Smooth Transitions in Volatility Dynamics: a Fuzzy Regime Approach

Giampiero M. Gallo

Dipartimento di Statistica, Informatica, Applicazioni "G. Parenti", Università di Firenze

Via G. B. Morgagni, 59 - 50134 Firenze, Italy (*gallog@disia.unifi.it*)

Edoardo Otranto

Dipartimento di Economia and CRENoS,

Università di Messina, Piazza Pugliatti, 1 - 98122 Messina, Italy (*eotranto@unime.it*)

Abstract

Volatility in financial markets is characterized by alternating persistent turmoil and quiet periods, but also by a slowly–varying average level. This slow moving component keeps open the question of whether some of its features are better represented as abrupt or smooth changes between local averages of volatility. We provide a new class of models with a set of parameters subject to abrupt changes in regime (Markov Switching – MS) and another set subject to smooth transition (ST) changes. These models capture the possibility that regimes may overlap with one another (*fuzzy*). The empirical application is carried out on the volatility of four US indices. It shows that the flexibility of the new model allows for a better overall performance over either MS or ST, and provides a Local Average Volatility measure as a parametric estimation of the low frequency component.

Keywords: Volatility modeling, Volatility forecasting, Multiplicative Error Model, Markov Switching, Smooth Transition, Common Trend

1 Introduction

The financial turmoil in the past ten-fifteen years has produced a rich empirical evidence about volatility behavior and evolution across markets. A short run dynamics around a slowly time-varying average level calls for an adequate econometric modelling: the derivation of measures of volatility based on ultra-high frequency data (Barndorff-Nielsen *et al.* (2008)) has allowed for the refinement of forecasting techniques relative to the GARCH approach to modeling conditional variances of returns. New models make direct use of realized volatility (Multiplicative Error Models MEM – Cipollini *et al.* (2013), Realized GARCH – Hansen *et al.* (2011), Heterogeneous Autoregressive HAR – Corsi (2009), among others); Andersen *et al.* (2007) provide an extension of the HAR model by inserting a proxy for jumps, whereas Caporin *et al.* (2015) add, in the same HAR model, a volatility jump term assumed to follow a compound Poisson process. A similar extension was adopted by Caporin *et al.* (2017) to accommodate jumps in the MEM.

Model diagnostics (e.g. residual ACFs) suggest that the low frequency component is not easily captured by standard models. The idea is that the long–range variance is not constant and that a flexible deterministic function of time can be fitted to that slowly moving average level of volatility. Explicit modeling within the GARCH approach was addressed by, for example, Engle and Rangel (2008) and Amado and Teräsvirta (2014). Within the MEM family, Brownlees and Gallo (2010) choose a basis of B-splines for such a function; Brownlees and Gallo (2011) refine its properties adopting shrinkage methods and penalized log–likelihood optimization. Such an approach aims at isolating the short–term dynamics around a flexible function which is individual series specific, and has very good in–sample properties. Barigozzi *et al.* (2014) use non parametric methods to extract a *common trend* from a large panel of N volatilities around which individual series have idiosyncratic MEM dynamics (SPvMEM).

Other authors consider this feature a result of the presence of long memory. A successful, simple, model is the HAR by Corsi (2009) which achieves a parsimonious representation with quasi long memory properties, aggregating past volatility into weekly and monthly variables used as regressors next to lagged volatility. Also in this case multivariate information can be added, as in Atak and Kapetanios (2013), who augment the HAR representation with common factors extracted from a large panel of N volatilities (F–HAR–RV model) to capture the influence of *common* movements in determining volatilities.

When modeling a slow moving component separately, the question remains as to whether some of its features could be better represented as abrupt or smooth changes between local averages of volatility. As an example, Gallo and Otranto (2015) extend the family of Asymmetric Multiplicative Error Models (AMEM) of Engle (2002) and Engle and Gallo (2006) to include Markovian changes in regime (abrupt changes) and smooth transitions to other regimes (gradual changes) with two separate models, called Markov Switching (MS)–AMEM and Smooth Transition (ST)– AMEM respectively. That paper builds on a large stream of literature that extends the study of volatility by introducing Markov switching features to the specification for the return conditional variance and, alternatively, Smooth Transition across different average levels; see, among others, Amado and Teräsvirta (2008), McAleer and Medeiros (2008), Maheu and McCurdy (2002).

The proper detection of various types of change turns out to be crucial for the fitting and the forecasting performances of the volatility models. A typical outcome is the trade–off between the in–sample and out–of–sample results: working with the realized kernel volatility (Barndorff-Nielsen *et al.*, 2008) of the S&P500 index, Gallo and Otranto (2015) analyze several alternative models within the AMEM and HAR classes, with a better overall diagnostics of the former over the latter. In particular, they detect a better in–sample fitting of the MS–AMEM and a superior out–of–sample performance of the ST–AMEM. This suggests the main motivation for this paper, a new model within the MEM class which borrows features from both models (abrupt changes from the MS; smooth changes from the ST), without being nested in or nesting either of them. Briefly put, it turns out that this model softens, so to speak, the abruptness of the MS changes and achieves such a trade–off when applied to the realized volatility of several US market indices.

The idea behind our new MEM is that abrupt and smooth changes need not be mutually exclusive in volatility dynamics: as such, we insert some model coefficients to be subject to discrete changes (between different regimes) and others as continuously time–varying as a function of a forcing variable. The former more directly regulate the average level of volatility, and the latter have more flexible features connected to the persistence of past volatility.

This specification has suitable properties and, rather than providing a constant average level of volatility within each regime as in the standard MS model, it gives a range of possible values for the *local average volatility* within each regime which may be overlapping across regimes. Although we lose the classical definition of regime linked to well distinct states (and therefore a clear–cut economic interpretability), the gain in performance can be attributed to a better (and more realistic) treatment of some periods which may not be clearly identified as belonging to a single regime. For this reason, we define this class of MEMs as the *Fuzzy Regime* AMEM with n states (FR(n)–AMEM).[1] As an important by–product of the model, we calculate the *Local Average Volatility* series, which is a novel parametric estimate of the low frequency component of volatility.

We verify the in–sample and out–of–sample performance of this new class of models, applying them to four US market index realized volatility series and comparing them with the MS– and ST–AMEMs, to verify that they are able to capture the better in–sample characteristics of the former and the out–of–sample ones of the latter. Moreover, we estimate also the HAR, the F–HAR–RV, the *classical* AMEM and the SPvMEM as useful benchmarks, showing that the new model improves significantly upon them.

The paper is organized as follows: in Section 2, we provide some preliminary empirical evidence on the data we use, as a support for what we discussed in the Introduction and lead the presentation of our theoretical contribution. In Section 3 we first introduce the notation for the models used, i.e. the four benchmark models, the MS–AMEM, the ST–AMEM and then we present the FR(n)–AMEM. The theoretical discussion of this new model includes the concept of Local Average Volatility which is suitable to capture a low frequency component of volatility. Section 4 is devoted to the presentation of the empirical results of the model. We first pay some special attention to the comparison of

[1] This model has some similarities with the Flexible MS model proposed by Otranto (2016), where coefficients may be time–varying within each regime, but states cannot overlap.

the information derived from the FR(2)–AMEM with those derived from MS(3)–AMEM and ST–AMEM for the S&P500. We then extend the number of models and the number of US realized kernel volatility time series including next to the S&P500, also the Russell 2000, the Dow Jones 30, and the NASDAQ 100. In– and out–of–sample performances are evaluated on the basis of single loss functions and a synthesis thereof to evaluate contemporaneously the in–sample and out–of–sample performance. A separate subsection (4.2) contains a discussion of the estimation of the LAV on our data and the comparison with the *common trend* estimated in the SPvMEM. Some final remarks conclude the paper.

2 The Empirical Motivation

Empirical applications on volatility dynamics abound, both when volatility is measured and modeled with GARCH models (Bollerslev, 1986) and when it is measured as one of the flavors of realized volatility (cf. Andersen *et al.* (2006) for a survey and references) and then modeled to derive forecasts (we will detail modeling approaches and provide some benchmark examples in what follows). In this paper we concentrate on the empirical behavior of volatility on the same geographical market (the US) measured on different indices whose components represent a wide coverage of individual stocks, in term of company size, economic sector, liquidity, etc. More specifically, we chose four main US market indices: S&P500, Russell 2000, Dow Jones 30, NASDAQ 100. As per the volatility variable, our choice goes to the realized kernel volatility, an estimator of the integrated volatility of a diffusion process, proposed by Barndorff-Nielsen *et al.* (2008), possessing the important property of robustness to market microstructure noise. The data relative to the series analyzed in this work are taken from the Oxford-Man Institute Realized Library version 0.2 (Heber *et al.*, 2009), and are annualized, i.e. we consider the square root of the realized kernel variance multiplied by $100\sqrt{252}$. We cover the period 2 January 2004–5 May 2015 with daily observations; in what follows we reserve the data until 31 December 2013 for estimation purposes, leaving the rest available for an out–of–sample forecasting evaluation.

In Figure 1 the four series are depicted: we notice that they all follow a similar (common?) low frequency time–varying component (a slowly evolving trend) and they exhibit a similar behavior around that, with peaks tending to occur on the same dates, in particular the bursts of volatility starting in September 2008. The last part of the period is characterized by a more stable local average with short term persistence but no high peaks.

This observed behavior is in line with the presence of Markov Switching behavior. To that end, we run the Otranto and Gallo (2002) nonparametric Bayesian approach to identify the number of switching regimes in the data. The results obtained under different choices for the A hyperparameter regulating the prior probabilities on the number of regimes are reported in Table 1. For all index volatilities, there is no clear evidence of a fixed number of regimes: for most series, results point toward 3 or 4 regimes, with somewhat nonnegligible posterior probabilities on five states for S&P 500 and DJ 30.

Building on previous results, Markov switching models are just one possibility to characterize the dynamics of these series; the evidence of the Bayesian approach is consistent

with other forms of capturing the presence of *time–varying average features* in the series: not only Markov Switching behavior, in fact, but also Smooth Transition dynamics, or ways to characterize the low frequency component as a flexible deterministic function of time, and so on.

We now turn to characterize the new model which is motivated by the need to combine the structure of switching regimes with the smooth transition behavior in order to achieve a trade–off between in–sample fitting capabilities and out–of–sample forecasting performance.

3 Combining Switching Regimes and Smooth Transitions

The Multiplicative Error Model (MEM) has many desirable properties for modelling and forecasting realized volatility. Among the reasons for a satisfactory performance are the capability of modeling volatility directly without resorting to log–trasformations, the possibility of including sign–related asymmetric effects, Markovian or smooth transition dynamics (Gallo and Otranto, 2015), presence of jumps (Caporin *et al.*, 2015), and the generally satisfactory residual diagnostics (Engle and Gallo, 2006). The model we propose is a combination of the MS–AMEM and the ST–AMEM proposed by Gallo and Otranto (2015), the main features of which are recalled hereafter for notational and comparison purposes.

3.1 Background Models

3.1.1 The Benchmark Models

Let y_t be the realized volatility of a certain asset or index at time t ($t = 1, \ldots, T$), μ_t a latent factor representing its conditional mean, ε_t a random disturbance (independent of μ_t) and D_t a dummy variable which is 1 when the return of the underlying asset at time t is negative. Moreover, we assume that the returns have zero median and that their sign and y_t are independent. We introduce here the models to be used as benchmarks:

- **HAR** (Corsi, 2009):

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t \\ \mu_t &= \omega + \alpha_D y_{t-1} + \alpha_W \bar{y}_{t-1}^{(5)} + \alpha_M \bar{y}_{t-1}^{(22)}, \end{aligned}$$

  where the three independent variables of the second equation express past volatilities aggregated at daily, weekly and monthly frequency, respectively. This is a model often used as a single benchmark, for its properties of quasi–long memory already mentioned.

- **F–HAR–RV** (Atak and Kapetanios, 2013):

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t \\ \mu_t &= \omega + \alpha_D y_{t-1} + \alpha_W \bar{y}_{t-1}^{(5)} + \alpha_M \bar{y}_{t-1}^{(22)} + \beta_D f_{t-1} + \beta_W \bar{f}_{t-1}^{(5)} + \beta_M \bar{f}_{t-1}^{(22)}, \end{aligned}$$

where f_t is the first Principal Component extracted from N series in a multivariate setting. In the context here of $N = 4$ index volatilities, it explains 97.3% of total variation; $\bar{f}_t^{(5)}$ and $\bar{f}_t^{(22)}$, are obtained by weekly, respectively, monthly aggregation of f_t. This is an interesting model in that it makes use of multivariate information from $N = 4$ volatilities together in the form of a common (lagged) factor conveniently aggregated to mirror the various HAR horizons.

- **AMEM** (Engle and Gallo, 2006):

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

$$\mu_t = \omega + \alpha y_{t-1} + \beta \mu_{t-1} + \gamma D_{t-1} y_{t-1},$$

The first equation shows the typical MEM structure suggested by Engle (2002), obtained as the product of two positive unobservable factors (several multiplicative components are possible Brownlees *et al.* (2011)). As in Engle and Gallo (2006), the choice of the Gamma distribution for the innovations depends only on one scale parameter. The conditional expectation has a GARCH–type dynamics with an asymmetric term. This model has a constant unconditional expected volatility.

- **SPvMEM** (Barigozzi *et al.*, 2014):

$$\begin{aligned} y_{i,t} &= b_i \phi_t \mu_{i,t} \varepsilon_{i,t}, \qquad \varepsilon_{i,t} \sim \text{Gamma}(a_i, 1/a_i) \text{ for each } t, \\ \mu_{i,t} &= \omega_i + \alpha_i \tfrac{y_{i,t-1}}{\phi_t} + \beta_i \mu_{i,t-1} + \gamma_i D_{i,t-1} \tfrac{y_{i,t-1}}{\phi_t}, \\ b_i &= \tfrac{\omega_i}{1-\alpha_i-\beta_i-\gamma_i/2} \end{aligned} \tag{3.1}$$

where each individual realized volatility is assumed to depend on a multiplicative common trend component, ϕ_t, estimated across the four volatilities as a smooth nonparametric function of time. This model exploits the information from $N = 4$ volatilities to derive a common low frequency component around which idiosyncratic (regime–free) dynamics is captured by index–specific AMEMs.[2]

3.1.2 MS–AMEM

Let s_t be a discrete latent variable ranging in $[1, \ldots, n]$, representing the regime at time t. The MS(n)–AMEM, where n indicates the number of regimes, is given by:

$$\begin{aligned} y_t &= \mu_{t,s_t} \varepsilon_t, \qquad \varepsilon_t | s_t \sim \text{Gamma}(a_{s_t}, 1/a_{s_t}) \text{ for each } t, \\ \mu_{t,s_t} &= \omega + \textstyle\sum_{i=1}^{n} k_i I_{s_t} + \alpha_{s_t} y_{t-1} + \beta_{s_t} \mu_{t-1,s_{t-1}} + \gamma_{s_t} D_{t-1} y_{t-1}. \end{aligned} \tag{3.2}$$

Here the scale parameter changes with the regime (a_{s_t}); thus innovations have a unit mean, no matter what the regime is, and a regime dependent variance, i.e. $1/a_{s_t}$. The second

[2] In the empirical estimation of this model, we used the GAUSS program coded by David Veredas, supplied by the authors of the paper, with our modification to allow for asymmetric effects as in the second equation of (3.1).

equation represents the dynamics of the mean, which follows a sort of GARCH(1,1)–type dynamics, with the intercepts $\omega + \sum_{i=1}^{n} k_i I_{s_t}$ (where I_{s_t} is the indicator of the state; $k_1 = 0$, $k_i \geq 0$ for $i > 1$), increasing with the state s_t. The coefficient ω is strictly positive and the switching coefficients α_{s_t}, β_{s_t} and γ_{s_t} are nonnegative for each s_t, to entail positiveness of μ_{t,s_t}. The dynamics of the latent state variable s_t is driven by a stationary, irreducible and aperiodic Markov chain, such that:

$$Pr(s_t = j | s_{t-1} = i, s_{t-2}, \dots) = Pr(s_t = j | s_{t-1} = i) = p_{ij}. \tag{3.3}$$

Considering the strict relationship between the MEM and the GARCH models (Engle, 2002), we can easily derive the stationarity and ergodicity properties of the MS–AMEM from the corresponding properties of the MS–GARCH, suggested by Francq *et al.* (2001), based on the Lyapunov exponent associated with the matrix of the coefficients of the vectorial regression GARCH representation. Extending the results of that work, the unique condition for (3.2) to be stationary and ergodic is given by:

$$\sum_{s_t=1}^{n} \pi_{s_t} Elog\left[(\alpha_{s_t} + \gamma_{s_t} D_t)\varepsilon_t + \beta_{s_t}\right] < 0 \tag{3.4}$$

where π_{s_t} is the ergodic probability of state s_t.[3] The condition is similar to the corresponding condition detected by Nelson (1990) for the GARCH(1,1) case. By Jensen's inequality, $Elog(X) \leq logE(X)$, the condition:

$$\sum_{s_t=1}^{n} \pi_{s_t} (\alpha_{s_t} + \gamma_{s_t}/2 + \beta_{s_t}) < 1 \tag{3.5}$$

satisfies (3.4). Note that $(\alpha_{s_t} + \gamma_{s_t}/2 + \beta_{s_t} < 1)$ for each s_t is a sufficient condition for (3.4), but turns out to be too strong, given that we could think of values of π_{s_t} satisfying the overall constraint even if $(\alpha_{s_t} + \gamma_{s_t}/2 + \beta_{s_t}) \geq 1$ for some s_t.

3.1.3 ST–AMEM

Gallo and Otranto (2015) define also the ST–AMEM as:

$$\begin{aligned} y_t &= \mu_t \varepsilon_t, \qquad \varepsilon_t \sim \text{Gamma}(a, 1/a) \text{ for each } t \\ \mu_t &= \omega + \alpha_t y_{t-1} + \beta_t \mu_{t-1} + \gamma_t D_{t-1} y_{t-1} \end{aligned} \tag{3.6}$$

The coefficients α_t, β_t, γ_t are smooth transition parameters,[4] which follow a simple linear dynamics depending on a smooth function $0 \leq g_t \leq 1$:

$$\begin{aligned} \alpha_t &= \alpha_0 + \alpha_* g_t \\ \beta_t &= \beta_0 - \beta_* g_t \\ \gamma_t &= \gamma_0 + \gamma_* g_t \\ g_t &= (1 + \exp(-\delta(x_t - c)))^{-1}. \end{aligned} \tag{3.7}$$

[3] See Hamilton (1994), ch. 22.

[4] In Gallo and Otranto (2015) γ_t was constant.

where x_t is a volatility–related forcing variable driving the smooth movements, with $\delta > 0$ and c unknown coefficients. Extreme values of δ make the g_t function very close to a step function and thus the smooth transition more akin to an abrupt change. The following constraints can be imposed to ensure that μ_t be positive:

1. $\omega > 0$;
2. nonnegativity of the coefficients α_0, α_*, β_0, β_*, γ_0, γ_*;
3. $\beta_0 \geq \beta_*$.

As a consequence, the time-varying coefficients α_t and γ_t increase with g_t, allowing for a stronger reaction to the lagged volatility term, while β_t decreases with g_t, inducing a faster reversal to the mean when volatility is high.

Stationarity conditions can be analyzed for the limiting behavior of the smooth transition function g_t, as in ST–GARCH processes (Gonzalez-Rivera, 1998). When $x_t \to -\infty$, the smooth function approaches zero and, following Bollerslev (1986), the sufficient stationarity condition is:

$$\alpha_0 + \beta_0 + \gamma_0/2 < 1 \tag{3.8}$$

whereas, when $x_t \to \infty$, it is:

$$\alpha_0 + \alpha_* + \beta_0 - \beta_* + \gamma_0/2 + \gamma_*/2 < 1 \tag{3.9}$$

Similarly, it is possible to derive the stationarity conditions for any value $\tilde{g}_t$ of the smooth transition function, as:

$$\alpha_0 + \beta_0 + \gamma_0/2 + \tilde{g}_t(\alpha_* - \beta_* + \gamma_*/2) < 1 \tag{3.10}$$

In the presence of the parameter constraints to ensure positiveness of μ_t, conditions (3.8)-(3.9) imply (3.10) for any $0 < \tilde{g}_t < 1$.

3.2 The Fuzzy Regime Model

The FR(n)–AMEM combines the characteristics of models (3.2) and (3.6) and it is defined as:

$$\begin{aligned} y_t &= \mu_{t,s_t}\varepsilon_t, \qquad \varepsilon_t|s_t \sim \text{Gamma}(a, 1/a) \text{ for each } t, \\ \mu_{t,s_t} &= \omega + \textstyle\sum_{i=1}^{n} k_i I_{s_t} + \alpha_t y_{t-1} + \beta_t \mu_{t-1,s_{t-1}} + \gamma_t D_{t-1} y_{t-1} \end{aligned} \tag{3.11}$$

As in the ST–AMEM, the Gamma distribution of the disturbances depends only on the constant parameter a, whereas the second equation represents the new specification of the dynamics of the mean, which is composed by one switching and three smooth transition coefficients. The switching coefficient is the intercept (better suited to capture the abrupt changes in the average level of volatility after a sizeable shock) and the coefficients α_t, β_t, γ_t are smooth transition parameters, following the (3.7) specification. In our applications we will consider the lagged VIX (Whaley, 2009) as forcing variable: it is a widely followed index derived as an average of implied volatilities of one–month at–the–money

call and put options on the S&P 500. As such, it is a market based measure of overall market volatility and it lends itself to the needed neutral role in this model.

Models MS(n)–AMEM and ST–AMEM are not nested in model (3.11). In fact for the MS–AMEM we hypothesize that $\alpha_t = \alpha_{s_t}$, $\beta_t = \beta_{s_t}$, $\gamma_t = \gamma_{s_t}$, which do not depend on g_t, so that c and δ are nuisance parameters. The ST–AMEM is not nested in the FR–AMEM because it considers a constant intercept or an intercept depending on g_t. By contrast, when all the coefficients are constant, we obtain the simple AMEM of Engle and Gallo (2006).

Note that the only switching parameters of the FR(n)–AMEM are the intercept terms $\omega + \sum_{i=1}^{n} k_i I_{s_t}$, which are never involved in the stationarity conditions (i.e. the Markov Switching features of the FR(n)–AMEM do not bear on stationarity); thus, the stationarity conditions of the FR–AMEM derive from the stationarity conditions of the ST–AMEM, given in (3.8)-(3.9).

Estimation of model (3.11) can be performed by Maximum Likelihood, deriving the likelihood function by the iterative Hamilton (1990) filter. As in MS–GARCH models, a typical path–dependence problem arises, due to the non–observability of the volatility and the dependence of the unobservable state on all the past values. As in Gallo and Otranto (2015) we adopt the Kim (1994) approximation to solve this problem. Briefly put, after step t of the Hamilton filter we obtain n^2 possible values of $\mu_{t,s_t,s_{t-1}}$; they are collapsed into n values μ_{t,s_t}, averaging for each one the n terms in correspondence with s_{t-1} with weights proportional to the conditional probabilities of (s_t, s_{t-1}); these n terms are used as an input for step $t+1$. For details we refer to Kim (1994). This solution seems to provide a good approximation, as shown in Gallo and Otranto (2015) by a Monte Carlo analysis for the MS–AMEM case. The work of Engle and Gallo (2006), extended to the MS case with the results of Kim (1994), provides a Quasi Maximum Likelihood interpretation of such estimator.

The Kim (1994) smoother provides an estimate of the probability of the state for each time t, conditional on the full information set. As in the other MS models, this probability can be used to make inference on the regime at each time t, assigning the corresponding observation to the state with highest smoothed probability. In the MS case, especially with more than two states, adopting this procedure we may encounter "close–call" cases (attributed to a regime by a narrow margin). The extension to the FR case allows for a formal recognition of these borderline attributions, with extra information provided to the economic analysis.

3.3 The Local Average Volatility – LAV

In a MS model the inference on whether an observation belongs to a regime is carried out as a by–product of the Hamilton filter, identifying the most likely regime for each time t (cf. Ch. 22 in Hamilton (1994)). In the FR-AMEM, inference is not so clear–cut, in that a peculiarity of this model is that the assignment of one observation to one regime or another depends on the value of the smooth transition function g_t. This is what modifies the meaning of the regime classification in this context, possibly limiting the economic interpretation of the results. To substantiate this feature, let us define the *local average*

volatility (LAV hereafter) of the series y_t, under stationarity:

$$\bar{\sigma}_{t,s_t,g_t} = E(\mu_{t,s_t}|s_t, s_{t-1}, g_t)$$

which represents an expected level of volatility conditional on the sequence s_{t-1}, s_t of adjacent regimes, and on the value of the transition function g_t. Note that, when all the coefficients are constant (as in the AMEM case), this indicator corresponds to the unconditional volatility. To illustrate its properties, let us consider, for the sake of simplicity, the 2-state case, since the results are easily extended to the general n-state model. LAV is derived from the second equation of (3.11) and evolves according to the recursion:

$$\bar{\sigma}_{t,s_t,g_t} = \omega + k_2 I_2 + (\alpha_t + \beta_t + \gamma_t/2)\bar{\sigma}_{t-1,s_{t-1},g_{t-1}}, \tag{3.12}$$

with $\bar{\sigma}_{0,s_0,g_0}$ set at a suitable initial condition. Adding to (3.9) the additional constraint $(\alpha_* + \gamma_*/2) \geq \beta_*$, $\bar{\sigma}_{t,s_t,g_t}$ will not decrease when g_t increases and will not decrease with the regime ($\bar{\sigma}_{t,2,g_t} \geq \bar{\sigma}_{t,1,g_t}$ because $k_2 \geq 0$). Also, when $s_t = s_{t-1}$ and $g_t = g_{t-1}$, equation (3.12) can be written in a manner similar to the classical unconditional volatility expression:

$$\bar{\sigma}_{t,s_t,g_t} = \frac{\omega + k_2 I_2}{1 - \alpha_t - \beta_t - \gamma_t/2}. \tag{3.13}$$

This implies that the minimum (respectively, maximum) of $\bar{\sigma}_{t,s_t,g_t}$ is reached in correspondence of both g_t and g_{t-1} equal to 0 (respectively, 1). As a consequence, we can derive the minimum and the maximum of the LAV in each regime s_t; calling $\bar{\sigma}^l_{s_t}$, respectively, $\bar{\sigma}^h_{s_t}$ such extremes, we have:

$$\begin{aligned}
&\bar{\sigma}^l_1 \equiv \min(\bar{\sigma}_{t,1,g_t}) = E(\mu_t|s_t = 1, s_{t-1} = 1, g_t = 0) = \tfrac{\omega}{1-\alpha_0-\beta_0-\gamma_0/2} \\
&\bar{\sigma}^h_2 \equiv \max(\bar{\sigma}_{t,2,g_t}) = E(\mu_t|s_t = 2, s_{t-1} = 2, g_t = 1) = \tfrac{\omega+k_2}{1-\alpha_0-\alpha_*-\beta_0+\beta_*-(\gamma_0+\gamma_*)/2} \\
&\bar{\sigma}^h_1 \equiv \max(\bar{\sigma}_{t,1,g_t}) = E(\mu_t|s_t = 1, s_{t-1} = 2, g_t = 1) = \omega + (\alpha_0 + \alpha_* + \beta_0 - \beta_* + \tfrac{\gamma_0+\gamma_*}{2})\bar{\sigma}^h_2 \\
&\bar{\sigma}^l_2 = min(\bar{\sigma}_{t,2,g_t}) = E(\mu_t|s_t = 2, s_{t-1} = 1, g_t = 0) = \omega + k_2 + (\alpha_0 + \beta_0 + \tfrac{\gamma_0}{2})\bar{\sigma}^l_1
\end{aligned} \tag{3.14}$$

The two ranges of the LAVs in different regimes may overlap, providing a sort of gray area in–between which we call *fuzzy zone*. In practice, the model (3.11) provides an inference on the regime as any other MS model, but the levels of the LAV are not clearly separated across regimes. Formally, if $\bar{\sigma}^h_1 > \bar{\sigma}^l_2$, then the set $[\bar{\sigma}^l_2; \bar{\sigma}^h_1]$ is the overlapping area and the local average volatilities can be classified (borrowing the set theory terminology, see Zadeh (1965)) as follows:

- crisp zone 1: local average volatilities ranging in $[\bar{\sigma}^l_1; \bar{\sigma}^l_2]$;
- fuzzy zone: local average volatilities ranging in $[\bar{\sigma}^l_2; \bar{\sigma}^h_1]$;
- crisp zone 2: local average volatilities ranging in $[\bar{\sigma}^h_1; \bar{\sigma}^h_2]$;

In Figure 2, we illustrate a graphical example to better visualize the fuzzy and crisp zones; we consider a simple FR(2)–AMEM with $\beta_t = \gamma_t = 0$); the other coefficients are $\omega = 1$, $k_2 = 3$, $\alpha_0 = 0.5$, $\alpha_* = 0.4$. The Figure shows the behavior of the LAV in the two regimes in correspondence of g_t. The two curves do not intersect but, in the fuzzy zone (gray area), they can assume the same value in correspondence of two different values of g_t; in other words, a certain level of the unconditional volatility can be reached in state 1 in correspondence with a certain g_t or in state 2 in correspondence with a lower g_t. In practice, the inference about observation t to be assigned to regime 1 or 2 depends on the value of g_t, when in a fuzzy zone, and the average value of the observations attributed to the regime 1–fuzzy could be higher than the corresponding average for the regime 2–fuzzy.

The extension to the $n-$state case is simple; for the state j, the range of the LAV is given by:

$$\begin{aligned} \bar{\sigma}_j^l &= \omega + \textstyle\sum_{i=1}^{j} k_i + (\alpha_0 + \beta_0 + \gamma_0/2)\bar{\sigma}_1^l \\ \bar{\sigma}_j^h &= \omega + \textstyle\sum_{i=1}^{j} k_i + (\alpha_0 + \alpha_* + \beta_0 - \beta_* + \gamma_0/2 + \gamma_*/2)\bar{\sigma}_n^h \end{aligned} \tag{3.15}$$

From (3.15), it is evident that the differences between the lowest (highest) LAV of two states is due exclusively to the changes in the intercept, so it is likely that the fuzzy zones are very large with respect to the crisp zones. This fact could imply some difficulty in the interpretation of regimes when the number of states is more than two, since several regimes will have overlapping areas. What we want to stress is that we gain in terms of flexibility of the model and goodness of fit, as empirically shown in what follows.

For reference purposes, let us introduce a corresponding formula for the MS–AMEM, given by:

$$\begin{aligned} \bar{\sigma}_{t,s_t} &= \omega + \textstyle\sum_{i=1}^{n} k_i I_{s_t} + (\alpha_{s_t} + \beta_{s_t} + \gamma_{s_t}/2)\bar{\sigma}_{t-1,s_{t-1}} \quad \text{if } s_t \neq s_{t-1} \\ \bar{\sigma}_{t,s_t} &= \tfrac{\omega + \sum_{i=1}^{n} k_i I_{s_t}}{1-\alpha_{s_t}-\beta_{s_t}-\gamma_{s_t}/2} \qquad\qquad\qquad\qquad \text{if } s_t = s_{t-1}, \end{aligned} \tag{3.16}$$

whereas the one for the ST–AMEM is:

$$\bar{\sigma}_{t,g_t} = \omega + (\alpha_t + \beta_t + \gamma_t/2)\bar{\sigma}_{t-1,g_{t-1}}. \tag{3.17}$$

We can use the strong relationship between the regime and g_t to assign the observations to a crisp or a fuzzy zone. Let us consider again the two-state case (the extension to a generic $n-$state case is trivial but notationally cumbersome); the conditional mean (second equation of (3.11)) is:

$$\mu_{s_t,t} = \omega + k_2 I(s_t) + \alpha_0 y_{t-1} + \beta_0 \mu_{s_{t-1},t-1} + \gamma_0 D_{t-1} y_{t-1} + g_t(\alpha_* y_{t-1} - \beta_* \mu_{s_{t-1},t-1} + \gamma_* D_{t-1} y_{t-1})$$

If the observation at time t falls in a fuzzy zone, then there will exist a value of g_t (call it g_t^*) such that:

$$g_t^* = \frac{\mu_{s_t,t} - \omega - k_2(1 - I_{s_t}) - \alpha_0 y_{t-1} - \beta_0 \mu_{s_{t-1},t-1} - \gamma_0 D_{t-1} y_{t-1}}{\alpha_1 y_{t-1} - \beta_1 \mu_{s_{t-1},t-1} + \gamma_1 D_{t-1} y_{t-1}}$$

and $0 \leq g_t^* \leq 1$.

If $g_t^* < 0$ or $g_t^* > 1$, it is a not admissible value and the observation at time t falls in a crisp zone.

4 Fuzzy Regimes in the US Markets

We estimate the FR(n)–AMEM, with $n = 2$ and 3, on the four realized kernel volatility series of the main US market indices: S&P 500, Russell 2000, Dow Jones 30, NASDAQ 100 illustrated in Section 2. Next to these, we have estimated a MS(3)–AMEM and a ST–AMEM, following the specifications recalled above.

In Table 2 we show the estimation results for the four indices using the FR(3)–AMEM, that, as we will see, shows in general the best performance. All the indices are characterized by a strong persistence in state 2 and a very high probability to switch from $s_{t-1} = 3$ to $s_t = 1$. For the Russell index, we follow Hamilton and Susmel (1994) in setting the parameters p_{12}, p_{21} and α_0 equal to zero to achieve overall convergence.

In order to gain deeper insights on the features of the FR–AMEM (in particular for the interpretation of the regimes), it is useful to contrast its results with those derived from the MS–AMEM and ST–AMEM. For illustration purposes, we keep on considering a 2-state FR model, with two crisp and one fuzzy zone; the natural comparison is made in terms of a 3-state MS model, because the fuzzy zone of an FR(2)–AMEM could be considered as an alternative to the second regime of an MS(3)–AMEM. In what follows, the inference on the Markovian regimes is obtained as usual from the smoothed probabilities in each model, calculated over the full span, assigning the observation at time t to the regime with the highest smoothed probability at that time.

For the sake of space, we will present some detailed outcomes with reference to the S&P 500 index; the corresponding results for the other indices are qualitatively similar. In Figure 3 we reconstruct the distributions of the realized kernel volatilities for the MS and the FR models, highlighting the regimes identified on the basis of the inference on the states. The different attribution of observations to regimes across models is more evident for lower values, while the higher values are uniformly assigned to the highest regime. In reference to the characterization of fuzzy regimes, we can notice that they are typical for volatility values under 40%. Looking at Table 3, the average value of the volatility when the observation falls in regime 2, but in a fuzzy zone, is lower than the average of the fuzzy zone in regime 1. From the same table, we notice that the regime 2–crisp within the FR model captures just the highest peaks: on average, this regime includes fewer observations and has a higher average value (by more than 12%) than the third regime within the MS model. By and large, the regime 2–fuzzy in the FR model matches the highest peaks of regime 1 in the MS(3)–AMEM, whereas observations in the regime 1–fuzzy fall into the second regime of MS(3). This regularity is confirmed by comparing Figure 4, where we show the inference about the regimes within the FR(2)–AMEM, with the first graph of Figure 5, where we derive the inference about the regimes within the MS(3)–AMEM.

Still for the S&P 500 index, in Figure 4, we distinguish between the fuzzy (thinner symbols) and crisp (thicker symbols) classification within either regime 1 or 2. Notice that the volatility values belonging to the regime 1–crisp are very small, isolating the periods of very low volatility, whereas the regime 2–crisp corresponds only to the highest episodes. The observations in the fuzzy area are obviously in–between cases: we can notice that the regime 1–fuzzy generally captures cases in which there is a smooth reduction in volatility whereas regime 2–fuzzy generally corresponds to a smooth increase in

volatility. Typically we reach regime 2–crisp by passing through regime 2–fuzzy (a burst from a volatility already increasing), while the downward evolution from regime 2–crisp is characterized by a change to regime 1 entering its fuzzy zone. For ease of reference, we report in Table 4 the reconstruction of turbulences occurring in the US and European markets which characterize the dates when the S&P 500 observations are assigned to the regime 2–crisp. Reassuringly, we find all major related events, from Bear Sterns, to Lehman Brothers to the US Budget crisis and the evolution of the Greek debt crisis with the inception of uncertainty about the solidity of the Euro institutional arrangements.

4.1 In–sample versus Out–of–Sample Performance

The comparison among the eight models (HAR, F–HAR–RV, AMEM, SPvMEM, ST–AMEM, MS(3)–AMEM, FR(2)–AMEM, FR(3)–AMEM) is conducted in terms of residual autocorrelation and Root Mean Squared Errors (RMSE) and Mean Absolute Errors (MAE) both for the in–sample and out–of–sample cases. In Table 5 we show the p-values of the Ljung–Box statistics for each model and each index in correspondence of several lags. The presence of residual correlation in the HAR and confirms what was found in Gallo and Otranto (2015) (also for the F–HAR–RV model), whereas the MEM family satisfies uncorrelatedness, with some exceptions. In particular we notice a disappointing performance of the SPvMEM in general and of the MS(3)–AMEM for the Russell 2000 and a general improvement of the test statistics for the FR model passing from 2 to 3 regimes.

The model performances are interesting both in fitting and forecasting terms. We have considered 1 and 10 steps (i.e. using intermediate predictions as inputs) ahead out–of–sample forecasts, evaluating 336 observations corresponding to the most recent period (2014 and some of 2015). For the out–of–sample forecasts we have used the model estimated in–sample, fixing the estimated parameters for the full out–of–sample period.[5] In Table 6, we show the RMSE and MAE loss functions for each index. In the same table we report the results of Diebold Mariano tests where we indicate with a star the cases in which the null hypothesis of equal performance relative to the best performing model according to that loss function cannot be rejected at 5% significance level.

In general, we confirm the Gallo and Otranto (2015) result that the MS(3)–AMEM generally performs better than ST–AMEM in–sample and the opposite is true out–of–sample. One interesting feature of the FR models is to attain the desired goal of striking a compromise between the two, as they reach a satisfactory performance across in– and out–of–sample. As a matter of fact, they are almost always the best or statistically undistinguishable from the best model.

To decide what model performs globally better in in– *and* out–of–sample terms, it is convenient to evaluate the indices in relative terms, considering their relative performance with respect to the HAR model adopted as a benchmark model. Each relative variation of the loss function l of model M is given by:

$$v_M = \frac{l_M - l_{HAR}}{l_{HAR}},$$

[5] As in Gallo and Otranto (2015), we use an AR(4) to generate predictions of the VIX.

with generally an expected negative sign (an improvement relative to HAR). To synthesize results, a global loss function G_M for model M may be derived as a weighted average of the resulting six loss functions (RMSE and MAE, in– and 1– and 10–steps ahead out–of–sample):

$$G_M = \sum_l w_l v_M, \qquad \sum_l w_l = 1 \tag{4.1}$$

We consider four possible global loss functions, differing by the set of weights used. We choose

1. *uniform*: each w_l equal to $1/6$;
2. *oos oriented*: the weights of the in–sample loss functions equal to 0.1 and the weights of the out–of–sample loss functions equal to 0.2;
3. *oos short term (st) oriented*: the weights of the 1–step ahead out–of–sample loss functions equal to 0.3 and the others equal to 0.1;
4. *oos long term (lt) oriented*: the weights of the 10–step ahead out–of–sample loss functions equal to 0.3 and the others equal to 0.1.

The performance in terms of global functions is better, the more G_M decreases away from the value $G_M = 0$ (which marks the same performance as the HAR model).

In Table 7 we show the evaluation derived from the four alternative global loss functions (4.1). For the S&P 500 index we notice a better behavior of the FR models considering the uniform loss function; when the out–of–sample performance has more importance in the evaluation, ST–AMEM generally comes on top, but FR(2)–AMEM wins in the short term horizon; also FR(3)–AMEM is almost always the second best (the best in the uniform case).

For the Russell index the MS(3)–AMEM seems to have the best performance, but, recalling the results in Table 5, it clearly suffers from the presence of autocorrelation in its residuals; again FR(3)–AMEM is also here the second best. For the DJ, ST–AMEM and FR(3)–AMEM alternate in the role of best and second best, whereas the NASDAQ index clearly favors the FR(3)–AMEM, which is the best model with respect to all the global loss functions, ST–AMEM being the second best almost always.

4.2 The Underlying Trend in the US Markets

For illustrative purposes, we show in Figure 5 the time series profile of the S&P 500 LAV calculated with the three models, starting from the MS(3)–AMEM (expression (3.16) in the top panel), then the ST–AMEM (expression (3.17) in the middle panel) and then the FR(2)–AMEM (expression (3.12) in the bottom panel). In all three panels the values of the local average volatilities are superimposed to the realized volatility series. For all models, we confirm the difficulty highlighted by Gallo and Otranto (2015) in capturing the burst of volatility of October 2008.

By construction, the MS(3)–AMEM LAV is subject to abrupt and discrete changes, depending on the specific values of s_t and of $\bar{\sigma}_{t-1,s_{t-1}}$: this accounts for having more

than three values in the figure (some are very close to one another), but it also shows how persistent the states are. Consider in fact that persistence in state 1 corresponds to a LAV equal to 9.06, in state 2 equal to 14.28 and in state 3 equal to 58.27. The other (intermediate) values correspond to a (short–lived) transition between states.

The ST–AMEM and FR(2)–AMEM, in turn, show similar dynamics of LAV relative to one another, but the latter better follows the profile of the realized volatility, especially for its highest values (the maximum level of the LAV in the ST–AMEM is equal to 32.3 versus a maximum of 41.7 in the FR case). In practice, therefore, the FR model has the same desirable features as the ST–AMEM in terms of flexibility, but exploits the MS properties to better identify the frequent jumps and abrupt changes.

For all indices, we show in Table 8 the extreme values of the LAV for the ST–AMEM, FR(2)–AMEM and FR(3)–AMEM and the regime–specific LAV in the MS model when there is state persistence (i.e. calculated by the second equation in (3.16)). It is interesting to notice as the highest values of the ST–AMEM are very low when compared to the highest peaks of the realized volatility. Also the presence of a third regime in the FR model provides the possibility to identify very high peaks, excluding the Russell case, where the maximum value of the LAV in the third regime is equal to the LAV of the third regime of the MS(3)–AMEM.

In Table 9 we show a classification of the dates, indicating the percentage of cases where a) the volatilities of the four indices at time t are all assigned to the same state, b) just three indices to the same state, c) two indices to the same state and two to another one (but the same), d) two indices to one state and the other two to two other separate states,[6] using FR(3)–AMEM, FR(2)–AMEM and MS(3)–AMEM. For a large part of the observations we assign three or four indices contemporaneously to the same state (cases a) and b)): in particular, more than 88% using the MS(3)–AMEM, more than 81% using FR(2)–AMEM and more than 94% using FR(3)–AMEM. The latter shows more than 61% of agreement of the indices within the same regime (case a)).

The univariate inference on the regimes is thus quite similar across the four indices, pointing to the similarity in general behavior of the four index volatilities. Recalling the discussion about the features of some benchmark models, the model that produces a low frequency component which can be interpreted as LAV and therefore can be compared with the Fuzzy Regime MEM is the SPvMEM. By construction, this model produces a common trend estimated across volatilities. In this case, we find it instructive to provide graphical evidence of the LAV estimated on the Fuzzy Regime MEM in comparison with the common trend (with loadings) $b_i\phi_t$ in the first equation of expression (3.1): this is done in Figure 6. The similar behavior is striking: apart from the capability of the common trend of the SPvMEM to better capture the burst of volatility in 2008 (recall its nonparametric nature as a smoothing function), the two estimates of the low–frequency component are quite similar. The correlations between the two components are, respectively, 0.93 for the S&P 500, 0.89 for the Russell 2000, 0.90 for the DJ 30 and 0.96 for the NASDAQ 100. Interestingly, the correlations between LAVs on different indices vary from 0.93 (between Russell 2000 and NASDAQ 100) and 0.98 (between S&P 500 and DJ 30). The LAV of the FR–AMEM is derived parametrically as a combination of sharp and

[6] In the FR(2)–AMEM, the last case is not possible.

smooth transition features: this similarity with the SPvMEM common trend confirms its low frequency interpretability in line with the theoretical discussion of Section 3.3.

5 Conclusions

The existence of waves of turbulence ensuing from events affecting the national and global economies has characterized the behavior of volatility in financial markets, posing a serious challenge to econometric modeling of the corresponding time series. The main feature that one can notice in the corresponding graphs is that the average level of volatility changes through time at low frequencies, with short term dynamics around such levels marked by slow mean reversion (persistence).

In this paper, we have concentrated on Multiplicative Error Models and their MS and ST extensions studied by Gallo and Otranto (2015) by introducing a new class of models that combines features of both. This Fuzzy Regime specification allows for a classification of the observations with an assignment to regimes as in the MS model, but with a further characterization within regimes between *crisp* (clear–cut) and *fuzzy* (more dubious) cases. From the empirical results, and for the case of two regimes, we interpret the regime 1–crisp as a regime of low volatility and the regime 2–crisp as a regime corresponding to major turmoil events. The fuzzy regimes are characterized by a slow increase in volatility (and hence a slow transition from low to high volatility, regime 1–fuzzy), respectively, by a slow decrease in volatility (and hence a slow transition from high to low volatility, regime 2–fuzzy).

The analysis was carried out on four major US indices which represent various degrees of market depth, capitalization and liquidity: we chose to specify models for the realized kernel volatility measures which can be seen as the best available realized measures of volatility (with robustness to jumps and market microstructure noise). We had several variants of the suggested models estimated, changing the number of regimes considered in the MS component, and keeping the HAR (Corsi (2009)), the F–HAR–RV (Atak and Kapetanios (2013)), the AMEM (Engle and Gallo (2006)) and the SPvMEM (Barigozzi *et al.* (2014)) as benchmark references. The estimation results were evaluated both in– and out–of–sample with a number of loss functions. We suggested a way to aggregate in– and out–of–sample performances in a global loss function with alternative choices of weights to mimic different goals a model could be geared toward. The general result is that the introduction of the fuzzy area allows for a convenient tradeoff between in–sample performance where MS models perform the best and out–of–sample, where the ST–AMEM comes out first, in line with Hansen and Timmerman (2015).

The fuzzy regime models also seem to capture the explosion in volatility following the bankruptcy of Lehman Brothers in Sep.–Oct. 2008; in the two regime case, we are capable of interpreting each observation assigned to the regime 2–crisp in terms of major events which affected the US markets with the evolution of the subprime mortgage crisis and then the uncertainty surrounding the US Budget and the Euro Area institutional arrangements.

We have devoted substantial attention to the comparison between the estimate of our *Local Average Volatility* as a parametric estimate of the low frequency component underlying the dynamics of the series and the nonparametric estimate of the *common trend*

estracted by the SPvMEM from all four realized volatility series. The fact that the two approaches deliver similar results, and that LAV across indices have a very high correlation is reassuring on the interpretability of the LAV as the long run volatility in the market.

Acknowledgments

Thanks are due to Fabrizio Cipollini, two anonymous referees and the Associate Editor for constructive remarks which helped us in refocusing the contribution of the paper. We acknowledge the comments by conference participants at the 2015 Macroeconometric Modeling Workshop Dec. 3–4 2015, Academia Sinica, Taipei, the Computational and Financial Econometrics Conference Dec. 12–14 2015, London and the 7th Italian Congress of Econometrics and Empirical Economics, Jan. 25–27 2017, Messina.

References

Amado, C. and Teräsvirta, T. (2008). Modelling conditional and unconditional heteroskedasticity with smoothly time-varying structure. Technical report, SSE/EFI Working Paper Series in Economics and Finance. No. 691.

Amado, C. and Teräsvirta, T. (2014). Modelling changes in the unconditional variance of long stock return series. *Journal of Empirical Finance*, **25**, 15–35.

Andersen, T. G., Bollerslev, T., Christoffersen, P. F., and Diebold, F. X. (2006). Volatility and correlation forecasting. In G. Elliott, C. W. J. Granger, and A. Timmermann, editors, *Handbook of Economic Forecasting*. North Holland.

Andersen, T. G., Bollerslev, T., and Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling and forecasting of return volatility. *Review of Economics and Statistics*, **89**, 701–720.

Atak, A. and Kapetanios, G. (2013). A factor approach to realized volatility forecasting in the presence of finite jumps and cross-sectional correlation in pricing errors. *Economics Letters*, **120**(2), 224 – 228.

Barigozzi, M., Brownlees, C., Gallo, G., and Veredas, D. (2014). Disentangling systematic and idiosyncratic dynamics in panels of volatility measures. *Journal of Econometrics*, **182**, 364–384.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008). Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica*, **76**, 1481–1536.

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

Brownlees, C. T. and Gallo, G. M. (2010). Comparison of volatility measures: a risk management perspective. *Journal of Financial Econometrics*, **8**, 29–56.

Brownlees, C. T. and Gallo, G. M. (2011). Shrinkage estimation of semiparametric multiplicative error models. *International Journal of Forecasting*, **27**, 365–378.

Brownlees, C. T., Cipollini, F., and Gallo, G. M. (2011). Intra-daily volume modeling and prediction for algorithmic trading. *Journal of Financial Econometrics*, **9**, 489–518.

Caporin, M., Rossi, E., and Santucci de Magistris, P. (2015). Volatility jumps and their economic determinants. *Journal of Financial Econometrics*, **14**, 29–80.

Caporin, M., Rossi, E., and Santucci De Magistris, P. (2017). Chasing volatility: a persistent multiplicative error model with jumps. *Journal of Econometrics*, **198**, 122–145.

Cipollini, F., Engle, R. F., and Gallo, G. M. (2013). Semiparametric vector mem. *Journal of Applied Econometrics*, **28**, 1067–1086.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, **7**, 174–196.

Engle, R. F. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics*, **17**, 425–446.

Engle, R. F. and Gallo, G. M. (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, **131**, 3–27.

Engle, R. F. and Rangel, J. G. (2008). The spline-GARCH model for low frequency volatility and its global macroeconomic causes. *Review of Financial Studies*, **21**, 1187–1222.

Francq, C., Roussignol, M., and Zakoïan, J.-M. (2001). Conditional heteroskedasticity driven by hidden markov chains. *Journal of Time Series Analysis*, **22**, 197–220.

Gallo, G. M. and Otranto, E. (2015). Forecasting realized volatility with changing average levels. *International Journal of Forecasting*, **31**, 620–634.

Gonzalez-Rivera, G. (1998). Smooth–transition garch models. *Studies in Non–Linear Dynamics and Econometrics*, **3**, 61–78.

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

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

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Hansen, P. R. and Timmerman, A. (2015). Discussion of "comparing predictive accuracy, twenty years later' ' by francis x. diebold. *Journal of Business and Economic Statistics*, **33**, 17–21.

Hansen, P. R., Huang, Z., and Shek, H. H. (2011). Realized GARCH: A complete model of returns and realized measures of volatility. *Journal of Applied Econometrics, forthcoming*.

Heber, G., Lunde, A., Shephard, N., and Sheppard, K. (2009). Omi' s realised library, version 0.1. Technical report, Oxford-Man Institute, University of Oxford.

Kim, C. (1994). Dynamic linear models with markov switching. *Journal of Econometrics*, **60**, 1–22.

Maheu, J. M. and McCurdy, T. H. (2002). Nonlinear features of realized fx volatility. *Review of Economics and Statistics*, **84**(4), 668–681.

McAleer, M. and Medeiros, M. (2008). A multiple regime smooth transition heterogeneous autoregressive model for long memory and asymmetries. *Journal of Econometrics*, **147**, 104–119.

Nelson, D. B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory*, **6**, 318–334.

Otranto, E. (2016). Adding flexibility to markov switching models. *Statistical Modelling*, **16**, 477–498.

Otranto, E. and Gallo, G. M. (2002). A nonparametric bayesian approach to detect the number of regimes in markov switching models. *Econometric Reviews*, **21**(4), 477–496.

Whaley, R. E. (2009). Understanding the vix. *The Journal of Portfolio Management*, **35**, 98105.

Zadeh, L. (1965). Fuzzy sets. *Information and Control*, **8**, 338–353.

Tables and Figures

Table 1: Nonparametric Bayesian approach to detect the number of in the volatility of four US financial indices. Empirical posterior distribution of the number of regimes as a function of the hyperparameter A. In bold the mode of the distribution.

	number of regimes							
A	2	3	4	5	6	7	8	9
	S&P 500							
0.22			**0.88**	0.10	0.02			
0.35			**0.70**	0.29	0.01			
0.47			**0.82**	0.17	0.01			
0.60			**0.51**	0.25	0.18	0.05	0.01	
	Russell 2000							
0.22		**0.83**	0.16	0.01				
0.35		**0.91**	0.09					
0.47	0.03	0.14	**0.71**	0.11	0.01			
0.60		**0.85**	0.15					
	DJ 30							
0.22			**0.50**	0.36	0.08	0.06		
0.35		0.15	**0.51**	0.23	0.05	0.01	0.02	0.03
0.47		**0.60**	0.33	0.05	0.02			
0.60	0.09	**0.46**	0.30	0.12	0.01	0.01	0.01	
	NASDAQ 100							
0.22			**0.88**	0.12				
0.35		**0.82**	0.10	0.08				
0.47		**0.54**	0.38	0.07	0.01			
0.60		0.07	**0.76**	0.16	0.01			

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 2: Estimation results (standard errors in parentheses) of a FR(3)–AMEM applied to four realized kernel volatilities of financial time series.

	ω	k_2	k_3	p_{11}	p_{12}	p_{21}	p_{22}	p_{31}	p_{32}
S&P 500	1.25	1.58	10.56	0.68	0.13	0.01	0.99	0.99	0.00
	(0.19)	(0.21)	(2.77)	(0.04)	(0.04)	(0.00)	(0.00)	(0.01)	(0.00)
Russell 2000	1.52	1.78	2.75	0.78			0.99	0.95	0.05
	(0.21)	(0.16)	(0.53)	(0.05)			(0.00)	(0.02)	(0.02)
DJ 30	1.39	1.32	14.88	0.79	0.10	0.00	0.99	0.99	0.00
	(0.22)	(0.08)	(4.01)	(0.03)	(0.02)	(0.00)	(0.00)	(0.02)	(0.00)
NASDAQ 100	0.42	1.26	5.93	0.60	0.15	0.01	0.98	0.86	0.00
	(0.94)	(0.50)	(1.17)	(0.10)	(0.05)	(0.00)	(0.00)	(0.07)	(0.00)
	α_0	α_*	β_0	β_*	γ_0	γ_*	δ	c	a
S&P 500	0.00	0.37	0.51	0.03	0.12	0.00	1.68	-0.28	16.72
	(0.00)	(0.03)	(0.04)	(0.00)	(0.01)	(0.00)	(0.12)	(0.05)	(0.66)
Russell 2000		0.39	0.49	0.03	0.07	0.06	1.15	-0.77	17.09
		(0.03)	(0.04)	(0.00)	(0.04)	(0.06)	(0.13)	(0.18)	(0.79)
DJ 30	0.00	0.35	0.52	0.03	0.08	0.04	1.57	-0.41	15.93
	(0.00)	(0.03)	(0.04)	(0.00)	(0.03)	(0.04)	(0.11)	(0.07)	(0.52)
NASDAQ 100	0.27	0.17	0.47	0.01	0.00	0.11	0.82	-0.09	21.29
	(0.07)	(0.10)	(0.05)	(0.01)	(0.01)	(0.03)	(0.22)	(0.59)	(1.11)

Table 3: Average values of the realized kernel volatility of S&P 500 in correspondence of the regimes derived from MS(3)–AMEM and FR(2)–AMEM.

	Regime from MS(3)			Regime from FR(2)				
	1	2	3	1–crisp	1–fuzzy	2–fuzzy	2–crisp	Total
Average	9.22	17.23	35.06	7.60	16.10	13.25	47.51	13.86
Frequency	1511	778	207	655	783	968	87	2496

Table 4: Timetable of events corresponding to the Regime 2–crisp dates in FR(2)–AMEM for S&P 500.

Dating	Related Events
17 Mar 2008	Collapse of Bear Stearns and contagion effect (Lehman shares fall as much as 48%)
17–18 Sep 2008	Aftermath of the Lehman Brothers collapse (15 Sep)
22 Sep-8 Dec 2008	Height of the financial crisis
6–7 May 2010	Flash crash (6 May)
18–19 May 2010	On May 18, Greece receives the first EUR 20 billion rescue package from EU and IMF
8–12 Aug 2011	Consequences of the Standard & Poor's downgrading of US credit rating from AAA to AA+ on 6 August 2011.
16 Aug 2011	Joint press conference Sarkozy–Merkel about European debt hinting at the possibility of Grexit.
18–30 Aug 2011	Contagion following market reactions to Euro debt crisis (on 18 Aug the European indices fall by 4.5-6%)

Table 5: P-values of the Ljung–Box statistics to test for autocorrelation at several lags: residuals and squared residuals from several models estimated on four US indices. The bold (resp. italics) numbers indicate no autocorrelation at 5% (resp. 1%) significance level.

	residuals				squared residuals			
	LB(1)	LB(5)	LB(10)	LB(20)	LB(1)	LB(5)	LB(10)	LB(20)
				S&P 500				
HAR	0.009	0.000	0.000	0.000	0.000	0.000	0.000	0.000
F–HAR–RV	*0.011*	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AMEM	**0.227**	**0.486**	**0.153**	0.000	**0.585**	**0.604**	**0.172**	0.001
SPvMEM	0.004	0.000	0.000	0.000	0.001	0.000	0.000	0.000
ST–AMEM	**0.457**	**0.515**	**0.056**	0.002	**0.339**	**0.495**	**0.061**	*0.011*
MS(3)–AMEM	**0.833**	**0.142**	**0.060**	0.001	**0.149**	*0.028*	0.009	0.000
FR(2)–AMEM	0.003	*0.020*	*0.045*	*0.022*	0.004	*0.034*	**0.078**	**0.057**
FR(3)–AMEM	**0.348**	**0.353**	*0.013*	0.000	**0.146**	**0.473**	*0.031*	0.006
				Russell 2000				
HAR	*0.011*	0.000	0.000	0.000	0.000	0.000	0.000	0.000
F–HAR–RV	*0.012*	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AMEM	**0.140**	**0.227**	**0.052**	0.004	**0.551**	**0.211**	**0.228**	0.004
SPvMEM	**0.260**	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ST–AMEM	**0.759**	**0.548**	*0.043*	0.001	**0.558**	**0.268**	**0.199**	0.007
MS(3)–AMEM	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
FR(2)–AMEM	**0.063**	**0.236**	*0.028*	0.000	*0.036*	**0.234**	**0.107**	0.008
FR(3)–AMEM	**0.101**	**0.428**	*0.048*	0.001	**0.090**	**0.343**	**0.082**	0.006
				DJ 30				
HAR	0.006	0.000	0.000	0.000	0.000	0.000	0.000	0.000
F–HAR–RV	0.008	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AMEM	**0.629**	**0.174**	**0.085**	0.002	**0.850**	**0.195**	*0.047*	*0.034*
SPvMEM	**0.281**	0.000	0.000	0.000	0.007	0.000	0.000	0.000
ST–AMEM	**0.179**	**0.117**	*0.026*	*0.014*	**0.134**	**0.092**	*0.036*	**0.127**
MS(3)–AMEM	**0.165**	0.003	*0.013*	**0.050**	*0.014*	0.001	0.003	*0.027*
FR(2)–AMEM	0.002	0.003	0.006	*0.016*	0.005	0.005	*0.015*	**0.077**
FR(3)–AMEM	*0.035*	*0.034*	0.002	0.001	*0.025*	**0.088**	*0.018*	**0.070**
				NASDAQ 100				
HAR	*0.025*	0.000	0.000	0.000	0.000	0.000	0.000	0.000
F–HAR–RV	*0.030*	0.000	0.000	0.000	0.000	0.000	0.000	0.000
AMEM	*0.010*	0.003	0.000	0.000	**0.119**	**0.128**	*0.021*	0.000
SPvMEM	**0.903**	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ST–AMEM	**0.447**	*0.018*	0.000	0.000	**0.984**	**0.324**	**0.065**	*0.011*
MS(3)–AMEM	*0.030*	0.000	0.000	0.000	**0.482**	*0.030*	0.000	0.000
FR(2)–AMEM	*0.020*	0.000	0.000	0.000	0.007	0.003	0.009	*0.022*
FR(3)–AMEM	**0.868**	0.006	0.004	0.000	**0.481**	**0.077**	**0.079**	*0.037*

Table 6: In–sample and 1-step and 10-steps out–of–sample performance of several models for four US indices, calculated in terms of RMSE and MAE. The bold numbers indicate the best performance; the star indicates that the Diebold–Mariano statistic leads to a non rejection, at a 5% significance level, of the null hypothesis of equal mean (squared and absolute) forecasting errors with respect to the best performing model.

	In–Sample		Out–of–Sample			
	RMSE	MAE	RMSE 1	MAE 1	RMSE 10	MAE 10
			S&P 500			
HAR	5.25	3.102	2.75	2.131	10.561	7.928*
F–HAR–RV	5.222	3.086	2.743	2.120*	20.455	16.413
AMEM	5.092	2.989	2.655*	2.072*	10.92	9.228
SPvMEM	4.714	2.755	2.653*	2.028*	10.249	7.745*
ST–AMEM	4.957	2.935	2.592*	2.044*	**9.111**	**7.010**
MS(3)–AMEM	4.464*	**2.615**	2.765	2.134	10.23	7.967
FR(2)–AMEM	4.825	2.773	**2.570**	**2.018**	9.54*	7.212*
FR(3)–AMEM	**4.407**	2.654*	2.633	2.079	9.435	7.364
			Russell 2000			
HAR	5.292	3.333	2.859	2.280	9.525*	7.324*
F–HAR–RV	5.275	3.334	2.886	2.305	10.835	8.318
AMEM	5.113	3.242	2.942	2.370	10.826	9.278
SPvMEM	4.772	3.07	2.93	2.331	9.974*	7.563*
ST–AMEM	5.023	3.220	2.788	2.234	9.201*	7.628*
MS(3)–AMEM	**4.420**	**2.776**	2.876	2.243	**8.797**	**6.650**
FR(2)–AMEM	4.887	3.114	2.745*	2.174*	10.005*	7.911
FR(3)–AMEM	4.761	2.903	**2.736**	**2.168**	9.259*	7.413
			DJ 30			
HAR	5.297	3.096	3.012	2.262	10.421	7.875*
F–HAR–RV	5.154	3.016	2.930	2.219	22.366	18.28
AMEM	5.153	2.983	2.840*	2.143*	10.408	8.665
SPvMEM	4.783	2.798	2.895*	2.124*	10.475	8.088
ST–AMEM	5.049	2.930	2.789*	2.087	**8.938**	**7.053**
MS(3)–AMEM	4.528*	**2.661**	2.998	2.224	10.180	7.984
FR(2)–AMEM	4.945	2.828	**2.779**	**2.044**	9.941*	7.542*
FR(3)–AMEM	**4.480**	2.735	2.818*	2.115	9.324	7.311
			NASDAQ 100			
HAR	4.304	2.545	2.683	2.152	10.081	7.982*
F–HAR–RV	4.300	2.543	2.686	2.157	14.202	11.194
AMEM	4.257	2.519	2.825	2.353	10.111	8.582
SPvMEM	3.905	2.356	2.625*	**1.972**	10.361	8.082*
ST–AMEM	4.142	2.477	2.534*	2.030*	**8.921**	**7.147**
MS(3)–AMEM	**3.474**	**2.114**	2.656	2.127	9.644*	7.637*
FR(2)–AMEM	3.994	2.316	**2.503**	1.973*	10.261	7.927*
FR(3)–AMEM	3.695	2.212	2.559*	2.055*	9.127*	7.450*

Table 7: Global Loss functions of several models for four US indices. The bold numbers indicate the best performance, the italic numbers the second best.

	uniform	oos oriented	oos st oriented	oos lt oriented
			S&P 500	
HAR	0.000	0.000	0.000	0.000
F–HAR–RV	0.331	0.399	0.197	0.600
AMEM	0.012	0.020	-0.006	0.047
SPvMEM	-0.058	-0.049	-0.052	-0.046
ST–AMEM	-0.077	**-0.081**	-0.066	**-0.097**
MS(3)–AMEM	-0.054	-0.035	-0.031	-0.038
FR(2)–AMEM	*-0.082*	*-0.080*	**-0.073**	-0.087
FR(3)–AMEM	**-0.092**	-0.079	*-0.068*	*-0.091*
			Russell 2000	
HAR	0.000	0.000	0.000	0.000
F–HAR–RV	0.048	0.058	0.033	0.084
AMEM	0.068	0.088	0.055	0.122
SPvMEM	-0.008	0.008	0.004	0.011
ST–AMEM	-0.020	-0.016	-0.021	-0.011
MS(3)–AMEM	**-0.085**	**-0.069**	**-0.053**	**-0.085**
FR(2)–AMEM	-0.016	-0.005	-0.027	0.016
FR(3)–AMEM	*-0.056*	*-0.045*	*-0.052*	*-0.037*
			DJ 30	
HAR	0.000	0.000	0.000	0.000
F–HAR–RV	0.429	0.520	0.248	0.793
AMEM	-0.011	-0.007	-0.029	0.015
SPvMEM	-0.043	-0.033	-0.046	-0.020
ST–AMEM	*-0.087*	**-0.094**	*-0.083*	**-0.106**
MS(3)–AMEM	-0.051	-0.034	-0.035	-0.033
FR(2)–AMEM	-0.070	-0.069	-0.077	-0.061
FR(3)–AMEM	**-0.098**	*-0.091*	**-0.085**	*-0.097*
			NASDAQ 100	
HAR	0.000	0.000	0.000	0.000
F–HAR–RV	0.135	0.163	0.082	0.244
AMEM	0.034	0.043	0.050	0.036
SPvMEM	-0.039	-0.030	-0.044	-0.015
ST–AMEM	-0.066	*-0.073*	*-0.062*	*-0.084*
MS(3)–AMEM	*-0.078*	-0.058	-0.051	-0.064
FR(2)–AMEM	-0.050	-0.044	-0.060	-0.028
FR(3)–AMEM	**-0.087**	**-0.078**	**-0.071**	**-0.085**

Table 8: Local average volatilities using alternative AMEM models for four US financial indices.

		S&P 500	Russell 2000	DJ 30	NASDAQ 100
MS(3)–AMEM	$\bar{\sigma}_1$	9.06	7.78	9.09	9.14
	$\bar{\sigma}_2$	14.28	11.91	14.37	14.77
	$\bar{\sigma}_3$	58.27	70.00	52.87	63.68
ST–AMEM	$\bar{\sigma}^l$	5.86	10.28	5.77	7.97
	$\bar{\sigma}^h$	32.61	25.79	30.84	29.56
FR(2)–AMEM	$\bar{\sigma}_1^l$	6.72	4.57	6.16	5.79
	$\bar{\sigma}_1^h$	40.69	41.36	33.13	39.10
	$\bar{\sigma}_2^l$	7.79	6.08	6.98	7.85
	$\bar{\sigma}_2^h$	41.76	42.86	33.95	41.16
FR(3)–AMEM	$\bar{\sigma}_1^l$	2.88	3.15	3.18	1.61
	$\bar{\sigma}_1^h$	127.60	65.53	162.90	184.74
	$\bar{\sigma}_2^l$	4.46	4.93	4.50	2.87
	$\bar{\sigma}_2^h$	129.18	67.31	164.22	186.00
	$\bar{\sigma}_3^l$	15.01	7.68	19.38	8.80
	$\bar{\sigma}_3^h$	139.74	70.05	179.10	191.93

For the MS(3)–AMEM we show the LAV when $s_t = s_{t-1}$; for the ST–AMEM we show the lowest and highest theoretical LAV; for the FR(2)–AMEM and FR(3)–AMEM the lowest and highest theoretical volatility within each regime.

Table 9: Inference on the regime: number of indices with the volatility falling in the same regime at the same time, using MS(3)–AMEM, FR(2)–AMEM and FR(3)–AMEM as alternative models.

	Number of indices with the same state			
Model	Four	Three	Two pairs	Two
MS(3)–AMEM	43.09	45.37	10.41	1.12
FR(2)–AMEM	42.97	38.25	18.78	
FR(3)–AMEM	61.67	32.64	5.09	0.60

Figure 1: Realized kernel volatility of four US market indices: Jan. 2, 2004 – May 5, 2015.

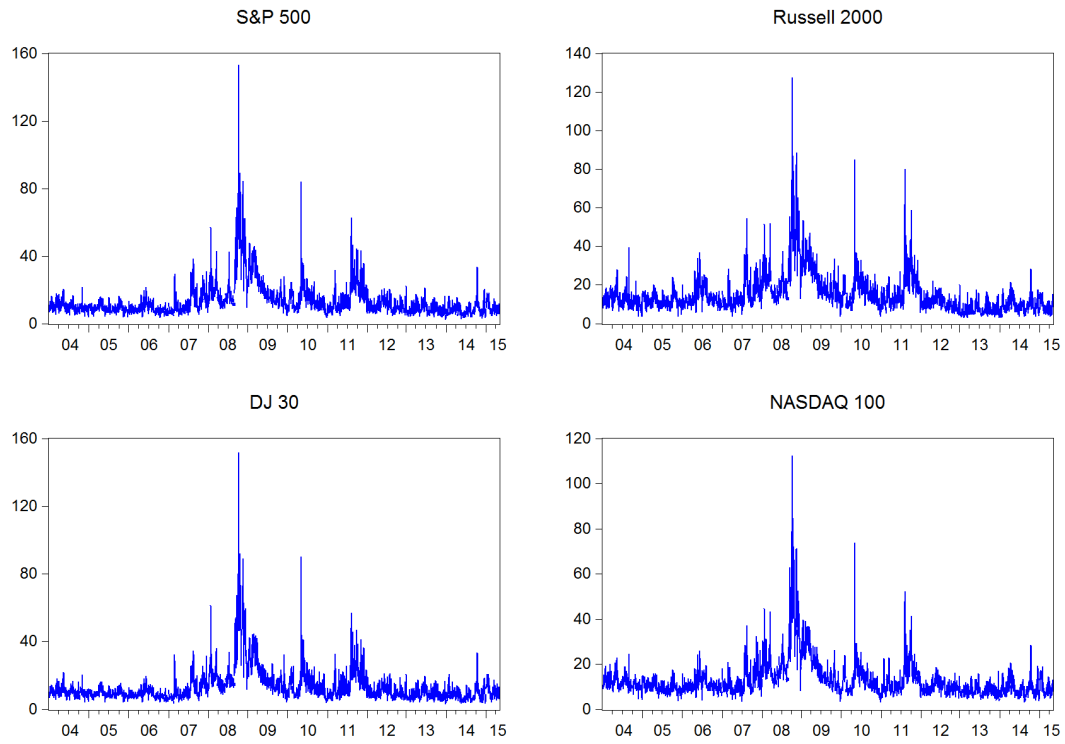

Figure 2: Theoretical behavior of the LAV in regime 1 (lower curve) and regime 2 (upper curve) as a function of the smooth transition g_t. The shaded area represents the fuzzy zone. Parameter values: $\omega = 1$, $k_2 = 3$, $\alpha_0 = 0.5$, $\alpha_* = 0.4$, $\beta_0 = 0$, $\beta_* = 0$, $\gamma_0 = 0$, $\gamma_* = 0$.

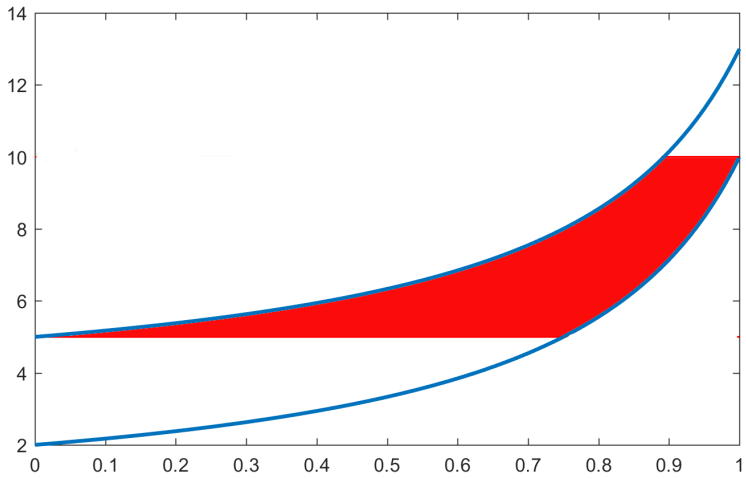

Figure 3: S&P 500 index: Distribution of the realized volatility in terms of regimes derived from the MS(3)–AMEM (top) and the FR(2)–AMEM (bottom). Absolute frequencies on the y-axis. Jan. 2, 2004 – Dec.31, 2013: $T = 2496$.

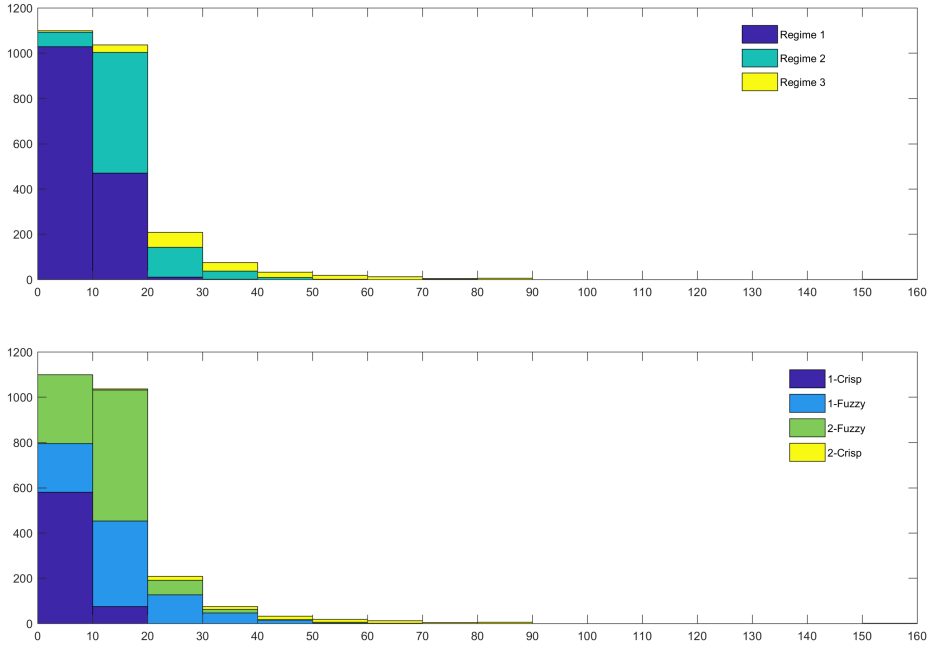

Figure 4: S&P 500 index: realized kernel volatility and inference on the regime (right axis), distinguishing crisp and fuzzy regimes derived from a FR(2)–AMEM. Jan. 2, 2004 – Dec.31, 2013.

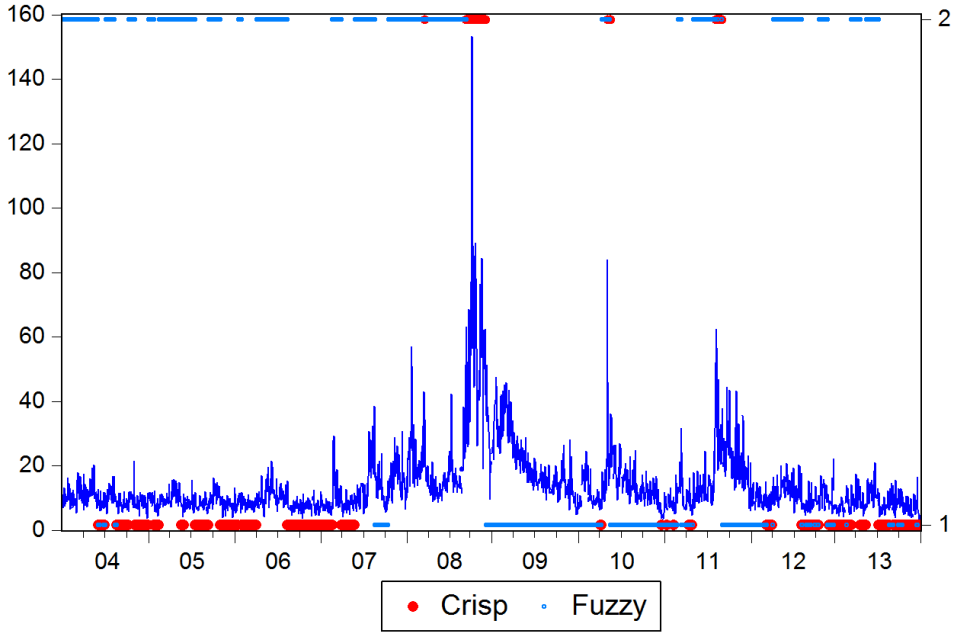

Figure 5: S&P 500 index: Realized kernel volatility and Local Average Volatilities according to MS(3)–AMEM, ST–AMEM and FR(2)–AMEM. Jan. 2, 2004 – Dec.31, 2013.

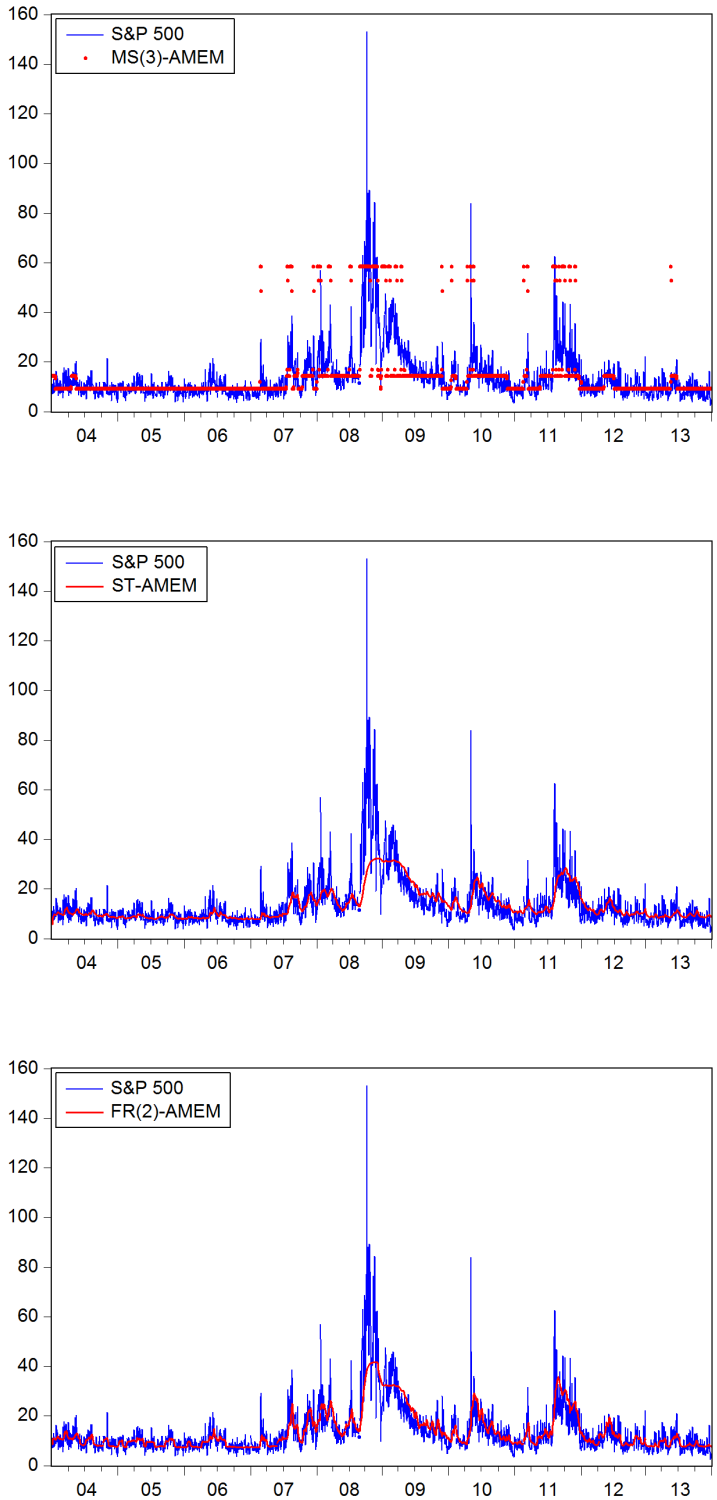

Figure 6: LAV (derived from the FR(2)–AMEM) and common trend (derived from the SPvMEM) of four US market indices. Jan. 2, 2004 – Dec.31, 2013.