Chasing volatility
A persistent multiplicative error model with jumps

Massimiliano Caporin * Eduardo Rossi † Paolo Santucci de Magistris ‡

This version: February 24, 2015

Abstract

The volatility of financial returns is characterized by persistence and unpredictable large increments. We propose the Multiplicative Error Model with volatility jumps (MEM-J) to characterize and predict the probability and the size of these extreme events. When a volatility jump component is included in the multiplicative specification, the conditional density of the realized measure is a countably infinite mixture of Gamma and *Kappa* distributions. Conditions for stationarity are derived. The MEM-J is fitted to the daily bipower measures of 7 stock indexes and 16 individual NYSE stocks. The estimates of the volatility jump component confirm that the probability of jumps dramatically increases during the financial crises. Compared to other realized volatility models with fat tails, the introduction of the volatility jump component provides a sensible improvement in the fit, as well as for in-sample and out-of-sample volatility tail forecasts.

Keywords: Multiplicative Error Model, Volatility Jumps, Bipower variation, Volatility-at-Risk

J.E.L. codes: C22, C58, G10

*Corresponding author: Dipartimento di Scienze Economiche e Aziendali "Marco Fanno", University of Padova, Via del Santo, 33, 35123, Padova; email: massimiliano.caporin@unipd.it; ph. +39-0498274258.

†Dipartimento di Scienze Economiche e Aziendali, University of Pavia, Italy.

‡Department of Economics and Business and CREATES, Aarhus University, Denmark.

1 Introduction

A great deal of the recent literature on volatility modeling exploits realized volatility measures as ex-post estimates of the return variation over a given horizon. Realized volatility series are characterized by large variations, often in correspondence of news arrival, thus challenging existing volatility models (see e.g. Caporin et al., 2014b). Such an evidence is further confirmed by the empirical analysis presented in this paper; indeed, currently available models, mostly designed to fit the dynamics of realized volatility sequences, fail in explaining the probability and the occurrence of volatility bursts. The inclusion of jumps in the volatility process, in addition to jumps in prices, is a step forward a more appropriate description of the volatility features. Several works, adopting a continuous-time framework, focus on the interactions between jumps in prices and volatility showing the importance of both components in fitting the observed dynamics of returns, see e.g. Chernov et al. (2003), Duffie et al. (2000), Pan (2002), Eraker (2004), Eraker et al. (2003), Jones (2003), Broadie et al. (2007), Todorov and Tauchen (2011), Andersen et al. (2012b), Bandi and Renò (2012, 2013).

Alternatively to the continuous time framework, where a jump is a discontinuity in the trajectory of price and/or volatility, a jump in discrete time takes the form of an extreme event, for example a very large value of the daily volatility. So far, the analysis of jumps in a discrete-time setting has focused on the role that jumps in prices have in predicting the future volatility. For example, Andersen et al. (2007), Corsi et al. (2010) and Caporin et al. (2014a) extend the Heterogeneous Autoregressive (HAR) model of Corsi (2009) to include past price jumps, i.e. the HAR-RV-J model. Differently, Caporin et al. (2014b) explicitly model the probability of volatility jumps in a HAR setup, leading to a significant increase in the model fit on the right tail. However, the use of the HAR model calls for a linear and additive specification of volatility jumps, that is more appropriate under a log-transformation of realized measure series, see Caporin et al. (2014b). But in this case obtaining the forecasts distribution of the volatility level can be problematic in non-Gaussian cases.

To overcome this problem, we propose the Asymmetric HAR-MEM-J (AHAR-MEM-J), an extension of the multiplicative error model (MEM) by Engle (2002) and Engle and Gallo (2006). The purpose of the model is to assign a probability to volatility boosts at each point in time. This is possible by the inclusion of a latent process labeled *volatility jump*, that generates infrequent large moves in the volatility. We can think of the AHAR-MEM-J as a three-factor model: first, a long-run factor, modeled by the Asymmetric HAR (AHAR), which accounts for the long-run dependence of volatility and improves the fit compared to less persistent GARCH-type specifications; second, a short-run factor, which represents the transitory component of the volatility process; and third, the volatility jump factor, which is responsible for the presence of realizations in the right tail of volatility distribution. This is close to the dynamic two-factor models of Ghysels et al. (2004) and Bauwens and Veredas (2004) for durations.

In order to obtain a closed-form expression for the conditional density, we assume that both volatility innovations and jump sizes are Gamma-distributed with different degrees of freedom. Despite the choice of a Gamma distribution may appear rather restrictive, the resulting conditional density of volatility is instead very flexible. It is a countably infinite mixture of two random variables: one distributed as a Gamma, when volatility jumps are absent, and the second distributed as a *Kappa*, henceforth K, when the number of volatility jumps is strictly positive. The K is a product distribution, known in physics and radar applications, but never used in econometrics, to the best of our knowledge. The K equals the product of two Gamma-distributed random variables with different degrees of freedom. The main advantage of this distributional choice is that the conditional moments of the dependent variable, the likelihood function and the quantiles can be obtained in closed-form, thus avoiding to rely on simulation-based methods to estimate the parameters. Moreover, in order to account for the empirical evidence of jump clustering, the intensity parameter, governing the jump occurrence in the compound Poisson process, can be made time-varying with an autoregressive specification, in the spirit of Hansen (1994) and Maheu and McCurdy (2004).

The AHAR-MEM-J parameters are estimated by maximum likelihood. We derive conditions for stationarity under the presence of jumps following the approach of Bougerol and Picard (1992) and Franq and Zakoian (2010). Unfortunately, given that the distribution of the innovations has an infinite mixture structure, it is not possible to verify the conditions for consistency and asymptotic normality of the maximum likelihood estimators, as in Engle and Gallo (2006) and Hansen et al. (2012). However, a Monte Carlo simulation experiment highlights that maximum likelihood estimates are unbiased and unimodal in finite samples.

In the empirical analysis is based on the high-frequency returns of 7 stock indexes and 16 NYSE stocks. We estimate the AHAR-MEM-J on the bipower variation series, which is an ex-post volatility measure robust to price jumps. In this way, we disentangle the price jumps from the volatility dynamics. A series of robustness checks confirms that the estimates of the AHAR-MEM-J are not strongly affected by the choice of the ex-post volatility measure and by the measurement error associated to it. The empirical application shows that the AHAR-MEM-J with time-varying jump intensity allows for a great flexibility in accommodating the probability of extremely large volatility realizations, dramatically improving the fit of the baseline MEM and other more sophisticated models. Potential sources of jump innovations to volatility can be important news, data releases, or unexpected events, which might induce market participants to suddenly revise their portfolios, thus producing large variations in the volatility level. By analogy to the Value-at-Risk (VaR), we introduce the *Volatility-at-Risk* (VolaR) which constitutes a natural measure of risk when designing volatility trading strategies. The evaluation of the VolaR strongly supports the MEM-J specification against models without jumps.

In summary, the contributions of the paper are at least three. Firstly, we generalize the baseline MEM of Engle and Gallo (2006) by including a volatility jump term, which captures the occasional boosts of volatility, and a pseudo long-memory component which is able to account for the observed persistence. Secondly, the conditional density of the dependent variable is derived in closed form. Thirdly, we provide empirical evidence on the relevance of jumps as a source of variation in the realized volatility measures and on their contribution to a correct estimation of the VolaR. On the contrary models without jumps are unable to fit the right tail of the volatility density, especially in periods of markets turmoils.

The paper is organized as follows. Section 2 sets the notation of the baseline MEM. Section 3 describes the MEM-J and the finite mixture distribution that characterizes the conditional density of the dependent variable. Conditional moments are also presented. Note that, in the rest of the paper, the term *jump*, will be always referred to volatility jumps, unless differently specified. Section 4 discusses both model extensions with HAR dynamics and time-varying parameters, and the model's properties, such as conditions for covariance stationarity and maximum likelihood estimation. Section 5 describes the dataset and illustrates the empirical results with stocks indexes and individual S&P 500 stocks under different model specifications. In Section 6 the results of the VolaR analysis are reported and discussed. Finally, Section 7 concludes. Proofs, selected derivations of relevant quantities and additional theoretical details are included in Appendices A and B.

2 The baseline MEM

In this section, we briefly present the MEM in its simplest form, as introduced by Engle and Gallo (2006), with the purpose of setting up the notation used throughout the rest of the paper. Let RM_t be an ex-post estimator of daily volatility at time t. We assume that RM_t is a.s. strictly positive and it follows a MEM, i.e.

$$RM_t = \mu_t \varepsilon_t \tag{1}$$

with

$$\mu_t = \omega + \alpha RM_{t-1} + \beta \mu_{t-1}.$$

with $\omega > 0, \alpha \geq 0, \beta \geq 0$. The innovation ε_t is a random variable with scale-shape Gamma density

$$\varepsilon_t | I_{t-1} \overset{iid}{\sim} \Gamma\left(\frac{1}{\nu}, \nu\right) \tag{2}$$

where $\frac{1}{\nu}$ is the scale and ν is the shape of the Gamma density, both driven by the common parameter ν, so that $\mathbb{E}_{t-1}\left[\varepsilon_t\right] = 1$ and $\mathbb{V}_{t-1}\left[\varepsilon_t\right] = \frac{1}{\nu}$. Let $\zeta = [\omega, \alpha, \beta, \nu]'$ be the parameter vector of the baseline MEM model.

If RM_t follows a MEM, with a Gamma disturbance, the conditional density, given I_{t-1}, is $\Gamma(\mu_t, \nu)$ in mean-shape form. Therefore, the conditional mean and variance are $\mathbb{E}\left[RM_t | I_{t-1}\right] = \mu_t$ and $\mathbb{V}\left[RM_t | I_{t-1}\right] = \mu_t^2 \nu^{-1}$, respectively. The form of μ_t is sufficiently flexible to include simple auto-regressive patterns, HAR terms as in Corsi (2009), asymmetry, or predetermined variables. Examples of possible specifications for μ_t are given, among others, in Engle and Gallo (2006) and Brownlees et al. (2012). Interestingly, the term μ_t induces the conditional variance of the realized measure to be time-varying, thus making the MEM consistent with the so-called volatility-of-volatility feature, studied in Corsi et al. (2008) among others. Recently, Gallo and Otranto (2012) extend the MEM to include time-varying parameters in the expression of μ_t as in the case of regime-switching MEM. The latter specification allows for changing parameters but requires to impose a priori structures on the form of the transition and on the number of underlying regimes. Differently, Haerdle et al. (2012) propose to adaptively estimate the MEM parameters based on a window of varying length thus providing updated parameter estimates at each point in time. The literature on multiplicative models of volatility already includes several extensions of the distributional assumptions of the baseline MEM. For example, Lanne (2006) proposes the mixture MEM to capture not only the long-memory dynamics of the realized volatility, but also its heavy tail marginal distribution as generated by the mixture of the two Gamma densities for the volatility innovations. Similarly, the possibility of fat tails has been considered by Lunde (1999) and Andres and Harvey (2012) that have adopted the generalized Gamma distribution in the ACD-MEM framework. Alternatively, the parameter ν can be made time-varying according to a GARCH-type law of motion, thus increasing the flexibility of the model in describing the higher moments of RM_t. As it will be shown in the empirical application, none of these distributional choices is able to assign the correct probability to the occurrence of the extremely high realizations that characterize the volatility dynamics. This further confirms the need for a novel modeling setup explicitly designed to capture the occurrence of such events.

3 A Multiplicative Error Model with Jumps

The baseline MEM with a Gamma distributed error term is poorly designated to assign the correct probability to the occurrence of large and abrupt movements, i.e. the jumps, that characterize the volatility dynamics. The presence and the effects of volatility jumps have been already documented in the literature either in a continuous time framework, as in Todorov and Tauchen (2011), among others, or in discrete time, see Caporin et al. (2014b). We propose a generalization of the MEM of Engle and Gallo (2006), which we call MEM-J. The new model introduces a multiplicative volatility jump term in the standard MEM of Engle and Gallo (2006). We also generalize the dynamic features of the MEM with the inclusion of HAR terms following Corsi (2009), we defer the discussion of this to Section 4. Under the MEM-J specification, the realized volatility measure RM_t equals the product of three elements

$$RM_t = \mu_t Z_t \varepsilon_t \tag{3}$$

where μ_t is a function measurable with respect to the information set at time $t-1$, Z_t is the volatility jump component, and the innovation ε_t is a scale-shape Gamma, $\varepsilon_t|I_{t-1} \sim \Gamma\left(\frac{1}{\nu}, \nu\right)$. Hereafter, to simplify the interpretation of the model outcome, the Gamma density of the innovation term is expressed in the mean-shape representation, i.e. $\varepsilon_t|I_{t-1} \sim \Gamma(1, \nu)$, which is, by construction, equivalent to the scale-shape representation. Hence, we require a number of assumptions on Z_t and ε_t to identify and separate the two sources of shocks. The jump term, Z_t, corresponds to

$$Z_t = \begin{cases} 1 & N_t = 0 \\ \sum_{j=1}^{N_t} Y_{j,t} & N_t > 0 \end{cases} \tag{4}$$

where N_t is a non-negative integer-valued random variable that represents the number of jumps occurring at time t. When $N_t = 0$, i.e. jumps are absent, the MEM-J reduces to the MEM. The random variable N_t determines the occurrence and the number of jumps; we model it as a Poisson with intensity λ,

$$P\left(N_t = m|I_{t-1}\right) = \frac{e^{-\lambda}\lambda^m}{m!}, \quad m = 0, 1, 2, ... \tag{5}$$

The second characterizing element of Z_t defines the size of the jumps. We set it equal to the sum of independent Gamma random variables, $Y_{j,t} \sim \Gamma(1, \varsigma)$ (in mean-shape form). Note that the jump density does not depend on time and the parameter characterizing the jump evolution is assumed to be time-invariant.

Assumption 1 *In the MEM-J in* (3)

i. ε_t *is an i.i.d. process defined on positive support with* $\mathbb{E}[\varepsilon_t] = 1$.

ii. ε_t, N_t *and the variables* $Y_{j,t}$, $j = 1, 2, \ldots, N_t$, *are assumed to be independent for any* t.

By the properties of the Gamma density, it follows that

$$Z_t | N_t = m > 0, I_{t-1} \sim \Gamma(m, m\varsigma) \tag{6}$$

in mean-shape representation. It is interesting to note that the mean and variance of the jump component depend on the number of jumps, i.e. $\mathbb{E}[Z_t | N_t = m > 0, I_{t-1}] = m$ and $\mathbb{V}[Z_t | N_t = m > 0, I_{t-1}] = \frac{m}{\varsigma}$. So far, all parameters are assumed to be time invariant. In Section 4 we discuss the possibility of time-varying parameters in the jump process, as a way to increase the capability of the model to adapt to the changing market conditions.

It follows from equation (3) that the MEM-J can be written as

$$RM_t = \mu_t \eta_t \tag{7}$$

where the innovation term $\eta_t = Z_t \varepsilon_t$ is the product of two shocks, one depending on jumps. In the next paragraphs we will study the properties of the conditional density of η_t and of RM_t which clearly depend on the distributional assumptions made on Z_t and ε_t.

3.1 The conditional density of η_t

The conditional density of η_t depends on N_t through Z_t. When $N_t = 0$, we have that $\eta_t | N_t = 0, I_{t-1}$ is simply equal to $\varepsilon_t | I_{t-1}$, since $Z_t = 1$. In this case, the conditional density of η_t, in mean-shape form, coincides with that of ε_t, i.e. $\Gamma(1, \nu)$. Differently, when $N_t = m > 0$, the conditional density of η_t given Z_t and I_{t-1} is Gamma in mean-shape form

$$\eta_t | (Z_t, N_t = m > 0, I_{t-1}) \sim \Gamma(Z_t, \nu). \tag{8}$$

In order to derive the conditional density of η_t given $N_t = m > 0$ and I_{t-1}, we have to evaluate the following integral:

$$\int_0^\infty f(\eta_t | N_t = m > 0, Z_t, I_{t-1}; \nu, \varsigma) f(z_t | N_t = m > 0, I_{t-1}; \nu, \varsigma) \mathrm{d}z, \tag{9}$$

where both conditional densities in the integral are Gamma. In the following proposition we present the conditional density of η_t given $N_t = m > 0, I_{t-1}$, i.e. the closed-form solution to the integral in (9).

Proposition 1 *Under Assumption 1, consider* $\eta_t = Z_t \varepsilon_t$ *where* Z_t *defined in* (4) *has the conditional density in* (6) *and* $\varepsilon_t | I_{t-1} \sim \Gamma(1, \nu)$*. Assuming that* Z_t *and* ε_t *are independent at all leads and lags, it follows that*

$$f(\eta_t | N_t = m > 0, I_{t-1}; \nu, \varsigma) = \frac{2}{\eta_t} (\eta_t \varsigma \nu)^{\frac{m\varsigma+\nu}{2}} \frac{1}{\Gamma(m\varsigma)\Gamma(\nu)} \mathbb{K}_{m\varsigma-\nu}\left(2\sqrt{\eta_t \varsigma \nu}\right), \tag{10}$$

where $\mathbb{K}_a(\cdot)$ *is the modified Bessel function of the second kind. Thus the innovation term* η_t*, conditional on* $N_t = m > 0$ *and* I_{t-1}*, has a* K *distribution, see Redding (1999), denoted as*

$$\eta_t | N_t = m > 0, I_{t-1} \sim K\big(m, m\varsigma, \nu\big).$$

The first two moments of η_t*, conditional on* $N_t = m > 0$ *and* I_{t-1}*, are*

$$\begin{aligned} \mathbb{E}\left[\eta_t | N_t = m > 0, I_{t-1}\right] &= m \\ \mathbb{V}\left[\eta_t | N_t = m > 0, I_{t-1}\right] &= m^2 \frac{m\varsigma + \nu + 1}{m\varsigma\nu} = m^2 \left(\frac{1}{\nu} + \frac{1}{m\varsigma} + \frac{1}{m\nu\varsigma}\right). \end{aligned}$$

Proof in Appendix 1.
The K density depends on three parameters which have specific meanings in the MEM-J. The first parameter is the mean of the K density, and it is equal to the number of jumps, m. The second parameter depends on the shape of the jump component Z_t, while the third is the shape parameter of the innovation term ε_t. Interestingly, the conditional variance of η_t is an increasing function of the number of jumps arrivals, m. Hence, periods with a larger number of jumps arrivals are characterized by a higher volatility-of-volatility. The Appendix A reports additional details on the K distribution.

The density of the innovation term, η_t, conditional only on the information set, I_{t-1}, is a countably infinite mixture

$$f(\eta_t | I_{t-1}; \nu, \varsigma, \lambda) = P\left(N_t = 0 | I_{t-1}\right) \Gamma(1, \nu) + \sum_{m=1}^{\infty} P\left(N_t = m | I_{t-1}\right) \times K\left(m, m\varsigma, \nu\right), \tag{11}$$

where

$$P\left(N_t = 0 | I_{t-1}\right) = e^{-\lambda}.$$

The mixing variable is the Poisson process N_t, which in turn depends on the parameter λ. As λ increases, more weight is given to the K distributions in the infinite sum, while when $\lambda = 0$ the density of η_t is $\Gamma(1, \nu)$ and the MEM-J reduces to the MEM.

3.2 The conditional density of RM_t

The conditional density of RM_t, given $N_t = m > 0$ and I_{t-1}, follows from the conditional distribution of η_t in equation (10). The following proposition presents the density of RM_t and the subsequent corollary introduces its conditional moments.

Proposition 2 *Consider model* (7) *where* $\eta_t = Z_t \varepsilon_t$ *with* Z_t *defined in equation* (4) *and* $\varepsilon_t | I_{t-1} \sim$

$\Gamma(1,\nu)$. *Assuming that Z_t and ε_t are independent at all leads and lags, it follows that*

$$f(RM_t|N_t = m > 0, I_{t-1}; \zeta, \varsigma) = \frac{2}{RM_t}\left(\frac{RM_t}{\mu_t}\varsigma\nu\right)^{\frac{m\varsigma+\nu}{2}} \frac{1}{\Gamma(m\varsigma)\Gamma(\nu)} \mathbb{K}_{m\varsigma-\nu}\left(2\sqrt{\frac{RM_t}{\mu_t}\varsigma\nu}\right), \quad (12)$$

Thus the realized measure RM_t, conditional on $N_t = m > 0$ and I_{t-1}, has a K distribution, denoted as

$$RM_t|N_t = m > 0, I_{t-1} \sim K\big(m\mu_t, m\varsigma, \nu\big).$$

The first two moments of RM_t, conditional on $N_t = m > 0$ and I_{t-1}, are

$$\begin{aligned} \mathbb{E}\left[RM_t|N_t = m > 0, I_{t-1}\right] &= \mu_t m, \\ \mathbb{V}\left[RM_t|N_t = m > 0, I_{t-1}\right] &= \mu_t^2 m^2 \frac{m\varsigma + \nu + 1}{m\varsigma\nu}. \end{aligned}$$

As a result, both the conditional mean and variance of RM_t are not only time-varying and driven by μ_t, as in the MEM, but also dependent on the realized number of jumps, m. On the other hand, when jumps are absent, i.e. $m = 0$, the conditional density $f(RM_t|N_t = 0, I_{t-1}; \zeta, \varsigma)$ is that of the MEM. Integrating out the realized number of jumps, the density of RM_t conditional on the information set I_{t-1} is a countably infinite mixture

$$f(RM_t|I_{t-1}; \theta) = P\left(N_t = 0|I_{t-1}\right)\Gamma\left(\mu_t, \nu\right) + \sum_{m=1}^{\infty} P\left(N_t = m|I_{t-1}\right) \times K\left(m\mu_t, m\varsigma, \nu\right), \quad (13)$$

where $\theta = [\zeta', \varsigma, \lambda]'$ is the vector of parameters of the MEM-J model. The conditional distribution of RM_t depends both on μ_t as well as on the jump intensity, λ. The expected value of Z_t can then be used to derive the expected value of the realized measure RM_t.[1] Integrating out the dependence on N_t, it is possible to obtain the expected value and the variance of RM_t with respect to the information set I_{t-1} only, see Section 4.2.

4 A persistent MEM-J with time-varying parameters

The volatility of financial returns are characterized by several dynamic and distributional features: high persistence, leverage effects, clusters of jumps and heteroskedastic effects in volatility. In this section, we show how the MEM-J presented in Section 3 can account for all these features.

4.1 Specification of μ_t

The importance of a correct specification of μ_t becomes clear when looking at the dynamics of the model residuals. Since volatility is characterized by a slow and hyperbolic decay of the autocorrelation function, it follows that a simple ARMA(1,1) specification, as implied by the MEM, is not enough to describe such a rich dynamic behaviour. As a consequence, the MEM residuals display significant autocorrelation. A successful and parsimonious approach to account for the (pseudo) long-memory property of the volatility series has been proposed by Corsi (2009) with the HAR model. The HAR is a long autoregressive model, subject to linear constraints,

[1]See Appendix B.3 for details on the derivation of the moments of Z_t.

designed to capture the persistence of the logarithm of realized volatility. We consider two alternative specifications for μ_t:

- Asymmetric HAR-MEM (AHAR-MEM):

$$\mu_t = \omega + \beta\mu_{t-1} + \alpha_1 RM_{t-1} + \alpha_2 RM_{t-1:t-5} + \alpha_3 RM_{t-1:t-21} + \gamma RM^-_{t-1} \tag{14}$$

  where $RM_{t-1:t-5} = \frac{1}{5}\sum_{j=1}^{5} RM_{t-j}$ and $RM_{t-1:t-21} = \frac{1}{21}\sum_{j=1}^{21} RM_{t-j}$ and all parameters are non-negative, and $RM^-_t \equiv RM_t \cdot \mathrm{I}\{r_t < 0\}$. This specification allows for the *leverage* effect, i.e. an asymmetric response of volatility to the sign of the returns (see Engle and Gallo, 2006). We can also consider a HAR structure on RM^-_{t-1}, but preliminary estimates do not lead to any significant improvement. It is obiviously possible to test whether the inclusion of the weekly and monthly volatility terms provide a significant improvement in fitting the volatility dynamics.

- Asymmetric MEM (A-MEM):

$$\mu_t = \omega + \alpha RM_{t-1} + \beta\mu_{t-1} + \gamma RM^-_{t-1} \tag{15}$$

  where all parameters are non negative. This model is nested in AHAR-MEM with the restriction $\alpha_2 = \alpha_3 = 0$. Further, the MEM is obtained simply setting $\gamma = 0$.

The inclusion of HAR dynamics into the specification of μ_t represents an alternative to the more sophisticated ways to model the long-range dependence in the MEM framework, as those in Lanne (2006) and Gallo and Otranto (2012). The main advantage of the HAR specification is that, despite it imposes *ad hoc* restrictions on the autocorrelation structure, it is able to account for the long memory behavior of the series with a limited number of free parameters and it has been proven to be successful in the log-linear context by Andersen et al. (2007), Bollerslev et al. (2009) and Ma et al. (2014). Recently, Audrino and Knaus (2014) has shown evidence, based on a LASSO regressions, that the HAR structure with daily, weekly and monthly factors may be subject to structural breaks during financial crises, but this issue is not addressed in the present paper.

4.2 Time-varying jump intensity

The specification of the AHAR-MEM-J is inherently limited given that the Poisson process governing the jumps arrival and the Gamma density characterizing the jump size are all driven by time invariant parameters. To increase the model flexibility we introduce time variation in the jump intensity parameter, λ, and define the AHAR-MEM-J-λ_t model. Instead, we maintain a time invariant jump size since preliminary evaluations of the proposed model show that letting this parameter to vary across time does not improve upon time-invariant specifications, but it only increases the computational burden associated with the model estimation. Nevertheless, if needed (and supported by the data), even the jump size can be made time-varying. We first specify the dynamic evolution of the parameter λ_t, for which we suggest the Auto Regressive Jump Intensity (ARJI) specification of Chan and Maheu (2002), Maheu and McCurdy (2004), within the GARCH-Jump context for stock returns, and Caporin et al. (2014b) for volatility jumps.

We show here that it is possible to adapt a similar modeling strategy in a MEM framework for the RM_t series. In particular, the jump intensity is assumed to follow:

$$\lambda_t = \phi_1 + \phi_2 \lambda_{t-1} + \phi_3 \xi_{t-1}, \tag{16}$$

where

$$\xi_t = \mathbb{E}\left[N_t | I_t\right] - \lambda_t = \sum_{m=0}^{\infty} mP\left(N_t = m | I_t\right) - \lambda_t. \tag{17}$$

The restrictions $\phi_1 > 0$ and $\phi_2 > \phi_3 > 0$ are sufficient to guarantee the positiveness of λ_t as in Chan and Maheu (2002). Note that the innovation term depends on the conditional probabilities of observing m jumps given the information set at time t, and those are determined following the hypothesis of having a Poisson process governing the number of jumps, see (5). However, as the conditioning set is different, those probabilities must be appropriately evaluated. We will discuss this issue in Section 4.4 when dealing with model estimation. From a distributional point of view, letting the mixing parameter λ to be dynamic implies that the conditional density of RM_t in (13) has a time-varying weight associated with the K densities. This provides an extremely flexible specification of the density of RM_t, which can be exploited to infer a precise probability of occurrence of tail events, see Section 6.[2] In a time-varying setup, the conditional moments of η_t and RM_t are given in the following proposition.

Proposition 3 *Consider model* (7) *where* $\eta_t = Z_t \varepsilon_t$ *with* Z_t *defined in equation* (4) *and* $\varepsilon_t | I_{t-1} \sim \Gamma(1, \nu)$*, with* λ_t *evolving as in* (16)*. Assuming that* Z_t *and* ε_t *are independent at all leads and lags, it follows that the first two moments of* RM_t *and* η_t *conditional on* I_{t-1} *are*

$$\mathbb{E}\left[\eta_t | I_{t-1}\right] = \left(e^{-\lambda_t} + \lambda_t\right), \tag{18}$$

$$\mathbb{E}\left[RM_t | I_{t-1}\right] = \mu_t \left(e^{-\lambda_t} + \lambda_t\right), \tag{19}$$

$$\mathbb{V}\left[\eta_t | I_{t-1}\right] = \left[\frac{\lambda_t}{\varsigma} + \lambda_t^2\right](1 + \nu^{-1}) + (e^{-\lambda_t} + \lambda_t)\left[1 + \nu^{-1} - e^{-\lambda_t} - \lambda_t\right], \tag{20}$$

$$\mathbb{V}\left[RM_t | I_{t-1}\right] = \mu_t^2 \left\{\left[\frac{\lambda_t}{\varsigma} + e^{-\lambda_t} + (\lambda_t + \lambda_t^2)\right](1 + \nu^{-1}) - (e^{-\lambda_t} + \lambda_t)^2\right\}. \tag{21}$$

Proof in Appendix 3.

The conditional expected value and variance of RM_t depend on the time-varying mean component as well as on the time-varying jump intensity, through the marginal moments of the jump term, Z_t. Compared to the baseline MEM, the conditional expectation of RM_t is inflated by a time-varying factor $\left(e^{-\lambda_t} + \lambda_t\right)$, which is never smaller than one by construction and acts as a boosting factor. Also the conditional variance of RM_t evolves over time as a function of both μ_t and λ_t, allowing for a larger degree of flexibility than the MEM to model the volatility-of-volatility features studied in Corsi et al. (2008) among others.

[2] Creal et al. (2013) derives the Generalized Autoregressive Score (GAS) representation for both the time-varying intensity Poisson process and the dynamic mixtures of models. We believe that an extension of the MEM-J model within the GAS framework is a natural advancement but this is left to future investigation.

4.3 Stationarity

We first provide the stationarity condition for the simple case of time-invariant jump intensity with μ_t specified as

$$\mu_t = \omega + \sum_{i=1}^{q} \alpha_i RM_{t-i} + \sum_{i=1}^{p} \beta_i \mu_{t-i}. \tag{22}$$

Theorem 1 *Let RM_t follow a strictly positive covariance stationary MEM-J in* (3) *with μ_t defined in* (22) *with $\omega > 0$. Under Assumption 1 and $\eta_t \sim$ i.i.d. with $\lambda > 0$, then*

$$(e^{-\lambda} + \lambda) \sum_{j=1}^{q} \alpha_j + \sum_{i=1}^{p} \beta_i < 1. \tag{23}$$

Conversely, if (23) *holds, the unique strictly stationary solution of* (3) *is a second-order stationary solution.*

Proof in Appendix B.5.

The introduction of the jump term in the MEM-J leads to a different stationarity condition than that relative to the baseline MEM of Engle and Gallo (2006). Indeed, a multiplicative term, depending on λ, appears in front of the *ARCH* coefficients, those capturing the impact of innovations on the mean-evolution. This is a consequence of the fact that the jump term, Z_t, is a constituent of the model innovations, η_t, and it is not persistent by construction when the jump intensity is constant. Interestingly, the larger the coefficient λ, the larger is the inflating factor, and the smaller is the stationarity region given the parameters α_i and β_i.

We now move to the more complex case of the time-varying jump intensity. In this case, we provide a sufficient condition for the stationarity of the MEM-J in (3) with density, conditional on η_t, defined in (12), where η_t is a sequence of random variables with time-varying intensity parameter λ_t defined as in (16). The process in (22) can be written in vector form (Markov representation):

$$z_t = b_t + A_t z_{t-1}. \tag{24}$$

with $A_t = A \odot E_t$ and E_t is a square matrix of dimension $p+q$, with η_t on the first row and ones elsewhere, see Appendix B.6.

Theorem 2 *Given the MEM-J in* (3) *with Assumption 1 and the processes for λ_t and μ_t specified as in* (16) *and* (22)*, respectively, a sufficient condition for the existence of a strictly stationary solution is*

$$\rho(A) < \exp(-E[\log[(p+q)(\eta_t + (p+q) - 1)]])$$

where $\rho(A)$ is the spectral radius of A (i.e. the greatest modulus of its eigenvalues).

Proof in Appendix B.6.

This second result is less intuitive than in the constant intensity case. Nevertheless, we stress that the sufficient condition depends both on the number of the parameters in the process for μ_t and on the expectation of the innovation η_t that combines the jump term and the error term ϵ_t.

4.4 Maximum likelihood estimation

The AHAR-MEM-J-λ_t can be estimated by maximum likelihood. Under the maintained assumption that $N_t|I_{t-1} \sim Poisson(\lambda_t)$ with N_t and ε_t independent processes, the conditional density of RM_t, $f(RM_t|I_{t-1})$, can be computed in closed form as in (13). Indeed, the density of the K distribution is known and it does not need to be simulated. Hence, the computation of the log-likelihood function is straightforward. The model parameters are estimated by maximizing the sample log-likelihood $\ell(\theta) = \sum_{t=1}^{T} \log f(RM_t|I_{t-1};\theta)$, where $\theta \in \Theta \subseteq \mathcal{R}_+^{11}$ is the vector of parameters for the AHAR-MEM-J-λ_t, $\theta = [\omega, \alpha_1, \alpha_2, \alpha_3, \beta, \gamma, \nu, \varsigma, \phi_1, \phi_2, \phi_3]'$. The log-likelihood function $\ell(\theta)$ is the log-transform of a mixture density. Since the conditional density of RM_t involves an infinite sum of densities, the sample log-likelihood function is computed for a finite number of jumps ($m = 0, 1, 2, \ldots, \bar{m}$). As it emerges from the Monte Carlo simulations, reported below, and the empirical application, any choice of $\bar{m}$ larger than 10 leads to almost identical parameter estimates. Indeed, for values of λ that are rarely larger than 1, the probability of observing more than 10 jumps is of order 10^{-8}. For the computation of log-likelihood of the AHAR-MEM-J-λ_t, we need the filtered probabilities $P(N_t = m|I_t)$ which are used to calculate the innovations ξ_t's in (17). The updating of the filtered probabilities is derived from Bayes rule as follows

$$P(N_t = m|I_t) = \frac{f(RM_t|N_t = m, I_{t-1}) \times P(N_t = m|I_{t-1})}{f(RM_t|I_{t-1})}, \quad m = 0, 1, 2, \ldots, \bar{m}. \tag{25}$$

In general, the mixture likelihood function can be unbounded, that is the function is characterized by the presence of singularities. Thus the ML estimator as global maximizer of the mixture likelihood function may not exist. Several local maximizers may exist for a given sample, and a major difficulty is to identify if the correct one has been found, see the discussion in Frühwirth-Schnatter (2006). Nevertheless, statistical theory outlined in Kiefer (1978) guarantees that a particular local maximizer of the mixture likelihood function is consistent, efficient, and asymptotically normal if the mixture is not overfitting. In the case of estimation of MixNormal-GARCH models, Ausín and Galeano (2007) and Bauwens et al. (2007) have devised a Bayesian estimation procedure to avoid such degenerated states. Alternatively, Broda et al. (2013) have proposed a method based on an augmented likelihood function. We don't adopt any of these computational devices since the Monte Carlo results, reported below, show that this problem is not a major concern in our case.

4.4.1 Monte Carlo Simulations

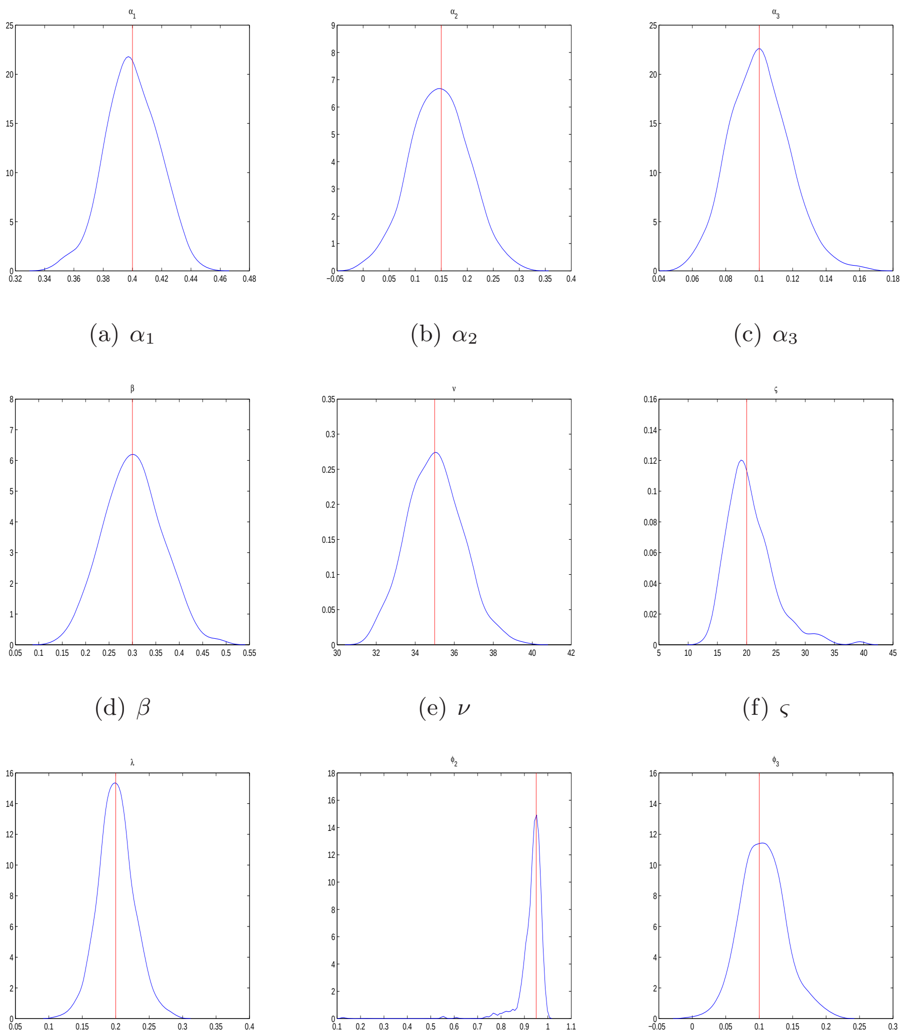
We run a set of Monte Carlo simulations to show that finite sample distribution of the ML estimates of the MEM-J parameters are centered on the true values and unimodal. We simulate three different specifications with the same μ_t as in (14) with no asymmetric effect: HAR-MEM (model in (1)), HAR-MEM-J with constant λ (model in (3)) and HAR-MEM-J-λ_t.

The algorithm to simulate pseudo-random variates from a K density is illustrated in Appendix A. The true parameter values and the corresponding Monte Carlo average estimates are reported in Table 1. The simulated sample size is set equal to 3000. Due to the computational burden in estimating the MEM-J the Monte Carlo replications are 500. We investigate the effects that the over-specification of the jump component can have on the maximum likelihood estimates. This can be a typical situation which arises when we have to specify nonlinear models with latent components. We estimate over-specified models (upper and middle panel of Table 1), i.e. the HAR-MEM-J-λ_t, when the data have been generated with either λ equal to zero (i.e. $\phi_1 = \phi_2 = \phi_3 = 0$) or with a constant λ. The infinite sum of densities required to compute the likelihood, see (11), is truncated at $\bar{m} = 10$. The same choice of $\bar{m}$ is later adopted in the empirical application. When the jumps are totally absent, the estimate of the unconditional mean of λ_t, i.e. $E[\lambda_t] = \phi_1/(1-\phi_2)$, is almost equal to zero on average, meaning that there is a very limited mixing effect in the conditional density of RM_t due to the estimated jump term. This means that the estimated model is very close to the HAR-MEM which is the DGP. Indeed, the parameters governing μ_t are correctly estimated and with small RMSE. It should also be noted that the parameter ς is not defined under the DGP, but is estimated when fitting the HAR-MEM-J-λ_t on the data as it determines the shape of the K distribution. This has no consequences on the parameters in μ_t, as they are all located around the true values. Indeed, as noted by Engle and Gallo (2006), the nuisance parameters governing the shape of the distribution of the innovation term do not impact on the estimates of the parameters of μ_t. When jumps are absent but the parameter ς is estimated, but bias and RMSE associated with this parameter are very high, which is a consequence of the lack of identification. The average estimate of ς is large, suggesting that the average size of the jumps is very small, as their expected size is the reciprocal of ς. Concluding, when the jumps are absent, and a MEM-J is fitted to the data (i.e. over-specification), the estimates of the parameters in μ_t are very close to the true values while the estimated jump component is negligible.

	ω	α_1	α_2	α_3	β	ν	ς	$\frac{\phi_1}{1-\phi_2}$	ϕ_2	ϕ_3
DGP: $\lambda = 0$										
	0.001	**0.4**	**0.15**	**0.1**	**0.3**	**20**	**-**	**-**	**-**	**-**
Mean	0.001	0.399	0.148	0.097	0.302	20.106	38.010	0.009	0.599	0.254
RMSE	0.000	0.021	0.069	0.023	0.076	0.577	40.972	0.033	0.636	0.334
DGP: $\lambda = 0.25$										
	0.001	**0.4**	**0.15**	**0.1**	**0.3**	**35**	**20**	**0.25**	**-**	**-**
Mean	0.001	0.400	0.152	0.099	0.296	34.957	20.625	0.250	0.479	0.019
RMSE	0.000	0.017	0.050	0.017	0.056	1.646	3.710	0.018	0.611	0.033
DGP: $\lambda_t > 0$										
	0.001	**0.4**	**0.15**	**0.1**	**0.3**	**35**	**20**	**0.2**	**0.95**	**0.1**
Mean	0.001	0.400	0.147	0.100	0.301	35.011	20.620	0.201	0.931	0.106
RMSE	0.000	0.018	0.056	0.018	0.062	1.394	4.051	0.027	0.060	0.034

Table 1: Monte Carlo results. The true parameter values used in simulation are in bold. Sample mean and Root mean squared error (RMSE) of maximum likelihood estimates of simulated HAR-MEM's models.

When λ is constant, but a HAR-MEM-J-λ_t is estimated, see the middle panel of Table 1, the distributions of the parameters in the DGP are centered on the true values, so that the impact of the over-specification is again very limited. For instance, if we look at the estimates of the HAR parameters, they seem unaffected by this over-specification. Moreover, the estimate of the parameter ϕ_3 is close to zero, meaning that the estimated variation in λ_t is almost absent as implied by the DGP. In the third case considered, i.e. the correctly specified model, the ML estimates have a very small finite sample bias and the RMSE's of $\phi_1/(1-\phi_2)$, ϕ_2 and ϕ_3 have the same order of magnitude of the HAR parameters. Figure 1 displays the kernel density estimates of the ML estimates based on the Monte Carlo simulations. The plots show that the finite sample distributions for all parameters are centered on the true values. Furthermore, the ML estimates of the HAR-MEM-J-λ_t model have Monte Carlo distributions that are well behaved with no evidence of multimodality that may be an indication of the presence of multiple local maxima.



(a) α_1 (b) α_2 (c) α_3

(d) β (e) ν (f) ς

(g) $\phi_1/(1-\phi_2)$ (h) ϕ_2 (i) ϕ_3

Figure 1: Kernel densities of the Monte Carlo estimates of the HAR-MEM-J-λ_t, where λ_t varies according to (16).

The Monte Carlo simulations confirm the validity of the ML estimation method, made possible by the knowledge of the closed-form expression of the K distribution, and show that the results are also valid when the jumps are absent, i.e. when the density of the innovation term is Gamma distributed. In the empirical application we rely on standard asymptotic results for the computation of the standard errors, as in Engle and Gallo (2006).

5 Empirical Results

5.1 Database and ex-post volatility estimation

Our purpose is to estimate the probability and the size of the volatility jumps once that *price jumps* have been disentangled from the volatility dynamics. Indeed, when price jumps are present, the total price variation, or *quadratic variation*, is equal to the sum of integrated variance plus the squared price jumps. The quadratic variation can be estimated by the realized variance (or realized volatility), as

$$RV_t = \sum_{j=1}^{M} r_{t,j}^2 \quad t = 1, ..., T \tag{26}$$

where $r_{t,j} \equiv p_{t,j} - p_{t,j-1}$ is the j-th intraday log-return on a fixed length grid with M intradaily observations. When $M \to \infty$ and microstructure noise is absent, the RV converges to the quadratic variation.

Disentangling the squared price jumps from the integrated variance is important when the focus is on the volatility dynamics. Indeed, as it has been noted by Huang and Tauchen (2005), jumps in prices account for approximately 7% of the total price variability. Barndorff-Nielsen and Shephard (2004) propose as an ex-post estimator of the integrated variance the bipower variation, defined as

$$BPV_t = \frac{\pi}{2} \sum_{j=2}^{M} |r_{t,j}||r_{t,j-1}| \quad t = 1, \dots, T. \tag{27}$$

BPV converges to the integrated variance as M diverges, also when the instantaneous volatility process has a jump component.

The empirical analysis reported in the following sections is conducted with the BPV series of two sets of assets. The first dataset includes seven stock indexes: S&P500, FTSE 100, DAX, DJIA, NASDAQ 100, CAC 40, Bovespa, sampled from January 3, 2000 through January 31, 2013, as made available by the Oxford-Man Institute's Realised Library. The second dataset consists of 16 large cap equities quoted on the New York market: Boeing, Bank of America, City Group, Caterpillar, Federal Express, Honeywell, Hewlett-Packard, IBM, JP Morgan, Kraft, Pepsi, Procter & Gamble, AT&T, Time Warner, Texas Instruments, and Wells Fargo. Prices are sampled at one minute frequency, from January 2, 2003 to June 30, 2012, and they are provided by TickData. The BPV is estimated from the 1-minute prices. In both datasets, the realized measure is expressed as daily volatility, i.e. the square root of BPV, $RM_t = \sqrt{BPV_t}$. The descriptive statistics of both datasets[3] highlight well known stylized facts such as high kurtosis and asymmetry, due to the long upper tail characterizing the empirical density of RM_t, and the presence of a strong serial correlation as suggested by the very high values of the auto-correlations at the selected lags 1, 5 and 22.

Several alternative multiplicative specifications are considered for modelling RM_t. We consider two alternative specifications of μ_t: the AMEM and the AHAR-MEM, where the former is nested in the latter. For what concerns the jump component, we consider the following cases:

[3] See Table 1 in the supplementary document.

- No jumps: $\phi_1 = \phi_2 = \phi_3 = \varsigma = 0$;
- Constant jump intensity: from equation (16) we set $\lambda_t = \phi_1$ and $\phi_2 = \phi_3 = 0$;
- Time-varying jump intensity: with λ_t evolving as in equation (16).

Due to space constraints, only the estimates of the parameter of the AHAR-MEM-J-λ_t are reported in the paper.[4] To compare the alternative models we consider two different approaches. Firstly, we pursuit a full-sample evaluation approach, where the MEM and MEM-J specifications are compared with respect to their fit on the empirical data and a series of statistical tests for restrictions on the parameters are performed. Secondly, we evaluate model abilities in fitting the upper tail of the realized measure both in-sample and out-of-sample. This is not only crucial for risk-management purposes, but also consistent with the expected ability of the MEM-J in capturing sudden and large increases in the volatility.

5.2 Estimation results

The alternative MEM specifications are first compared in terms of their ability in fitting the dynamics of the series. To this end, we analyze the dynamic properties of the residuals[5]

$$\hat{\varepsilon}_t = \frac{RM_t}{\mathbb{E}\left[RM_t | I_{t-1}\right]}. \tag{29}$$

Since the standard diagnostic statistics are designed for residuals that are assumed normally distributed, we normalize the residuals with $\widehat{\varepsilon}_t^* = F_N^{-1}\left[F_\Gamma(\widehat{\varepsilon}_t)\right]$, for $t = 1, \ldots, n$, where $F_N()$ and $F_\Gamma()$ are the cumulative density functions of the standard normal and Gamma distributions, respectively. Table 2 reports the diagnostic statistics and tests for the estimated models: the AMEM, AHAR-MEM, AHAR-MEM-J and AHAR-MEM-J-λ_t. The AMEM does not account for the persistence present in RM_t, as the Ljung-Box tests on the residuals strongly reject the null hypothesis in nearly all cases. On the contrary, the Ljung-Box statistics of the AHAR-MEM residuals do not reject the null of no residual autocorrelation in 4 out of the 7 stock indexes considered, and only when we focus on lags up to the 22-nd. Looking at the individual stocks, at the 5% confidence level we have only 4 out of the 16 equities with some evidence of residual serial correlation, and only over 22 lags. The number of stocks with autocorrelated residuals of the AHAR-MEM decreases to 1 at 1% significance level.

From the theoretical analysis in Section 3, the inclusion of jumps in the MEM specification is mainly designed to provide a high degree of flexibility to the conditional density of RM_t while the dynamic features of the model are not affected. This is confirmed by the the Ljung-Box statistics that are mainly unaffected by the inclusion of the jump component in the AHAR-MEM model.

[4] The parameter estimates of the other sub-models are in the Tables 2-4 in the supplementary document.

[5] As an alternative, model residuals might be computed by standardization of RM_t with respect to its expected value, involving the impact of μ_t and (when present) Z_t. In this case, innovations are defined as Pearson's residuals

$$\frac{RM_t - \mathbb{E}\left[RM_t | I_{t-1}\right]}{\mathbb{V}\left[RM_t | I_{t-1}\right]^{1/2}}. \tag{28}$$

	AMEM			AHAR-MEM				AHAR-MEM-J				AHAR-MEM-J-λ_t			
	Q_1	Q_{10}	Q_{22}	Q_1	Q_{10}	Q_{22}	LR_{α_2,α_3}	Q_1	Q_{10}	Q_{22}	$LR_{\lambda=0}$	Q_1	Q_{10}	Q_{22}	LR_{ϕ_2,ϕ_3}
SP500	0.004	0.000	0.000	0.9042	0.4782	0.3882	70.29^a	0.7856	0.4161	0.3145	160.7	0.3521	0.2656	0.2197	38.14
FTSE 100	0.001	0.000	0.000	0.9106	0.6457	0.0119	60.46^a	0.1217	0.6439	0.0016	356.1	0.0099	0.2004	0.0003	46.66
DAX	0.000	0.000	0.000	0.2201	0.1974	0.0000	60.40^a	0.0040	0.0420	0.0000	156.0	0.0007	0.0021	0.0000	40.44
DJIA	0.003	0.000	0.000	0.9133	0.4444	0.5854	57.96^a	0.4610	0.3420	0.4822	162.3	0.1050	0.1101	0.2587	47.00
NSDQ	0.001	0.000	0.000	0.7336	0.1068	0.0944	99.88^a	0.4407	0.1010	0.0563	148.3	0.1816	0.0700	0.0283	38.09
CAC	0.000	0.000	0.000	0.3166	0.0889	0.0000	53.05^a	0.0130	0.0475	0.0000	187.0	0.0001	0.0015	0.0000	56.81
BOVESPA	0.007	0.000	0.000	0.7588	0.6229	0.2247	80.81^a	0.0860	0.2011	0.0660	153.4	0.0232	0.0766	0.0298	27.41
BA	0.0387	0.0003	0.0002	0.7975	0.7728	0.0449	37.44^a	0.6809	0.9054	0.0349	221.3	0.7242	0.9225	0.0431	26.30
BAC	0.0832	0.0000	0.0005	0.7118	0.6958	0.6892	59.75^a	0.2802	0.8761	0.7291	417.4	0.7242	0.9225	0.0431	26.30
C	0.0370	0.0000	0.0005	0.6971	0.7133	0.2519	47.02^a	0.0187	0.0695	0.0834	255.6	0.0004	0.0028	0.0112	106.29
CAT	0.0046	0.0000	0.0020	0.8485	0.6575	0.8894	44.30^a	0.6661	0.5964	0.7196	210.3	0.2384	0.4199	0.6357	23.97
FDX	0.0227	0.0000	0.0000	0.7435	0.1739	0.4699	60.45^a	0.8995	0.2012	0.3635	263.6	0.5496	0.1314	0.3446	25.09
HON	0.0516	0.0022	0.0033	0.7512	0.8410	0.1363	23.80^a	0.1232	0.4163	0.0273	177.2	0.0257	0.2464	0.0204	48.14
HPQ	0.0390	0.0000	0.0004	0.7678	0.4510	0.4235	56.44^a	0.9266	0.6977	0.3117	341.4	0.7540	0.6728	0.3541	63.04
IBM	0.0968	0.0071	0.0260	0.9893	0.9638	0.5213	22.40^a	0.2993	0.7714	0.3067	165.1	0.0218	0.2105	0.1295	62.64
JPM	0.0039	0.0000	0.0000	0.6510	0.8022	0.3092	53.64^a	0.0524	0.2157	0.1139	223.6	0.0524	0.2157	0.1139	31.39
KFT	0.0030	0.0000	0.0003	0.6344	0.5964	0.9161	59.10^a	0.1221	0.3387	0.6590	431.5	0.0457	0.2383	0.6070	35.68
PEP	0.3856	0.0229	0.0003	0.6252	0.3843	0.0052	13.83^a	0.5895	0.3141	0.0028	260.1	0.0208	0.0455	0.0007	64.31
PG	0.0953	0.0005	0.0003	0.4232	0.9140	0.0508	41.37^a	0.6694	0.8454	0.0130	266.9	0.3166	0.5629	0.0057	37.96
T	0.0234	0.0000	0.0001	0.8620	0.6676	0.0327	34.94^a	0.5551	0.6199	0.0105	225.5	0.0960	0.2222	0.0051	63.91
TWX	0.0698	0.0024	0.0044	0.7001	0.9871	0.4933	40.95^a	0.6941	0.8620	0.3519	310.8	0.2432	0.6658	0.3053	37.17
TXN	0.0527	0.0018	0.0006	0.4013	0.8918	0.1558	40.59^a	0.2234	0.7120	0.1184	148.7	0.4644	0.8732	0.1554	26.78
WFC	0.4878	0.0000	0.0002	0.0834	0.4063	0.3548	65.20^a	0.3035	0.7847	0.5023	459.9	0.7835	0.9313	0.5639	40.06

Table 2: Residual Statistics and Tests. The upper part of the table reports the results for several stock indexes while the lower part refers to 16 NYSE stocks. a, b and c stand for significance at 1%, 5% and 10% respectively. Q_1, Q_{10} and Q_{22} are the p-values of the Ljung-Box test for absence of autocorrelation in the normalize residuals, where the latter are computed as $\widehat{\varepsilon}_t^* = F_N^{-1}\left[F_\Gamma(\widehat{\varepsilon}_t)\right]$ with $\hat{\varepsilon}_t = \frac{RM_t}{E[RM_t|I_{t-1}]}$. LR_{α_2,α_3} is the value of the LR test for the nullity of α_2 and α_3 in the AHAR-MEM model. $LR_{\lambda=0}$ is the value of the LR test of the AHAR-MEM-J model with constant λ vs the model without jumps (i.e. with $\lambda = 0$). Since the parameter ς is not defined under the null, it is not possible to associate to the test a significance level. LR_{ϕ_2,ϕ_3} is the likelihood ratio test for the nullity of ϕ_2 and ϕ_3 in the AHAR-MEM-J-λ_t.

The estimated parameters in μ_t of the AHAR-MEM-J are close to those of the AHAR-MEM even though the parameters associated with the jumps, ς and λ are statistically significant for all series considered.[6] Since the parameter ς is not defined in the AHAR-MEM, i.e. under the null hypothesis ($\lambda = 0$), it is not possible to evaluate the significance of the jump term using the quantiles of the χ^2 distribution as in the standard LR test, see the discussion in Hansen (1996).[7] However, from a comparison of the likelihood values in Table 2 ($LR_{\lambda=0}$), it clearly emerges that the log-likelihood functions of the AHAR-MEM-J are much larger than those of the AHAR-MEM, as their difference is often larger than 100. This evidence supports the hypothesis of a mixture of Gamma and K distributions in the innovation term as originated by the presence of jumps.

Table 3 reports the parameter estimates of AHAR-MEM-J-λ_t model. The parameters in μ_t are strongly significant in almost all cases, similarly to the estimates of AHAR-MEM and AHAR-MEM-J. In particular, if we compare the estimated parameters of the stocks to those of the indexes, we note that the stocks are characterized by a somewhat higher impact of previous day RM_t levels, i.e. coefficient α_1. Differently, the impact of last week and last month average of RM_t is more heterogeneous across stocks, with some cases of reduced significance.

For what concerns the estimates of the parameters in the jump component, the unconditional mean of λ_t is between 0.15 and 0.20 in most cases, and there are not relevant differences between stock indexes and individual stocks. Indeed, obtaining significant coefficients for the jump intensity, either in the constant or in the dynamic specification, is a first evidence that jumps in volatility are a significant component of the variability of RM_t. Interestingly, most markets and stocks, among those considered, display estimates of ϕ_2 larger than 0.8, suggesting persistence in jump arrivals. Notable exceptions are FTSE-100, BOVESPA and CAT. For BOVESPA the sensitivity to the news arrival, measured by the parameter ϕ_3, is close to the persistence parameter, i.e. ϕ_2. This might suggest that the time-varying jump intensity specification is not needed in this case. The LR tests for the joint nullity of ϕ_2 and ϕ_3 in the last column of Table 2 take always very large values. Even though an asymptotic theory for the LR test is not available, when ϕ_2 is not defined under the null, we believe that the observed values of the test statistic can reasonably lead to the rejection of the null hypothesis in all cases considered. Introducing dynamics in the jump intensity is empirically important as it provides the necessary degree of flexibility in the conditional density of RM_t.

Looking at the scale parameter estimates, ν, they are much larger for the AHAR-MEM-J-λ_t than for the AHAR-MEM model without jumps. The variance of ε_t, equal to $\frac{1}{\nu}$, sensibly reduces when jumps are included. As shown in (20), for a given level of the conditional variance of η_t and a given arrival probability λ_t, there is an inverse relationship between ν and ς. Moreover, since ς is much smaller than ν in all cases, it follows that the variance of the jump terms is several times larger than that of ε_t. As a consequence, both a smooth and a discontinuous component, responsible for big moves, must be included when modeling volatility in a discrete time setting. This is in accordance with the findings of Todorov and Tauchen (2011) who show

[6]See the supplementary document.

[7]In this case simulation based approaches can be used to recover likelihood ratio test critical values following Hansen (1996). We don't pursue that strategy due to the computational burden implied in the estimation of the MEM-J.

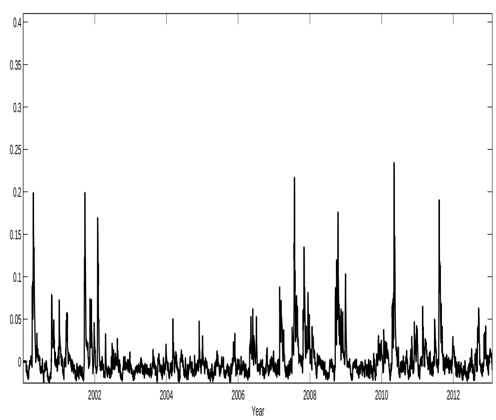
	ω	α_1	α_2	α_3	β	γ	ν	ς	$\frac{\phi_1}{1-\phi_2}$	ϕ_2	ϕ_3
SP500	0.0003^a	0.2913^a	0.1694^a	0.1182^a	0.3162^a	0.1061^a	22.9479^a	13.9452^a	0.1757^a	0.9408^a	0.1930^a
FTSE 100	0.0002^a	0.2748^a	0.2189^a	0.1648^a	0.2705^a	0.0564^a	26.5095^a	9.7599^a	0.0424^a	0.7145^c	0.3581^b
DAX	0.0002^a	0.2886^a	0.1415^b	0.1356^a	0.3653^a	0.0699^a	26.2173^a	12.1181^a	0.1760^a	0.9193^a	0.2146^a
DJIA	0.0003^a	0.2868^a	0.1688^a	0.1127^a	0.3307^a	0.0946^a	22.4206^a	12.9378^a	0.1761^a	0.9489^a	0.2248^a
NSDQ	0.0003^a	0.3003^a	0.1888^a	0.1587^a	0.2597^a	0.0973^a	23.6406^a	10.2317^a	0.1585^a	0.9634^a	0.1665^a
CAC	0.0002^a	0.2774^a	0.1365^b	0.1152^a	0.3978^a	0.0770^a	27.5657^a	10.8710^a	0.1630^a	0.9012^a	0.2313^a
BOVESPA	0.0009^a	0.3127^a	0.1226^b	0.1705^a	0.2687^a	0.0700^a	22.6130^a	17.4224^a	0.2147^a	0.3461^a	0.3455^a
BA	0.0006^a	0.3664^a	0.0893	0.1424^a	0.3322^a	0.0245^a	35.8425^a	19.0657^a	0.1904^a	0.8863^a	0.1048^b
BAC	0.0004^a	0.4785^a	0.0937^c	0.1371^a	0.2334^a	0.0346^a	33.6804^a	19.4741^a	0.1945^a	0.9814^a	0.0986^a
C	0.0003^a	0.4483^a	0.1164^b	0.1069^a	0.2876^a	0.0218^a	41.6322^a	13.6892^a	0.1541^a	0.9911^a	0.1105^a
CAT	0.0008^a	0.3901^a	0.2122^a	0.1395^a	0.1718^a	0.0419^a	33.8332^a	21.4409^a	0.1696^a	0.5204	0.2761^b
FDX	0.0005^a	0.3742^a	0.1433^a	0.1946^a	0.2224^a	0.0348^a	33.6948^a	22.4715^a	0.2008^a	0.8536^a	0.1597^a
HON	0.0007^a	0.3665^a	0.0947	0.1054^a	0.3505^a	0.0493^a	33.8111^a	19.3674^a	0.1631^a	0.9065^a	0.2174^a
HPQ	0.0008^a	0.3892^a	0.1621^a	0.1445^a	0.2137^a	0.0437^a	33.9732^a	17.4021^a	0.1697^a	0.8364^a	0.1549^a
IBM	0.0006^a	0.3947^a	0.1263^c	0.0791^a	0.3220^a	0.0298^a	36.6859^a	17.7869^a	0.1490^a	0.9819^a	0.1292^c
JPM	0.0005^a	0.4541^a	0.1832^a	0.1361^a	0.1824^a	0.0322^a	33.1899^a	26.7587^a	0.1918^a	0.9878^a	0.0548
KFT	0.0011^b	0.3065^a	0.2625^a	0.2090^a	0.0846	0.0012	30.0529^a	27.2279^a	0.2694^a	0.9673^a	0.0791^a
PEP	0.0003^a	0.2846^a	0.0275	0.0765	0.5545^a	0.0341^a	34.5631^a	21.7868^a	0.1710^a	0.8060^a	0.2754^a
PG	0.0007^a	0.3705^a	0.2044^a	0.1263^a	0.1975^a	0.0298^a	36.6526^a	15.1394^a	0.1541^a	0.9483^a	0.1342^a
T	0.0005^a	0.3682^a	0.1356^a	0.1256^a	0.3060^a	0.0246^a	37.1412^a	16.5691^a	0.1539^a	0.9723^a	0.1493^a
TWX	0.0006^a	0.3757^a	0.1253^c	0.1431^a	0.2888^a	0.0261^a	45.4493^a	23.1275^a	0.1539^a	0.7662^a	0.2464^a
TXN	0.0009^a	0.3625^a	0.1785^b	0.1390^a	0.2463^a	0.0296^a	34.3740^a	19.1028^a	0.1306^a	0.9704^a	0.1152^a
WFC	0.0004^a	0.4281^a	0.1665^b	0.1531^a	0.1926^a	0.0201^b	35.7612^a	16.2251^a	0.2086^a	0.9266^a	0.1493^a

Table 3: Estimates of the Asymmetric HAR-MEM-J with time-varying λ, see (16). The upper part of the table reports the results for several stock indexes while the lower part refers to 16 NYSE stocks. a, b and c stand for significance at 1%, 5% and 10% respectively.

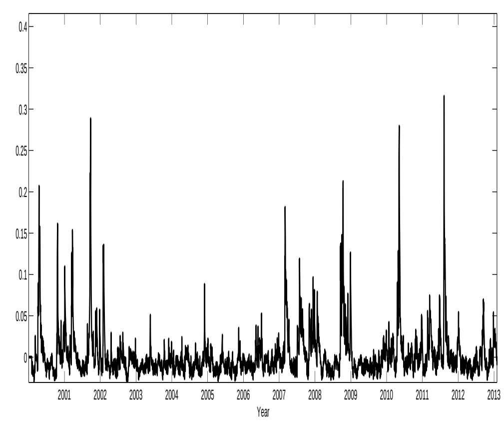
that volatility can be well approximated by a pure jump process with infinite variation. Finally, The estimates of ν for the individual stocks are sensibly higher than those of the indexes. This reflects differences in the volatility-of-volatility as that of the indexes is higher than that of the individual stocks. This may be due to the different sample periods under exam.

Figures 2 report two examples of fitted expected jump component $\mathbb{E}\left[Z_t|I_{t-1}\right] = (e^{-\lambda_t} + \lambda_t) \geq 1$. The expected jump (which is a non-linear function of λ_t) is very close to 1, i.e. there are no jumps, when the markets experience no major shocks, e.g. from 2003 through 2007. Instead, it sharply increases during market turmoils, like: the end of technology market bubble in 2001-2002, the sub-prime crisis in 2007-2008 and the European sovereign crisis in 2010. Notably, the most recent crisis seem to be more relevant in France compared to the US market, a somewhat expected result.

Figure 3 dislpay the impact of a change in the model structure, moving from an AHAR-MEM without jumps to a specification including jumps, AHAR-MEM-J-λ_t. The figure shows the relative difference between the expected realized measure, i.e. $\mathbb{E}\left[RM_t|I_{t-1}\right]$, of AHAR-MEM-J-$\lambda_t$ and AHAR-MEM. Again, during market turmoils, the model with jumps has conditional expected values of RM_t that are from 5% to 25% larger than those obtained with a model without jumps.



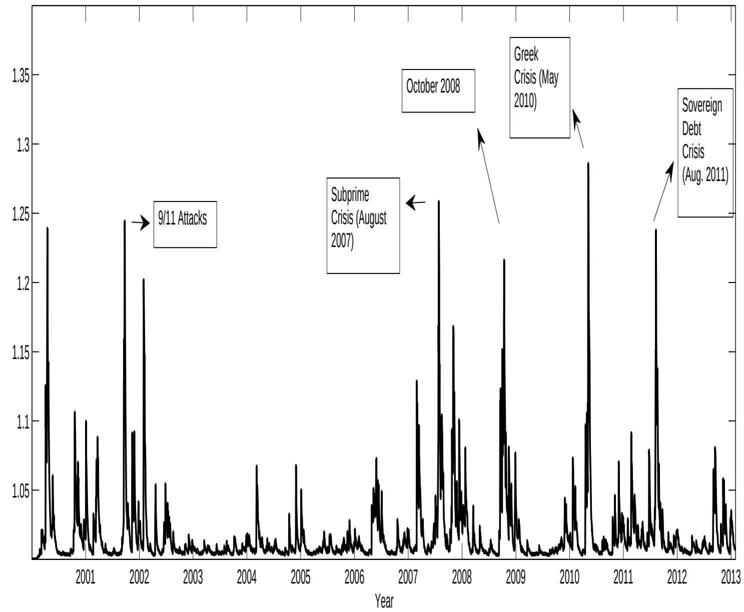
(a) $S\&P500$



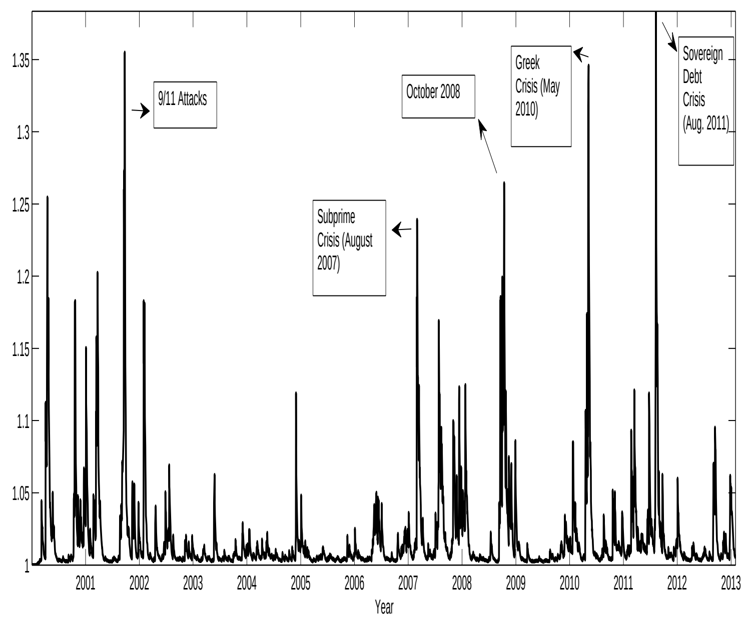
(b) CAC

Figure 3: Relative expected jump contribution. The figures display the ratio $\frac{\mathbb{E}_J[RM_t|I_{t-1}]-\mathbb{E}_0[RM_t|I_{t-1}]}{\mathbb{E}_0[RM_t|I_{t-1}]}$, where $\mathbb{E}_J[RM_t|I_{t-1}]$ is the conditional expectation under AHAR-MEM-J-λ_t, and $\mathbb{E}_0[RM_t|I_{t-1}]$ is the conditional expectation under AHAR-MEM.

The same evidence arises from Figure 4 which reports the ratio between the conditional variance of RM_t of the model with jumps and without jumps. The ratio is generally larger than 1, meaning that the conditional variance generated by jumps is by larger than that generated without jumps. Especially during periods of high volatility the ratio takes extremely large values, thus confirming that jumps are an important factor of variation in these periods. This can be also interpreted as a further evidence in favour that the variation in volatility can be attributed to a combination of continuous and discontinuous processes.

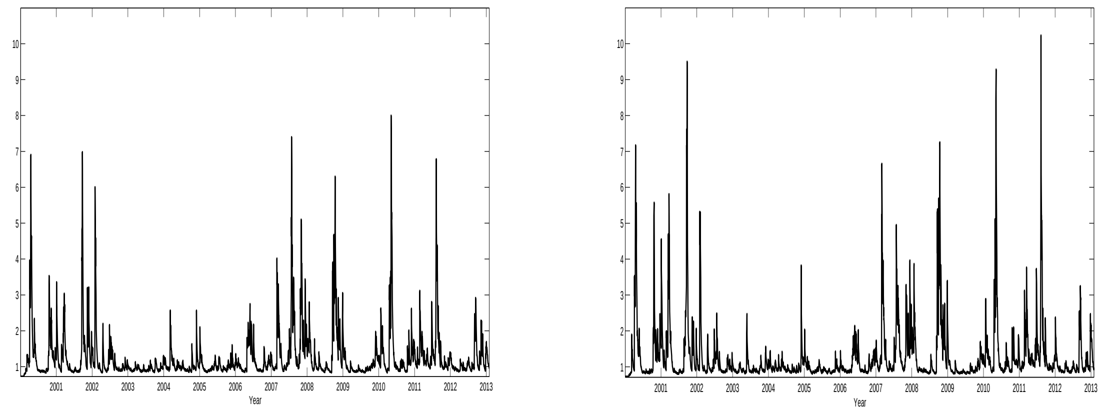


(a) $E(Z_t|I_{t-1})$ of S&P500



(b) $E(Z_t|I_{t-1})$ of CAC

Figure 2: Expected jump component in the AHAR-MEM-J-λ_t model for S&P500 and CAC 40.



(a) $S\&P500$ (b) CAC

Figure 4: Volatility-of-volatility ratio. The figures display the ratio $\frac{\mathbb{V}_J[RM_t|I_{t-1}]}{\mathbb{V}_0[RM_t|I_{t-1}]}$, where $\mathbb{V}_J[RM_t|I_{t-1}]$ is the conditional variance under AHAR-MEM-J-λ_t, and $\mathbb{V}_0[RM_t|I_{t-1}]$ is the conditional variance under AHAR-MEM.

5.3 Robustness checks

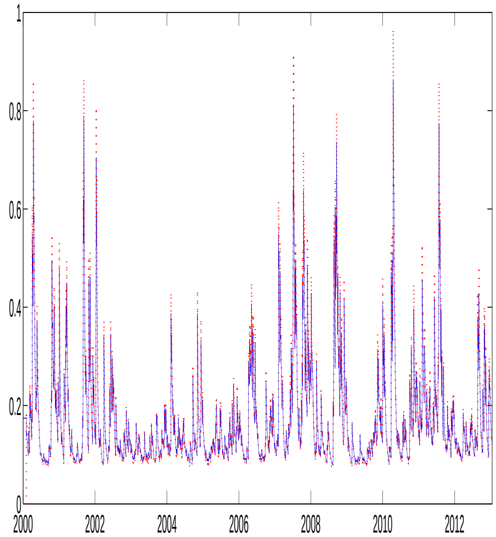
A series of robustness checks have been carried out to evaluate to what extent the empirical results are affected by the measurement error and by the potential low power of BPV in disentangling jump in prices from the volatility dynamics. First, the presence of the measurement error can be accounted for by means of a parametric conditional model that includes an equation for the returns and a measurement equation.[8] The estimates of the MEM-J equation parameters change up to a very limited extent. This is perhaps due to the very liquid nature of the indexes and individual stocks used in the empirical application. In Figure 5, it is evident the two estimated seeequences of λ_t are very close, meaning that the measurement error is not a major concern and it can be considered negligible for the purposes of this paper.

Furthermore, another obvious concern is whether the results on the estimation of volatility jumps depend on the particular realized measure employed in the analysis. Indeed, Christensen et al. (2014) show that BPV may be biased when the underlying volatility process is characterized by high volatility-of-volatility, so that volatility jumps can be easily confused as price jumps. We have therefore estimated the MEM-J model on another set of volatility series that is robust to price jumps. In particular, we estimated AHAR-MEM-J-λ_t with $RM_t = \sqrt{MedRV_t}$, i.e. the median realized volatility computed as follows ,

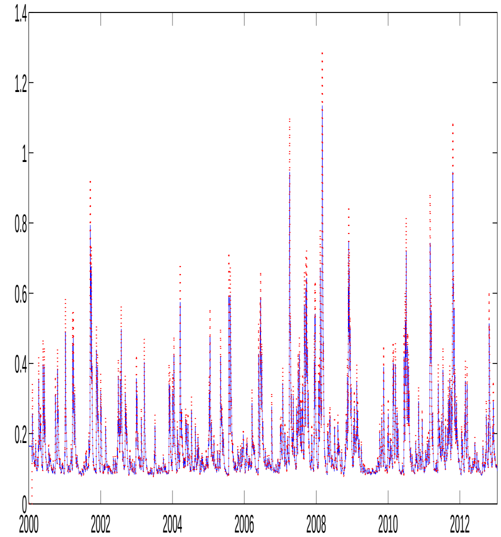
$$MedRV_t = \frac{\pi}{6 - 4\sqrt{3} + \pi}\left(\frac{M}{M-2}\right) \times \sum_{j=2}^{M-1} med(|r_{t,j-1}|, |r_{t,j}|, |r_{t,j+1}|)^2$$

that is less efficient than BPV but more powerful in filtering the price jumps, see Andersen

[8]Due to space constraints, an exhaustive description of this model, that can be seen as an extension of the realized GARCH model of Hansen et al. (2012), with volatility and price jumps and multiplicative measurement equation, can not be outlined here. Therefore, a full discussion of the properties of this multivariate model is left to future research. The estimation results are in Table 5 of the document with supplementary material.



(a) SP500



(b) CAC

Figure 5: Estimated time-varying volatility jump intensity, λ_t, obtained with the univariate AHAR-MEM-J-λ_t, solid line, and with the bivariate AHAR-MEM-J-λ_t based on daily returns and $\sqrt{BPV}$, dotted line, (see details in the Supplementary document).

et al. (2012a). Notably, this replacement does not lead to any significant change in the AHAR-MEM-J-λ_t parameters and in the estimates of the dynamics of the volatility jump intensity, as shown in Figure 6.[9]

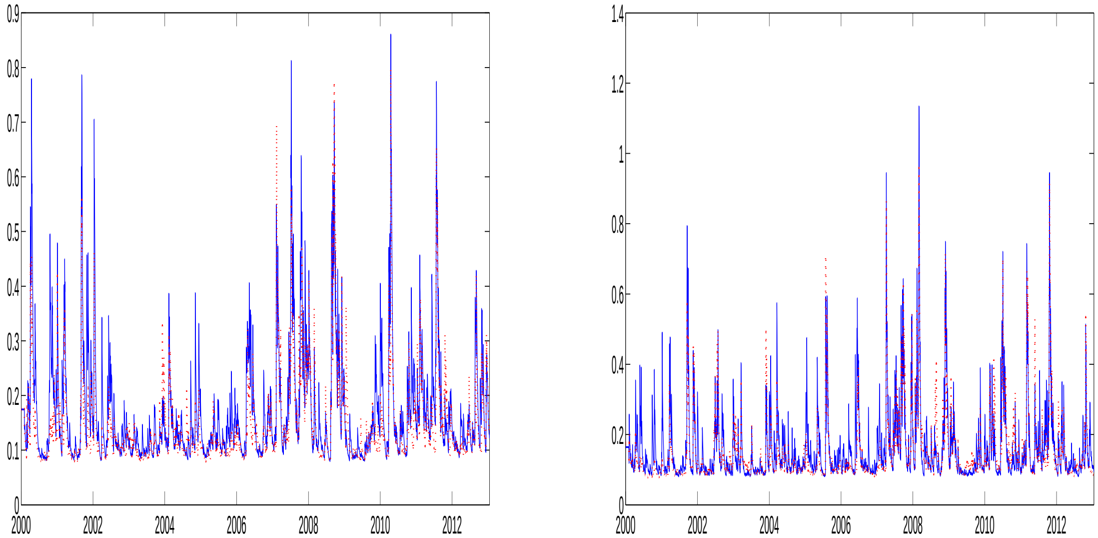
6 Volatility-at-Risk

In this section, we evaluate the ability of the MEM specifications considered in this study to correctly predict the probability of tail events. The model with volatility jumps is expected to provide a better description of tail events, i.e. extreme volatility realizations, as it is able to generate large and sudden increases in the conditional volatility-of-volatility levels, thus providing a better fitting of the upper tail. By analogy to the Value-at-Risk introduced for quantifying the risk of extremely negative returns, we define the VolaR, i.e. the risk of extreme high volatility as

$$\Pr\{RM_t > v(\alpha)|I_{t-1}\} = \alpha$$

where $\Pr\{\cdot|I_{t-1}\}$ denotes the conditional distribution at date t of the one-step-ahead volatility, whereas $v(\alpha)$ is the realized volatility level that may occur with probability α. The VolaR might be of interest for investors who trade in volatility, see Zhang et al. (2010), Euan (2013) and therein cited references. In fact, the knowledge of the probability that volatility will exceed a given threshold is useful both in designing volatility trading strategies based on options (allowing for an optimal calibration of the option maturity as well as the option strike) and for strategies

[9] See Table 6 in the Supplementary document.



(a) SP500 (b) CAC

Figure 6: Estimated time-varying volatility jump intensity, λ_t, obtained estimating the AHAR-MEM-J-λ_t with $RM_t = \sqrt{BPV_t}$, solid line, and with $RM_t = \sqrt{MedRV_t}$, dotted line.

based on volatility indices or exchange traded volatility products (having an impact on the choice of the investment direction as well as on the size of the position). In addition, the evaluation of volatility risk might be of interest for options traders and market makers to define optimal prices and order execution, and, finally, to portfolio managers willing to determine the need and the amount of a volatility hedge.

In order to evaluate the estimation of the VolaR (i.e. the right tail coverage) obtained with models with and without jumps, we consider the method introduced by Berkowitz (2001), which allows to test for the adequacy of the proposed density with the realization of the modeled variable. The test is flexible and can be applied to the fit of the entire density as well as over specific segments of the density support. For our purposes, we apply the test over the upper $q\%$ tail of the RM_t density. In details, given the density of the RM_t, we compute the conditional CDF of RM_t

$$y_t = F\left(RM_t | I_{t-1}\right) = \int_0^{RM_t} f\left(x | I_{t-1}\right) \mathrm{d}x,$$

where $F(RM_t|I_{t-1})$ for the AHAR-MEM-J is given by the mixture of Gamma and K conditional CDFs. Then, under correct model specification, the empirical CDF values should be distributed according to the standard uniform, i.e. $y_t \sim U\left(0,1\right)$, which are further transformed as

$$z_t = \Phi^{-1}\left(y_t\right)$$

where $\Phi\left(\cdot\right)$ is the standard normal CDF, so that z_t are distributed as a standardized normal. To test the correct tail coverage, we choose a VolaR level of 1% (i.e. VolaR = 2.3263), and calculate a new truncated variable

$$z_t^* = \begin{cases} \text{VolaR} & \text{if} \quad z_t \leq \text{VolaR} \\ z_t & \text{if} \quad z_t > \text{VolaR}. \end{cases} \tag{30}$$

A tail coverage test can be derived using the LR principle. Under the null, the mean and the variance of z_t^* are 0 and 1, respectively, while under the alternative they are unrestricted. Under the null of correct tail coverage the test statistic is distributed as $\chi^2(2)$. See Berkowitz (2001) for further details on this test.

Table 4 reports the p-values of the Berkowitz test relative to the in-sample estimates of different volatility model specifications. Beyond the MEM specifications seen so far, we also include in the comparison models which are based on more flexible innovation density specifications than the Gamma distribution adopted in the baseline MEM. These are characterized by fatter right tail than the Gamma distribution and are expected to accommodate extreme realizations observed in the realized measures series. The models considered are:

- AHAR-MEM-GG: where ϵ_t follows a Generalized Gamma, as in Lunde (1999) and Andres and Harvey (2012).
- AHAR-MEM-$\bar{\nu}_t$: where the variance of ε_t, i.e. the parameter $\bar{\nu}_t = \frac{1}{\nu_t}$, follows a GARCH(1,1) process.
- M-AHAR-MEM: that is the mixture model of Lanne (2006) with AHAR dynamics in each volatility component.

Due to space constraints, the specifications and the parameter estimates for these models are reported in the Supplementary document.[10] We also estimate the HAR-V-J of Caporin et al. (2014b) to have a model with jumps which is not based on a multiplicative structure.

It clearly emerges that the AHAR-MEM with jumps outperforms the corresponding specification without jumps in estimating the VolaR. All the MEM specifications without jumps and Gamma distribution for ε_t strongly reject the null hypothesis of correct specification of the upper quantiles. It should be noted that also the AHAR-MEM-$\bar{\nu}_t$ model is poorly designed to capture tail events. Letting the conditional variance of ε_t to be time varying is not sufficient for a proper characterization of VolaR. This suggests a distinct role of the jumps from pure heteroskedastic effects in ε_t. Interestingly, the M-AHAR-MEM of Lanne (2006) provides some evidence of correct specification of the upper tail, as the Berkowitz test cannot reject the null hypothesis in 8 cases. Conversely, the AHAR-MEM-GG model fails to give the correct probability mass on the right tail. In other words, despite the generalized Gamma distribution provides a good fitting for the entire distribution, it fails to properly account for the probability of tail events.

It is noteworthy that the introduction of jumps, with constant and time-varying λ, provides a good fit of the VolaR. In only two cases, BOVESPA and PG, the presence of jumps does not succeed in correctly estimating the VolaR. The introduction of the time-varying jump intensity sensibly improves the performances for the CAC40 index, BA, HPQ and KRF. A possible explanation for the good performance of the model with jumps can be derived from Figure 7.

[10]See Tables 7-9 in the Suppplementary document

	I	II	III	IV	V	VI	VII	VIII
SP500	0.0000	0.0000	0.0000	0.0080	0.0000	0.5181	0.6749	0.3456
FTSE 100	0.0000	0.0000	0.0000	0.4882	0.0000	0.7394	0.3289	0.4896
DAX	0.0000	0.0000	0.0000	0.0241	0.0001	0.6660	0.1415	0.1337
DJIA	0.0000	0.0000	0.0000	0.0010	0.0000	0.4189	0.4334	0.5880
NSDQ	0.0000	0.0000	0.0000	0.6016	0.0000	0.6651	0.5216	0.5999
CAC	0.0000	0.0000	0.0000	0.0013	0.0004	0.0984	0.0677	0.3594
BOVESPA	0.0000	0.0000	0.0000	0.5952	0.0519	0.8380	0.0055	0.0145
BA	0.0000	0.0000	0.0000	0.0229	0.0000	0.5486	0.0423	0.1483
BAC	0.0000	0.0000	0.0000	0.0268	0.0000	0.0008	0.1095	0.2011
C	0.0000	0.0000	0.0000	0.0045	0.0000	0.0217	0.1608	0.6656
CAT	0.0000	0.0000	0.0000	0.0921	0.0000	0.5137	0.2131	0.2528
FDX	0.0000	0.0000	0.0000	0.8566	0.0000	0.5309	0.0746	0.0815
HON	0.0000	0.0000	0.0000	0.0013	0.0000	0.0369	0.2604	0.2923
HPQ	0.0000	0.0000	0.0000	0.5133	0.0000	0.0621	0.0601	0.1515
IBM	0.0000	0.0000	0.0000	0.0038	0.0000	0.5957	0.3150	0.0909
JPM	0.0000	0.0000	0.0000	0.0856	0.0000	0.0059	0.1312	0.1218
KFT	0.0000	0.0000	0.0000	0.0112	0.0000	0.6487	0.0000	0.2042
PEP	0.0000	0.0000	0.0000	0.0000	0.0000	0.6652	0.1350	0.9984
PG	0.0000	0.0000	0.0000	0.0123	0.0000	0.9482	0.0267	0.0290
T	0.0000	0.0000	0.0000	0.0078	0.0000	0.5549	0.5883	0.7226
TWX	0.0000	0.0000	0.0000	0.0023	0.0000	0.0487	0.4656	0.3480
TXN	0.0000	0.0000	0.0000	0.0001	0.0000	0.5385	0.1078	0.2782
WFC	0.0000	0.0000	0.0000	0.4959	0.0000	0.0128	0.1071	0.0930

Table 4: P-values of the Berkowitz (2001) test for the in-sample VolaR at 1% level, corresponding to a value of 2.3263. The models considered are AMEM (I), AHAR-MEM (II), AHAR-MEM-ν_t (III), M-AHAR-MEM (IV), AHAR-MEM-GG (V), HAR-V-J (VI), AHAR-MEM-J (VII) and AHAR-MEM-J-λ_t (VIII).

The model with jumps is able to generate large and sudden spikes in the conditional variance of RM_t, as generated by the jump component Z_t, while the model with time-varying ν_t, can only generate smooth trajectories, and hence it is not able to assign enough probability to extreme volatility events. Interestingly, also the HAR-V-J model of Caporin et al. (2014b), which is an HAR specification with time-varying jump intensity on $\log RM_t$, provides a good fitting of the tails of the volatility distribution.[11] In this case, the null hypothesis cannot be rejected at 5% significance level for all the indexes and many of the individual stocks.

Table 5 reports the p-values of the Berkowitz test based on the out-of-sample forecasts, for a total of 1,000 observations (for the indexes the holdout sample starts on February 2, 2009, while for the individual stocks it starts in July 14, 2008). The sample size of the rolling window used to estimate the parameters in the cases of stock indexes is about 2,200 while for the individual stocks is approximately 1,400. Given the sample dimensions used in the estimation, the uncertainty in the parameters estimates is expected a fairly limited effect on the test results. When AMEM and AHAR are used, the null hypothesis is always rejected, at 5% significance level, with the exception of FTSE 100. Better results are obtained when the Generalized Gamma is adopted. In 5 out of 23 cases, the null hypothesis cannot be rejected. A similar performance is also achieved

[11] Due to space constraints, we do not report a complete discussion of the HAR-V-J model, which can be found in Caporin et al. (2014b). It is however important to note, that, since the HAR-V-J model is linear in the logarithms of realized volatility, this implies a multiplicative structure for the latter, similar to that obtained under the AHAR-MEM-J.

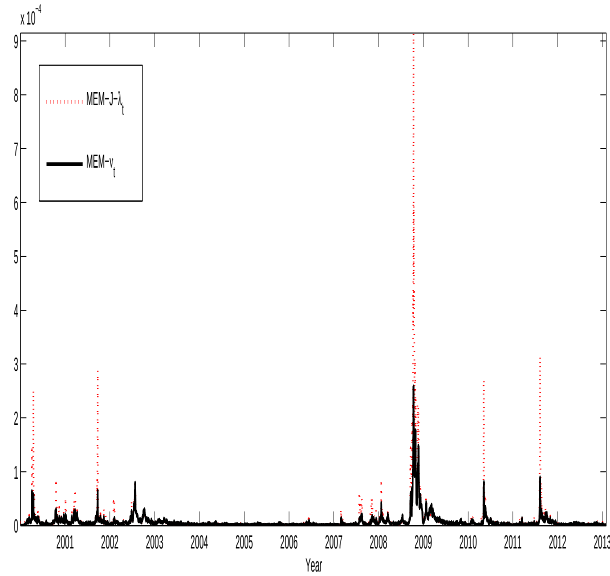


Figure 7: Conditional variance of RM_t of $S\&P500$ obtained with the AHAR-MEM-J-λ_t, dashed line, and with the AHAR-MEM-ν_t, solid line.

with M-AHAR-MEM and AHAR-MEM-J with constant λ. Indeed, by including the jumps (with constant intensity) the null is not rejected for 11 cases. A slightly better performance is achieved with the HAR-V-J model as the null hypothesis cannot be rejected in 13 out of 23 cases. An impressive improvement is instead associated with the full model with persistence and time-varying jump intensity. For the AHAR-MEM-J-λ_t the lowest p-value is associated with the DAX index and equals 12%. The out-of-sample results confirm the adequacy of AHAR-MEM-J-λ_t in predicting the presence of volatility jumps which turn out to be of crucial importance in forecasting the VolaR.

7 Concluding remarks

We have introduced a new model for realized volatility measures, the AHAR-MEM-J. Our model generalizes the MEM of Engle and Gallo (2006) by adding persistence (through HAR terms, see Corsi, 2009) and multiplicative volatility jumps. A volatility jump takes the form of an extreme event, for example a very large value of the daily volatility such as those observed in the last years. By specifying the volatility process as a combination of a continuous volatility component and a discrete compound Poisson process for the jumps, the conditional density of the realized measure becomes a countably infinite mixture constituted by a Gamma random variable and a weighted sum of K distributed random variables. We add further flexibility considering both time-varying jump intensity. This flexible parametrization of the dynamics of the realized measure allow capturing the extreme or abnormal movements in the volatility level. We discuss the model estimation in finite samples and the effects of misspecification by resorting to a Monte Carlo simulation. The empirical application shows that the AHAR-MEM-J-λ_t captures the extreme moves registered in the last years in the volatilities of individual stocks and equity indexes. We provide statistical evidence that, for the sample period analyzed,

	I	II	III	IV	V	VI	VII	VIII
S&P500	0.0000	0.0000	0.0000	0.0035	0.0002	0.3503	0.0048	0.5133
FTSE 100	0.5577	0.1526	0.0000	0.0615	0.5078	0.0000	0.3551	0.4917
DAX	0.0010	0.0002	0.0000	0.0436	0.4529	0.1246	0.0171	0.1212
DJIA	0.0000	0.0000	0.0000	0.0009	0.0001	0.0260	0.0358	0.5111
NSDQ	0.0000	0.0000	0.0000	0.0025	0.0000	0.0010	0.0293	0.9532
CAC	0.0000	0.0000	0.0000	0.0696	0.1039	0.2384	0.0711	0.4825
BOVESPA	0.0279	0.0117	0.0000	0.3168	0.3036	0.8389	0.4567	0.2591
BA	0.0000	0.0000	0.0000	0.0477	0.0002	0.3503	0.2482	0.4850
BAC	0.0000	0.0000	0.0000	0.9360	0.0000	0.0010	0.0185	0.0768
C	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0043	0.1416
CAT	0.0000	0.0000	0.0000	0.0231	0.0003	0.0631	0.0262	0.3891
FDX	0.0000	0.0000	0.0000	0.0362	0.0000	0.1076	0.1889	0.6195
HON	0.0000	0.0000	0.0000	0.0002	0.0006	0.6167	0.6935	0.1595
HPQ	0.0000	0.0000	0.0000	0.0416	0.0000	0.1246	0.0142	0.3614
IBM	0.0000	0.0000	0.0000	0.0000	0.0000	0.8587	0.7541	0.7128
JPM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0260	0.0453	0.1329
KFT	0.0000	0.0000	0.0000	0.3501	0.0000	0.0007	0.7700	0.6569
PEP	0.0000	0.0000	0.0000	0.0356	0.0001	0.0010	0.5746	0.5788
PG	0.0000	0.0000	0.0000	0.0119	0.0000	0.0000	0.3249	0.9246
T	0.0000	0.0000	0.0000	0.0140	0.0000	0.2384	0.7679	0.1859
TWX	0.0000	0.0000	0.0000	0.1001	0.0000	0.0409	0.0117	0.2544
TXN	0.0000	0.0000	0.0000	0.0031	0.0069	0.8389	0.0000	0.0055
WFC	0.0000	0.0000	0.0000	0.0000	0.0000	0.5799	0.0121	0.0002

Table 5: P-values of the Berkowitz (2001) test for the out-of-sample VolaR at 1% level, corresponding to a value of 2.3263. The out-of-sample forecasts are computed with a rolling window starting in February 2, 2009 for the stock indexes and July 14, 2008 for the individual stocks. The total number of forecasts is 1,000 for both data sets. The models considered are AMEM (I), AHAR-MEM (II), AHAR-MEM-ν_t (III), M-AHAR-MEM (IV), AHAR-MEM-GG (V), HAR-V-J (VI), AHAR-MEM-J (VII) and AHAR-MEM-J-λ_t (VIII).

the model correctly predicts the probability of occurrence of abnormal volatility levels, i.e. of jumps. We compare alternative models by means of a new measure called the volatility-at-risk, i.e. the risk of extreme high volatility. The empirical analysis put in evidence how models that cannot generate sudden and large movements in the realized measures, i.e. without jumps, fail in fitting the extreme right tail of the distribution. Moreover, the recent empirical evidence on the contemporaneous correlation between jumps in price and volatility would suggest an extension of our set up to include the presence of price jumps. This is left for future research. Finally, the potential application of this model is not limited to the study of volatility but it can be employed in the analysis of any positive time series that features persistence and sudden large variations, e.g. trading volume and durations.

Acknowledgements. We wish to thank Eric Ghysels, Christian Gourieroux, Andrew Harvey, Roberto Renò, Esther Ruiz, and Jean-Michel Zakoïan, for stimulating discussions and comments. We also thank the participants to the seminar at CREST, the SoFiE-INET conference on Skewness, Heavy Tails, Market Crashes, and Dynamics (Cambridge, 2014), the XV Workshop in Quantitative Finance (Florence, 2014), the 13th OxMetrics User Conference (Aarhus, 2013), the IAAE 2014 Conference (London, 2014), the International Symposium on Forecasting (Rotterdam, 2014), for their valuable suggestions. All errors and omissions are our own. Massimiliano Caporin acknowledges financial support from: the European Union, Seventh Framework Program FP7/2007-2013 under grant agreement SYRTO-SSH-2012-320270; the MIUR PRIN project MISURA - Multivariate Statistical Models for Risk Assessment; the Global Risk Institute in Financial Services and the Louis Bachelier Institute under the project *Systemic Risk*. Eduardo Rossi acknowledges financial support from the MIUR PRIN project MISURA - Multivariate Statistical Models for Risk Assessment. Paolo Santucci de Magistris acknowledges financial support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

References

Andersen, T., Dobrev, D., and Schaumbur, E. (2012a). Robust volatility estimation using nearest-neighbor truncation. *Journal of Econometrics*, 169:75–93.

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. *The Review of Economics and Statistics*, 2007:701–721.

Andersen, T. G., Bondarenko, O., and Gonzalez-Perez, M. T. (2012b). Uncovering novel features of equity-index return dynamics via corridor implied volatility. Technical report.

Andres, P. and Harvey, A. (2012). The dynamic location/scale model: with applications to intra-day financial data. Technical report, Faculty of Economics, University of Cambridge.

Audrino, F. and Knaus, S. (2014). Lassoing the HAR model: A model selection perspective on realized volatility dynamics. *Econometric Reviews*, Forthcoming.

Ausín, M.-C. and Galeano, P. (2007). Bayesian estimation of the gaussian mixture GARCH model. *Computational Statistics & Data Analysis*, 51(5):2636 – 2652.

Bandi, F. and Renò, R. (2012). Time-varying leverage effects. *Journal of Econometrics*, 169:94–113.

Bandi, F. and Renò, R. (2013). Price and volatility co-jumps. Technical report, University of Siena.

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2:1–37.

Bauwens, L., Hafner, C., and Rombouts, J. (2007). Multivariate mixed normal conditional heteroskedasticity. *Computational Statistics & Data Analysis*, 51(7):3551 – 3566.

Bauwens, L. and Veredas, D. (2004). The stochastic conditional duration model: a latent factor model for the analysis of financial durations. *Journal of Econometrics*, 119:381–412.

Berkowitz, J. (2001). The accuracy of density forecasts in risk management. *Journal of Business and Economic Statistics*, 19:465–474.

Bollerslev, T., Kretschmer, U., Pigorsch, C., and Tauchen, G. E. (2009). A discrete-time model for daily S&P 500 returns and realized variations: Jumps and leverage effects. *Journal of Econometrics*, 150:151–166.

Bougerol, P. and Picard, N. (1992). Stationarity of GARCH processes and some nonnegative time series. *Journal of Econometrics*, 52:115–127.

Broadie, M., Chernov, M., and Johannes, M. (2007). Model specification and risk premia: Evidence from futures options. *The Journal of Finance*, 62:1453–1490.

Broda, S. A., Haas, M., Krause, J., Paolella, M. S., and Steude, S. C. (2013). Stable mixture GARCH models. *Journal of Econometrics*, 172(2):292 – 306.

Brownlees, C. T., Cipollini, F., and Gallo, G. M. (2012). Multiplicative error models. In Bauwens, L., Hafner, C., and Laurent, S., editors, *Handbook of Volatility Models and Their Applications*, pages 223–247. John Wiley & Sons, Inc.

Caporin, M., Kolokolov, A., and Renò, R. (2014a). Multi-jumps. Technical report, MPRA.

Caporin, M., Rossi, E., and Santucci de Magistris, P. (2014b). Volatility jumps and their economic determinants. *CREATES RP2014-25, Forthcoming on Journal of Financial Econometrics*.

Chan, W. H. and Maheu, J. M. (2002). Conditional jump dynamics in stock market returns. *Journal of Business and Economic Statistics*, 20(3):377–389.

Chernov, M., Gallant, R., Ghysels, E., and Tauchen, G. (2003). Alternative models for stock price dynamics. *Journal of Econometrics*, 116:225–257.

Christensen, K., Oomen, R. C., and Podolskij, M. (2014). Fact or friction: Jumps at ultra high frequency. *Journal of Financial Economics*, 114(3):576 – 599.

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

Corsi, F., Mittnik, S., Pigorsch, C., and Pigorsch, U. (2008). The volatility of realized volatility. *Econometric Reviews*, 27:46–78.

Corsi, F., Pirino, D., and Renò, R. (2010). Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics*, 159(2):276 – 288.

Creal, D., Koopman, S. J., and Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, 28(5):777–795.

Duffie, D., Pan, J., and Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68:1343–1376.

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(12):3–27.

Eraker, B. (2004). Do stock prices and volatility jump? Reconciling evidence from spot and option prices. *Journal of Finance*, 59(2):1367–1403.

Eraker, B., Johannes, M. S., and Polson, N. G. (2003). The impact of jumps in returns and volatility. *Journal of Finance*, 53(3):1269–1300.

Euan, S. (2013). *Volatility Trading*. Wiley Trading series. John Wiley & Sons, Inc., Hoboken (NJ), second edition.

Franq, C. and Zakoian, J.-M. (2010). *GARCH Models*. John Wiley & Sons, Inc., Chichester, UK.

Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Spriger, New York.

Gallo, G. M. and Otranto, E. (2012). Realized volatility and change of regimes. Technical report, Dipartimento di statistica "G. Parenti",Università di Firenze.

Ghysels, E., Gouriéroux, C., and Jasiak, J. (2004). Stochastic volatility duration models. *Journal of Econometrics*, 119(2):413 – 433.

Haerdle, W., Hautsch, N., and Mihoci, A. (2012). Local adaptive multiplicative error models for high-frequency forecasts. Technical Report 31, Humboldt -Universitat zu Berlin.

Hansen, B. (1994). Autoregressive conditional density estimation. *International Economic Review*, 35:705–730.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, 64(2):413–30.

Hansen, P. R., Huang, Z., and Shek, H. H. (2012). Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, 27(6):877–906.

Huang, X. and Tauchen, G. (2005). The relative contribution of jumps to total price variance. *Journal of Financial Econometrics*, 3:456–499.

Jones, C. (2003). Nonlinear mean reversion in the short-term interest rate. *Review of Financial Studies*, 16:765–791.

Kiefer, N. (1978). Discrete parameter variation: Efficient estimation of a switching regression model. *Econometrica*, 46:427–434.

Lanne, M. (2006). A mixture multiplicative error model for realized volatility. *Journal of Financial Econometrics*, 4(4):594–616.

Lunde, A. (1999). A generalized Gamma Autoregressive Conditional Duration model. Technical report, Department of Economics, Politics and Public Administration, Aalborg University.

Ma, F., Wei, Y., Huang, D., and Chen, Y. (2014). Which is the better forecasting model? A comparison between HAR-RV and multifractality volatility. *Physica A: Statistical Mechanics and its Applications*, 405(C):171–180.

Maheu, J. M. and McCurdy, T. H. (2004). News arrival, jump dynamics, and volatility components for individual stock returns. *Journal of Finance*, LIX:755–793.

Pan, J. (2002). The jump-risk premia implicit in options: evidence from an integrated time-series study. *Journal of Financial Economics*, 63(1):3 – 50.

Redding, N. J. (1999). Estimating the parameters of the K distribution in the intensity domain. Technical report, DSTO Electronics and Surveillance Laboratory, South Australia.

Todorov, V. and Tauchen, G. (2011). Volatility jumps. *Journal of Business and Economic Statistics*, 29:356–371.

Zhang, J. E., Shu, J., and Brenner, M. (2010). The new market for volatility trading. *Journal of Futures Markets*, 30(9):809–833.

A The K distribution

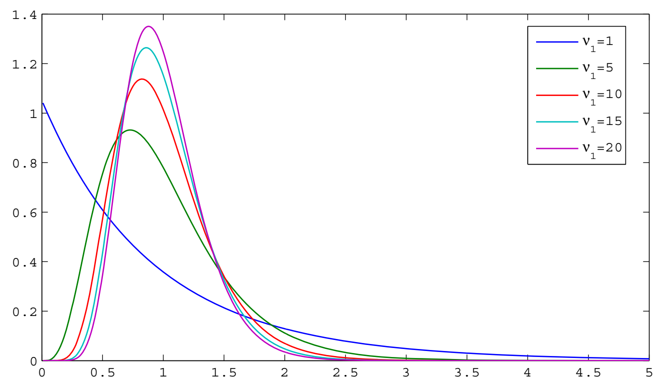
If $X \sim \Gamma(\mu, \nu_1)$, in mean-shape form, and $Y|X \sim \Gamma(X, \nu_2)$ then we have $Y \sim K(\mu, \nu_1, \nu_2)$ such that

$$f(y) = \frac{2}{y}\left(y\frac{\nu_1\nu_2}{\mu}\right)^{\frac{\nu_1+\nu_2}{2}} \frac{1}{\Gamma(\nu_1)\Gamma(\nu_2)}\mathbb{K}_{\nu_1-\nu_2}\left(2\sqrt{y\frac{\nu_1\nu_2}{\mu}}\right), \quad y \geq 0$$

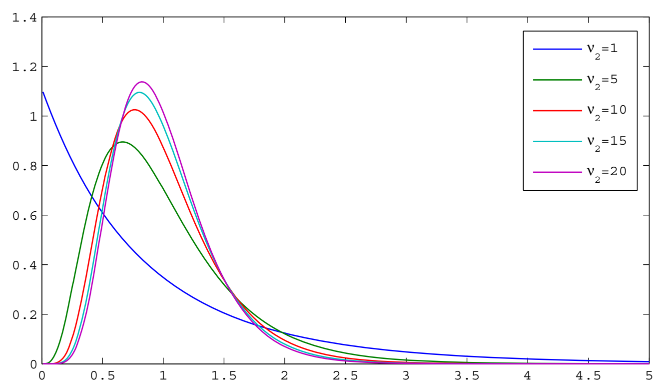
where $\mathbb{K}_a(\cdot)$ is the modified Bessel function of the second kind. Moments of the K density are given by

$$\mathbb{E}[y^s] = \frac{\mu^s\Gamma(\nu_1+s)\Gamma(\nu_2+s)}{\nu_1^s\nu_2^s\Gamma(\nu_1)\Gamma(\nu_2)} \tag{31}$$

so that $\mathbb{E}[Y] = \mu$ and $\mathbb{V}[Y] = \mu^2\left(\frac{\nu_1+\nu_2+1}{\nu_1\nu_2}\right)$. Furthermore, if $Y \sim K(\mu, \nu_1, \nu_2)$ then $\alpha Y \sim K(\alpha\mu, \nu_1, \nu_2)$, $\mathbb{E}[\alpha Y] = \alpha\mu$ and $\mathbb{V}[\alpha Y] = \alpha^2\mu^2\left(\frac{\nu_1+\nu_2+1}{\nu_1\nu_2}\right)$. Different K densities, corresponding to different values of ν_1 and ν_2, are plotted in Figure 8. Increasing both parameters reduces the variance, as it is apparent in both plots.



(a) K density calculated with $\nu_1 = \{1, 5, 10, 15, 20\}$ and $\nu_2 = 10$.



(b) K density calculated with $\nu_2 = \{1, 5, 10, 15, 20\}$ and $\nu_1 = 10$.

Figure 8: K density computed for different values of ν_1 (upper panel) and ν_2 (lower panel).

Further, as noted by Redding (1999, p.3), the product of two independent Gamma random variables, $Z \sim \Gamma(1, \nu_2)$ and $X \sim \Gamma(\mu, \nu_1)$, is

$$Y = Z \cdot X \sim K(\mu, \nu_1, \nu_2)$$

with density given by the following integral

$$f_Y(y) = \int_{-\infty}^{\infty} \frac{1}{x} f_Z\left(\frac{y}{x}\right) f_X(x) \, \mathrm{d}\, x.$$

Pseudo random numbers with K distribution can be generated from $Z \sim \Gamma(1, L)$ and $X \sim \Gamma(\mu, \nu)$, since $Y = Z \cdot X$ is distributed as $K(\mu, \nu, L)$.

Setting $\theta_1 = \nu_1, \theta_2 = \nu_2$ or viceversa, the cumulative distribution function (CDF) of a K-distributed random variable can be written as

$$F(y; \mu, \theta_1, \theta_2) = \frac{2^{2-\theta_1-\theta_2}}{\Gamma(\theta_1)\Gamma(\theta_1)} \int_0^{2\sqrt{\theta_1\theta_2 y/\mu}} t^{\theta_1+\theta_2-1} \mathbb{K}_{\theta_1-\theta_2}(t) \mathrm{d}t, y \geq 0. \tag{32}$$

The hypothesis of $\theta_2 \in \mathbb{N}$ instead of $\theta_2 \in \mathbb{R}_+$ is required in order to obtain the previous expression in closed form. Writing $\zeta = \theta_1 - \theta_2$, $k = 2\nu_2 - 1$ and $z = 2\sqrt{\theta_1\theta_2 y/\mu}$

$$F(y; \mu, \theta_1, \theta_2) = 1 + \frac{2^{2-\theta_1-\theta_2}}{\Gamma(\theta_1)\Gamma(\theta_2)} g(z, \zeta, k) \tag{33}$$

where

$$g(y, \zeta, k) = \begin{cases} -z^{\zeta+1}\mathbb{K}_{(\zeta+1)}(z) & k = 1 \\ (k-1)(2\zeta + k - 1) g(y, \zeta, k-2) - z^{\zeta+k}\mathbb{K}_{\zeta+1}(z) & \text{elsewhere} \\ -(k-1) z^{(\zeta+k-1)}\mathbb{K}_{\zeta}(z) & \end{cases}$$

The number of required recursions to compute a single value $F_Y(y; \mu, \theta_1, \theta_2)$ is θ_2. The best parametrization, in terms of computational speed, is $\theta_1 = \max\{\nu_1, \nu_2\}$ and $\theta_1 = \min\{\nu_1, \nu_2\}$.

B Proofs and results

B.1 Proof of Proposition 1

Given $\varepsilon_t | I_{t-1} \sim \Gamma(1, \nu)$ and $Z_t | N_t = m > 0, I_{t-1} \sim \Gamma(m, m\varsigma)$, integrating out z_t we have the conditional density of η_t

$$f(\eta_t | N_t = m > 0, I_{t-1}; \nu, \varsigma) = \frac{2}{\eta_t} \left(\eta_t \varsigma \nu\right)^{\frac{m\varsigma+\nu}{2}} \frac{1}{\Gamma(m\varsigma)\Gamma(\nu)} \mathbb{K}_{m\varsigma-\nu}\Big(2\sqrt{\eta_t \varsigma \nu}\Big).$$

which is the K density, see Redding (1999). $\mathbb{K}_a(\cdot)$ is the modified Bessel function of the second kind. The moments of η_t, conditional on $N_t = m > 0$ and I_{t-1}, are derived from the moments of the K density in (31). ■

B.2 Proof of Proposition 2

From equation (10), the conditional density of RM_t is derived as

$$\begin{aligned} f(RM_t | N_t = m > 0, I_{t-1}; \zeta, \varsigma) &= f\Big(\frac{RM_t}{\mu_t} | N_t = m > 0, I_{t-1}\Big) \Big|\frac{1}{\mu_t}\Big| \\ &= \frac{2\mu_t}{RM_t} \Big(\frac{RM_t}{\mu_t} \varsigma \nu\Big)^{\frac{m\varsigma+\nu}{2} u} \frac{1}{\Gamma(m\varsigma)\Gamma(\nu)} \mathbb{K}_{m\varsigma-\nu}\left(2\sqrt{\frac{RM_t}{\mu_t}\varsigma\nu}\right) \Big|\frac{1}{\mu_t}\Big| \\ &= \frac{2}{RM_t} \Big(\frac{RM_t}{\mu_t} \varsigma \nu\Big)^{\frac{m\varsigma+\nu}{2}} \frac{1}{\Gamma(m\varsigma)\Gamma(\nu)} \mathbb{K}_{m\varsigma-\nu}\left(2\sqrt{\frac{RM_t}{\mu_t}\varsigma\nu}\right) \end{aligned}$$

Similarly to the case of η_t, the moments of RM_t conditional on $N_t = m > 0$ and I_{t-1} are derived from the moments function of the K density. ■

B.3 Moments of Z_t

Lemma B.1 *Given the MEM-J in* (3) *with Assumption 1 and the processes for* λ_t *specified as in* (16)*, the conditional moments of* Z_t *are*

$$\mathbb{E}\big[Z_t|I_{t-1}\big] = e^{-\lambda_t} + \lambda_t \tag{34}$$

$$\mathbb{V}\big[Z_t|I_{t-1}\big] = \frac{\lambda_t}{\varsigma} + e^{-\lambda_t} + \big(\lambda_t + \lambda_t^2\big) - \Big(e^{-\lambda_t} + \lambda_t\Big)^2. \tag{35}$$

The filtered expected jumps are

$$\mathbb{E}\left[Z_t|I_t\right] = \sum_{m=0}^{\infty} P\left(N_t = m|I_t\right) \times \mathbb{E}\left[Z_t|N_t = m, I_{t-1}\right] \tag{36}$$

Proof To prove this result, we start by computing the expectation (and for completeness the variance) of the jump term Z_t. We distinguish two cases depending on the presence of jumps. When we have no jumps, the mean and variance of Z_t are

$$\begin{aligned}\mathbb{E}\left[Z_t|N_t = 0, I_{t-1}\right] &= 1 \\ \mathbb{V}\left[Z_t|N_t = 0, I_{t-1}\right] &= 0\end{aligned}$$

Differently, when $N_t > 0$ we have

$$\begin{aligned}\mathbb{E}\left[Z_t|N_t = m > 0, I_{t-1}\right] &= m \\ \mathbb{V}\left[Z_t|N_t = m > 0, I_{t-1}\right] &= \frac{m}{\varsigma}.\end{aligned}$$

Integrating out the dependence on N_t, we obtain

$$\begin{aligned}\mathbb{E}\left[Z_t|I_{t-1}\right] &= \sum_{m=0}^{\infty} P\left(N_t = m|I_{t-1}\right) \times \mathbb{E}\left[Z_t|N_t = m, I_{t-1}\right] \\ &= P\left(N_t = 0|I_{t-1}\right) \times 1 + \sum_{m=1}^{\infty} P\left(N_t = m|I_{t-1}\right) \times m \\ &= e^{-\lambda_t} + \lambda_t \end{aligned} \tag{37}$$

as, for the Poisson process governing the jumps number we have $P\left(N_t = 0|I_{t-1}\right) = e^{-\lambda_t}$ and $\mathbb{E}\left[N_t|I_{t-1}\right] = \sum_{m=0}^{\infty} P\left(N_t = m|I_{t-1}\right) \times m = \lambda_t = \sum_{m=1}^{\infty} P\left(N_t = m|I_{t-1}\right) \times m$ since for $m = 0$ we do not have a contribution to the expected value.

For the variance we have

$$\mathbb{V}\left[Z_t|I_{t-1}\right] = \mathbb{E}\left[\mathbb{V}\left[Z_t|N_t, I_{t-1}\right]|I_{t-1}\right] + \mathbb{V}\left[\mathbb{E}\left[Z_t|N_t, I_{t-1}\right]|I_{t-1}\right].$$

Separately evaluating the two components, we obtain first

$$\begin{aligned}\mathbb{E}\left[\mathbb{V}\left[Z_t|N_t, I_{t-1}\right]|I_{t-1}\right] &= \sum_{m=0}^{\infty} P\left(N_t = m|I_{t-1}\right) \times \mathbb{V}\left[Z_t|N_t = m, I_{t-1}\right] \\ &= P\left(N_t = 0|I_{t-1}\right) \times 0 + \sum_{m=1}^{\infty} P\left(N_t = m|I_{t-1}\right) \times \frac{m}{\varsigma} \\ &= \frac{\lambda_t}{\varsigma}. \qquad (38)\end{aligned}$$

For the second element we can write

$$\mathbb{V}\left[\mathbb{E}\left[Z_t|N_t, I_{t-1}\right]|I_{t-1}\right] = \mathbb{E}\left[\mathbb{E}\left[Z_t|N_t, I_{t-1}\right]^2|I_{t-1}\right] - \left(\mathbb{E}\left[\mathbb{E}\left[Z_t|N_t, I_{t-1}\right]|I_{t-1}\right]\right)^2,$$

where the first term is given as

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}\left[Z_t|N_t, I_{t-1}\right]^2|I_{t-1}\right] &= \sum_{m=0}^{\infty} P\left(N_t = m|I_{t-1}\right) \times \mathbb{E}\left[Z_t|N_t = m, I_{t-1}\right]^2 \\ &= P\left(N_t = 0|I_{t-1}\right) \times 1 + \sum_{m=1}^{\infty} P\left(N_t = m|I_{t-1}\right) \times m^2 \\ &= e^{-\lambda_t} + \left(\lambda_t + \lambda_t^2\right),\end{aligned}$$

from the second order moment of a Poisson and using the fact that the contribution of the zero-jump component to the second order moment is equal to zero. By the law of iterated expectations,

$$\mathbb{E}\left[\mathbb{E}\left[Z_t|N_t, I_{t-1}\right]|I_{t-1}\right] = \mathbb{E}\left[Z_t|I_{t-1}\right] = e^{-\lambda_t} + \lambda_t.$$

■

The filtered expected jumps are given by

$$\begin{aligned}\mathbb{E}\left[Z_t|I_t\right] &= \sum_{m=0}^{\infty} P\left(N_t = m|I_t\right) \times \mathbb{E}\left[Z_t|N_t = m, I_t\right] \\ &= \sum_{m=0}^{\infty} P\left(N_t = m|I_t\right) \times \mathbb{E}\left[Z_t|N_t = m, I_{t-1}\right] \qquad (39)\end{aligned}$$

where $P\left(N_t = m|I_t\right)$ is derived in equation (25).

B.4 Proof of Proposition 3

The expected value of RM_t conditional on I_{t-1} is obtained, noting that μ_t is measurable on I_{t-1}, so that

$$\begin{aligned}\mathbb{E}\left[RM_t|I_{t-1}\right] &= \mu_t \mathbb{E}\left[\eta_t|I_{t-1}\right] \\ &= \mu_t \sum_{m=0}^{\infty} \mathbb{E}\left[\eta_t|N_t = m, I_{t-1}\right] \times P\left(N_t = m|I_{t-1}\right) \\ &= \mu_t \mathbb{E}\left[Z_t|I_{t-1}\right].\end{aligned}$$

It follows that

$$\mathbb{E}\left[RM_t|I_{t-1}\right] = \mu_t\left(e^{-\lambda_t} + \lambda_t\right), \qquad (40)$$

where the conditional expected value of Z_t is derived in (37). The conditional variance of RM_t is obtained as

$$\begin{aligned}\mathbb{V}\left[RM_t|I_{t-1}\right] &= \mathbb{E}\left[RM_t^2|I_{t-1}\right]-\mathbb{E}\left[RM_t|I_{t-1}\right]^2 \\ &= \mu_t^2\left\{\left[\frac{\lambda_t}{\varsigma}+e^{-\lambda_t}+(\lambda_t+\lambda_t^2)\right](1+\nu^{-1})-(e^{-\lambda_t}+\lambda_t)^2\right\}. \qquad (41)\end{aligned}$$

■

B.5 Proof of Theorem 1

The proof follows closely that of Theorem 2.5 in Franq and Zakoian (2010). First, we show that condition (23) is necessary. Since RM_t is a strictly positive covariance stationary process with $\lambda > 0$

$$E[RM_t] = (e^{-\lambda}+\lambda)E[\mu_t]$$

is a positive real number which does not depend on t, with $E[\mu_t] > 0$. Taking the expectation of both sides of (22)

$$E[\mu_t] = \omega + \sum_{i=1}^{q}(e^{-\lambda}+\lambda)E[\mu_t] + \sum_{i=1}^{p}\beta_i E[\mu_t]$$

that is

$$\left(1-(e^{-\lambda}+\lambda)\sum_{i=1}^{q}\alpha_i-\sum_{i=1}^{p}\beta_i\right)E[\mu_t] = \omega$$

Since $\omega > 0$, we must have condition (23).

Now we turn to the sufficient condition. Let $y_t \equiv RM_t$, the vector form of the process in (22) is

$$z_t = b_t + A_t z_{t-1} \qquad (42)$$

where

$$z_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-q+1} \\ \mu_t \\ \vdots \\ \mu_{t-p+1}\end{bmatrix}, \quad b_t = \begin{bmatrix} \omega\eta_t \\ \vdots \\ 0 \\ \omega \\ \vdots \\ 0\end{bmatrix}$$

and

$$A_t = \begin{bmatrix} \alpha_1\eta_t & \dots & \alpha_{q-1}\eta_t & \alpha_q\eta_t & \beta_1\eta_t & \dots & \beta_{p-1}\eta_t & \beta_p\eta_t \\ 1 & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \ddots & 1 & 0 & 0 & \dots & 0 & 0 \\ \alpha_1 & \dots & \alpha_{q-1} & \alpha_q & \beta_1 & \dots & \beta_{p-1} & \beta_p \\ 0 & \dots & 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix}$$

A_t is a $(p+q)\times(p+q)$ matrix with positive and independent coefficients. Further, $A = \mathbb{E}[A_t]$ and $b = \mathbb{E}[b_t]$ do not depend on t.

Given the condition in (23), we can construct a stationarity solution. For $t, k \in \mathbb{Z}$, we define

$\mathbb{R}^d$-valued vectors as follows:

$$Z_k(t) = \begin{cases} 0 & \text{if } k < 0 \\ b_t + A_t Z_{k-1}(t-1) & \text{if } k \geq 0 \end{cases}$$

With a multiplicative norm, i.e. $\|A\| = \sum |a_{ij}|$, we have, for any random matrix A with positive coefficients, $\mathbb{E}\|A\| = \mathbb{E}\sum_{i,j} |a_{i,j}| = \|\mathbb{E}[A]\|$. For $k > 0$

$$\mathbb{E}\|Z_k(t) - Z_{k-1}(t-1)\| = \|\mathbb{E}[A_t A_{t-1} \dots A_{t-k+} b_{t-k}]\|$$

because the matrix $A_t A_{t-1} \dots A_{t-k+} b_{t-k}$ is positive. All the terms of the product are independent (because the process $\{\eta_t\}$ is i.i.d. and every term is function of a variable η_{t-j}). Provided that $A \equiv \mathbb{E}[A_t]$ and $b = \mathbb{E}[b_t]$ do not depend on t, it follows that

$$\mathbb{E}\|Z_k(t) - Z_{k-1}(t-1)\| = \|A^k b\| = \iota' A^k b$$

where $\iota = (1, \dots, 1)'$, because all the elements in the vector $A^k b$ are positive. The condition (23) implies that the eignevalues of A are strictly less than one. The characteristic polynomial can be expressed as

$$\det(\lambda I_{p+q} - A) = \lambda^{p+q}\Big(1 - (e^{-\lambda} + \lambda)\sum_{j=1}^{q} \alpha_j \lambda^{-j} - \sum_{i=1}^{p} \beta_i \lambda^{-i}\Big)$$

When $|\lambda| \geq 1$, using the inequality $|a - b| \geq |a| - |b|$ we have

$$\begin{aligned} |\det(\lambda I_{p+q} - A)| &\geq \left|\Big(1 - (e^{-\lambda} + \lambda)\sum_{j=1}^{q} \alpha_j \lambda^{-j} - \sum_{i=1}^{p} \beta_i \lambda^{-i}\Big)\right| \\ &\geq 1 - (e^{-\lambda} + \lambda)\sum_{j=1}^{q} \alpha_j - \sum_{i=1}^{p} \beta_i > 0 \end{aligned}$$

it follows that the spectral radius of A is less than one, i.e. $\rho(A) < 1$. This implies that $A^k \to 0$ at exponential rate as $k \to \infty$. For any fixed t, $Z_k(t)$ converges almost surely as $k \to \infty$. Let z_t denote the limit of $\{Z_k(t)\}_{k\in\mathbb{Z}}$. For fixed k, the process $\{Z_k(t)\}_{t\in\mathbb{Z}}$ is strictly stationary. It follows that the limit process $z(t)$ is strictly stationary and it is a solution of (42).

■

B.6 Proof of Theorem 2

Differently from Theorem 1, now A_t is a $(p+q)\times(p+q)$ matrix with positive coefficients but not independent, since η_t is correlated through λ_t. The matrix A_t can be written as the Hadamard product of two matrices, i.e.

$$A_t = A \odot E_t$$

with

$$A = \begin{bmatrix} \alpha_1 & \dots & \alpha_{q-1} & \alpha_q & \beta_1 & \dots & \beta_{p-1} & \beta_p \\ 1 & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \ddots & 1 & 0 & 0 & \dots & 0 & 0 \\ \alpha_1 & \dots & \alpha_{q-1} & \alpha_q & \beta_1 & \dots & \beta_{p-1} & \beta_p \\ 0 & \dots & 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix}$$

and a $(p+q) \times (p+q)$ matrix

$$E_t = \begin{bmatrix} \eta_t & \eta_t & \dots & \eta_t \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \dots & 1 \end{bmatrix}.$$

The conditional mean of η_t is

$$E[\eta_t | I_{t-1}] = e^{-\lambda_t} + \lambda_t.$$

whereas, the unconditional mean satisfies

$$E[\eta_t] = E[e^{-\lambda_t}] + E[\lambda_t] = E[e^{-\lambda_t}] + \frac{\phi_1}{1-\phi_2}.$$

In the last equality, since $0 < E[e^{-\lambda_t}] < 1$, it follows that $E[e^{-\lambda_t}] + \frac{\phi_1}{1-\phi_2} < 1 + \frac{\phi_1}{1-\phi_2}$.

The sequence $\{A_t, t \in \mathbb{Z}\}$ is ergodic and strictly stationary. With a multiplicative norm, i.e. $\|A\| = \sum |a_{ij}|$, $\log \|A_t\| \leq \log \|A\| + \log \|E_t\|$, with $\log \|E_t\| = \log [(p+q)(\eta_t + (p+q) - 1)]$. Therefore $\log^+ \|A_t\| \leq \log \|A\| + \log^+ \|E_t\|$, where $\log^+(x) = \max(\log(x), 0)$. Given that the Lyapunov exponent γ is equal to, see Franq and Zakoian (2010, Theorem 2.3),

$$\gamma = \lim_{t \to \infty} a.s. \frac{1}{t} \log \|A_t A_{t-1} \dots A_1\| \tag{43}$$

and

$$\log (\|A_t A_{t-1} \dots A_1\|) \leq \log \|A^t\| + \sum_{i=1}^{t} \log \|E_i\|$$

Since, $\lim_{t \to \infty} \frac{1}{t} \log \|A^t\| = \log\{\rho(A)\}$, $\gamma < 0$ if and only if

$$\rho(A) < \exp(-E[\log \|E_t\|]). \tag{44}$$

Now, we turn to the proof of the existence of a stationary and ergodic solution if the condition in (44) is satisfied, i.e. $\gamma < 0$. Since the random variable η_t has finite variance, the components of the matrix A_t are integrable. Hence,

$$\mathbb{E}[\log^+ \|A_t\|] \leq \mathbb{E}\|A_t\| < \infty.$$

With $\gamma < 0$ it follows from (43) that

$$\tilde{z}_t(N) = b_t + \sum_{n=0}^{N} A_t A_{t-1} \dots A_{t-n} b_{t-n-1}$$

converges a.s. when N goes to infinity, to some limit $\tilde{z}_t$. Using the multiplicative norm

$$\|\tilde{z}_t(N)\| \leq \|b_t\| + \sum_{n=0}^{\infty} \|A_t A_{t-1} \dots A_{t-n}\| \|b_{t-n-1}\|$$

and

$$\begin{aligned} \|A_t A_{t-1} \dots A_{t-n}\|^{1/n} \|b_{t-n-1}\|^{1/n} &= \exp\left[\frac{1}{n}\|A_t A_{t-1} \dots A_{t-n}\| + \frac{1}{n}\|b_{t-n-1}\|\right] \\ &\overset{a.s.}{\to} \exp(\gamma) < 1. \end{aligned}$$

To show that $n^{-1} \log \|b_{t-n-1}\| \to 0$ we have used the result that for a sequence X_n of identically distributed random variables admitting an expectation holds that $X_n/n \overset{a.s.}{\to} 0$ when $n \to \infty$. In our case this can be applied because $\mathbb{E}|\log \|b_{t-n-1}\|| < \infty$, see Franq and Zakoian (2010, Proof of Theorem 2.4, p.31). Let $\tilde{z}_{q+1,t}$ denote the $(q+1)$-th element of $\tilde{z}_t$. Setting $y_t = \tilde{z}_{q+1,t}\eta_t$, we define a solution of model (3). This solution is nonanticipative because y_t can be expressed as a measurable function of $\eta_t, \eta_{t-1}, \dots$. By the ergodicity of η_t this solution is also strictly stationary and ergodic.

■