



AARHUS  
UNIVERSITY  
BUSINESS AND SOCIAL SCIENCES  
DEPARTMENT OF ECONOMICS AND BUSINESS



**CREATES**

Center for Research in Econometric Analysis of Time Series

# **Chasing volatility - A persistent multiplicative error model with jumps**

**Massimiliano Caporin, Eduardo Rossi and  
Paolo Santucci de Magistris**

**CREATES Research Paper 2014-29**

# Chasing volatility

## A persistent multiplicative error model with jumps

Massimiliano Caporin \*    Eduardo Rossi †    Paolo Santucci de Magistris ‡

This version: December 27, 2015

### Abstract

Persistence and unpredictable large increments characterize the volatility of financial returns. We propose the Multiplicative Error Model with volatility jumps (MEM-J) to describe and predict the probability and the size of these extreme events. The probabilistic properties of the MEM-J model are studied. Under the MEM-J, the conditional density of the realized measure is a countably infinite mixture of Gamma and *Kappa* distributions, and the conditional moments can be derived in closed form. Stationarity conditions and the asymptotic theory for the maximum likelihood estimation are also derived. The MEM-J is fitted to the daily bipower measures of 7 stock indexes and 16 individual NYSE stocks. 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 of the volatility tail.

**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 associated with 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), Jacod and Todorov (2010), Todorov and Tauchen (2011), Andersen et al. (2012b), Bandi and Renò (2012, 2015), Jacod et al. (2013a), Jacod et al. (2013b), and Ait-Sahalia and Jacod (2014), which discuss a direct test for volatility jumps based on high frequency returns, and therein cited references. 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 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. We note that 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 the realized volatility series, see Caporin et al. (2014b). However, in this case obtaining the forecasts distribution of the volatility level can be problematic in non-Gaussian cases. In order to jointly model both the continuous and the jump component present in the daily volatility process, we propose the Asymmetric HAR-Multiplicative Error Model with jumps (AHAR-MEM-J) that extends 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 persistent 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 realizations in the right tail of volatility distribution. Consequently, the MEM-J model allows capturing the observed probability mass associated with large, or even extreme, realizations. The model is analogous to the dynamic two-factor models of Russell (1999), Ghysels et al. (2004) and

Bauwens and Veredas (2004) for durations. We carefully analyze the probabilistic properties of the model. 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 indeed a countably infinite mixture, governed by a compound Poisson process, 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). Given that the conditional density has a closed form expression, the AHAR-MEM-J parameters can be estimated by maximum likelihood (ML). A number of theoretical results are derived to support the adequacy of the ML approach. First, we derive stationarity conditions under the presence of jumps following the approach of Bougerol and Picard (1992) and Franq and Zakoian (2010). Second, we prove that the infinite mixture is bounded and identified, meaning that different values of the conditional density of the innovations are always associated with distinct parameter values. Indeed, the mixture is governed by a compound Poisson process which guarantees that the ML estimator exists and is unique, see also Simar (1976). Finally, we derive the asymptotic properties of the ML estimator, showing that the estimates are consistent and asymptotically normal. A number of Monte Carlo simulations support the theoretical results and show that the ML estimates are indeed unbiased and unimodal in finite samples also under over-specification of the conditional density. We also propose a series of diagnostic tests, based on the moments of the infinite mixture and on the dynamics in the jump innovations, to statistically evaluate the correctness of the model specification. 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, thus disentangling 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 four. 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 discuss model stationarity conditions and the asymptotic properties of the ML estimator. Fourthly, 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. We compare MEM-J to alternative specifications without jumps which turn out to be unable to fit the right tail of the volatility density, especially in periods of market turmoil. Consequently, the MEM-J model might turn out to be empirically relevant for the analysis of positive valued series affected by the occurrence of spikes, such as trading volume, liquidity measures, inter-trade durations and commodity prices. Outside the financial framework, the MEM-J can be adopted to the model the wind speed, the barometric pressure or the concentration of gases and pollutants in the atmosphere.

The paper is organized as follows. Section 2 sets the notation of the baseline MEM and briefly review recent generalizations of the model. Section 3 describes the MEM-J and the infinite mixture distribution that characterizes the conditional density of the dependent variable  $X_t$ . Conditional moments are also presented. Section 4 introduces model extensions with HAR dynamics and time-varying parameters, and presents the conditions for stationarity. Section 5 discusses the model identification and provides theorems for consistency and asymptotic normality of the ML estimator. Moreover, it includes a Monte Carlo experiment and introduces three diagnostic tests. Section 6 describes the dataset and illustrates the empirical results with stocks indexes and individual S&P 500 stocks under different model specifications. In Section 7 the results of the VolaR analysis are reported and discussed. Finally, Section 8 concludes. Proofs, selected derivations of relevant quantities and additional theoretical details are included in Appendices A and B.<sup>1</sup> Note that, in the rest of the paper, the term *jumps*, will be always referred to *volatility jumps*, unless differently specified.

## 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  $X_t$  an almost surely (a.s.) strictly positive random variable, for instance an ex-post estimator of daily integrated volatility at time  $t$ . The variable  $X_t$  follows a MEM model if

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

---

<sup>1</sup>An on-line web appendix containing further theoretical elements and additional results complements the paper.

with

$$\mu_t = \omega + \alpha X_{t-1} + \beta \mu_{t-1},$$

and  $\omega > 0, \alpha \geq 0, \beta \geq 0$ . The innovation  $\varepsilon_t$  is an i.i.d. random variable with strictly positive support that can be assumed to be Gamma distributed, whose density in scale-shape is

$$f_\varepsilon(\varepsilon_t | \mathcal{F}_{t-1}) = \frac{1}{\varepsilon_t} (\varepsilon_t \nu)^\nu \frac{1}{\Gamma(\nu)} e^{-\nu \varepsilon_t}, \quad \varepsilon_t \geq 0 \quad (2)$$

denoted  $\Gamma(\frac{1}{\nu}, \nu)$  where  $\frac{1}{\nu}$  is the scale and  $\nu$  is the shape of the Gamma density, both driven by the common parameter  $\nu$ . Consequently,  $\mathbb{E}[\varepsilon_t | \mathcal{F}_{t-1}] = 1$  and  $\mathbb{V}[\varepsilon_t | \mathcal{F}_{t-1}] = \frac{1}{\nu}$ , where  $\mathcal{F}_{t-1}$  is the information set at time  $t-1$  and  $\varepsilon_t$  is independent of  $\mathcal{F}_{t-1}$ . If  $X_t$  follows a MEM with a Gamma disturbance, its conditional density to  $\mathcal{F}_{t-1}$  is  $\Gamma(\frac{\mu_t}{\nu}, \nu)$ . Therefore, the conditional mean and variance are  $\mathbb{E}[X_t | \mathcal{F}_{t-1}] = \mu_t$  and  $\mathbb{V}[X_t | \mathcal{F}_{t-1}] = \mu_t^2 \nu^{-1}$ , respectively. Interestingly, the conditional variance of  $X_t$  is function of  $\mu_t^2$ , thus making the MEM potentially able to account for heteroskedastic effects that characterize many financial time series. With this purpose, the parameter  $\nu$  in the baseline MEM specification can be easily made time-varying according to a GARCH-type law of motion, thus increasing the flexibility of the model in describing the higher moments of  $X_t$ .

The form of  $\mu_t$  is sufficiently flexible to include simple auto-regressive patterns, HAR terms as in Corsi (2009), asymmetry, or predetermined variables. Examples of possible specifications for  $\mu_t$  are given, among others, in Engle and Gallo (2006) and Brownlees et al. (2012). Recently, Gallo and Otranto (2012) extend the MEM to include time-varying parameters in the expression of  $\mu_t$  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. (2015) 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, in particular for applications on realized volatility sequences, 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. Finally, Hautsch et al. (2013) introduce a mixture in a multiplicative model to capture the probability mass at zero, an empirical evidence in some high frequency series such as trading volumes.

As it will be shown in the empirical application, several of these distributional choices are not able to assign the correct probability to the occurrence of the extremely high realizations that characterize the volatility dynamics, as proxied by a realized volatility estimator. This further confirms the need for a novel modeling setup explicitly designed to capture the occurrence of such events. We also stress this model feature might be of relevant interest in other research areas where the series of interest is positively valued characterized by sudden movements, i.e. jumps.

### 3 A Multiplicative Error Model with Jumps

The baseline MEM with a Gamma distributed error term might be poorly designed to assign the correct probability to the occurrence of large and abrupt movements, i.e. the jumps, that characterize, for instance, 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,  $X_t$  equals the product of three stochastic elements

$$X_t = \mu_t Z_t \varepsilon_t \quad (3)$$

where  $\mu_t$  is a function measurable with respect to  $\mathcal{F}_{t-1}$ .  $Z_t$  is the volatility jump component, and the innovation  $\varepsilon_t | \mathcal{F}_{t-1} \sim i.i.d. \Gamma(\frac{1}{\nu}, \nu)$ . Hereafter, to simplify the interpretation of the model outcome, the Gamma density of the innovation term is expressed in the mean-shape form, i.e.  $\varepsilon_t | \mathcal{F}_{t-1} \sim i.i.d. \Gamma(1, \nu)$ , which is, by construction, equivalent to the scale-shape representation. Hence, we require a number of assumptions on  $Z_t$  and  $\varepsilon_t$  to identify and separate the two sources of shocks. The jump term,  $Z_t$ , is defined as

$$Z_t = \begin{cases} d_\lambda & N_t = 0 \\ \sum_{j=1}^{N_t} Y_{j,t} & N_t > 0 \end{cases} \quad (4)$$

where  $N_t$  is a discrete counting process governing the number of jumps that arrive between  $t-1$  and  $t$ . We model it as a Poisson process with intensity  $\lambda > 0$ ,

$$P(N_t = m | \mathcal{F}_{t-1}) = \frac{e^{-\lambda} \lambda^m}{m!}, \quad m = 0, 1, 2, \dots \quad (5)$$

The parameter  $\lambda$  is the expected number of jumps in the interval  $(t-1, t]$ . 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,  $Y_{j,t} \sim \Gamma(d_\lambda, \varsigma)$  (in mean-shape form). Note that the jump density does not depend on time and the parameter characterizing the jump size law is assumed to be time-invariant. The scalar  $d_\lambda$  is a scaling factor ensuring  $\mathbb{E}[Z_t \varepsilon_t | \mathcal{F}_{t-1}] = 1$ , and it is a function of the jump intensity  $\lambda$ :<sup>2</sup>

$$d_\lambda := \left( e^{-\lambda} + \lambda \right)^{-1}. \quad (6)$$

In the following assumption we specify the elements which are needed to derive the conditional density of  $X_t$  in closed form.

**Assumption 1** *In the MEM-J in (3),  $X_t = \mu_t \eta_t$  with  $\eta_t = Z_t \varepsilon_t$  and*

<sup>2</sup>The supplementary material provides a justification for this specific choice of the scaling factor.

- i.  $\varepsilon_t | \mathcal{F}_{t-1} \sim i.i.d. \Gamma(1, \nu)$ ;
- ii.  $N_t | \mathcal{F}_{t-1} \sim i.i.d. Poi(\lambda)$  with  $\lambda \geq 0$ ;
- iii.  $Y_{j,t} \sim i.i.d. \Gamma(d_\lambda, \varsigma)$ , for  $j = 1, \dots, N_t$ ;
- iv.  $\varepsilon_t$ ,  $N_t$  and the variables  $Y_{j,t}$ ,  $j = 1, 2, \dots, N_t$ , are assumed to be independent for any  $t$ .

By Assumption 1, it follows that  $Z_t | (N_t = m > 0, \mathcal{F}_{t-1})$  is distributed as a Gamma with mean  $md_\lambda$  and shape  $m\varsigma$ , with density given by

$$g_Z(z_t | N_t = m > 0, \mathcal{F}_{t-1}; \varsigma, \lambda) = \frac{1}{z_t} \left( \frac{\varsigma z_t}{d_\lambda} \right)^{m\varsigma} \frac{e^{-\frac{\varsigma z_t}{d_\lambda}}}{\Gamma(m\varsigma)}. \quad (7)$$

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.

### 3.1 The conditional density of $\eta_t$

The conditional density of  $\eta_t = Z_t \varepsilon_t$  depends on  $N_t$  through  $Z_t$ . When  $N_t = 0$ ,  $\eta_t = d_\lambda \varepsilon_t$  the conditional density is

$$g_\eta(\eta_t | N_t = 0, \mathcal{F}_{t-1}; \nu, \lambda) = \frac{1}{\eta_t} \left( \frac{\nu \eta_t}{d_\lambda} \right)^\nu \frac{e^{-\frac{\nu \eta_t}{d_\lambda}}}{\Gamma(\nu)}. \quad (8)$$

with mean  $d_\lambda$ . Differently, when  $N_t = m > 0$ ,  $\eta_t | (Z_t, N_t = m > 0, \mathcal{F}_{t-1}) \sim \Gamma(Z_t, \nu)$ . Hence, to derive the conditional density of  $\eta_t$  given  $N_t = m > 0$  and  $\mathcal{F}_{t-1}$ , we have to evaluate the following integral:

$$\int_0^\infty g_\eta(\eta_t | N_t = m > 0, Z_t, \mathcal{F}_{t-1}; \nu) g_Z(z_t | N_t = m > 0, \mathcal{F}_{t-1}; \varsigma, \lambda) dz, \quad (9)$$

where  $g_\eta(\eta_t | N_t = m > 0, Z_t, \mathcal{F}_{t-1}; \nu)$  is the density of a Gamma random variable with mean  $Z_t$  and shape  $\nu$ . In the following proposition we present the conditional density of  $\eta_t$  which includes the closed-form solution to the integral in (9).

**Proposition 1** Consider model (3) where  $\eta_t = Z_t \varepsilon_t$  with  $Z_t$  defined in (4). Under Assumption 1, it follows that

$$k_\eta(\eta_t | N_t = m > 0, \mathcal{F}_{t-1}; \nu, \varsigma, \lambda) = \frac{2}{\eta_t} \left( \eta_t \frac{\varsigma \nu}{d_\lambda} \right)^{\frac{m\varsigma + \nu}{2}} \frac{1}{\Gamma(m\varsigma) \Gamma(\nu)} \mathbb{K}_{m\varsigma - \nu} \left( 2 \sqrt{\eta_t \frac{\varsigma \nu}{d_\lambda}} \right), \quad (10)$$

where:  $\mathbb{K}_a(\cdot)$  is the modified Bessel function of the second kind. Hence, the innovation term  $\eta_t$ , conditional on  $N_t = m > 0$  and  $\mathcal{F}_{t-1}$ , has a Kappa, or  $K$ , distribution, see Redding (1999), denoted as  $K(md_\lambda, m\varsigma, \nu)$ . Conditional on  $\mathcal{F}_{t-1}$  the density of  $\eta_t$  is a countably infinite mixture

$$f_\eta(\eta_t | \mathcal{F}_{t-1}; \nu, \varsigma, \lambda) = P(N_t = 0 | \mathcal{F}_{t-1}; \lambda) g_\eta(\eta_t | N_t = 0, \mathcal{F}_{t-1}; \nu, \lambda) + \sum_{m=1}^{\infty} P(N_t = m | \mathcal{F}_{t-1}; \lambda) k_\eta(\eta_t | N_t = m > 0, \mathcal{F}_{t-1}; \nu, \varsigma, \lambda), \quad (11)$$



where  $P(N_t = 0|\mathcal{F}_{t-1}; \lambda) = e^{-\lambda}$  and  $P(N_t = m > 0|\mathcal{F}_{t-1}; \lambda) = (e^{-\lambda}\lambda^m)/m!$ .

Proof in Appendix A.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 depends on the realized 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  $\varepsilon_t$ . Interestingly, the conditional variance of  $\eta_t$  is an increasing function of the number of jump arrivals,  $m$ . Hence, periods with a larger number of jump arrivals are characterized by a higher volatility-of-volatility. When integrating out the dependence of the mixing variable, that is the Poisson process  $N_t$ , the density of  $\eta_t$  conditional on  $\mathcal{F}_{t-1}$  is a countably infinite mixture of Gamma and Kappa densities, whose weights depend on  $\lambda$ . As  $\lambda$  increases, more weight is given to the  $k_\eta$ 's in the infinite sum, while when  $\lambda = 0$  the conditional density of  $\eta_t$  is  $g_\eta$  and the MEM-J reduces to the MEM. Appendix B reports additional details on the  $K$  distribution.

### 3.2 The conditional density of $X_t$

The conditional density of  $X_t$ , given  $N_t = m > 0$  and  $\mathcal{F}_{t-1}$ , follows from the conditional distribution of  $\eta_t$  in equation (10). The following propositions present the density of  $X_t$  and introduce its conditional moments.

**Proposition 2** Consider model (3) where  $\eta_t = Z_t\varepsilon_t$  with  $Z_t$  defined in equation (4). Under Assumption 1 and given Proposition 1, it follows that

$$g_X(x_t|N_t = 0, \mathcal{F}_{t-1}; \theta_1, \nu, \lambda) = \frac{1}{x_t} \left( \frac{\nu x_t}{d_\lambda \mu_t} \right)^\nu \frac{e^{-\frac{\nu x_t}{d_\lambda \mu_t}}}{\Gamma(\nu)} \quad (12)$$

$$k_X(x_t|N_t = m > 0, \mathcal{F}_{t-1}; \theta_1, \nu, \varsigma, \lambda) = \frac{2}{x_t} \left( \frac{x_t \varsigma \nu}{\mu_t d_\lambda} \right)^{\frac{m\varsigma + \nu}{2}} \frac{1}{\Gamma(m\varsigma)\Gamma(\nu)} \mathbb{K}_{m\varsigma - \nu} \left( 2\sqrt{\frac{x_t \varsigma \nu}{\mu_t d_\lambda}} \right), \quad (13)$$

where  $\theta_1$  contains the parameters of  $\mu_t$ . Thus  $X_t$ , conditional on  $N_t = m > 0$  and  $\mathcal{F}_{t-1}$ , has a  $K$  distribution, denoted as  $K(\mu_t m d_\lambda, m\varsigma, \nu)$ . The first two conditional moments of  $X_t$  are

$$\begin{aligned} \mathbb{E}[X_t|(N_t = 0, \mathcal{F}_{t-1})] &= \mu_t d_\lambda, & \mathbb{E}[X_t|(N_t = m > 0, \mathcal{F}_{t-1})] &= \mu_t m d_\lambda, \\ \mathbb{V}[X_t|(N_t = 0, \mathcal{F}_{t-1})] &= \frac{\mu_t^2 d_\lambda^2}{\nu}, & \mathbb{V}[X_t|(N_t = m > 0, \mathcal{F}_{t-1})] &= \mu_t^2 m^2 d_\lambda^2 \frac{m\varsigma + \nu + 1}{m\varsigma\nu}. \end{aligned}$$

Proof in Appendix A.2.

As a result, both the conditional mean and variance of  $X_t$  are not only time-varying and driven by  $\mu_t$ , as in the MEM, but also dependent on the realized number of jumps,  $m$ . Integrating out the dependence on the realized number of jumps, it follows that

**Proposition 3** Under Assumption 1 and given Proposition 2, the density of  $X_t$  conditional on

$\mathcal{F}_{t-1}$  is a countably infinite mixture

$$f_X(x_t|\mathcal{F}_{t-1};\theta) = e^{-\lambda}g_X(x_t;\theta_1,\nu,\lambda) + \sum_{m=1}^{\infty} \frac{e^{-\lambda}\lambda^m}{m!} \times k_X(x_t|m,\theta_1,\nu,\varsigma,\lambda), \quad (14)$$

where  $g_X(x_t;\theta_1,\nu,\lambda) \equiv g_X(x_t|N_t = 0,\mathcal{F}_{t-1};\theta_1,\nu,\lambda)$ ,  $k_X(x_t|m,\theta_1,\nu,\varsigma,\lambda) \equiv k_X(x_t|N_t = m > 0,\mathcal{F}_{t-1};\theta_1,\nu,\varsigma,\lambda)$  and  $\theta = (\theta'_1,\varsigma,\nu,\lambda)'$ . It follows that the mean and variance of  $X_t$  conditional on  $\mathcal{F}_{t-1}$  are

$$\mathbb{E}[X_t|\mathcal{F}_{t-1}] = \mu_t, \quad (15)$$

$$\mathbb{V}[X_t|\mathcal{F}_{t-1}] = \mu_t^2 \left\{ \left[ \frac{\lambda}{\varsigma} + e^{-\lambda} + (\lambda + \lambda^2) \right] d_\lambda^2 (1 + \nu^{-1}) - 1 \right\}. \quad (16)$$

Proof in Appendix A.3.

Notably, while the conditional mean of  $X_t$  is exactly equal to  $\mu_t$ , the conditional variance of  $X_t$  is the product between the squared conditional mean and a factor depending on the parameters of the mixture. By a straightforward extension of the conditional moments of  $\eta_t$  (see the Supplementary material) it is possible to show that the parameters of the mixture also influence the higher conditional moments of  $X_t$ . In the following, we generalize the model allowing for a richer dynamic specification for  $\mu_t$ , as opposed to the baseline case of the MEM model, and introducing time-varying jump arrival intensity,  $\lambda_t$ .

## 4 A persistent MEM-J with time-varying parameters

In this section we provide generalizations of the MEM-J model aimed at capturing dynamic and distributional features characterizing sequences of ex-post estimators of the daily integrated volatility.

### 4.1 Specification of $\mu_t$

Looking at the dynamics of the model residuals reveals the importance of a correct specification of  $\mu_t$ . Since time series like those of integrated volatility estimators are characterized by a slow and hyperbolic decay of the autocorrelation function, it follows that a simple ARMA(1,1) specification, as implied by the baseline MEM, is not enough to describe such a rich dynamic behaviour. A 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, designed to capture the persistence of the logarithm of realized volatility. Building on the latter contribution, we consider two alternative specifications for  $\mu_t$ , the time-varying mean of  $X_t$ , an ex-post estimator of the daily integrated volatility for a given asset:

- Asymmetric HAR-MEM (AHAR-MEM):

$$\mu_t = \omega + \beta\mu_{t-1} + \alpha_1 X_{t-1} + \alpha_2 X_{t-1:t-5} + \alpha_3 X_{t-1:t-21} + \gamma X_{t-1}^- + \gamma_2 X_{t-1:t-5}^- + \gamma_3 X_{t-1:t-21}^- \quad (17)$$

where  $X_{t-1:t-5} = \frac{1}{5} \sum_{j=1}^5 X_{t-j}$  and  $X_{t-1:t-21} = \frac{1}{21} \sum_{j=1}^{21} X_{t-j}$ ,  $X_t^- \equiv X_t \cdot \mathbb{I}\{r_t < 0\}$ , with  $r_t$  being the daily return of the asset whose daily integrated volatility is being analyzed,  $X_{t-1:t-5}^- = \frac{1}{5} \sum_{j=1}^5 X_{t-j}^-$  and  $X_{t-1:t-21}^- = \frac{1}{21} \sum_{j=1}^{21} X_{t-j}^-$ ; finally, all parameters are non-negative. 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). Within this specification, it is possible to test whether the inclusion of the weekly and monthly volatility and asymmetry terms provides a significant improvement in fitting the volatility dynamics.<sup>3</sup>

- Asymmetric MEM (A-MEM):

$$\mu_t = \omega + \alpha_1 X_{t-1} + \beta \mu_{t-1} + \gamma_1 X_{t-1}^- \quad (18)$$

where all parameters are non negative. This model is nested in AHAR-MEM with the restriction  $\alpha_2 = \alpha_3 = \gamma_2 = \gamma_3 = 0$ . Further, the MEM is obtained simply setting  $\gamma_1 = 0$ .

The inclusion of HAR dynamics into the specification of  $\mu_t$  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. Finally, note that the inclusion of lagged values of  $X_t$  makes the conditional mean  $\mu_t$  dependent on past jumps which are embedded in the past values of  $X_t$ . Further generalizations might introduce additional covariates in  $\mu_t$ , as in Caporin et al. (2014b).

## 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,  $\lambda$ , and define the AHAR-MEM-J- $\lambda_t$  model. We first specify the dynamic evolution of the parameter  $\lambda_t$ , 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 also in the MEM framework. Under time-varying jump intensity, the conditional variance of  $X_t$  in (16) evolves over time as a function of both  $\mu_t$  and  $\lambda_t$ . This allows for a great degree of flexibility in the MEM-J, making the model capable of a flexible dynamic specification for high order moments. In particular, the

---

<sup>3</sup>If the asymmetry component is not present, we refer to the model as the HAR-MEM.

jump intensity is assumed to follow:

$$\lambda_t = \phi_1 + \phi_2 \lambda_{t-1} + \phi_3 \xi_{t-1}, \quad (19)$$

where

$$\xi_t = \mathbb{E}[N_t | \mathcal{F}_t] - \lambda_t = \sum_{m=0}^{\infty} m P(N_t = m | \mathcal{F}_t, \theta_2) - \lambda_t, \quad (20)$$

and  $\theta_2 = (\phi_1, \phi_2, \phi_3)'$ . The restrictions  $\phi_1 > 0$  and  $\phi_2 > \phi_3 > 0$  are sufficient to guarantee the positiveness of  $\lambda_t$  as in Chan and Maheu (2002). As noted by Chan and Maheu (2002),  $\xi_t$  represents the change in the econometrician's conditional forecast of  $N_{t-1}$ , as the information set is updated. Note from the definition in (20) we have that  $\xi_t$  is a real-valued martingale difference sequence with respect to  $\mathcal{F}_{t-1}$ , since  $E(\xi_t | \mathcal{F}_{t-1}) = 0$  so that  $E(\xi_t) = 0$  and  $\text{Cov}(\xi_t, \xi_{t-h}) = 0$  for  $h > 0$ . We also assume that  $\xi_t$  is strictly stationary and ergodic. In particular, the innovation term of the jump intensity depends on the conditional probabilities of observing  $m$  jumps given  $\mathcal{F}_t$ , 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 updated by applying the Bayes rule as follows

$$P(N_t = m | \mathcal{F}_t; \theta_2) = \frac{f_X(x_t | N_t = m, \mathcal{F}_{t-1}, \theta) \times P(N_t = m | \mathcal{F}_{t-1}; \theta_2)}{f_X(x_t | \mathcal{F}_{t-1}, \theta)}, \quad m = 0, 1, 2, \dots \quad (21)$$

where  $\theta = (\theta'_1, \theta'_2, \nu, \varsigma)'$ , the terms on the right hand side are functions of  $\mathcal{F}_{t-1}$ , and the functional expression of  $f_X(x_t | N_t = m, \mathcal{F}_{t-1}, \theta)$  is defined in (12) for  $m = 0$ , and in (13) for  $m > 0$ . From a distributional point of view, letting the mixing parameter  $\lambda$  to be dynamic implies that the conditional density of  $X_t$  in (14) has time-varying weights associated with the mixture. This provides an extremely flexible specification of the density of  $X_t$ , which can be exploited to infer a precise probability of occurrence of tail events, see Section 7. Moreover, the results outlined in Propositions 1-3 on the densities of  $\eta_t$  and  $X_t$  conditional on  $\mathcal{F}_{t-1}$  are still valid under the time-varying jump intensity specification, since also the dynamics of  $\lambda_t$  are conditional on  $\mathcal{F}_{t-1}$ .<sup>4</sup> Finally, alternative specifications for  $\mu_t$  and  $\lambda_t$  are viable but not covered in the present work. For instance, the use of a log-linear specification for the jump intensity might be adopted to avoid the positivity constraints on the parameters, or a Generalized Autoregressive Score (GAS) representation, see Creal et al. (2013). Differently, even the jump size might be made dynamic, similarly to Chan and Maheu (2002).

### 4.3 Stationarity

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

$$\mu_t = \omega + \sum_{i=1}^q \alpha_i x_{t-i} + \sum_{i=1}^p \beta_i \mu_{t-i}. \quad (22)$$

---

<sup>4</sup>Assumption 1.iii need a minor change as the density of  $Y_{i,t}$  becomes conditional to  $\mathcal{F}_{t-1}$ .

with  $\omega > 0$ ,  $\alpha_i \geq 0, i = 1, 2, \dots, q$  and  $\beta_j \geq 0, j = 1, 2, \dots, p$  to ensure the positivity of  $\mu_t$ .

**Theorem 1** *Let  $X_t$  follow a strictly positive and covariance stationary MEM-J defined as in (3), with  $\mu_t$  defined in (22) and with  $\omega > 0$ . Under  $\mathbb{E}[\varepsilon_t | \mathcal{F}_{t-1}] = 1$  and Assumption 1.iv, then*

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

*Conversely, if (23) holds, the unique strictly stationary solution of the MEM-J model defined in (3) and (22) is a second-order stationary solution.*

Proof in Appendix A.4.

Deriving the stationarity conditions when the jump intensity is time-varying is more involved compared to the constant  $\lambda_t$  specification. In this case, we provide a sufficient condition for the stationarity of the MEM-J in (3) with density, conditional on  $\eta_t$ , defined in (13), where  $\eta_t$  is a sequence of random variables with time-varying intensity parameter  $\lambda_t$  defined as in (19).

**Theorem 2** *Under Assumption 1 with the processes  $\lambda_t$  and  $\mu_t$  specified as in (19) and (22) respectively, a sufficient condition for the existence of a strictly stationary solution of the MEM-J- $\lambda_t$ , defined by (3), (19) and (22), is*

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

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

$$A = \begin{bmatrix} \alpha_1 & \dots & \alpha_{q-1} & \alpha_q & \beta_1 & \dots & \beta_{p-1} & \beta_p \\ & I_{q-1} & & & & \mathbf{0}_{(q-1) \times (p+1)} & & \\ \alpha_1 & \dots & \alpha_{q-1} & \alpha_q & \beta_1 & \dots & \beta_{p-1} & \beta_p \\ & \mathbf{0}_{(p-1) \times q} & & & I_{p-1} & & \mathbf{0}_{(p-1) \times 1} & \end{bmatrix},$$

where  $I_m$  is an  $m$ -dimensional identity matrix and  $\mathbf{0}_{m \times n}$  is an  $m \times n$  matrix of zeros. When the strictly stationary solution exists, it is unique and ergodic.

Proof in Appendix A.5.

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  $\mu_t$  and on the expectation of the innovation  $\eta_t$  that combines the jump term and the error term  $\varepsilon_t$ .

## 5 Maximum Likelihood Estimation

The MEM-J- $\lambda_t$  can be estimated by maximum likelihood. Under the maintained assumption that  $N_t | \mathcal{F}_{t-1} \sim \text{Poisson}(\lambda_t)$  with  $N_t$  and  $\varepsilon_t$  independent processes, the conditional density of  $X_t$ ,  $f_X(x_t | \mathcal{F}_{t-1}; \theta)$ , can be computed in closed form as in (14). Indeed, the density of the  $K$  distribution is known and it does not need to be simulated, so that the computation of the

conditional log-likelihood function is straightforward. Let  $\mu_t$  be specified as in (22) and  $\theta = (\theta_1, \theta_2, \theta_3)$  be the parameter vector of the MEM-J- $\lambda_t$  model, with  $\theta_1 = (\omega, \alpha_1, \dots, \alpha_q, \beta_1, \dots, \beta_p)$ ,  $\theta_2 = (\phi_1, \phi_2, \phi_3)$  and  $\theta_3 = (\nu, \varsigma)$ . The conditional density of  $X_t = \mu_t Z_t \varepsilon_t$  is given by the infinite countably mixture in (14) and the sample log-likelihood is

$$L_T(\theta) = \sum_{t=1}^T l_t(\theta)$$

where  $l_t(\theta) = \log f_X(x_t | \mathcal{F}_{t-1}; \theta)$ . In the following, we also suppress the dependence on  $\mathcal{F}_{t-1}$  in the notation, so that  $f_X(x_t; \theta) \equiv f_X(x_t | \mathcal{F}_{t-1}; \theta)$ . The maximum likelihood estimator is defined as the maximum of the the log-likelihood function computed conditional on some initial r.v's  $\tilde{X}_0, \dots, \tilde{X}_{1-q}, \tilde{\xi}_0$  and  $\tilde{\mu}_0, \dots, \tilde{\mu}_{1-p}$ , i.e.

$$\hat{\theta}_T = \arg \max_{\theta \in \Theta} \tilde{L}_T(\theta).$$

The identifiability of a mixture distribution is essential for parameter estimation. In the following theorem we state that the MEM-J is identifiable.

**Theorem 3 (Identifiability)** *Given a sample  $\{x_t\}_{t=1}^T$  and assuming that  $g_X(x_t; \theta_1, \theta_2, \nu)$  and  $k_X(x_t; m, \theta)$  are uniquely identified by a parameter vector  $\theta$ , the MEM-J- $\lambda_t$  model is identifiable, that is*

$$\sum_{t=1}^T f_X(x_t; \theta^{(1)}) = \sum_{t=1}^T f_X(x_t; \theta^{(2)}) \text{ a.s. } \rightarrow \theta^{(1)} = \theta^{(2)}.$$

Proof in Appendix A.6.

For the consistency of the MLE for the MEM-J- $\lambda_t$  model, we assume:

**Assumption 2 (Consistency)**

- i.  $\{\theta_0\} \in \text{int}(\Theta)$  with  $\Theta$  compact.
- ii. The MEM-J is identifiable.
- iii. The top Lyapounov exponent  $\gamma_{\theta_0}$  at  $\theta_0$  is strictly negative.
- iv.  $\mathbb{E} \{ \sup_{\theta \in \Theta} [\log f_X(x_t; \theta)] \} < \infty$  and  $\mathbb{E}[l_t(\theta)]$  has a unique maximizer at  $\theta_0$ .
- v.  $\mathbb{E} \left\{ \sup_{\theta \in \Theta} |l_t(\theta) - \tilde{l}_t(\theta)| \right\} = O\left(\frac{1}{t^v}\right)$  for some  $v > 0$ , where  $\tilde{l}_t(\theta)$  is  $l_t(x_t; \theta)$  with the initial value  $\tilde{x}_0$ .

**Theorem 4** *Under Assumption 2, the ML estimator  $\hat{\theta} \xrightarrow{\text{a.s.}} \theta$ , as  $T \rightarrow \infty$ .*

Proof in Appendix A.7.

The crucial assumption for consistency is the identifiability of the model stated in Theorem 3. In addition to the assumptions for Theorem 4, we consider the following set of assumptions to obtain the asymptotic normality of the MLE of the MEM-J- $\lambda_t$ . Denote  $\mathcal{X}_t$  the range of  $X_t$  where  $\mathcal{X}_t$  is a subset of a finite dimensional Euclidean space.

**Assumption 3** (*Asymptotic Normality*)

i.  $l_t(\theta)$  is twice continuously differentiable on  $\mathcal{X}_t \times \Theta$ .

ii. For each  $\theta \in \Theta$ , there exist functions

$$\left| \frac{\partial f_X(x_t; \theta)}{\partial \theta} \right| \leq h_i(x), \quad \left| \frac{\partial^2 f_X(x_t; \theta)}{\partial \theta_i \partial \theta_j} \right| \leq h_{i,j}(x), \quad \left| \frac{\partial^3 f_X(x_t; \theta)}{\partial \theta_i \partial \theta_j \partial \theta_k} \right| \leq h_{i,j,k}(x) \quad i, j, k = 1, \dots, p \quad (24)$$

where  $h_i(x), h_{i,j}(x)$  are integrable and  $h_{i,j,k}(x)$  satisfies

$$\int_{\mathcal{X}} h_{i,j,k}(x) f_X(x; \theta) dx < \infty$$

iii.  $\mathcal{H}(\theta_0) \equiv \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \mathbb{E} \left[ -\frac{\partial^2 l_t(\theta_0)}{\partial \theta \partial \theta'} \right]$  is positive definite.

iv.  $\frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'}$  satisfies the uniform weak law of large numbers (UWLLN).

v. the score  $\{\partial l_t(\theta_0)/\partial \theta, t = 1, 2, \dots\}$  satisfies the CLT with asymptotic variance given by

$$\mathcal{I}(\theta_0) = \lim_{T \rightarrow \infty} \text{Var} \left( T^{-1/2} \sum_{t=1}^T \frac{\partial l_t(\theta_0)}{\partial \theta} \right)$$

vi. the score satisfies  $\mathbb{E} \left[ \frac{\partial l_t(\theta_0)}{\partial \theta} \frac{\partial l_{t+j}(\theta_0)}{\partial \theta'} \right] = 0, j \neq 0$

**Theorem 5** Under Assumption 3, the ML estimator of  $\theta$  of the MEM-J- $\lambda_t$

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \xrightarrow{d} N(0, \mathcal{J}^{-1}(\theta_0))$$

where  $\mathcal{J}(\theta_0) = \lim_{T \rightarrow \infty} \sum_{t=1}^T \mathbb{E} \left[ \frac{\partial l_t(\theta_0)}{\partial \theta} \frac{\partial l_t(\theta_0)}{\partial \theta'} \right]$ , with  $\mathcal{J}(\theta_0) = \mathcal{I}(\theta_0) = \mathcal{H}(\theta_0)$ .

Proof in Appendix A.8.

This is a standard result given the characteristics of model (i.e. stationarity and ergodicity). The differentiability of the Bessel function cannot be directly checked, but we provide some simulation evidence in the following section.

## 5.1 Monte Carlo Simulations

We run a set of Monte Carlo simulations to show that the finite sample distributions of the ML estimates of the MEM-J parameters, which involve numerical evaluations of the derivatives of the modified Bessel function of the second kind with respect to its order, are centered on the true values and they are unimodal, coherently with Theorem 5. We therefore simulate three different specifications with the same  $\mu_t$  as in (17) with no asymmetric effect: HAR-MEM (model in (1)), HAR-MEM-J with constant  $\lambda$  (model in (3)) and HAR-MEM-J- $\lambda_t$ . The algorithm to simulate



pseudo-random variates from a  $K$  density is illustrated in Appendix B. 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 observations. 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- $\lambda_t$ , when the data have been generated with either  $\lambda$  equal to zero (i.e.  $\phi_1 = \phi_2 = \phi_3 = 0$ ) or with a constant  $\lambda$ . The infinite sum of densities required to compute the likelihood, see (35), is truncated at  $\bar{m} = 10$ . Indeed, any choice of  $\bar{m}$  larger than 10 leads to almost identical parameter estimates since for  $\lambda = 0.25$  (which is a realistic value for the jump intensity) the probability of observing more than 10 jumps is of order  $10^{-15}$ . 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  $\lambda_t$ , 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  $X_t$  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  $\mu_t$  are correctly estimated with a small RMSE. It should also be noted that the parameter  $\varsigma$  is not defined under the DGP, but it is estimated when fitting the HAR-MEM-J- $\lambda_t$  on the data as it determines the shape of the  $K$  distribution. This has no consequences on the parameters in  $\mu_t$ , 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  $\mu_t$ . When jumps are absent but the parameter  $\varsigma$  is estimated, the bias and RMSE associated with this parameter are very high, which is a consequence of the lack of identification. The average estimate of  $\varsigma$  is large, suggesting that the average size of the jumps is very small, as their expected size is the reciprocal of  $\varsigma$ . Concluding, when the jumps are absent, and a HAR-MEM-J- $\lambda_t$  is fitted to the data (i.e. over-specification), the estimates of the parameters in  $\mu_t$  are very close to the true values while the estimated jump component is negligible.

When  $\lambda$  is constant, but a HAR-MEM-J- $\lambda_t$  is estimated, see the second 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  $\phi_3$  is close to zero, meaning that the estimated variation in  $\lambda_t$  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  $\phi_1/(1 - \phi_2)$ ,  $\phi_2$  and  $\phi_3$  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- $\lambda_t$  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.

The Monte Carlo simulations confirm the validity of the ML estimation method, made pos-



	$\omega$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\beta$	$\nu$	$\varsigma$	$\frac{\phi_1}{1-\phi_2}$	$\phi_2$	$\phi_3$
DGP: $\lambda = 0$										
	<b>0.001</b>	<b>0.4</b>	<b>0.15</b>	<b>0.1</b>	<b>0.3</b>	<b>20</b>	-	-	-	-
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: $\lambda = 0.25$										
	<b>0.001</b>	<b>0.4</b>	<b>0.15</b>	<b>0.1</b>	<b>0.3</b>	<b>35</b>	<b>20</b>	<b>0.25</b>	-	-
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: $\lambda_t > 0$										
	<b>0.001</b>	<b>0.4</b>	<b>0.15</b>	<b>0.1</b>	<b>0.3</b>	<b>35</b>	<b>20</b>	<b>0.2</b>	<b>0.95</b>	<b>0.1</b>
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
DGP: Log-Normal ( $\mu = -2\sigma^2, \sigma^2 = 0.0625$ )										
	<b>0.001</b>	<b>0.4</b>	<b>0.15</b>	<b>0.1</b>	<b>0.3</b>	-	-	-	-	-
Mean	0.001	0.398	0.150	0.097	0.301	16.967	62.244	0.080	-	-
RMSE	0.000	0.021	0.063	0.023	0.069	0.508	11.885	0.029	-	-

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. If the value of the true parameter is not defined under the DGP, the corresponding RMSE is replaced by the Monte Carlo standard deviation.

sible 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). Finally, the bottom panel of Table 1 reports the Monte Carlo mean and RMSE of the HAR-MEM-J estimates with constant  $\lambda$  under misspecification, i.e. when the innovations follow a log-normal distribution. The parameters  $\mu$  and  $\sigma^2$  are calibrated to guarantee that innovations are centered around 1 with the same variance of the HAR-MEM-J model with  $\lambda = 0.2$ ,  $\nu = 35$  and  $\varsigma = 20$ . Two interesting clues emerge from the Monte Carlo estimates. First, the parameters of  $\mu_t$  are not affected by the misspecification and are centered around the true values. Second, the estimated mixture parameters are such that there is limited mixing effect coming from the  $K$  distribution ( $\lambda \approx 0.08$  and  $\varsigma \approx 62$ ), i.e. the jump component. The Monte Carlo results also show that in case of misspecification the parameter estimates of HAR-MEM-J mixture density are such that the presence of jumps is unlikely. The mean of the  $\hat{\varsigma}$  is three times larger those obtained under HAR-MEM-J DGP cases of comparable variance which indicates that the jump size distribution is much more concentrated and accompanied by a larger variability of the  $\varepsilon_t$  shock. This result suggests that, when the HAR-MEM-J is fitted on leptokurtic time series, the estimated parameters seem to exclude that the excess kurtosis is generated by jumps.

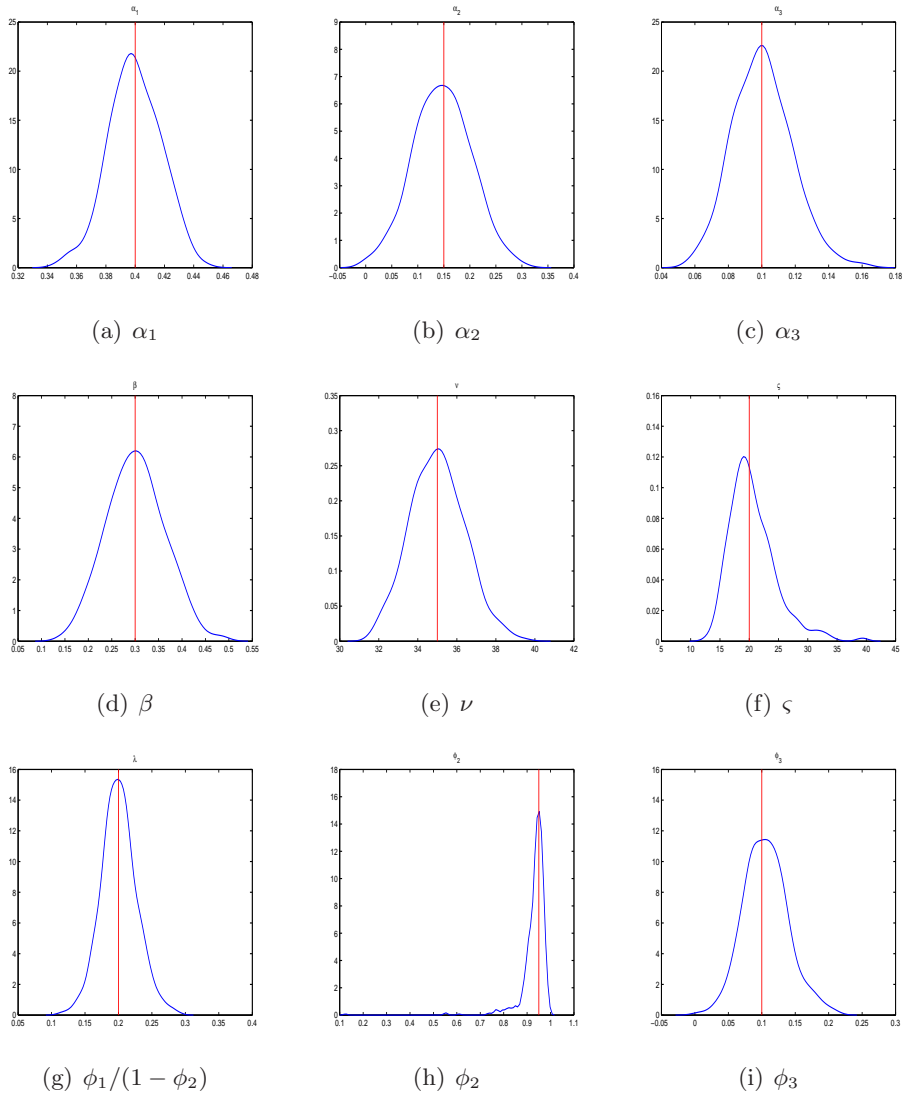


Figure 1: Kernel densities of the Monte Carlo estimates of the HAR-MEM-J- $\lambda_t$ , where  $\lambda_t$  varies according to (19).

## 5.2 Diagnostic Tests

A direct test for a MEM model with jumps as opposed to a model without jumps cannot be easily performed by a likelihood ratio test because the parameter  $\zeta$  is not defined under the null hypothesis.<sup>5</sup> Therefore, similarly to Richardson and Smith (1993) and Harvey (1995), we define a test based on the idea that, the maintained model is the MEM with gamma distributed disturbances, i.e.  $\varepsilon_t \sim \Gamma(1, \nu)$ . The test statistics is computed as

$$J_\Gamma = T \widehat{M}'_\varepsilon \Omega_\varepsilon^{-1} \widehat{M}_\varepsilon \quad (25)$$

<sup>5</sup>In this case simulation based approaches can be used to recover the critical values of the likelihood ratio test, as suggested Hansen (1996). However, this strategy is subject to such a high computational burden that it would nullify the practical usefulness of the MEM-J.

where

$$\widehat{M}_\varepsilon = \frac{1}{T} \sum_{t=1}^T \widehat{m}_{t,\varepsilon}$$

with  $\widehat{m}_{t,\varepsilon}$  containing deviations between the observed values and the theoretical moments:

$$\widehat{m}_{t,\varepsilon} = \begin{pmatrix} \widehat{\varepsilon}_t^3 - \frac{2}{\widehat{\nu}^2} - \frac{3}{\widehat{\nu}} - 1 \\ \widehat{\varepsilon}_t^4 - \frac{6}{\widehat{\nu}^3} - \frac{11}{\widehat{\nu}^2} - \frac{6}{\widehat{\nu}} - 1 \end{pmatrix}.$$

For the weighting matrix  $\Omega_\varepsilon$  we employ the Newey-West estimator

$$\Omega_\varepsilon = \frac{1}{T} \sum_t \widehat{m}_{t,\varepsilon} \widehat{m}'_{t,\varepsilon} + \sum_{i=1}^{b_T} \omega_i (\widehat{C}_i + \widehat{C}'_i)$$

with  $\text{Cov}(\widehat{m}_{t-j,\varepsilon}, \widehat{m}_{t,\varepsilon}) = C_j \neq 0$ , and  $\omega_i$  is a kernel, which must ensure consistency and positive semi-definiteness. The Newey-West estimator normally adopts a Bartlett kernel, i.e.  $\omega_i = 1 - \frac{i}{b_T+1}$ , with  $b_T = \lfloor 0.75T^{1/3} \rfloor$ , where the operator  $\lfloor x \rfloor$  rounds to the largest previous integer value of  $x$ . Under the null hypothesis that the true model is a MEM with Gamma innovations, or, alternatively, there are no jumps in the MEM model, the test statistic is asymptotically distributed as a  $\chi^2$  with two degrees of freedom. Clearly, the rejection of the null hypothesis could signal either the possible presence of jumps or, more generally, the inappropriateness of the Gamma distribution for the innovations.

A similar moment-based-test can be developed under the MEM-J model. In this case, we test if the third and fourth moments of the MEM-J standardized residuals  $\widehat{\eta}_t = X_t/\widehat{\mu}_t$  coincide with those prescribed by the mixture of Gamma and *Kappa* distributions. In this case, the test is

$$J_{mix} = T \widehat{M}_\eta' \Omega_\eta^{-1} \widehat{M}_\eta \quad (26)$$

$$\widehat{M}_\eta = \frac{1}{T} \sum_{t=1}^T \widehat{m}_{t,\eta}$$

with

$$\widehat{m}_{t,\eta} = \begin{pmatrix} \widehat{\eta}_t^3 - E(\eta_t^3) \\ \widehat{\eta}_t^4 - E(\eta_t^4) \end{pmatrix}.$$

where the parametric expressions for  $E(\eta_t^3)$  and  $E(\eta_t^4)$  are provided in the Supplementary material. The weighting matrix has the same form of the previous test, but with  $\widehat{m}_{t,\eta}$  in place of  $\widehat{m}_{t,\varepsilon}$ . Again, under the null hypothesis of correct model specification, the true model is a MEM-J, the test statistics equals, asymptotically, a  $\chi^2_2$ . The joint use of the two tests  $J_\Gamma$  and  $J_{mix}$  allow verifying the appropriateness of the MEM-J model specification and, consequently, of the presence of jumps in the modeled variable. Indeed, if we reject the null under MEM and we do not reject the null under the MEM-J, we are implicitly validating the inclusion of jumps in the MEM.

Finally, we provide a diagnostic test for the jump specification of the MEM-J. If the jump innovations  $\xi_t = E(N_t|\mathcal{F}_t) - \lambda_t$  are serially dependent, then a proper dynamic specification of  $\lambda_t$  is needed. Indeed, if the model is well specified, the  $\xi_t$  should be orthogonal to  $\mathcal{F}_{t-1}$ , and should

also display no serial correlation. Hence, given  $\hat{\xi}_t$ , the test of no autocorrelation is the test of the null hypothesis  $\mathcal{H}_0 : \delta_1 = \delta_2 = \dots = \delta_r = 0$  in the following auxiliary regression

$$\hat{\xi}_t = \delta_0 + \delta_1 \hat{\xi}_{t-1} + \dots + \delta_m \hat{\xi}_{t-r} + w_t. \quad (27)$$

We interpret the rejection of the null hypothesis as a consequence of a misspecification in the jump intensity dynamic.

## 6 Empirical Results

The MEM-J aims at capturing the occurrence of large and unpredictable movements of the dependent variable. This corresponds to one of the features of the realized volatility sequences, a variable commonly analyzed in the literature dealing with multiplicative error models, see Engle and Gallo (2006), Lanne (2006), and Andres and Harvey (2012) among many others.

### 6.1 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 daily quadratic variation can be estimated by the realized variance (or realized volatility), as

$$RV_t = \sum_{j=1}^M r_{t,j}^2 \quad t = 1, \dots, T \quad (28)$$

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  $M \rightarrow \infty$  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 daily integrated variance, the bipower variation, defined as

$$BPV_t = \frac{\pi}{2} \sum_{j=2}^M |r_{t,j}| |r_{t,j-1}| \quad t = 1, \dots, T. \quad (29)$$

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

The empirical analysis reported in the following sections is conducted with the daily 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,  $RM_t$ , is expressed as daily volatility, i.e. the square root of BPV,  $RM_t = \sqrt{BPV_t}$ . The descriptive statistics of both datasets<sup>6</sup> 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  $\mu_t$ : 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:

- No jumps:  $\phi_1 = \phi_2 = \phi_3 = \varsigma = 0$ ;
- Constant jump intensity: from equation (19) we set  $\lambda_t = \phi_1$  and  $\phi_2 = \phi_3 = 0$ ;
- Time-varying jump intensity: with  $\lambda_t$  evolving as in equation (19).

Due to space constraints, only the estimates of the parameter of the AHAR-MEM-J- $\lambda_t$  are reported in the paper.<sup>7</sup> To compare the alternative models we consider two different approaches. Firstly, we pursue a full-sample evaluation approach, where the MEM and MEM-J specifications are compared with respect to their fit on the empirical data by using both parameter restrictions tests and the diagnostic tests introduced in Section 5.2. 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.

## 6.2 Estimation results

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<sup>8</sup>

$$\hat{\varepsilon}_t = \frac{RM_t}{\hat{\mu}_t}. \quad (30)$$

Beside the tests of Section 5.2, we also make use of more traditional tests, in particular for the detection of residual serial correlation. However, since the standard diagnostic  $Q$  statistic is designed for residuals that are assumed normally distributed, we normalize the residuals with

<sup>6</sup>See Table 1 in the supplementary document.

<sup>7</sup>The parameter estimates of the other sub-models are in the Tables 2-4 in the supplementary document.

<sup>8</sup>As an alternative, model residuals might be computed by standardization of  $RM_t$  with respect to its expected value, involving the impact of  $\mu_t$  and (when present)  $Z_t$ . In this case, innovations are defined as Pearson's residuals

$$\frac{RM_t - \mathbb{E}[RM_t|\mathcal{F}_{t-1}]}{\sqrt{\mathbb{V}[RM_t|\mathcal{F}_{t-1}]^{1/2}}}.$$

$\hat{\varepsilon}_t^* = F_N^{-1}[F_\Gamma(\hat{\varepsilon}_t)]$ , for  $t = 1, \dots, T$ , where  $F_N(\cdot)$  and  $F_\Gamma(\cdot)$  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- $\lambda_t$ . Note that for the models with asymmetry, preliminary estimates lead to the exclusion of the HAR terms for asymmetry. Thus, the HAR applies only to the lagged values of the realized volatility. We start by looking at the presence of residual correlation. 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 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. Indeed, the estimated parameters in  $\mu_t$  of the AHAR-MEM-J are close to those of the AHAR-MEM even though the parameters associated with the jumps,  $\varsigma$  and  $\lambda$  are statistically significant for all series considered.<sup>9</sup> We also consider the tests of correct specification introduced in Section 5.2. Table 2 shows that, if we test for the appropriateness of the Gamma distribution for the MEM innovations, the null is rejected in 16 out of 23 cases at 10% significance level. On the contrary, the  $J_{mix}$  test, based on the AHAR-MEM-J residuals, leads to a rejection of the null in only 7 cases. Moreover, the  $D_r$  test, based on the jump innovations of the AHAR-MEM-J with constant  $\lambda$ , shows evidence of serial correlations for all series (with the exception of TXN). Differently, when we adopt the AHAR-MEM-J- $\lambda_t$  specification, the jump innovations with serial dependence decrease to only 6 out of 23. Therefore, the diagnostic procedures provide a strong support to the assumption that that the AHAR-MEM-J- $\lambda_t$  is correctly specified.

Table 3 reports the parameter estimates of AHAR-MEM-J- $\lambda_t$  model. The parameters in  $\mu_t$  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  $\alpha_1$ . Differently, the impact of last week and last month average of  $RM_t$  is more heterogeneous across stocks, with some cases of reduced significance. Regarding the estimates of the parameters in the jump component, the unconditional mean of  $\lambda_t$  is between 0.15 and 0.20 in most cases, and there are not relevant differences between stock indexes and individual stocks. Notably, the unconditional values of  $\lambda_t$  are always significant and they are very close to the estimates of  $\lambda$  obtained fitting the AHAR-MEM-J with constant intensity. 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  $RM_t$ .

---

<sup>9</sup>See Tables 3 and 4 in the supplementary document.

	AMEM			AHAR-MEM					AHAR-MEM-J					AHAR-MEM-J- $\lambda_t$			
	$Q_1$	$Q_{10}$	$Q_{22}$	$Q_1$	$Q_{10}$	$Q_{22}$	$LR$	$J_\Gamma$	$Q_1$	$Q_{10}$	$Q_{22}$	$J_{mix}$	$D_5$	$Q_1$	$Q_{10}$	$Q_{22}$	$D_5$
SP500	0.004	0.000	0.000	0.904	0.478	0.388	70.29 <sup>a</sup>	7.46 <sup>b</sup>	0.881	0.444	0.443	2.48	17.28 <sup>a</sup>	0.795	0.347	0.417	5.08 <sup>b</sup>
FTSE	0.001	0.000	0.000	0.911	0.646	0.012	60.46 <sup>a</sup>	2.15	0.225	0.001	0.0000	1.15	37.00 <sup>a</sup>	0.132	0.000	0.000	2.81
DAX	0.000	0.000	0.000	0.220	0.197	0.000	60.40 <sup>a</sup>	13.83 <sup>a</sup>	0.254	0.008	0.000	1.28	17.45 <sup>a</sup>	0.184	0.003	0.000	0.97
DJIA	0.003	0.000	0.000	0.913	0.444	0.585	57.96 <sup>a</sup>	6.05 <sup>b</sup>	0.475	0.544	0.503	1.08	16.05 <sup>a</sup>	0.618	0.290	0.305	1.59
NSDQ	0.001	0.000	0.000	0.734	0.107	0.094	99.88 <sup>a</sup>	2.12	0.625	0.521	0.071	0.11	14.77 <sup>a</sup>	0.744	0.583	0.069	1.48
CAC	0.000	0.000	0.000	0.317	0.089	0.000	53.05 <sup>a</sup>	4.74 <sup>c</sup>	0.228	0.000	0.000	4.69 <sup>c</sup>	34.04 <sup>a</sup>	0.079	0.000	0.000	2.51
BOVSP	0.007	0.000	0.000	0.759	0.623	0.225	80.81 <sup>a</sup>	9.95 <sup>a</sup>	0.420	0.110	0.089	39.51 <sup>a</sup>	17.40 <sup>a</sup>	0.792	0.082	0.068	3.87 <sup>b</sup>
BA	0.039	0.000	0.000	0.798	0.773	0.045	37.44 <sup>a</sup>	10.55 <sup>a</sup>	0.445	0.077	0.030	28.13 <sup>a</sup>	9.35 <sup>a</sup>	0.968	0.090	0.038	0.19
BAC	0.083	0.000	0.001	0.712	0.696	0.689	59.75 <sup>a</sup>	1.32	0.573	0.744	0.646	1.33	7.93 <sup>a</sup>	0.314	0.741	0.700	0.07
C	0.037	0.000	0.001	0.697	0.713	0.252	47.02 <sup>a</sup>	6.15 <sup>b</sup>	0.623	0.372	0.274	6.39 <sup>b</sup>	47.13 <sup>a</sup>	0.213	0.378	0.225	5.98 <sup>b</sup>
CAT	0.005	0.000	0.002	0.849	0.658	0.889	44.30 <sup>a</sup>	4.22	0.577	0.687	0.477	2.38	7.72 <sup>a</sup>	0.855	0.747	0.593	0.00
FDX	0.023	0.000	0.000	0.744	0.174	0.470	60.45 <sup>a</sup>	11.67 <sup>a</sup>	0.386	0.464	0.179	21.39 <sup>a</sup>	7.503 <sup>a</sup>	0.815	0.692	0.402	4.11 <sup>b</sup>
HON	0.052	0.002	0.003	0.751	0.841	0.136	23.80 <sup>a</sup>	10.92 <sup>a</sup>	0.223	0.730	0.039	8.78 <sup>b</sup>	12.58 <sup>a</sup>	0.264	0.571	0.029	0.07
HPQ	0.039	0.000	0.000	0.768	0.451	0.424	56.44 <sup>a</sup>	1.77	0.624	0.878	0.448	1.02	4.55 <sup>b</sup>	0.693	0.979	0.591	1.05
IBM	0.097	0.007	0.026	0.989	0.964	0.521	22.40 <sup>a</sup>	5.99 <sup>b</sup>	0.288	0.686	0.373	1.57	13.96 <sup>a</sup>	0.231	0.726	0.306	0.19
JPM	0.004	0.000	0.000	0.651	0.802	0.309	53.64 <sup>a</sup>	4.71 <sup>c</sup>	0.395	0.297	0.166	1.42	13.91 <sup>a</sup>	0.230	0.197	0.122	0.94
KFT	0.003	0.000	0.000	0.634	0.596	0.916	59.10 <sup>a</sup>	14.44 <sup>a</sup>	0.065	0.033	0.195	1.01	3.95 <sup>b</sup>	0.287	0.113	0.365	6.34 <sup>b</sup>
PEP	0.386	0.023	0.000	0.625	0.384	0.005	13.83 <sup>a</sup>	4.34	0.138	0.001	0.004	0.19	13.65 <sup>a</sup>	0.424	0.000	0.001	0.88
PG	0.095	0.001	0.000	0.423	0.914	0.051	41.37 <sup>a</sup>	5.19 <sup>c</sup>	0.746	0.810	0.162	0.01	11.50 <sup>a</sup>	0.781	0.764	0.178	0.28
T	0.023	0.000	0.000	0.862	0.668	0.033	34.94 <sup>a</sup>	5.48 <sup>c</sup>	0.456	0.198	0.000	0.25	20.67 <sup>a</sup>	0.321	0.156	0.000	0.34
TWX	0.070	0.002	0.004	0.700	0.987	0.493	40.95 <sup>a</sup>	9.68 <sup>a</sup>	0.202	0.434	0.207	1.87	16.97 <sup>a</sup>	0.518	0.504	0.287	0.01
TXN	0.053	0.002	0.001	0.401	0.892	0.156	40.59 <sup>a</sup>	8.39 <sup>b</sup>	0.719	0.640	0.138	9.88 <sup>a</sup>	7.61	0.532	0.562	0.153	4.89 <sup>b</sup>
WFC	0.488	0.000	0.000	0.083	0.406	0.355	65.20 <sup>a</sup>	1.98	0.927	0.259	0.298	1.30	34.68 <sup>a</sup>	0.837	0.352	0.390	1.22

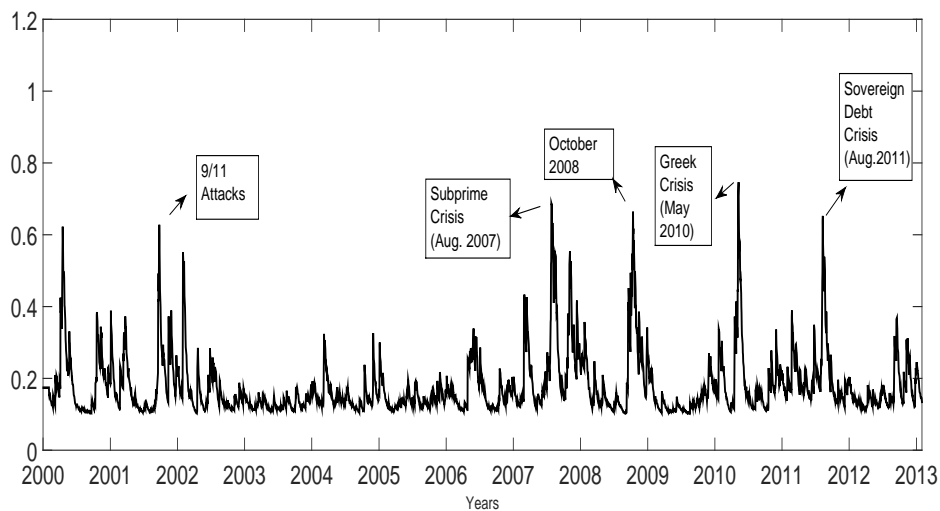
Table 2: Residual Statistics and Tests. The upper part of the table reports the results for several stock indexes while the lower part refers to 16 NYSE stocks.  $a$ ,  $b$  and  $c$  stand for significance at 1%, 5% and 10% respectively.  $Q_1$ ,  $Q_{10}$  and  $Q_{22}$  are the p-values of the Ljung-Box test for absence of autocorrelation in the normalized residuals, where the latter are computed as  $\hat{\varepsilon}_t^* = F_N^{-1}[F_\Gamma(\hat{\varepsilon}_t)]$  with  $\hat{\varepsilon}_t = \frac{RM_t}{\hat{\mu}_t}$ .  $J_\Gamma$  and  $J_{mix}$  are the test statistics for the null hypotheses that the errors have a Gamma distribution and a mixture of Gamma and Kappa, respectively.  $D_5$  denotes the test statistic for the null of no autocorrelation up to the 5-th lag in the jump innovations (both for the case of constant and dynamic jump intensity).

	$\omega$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\beta$	$\gamma$	$\nu$	$\varsigma$	$\frac{\phi_1}{1-\phi_2}$	$\phi_2$	$\phi_3$
SP500	0.0003 <sup>a</sup>	0.3041 <sup>a</sup>	0.1727 <sup>a</sup>	0.1098 <sup>a</sup>	0.3235 <sup>a</sup>	0.1087 <sup>a</sup>	23.1069 <sup>a</sup>	15.3934 <sup>a</sup>	0.1739 <sup>a</sup>	0.9379 <sup>a</sup>	0.1275 <sup>c</sup>
FTSE100	0.0002 <sup>a</sup>	0.2814 <sup>a</sup>	0.2082 <sup>b</sup>	0.1383 <sup>a</sup>	0.3130 <sup>a</sup>	0.0590 <sup>a</sup>	26.4544 <sup>a</sup>	10.6540 <sup>a</sup>	0.1448 <sup>a</sup>	0.8550 <sup>b</sup>	0.1919 <sup>a</sup>
DAX	0.0003 <sup>a</sup>	0.2985 <sup>a</sup>	0.1476 <sup>a</sup>	0.1179 <sup>a</sup>	0.3749 <sup>a</sup>	0.0721 <sup>a</sup>	26.2707 <sup>a</sup>	12.7932 <sup>a</sup>	0.1691 <sup>a</sup>	0.9391 <sup>a</sup>	0.1303 <sup>b</sup>
DJIA	0.0003 <sup>a</sup>	0.2984 <sup>a</sup>	0.1811 <sup>a</sup>	0.0988 <sup>a</sup>	0.3360 <sup>a</sup>	0.0964 <sup>a</sup>	22.4470 <sup>a</sup>	14.0424 <sup>a</sup>	0.1616 <sup>a</sup>	0.9146 <sup>a</sup>	0.2006 <sup>a</sup>
NSDQ	0.0003 <sup>a</sup>	0.3053 <sup>a</sup>	0.1930 <sup>a</sup>	0.1583 <sup>a</sup>	0.2637 <sup>a</sup>	0.0996 <sup>a</sup>	23.8215 <sup>a</sup>	10.7619 <sup>a</sup>	0.1490 <sup>a</sup>	0.9698 <sup>a</sup>	0.1006
CAC	0.0002 <sup>a</sup>	0.2809 <sup>a</sup>	0.1377 <sup>b</sup>	0.0978 <sup>a</sup>	0.4192 <sup>a</sup>	0.0799 <sup>a</sup>	27.5802 <sup>a</sup>	11.8415 <sup>a</sup>	0.1581 <sup>a</sup>	0.9064 <sup>a</sup>	0.1655 <sup>c</sup>
BOVESPA	0.0009 <sup>a</sup>	0.3300 <sup>a</sup>	0.1298 <sup>b</sup>	0.1564 <sup>a</sup>	0.2780 <sup>a</sup>	0.0725 <sup>a</sup>	22.1476 <sup>a</sup>	20.9309 <sup>a</sup>	0.2147 <sup>a</sup>	0.8663 <sup>a</sup>	0.1238 <sup>b</sup>
BA	0.0006 <sup>a</sup>	0.3882 <sup>a</sup>	0.1106 <sup>a</sup>	0.1299 <sup>a</sup>	0.3183 <sup>a</sup>	0.0243 <sup>a</sup>	36.1310 <sup>a</sup>	20.3827 <sup>a</sup>	0.1999 <sup>a</sup>	0.8179 <sup>a</sup>	0.1686 <sup>a</sup>
BAC	0.0002 <sup>a</sup>	0.4837 <sup>a</sup>	0.0942 <sup>c</sup>	0.1256 <sup>a</sup>	0.2614 <sup>a</sup>	0.0384 <sup>a</sup>	33.6480 <sup>a</sup>	21.4976 <sup>a</sup>	0.1779 <sup>b</sup>	0.9719 <sup>a</sup>	0.1053 <sup>a</sup>
C	0.0002 <sup>a</sup>	0.4469 <sup>a</sup>	0.1037 <sup>c</sup>	0.1004 <sup>a</sup>	0.3231 <sup>a</sup>	0.0253 <sup>a</sup>	41.0615 <sup>a</sup>	14.6146 <sup>a</sup>	0.1034 <sup>b</sup>	0.9826 <sup>a</sup>	0.0823 <sup>a</sup>
CAT	0.0008 <sup>a</sup>	0.4082 <sup>a</sup>	0.2305 <sup>a</sup>	0.1206 <sup>a</sup>	0.1712 <sup>a</sup>	0.0407 <sup>a</sup>	33.5859 <sup>a</sup>	28.3271 <sup>a</sup>	0.1849 <sup>a</sup>	0.7187 <sup>a</sup>	0.3344 <sup>a</sup>
FDX	0.0005 <sup>a</sup>	0.4009 <sup>a</sup>	0.1516 <sup>a</sup>	0.1593 <sup>a</sup>	0.2402 <sup>a</sup>	0.0364 <sup>a</sup>	33.1796 <sup>a</sup>	30.6452 <sup>a</sup>	0.2181 <sup>a</sup>	0.8652 <sup>a</sup>	0.2091 <sup>a</sup>
HON	0.0006 <sup>a</sup>	0.3758 <sup>a</sup>	0.1052	0.0892 <sup>a</sup>	0.3625 <sup>a</sup>	0.0491 <sup>a</sup>	33.7423 <sup>a</sup>	22.9547 <sup>a</sup>	0.1654 <sup>a</sup>	0.8150 <sup>a</sup>	0.2750 <sup>a</sup>
HPQ	0.0007 <sup>a</sup>	0.3992 <sup>a</sup>	0.1709 <sup>a</sup>	0.1170 <sup>a</sup>	0.2420 <sup>a</sup>	0.0454 <sup>a</sup>	34.2396 <sup>a</sup>	20.3591 <sup>a</sup>	0.1877 <sup>a</sup>	0.7976 <sup>a</sup>	0.2495 <sup>a</sup>
IBM	0.0004 <sup>a</sup>	0.3994 <sup>a</sup>	0.1400 <sup>a</sup>	0.0728 <sup>a</sup>	0.3304 <sup>a</sup>	0.0319 <sup>a</sup>	36.8106 <sup>a</sup>	17.9724 <sup>a</sup>	0.1269 <sup>a</sup>	0.9417	0.1669 <sup>c</sup>
JPM	0.0004 <sup>a</sup>	0.4373 <sup>a</sup>	0.2076 <sup>a</sup>	0.1199 <sup>a</sup>	0.1949 <sup>a</sup>	0.0343 <sup>a</sup>	32.7366 <sup>a</sup>	32.0963 <sup>a</sup>	0.1820 <sup>a</sup>	0.9544 <sup>a</sup>	0.1001
KFT	0.0007 <sup>a</sup>	0.3293 <sup>a</sup>	0.2987 <sup>a</sup>	0.1610 <sup>a</sup>	0.1476 <sup>a</sup>	0.0050 <sup>a</sup>	30.9573 <sup>a</sup>	29.4368 <sup>a</sup>	0.2901 <sup>a</sup>	0.8890 <sup>a</sup>	0.1894 <sup>b</sup>
PEP	0.0003 <sup>a</sup>	0.3214	0.0752	0.0620 <sup>c</sup>	0.4980 <sup>a</sup>	0.0358 <sup>a</sup>	33.7365 <sup>a</sup>	37.9719 <sup>a</sup>	0.1929 <sup>a</sup>	0.8388 <sup>a</sup>	0.3527 <sup>a</sup>
PG	0.0006 <sup>a</sup>	0.3773 <sup>a</sup>	0.2128 <sup>a</sup>	0.1070 <sup>a</sup>	0.2269 <sup>a</sup>	0.0320 <sup>a</sup>	36.2161 <sup>a</sup>	19.5642 <sup>a</sup>	0.1655 <sup>a</sup>	0.7994 <sup>a</sup>	0.2869 <sup>b</sup>
T	0.0004 <sup>a</sup>	0.3720 <sup>a</sup>	0.1480 <sup>a</sup>	0.1078 <sup>a</sup>	0.3226 <sup>a</sup>	0.0264 <sup>a</sup>	37.4439 <sup>a</sup>	19.2182 <sup>a</sup>	0.1364 <sup>a</sup>	0.8306 <sup>a</sup>	0.2834 <sup>c</sup>
TWX	0.0005 <sup>a</sup>	0.3894 <sup>a</sup>	0.1668 <sup>b</sup>	0.1220 <sup>a</sup>	0.2714 <sup>a</sup>	0.0264 <sup>a</sup>	44.4043 <sup>a</sup>	31.0669 <sup>a</sup>	0.1638	0.8284 <sup>a</sup>	0.2562 <sup>a</sup>
TXN	0.0009 <sup>a</sup>	0.3836 <sup>a</sup>	0.2001 <sup>b</sup>	0.1283 <sup>a</sup>	0.2260 <sup>b</sup>	0.0323 <sup>a</sup>	33.0913 <sup>a</sup>	35.8256 <sup>b</sup>	0.1488 <sup>a</sup>	0.7406 <sup>a</sup>	0.4056 <sup>a</sup>
WFC	0.0003 <sup>a</sup>	0.4465 <sup>a</sup>	0.1782 <sup>a</sup>	0.1127 <sup>a</sup>	0.2320 <sup>a</sup>	0.0254 <sup>a</sup>	34.6772 <sup>a</sup>	21.6895 <sup>a</sup>	0.2142	0.8912 <sup>a</sup>	0.2072 <sup>a</sup>

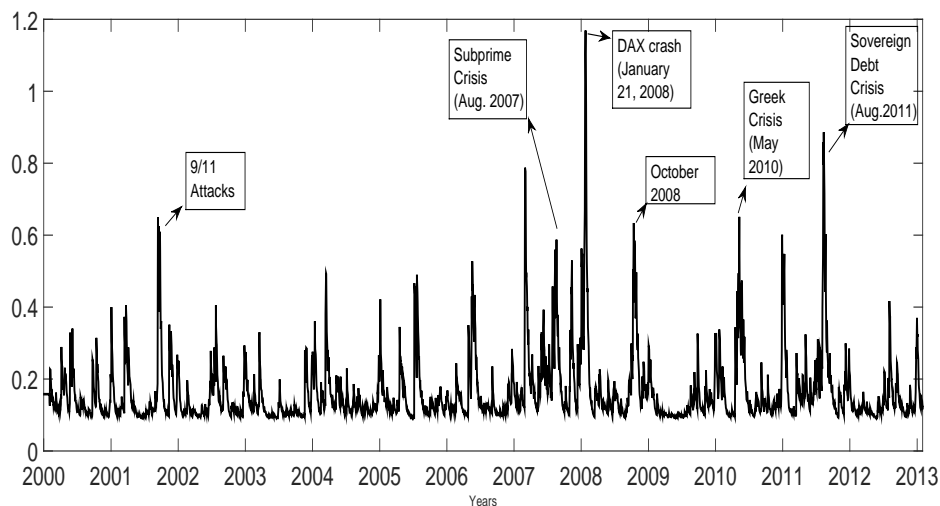
Table 3: Estimates of the Asymmetric HAR-MEM-J with time-varying  $\lambda$ , see (19). The upper part of the table reports the results for several stock indexes while the lower part refers to 16 NYSE stocks.  $a$ ,  $b$  and  $c$  stand for significance at 1%, 5% and 10% respectively.



Interestingly, most markets and stocks, among those considered, display estimates of  $\phi_2$  larger than 0.8, suggesting persistence in jump arrivals. Indeed, out of 23 stocks and indexes, only two present values of  $\phi_2$  below 0.8 and only one below 0.75. The sensitivity to news arrivals, measured by the parameter  $\phi_3$ , is statistically significant for most series. The only exceptions are NASDAQ and JPM. This might challenge the appropriateness of the time-varying jump intensity specification for those two cases. Looking at the shape parameter estimates,  $\nu$ , they are much larger for the AHAR-MEM-J- $\lambda_t$  than for the AHAR-MEM model without jumps. Indeed, the variance of  $\varepsilon_t$ , that is equal to  $\frac{1}{\nu}$ , sensibly reduces when jumps are included. As shown in (16), for a given level of the conditional variance of  $\eta_t$  and a given arrival probability  $\lambda_t$ , there is an inverse relationship between  $\nu$  and  $\varsigma$ .



(a)  $\hat{\lambda}_t$  of S&P500



(b)  $\hat{\lambda}_t$  of CAC 40

Figure 2: Expected jump intensity in the AHAR-MEM-J- $\lambda_t$  model for S&P500 and CAC 40.

Moreover, since  $\varsigma$  is much smaller than  $\nu$  in all cases, it follows that the variance of the jump terms is several times larger than that of  $\varepsilon_t$ . Therefore variation in volatility can be attributed to a combination of a component with relatively low variability,  $\varepsilon_t$ , and another one able to generate potentially large increments,  $Z_t$ . This is in accordance with the findings of Todorov and Tauchen (2011) who show that volatility can be well approximated by a pure jump process with infinite variation. Finally, the estimates of  $\nu$  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.

The plots in Figure 2 report two examples of estimated expected jump intensity,  $\hat{\lambda}_t$ . The expected jump is very close to 0, 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 debt crisis in 2011. Notably, the DAX crash on January 21, 2008 seem to be more relevant in France compared to the US market, a somewhat expected result. This confirms the model reacts to the occurrence of extreme events, this is clear for known and globally relevant facts, but is not limited to them.

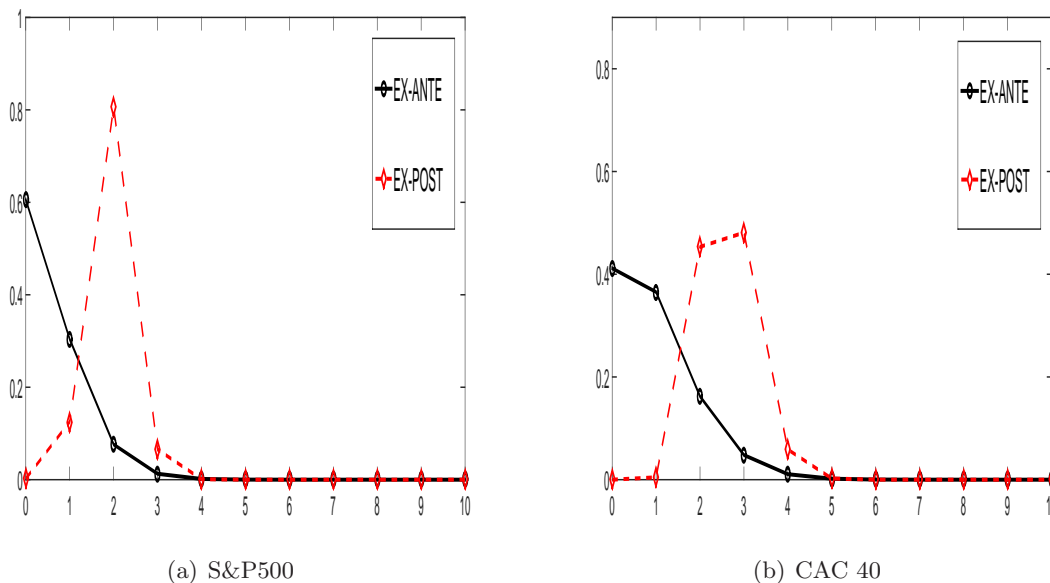


Figure 3: Ex-ante (in black-dots) and ex-post (in red-diamonds) jump probabilities obtained from the AHAR-MEM-J- $\lambda_t$  for S&P500 on October 10, 2008 and for CAC 40 on January 22, 2008.

With the AHAR-MEM-J- $\lambda_t$ , we can study the difference between the ex-ante and ex-post probabilities of jumps during a given day. This can be done comparing  $P(N_t = m|\mathcal{F}_{t-1}; \theta)$  with  $P(N_t = m|\mathcal{F}_t; \theta)$ , where the latter is obtained by the Bayes law in (21). Figure 3 reports these probabilities for different  $m$  values (i.e. the number of jumps) on days characterized by large volatility increments such as on October 10, 2008 for S&P500 and on January 22, 2008 for CAC 40 (DAX crash). In particular, the *ex-ante* probability of observing at least one jump is 40% for

S&P500 and 60% for CAC 40, while *ex-post* this probability increases to almost 100% in both cases, since  $P(N_t = 0|\mathcal{F}_t) \approx 0$  and the jump probability distribution is shifted to the right.

### 6.3 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 jumps in prices from the volatility dynamics. For example, an 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 when sampling too sparsely. 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- $\lambda_t$  with  $RM_t = \sqrt{MedRV_t}$ , i.e. the median realized volatility computed as follows ,

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

where  $\text{med}(\cdot)$  is the median. This is less efficient than *BPV* but more powerful in filtering the price jumps, see Andersen et al. (2012a). Notably, this replacement does not lead to any significant change in the estimates of the AHAR-MEM-J- $\lambda_t$  parameters<sup>10</sup> and in the estimates of the dynamics of the volatility jump intensity. Finally, the significance in the jump intensity could be influenced by possible misspecifications in  $\mu_t$ . For example, Corsi et al. (2010) show that lagged price jumps significantly affect future values of volatility. Therefore, we also consider a model specification where lagged price jumps are included in the conditional dynamics of  $\mu_t$ . The document with Supplementary material presents the details about this model specification and Table 7 shows that the parameters driving the jump intensity are substantially unaffected and the coefficients associated with the lagged price jumps in  $\mu_t$  are often insignificant.

## 7 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 the value  $v(\alpha)$  satisfying

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

---

<sup>10</sup>See Table 6 in the Supplementary document.

where  $\Pr\{\cdot|\mathcal{F}_{t-1}\}$  denotes the conditional distribution at date  $t$  of the one-step-ahead volatility. Therefore,  $v(\alpha)$  is the realized volatility level that may be exceeded with probability  $\alpha$ . 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 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). For example, VIX options associated with large strikes (e.g. above 70%) are generally traded also in non-turbulent periods, meaning that the market assigns a non-null probability to the occurrence of tail events in volatility.

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 evaluate 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  $RM_t$  as

$$y_t = F(RM_t|\mathcal{F}_{t-1}) = \int_0^{RM_t} f_X(x_t|\mathcal{F}_{t-1}) dx_t,$$

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

$$s_t = \Phi^{-1}(y_t)$$

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

$$s_t^* = \begin{cases} \text{VolaR} & \text{if } s_t \leq \text{VolaR} \\ s_t & \text{if } s_t > \text{VolaR}. \end{cases} \quad (31)$$

A tail coverage test can be derived using the LR principle based on the censored normal density of  $s_t^*$ . Under the null of correct tail coverage the test statistic is distributed as  $\chi^2(2)$  as it corresponds to a test for comparing the mean and variance estimated from the truncated likelihood of  $s_t^*$  to those expected under correct coverage, i.e. zero mean and unit variance. 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  $\varepsilon_t$  follows a Generalized Gamma, as in Lunde (1999) and Andres and Harvey (2012).
- AHAR-MEM- $\bar{\nu}_t$ : where the variance of  $\varepsilon_t$ , i.e. the parameter  $\bar{\nu}_t = \frac{1}{\nu_t}$ , follows a GARCH(1,1) process. To the best of our knowledge, this is a new specification for the innovation term in the MEM framework.
- 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.<sup>11</sup> We also estimate the HAR-V-J of Caporin et al. (2014b) to consider a model with jumps which is not based on a multiplicative structure. It clearly emerges that the AHAR-MEM specifications with jumps outperform the corresponding models without jumps in estimating the VolaR. Indeed, all the MEM specifications without jumps and Gamma distribution for  $\varepsilon_t$  strongly reject the null hypothesis of correct specification of the upper quantiles. Interestingly, also the AHAR-MEM- $\bar{\nu}_t$  model results to be poorly designed to capture tail events. Letting the conditional variance of  $\varepsilon_t$  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  $\varepsilon_t$ . 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  $\lambda$ , provides a good fit of the VolaR. In only two cases, BOVESPA and BA, the presence of jumps with time-invariant intensity does not succeed in correctly estimating the VolaR. The introduction of the time-varying jump intensity decisively improves the performances for the CAC40 index, and for BA, FDX, HPQ, JPM, PG and WFG; nevertheless, we also have a case with a decrease in performances, TXN; we interpret this as a possible overfitting issue. Overall, only for BOVESPA, a model without jumps seems to be appropriate. Figure 4 provides a possible explanation for the good performance of the model with jumps. 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  $Z_t$ . On the contrary, the model with time-varying  $\bar{\nu}_t$  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 a HAR specification with time-varying jump

---

<sup>11</sup>See Tables 7-9 in the Supplementary document

	I	II	III	IV	V	VI	VII	VIII
SP500	0.0000	0.0000	0.0000	0.0080	0.0000	0.5181	0.4046	0.3651
FTSE	0.0000	0.0000	0.0000	0.4882	0.0000	0.7394	0.2102	0.0586
DAX	0.0000	0.0000	0.0000	0.0241	0.0001	0.6660	0.1850	0.1601
DJIA	0.0000	0.0000	0.0000	0.0010	0.0000	0.4189	0.5941	0.7272
NSDQ	0.0000	0.0000	0.0000	0.6016	0.0000	0.6651	0.3885	0.8014
CAC	0.0000	0.0000	0.0000	0.0013	0.0004	0.0984	0.0657	0.3590
BVSP	0.0000	0.0000	0.0000	0.5952	0.0519	0.8380	0.0041	0.0013
BA	0.0000	0.0000	0.0000	0.0229	0.0000	0.5486	0.0423	0.2172
BAC	0.0000	0.0000	0.0000	0.0268	0.0000	0.0008	0.1048	0.9663
C	0.0000	0.0000	0.0000	0.0045	0.0000	0.0217	0.2208	0.4735
CAT	0.0000	0.0000	0.0000	0.0921	0.0000	0.5137	0.1771	0.2727
FDX	0.0000	0.0000	0.0000	0.8566	0.0000	0.5309	0.0564	0.2371
HON	0.0000	0.0000	0.0000	0.0013	0.0000	0.0369	0.2038	0.1456
HPQ	0.0000	0.0000	0.0000	0.5133	0.0000	0.0621	0.0753	0.7876
IBM	0.0000	0.0000	0.0000	0.0038	0.0000	0.5957	0.4022	0.2478
JPM	0.0000	0.0000	0.0000	0.0856	0.0000	0.0059	0.0740	0.4750
KFT	0.0000	0.0000	0.0000	0.0112	0.0000	0.6487	0.3190	0.1899
PEP	0.0000	0.0000	0.0000	0.0000	0.0000	0.6652	0.1531	0.7471
PG	0.0000	0.0000	0.0000	0.0123	0.0000	0.9482	0.0510	0.1457
T	0.0000	0.0000	0.0000	0.0078	0.0000	0.5549	0.5906	0.1448
TWX	0.0000	0.0000	0.0000	0.0023	0.0000	0.0487	0.4681	0.8184
TXN	0.0000	0.0000	0.0000	0.0001	0.0000	0.5385	0.2756	0.0164
WFC	0.0000	0.0000	0.0000	0.4959	0.0000	0.0128	0.0691	0.6529

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- $\bar{\nu}_t$  (III), M-AHAR-MEM (IV), AHAR-MEM-GG (V), HAR-V-J (VI), AHAR-MEM-J (VII) and AHAR-MEM-J- $\lambda_t$  (VIII).

intensity on  $\log RM_t$ , provides a good fitting of the tails of the volatility distribution.<sup>12</sup> In this case, the null hypothesis cannot be rejected at 5% significance level for all indexes and for many individual stocks.

<sup>12</sup>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.

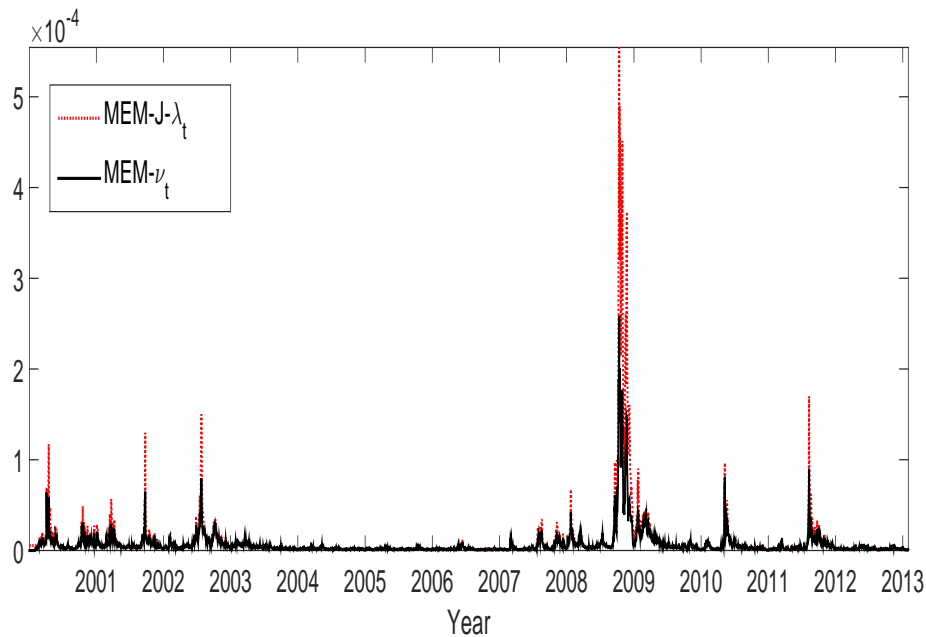


Figure 4: Conditional variance of  $RM_t$  of *S&P500* obtained with the AHAR-MEM-J- $\lambda_t$ , dashed line, and with the AHAR-MEM- $\nu_t$ , solid line.

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 observations while for the individual stocks is approximately 1,400 observations. Given the sample dimensions used in the estimation, the uncertainty in the parameters estimates is expected to lead to a fairly limited effect on the testing results. When the AMEM and AHAR-MEM are used, the null hypothesis is always rejected at 5% significance level, with the exception of FTSE 100. The AHAR-MEM- $\nu_t$  is similar, with no rejections of the null. Slightly better results are obtained when the Generalized Gamma is adopted. In 4 out of 23 cases, the null hypothesis cannot be rejected. A similar performance is also achieved with the M-AHAR-MEM. Things are partially different when moving to the AHAR-MEM-J with constant  $\lambda$ . Indeed, by including the jumps (with constant intensity) the null is not rejected in 11 cases. A slightly better performance is achieved with the HAR-V-J model as the null hypothesis cannot be rejected in 12 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- $\lambda_t$  we observe only one rejection at the 1% confidence level, 3 at the 5% level and 6 at the 10% level. The out-of-sample results confirm the adequacy of AHAR-MEM-J- $\lambda_t$  in predicting the presence of volatility jumps which turn out to be of crucial importance in forecasting the *VolaR*.

	I	II	III	IV	V	VI	VII	VIII
S&P500	0.0000	0.0000	0.0000	0.0035	0.0002	0.3503	0.0048	0.0811
FTSE 100	0.5577	0.1526	0.0000	0.0615	0.5078	0.0000	0.3551	0.6993
DAX	0.0010	0.0002	0.0000	0.0436	0.4529	0.1246	0.0171	0.5194
DJIA	0.0000	0.0000	0.0000	0.0009	0.0001	0.0260	0.0358	0.0381
NSDQ	0.0000	0.0000	0.0000	0.0025	0.0000	0.0010	0.0293	0.3157
CAC	0.0000	0.0000	0.0000	0.0696	0.1039	0.2384	0.0711	0.4844
BOVESPA	0.0279	0.0117	0.0000	0.3168	0.3036	0.8389	0.4567	0.7577
BA	0.0000	0.0000	0.0000	0.0477	0.0002	0.3503	0.2482	0.2336
BAC	0.0000	0.0000	0.0000	0.9360	0.0000	0.0010	0.0185	0.0745
C	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0043	0.0111
CAT	0.0000	0.0000	0.0000	0.0231	0.0003	0.0631	0.0262	0.3305
FDX	0.0000	0.0000	0.0000	0.0362	0.0000	0.1076	0.1889	0.5233
HON	0.0000	0.0000	0.0000	0.0002	0.0006	0.6167	0.6935	0.0835
HPQ	0.0000	0.0000	0.0000	0.0416	0.0000	0.1246	0.0142	0.3798
IBM	0.0000	0.0000	0.0000	0.0000	0.0000	0.8587	0.7541	0.5032
JPM	0.0000	0.0000	0.0000	0.0000	0.0000	0.0260	0.0453	0.2990
KFT	0.0000	0.0000	0.0000	0.3501	0.0000	0.0007	0.7700	0.3471
PEP	0.0000	0.0000	0.0000	0.0356	0.0001	0.0010	0.5746	0.6757
PG	0.0000	0.0000	0.0000	0.0119	0.0000	0.0000	0.3249	0.7466
T	0.0000	0.0000	0.0000	0.0140	0.0000	0.2384	0.7679	0.1772
TWX	0.0000	0.0000	0.0000	0.1001	0.0000	0.0409	0.0117	0.1517
TXN	0.0000	0.0000	0.0000	0.0031	0.0069	0.8389	0.0000	0.1273
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- $\nu_t$  (III), M-AHAR-MEM (IV), AHAR-MEM-GG (V), HAR-V-J (VI), AHAR-MEM-J (VII) and AHAR-MEM-J- $\lambda_t$  (VIII).

## 8 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 random variable for the jumps, the conditional density of  $RM_t$  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 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 illustrate the stationarity conditions and show that the maximum likelihood estimator of model's parameters is consistent and asymptotically normal. We also provide evidence of the performance in finite samples and the effects of misspecification on maximum likelihood estimation by resorting to Monte Carlo simulations. The empirical application shows that the AHAR-MEM-J- $\lambda_t$  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, 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 volatility-at-risk, i.e. the risk of extremely high volatility. The empirical analysis highlights 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, durations or energy prices.

**Acknowledgements.** We wish to thank Eric Ghysels, Christian Gouriéroux, 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

- Ait-Sahalia, Y. and Jacod, J. (2014). *High-frequency financial econometrics*. Princeton University Press.
- Andersen, T., Dobrev, D., and Schaumburg, 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.

- Bandi, F. and Renò, R. (2012). Time-varying leverage effects. *Journal of Econometrics*, 169:94–113.
- Bandi, F. and Renò, R. (2015). Price and volatility co-jumps. *Journal of Financial Economics*, *Forthcoming*.
- 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. 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.
- Billingsley, P. (1995). *Probability and Measure*. John Wiley & Sons.
- 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.
- Brandt, A. (1986). The stochastic equation  $Y_{n+1} = A_n Y_n + B_n$  with stationary coefficients. *Advances in Applied Probability*, 18(1):pp. 211–220.
- Broadie, M., Chernov, M., and Johannes, M. (2007). Model specification and risk premia: Evidence from futures options. *The Journal of Finance*, 62:1453–1490.
- 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., 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.
- Franco, C. and Zakoian, J.-M. (2004). Maximum likelihood estimation of pure garch and arma-garch processes. *Bernoulli*, 10(4):605–637.
- Franq, C. and Zakoian, J.-M. (2010). *GARCH Models*. John Wiley & Sons, Inc., Chichester, UK.
- 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. (2015). Local adaptive multiplicative error models for high-frequency forecasts. *Journal of Applied Econometrics*, 30(4):529–550.
- 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.
- Harvey, C. R. (1995). Predictable risk and returns in emerging markets. *Review of Financial Studies*, 8(3):773–816.
- Hautsch, N., Malec, P., and Schienle, M. (2013). Capturing the zero: a new class of zero-augmented distributions and multiplicative error processes. *Journal of Financial Econometrics*, 12(1):89–121.
- Huang, X. and Tauchen, G. (2005). The relative contribution of jumps to total price variance. *Journal of Financial Econometrics*, 3:456–499.
- Jacod, J., Kluppelberg, C., and Muller, G. (2013a). Functional relationship between price and volatility jumps and their consequences for discretely observed data. *Journal of Applied Probability*, 49:901–914.
- Jacod, J., Kluppelberg, C., and Muller, G. (2013b). Testing for non-correlation between price and volatility jumps. Technical report, TUM Munich.

- Jacod, J. and Todorov, V. (2010). Do price and volatility jump together? *Annals of applied probability*, 20:1425–1469.
- Jones, C. (2003). Nonlinear mean reversion in the short-term interest rate. *Review of Financial Studies*, 16:765–791.
- Lanne, M. (2006). A mixture multiplicative error model for realized volatility. *Journal of Financial Econometrics*, 4(4):594–616.
- Lee, T. and Lee, S. (2009). Normal mixture quasi-maximum likelihood estimator for GARCH models. *Scandinavian Journal of Statistics*, 36(1):157–170. cited By 12.
- 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.
- Richardson, M. and Smith, T. (1993). A Test for Multivariate Normality in Stock Returns. *The Journal of Business*, 66(2):295–321.
- Russell, J. R. (1999). Econometric modeling of multivariate irregularly-spaced high-frequency data. Technical report, Working paper.
- Simar, L. (1976). Maximum likelihood estimation of a compound poisson process. *The Annals of Statistics*, 4:1200–1209.
- Todorov, V. and Tauchen, G. (2011). Volatility jumps. *Journal of Business and Economic Statistics*, 29:356–371.
- Wooldridge, J. (1994). Estimation and inference for dependent processes. In Engle, R. and McFadden, D., editors, *Handbook of Econometrics*, volume 4, chapter 45. Elsevier Science.
- Zhang, J. E., Shu, J., and Brenner, M. (2010). The new market for volatility trading. *Journal of Futures Markets*, 30(9):809–833.

## A Proofs and results

This Appendix contains the proofs of all Propositions and Theorems, together with Lemmas summarizing further theoretical elements.

### A.1 Proof of Proposition 1

Given  $\varepsilon_t | \mathcal{F}_{t-1} \sim \Gamma(1, \nu)$  and  $Z_t | N_t = m > 0, \mathcal{F}_{t-1} \sim \Gamma(md_\lambda, m\varsigma)$ , integrating out  $Z_t$  in (9) we have the conditional density of  $\eta_t$

$$f_\eta(\eta_t | N_t = m > 0, \mathcal{F}_{t-1}; \nu, \varsigma, \lambda) = \frac{2}{\eta_t} \left( \eta_t \frac{\varsigma \nu}{d_\lambda} \right)^{\frac{m\varsigma + \nu}{2}} \frac{1}{\Gamma(m\varsigma)\Gamma(\nu)} \mathbb{K}_{m\varsigma - \nu} \left( 2\sqrt{\eta_t \frac{\varsigma \nu}{d_\lambda}} \right),$$

which is the  $K$  density, see Redding (1999).  $\mathbb{K}_a(\cdot)$  is the modified Bessel function of the second kind. The moments of  $\eta_t$ , conditional on  $N_t = m > 0$  and  $\mathcal{F}_{t-1}$ , can be derived from the moments of the  $K$  density in (52). The countable infinite mixture of densities in (11) is obtained by integrating out the dependence on  $N_t$ , i.e. multiplying the density associated to each jump realization with the corresponding probability of the Poisson distribution. ■

### A.2 Proof of Proposition 2

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

$$\begin{aligned} f_X(x_t | N_t = m > 0, \mathcal{F}_{t-1}; \zeta, \varsigma) &= f_\eta \left( \frac{x_t}{\mu_t} | N_t = m > 0, \mathcal{F}_{t-1} \right) \left| \frac{1}{\mu_t} \right| \\ &= \frac{2\mu_t}{x_t} \left( \frac{x_t \varsigma \nu}{\mu_t d_\lambda} \right)^{\frac{m\varsigma + \nu}{2}} \frac{1}{\Gamma(m\varsigma)\Gamma(\nu)} \mathbb{K}_{m\varsigma - \nu} \left( 2\sqrt{\frac{x_t \varsigma \nu}{\mu_t d_\lambda}} \right) \left| \frac{1}{\mu_t} \right| \\ &= \frac{2}{x_t} \left( \frac{X_t \varsigma \nu}{\mu_t d_\lambda} \right)^{\frac{m\varsigma + \nu}{2}} \frac{1}{\Gamma(m\varsigma)\Gamma(\nu)} \mathbb{K}_{m\varsigma - \nu} \left( 2\sqrt{\frac{x_t \varsigma \nu}{\mu_t d_\lambda}} \right) \end{aligned}$$

Similarly to the case of  $\eta_t$ , the moments of  $X_t$  conditional on  $N_t = m > 0$  and  $\mathcal{F}_{t-1}$  are derived from the moment function of the  $K$  density. ■

### A.3 Proof of Proposition 3

Similarly to the derivation of the conditional density of  $\eta_t$ , the countable infinite mixture of densities for  $X_t$  reported in (14) is obtained by integrating out the dependence on  $N_t$ , i.e. multiplying the density associated to each jump realization with the corresponding probability of the Poisson distribution. The expected value of  $X_t$  conditional on  $\mathcal{F}_{t-1}$  is obtained, noting that  $\mu_t$  is measurable w.r.t.  $\mathcal{F}_{t-1}$ , so that

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

It follows that

$$\mathbb{E}[X_t | \mathcal{F}_{t-1}] = \mu_t d_\lambda, \tag{32}$$

where the conditional expected value of  $Z_t$  is  $d_\lambda$ , see the Supplementary material. The conditional variance of  $X_t$  is obtained as

$$\begin{aligned}\mathbb{V}[X_t|\mathcal{F}_{t-1}] &= \mathbb{E}[X_t^2|\mathcal{F}_{t-1}] - \mathbb{E}[X_t|\mathcal{F}_{t-1}]^2 \\ &= \mu_t^2 \left\{ \left[ \frac{\lambda}{\varsigma} + e^{-\lambda} + (\lambda + \lambda^2) \right] (1 + \nu^{-1}) d_\lambda^2 - 1 \right\}.\end{aligned}\quad (33)$$

■

**Lemma A.1** *When  $\lambda$  is constant in time,  $\{\eta_t\}_{t \in \mathbb{Z}}$  is an i.i.d. random sequence.*

**Proof** Consider the conditional density of  $\eta_t$  in (35). Note that

$$P(N_t = m | \mathcal{F}_{t-1}) = \frac{e^{-\lambda} \lambda^m}{m!} = P(N_t = m)$$

Replacing  $P(N_t = m)$  in (35), it follows that the conditional density of  $\eta_t$  is not a function of any element of  $\mathcal{F}_{t-1}$ , hence  $f_\eta(\eta_t | \mathcal{F}_{t-1}; \theta) = f(\eta_t; \theta)$ , i.e. the marginal density of  $\eta_t$ . ■

**Lemma A.2** *The process*

$$\lambda_t = \phi_1 + \phi_2 \lambda_{t-1} + \phi_3 \xi_{t-1} \quad t = 1, 2, \dots, T$$

*with  $\phi_1 > 0, \phi_2, \phi_3 \geq 0$  and  $\xi_t$  a strictly stationary ergodic m.d.s. has a convergent (almost surely) strictly stationary and ergodic solution, if  $\phi_2 < 1$ .*

**Proof** Since the process for  $\lambda_t$  can be rewritten as

$$X_t = B_t + A_t X_{t-1}$$

where  $X_t = \lambda_t$ ,  $A_t = \phi_2$  and  $B_t = \phi_1 + \phi_3 \xi_{t-1}$ , with  $E[\log B_t] = E[\log(\phi_1 + \phi_3 \xi_{t-1})] < \infty$ . By recursive substitutions, the process of  $\lambda_t$  can be written as

$$\begin{aligned}X_t &= B_t + \sum_{n=1}^{\infty} \prod_{j=1}^n A_{t-j} B_{t-n-1} \\ \lambda_t &= (\phi_1 + \phi_3 \xi_{t-1}) + \sum_{n=1}^{\infty} \phi_2^n \prod_{j=1}^n A_{t-j} (\phi_1 + \phi_3 \xi_{t-n-1})\end{aligned}\quad (34)$$

If  $0 \leq \phi_2 < 1$  then the condition  $\log(|A_t|) = \log(\phi_2) < 0$  is satisfied, see Brandt (1986), then,  $\lambda_t$  converges almost surely and (34) is the unique strictly stationary causal solution, since  $\sum_{n=1}^{\infty} \phi_2^n \phi_3 \xi_{t-n-1}$  is geometrically bounded. Furthermore,  $\lambda_t$  is a measurable function of  $\{\xi_{n-1}\}_{n=-\infty}^{n=t}$ , hence from Franq and Zakoian (2010, Theorem A.1) the strictly stationary solution in (34) is also ergodic. ■

**Lemma A.3** *Given Assumption 1 with  $\lambda_t \geq 0$  a.s. and strictly stationary*

$$\eta_t = Z_t \varepsilon_t$$

*is a strictly stationary and ergodic process.*

**Proof** Since the conditional density of  $\eta_t$

$$f_\eta(\eta_t|\mathcal{F}_{t-1}; \nu, \varsigma, \phi_1, \phi_2, \phi_3) = P(N_t = 0|\mathcal{F}_{t-1}; \lambda_t)g_\eta(\eta_t|N_t = 0, \mathcal{F}_{t-1}; \nu, \lambda_t) \\ + \sum_{m=1}^{\infty} P(N_t = m|\mathcal{F}_{t-1}; \lambda_t)k_\eta(\eta_t|N_t = m > 0, \mathcal{F}_{t-1}; \nu, \varsigma, \lambda_t). \quad (35)$$

is a function of  $\lambda_t$ , and  $\lambda_t$  has a strictly stationary and ergodic solution in (34), it follows that  $\eta_t$ , that is a measurable function of  $\lambda_t$ , is a strictly stationary and ergodic process, see Billingsley (1995, Theorem 36.4).

#### A.4 Proof of Theorem 1

Since  $\{\eta_t\}$  is a sequence of i.i.d. random variables, see Lemma A.1, the proof is identical to that of Theorem 2.5 in Franq and Zakoian (2010).■

#### A.5 Proof of Theorem 2

The vector form of the process in (22) is

$$z_t = b_t + A_t z_{t-1} \quad (36)$$

where

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

$A_t$  is a  $(p+q) \times (p+q)$  matrix with positive coefficients but not independent, since  $\eta_t$  is correlated through  $\lambda_t$ . 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

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

Now, since  $\eta_t$  from Lemma A.3 is ergodic and strictly stationary, the sequence of matrices  $\{A_t, t \in \mathbb{Z}\}$  is ergodic and strictly stationary. With a multiplicative norm, i.e.  $\|A\| = \sum |a_{ij}|$ ,  $\|AB\| \leq \|A\|\|B\|$  for all matrices  $A$  and  $B$  such that  $AB$  exists, then  $\log \|A_t\| \leq \log \|A\| + \log \|E_t\|$ , with  $\log \|E_t\| = \log [(p+q)(\eta_t + (p+q) - 1)]$ . Given that the Lyapunov exponent  $\gamma$  is equal to, see Franq and Zakoian (2010, Theorem 2.3),

$$\gamma = \lim_{t \rightarrow \infty} a.s. \frac{1}{t} \log \|A_t A_{t-1} \dots A_1\| \quad (37)$$

and

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

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

$$\rho(A) < \exp (-E[\log \|E_t\|]). \quad (38)$$

Now, we turn to the proof of the existence of a stationary and ergodic solution if the condition in (38) is satisfied, i.e.  $\gamma < 0$ . Since the random variable  $\eta_t$  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 (37) 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

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

To show that  $n^{-1} \log \|b_{t-n-1}\| \rightarrow 0$  we have used the result that for a sequence  $X_n$  of identically distributed random variables admitting an expectation holds that  $X_n/n \xrightarrow{a.s.} 0$  when  $n \rightarrow \infty$ . 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). By the Cauchy rule,  $\tilde{z}_t$  is well defined in  $(\mathbb{R}^{*+})^{p+q}$ . Let  $\tilde{z}_{q+1,t}$  denote the  $(q+1)$ -th element of  $\tilde{z}_t$ . Setting  $X_t = \tilde{z}_{q+1,t} \eta_t$ , we define a solution of model (3). This solution is nonanticipative because  $X_t$  can be expressed as a measurable function of  $\eta_t, \eta_{t-1}, \dots$ . By the ergodicity of  $\eta_t$  this solution is also strictly stationary and ergodic.

The proof of the uniqueness parallels the arguments in the proof of Theorem 2.4 of Franq and Zakoian (2010). ■



## A.6 Proof of Theorem 3

**Proof** Let  $\theta^{(1)} = (\theta'_{1,1}, \theta'_{2,1}, \theta'_{3,1})'$  and  $\theta^{(2)} = (\theta'_{1,2}, \theta'_{2,2}, \theta'_{3,2})'$ . Suppose that

$$f_X(x_t; \theta^{(1)}) = f_X(x_t; \theta^{(2)})$$

then

$$\begin{aligned} e^{-\lambda_t(\theta_{2,1})} g_X(x_t; \theta_{1,1}, \theta_{2,1}, \nu_1) + \sum_{m=1}^{\infty} \frac{e^{-\lambda_t(\theta_{2,1})} \lambda_t(\theta_{2,1})^m}{m!} \times k_X(x_t; m, \theta^{(1)}) = \\ e^{-\lambda_t(\theta_{2,2})} g_X(x_t; \theta_{1,2}, \theta_{2,2}, \nu_2) + \sum_{m=1}^{\infty} \frac{e^{-\lambda_t(\theta_{2,2})} \lambda_t(\theta_{2,2})^m}{m!} \times k_X(x_t; m, \theta^{(2)}) \end{aligned}$$

this implies that

$$\begin{aligned} g_X(x_t; \theta_{1,1}, \theta_{2,1}, \nu_1) &= e^{\lambda_t(\theta_{2,1})} e^{-\lambda_t(\theta_{2,2})} g_X(x_t; \theta_{1,2}, \theta_{2,2}, \nu_2) \\ &+ e^{\lambda_t(\theta_{2,1})} e^{-\lambda_t(\theta_{2,2})} \sum_{m=1}^{\infty} \frac{\lambda_t(\theta_{2,2})^m}{m!} \times k_X(x_t; m, \theta^{(1)}) \\ &- \sum_{m=1}^{\infty} \frac{\lambda_t(\theta_{2,1})^m}{m!} \times k_X(x_t; m, \theta^{(2)}) \end{aligned} \quad (39)$$

where the equality is satisfied only if  $\theta^{(1)} = \theta^{(2)}$ . To see this, recall that the RHS of (39) is a mixture density if the weights are positive and sum to 1. The condition is satisfied only if  $\lambda_t(\theta_{2,1}) = \lambda_t(\theta_{2,2})$ . Further, both sides of (39) must integrate to one, which in turn implies that

$$k_X(x_t; m, \theta^{(1)}) = k_X(x_t; m, \theta^{(2)}) \quad \forall m. \quad (40)$$

so that  $\theta_{1,1} = \theta_{1,2}$  and  $\theta_{3,1} = \theta_{3,2}$ . We have shown that  $f_X(x_t; \theta^{(1)}) = f_X(x_t; \theta^{(2)})$  implies  $\theta^{(1)} = \theta^{(2)}$ . ■

## A.7 Proof of Theorem 4

**Proof** The proof consists of the verification of Assumption 2. Given Theorem 3 which shows that Assumption 2.ii is satisfied, Theorem 2 which for Assumption 2.iii and Lemma's, A.5, A.7 below which prove that Assumptions 2.iv-v are satisfied, the theorem follows. ■

**Lemma A.4** *Under Assumptions 2.i and 2.iii there exist two random variables  $V_\mu$  and  $V_\lambda$  and two constants  $\rho_\mu, \rho_\lambda \in (0, 1)$  such that for all  $t$*

$$\sup_{\theta_1 \in \Theta_1} |\mu_t(\theta_1) - \tilde{\mu}_t(\theta_1)| \leq V_\mu \rho_\mu^t \quad (41)$$

$$\sup_{\theta_2 \in \Theta_2} |\lambda_t(\theta_2) - \tilde{\lambda}_t(\theta_2)| \leq V_\lambda \rho_\lambda^t \quad (42)$$

where  $\tilde{\mu}_t(\theta_1)$  replaces unavailable observations of the  $\mu_t$  expansion with initial values and a similar replacement applies to  $\tilde{\lambda}_t(\theta_2)$ .

**Proof** The proof is a consequence of Proposition 1 in Francq and Zakoian (2004) and builds on strict stationarity of both  $\mu_t$  and  $\lambda_t$ . ■

Assumption 2.v is then verified by the following lemma.

**Lemma A.5** *Under Assumptions 2.i and 2.iii*

$$\lim_{T \rightarrow \infty} \sup_{\theta \in \Theta} |L_T(\theta)/T - \tilde{L}_t(\theta)/T| = 0 \quad a.s. \quad (43)$$

**Proof** We adapt to the presence of a mixture of Gamma and  $K$  densities the proof of Lee and Lee (2009). To this end we must prove the following convergence

$$\sup_{\theta \in \Theta} \left| \frac{f_X(x; \mu_t, \lambda_t; \theta) - f_X(x; \tilde{\mu}_t, \tilde{\lambda}_t; \theta)}{f_X(x; \tilde{\mu}_t, \tilde{\lambda}_t; \theta)} \right| \rightarrow 0 \quad a.s. \quad (44)$$

We can separately look at each summand of the infinite mixture in (13). Take the first component of the mixture

$$P(N_t = 0 | \mathcal{F}_{t-1}) g_X(x_t | N_t = 0, \mathcal{F}_{t-1}; \theta_1, \theta_2, \nu) = e^{-\lambda_t} \frac{1}{\Gamma(\nu)} \frac{1}{x_t} \left( \frac{\nu}{\mu_t} (e^{-\lambda_t} + \lambda_t) x_t \right)^\nu e^{-\frac{\nu}{\mu_t} (e^{-\lambda_t} + \lambda_t) x_t} \quad (45)$$

First, we have to show that

$$\left| \exp(\tilde{\lambda}_t - \lambda_t) \exp\left( \nu x_t \frac{\mu_t (e^{-\tilde{\lambda}_t} + \tilde{\lambda}_t) - \tilde{\mu}_t (e^{-\lambda_t} + \lambda_t)}{\tilde{\mu}_t \mu_t} \right) \left( \frac{\tilde{\mu}_t (e^{-\lambda_t} + \lambda_t)}{\mu_t (e^{-\tilde{\lambda}_t} + \tilde{\lambda}_t)} \right)^\nu - 1 \right| \quad (46)$$

is bounded over  $\Theta$ . For simplicity we rewrite it as  $|A - 1|$ , and further decompose  $A$  into three terms,  $A_1, A_2, A_3$  representing the two exponentials and the power. We first note that  $A$  is, by construction, positive. In fact, the jump intensity and mean dynamic are constrained to assume only positive values. Consequently,

$$\sup_{\theta \in \Theta} |A - 1| \leq \sup_{\theta \in \Theta} (A - 1)^+, \quad (47)$$

where  $(x)^+ = \max\{x, 0\}$ . We also have

$$\sup_{\theta \in \Theta} (A - 1)^+ = \left( \sup_{\theta \in \Theta} A - 1 \right)^+ = \left( \sup_{\theta \in \Theta} A_1 \sup_{\theta \in \Theta} A_2 \sup_{\theta \in \Theta} A_3 - 1 \right)^+ \quad (48)$$

and now we separately consider the terms of  $A$ . For the first term,

$$\sup_{\theta \in \Theta} A_1 = \sup_{\theta \in \Theta} \exp(\tilde{\lambda}_t - \lambda_t) = \exp\left( \sup_{\theta \in \Theta} (\tilde{\lambda}_t - \lambda_t) \right) \leq \exp(V_\lambda \rho_\lambda^t)$$

where we applied Lemma A.4 and the fact that  $\sup_{\theta \in \Theta} (\tilde{\lambda}_t - \lambda_t) \leq \sup_{\theta \in \Theta} |\tilde{\lambda}_t - \lambda_t|$ . For the

second term we have

$$\begin{aligned}
\sup_{\theta \in \Theta} A_2 &= \exp \left\{ \nu x_t \frac{\mu_t (e^{-\tilde{\lambda}_t} + \tilde{\lambda}_t) - \tilde{\mu}_t (e^{-\lambda_t} + \lambda_t)}{\tilde{\mu}_t \mu_t} \right\} \\
&= \exp \left\{ C_1 \sup_{\theta \in \Theta} \left( \mu_t (e^{-\tilde{\lambda}_t} + \tilde{\lambda}_t) - \tilde{\mu}_t (e^{-\lambda_t} + \lambda_t) \right) \right\} \\
&= \exp \left\{ C_1 \sup_{\theta \in \Theta} \left( (\mu_t - \tilde{\mu}_t) (e^{-\tilde{\lambda}_t} + \tilde{\lambda}_t) - \tilde{\mu}_t (e^{-\lambda_t} + \lambda_t - e^{-\tilde{\lambda}_t} - \tilde{\lambda}_t) \right) \right\} \\
&\leq \exp \left\{ C_{1,a} \sup_{\theta \in \Theta} (\mu_t - \tilde{\mu}_t) + C_{1,b} \sup_{\theta \in \Theta} \left( \lambda_t - \tilde{\lambda}_t + e^{-\lambda_t} (1 - e^{\lambda_t - \tilde{\lambda}_t}) \right) \right\} \\
&\leq \exp \left\{ C_{1,a} \sup_{\theta \in \Theta} (\mu_t - \tilde{\mu}_t) + C_{1,b} \sup_{\theta \in \Theta} (\lambda_t - \tilde{\lambda}_t) + C_{1,b} \sup_{\theta \in \Theta} e^{-\lambda_t} (1 - e^{\lambda_t - \tilde{\lambda}_t}) \right\} \\
&\leq \exp \left\{ C_{1,a} \sup_{\theta \in \Theta} (\mu_t - \tilde{\mu}_t) + C_{1,b} \sup_{\theta \in \Theta} (\lambda_t - \tilde{\lambda}_t) + C_{1,c} (1 - e^{\sup_{\theta \in \Theta} \{\lambda_t - \tilde{\lambda}_t\}}) \right\} \\
&\leq \exp \left\{ C_{1,a} V_\mu \rho_\mu^t + C_{1,b} V_\lambda \rho_\lambda^t + C_{1,c} (1 - \exp(V_\lambda \rho_\lambda^t)) \right\},
\end{aligned}$$

where we have used Lemma A.4,  $\sup_{\theta \in \Theta} (\lambda_t - \tilde{\lambda}_t) \leq \sup_{\theta \in \Theta} |\lambda_t - \tilde{\lambda}_t|$ , and  $\sup_{\theta \in \Theta} (\mu_t - \tilde{\mu}_t) \leq \sup_{\theta \in \Theta} |\mu_t - \tilde{\mu}_t|$ . Finally, for the third term

$$\begin{aligned}
\sup_{\theta \in \Theta} A_3 &= \sup_{\theta \in \Theta} \left( \frac{\tilde{\mu}_t (e^{-\lambda_t} + \lambda_t)}{\mu_t (e^{-\tilde{\lambda}_t} + \tilde{\lambda}_t)} \right)^\nu = \sup_{\theta \in \Theta} \left( \exp \left\{ \log \left( \frac{\tilde{\mu}_t e^{-\lambda_t} + \lambda_t}{\mu_t e^{-\tilde{\lambda}_t} + \tilde{\lambda}_t} \right)^\nu \right\} \right) \\
&= \exp \left( \sup_{\theta \in \Theta} \nu \log \left( \frac{\tilde{\mu}_t (e^{-\lambda_t} + \lambda_t)}{\mu_t (e^{-\tilde{\lambda}_t} + \tilde{\lambda}_t)} \right) \right) \\
&\leq \exp \left( C_2 \log \left( \sup_{\theta \in \Theta} \frac{\tilde{\mu}_t}{\mu_t} \sup_{\theta \in \Theta} \left( \frac{e^{-\lambda_t} + \lambda_t}{e^{-\tilde{\lambda}_t} + \tilde{\lambda}_t} \right) \right) \right) \\
&\leq \exp \left( C_2 \log \left( \sup_{\theta \in \Theta} \left( \frac{\tilde{\mu}_t - \mu_t}{\mu_t} + 1 \right) \sup_{\theta \in \Theta} \left( \frac{e^{-\lambda_t} + \lambda_t - e^{-\tilde{\lambda}_t} - \tilde{\lambda}_t}{e^{-\tilde{\lambda}_t} + \tilde{\lambda}_t} + 1 \right) \right) \right) \\
&\leq \exp \left( C_{2,a} \log \left( \left( 1 + C_{2,b} \sup_{\theta \in \Theta} (\tilde{\mu}_t - \mu_t) \right) \left( 1 + C_{2,c} \sup_{\theta \in \Theta} (e^{\lambda_t} + \lambda_t - e^{\tilde{\lambda}_t} - \tilde{\lambda}_t) \right) \right) \right) \\
&\leq \exp (C_{2,a} \log ((1 + C_{2,b} V_\mu \rho_\mu^t) (1 + C_{2,c} V_\lambda \rho_\lambda^t + C_{2,d} (1 - \exp(V_\lambda \rho_\lambda^t))))))
\end{aligned}$$

Therefore, combining the previous results we can verify that (46) is bounded as follows

$$\begin{aligned}
\sup_{\theta \in \Theta} |A - 1| &\leq (\exp (V_\lambda \rho_\lambda^t + C_{1,a} V_\mu \rho_\mu^t + C_{1,b} V_\lambda \rho_\lambda^t + C_{1,c} (1 - \exp (V_\lambda \rho_\lambda^t))) + \\
&\quad C_{2,a} \log ((1 + C_{2,b} V_\mu \rho_\mu^t) (1 + C_{2,c} V_\lambda \rho_\lambda^t + C_{2,d} (1 - \exp (V_\lambda \rho_\lambda^t)))) - 1)^+.
\end{aligned}$$

By Proposition 1 of Francq and Zakoïan (2004), the bound converges almost surely to zero. We move to the components of the mixture involving the  $K$  density. The proof corresponds to that

of the previous case but with  $A$  taking the following value

$$A = \exp(\tilde{\lambda}_t - \lambda_t) \left( \frac{\lambda_t}{\tilde{\lambda}_t} \right)^m \left( \frac{\mu_t e^{-\tilde{\lambda}_t} + \tilde{\lambda}_t}{\tilde{\mu}_t e^{-\lambda_t} + \lambda_t} \right)^{\frac{m\zeta - \nu}{2}} \frac{\mathbb{K}_{m\zeta - \nu} \left( 2\sqrt{\frac{X_t}{\mu_t}} \zeta \nu (e^{-\lambda_t} + \lambda_t) \right)}{\mathbb{K}_{m\zeta - \nu} \left( 2\sqrt{\frac{X_t}{\tilde{\mu}_t}} \zeta \nu (e^{-\tilde{\lambda}_t} + \tilde{\lambda}_t) \right)}$$

Therefore we have

$$\sup_{\theta \in \Theta} A = \sup_{\theta \in \Theta} \exp \left\{ \tilde{\lambda}_t - \lambda_t + m \log \left( \frac{\lambda_t}{\tilde{\lambda}_t} \right) + \frac{m\zeta - \nu}{2} \log \left( \frac{\mu_t e^{-\tilde{\lambda}_t} + \tilde{\lambda}_t}{\tilde{\mu}_t e^{-\lambda_t} + \lambda_t} \right) + \log \frac{\mathbb{K}_{m\zeta - \nu} \left( 2\sqrt{\frac{X_t}{\mu_t}} \zeta \nu (e^{-\lambda_t} + \lambda_t) \right)}{\mathbb{K}_{m\zeta - \nu} \left( 2\sqrt{\frac{X_t}{\tilde{\mu}_t}} \zeta \nu (e^{-\tilde{\lambda}_t} + \tilde{\lambda}_t) \right)} \right\}.$$

The last term involving the ratio of the Bessel functions simplifies only partially using the integral representation of the Bessel function of the second kind. We rewrite the ratio as follows

$$\frac{\mathbb{K}_a(b)}{\mathbb{K}_a(\tilde{b})} = \frac{-\frac{2(0.5b)^{-a}}{\sqrt{\pi}\Gamma(0.5-a)} \int_1^\infty \frac{\cos(bt)}{(t^2-1)^{a+0.5}} dt}{-\frac{2(0.5\tilde{b})^{-a}}{\sqrt{\pi}\Gamma(0.5-a)} \int_1^\infty \frac{\cos(\tilde{b}t)}{(t^2-1)^{a+0.5}} dt}.$$

Then, we have

$$\frac{-\frac{2(0.5b)^{-a}}{\sqrt{\pi}\Gamma(0.5-a)}}{-\frac{2(0.5\tilde{b})^{-a}}{\sqrt{\pi}\Gamma(0.5-a)}} = \left( \frac{b}{\tilde{b}} \right)^{-a} = \left( \frac{\tilde{\mu}_t e^{-\lambda_t} + \lambda_t}{\mu_t e^{-\tilde{\lambda}_t} + \tilde{\lambda}_t} \right)^{-\frac{m\zeta - \nu}{2}} = \left( \frac{\mu_t e^{-\lambda_t} + \lambda_t}{\tilde{\mu}_t e^{-\tilde{\lambda}_t} + \tilde{\lambda}_t} \right)^{\frac{m\zeta - \nu}{2}}$$

while for the ratio of integrals we rename it as  $\mathbb{B}$  and numerically verify it is bounded. Therefore, again using Lemma A.4 and Proposition 1 of Francq and Zakoïan (2004), we can verify the following upper bound converges to 0 almost surely.

$$\sup_{\theta \in \Theta} |A - 1| \leq (\exp(V_\lambda \rho_\lambda^t + C_{1,a} \log(1 + c_{1,b} V_\lambda \rho_\lambda^t)) + C_{2,a} \log((1 + C_{2,b} V_\mu \rho_\mu^t)(1 + C_{2,c} V_\lambda \rho_\lambda^t + C_{2,d}(1 - \exp(V_\lambda \rho_\lambda^t)))) + \log \mathbb{B} - 1)^+.$$

■

**Lemma A.6** *Given the MEM-J model, the mixture density satisfies*

$$E[f_X(x_t | \mathcal{F}_{t-1}; \theta)] < \infty \quad \forall \theta \in \text{int}(\Theta)$$

**Proof** Given

$$f_X(x_t | \mathcal{F}_{t-1}; \theta) = e^{-\lambda_t} g_X(x_t; \theta_1, \theta_2, \nu) + \sum_{m=1}^{\infty} \frac{e^{-\lambda_t} \lambda_t^m}{m!} \times k_X(x_t; m, \theta),$$

$E[e^{-\lambda_t} g_X(x_t; \theta_1, \theta_2, \nu)]$  is finite since  $e^{-\lambda_t}$  is finite for all  $\lambda_t > 0$  a.s. and the  $g_X(\cdot)$  is finite for all

$(\theta'_1, \theta'_2, \nu) \in \text{int}(\Theta)$ . Let

$$A(x_t; m, \theta) = \frac{e^{-\lambda_t} \lambda_t^m}{m!} \times k_X(x_t; m, \theta)$$

$$E[A(x_t; m, \theta) | \lambda_t = \lambda] = \frac{e^{-\lambda} \lambda^m}{m!} E[k_X(x_t; m, \theta)]$$

where  $E[k_X(x_t; m, \theta)]$  is bounded in  $m$ , such that

$$E \left[ \sum_{m=1}^{\infty} \frac{e^{-\lambda_t} \lambda_t^m}{m!} \times k_X(x_t; m, \theta) \right] = O(1)$$

The Lemma follows. ■

**Lemma A.7** *Under Assumptions 2.i-ii*

$$\mathbb{E}_{\theta_0}[l_t(\theta_0)] < \infty \quad \text{and if } \theta \neq \theta_0, \quad \text{then } \mathbb{E}_{\theta_0}[l_t(\theta)] < \mathbb{E}_{\theta_0}[l_t(\theta_0)].$$

**Proof** To show that  $\mathbb{E}_{\theta_0}[l_t(\theta_0)] < \infty$ , it is sufficient to note that

$$\mathbb{E}_{\theta_0}[\log(f_X(x_t; \theta_0))] \leq \log \mathbb{E}_{\theta_0}[f_X(x_t; \theta_0)]$$

and Lemma A.6. Note that

$$\mathbb{E}_{\theta_0}[l_t(\theta_0)] - \mathbb{E}_{\theta_0}[l_t(\theta)] = \mathbb{E}_{\theta_0} \left[ \log \frac{f_X(x_t; \theta_0)}{f_X(x_t; \theta)} \right]$$

$$\geq 0$$

since for Theorem 3  $\Pr[(f_X(x_t; \theta_0)/f_X(x_t; \theta)) \neq 1] > 0$ . Under global identification the inequality is strict. ■

**Proof A.8 Proof of Theorem 5**

The proof consists of the verification of Assumption 3. In the following we discuss the assumptions for Theorem 5.

**Assumptions i-ii** The derivatives of the density of the mixture with respect to  $\theta$ , are given by

$$\frac{\partial f_X(x_t; \theta)}{\partial \theta_1} = \frac{\partial \mu_t(\theta_1)}{\partial \theta_1} \left[ e^{-\lambda_t} \times \frac{\partial g_X(x_t; \theta_1, \theta_2, \nu)}{\partial \mu_t} + \sum_{m=1}^{\infty} \frac{e^{-\lambda_t} \lambda_t^m}{m!} \times \frac{\partial k_X(x_t|m, \theta)}{\partial \mu_t} \right] \quad (49)$$

$$\frac{\partial f_X(x_t; \theta)}{\partial \theta_2} = \frac{\partial e^{-\lambda_t} g_X(x_t; \theta_1, \theta_2, \nu)}{\partial \theta_2} + \sum_{m=1}^{\infty} \frac{\partial \frac{e^{-\lambda_t} \lambda_t^m}{m!} k_X(x_t|m, \theta)}{\partial \theta_2} \quad (50)$$

$$\frac{\partial f_X(x_t; \theta)}{\partial \theta_3} = e^{-\lambda_t} \frac{\partial g_X(x_t; \theta_1, \theta_2, \nu)}{\partial \theta_3} + \sum_{m=1}^{\infty} \frac{e^{-\lambda_t} \lambda_t^m}{m!} \frac{\partial k_X(x_t|m, \theta)}{\partial \theta_3} \quad (51)$$

In particular, the derivatives with respect to  $\omega$ ,  $\alpha$ ,  $\beta$  are

$$\begin{aligned}\frac{\partial f_X(x_t; \theta)}{\partial \omega} &= e^{-\lambda_t} \left[ \frac{\partial g_X(x_t; \theta_1, \theta_2, \nu)}{\partial \mu_t} + \sum_{m=1}^{\infty} \frac{\lambda_t^m}{m!} \cdot \frac{\partial k_X(x_t|m, \theta)}{\partial \mu_t} \right] \\ \frac{\partial f_X(x_t; \theta)}{\partial \alpha} &= e^{-\lambda_t} x_{t-1} \left[ \frac{\partial g_X(x_t; \theta_1, \theta_2, \nu)}{\partial \mu_t} + \sum_{m=1}^{\infty} \frac{\lambda_t^m}{m!} \cdot \frac{\partial k_X(x_t|m, \theta)}{\partial \mu_t} \right] \\ \frac{\partial f_X(x_t; \theta)}{\partial \beta} &= e^{-\lambda_t} \mu_{t-1} \left[ \frac{\partial g_X(x_t; \theta_1, \theta_2, \nu)}{\partial \mu_t} + \sum_{m=1}^{\infty} \frac{\lambda_t^m}{m!} \cdot \frac{\partial k_X(x_t|m, \theta)}{\partial \mu_t} \right]\end{aligned}$$

The terms  $\frac{\partial g_X(x_t; \theta_1, \theta_2, \nu)}{\partial \mu_t}$  and  $\frac{\partial k_X(x_t|m, \theta)}{\partial \mu_t}$  are finite ( $\mu_t$  appears in the argument of the Bessel function, which is bounded when its argument is larger than 0). The derivatives with respect to  $\phi_1$ ,  $\phi_2$  and  $\phi_3$  are

$$\begin{aligned}\frac{\partial f_X(x_t; \theta)}{\partial \phi_1} &= \frac{\partial f_X(x_t; \theta)}{\partial \lambda_t} \\ \frac{\partial f_X(x_t; \theta)}{\partial \phi_2} &= \lambda_{t-1} \cdot \frac{\partial f_X(x_t; \theta)}{\partial \lambda_t} \\ \frac{\partial f_X(x_t; \theta)}{\partial \phi_3} &= \xi_{t-1} \cdot \frac{\partial f_X(x_t; \theta)}{\partial \lambda_t}\end{aligned}$$

where

$$\begin{aligned}\frac{\partial f_X(x_t; \theta)}{\partial \lambda_t} &= e^{-\lambda_t} \left( \frac{\partial g_X(x_t; \theta_1, \theta_2, \nu)}{\partial \lambda_t} - g_X(x_t; \theta_1, \theta_2, \nu) \right) \\ &\quad + \sum_{m=1}^{\infty} \left[ \frac{\partial g_m(\lambda_t)}{\partial \lambda_t} k_X(x_t|m, \theta) + g_m(\lambda_t) \frac{\partial k_X(x_t|m, \theta)}{\lambda_t} \right]\end{aligned}$$

with  $g_m(\lambda_t) = \frac{e^{-\lambda} \lambda_t^m}{m!}$ . The terms  $\frac{\partial g_X(x_t; \theta_1, \theta_2, \nu)}{\partial \lambda_t}$  and  $\frac{\partial k_X(x_t|m, \theta)}{\partial \lambda_t}$  are finite ( $\lambda_t$  appears in the argument of the Bessel function, which is bounded when its argument is larger than 0).

Finally, the first derivatives with respect to  $\nu$  and  $\varsigma$  are

$$\begin{aligned}\frac{\partial f_X(x_t; \theta)}{\partial \nu} &= e^{-\lambda_t} \frac{\partial g_X(x_t; \theta_1, \theta_2, \nu)}{\partial \nu} + \sum_{m=1}^{\infty} \frac{e^{-\lambda_t} \lambda_t^m}{m!} \frac{\partial k_X(x_t|m, \theta)}{\partial \nu} \\ \frac{\partial f_X(x_t; \theta)}{\partial \varsigma} &= \sum_{m=1}^{\infty} \frac{\lambda_t^m}{m!} \cdot \frac{\partial k_X(x_t|m, \theta)}{\partial \varsigma}\end{aligned}$$

where  $\frac{\partial g_X(x_t; \theta_1, \theta_2, \nu)}{\partial \nu}$  is surely bounded for  $\nu$  larger than 0, while  $\frac{\partial k_X(x_t|m, \theta)}{\partial \nu}$  and  $\frac{\partial k_X(x_t|m, \theta)}{\partial \varsigma}$  are bounded given the following assumption

**Assumption 4** *The modified Bessel function of the second type written as*

$$\mathbb{K}_{m\varsigma-\nu} \left( 2\sqrt{\frac{\varsigma\nu x_t}{\mu_t(\theta_1)d_\lambda(\theta_2)}} \right) = \mathbb{K}_{g(\theta_3, m)} \left( 2\sqrt{h(\theta_2, \theta_3, m)} \right)$$

where

$$g(\theta_3, m) = m\varsigma - \nu \quad h(\theta_3, m) = m!\varsigma\nu\eta_t$$

is continuously differentiable with respect to  $\theta_3$  and the derivatives are assumed to be finite

up to the third order.

Therefore, Assumption 3.i is satisfied since

$$\begin{aligned}\frac{\partial l_t(\theta)}{\partial \theta_i} &= \frac{\partial f_X(x_t; \theta)}{\partial \theta_i} \cdot \frac{1}{f_X(x_t; \theta)} \\ \frac{\partial^2 l_t(\theta)}{\partial \theta_i \partial \theta_j} &= \frac{1}{f_X(x_t; \theta)} \cdot \frac{\partial^2 f_X(x_t; \theta)}{\partial \theta_i \partial \theta_j} - \frac{\partial f_X(x_t; \theta)}{\partial \theta_i} \cdot \frac{\partial f_X(x_t; \theta)}{\partial \theta_j} \cdot \frac{1}{f_X(x_t; \theta)^2},\end{aligned}$$

only involve quantities that are continuous for all  $\theta \in \Theta$ .

Similarly to the previous point, it remains to prove that  $\frac{\partial f_X(x_t; \theta)}{\partial \theta_i}$ ,  $\frac{\partial^2 f_X(x_t; \theta)}{\partial \theta_i \partial \theta_j}$  and  $\frac{\partial^3 f_X(x_t; \theta)}{\partial \theta_i \partial \theta_j \partial \theta_k}$  are finite for all  $\theta \in \Theta$ . Note that the first, second and third order derivatives of  $g_X(x_t; \theta_1, \theta_2, \nu)$ ,  $k_X(x_t | m, \theta)$  and  $g_m(\lambda_t)$  with respect to  $\mu_t$  and  $\lambda_t$  are bounded under stationarity of  $\mu_t$  and  $\lambda_t$ . Moreover, the derivatives of  $k_X(x_t | m, \theta)$  with respect to  $\nu$  and  $\varsigma$  are bounded under assumption 4. Therefore  $\frac{\partial f_X(x_t; \theta)}{\partial \theta_i}$ ,  $\frac{\partial^2 f_X(x_t; \theta)}{\partial \theta_i \partial \theta_j}$  and  $\frac{\partial^3 f_X(x_t; \theta)}{\partial \theta_i \partial \theta_j \partial \theta_k}$  are finite for all  $\theta \in \Theta$ .

**Assumption iii.** We know from Theorem 3 that  $l_t(\theta)$  is identified and from Lemma A.7 that has only one maximum in  $\theta_0 \in \Theta$ , so  $\mathbb{E} \left[ \frac{\partial^2 l_t(\theta_0)}{\partial \theta_i \partial \theta_j} \right]$  is negative definite in  $\theta_0$ , i.e.  $\mathcal{H}(\theta_0)$  must be positive definite.

**Assumption iv.** Given that  $x_t$  is strictly stationary and ergodic, it follows that  $\frac{1}{T} \sum_{t=1}^T x_t \xrightarrow{a.s.} \mathbb{E}[x_t]$  and  $\mathbb{V} \left[ \frac{1}{T} \sum_{t=1}^T x_t \right] < \infty$ . Since the functions  $f_X(x_t; \theta)$ ,  $\frac{\partial f_X(x_t; \theta)}{\partial \theta}$  and  $\frac{\partial^2 f_X(x_t; \theta)}{\partial \theta^2}$  are continuous and their expected value is finite, then the results for  $x_t$  can be extended to  $f_X(x_t; \theta)$ ,  $\frac{\partial f_X(x_t; \theta)}{\partial \theta}$  and  $\frac{\partial^2 f_X(x_t; \theta)}{\partial \theta^2}$ . Hence,  $\frac{\partial^2 l_t(x_t; \theta)}{\partial \theta_i \partial \theta_j}$  also follows the UWLLN, so that Assumption iv is satisfied.

**Assumption v.** To prove Assumption v, it is sufficient to show that

- A  $T^{-1/2} \frac{\partial f_X(x_t; \theta_0)}{\partial \theta}$  is strictly stationary;
- B  $\frac{1}{T} \sum_{t=1}^T \left( \frac{\partial f_X(x_t; \theta_0)}{\partial \theta} \cdot \frac{\partial f_X(x_t; \theta_0)}{\partial \theta'} \right) \xrightarrow{a.s.} \mathbb{E} \left[ \frac{\partial f_X(x_t; \theta_0)}{\partial \theta} \cdot \frac{\partial f_X(x_t; \theta_0)}{\partial \theta'} \right]$ .

For what concerns the strict stationarity of the score, note that  $\frac{\partial f_X(x_t; \theta_0)}{\partial \theta}$  in (49)-(51) involve continuous functions of  $\lambda_t$ ,  $\mu_t$ ,  $x_t$  and  $\eta_t$ , that are all strictly stationary processes by Lemma A.6. Therefore, given that a continuous transformation of a strictly stationary process is still strictly stationary, the Assumption A follows. Moreover, the Assumption B follows from the fact that the UWLLN of  $\frac{\partial f_X(x_t; \theta_0)}{\partial \theta} \cdot \frac{\partial f_X(x_t; \theta_0)}{\partial \theta'}$  is implied by the UWLLN of  $\frac{\partial f_X(x_t; \theta_0)}{\partial \theta}$  which follows from Assumption v.

**Assumption vi.** Assumption 3.vi corresponds to the assumption that the MEM-J- $\lambda_t$  captures all of the dynamics in  $x_t$ , i.e. is dynamically complete (see Wooldridge, 1994) which follows under the assumption of correct model specification.

## B The $K$ distribution

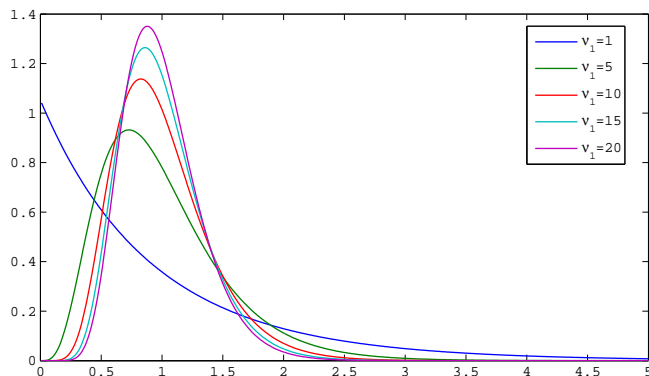
If  $X \sim \Gamma(\mu, \nu_1)$ , in mean-shape form, and  $Y|X \sim \Gamma(X, \nu_2)$  then we have  $Y \sim K(\mu, \nu_1, \nu_2)$  such that

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

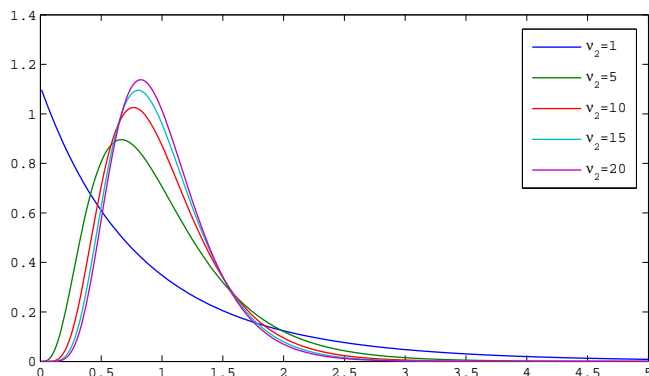
where  $\mathbb{K}_a(\cdot)$  is the modified Bessel function of the second kind. Moments of the  $K$  density are given by

$$\mathbb{E}[y^s] = \frac{\mu^s \Gamma(\nu_1 + s) \Gamma(\nu_2 + s)}{\nu_1^s \nu_2^s \Gamma(\nu_1) \Gamma(\nu_2)} \quad (52)$$

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  $\nu_1$  and  $\nu_2$ , are plotted in Figure 5. 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 5:  $K$  density computed for different values of  $\nu_1$  (upper panel) and  $\nu_2$  (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

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



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

$$F(y; \mu, \theta_1, \theta_2) = \frac{2^{2-\theta_1-\theta_2}}{\Gamma(\theta_1)\Gamma(\theta_2)} \int_0^{2\sqrt{\theta_1\theta_2 y/\mu}} t^{\theta_1+\theta_2-1} \mathbb{K}_{\theta_1-\theta_2}(t) dt, y \geq 0. \quad (53)$$

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  $z = 2\sqrt{\theta_1\theta_2 y/\mu}$

$$F(y; \mu, \theta_1, \theta_2) = 1 + \frac{2^{2-\theta_1-\theta_2}}{\Gamma(\theta_1)\Gamma(\theta_2)} g(z, \zeta, k) \quad (54)$$

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  $\theta_2$ . The best parametrization, in terms of computational speed, is  $\theta_1 = \max\{\nu_1, \nu_2\}$  and  $\theta_2 = \min\{\nu_1, \nu_2\}$ .

# Research Papers 2013



- 2014-13: Hossein Asgharian, Charlotte Christiansen and Ai Jun Hou: Macro-Finance Determinants of the Long-Run Stock-Bond Correlation: The DCC-MIDAS Specification
- 2014-14: Mikko S. Pakkanen and Anthony Réveillac: Functional limit theorems for generalized variations of the fractional Brownian sheet
- 2014-15: Federico Carlini and Katarzyna Łasak: On an Estimation Method for an Alternative Fractionally Cointegrated Model
- 2014-16: Mogens Bladt, Samuel Finch and Michael Sørensen: Simulation of multivariate diffusion bridges
- 2014-17: Markku Lanne and Henri Nyberg: Generalized Forecast Error Variance Decomposition for Linear and Nonlinear Multivariate Models
- 2014-18: Dragan Tevdovski: Extreme negative coexceedances in South Eastern European stock markets
- 2014-19: Niels Haldrup and Robinson Kruse: Discriminating between fractional integration and spurious long memory
- 2014-20: Martyna Marczak and Tommaso Proietti: Outlier Detection in Structural Time Series Models: the Indicator Saturation Approach
- 2014-21: Mikkel Bennedsen, Asger Lunde and Mikko S. Pakkanen: Discretization of Lévy semistationary processes with application to estimation
- 2014-22: Giuseppe Cavaliere, Morten Ørregaard Nielsen and A.M. Robert Taylor: Bootstrap Score Tests for Fractional Integration in Heteroskedastic ARFIMA Models, with an Application to Price Dynamics in Commodity Spot and Futures Markets
- 2014-23: Maggie E. C. Jones, Morten Ørregaard Nielsen and Michael Ksawery Popiel: A fractionally cointegrated VAR analysis of economic voting and political support
- 2014-24: Sepideh Dolatabadim, Morten Ørregaard Nielsen and Ke Xu: A fractionally cointegrated VAR analysis of price discovery in commodity futures markets
- 2014-25: Matias D. Cattaneo and Michael Jansson: Bootstrapping Kernel-Based Semiparametric Estimators
- 2014-26: Markku Lanne, Jani Luoto and Henri Nyberg: Is the Quantity Theory of Money Useful in Forecasting U.S. Inflation?
- 2014-27: Massimiliano Caporin, Eduardo Rossi and Paolo Santucci de Magistris: Volatility jumps and their economic determinants
- 2014-28: Tom Engsted: Fama on bubbles
- 2014-29: Massimiliano Caporin, Eduardo Rossi and Paolo Santucci de Magistris: Chasing volatility - A persistent multiplicative error model with jumps