Dipartimento di Statistica
"Giuseppe Parenti"

WORKING PAPER 2011/03

Multiplicative Error Models

C. T. Brownlees, F. Cipollini,
G. M. Gallo

Università degli Studi
di Firenze

Dipartimento di Statistica "G. Parenti" – Viale Morgagni 59 – 50134 Firenze - www.ds.unifi.it

Multiplicative Error Models

Prepared for **Volatility Models and Their Applications**
(*Luc Bauwens, Christian Hafner, Sébastien Laurent* eds.)

Christian T. Brownlees* Fabrizio Cipollini† Giampiero M. Gallo‡

2011.03.31

Abstract

Financial time series analysis has focused on data related to market trading activity. Next to the modeling of the conditional variance of returns within the GARCH family of models, recent attention has been devoted to other variables: first, and foremost, volatility measured on the basis of ultra–high frequency data, but also volumes, number of trades, durations. In this paper, we examine a class of models, named Multiplicative Error Models, which are particularly suited to model such non-negative time series. We discuss the univariate specification, by considering the base choices for the conditional expectation and the error term. We provide also a general framework, allowing for richer specifications of the conditional mean. The outcome is a novel MEM (called *Composite MEM*) which is reminiscent of the short– and long–run component GARCH model by Engle and Lee (1999). Inference issues are discussed relative to Maximum Likelihood and Generalized Method of Moments estimation. In the application, we show the regularity in parameter estimates and forecasting performance obtainable by applying the MEM to the realized kernel volatility of components of the S&P100 index. We suggest extensions of the base model by enlarging the information set and adopting a multivariate specification.

*Department of Finance, Stern School of Business, NYU and Dipartimento di Statistica "G. Parenti", Università di Firenze, Italy. e-mail: ctb@stern.nyu.edu

†Dipartimento di Statistica "G. Parenti", Università di Firenze, Italy. e-mail: cipollini@ds.unifi.it

‡Corresponding author: Dipartimento di Statistica "G. Parenti", Università di Firenze, Viale G.B. Morgagni, 59 - 50134 Firenze – Italy. e-mail: gallog@ds.unifi.it

1 Introduction

Multiplicative Error Models (MEMs) are born out of the availability of financial data at a very fine detail. Examples are: durations between trades or quotes, number of trades and volumes, number buys and sells within (possibly fine) intervals, volatility measures derived from ultra–high frequency data (and measures of jumps) as detailed elsewhere in this Handbook. All these variables share the feature of being non–negative valued and of exhibiting persistence features similarly to returns' variance modeled as GARCH. MEMs were first introduced as Autoregressive Conditional Duration (ACD) models by Engle and Russell (1998) and were generalized to any non–negative valued process by Engle (2002). Early work is by Chou (2005) (for range) and Manganelli (2005) (in a multivariate setting).

In a univariate setting, the simplest MEM expresses the dynamics of the variable of interest as the product of two non–negative terms, namely a scale factor (evolving in a conditionally autoregressive way paralleling GARCH specifications) times an i.i.d. error term with unit mean. The scale factor thus represents the expectation of the process conditionally on the available information, and, as such, it can be used for forecasting purposes.

There are several reasons why a direct specification of the dynamics of the variables is preferable to taking logarithms and adopt a linear model. There could be zeros in the data; even nonzero but small values may have a severe impact on estimation in a logarithmic model; deriving direct forecasts for the variable of interest is better than transforming forecasts expressed in logs; finally, as we will see in what follows, under non restrictive assumptions, the proposed estimator of the parameters in the conditional mean can be interpreted as a Quasi Maximum Likelihood estimator.

This Chapter is structured as follows. In Section 2 we introduce the univariate specification, discussing the base choices for the conditional expectation and the error term. In the same section we provide a general framework, allowing for richer specifications of the conditional mean. The outcome is a novel MEM (called *Composite MEM*) which is reminiscent of the short– and long–run component GARCH model by Engle and Lee (1999). Inference issues are discussed relative to Maximum Likelihood and Generalized Method of Moments estimation. Section 3 handles a univariate application with several kernel realized volatilities computed on Blue Chips traded on the NYSE. The estimated models are evaluated on the basis of the residual diagnostics and of forecast performance. Section 4 briefly discusses some extensions handling the presence of components in the conditional mean with a different dynamics and multivariate extensions. Concluding remarks follow.

2 Theory and Methodology

2.1 Model Formulation

Let $\{x_t\}$ be a discrete time process defined on $[0, +\infty)$, $t \in \mathbb{N}$, and let $\mathcal{F}_{t-1}$ the information available for forecasting x_t.[1] $\{x_t\}$ follows a MEM if it can be expressed as

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

where, conditionally on $\mathcal{F}_{t-1}$: μ_t is a positive quantity that evolves deterministically according to a parameter vector $\boldsymbol{\theta}$,

$$\mu_t = \mu(\boldsymbol{\theta}, \mathcal{F}_{t-1}); \tag{2}$$

ε_t is a random variable (rv) with probability density function (pdf) defined over a $[0, +\infty)$ support, unit mean and unknown constant variance,

$$\varepsilon_t | \mathcal{F}_{t-1} \sim D^+(1, \sigma^2). \tag{3}$$

Irrespective of the specification of the function $\mu(.)$ and of the distribution D^+, equations (1)-(2)-(3) entail (see Engle (2002))

$$E(x_t | \mathcal{F}_{t-1}) = \mu_t \tag{4}$$
$$V(x_t | \mathcal{F}_{t-1}) = \sigma^2 \mu_t^2. \tag{5}$$

2.1.1 Specifications for μ_t

Any practical specification of $\mu(.)$ in (2) depends on the available information $\mathcal{F}_{t-1}$: in what follows, we list some examples. In order to simplify the exposition, we discuss formulations including only one lag, at most, of the lagged effects appearing in the right hand sides. Richer structures, useful in some applications, can be obtained in a trivial way.

When $\mathcal{F}_{t-1}$ includes only past values of the series, μ_t can be specified as

$$\mu_t = \omega + \beta_1 \mu_{t-1} + \alpha_1 x_{t-1}, \tag{6}$$

which we label as baseline MEM model. The term $\beta_1 \mu_{t-1}$ represent an inertial component, whereas $\alpha_1 x_{t-1}$ stands for the contribution of the more recent observation. Alternative representations are possible: defining the forecast error in the t-th observation as

[1] The occurrence of zeros is relevant for some financial variables observed at a very high frequency; for example, fixed short intervals or illiquid assets may deliver zero volumes; large trades against small orders may correspond to zero durations; absolute returns may also be zero. Hautsch *et al.* (2010) discuss the issue at length.

$v_t = x_t - \mu_t$, and $\beta_1^* = \beta_1 + \alpha_1$, we can write

$$x_t = \omega + \beta_1^* x_{t-1} + v_t - \beta_1 v_{t-1}$$

which is an ARMA representation (with heteroskedastic errors).

Sometimes an observed signed variable determines different dynamics in x_t according to its (lagged) sign. For example, we may want to consider lagged returns (e.g. r_{t-1}) in $\mathcal{F}_{t-1}$, and define $x_t^{(-)} \equiv x_t \, \mathrm{I}(r_t < 0)$, where $\mathrm{I}(\cdot)$ denotes the indicator function. To this end, we will assume that, conditionally on $\mathcal{F}_{t-1}$, r_t has a zero median and is uncorrelated with x_t: this implies $E\left(x_t^{(-)} | \mathcal{F}_{t-1}\right) = \mu_t/2$. Accordingly, μ_t can be specified as

$$\mu_t = \omega + \beta_1 \mu_{t-1} + \alpha_1 x_{t-1} + \gamma_1 x_{t-1}^{(-)}, \tag{7}$$

which we refer to as the Asymmetric MEM. In applications where market microstructure is of interest, another relevant variable is signed trades where the sign can be attributed on the basis of whether they are 'buy' or 'sell'.

If $\{x_t\}$ is mean-stationary, then $E(x_t) = E(\mu_t) \equiv \mu$. In such a case, by taking the expectation of both members of (7) one obtains

$$\omega = \mu - (\beta_1 + \alpha_1 + \gamma_1/2)\, \mu, \tag{8}$$

which is interesting for at least two reasons. The first one is related to inference. If μ is estimated by means of the unconditional mean $\overline{x}$, then ω is removed from the optimization algorithm and can be estimated in a second step by means of (8), once estimates of α_1, γ_1 and β_1 are obtained. This strategy, that we can name *expectation targeting*, parallels the *variance targeting* approach within the GARCH framework proposed by Engle and Mezrich (1996) and investigated by Kristensen and Linton (2004) and Francq *et al.* (2009) from a theoretical point of view. The second reason deals with model interpretation. By means of (8), equation (7) can be rewritten as

$$\begin{aligned} \mu_t &= \mu + \xi_t \\ \xi_t &= \beta_1 \xi_{t-1} + \alpha_1 x_{t-1}^{(\xi)} + \gamma_1 x_{t-1}^{(\xi-)}, \end{aligned} \tag{9}$$

where

$$x_t^{(\xi)} = x_t - \mu \qquad x_t^{(\xi-)} = x_t^{(-)} - \mu/2,$$

which shows μ_t as the sum of a long-run average level, μ, and a zero-mean-stationary component, ξ_t, driven by the past values of the series (with asymmetries).

Such a representation inspires further formulations of the conditional mean, constructed by replacing the constant average level μ by time varying structures. Considering the simplest case of just one time varying component χ_t in the place of μ, the dynamics is given by

$$\mu_t = \chi_t + \xi_t \tag{10}$$

with ξ_t defined in (9) and, this time,

$$x_t^{(\xi)} = x_t - \chi_t \qquad x_t^{(\xi-)} = x_t^{(-)} - \chi_t/2. \tag{11}$$

Let us consider some possible choices for χ_t. In the presence of deterministic or predetermined variables ($\mathcal{F}_{t-1}$ is extended to include other indicators, calendar variables, etc. denoted as $\boldsymbol{z}_t$), we can write

$$\chi_t = \omega^{(\chi)} + \boldsymbol{\delta}'\boldsymbol{z}_t. \tag{12}$$

A direct inclusion of the $\boldsymbol{\delta}'\boldsymbol{z}_t$ term into the expression (7) of the conditional mean may entail unwanted persistence effects, whereby a large value of $\boldsymbol{\delta}'\boldsymbol{z}_t$ would increase μ_t, driving the following values of the conditional mean. By contrast, expression (12) allows for the distinct identification of the contribution of the past observations of x_t (together with asymmetric effects), keeping it separate from the predetermined variable(s), and preserving also its mean-stationarity.

The second extension, inspired to Brownlees and Gallo (2010), is useful when the data show a pattern evolving around some slow moving trend. In such a case, χ_t can be structured as a smooth spline function, say (omitting technical details) $s(t)$. The methodology introduced by Engle and Rangel (2008) in the GARCH context lends itself to capturing an underlying average volatility that is not constant.

As an alternative to splines, but in the same vein as the Component GARCH model proposed by Engle and Lee (1999) (see also Andersen *et al.* (2006, p.806)), we suggest here a third type of specification, in which we specify the long run component as

$$\chi_t = \omega^{(\chi)} + \beta_1^{(\chi)}\chi_{t-1} + \alpha_1^{(\chi)}x_{t-1}^{(\chi)}, \tag{13}$$

where

$$x_t^{(\chi)} = x_t - \xi_t. \tag{14}$$

We can reparameterize the short and the long run dynamics of the model, respectively (9) and (13), as

$$\begin{aligned} \xi_t &= \beta_1^*\xi_{t-1} + \alpha_1 v_{t-1} + \gamma_1 v_{t-1}^{(-)} \\ \chi_t &= \omega^{(\chi)} + \beta_1^{(\chi)*}\chi_{t-1} + \alpha_1^{(\chi)}v_{t-1}, \end{aligned} \tag{15}$$

where

$$v_t = x_t - \mu_t \qquad v_t^{(-)} = x_t^{(-)} - \mu_t/2. \tag{16}$$

$\beta_1^* = \beta_1 + \alpha_1 + \gamma_1/2$ and $\beta_1^{(\chi)*} = \beta_1^{(\chi)} + \alpha_1^{(\chi)}$. This expression of the conditional mean emphasizes the role of the innovations v_t as the driving force behind μ_t, in both the short and the long run components ξ_t and χ_t which can be interpreted as such only if we identify the model (on the interchangeability of the two components, cf. Engle and Lee (1999, p. 478)) by imposing

$$\beta_1^* < \beta_1^{(\chi)*}. \tag{17}$$

We name this model an Asymmetric *Composite* MEM[2] with the possibility to constrain

[2]We prefer the term *Composite*, since the *Component* MEM is one in which the dynamics of the con-

γ_1 to be zero.

By following Lütkepohl (2005, Section 11.6), the conditional mean of such a model has the MEM-like representation

$$\begin{aligned}\mu_t = \omega^{(\mu)} &+ \beta_1^{(\mu)}\mu_{t-1} + \beta_2^{(\mu)}\mu_{t-2} \\ &+ \alpha_1^{(\mu)}x_{t-1} + \alpha_2^{(\mu)}x_{t-2} + \gamma_1^{(\mu)}x_{t-1}^{(-)} + \gamma_2^{(\mu)}x_{t-2}^{(-)}\end{aligned} \tag{18}$$

where

$$\begin{aligned}&\omega^{(\mu)} = (1-\beta_1^*)\omega^{(\chi)} \qquad \alpha_1^{(\mu)} = \alpha_1 + \alpha_1^{(\chi)} \qquad \gamma_1^{(\mu)} = \gamma_1 \\ &\alpha_2^{(\mu)} = -(\beta_1^*\alpha_1^{(\chi)} + \beta_1^{(\chi)*}\alpha_1) \qquad \gamma_2^{(\mu)} = -\beta_1^{(\chi)*}\gamma_1 \\ &\beta_1^{(\mu)} = \beta_1^* + \beta_1^{(\chi)*} - \alpha_1^{(\mu)} - \gamma_1^{(\mu)}/2 \qquad \beta_2^{(\mu)} = -\beta_1^*\beta_1^{(\chi)*} - \alpha_2^{(\mu)} - \gamma_2^{(\mu)}/2.\end{aligned}$$

By means of (18), necessary and sufficient conditions for $\mu_t \geq 0$ can be expressed, by following Nelson and Cao (1992, Section 2.2), as $\omega^{(\mu)} \geq 0$, $\alpha_1^* \geq 0$,

$$-1 \leq \beta_2^{(\mu)} < 0, \quad 2\sqrt{-\beta_2^{(\mu)}} \leq \beta_1^{(\mu)} < 1 - \beta_2^{(\mu)}, \quad \beta_1^{(\mu)}\alpha_2^* + \left(\beta_1^{(\mu)2} + \beta_2^{(\mu)}\right)\alpha_1^* \geq 0$$

or

$$0 \leq \beta_2^{(\mu)} < 1, \quad 0 \leq \beta_1^{(\mu)} < 1 - \beta_2^{(\mu)}, \quad \beta_1^{(\mu)}\alpha_2^* + \alpha_1^* \geq 0$$

where α_l^* indicate both $\alpha_l^{(\mu)}$ and $\alpha_l^{(\mu)} + \gamma_l^{(\mu)}$, $l = 1, 2$.

Stationarity conditions can be obtained by adapting the Corollary 2.2 (and the subsequent Remark) in Bougerol and Picard (1992): if all the coefficients into (18) are non-negative and $\beta_1^{(\mu)*} + \beta_2^{(\mu)*} \leq 1$ (where $\beta_l^{(\mu)*} = \beta_l^{(\mu)} + \alpha_l^{(\mu)} - \gamma_l^{(\mu)}/2$, $l = 1, 2$) then the Composite MEM is strictly stationary. A better insight can be gained by rewriting such a condition as $\left(1 - \beta_1^{(\chi)}\right)(1 - \beta_1^*) \geq 0$: we note immediately that the Composite MEM is strictly stationary if $\beta_1^{(\chi)*}$, $\beta_1^* \leq 1$. If $\beta_1^{(\chi)}$, $\beta_1^{(\xi)} < 1$, then the process is also second order stationary; otherwise if $\beta_1^{(\chi)*} = 1$ (remembering the identification condition (17)) then the model is strictly stationary but not second order stationary. Note that conditions in this discussion are weaker than in Engle and Lee (1999).

2.1.2 Specifications for ε_t

In principle, the conditional distribution of the error term ε_t can be specified by means of any pdf having the characteristics in (3). Examples include Gamma, Log-Normal, Weibull, Inverted-Gamma and mixtures of them. Engle and Gallo (2006) favor a $Gamma(\phi, \phi)$ (implying $\sigma^2 = 1/\phi$); Bauwens and Giot (2000), in an ACD framework, consider a $Weibull\left(\Gamma(1+\phi)^{-1}, \phi\right)$ (in this case, $\sigma^2 = \Gamma(1+2\phi)/\Gamma(1+\phi)^2 - 1$). De Luca and Gallo (2010) investigate (possibly time-varying) mixtures, while Lanne (2006) adopts

ditional mean originates from elements entering multiplicatively in the specification: cf. for example, the MEM for intra–daily volume in Brownlees *et al.* (2011), discussed later in Section 4.

mixtures and a conditional expectation specification with time varying parameters.

An alternative strategy, leading to a semiparametric specification of the model, is to leave the distribution unspecified, excepts for the two conditional moments in (3).

2.2 Inference

2.2.1 Maximum Likelihood Inference

We describe Maximum Likelihood (ML) inference by assuming a MEM formulated as in Section 2.1, where ε_t has a conditional pdf expressed in generic terms by $f_\varepsilon(\varepsilon_t|\mathcal{F}_{t-1})$ (we suppress the dependence on parameters).

Because of such assumptions, $f_x(x_t|\mathcal{F}_{t-1}) = f_\varepsilon(x_t/\mu_t|\mathcal{F}_{t-1})/\mu_t$ so that the average log-likelihood function is

$$\bar{l}_T = T^{-1}\sum_{t=1}^{T} l_t = T^{-1}\sum_{t=1}^{T}\left[\ln f_\varepsilon(\varepsilon_t|\mathcal{F}_{t-1}) + \ln\varepsilon_t - \ln x_t\right],$$

where $\varepsilon_t = x_t/\mu_t$. The portions relative to $\boldsymbol{\theta}$ of the average score function, of the average outer product of gradients (OPG) and of the average Hessian are thus given by, respectively,

$$\overline{\boldsymbol{s}}_T = T^{-1}\sum_{t=1}^{T}\nabla_{\boldsymbol{\theta}} l_t = -T^{-1}\sum_{t=1}^{T}(\varepsilon_t b_t + 1)\boldsymbol{a}_t, \tag{19}$$

$$\overline{\boldsymbol{I}}_T = T^{-1}\sum_{t=1}^{T}\nabla_{\boldsymbol{\theta}} l_t \nabla_{\boldsymbol{\theta}'} l_t = T^{-1}\sum_{t=1}^{T}(\varepsilon_t b_t + 1)^2\boldsymbol{a}_t\boldsymbol{a}_t', \tag{20}$$

$$\overline{\boldsymbol{H}}_T = T^{-1}\sum_{t=1}^{T}\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'} l_t = T^{-1}\sum_{t=1}^{T}\left[\varepsilon_t\left(b_t + \varepsilon_t\nabla_{\varepsilon_t} b_t\right)\boldsymbol{a}_t\boldsymbol{a}_t' - (\varepsilon_t b_t + 1)\nabla_{\boldsymbol{\theta}}\boldsymbol{a}_t'\right], \tag{21}$$

where

$$\begin{aligned}\boldsymbol{a}_t &= \frac{1}{\mu_t}\nabla_{\boldsymbol{\theta}}\mu_t\\ b_t &= \nabla_{\varepsilon_t}\ln f_\varepsilon(\varepsilon_t|\mathcal{F}_{t-1}).\end{aligned}$$

It is worth discussing the case when ε_t is taken to be conditionally distributed as $Gamma(\phi,\phi)$ (see Section 2.1.2). In such a case,

$$b_t = \frac{\phi-1}{\varepsilon_t} - \phi \qquad \Rightarrow \qquad \varepsilon_t b_t + 1 = \phi(1-\varepsilon_t), \tag{22}$$

and

$$\nabla_{\boldsymbol{\theta}} l_t = \phi(\varepsilon_t - 1)\boldsymbol{a}_t. \tag{23}$$

Plugging this result into (19), (20) and (21), one obtains the corresponding expressions for the three objects. In particular, the resulting average score

$$\overline{\boldsymbol{s}}_T = \phi T^{-1} \sum_{t=1}^{T} (\varepsilon_t - 1)\boldsymbol{a}_t, \tag{24}$$

leads to the first order condition

$$\sum_{t=1}^{T} (\varepsilon_t - 1)\boldsymbol{a}_t = \mathbf{0}. \tag{25}$$

As a consequence, the solution $\widehat{\boldsymbol{\theta}}_T^{(ML)}$ does not depend on the dispersion parameter ϕ. This result also allows for a computational trick based on the similarity between a MEM and a GARCH model and the fact that any choice for ϕ leads to identical point estimates. Given that μ_t is the conditional expectation of x_t, its parameters can be estimated by specifying a GARCH for the conditional second moment of $\sqrt{x_t}$ while imposing its conditional mean to be zero. The wide availability of GARCH estimation routines (usually based on a normal distribution assumption) made this shortcut fairly convenient at the early stage of diffusion of MEMs: in fact, the point was raised by Engle and Russell (1998) for the case of the exponential ($\phi = 1$) in the context of ACDs, but was then shown by Engle and Gallo (2006) to be valid for any other value ϕ.[3]

Moreover, under a correct specification of μ_t, (24) guarantees that the expected score is zero even when the $Gamma(\phi, \phi)$ is not the true distribution of the error term. Note also that

$$E\left(\nabla_{\boldsymbol{\theta}} l_t \nabla_{\boldsymbol{\theta}'} l_t\right) = E\left(\phi^2 (\varepsilon_t - 1)^2 \boldsymbol{a}_t \boldsymbol{a}_t'\right) = \phi^2 \sigma^2 \boldsymbol{a}_t \boldsymbol{a}_t'$$

and

$$E\left(\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'} l_t\right) = E\left(-\phi \varepsilon_t \boldsymbol{a}_t \boldsymbol{a}_t' + \phi(\varepsilon_t - 1)\nabla_{\boldsymbol{\theta}} \boldsymbol{a}_t\right) = -\phi \boldsymbol{a}_t \boldsymbol{a}_t'$$

As per (20) and (21), (22) leads to the corresponding limiting expressions

$$\begin{aligned}
\overline{\boldsymbol{I}}_\infty &= \lim_{T\to\infty} \left[T^{-1} \sum_{t=1}^{T} E\left(\nabla_{\boldsymbol{\theta}} l_t \nabla_{\boldsymbol{\theta}'} l_t\right) \right] = \phi^2 \sigma^2 \boldsymbol{A} \\
\overline{\boldsymbol{H}}_\infty &= \lim_{T\to\infty} \left[T^{-1} \sum_{t=1}^{T} E\left(\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}'} l_t\right) \right] = -\phi \boldsymbol{A},
\end{aligned}$$

where

$$\boldsymbol{A} = \lim_{T\to\infty} \left[T^{-1} \sum_{t=1}^{T} E\left(\boldsymbol{a}_t \boldsymbol{a}_t'\right) \right].$$

Correspondingly, the OPG, Hessian and Sandwich versions of the asymptotic variance

[3] However, the code for a direct estimation of the parameters assuming any conditional pdf for ε_t is fairly easy to implement.

matrix are, respectively,

$$\begin{aligned}\mathrm{Avar}_I(\widehat{\boldsymbol{\theta}}_T^{(ML)}) &= \phi^{-2}\sigma^{-2}\boldsymbol{A}^{-1}\\ \mathrm{Avar}_H(\widehat{\boldsymbol{\theta}}_T^{(ML)}) &= \phi^{-1}\boldsymbol{A}^{-1}\\ \mathrm{Avar}_S(\widehat{\boldsymbol{\theta}}_T^{(ML)}) &= \sigma^{2}\boldsymbol{A}^{-1}.\end{aligned} \tag{26}$$

Following the discussion above, the value of ϕ is irrelevant for point estimation; by the same token, from an inferential point of view, taking ϕ as fixed (e.g. $\phi = 0.5$ as in GARCH routines) leads to three different versions of the asymptotic variance matrix, up to a scale coefficient. Equivalence is ensured by taking $\phi = \sigma^{-2}$; hence, a consistent estimator is

$$\widehat{\mathrm{Avar}}(\widehat{\boldsymbol{\theta}}_T^{(ML)}) = \widehat{\sigma}_T^2 \left[T^{-1} \sum_{t=1}^{T} \widehat{\boldsymbol{a}}_t \widehat{\boldsymbol{a}}_t' \right]^{-1},$$

where $\widehat{\sigma}_T^2$ is a consistent estimator of σ^2 and $\widehat{\boldsymbol{a}}_t$ means $\boldsymbol{a}_t$ evaluated at $\widehat{\boldsymbol{\theta}}_T^{(ML)}$.

As per σ^2, in general its ML estimator depends on the way it is related to the natural parameterization of $f_\varepsilon(\varepsilon_t|\mathcal{F}_{t-1})$. Considering again the $Gamma(\phi,\phi)$ case, the ML estimator of ϕ solves

$$\ln\phi + 1 - \psi(\phi) + T^{-1}\sum_{t=1}^{T}[\ln\widehat{\varepsilon}_t - \widehat{\varepsilon}_t] = 0,$$

where, $\psi(\cdot)$ denotes the *digamma* function and $\widehat{\varepsilon}_t$ indicates x_t/μ_t when μ_t is evaluated at $\widehat{\boldsymbol{\theta}}_T^{(ML)}$. This estimator, of course, is efficient if the true distribution is Gamma but it is unfeasible if zeros are present in the data, given that $\ln\varepsilon_t = \ln x_t - \ln\mu_t$.

2.2.2 Generalized Method of Moments Inference

A different way to estimate the model, which does not need an explicit choice of the error term distribution, is to resort to Generalized Method of Moments (GMM). Let

$$u_t = \frac{x_t}{\mu_t} - 1. \tag{27}$$

Under model assumptions, u_t is a conditionally homoskedastic martingale difference, with conditional expectation zero and conditional variance σ^2. As a consequence, let us consider any $(M,1)$ vector $\boldsymbol{G}_t$ depending deterministically on the information set $\mathcal{F}_{t-1}$ and write $\boldsymbol{G}_t u_t \equiv \boldsymbol{g}_t$. We have

$$E(\boldsymbol{g}_t|\mathcal{F}_{t-1}) = \boldsymbol{0} \quad \forall\, t \;\Rightarrow E(\boldsymbol{g}_t) = \boldsymbol{0},$$

by the law of iterated expectations. This establishes that $\boldsymbol{G}_t$ and u_t are uncorrelated and that $\boldsymbol{g}_t$ is also a martingale difference. $\boldsymbol{G}_t$ satisfies the requirements to play the instrumental role in the GMM estimation of $\boldsymbol{\theta}$ (Newey and McFadden (1994)): should it depend on a nuisance parameter σ^2, we can assume for the time being that σ^2 is a known constant,

postponing any further discussion about it to the end of this section.

If $M = p$, we have as many equations as the dimension of $\boldsymbol{\theta}$, and hence the moment criterion is

$$\overline{\boldsymbol{g}}_T = T^{-1} \sum_{t=1}^{T} \boldsymbol{g}_t = \mathbf{0}. \tag{28}$$

Under correct specification of μ_t and some regularity conditions, the GMM estimator $\widehat{\boldsymbol{\theta}}_T^{(GMM)}$, obtained solving (28) for $\boldsymbol{\theta}$, is consistent (Wooldridge (1994, th. 7.1)). Furthermore, under additional regularity conditions, we have asymptotic normality of $\widehat{\boldsymbol{\theta}}_T^{(GMM)}$, with asymptotic covariance matrix (Wooldridge (1994, th. 7.2))

$$\mathrm{Avar}(\widehat{\boldsymbol{\theta}}_T^{(GMM)}) = (\boldsymbol{S}'\boldsymbol{V}^{-1}\boldsymbol{S})^{-1}, \tag{29}$$

where

$$\boldsymbol{S} = \lim_{T\to\infty} \left[T^{-1} \sum_{t=1}^{T} E\left(\nabla_{\boldsymbol{\theta}'} \boldsymbol{g}_t\right) \right] \tag{30}$$

$$\boldsymbol{V} = \lim_{T\to\infty} \left[T^{-1} V \left(\sum_{t=1}^{T} \boldsymbol{g}_t \right) \right] = \lim_{T\to\infty} \left[T^{-1} \sum_{t=1}^{T} E\left(\boldsymbol{g}_t \boldsymbol{g}_t'\right) \right]. \tag{31}$$

The last equality for $\boldsymbol{V}$ comes from the fact that the $\boldsymbol{g}_t$'s are serially uncorrelated given that they are a martingale difference. Moreover, the same property allows one to exploit a martingale Central Limit Theorem and hence to some simplifications in the assumptions needed for asymptotic normality.

The martingale difference structure of u_t leads to a simple formulation of the *efficient* choice of the instrument $\boldsymbol{G}_t$, i.e. having the 'smallest' asymptotic variance within the class of GMM estimators considered here. Such an efficient choice is

$$\boldsymbol{G}_t^* = -E(\nabla_{\boldsymbol{\theta}} u_t | \mathcal{F}_{t-1}) E(u_t^2 | \mathcal{F}_{t-1})^{-1}. \tag{32}$$

Inserting $E\left(\boldsymbol{g}_t \boldsymbol{g}_t'\right)$ into (31) and $E\left(\nabla_{\boldsymbol{\theta}'} \boldsymbol{g}_t\right)$ into (30), we obtain

$$E\left(\boldsymbol{g}_t \boldsymbol{g}_t'\right) = -E\left(\nabla_{\boldsymbol{\theta}'} \boldsymbol{g}_t\right) = \sigma^2 E\left(\boldsymbol{G}_t^* \boldsymbol{G}_t^{*\prime}\right),$$

so that

$$\boldsymbol{V} = -\boldsymbol{S} = \sigma^2 \lim_{T\to\infty} \left[T^{-1} \sum_{t=1}^{T} E\left(\boldsymbol{G}_t^* \boldsymbol{G}_t^{*\prime}\right) \right] \tag{33}$$

and (29) specializes to

$$\mathrm{Avar}(\widehat{\boldsymbol{\theta}}_T^{(GMM)}) = (\boldsymbol{S}'\boldsymbol{V}^{-1}\boldsymbol{S})^{-1} = -\boldsymbol{S}^{-1} = \boldsymbol{V}^{-1}.$$

Considering the specific form of u_t in this model (equation (27)), we have

$$\nabla_{\boldsymbol{\theta}} u_t = -(u_t + 1)\boldsymbol{a}_t,$$

so that (32) becomes

$$\boldsymbol{G}_t^* = \sigma^{-2}\boldsymbol{a}_t.$$

Substituting it into $\boldsymbol{g}_t = \boldsymbol{G}_t u_t$ (and this, in turn, into (28)) and (33), we obtain that the GMM estimator of $\boldsymbol{\theta}$ solves the criterion equation (25) and has the asymptotic variance matrix given in (26). In practice, the same properties of $\widehat{\boldsymbol{\theta}}_T^{(ML)}$ assuming Gamma distributed errors.

In the spirit of a semiparametric approach, a straightforward estimator for σ^2 is

$$\widehat{\sigma}_T^2 = T^{-1}\sum_{t=1}^{T}\widehat{u}_t^2$$

where $\widehat{u}_t$ represents u_t evaluated at $\widehat{\boldsymbol{\theta}}_T^{(GMM)}$. Note that while this estimator does not suffer from the presence of zeros in the data, the corresponding model is not capable of predicting zero values. If such a property is needed, the model must be changed specifying a distribution of the error term as a mixture of a point mass at zero and an absolutely continuous component, following, for example, the suggestion of Hautsch *et al.* (2010). In our volatility framework such a feature is irrelevant.

3 MEMs for realized volatility

We illustrate some features of a few MEM specifications by means of an application on *realized volatility*, that is an ultra high frequency data based measure of daily return variability. The list of estimators proposed in the literature is quite extensive (among others Andersen *et al.* (2003), Aït-Sahalia *et al.* (2005), Bandi and Russell (2006), Barndorff-Nielsen *et al.* (2008)). In this work we use realized kernels (Barndorff-Nielsen *et al.* (2008)), a family of heteroskedastic and autocorrelation consistent (HAC) type estimators robust to various forms of market microstructure frictions, known to contaminate high frequency data. For the sake of space, we will not discuss the important but somewhat intricate issues about data handling and the construction of the estimator (cf. Brownlees and Gallo (2006) and Barndorff-Nielsen *et al.* (2009)).

Realized volatility measurement delivers a positive valued time series characterized by empirical regularities of persistence and asymmetric behavior according to the sign of the lagged return of the corresponding asset (cf. Andersen *et al.* (2001), among others, for equities).

We analyze volatility dynamics of ten stocks (constituents of the DJ30, as of December 2008, for which data exist between January 2, 2001 and December 31, 2008 corresponding to 1989 observations): BA (Boeing), CSCO (CISCO), DD (du Pont de Nemours), IBM (International Business Machines), JPM (JP Morgan Chase), MCD (McDonald's), PFE (Pfizer), UTX (United Technologies), WMT (Wal-Mart Stores) and XOM (Exxon Mobil). For illustration purposes we choose BA as a leading example, synthesizing the results for the other tickers. We start by fitting the realized volatility series prior to the be-

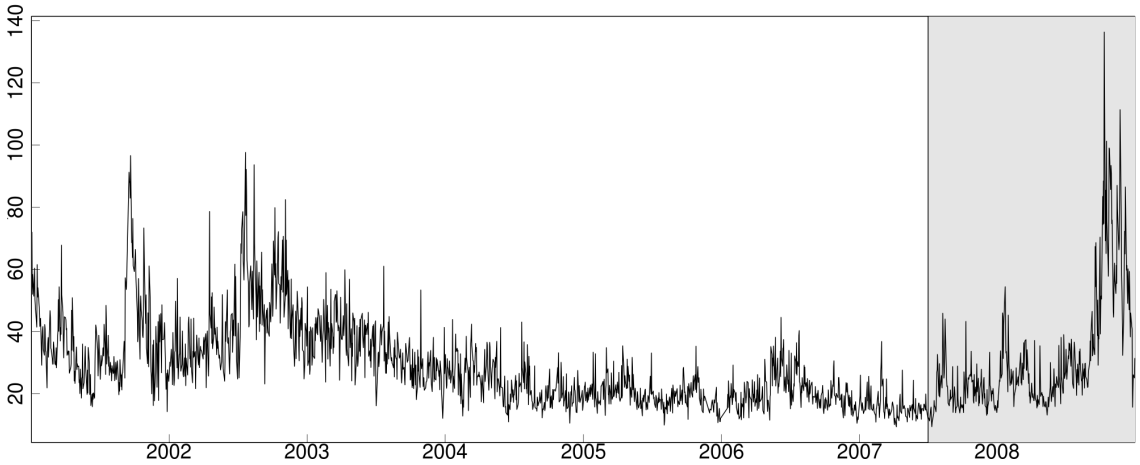

Figure 1: BA realized volatility expressed on a percent annualized scale. Jan. 2, 2001 to Dec. 31, 2008 (out–of–sample period starting Jul. 2 2007 is shaded).

Ticker	Mean	St.dev.	Kurt	ρ_{1d}	ρ_{1w}	ρ_{1m}
BA	28.82	13.42	5.80	0.81	0.71	0.58
CSCO	31.01	16.71	14.20	0.86	0.75	0.61
DD	26.53	11.54	7.93	0.79	0.70	0.50
IBM	22.14	11.39	9.02	0.86	0.77	0.63
JPM	30.20	20.98	11.23	0.86	0.76	0.64
MCD	28.35	14.06	6.69	0.72	0.60	0.51
PFE	25.21	12.08	8.96	0.74	0.62	0.52
UTX	25.59	10.76	10.29	0.76	0.63	0.48
WMT	24.37	12.63	5.93	0.86	0.78	0.67
XOM	22.25	10.05	9.01	0.80	0.73	0.54

Table 1: Realized volatility for 10 constituents of the DJ30 index. Descriptive statistics: means, standard deviations, kurtosis, as well as autocorrelations at daily, weekly and monthly levels.

ginning of the credit crunch period (January 2, 2001 to June 29, 2007, 1612 observations) and we then perform a static out–of–sample forecasting exercise for the period between July 2, 2007 and December 31, 2008 (377 observations).

The time series plot of BA realized volatility in Figure 1 shows the customary strong persistence (the widely documented *volatility clustering*), alternating periods of great turmoil to others of great calm. We discern the steep surge in volatility following the 9/11 terrorist attacks, the crisis of confidence in 2002, the protracted period of low volatility between 2003 and 2007, and the financial crisis starting in August 2007 peeking in March 2008 with the Bear Sterns demise and in September 2008 with the collapse of Lehman Brothers.

Focusing for the moment on just the two top panels of Figure 2 (histogram and autocorrelogram of the raw series), the unconditional realized volatility distribution exhibits

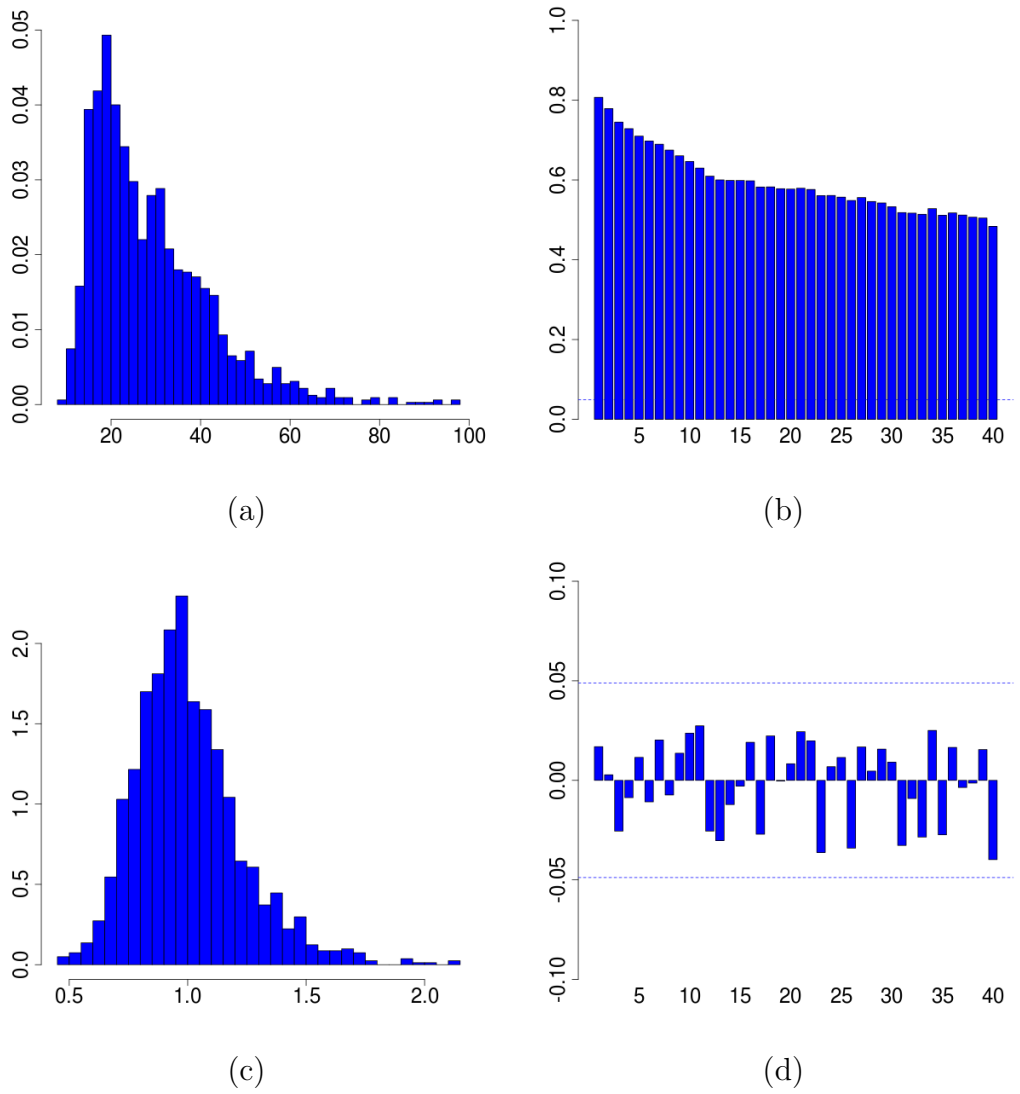

Figure 2: BA realized volatility: Histogram and autocorrelogram of the raw data (top level) and of the ACM residuals (bottom level).

a strong positive asymmetry slowly tapering off. Table 3 reports summary descriptive statistics for the various tickers, with an average volatility ranging from 22.14 to 31.01, a volatility of volatility typically around 12, and a varying degree of occurrence of extreme values as indicated by the kurtosis. Interestingly, the inspection of the autocorrelations at daily, weekly, and monthly levels reveals a common slowly decaying pattern.

We employ four different specifications for the conditional expected volatility: the (baseline) MEM (M), the Asymmetric MEM (AM), the Composite MEM (CM) and the Asymmetric Composite MEM (ACM). For each ticker and specification, the coefficients estimated over the period January 2, 2001 to June 29, 2007 are reported on the left–hand side of Table 2.

All coefficients are statistically significant at 5% (details not reported) except for the

Ticker	Model	α_1	γ_1	β_1^*	$\alpha_1^{(\chi)}$	$\beta_1^{(\chi)*}$	Q_{1d}	Q_{1w}	Q_{1m}
BA	M	0.301		0.982			0.046	0.070	0.257
	AM	0.249	0.049	0.982			0.044	0.024	0.131
	CM	0.215		0.802	0.1154	0.9957	0.733	0.926	0.975
	ACM	0.151	0.100	0.830	0.1083	0.9957	0.489	0.864	0.947
CSCO	M	0.433		0.970			0.005	0.000	0.000
	AM	0.409	0.032	0.971			0.005	0.000	0.000
	CM	0.324		0.674	0.1554	0.9917	0.864	0.103	0.509
	ACM	0.302	0.044	0.682	0.1501	0.9917	0.941	0.099	0.514
DD	M	0.306		0.972			0.100	0.040	0.098
	AM	0.243	0.044	0.973			0.038	0.039	0.127
	CM	0.193		0.765	0.1335	0.9917	0.676	0.953	0.574
	ACM	0.186	0.073	0.868	0.0772	0.9948	0.240	0.532	0.347
IBM	M	0.349		0.979			0.008	0.001	0.038
	AM	0.285	0.059	0.980			0.010	0.000	0.011
	CM	0.248		0.863	0.1235	0.9932	0.206	0.047	0.324
	ACM	0.166	-0.032	0.380	0.2602	0.9864	0.980	0.420	0.525
JPM	M	0.323		0.988			0.038	0.011	0.009
	AM	0.278	0.049	0.988			0.058	0.004	0.004
	CM	0.256		0.853	0.1001	0.9962	0.707	0.785	0.341
	ACM	0.245	0.029	0.850	0.0947	0.9962	0.764	0.787	0.332
MCD	M	0.213		0.990			0.001	0.000	0.001
	AM	0.197	0.026	0.988			0.002	0.000	0.001
	CM	0.191		0.642	0.0955	0.9967	0.880	0.871	0.937
	ACM	0.164	0.045	0.644	0.0948	0.9968	0.913	0.879	0.934
PFE	M	0.256		0.979			0.000	0.000	0.000
	AM	0.252	0.004	0.979			0.000	0.000	0.000
	CM	0.233		0.385	0.1376	0.9920	0.679	0.677	0.376
	ACM	0.255	-0.043	0.376	0.1396	0.9920	0.722	0.697	0.383
UTX	M	0.311		0.976			0.464	0.803	0.046
	AM	0.261	0.053	0.978			0.497	0.838	0.028
	CM	0.229		0.849	0.0900	0.9956	0.924	0.689	0.226
	ACM	0.212	0.029	0.851	0.0880	0.9955	0.928	0.691	0.219
WMT	M	0.272		0.984			0.008	0.021	0.037
	AM	0.256	0.037	0.984			0.024	0.041	0.028
	CM	0.206		0.814	0.0978	0.9947	0.369	0.382	0.575
	ACM	0.205	0.003	0.815	0.0976	0.9947	0.364	0.380	0.574
XOM	M	0.281		0.978			0.134	0.475	0.286
	AM	0.237	0.052	0.974			0.362	0.651	0.435
	CM	0.200		0.930	0.0909	0.9940	0.306	0.439	0.376
	ACM	0.109	0.118	0.922	0.1064	0.9912	0.264	0.424	0.364

Table 2: Estimation results for the four models analyzed: the (baseline) MEM (M), the Asymmetric MEM (AM), the Composite MEM (CM) and the Asymmetric Composite MEM (ACM). Coefficient estimates reported on the left–hand side (first five columns); Ljung-Box statistics at lag 1, 5 and 22 on the right–hand side. In–sample period: January 2, 2001 to June 29, 2007.

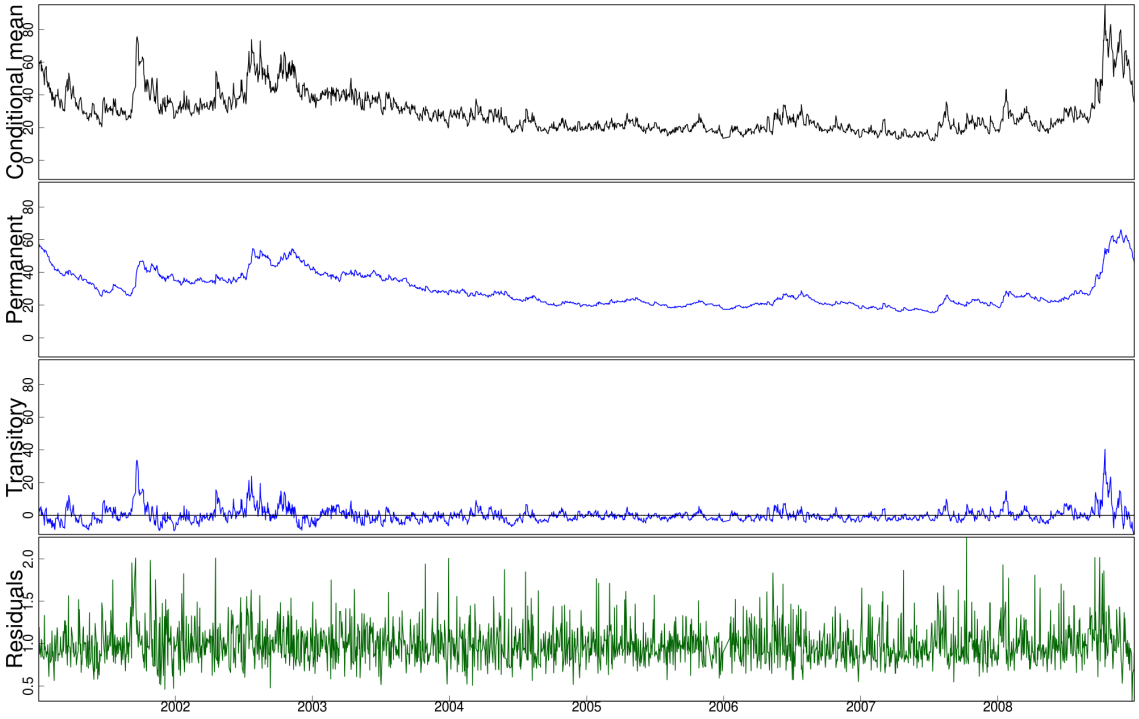

Figure 3: Panels on estimated components from the Asymmetric Composite MEM for BA (from top to bottom): overall conditional expectation; permanent component; transitory component; multiplicative residuals. Jan. 2, 2001 to Dec. 31, 2008.

asymmetric effect, for which there is a difference between the MEM and its Composite version: γ_1 is almost always significant in the AM (PFE is the only exception), but it is usually not significant in the ACM (for which only BA and XOM are have a significant γ_1). We may notice that the estimated coefficient values by specification are within a fairly narrow range providing evidence that volatility dynamics have a certain degree of homogeneity (see, among others, Bauwens and Rombouts (2007), Shephard and Sheppard (2010), Pakel *et al.* (2010), Barigozzi *et al.* (2010) and Brownlees (2010)). The baseline MEM exhibits a β_1^* very close to one, showing a high degree of persistence, and an α_1 coefficient usually between 0.25 and 0.30, which makes the dynamics more reactive to past realizations of the process than what is typically implied by past squared returns in a GARCH framework (cf. Shephard and Sheppard (2010)). Similar comments hold for the AM estimates where the asymmetric effects have an impact on the corresponding α_1's (which become smaller), while not changing the persistence parameter β_1^*. The Composite MEM enriches the specification in the direction of allowing us to confirm that the long run component has a high degree of persistence ($\beta_1^{(\chi)*}$ is close to unity) which is coherent with its slow moving underlying pattern. The short run component provides a very varied asset–specific response in the dynamics and, in particular, in the persistence β_1^* (which is lower than the other one as required by the identification condition 17). Finally, in view of the generalized non significance of the asymmetric effects, the Asymmetric Composite MEM does not seem to add much to the CM.

Figure 3 illustrates the features of the Asymmetric Composite MEM in producing the two

components, and the size of the multiplicative residual for the BA ticker (for which asymmetry is significant): the overall estimated conditional expectation reproduced in the top panel is split between the permanent component (2nd panel) and the transitory component (3rd panel); the estimated multiplicative residuals are added in the bottom panel. We can notice that the most important events (e.g. 9/11 in 2001, the WorldCom scandal in mid 2002, Bear Sterns' demise in March 2008 and Lehman Brothers' bankruptcy in Sep. 2008) are well recognizable in all panels being captured by the expected volatility divided between a sudden change in the permanent component, some turbulence in the temporary component and relatively high values of the residuals.

The right part of Table 2 reports the p–values associated with the 1 day, 1 week and 1 month Ljung–Box test statistics for each ticker and specification. Only for UTX and XOM are the test statistics non significant across specifications. For all the others, autocorrelation is a problem for the base specifications and is not solved by considering asymmetric effects. By contrast, the empirical significance levels are very high for the Composite MEMs (a typical correlogram for such a class is the one reported in the bottom right panel of Figure 2, next to a typical residual histogram – bottom left – which shows a reduction in asymmetry relative to the raw data).

It is interesting to note that the consideration of a Composite specification solves the autocorrelation problem revealed in the base specification while, at the same time, it belittles the evidence about the presence of asymmetric effects in the short run dynamics. Thus, taking the absence of autocorrelation as a lead into an improved specification, the Composite MEM would come out as the favored model. Our model provides yet another plausible 'short–memory' process capable of reproducing the 'long–memory' empirical features in the realized volatility data: for log–volatility, see Corsi (2009) who exploits dynamic components moving at different speeds as an alternative to the ARFIMA modeling of Giot and Laurent (2004).[4] Moreover, whether asymmetric effects are relevant or not in this context, becomes an empirical question (in our case, in seven out of nine tickers we lose relevance, for PFE it is never relevant). For the BA ticker, the bottom panel of Figure 3 allows for the appreciation of the uncorrelatedness of the residuals as providing no information on the volatility dynamics.

This strong result based on statistical significance is mitigated by the outcome of an out of sample static forecasting exercise of realized volatility between July 2007 and December 2008. For each day we compute the predicted realized volatility for the next period using data up to the current observation and the in–sample parameter estimates. We choose the Quasi Loglikelihood loss function, defined as

$$\mathsf{QL} = \frac{1}{T^*} \sum_{t=T+1}^{T+T^*} \left(\frac{x_t}{\hat{\mu}_t} - \ln \frac{x_t}{\hat{\mu}_t} \right) - 1,$$

for its appealing theoretical properties (see Hansen and Lunde (2005), Patton (2010)). In fact, it is a *consistent* loss function for ranking volatility models: as T^* grows, the model ranking based on the volatility proxy x_t approximates well the one based on the actual

[4]For a comparison between the Component GARCH and the FIGARCH models having similar decay of the volatility shocks see Andersen *et al.* (2006, p.807).

Ticker	M	CM	ACM
BA	3.27	5.07	-0.61
CSCO	1.91	3.67	1.63
DD	0.62	3.82	-0.43
IBM	4.61	6.46	7.38
JPM	4.07	3.74	0.53
MCD	1.24	8.06	7.60
PFE	0.24	0.90	3.66
UTX	6.16	8.91	4.74
WMT	2.64	6.06	5.81
XOM	1.18	1.33	-0.06

Table 3: Static forecasting results: July 2007 to December 2008. The table reports the percentage difference in QL loss against the AM model (equation 34).

unobserved volatility. It reaches the value 0 when all forecasts coincide with the observed values and it is greater otherwise. Furthermore, the QL loss function is a natural choice in the MEM framework in that it is related to the out–of–sample log–likelihood function under the assumption of Gamma disturbances. It turns out that the QL values for each model are fairly similar to one another, with the AM model corresponding to the lowest value in seven out of ten cases and the ACM model in the remaining three. To appreciate the differences for comparison purposes, Table 3 shows a relative index constructed as follows

$$\left(\frac{\mathsf{QL}_m}{\mathsf{QL}_{AM}} - 1\right) \times 100 \qquad m = M, CM, ACM. \tag{34}$$

The table indicates that the gains from the Composite model are of little relevance, while the gains from the AM are more substantial. Thus, judging upon these results, in choosing a single best model in forecasting, that model should be the AM model. Not surprisingly, in the crisis period the role of asymmetries becomes crucial for forecasting: in all cases the best performing model contains an asymmetric component. On the other hand, even if the Composite specifications appear to improve significantly in–sample fit, in the crisis period they improve performance only in three cases (only marginally so). This result is also in line with the findings (Brownlees *et al.* (2010)) that the TGARCH model is the best performing specification in forecasting among several GARCH alternatives. The appealing features reported in the two mid–panels of Figure 3 make the Composite model one in which some additional insights on the evolution of the dynamics are possible.

4 MEM Extensions

4.1 Component Multiplicative Error Model

The *Component* MEM, proposed by Brownlees *et al.* (2011), is an interesting representation of the model illustrated in Section 2.1. Built to reproduce the dynamics of (regularly

spaced) intra-daily volumes, the Component MEM is motivated by the salient stylized facts of such series: a clustering of daily averages, clear evidence that the overall series moves around a daily dynamic component; a regular U-shaped intra-daily pattern which emerges once daily averages are removed; and, yet, a distinctive non periodic intra-daily dynamics. The same framework can be adopted in all other contexts (volatility, number of trades, average durations) where data aggregated by intra-daily bins show some periodic features together with overall dynamics having components at different frequencies.

According to such empirical evidence, a Component MEM for $x_{t,i}$ (where $t \in \{1, \ldots, T\}$ indicates the day and $i \in \{1, \ldots, I\}$ denotes one of the I equally spaced time bins between market opening and closing times) is thus given by

$$x_{t,i} = \eta_t \, \phi_i \, \mu_{t,i} \, \varepsilon_{t,i},$$

where: η_t is a daily component; ϕ_i is an intra-daily periodic component, aimed at reproducing the time–of–day pattern; $\mu_{t,i}$ is an intra-daily dynamic (non-periodic) component; again, $\varepsilon_{t,i}$ is a non-negative disturbance term assumed i.i.d. and, conditionally on the relevant information set $\mathcal{F}_{t,i-1}$, with mean 1 and constant variance σ^2. More specifically, the components entering in the conditional mean $E\left(x_{t,i}|\mathcal{F}_{t,i-1}\right) = \eta_t \, \phi_i \, \mu_{t,i}$ can be structured according to the following, relatively simple, specifications.

The daily component is specified as

$$\eta_t = \omega^{(\eta)} + \beta_1^{(\eta)} \eta_{t-1} + \alpha_1^{(\eta)} x_{t-1}^{(\eta)} + \gamma_1^{(\eta)} x_{t-1}^{(\eta-)} \tag{35}$$

where

$$x_t^{(\eta)} = \frac{1}{I} \sum_{i=1}^{I} \frac{x_{t,i}}{\phi_i, \mu_{t,i}}, \tag{36}$$

is the *standardized daily volume*, that is the daily average of the intra-daily volumes normalized by the intra-daily components ϕ_i and $\mu_{t,i}$, and $x_t^{(\eta-)} \equiv x_t^{(\eta)} \, \mathrm{I}(r_{t,\cdot} < 0)$ (where $r_{t,\cdot}$ is the total return in day t) is a term built to capture the asymmetric effect.

The intra-daily dynamic component is formulated as

$$\mu_{t,i} = \omega^{(\mu)} + \beta_1^{(\mu)} \mu_{t,i-1} + \alpha_1^{(\mu)} x_{t,i-1}^{(\mu)} + \gamma_1^{(\mu)} x_{t,i-1}^{(\mu-)} \tag{37}$$

where

$$x_{t,i}^{(\mu)} = \frac{x_{t,i}}{\eta_t, \phi_i} \tag{38}$$

is the *standardized intra-daily volume* and $x_{t,i}^{(\mu-)} \equiv x_{t,i}^{(\mu)} \, \mathrm{I}(r_{t,i} < 0)$ (where $r_{t,i}$ is the return at bin i of day t) is the corresponding asymmetric term. $\mu_{t,i}$ is constrained to have unconditional expectation equal to 1 in order to make the model identifiable, allowing us to interpret it as a pure intra-daily dynamic component. If $r_{t,i}$ is assumed conditionally uncorrelated with $x_{t,i}$ and to have zero conditional mean (cf. Section2.1.1), such a constraint implies $\omega^{(\mu)} = 1 - (\beta_1^{(\mu)} + \alpha_1^{(\mu)} + \gamma_1^{(\mu)}/2)$.

In synthesis, the system nests the daily and the intra-daily dynamic components by al-

ternating the update of the former (from η_{t-1} to η_t) and of the latter (from $\mu_{t-1,I} = \mu_{t,0}$ to $\mu_{t,I}$): the time-varying η_t adjusts the mean level of the series, whereas the intra-daily component $\phi_i \, \mu_{t,i}$ captures bin–specific (periodic, respectively non–periodic) departures from such an average level.

Finally, the intra-daily periodic component ϕ_i can be specified in various ways. In Brownlees *et al.* (2011) a Fourier (sine/cosine) representation (whose details are omitted here for the sake of space) is supported.

Parameter estimation of the Component MEM can be done via ML or GMM, along the lines outlined in Section 2.2. In Brownlees *et al.* (2011) the GMM approach is discussed in detail.

There are two major differences in the Component MEM relative to the model defined in Section 2.1 above. The first is general: several components, each with a different meaning, are combined together multiplicatively, rather than additively as in the Composite MEM. The second aspect is related to the specific analysis of intra–daily data: when markets close, the data are not equally spaced (the time lag between two contiguous bins of the same day is different from the lag occurring between the first bin of a day and the last one of the previous day). As a consequence, some adjustments may be needed in the formulation of the dynamic component for the first daily bin (see Brownlees *et al.* (2011) for details).

4.2 Vector Multiplicative Error Model

There are many instances in which the joint consideration of several non–negative processes is of interest. For example, different measures (realized volatility, daily range, absolute returns) summarize information on return volatility but no individual one appears to be a sufficient measure (i.e. depending solely on its own past). Analyzing their joint dynamics may thus be of interest.

A second example concerns the dynamic interactions among volatilities in different markets (evaluated by means of a proxy, e.g. the daily range of the market indices) for analyzing the transmission mechanisms (spillovers, contagion) across markets (Engle *et al.* (2011)).

A third example involves order-driven markets, in which there is a tradeoff between the potential payoff of placing orders at a better price, against the risk of these orders not executing. Therefore it is relevant to investigate the dynamics of the *quantity* of stock to be executed at a given distance from the current price in itself, but also in the interaction with one another at different distances. In this framework, zeros are relevant because there are times when the quantity which could be executed at a certain distance from current price can be zero. Forecasts can be used for a trading strategy (Noss (2007)).

In a ultra-high frequency framework, the market activity is evaluated by different indicators, like the time elapsed since the last trade, the (possibly signed) volume and the return associated with the trade. A model for the dynamic interrelationship between such vari-

ables can reveal the speed (in market and calendar time) at which private information is incorporated into prices (cf. Manganelli (2005), Hautsch (2008)).

The MEM as defined in Section 4.2 can be extended to handle these situations (cf. Cipollini *et al.* (2007) and Cipollini *et al.* (2009)). Let $\{\boldsymbol{x}_t\}$ be the corresponding K–dimensional process with non–negative elements[5]. Paralleling (1)-(2)-(3), $\{\boldsymbol{x}_t\}$ follows a vector MEM (or vMEM for shortly) if it can be expressed as

$$\boldsymbol{x}_t = \boldsymbol{\mu}_t \odot \boldsymbol{\varepsilon}_t = \operatorname{diag}(\boldsymbol{\mu}_t)\boldsymbol{\varepsilon}_t,$$

where $\odot$ indicates the Hadamard (element–by–element) product and $\operatorname{diag}(\cdot)$ denotes a diagonal matrix with the vector in the argument on the main diagonal. Conditionally on the information set $\mathcal{F}_{t-1}$, $\boldsymbol{\mu}_t$ can be defined as

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}(\mathcal{F}_{t-1}; \boldsymbol{\theta}),$$

except that now we are dealing with a K– dimensional vector depending on a (larger) vector of parameters $\boldsymbol{\theta}$. The innovation vector $\boldsymbol{\varepsilon}_t$ is a K–dimensional random variable defined over a $[0, +\infty)^K$ support, with unit vector $\mathbb{1}$ as its expectation and a general variance–covariance matrix $\boldsymbol{\Sigma}$,

$$\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1} \sim D_K^+(\mathbb{1}, \boldsymbol{\Sigma}). \tag{39}$$

From the previous conditions we have

$$\begin{aligned} E(\boldsymbol{x}_t|\mathcal{F}_{t-1}) &= \boldsymbol{\mu}_t \\ V(\boldsymbol{x}_t|\mathcal{F}_{t-1}) &= \boldsymbol{\mu}_t\boldsymbol{\mu}_t' \odot \boldsymbol{\Sigma} = \operatorname{diag}(\boldsymbol{\mu}_t)\boldsymbol{\Sigma}\operatorname{diag}(\boldsymbol{\mu}_t), \end{aligned}$$

where the latter is a positive definite matrix by construction (cf. the parallel equations (4) and (5) in the univariate case).

As far as the conditional mean is concerned, the generalization of (7) becomes

$$\boldsymbol{\mu}_t = \boldsymbol{\omega} + \boldsymbol{\beta}_1\boldsymbol{\mu}_{t-1} + \boldsymbol{\alpha}_1\boldsymbol{x}_{t-1} + \boldsymbol{\gamma}_1\boldsymbol{x}_{t-1}^{(-)}.$$

Among the parameters (whose nonzero elements are arranged in the vector $\boldsymbol{\theta}$) $\boldsymbol{\omega}$ has dimension $(K, 1)$, whereas $\boldsymbol{\alpha}_1$, $\boldsymbol{\gamma}_1$ and $\boldsymbol{\beta}_1$ have dimension (K, K). As above, the term $\boldsymbol{\gamma}_1\boldsymbol{x}_{t-1}^{(-)}$ aims to capture asymmetric effects associated with the sign of an observed variable. For example, when different volatility indicators of the same asset are considered, such an indicator assumes value one when its previous day's return r_{t-1} is negative. In a market volatility spillover study, each market i would have its own indicator function built from the sign of its own returns $r_{t-1,i}$. Finally, in a microstructure context, we can think of assigning positive or negative values to volumes according to whether the trade was a buy or a sell.

[5]In what follows we will adopt the following conventions: if $\boldsymbol{x}$ is a vector or a matrix and a is a scalar, then the expressions $\boldsymbol{x} \geq \mathbf{0}$ and $\boldsymbol{x}^a$ are meant element by element; if $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K$ are (m, n) matrices then $(\boldsymbol{x}_1; \ldots; \boldsymbol{x}_K)$ means the (mK, n) matrix obtained stacking the matrices $\boldsymbol{x}_i$'s columnwise.

As for the error term $\boldsymbol{\varepsilon}_t$, a completely parametric formulation of the vMEM requires a full specification of its conditional distribution: Ahoniemi and Lanne (2009) adopt a bivariate Gamma (and mixtures) for call and put volatilities. In Cipollini *et al.* (2007) marginals for the components $\varepsilon_{t,i}$ satisfying the unit mean constraint are linked together using copulas. Relying on the considerations of Section 2.2.1, a reasonable choice is to take $Gamma(\phi_i, \phi_i)$ $(i = 1, \ldots, K)$ marginals whereas, for what concerns copulas, the Normal or Student's t represent rather flexible choices. Alternatively, Cipollini *et al.* (2009) suggest a semiparametric formulation relying on the first two moments in (39). In such a case, the estimation can be done via GMM along the lines illustrated in section 2.2.2. The resulting estimator generalizes the one obtained in the univariate case (equation (25)), in the sense of solving the criterion equation

$$\sum_{t=1}^{T} \boldsymbol{a}_t \boldsymbol{\Sigma}^{-1} (\boldsymbol{\varepsilon}_t - \mathbb{1}) = \mathbf{0} \tag{40}$$

and having the asymptotic variance matrix

$$\operatorname{Avar}(\widehat{\boldsymbol{\theta}}_T^{(GMM)}) = \left[\lim_{T \to \infty} T^{-1} \sum_{t=1}^{T} E\left[\boldsymbol{a}_t \boldsymbol{\Sigma}^{-1} \boldsymbol{a}_t' \right] \right]^{-1},$$

where

$$\boldsymbol{\varepsilon}_t = \boldsymbol{x}_t \oslash \boldsymbol{\mu}_t - \mathbb{1}$$

($\oslash$ denotes the element by element division) and

$$\boldsymbol{a}_t = \nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_t' \operatorname{diag}(\boldsymbol{\mu}_t)^{-1}.$$

The main differences with respect to the univariate case are the dependence of the criterion equation (40) on $\boldsymbol{\Sigma}$ and the fact that the same equation cannot be derived as a score function based on a known parametric distribution of the error term.

As a possible extension to univariate volatility modeling (cf. Section 3), we can analyze the joint dynamics of different measures within the vMEM framework. For example, let us consider absolute returns ($|r_t|$) and realized kernel volatility (rv_t), so that $\boldsymbol{x}_t = (x_{t,1}; x_{t,2}) = (|r_t|; rv_t)$ has conditional mean given by

$$\begin{pmatrix} \mu_{t,1} \\ \mu_{t,2} \end{pmatrix} = \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} + \begin{pmatrix} \alpha_{1,1} & \alpha_{1,2} \\ \alpha_{2,1} & \alpha_{2,2} \end{pmatrix} \begin{pmatrix} x_{t-1,1} \\ x_{t-1,2} \end{pmatrix} + \begin{pmatrix} \beta_{1,1} & \beta_{1,2} \\ \beta_{2,1} & \beta_{2,2} \end{pmatrix} \begin{pmatrix} \mu_{t-1,1} \\ \mu_{t-1,2} \end{pmatrix},$$

where asymmetric effects are not included for the sake of space. This equation shows that the vMEM encompasses the GARCH model, when $\alpha_{1,2} = \beta_{1,2} = 0$; the GARCH-X model of Engle (2002), when $\beta_{1,2} = 0$; the HEAVY model of Shephard and Sheppard (2010), when $\alpha_{1,1} = \alpha_{2,1} = \beta_{1,2} = \beta_{2,1} = 0$. A vMEM formulation is thus preferable: substantial efficiency may be gained from the joint estimation of the equations (given the high contemporaneous correlation of errors) and from modeling the possible dynamic interdependence related to a non diagonal $\boldsymbol{\beta}$; more specific models will result if testable restrictions are satisfied.

Other approaches for non negative joint processes are present in the literature. Manganelli (2005) proposes a model for durations, volumes and volatility which rests on factoring the conditional distribution

$$f(\boldsymbol{x}_t|\mathcal{F}_{t-1}) = f(x_{t,1}|\mathcal{F}_{t-1})f(x_{t,2}|x_{t,1},\mathcal{F}_{t-1})\dots f(x_{t,K}|x_{t,1},\dots,x_{t,K-1},\mathcal{F}_{t-1}).$$

Assuming uncorrelated errors, each factor can be formulated as a univariate MEM, but is driven by contemporaneous information as well. Hautsch (2008) adopts a similar strategy for modeling the intra–day dynamics of volatility, average volume per trade and number of trades evaluated at equally spaced time intervals; in comparison with Manganelli (2005), the essential difference lies in the dependence of the univariate conditional distributions on a common latent component (assumed to represent the information flow). Hansen *et al.* (2010) propose the Realized-GARCH model, where the unobservable conditional variance of returns is assumed to be driven just by the realized variance as in the HEAVY model. Since such a quantity is considered a noisy measure of the true latent variance, it ends up being indirectly dependent on the contemporaneous return as well.

A drawback from vMEMs is that the number of parameters tends to increase very rapidly with K, a fact which could potentially hinder the application domain. One possible solution is to resort to model selection techniques to isolate the elements of the coefficient matrices which could be zero (off-diagonal or $\boldsymbol{\gamma}_1$ elements, in particular). In Cipollini and Gallo (2010) an automated general to specific selection procedure is proposed and investigated. Another solution is the one investigated by Barigozzi *et al.* (2010) for a vector of volatilities where each series follows univariate MEM dynamics around a common (systematic) component (estimated nonparametrically).

5 Concluding Remarks

In this Chapter we have presented the main theoretical features of a class of models, called Multiplicative Error Models, which are particularly suitable to represent the dynamics of non–negative financial time series observed at daily or intra–daily frequency. The main advantages are a direct modeling of the persistence in the conditional mean, without resorting to logarithmic transformations which could be unfeasible in the presence of zeros or introduce numerical problems, and the possibility of forecasting the variable of interest directly. In a univariate context, several components evolving at different speeds can be considered (e.g. long– versus short–run, or daily versus intra–daily components): in most cases estimation proceeds in a fairly straightforward manner, either via a Maximum Likelihood or a Generalized Method of Moments approach. Multivariate specifications are possible, including those where some (factor–like) common features in the data can be exploited, allowing for full dynamic interdependence across several variables.

The application run for illustration purposes shows that an asymmetric specification is well suited in one-step ahead forecasting of volatility even if more sophisticated representations allow to better capture the dependence structure in the data.

References

Ahoniemi, K. and Lanne, M. (2009). Joint modeling of call and put implied volatility. *International Journal of Forecasting*, **25**, 239–258.

Aït-Sahalia, Y., Mykland, P. A., and Zhang, L. (2005). How often to sample a continuous–time process in the presence of market microstructure noise. *The Review of Financial Studies*, **28**, 351–416.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, **61**, 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, **71**, 579–625.

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

Bandi, F. M. and Russell, J. R. (2006). Separating microstructure noise from volatility. *Journal of Financial Economics*, **79**, 655–692.

Barigozzi, M., Brownlees, C. T., Gallo, G. M., and Veredas, D. (2010). Disentangling systemic and idiosyncratic volatility for large panels of assets. a seminonparametric vector multiplicative error model. Technical report, ECARES.

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

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2009). Realised kernels in practice: trades and quotes. *Econometrics Journal*, **12**, 1–32.

Bauwens, L. and Giot, P. (2000). The logarithmic acd model: An application to the bid-ask quote process of three nyse stocks. *Annales d'Économie et de Statistique*, **60**, 117–149.

Bauwens, L. and Rombouts, J. V. K. (2007). Bayesian clustering of many GARCH models. *Econometric Reviews*, **26**, 365–386.

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

Brownlees (2010). On the relation between firm characteristics and volatility dynamics with an application to the 2007-2009 financial crisis. Technical report.

Brownlees, C. T. and Gallo, G. M. (2006). Financial econometric analysis at ultra–high frequency: Data handling concerns. *Computational Statistics and Data Analysis*, **51**, 2232–2245.

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

Brownlees, C. T., Engle, R., and Kelly, B. T. (2010). A practical guide to volatility forecasting through calm and storm. Technical report, Department of Finance, NYU.

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

Chou, R. Y. (2005). Forecasting financial volatilities with extreme values: The conditional autoregressive range (CARR) model. *Journal of Money, Credit and Banking*, **37**(3), 561–582.

Cipollini, F. and Gallo, G. M. (2010). Automated variable selection in vector multiplicative error models. *Computational Statistics & Data Analysis*, **54**(11), 2470–2486.

Cipollini, F., Engle, R. F., and Gallo, G. M. (2007). A model for multivariate non-negative valued processes in financial econometrics. Technical Report 2007/16.

Cipollini, F., Engle, R. F., and Gallo, G. M. (2009). Semiparametric vector MEM. Econometrics Working Papers Archive wp200903, Università degli Studi di Firenze, Dipartimento di Statistica "G. Parenti".

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

De Luca, G. and Gallo, G. M. (2010). A time-varying mixing multiplicative error model for realized volatility. Econometrics Working Papers Archive wp2010_03, Università degli Studi di Firenze, Dipartimento di Statistica "G. Parenti".

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

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

Engle, R. F. and Lee, G. J. (1999). A permanent and transitory component model of stock return volatility. In R. F. Engle and H. White, editors, *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W. J. Granger*, pages 475–497. Oxford University Press, Oxford.

Engle, R. F. and Mezrich, J. (1996). GARCH for groups. *Risk*, **9**, 36–40.

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

Engle, R. F. and Russell, J. R. (1998). Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica*, **66**, 1127–62.

Engle, R. F., Gallo, G. M., and Velucchi, M. (2011). Volatility spillovers in East Asian financial markets: A MEM based approach. *Review of Economics and Statistics*, (forthcoming).

Francq, C., Horvath, L., and Zakoïan, J.-M. (2009). Merits and drawbacks of variance targeting in GARCH models. Technical report.

Giot, P. and Laurent, S. (2004). Modelling daily value–at–risk using realized volatility and ARCH type models. *Journal of Empirical Finance*, **11**, 379–398.

Hansen, P. R. and Lunde, A. (2005). Consistent ranking of volatility models. *Journal of Econometrics*, **131**(1–2), 97–121.

Hansen, P. R., Huang, Z., and Shek, H. H. (2010). Realized GARCH: A complete model of returns and realized measures of volatility. Technical report.

Hautsch, N. (2008). Capturing common components in high-frequency financial time series: A multivariate stochastic multiplicative error model. *Journal of Economic Dynamics and Control*, **32**, 3978 – 4015.

Hautsch, N., Malec, P., and Schienle, M. (2010). Capturing the zero: a new class of zero–augmented distributions and multiplicative error processes. Technical report, Humboldt Universität zu Berlin. SFB 649, DP 2010-055.

Kristensen, D. and Linton, O. (2004). Consistent standard errors for target variance approach to GARCH estimation. *Econometric Theory*, **20**, 990–993.

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

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer, Berlin Heidelberg.

Manganelli, S. (2005). Duration, volume and volatility impact of trades. *Journal of Financial Markets*, **8**, 377–399.

Nelson, D. B. and Cao, C. Q. (1992). Inequality constraints in the univariate GARCH model. *Journal of Business and Economic Statistics*, **10**, 229–235.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. McFadden, editors, *Handbook of Econometrics*, volume 4, chapter 36, pages 2111–2245. Elsevier.

Noss, J. (2007). The econometrics of optimal execution in an order-driven market. Technical report, Nuffield College, Oxford University, Master's Thesis.

Pakel, C., Shephard, N., and Sheppard, K. (2010). Nuisance parameters, composite likelihoods and a panel of GARCH models. *Statistica Sinica*. forthcoming.

Patton, A. (2010). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*. forthcoming.

Shephard, N. and Sheppard, K. (2010). Realising the future: forecasting with high frequency based volatility (heavy) models. *Journal of Applied Econometrics*, **25**, 197–231.

Wooldridge, J. M. (1994). Estimation and inference for dependent processes. In R. F. Engle and D. McFadden, editors, *Handbook of Econometrics*, volume 4, chapter 45, pages 2639–2738. Elsevier.