Modeling financial time series with multiplicative errors

Giampiero M. Gallo

NYU in Florence and CRENOS

8th Gretl Conference – Gdańsk, Poland
June 15, 2023
A bird’s eye view on Multiplicative Error Models

- Focus on the evolution of this class of models
- Take the univariate MEM as a leading case of representation and inference issues
- How data structure suggests refinements
- Show a representation of a vector MEM
- Open issues (dimensionality, model selection, etc.)
In financial econometrics, returns as the main object of analysis.

- Financial volatility has been extensively investigated for more than twenty-five years
- Risk-related motivations
- Conditional density evaluation for VaR, ES
- Strong empirical regularities about GARCH models
- Ultra-high frequency data have allowed for more detailed analysis of market activity
- Clustering spreads over to other financial time series
GARCH conditional variance is the expectation of squared returns (if zero mean return): autoregressive dynamics

- A lot of information available in financial markets is positive valued:
  - ultra-high frequency data provides intra–daily time intervals: range, volume, number of trades, number of buys/sells, durations)
  - daily volatility estimators (realized volatility, daily range, absolute returns)

- Time series exhibit persistence which can be modeled à la GARCH
Abs Returns

Autocorrelation 0.29
Daily Range – Parkinson (1980)

\[ h_l^P = \frac{1}{4 \log(2)} (\log H_t - \log L_T) \]

Autocorrelation 0.59

\[(h_t^{GK})^2 = 0.511 \log(H_t/L_t)^2 - 0.019\{\log(C_t/O_t)(\log H_t + \log L_t - 2 \log O_t) - 2(\log(H_t/O_t) \log(L_t/O_t))\} - 0.383 \log(C_t/O_t)^2\]

Autocorrelation 0.67 – Correlation with Parkinson measure 0.96
Volume in million shares

Autocorrelation 0.58 – Correlation with daily range 0.51
Multiplicative Error Models

- Extension of the GARCH approach to modeling the expected value of processes with positive support (Engle, 2002; Engle and Gallo, 2006)
- Autoregressive Conditional Duration (Engle and Russell, 1998) is a special case. Absolute returns, high-low, number of trades in a certain interval, volume, realized volatility can be modeled with the same structure.
- Rather than calling the models Autoregressive Conditional Volatility, Autoregressive Conditional Volume, etc. call them MEM
- Ease of estimation (more later)
- Expand the information set or introduce different components (main interesting results)
The Base Model

Consider

- $x_t$, a non-negative univariate process,
- $\mathcal{F}_{t-1}$ the information about the process up to time $t - 1$.

A MEM for $x_t$ is specified as

$$x_t = \mu_t \varepsilon_t$$
The Base Model

Consider

- $x_t$, a non-negative univariate process,
- $\mathcal{F}_{t-1}$ the information about the process up to time $t - 1$.

A MEM for $x_t$ is specified as

$$x_t = \mu_t \varepsilon_t$$
The Base Model

Consider

- $x_t$, a non-negative univariate process,
- $\mathcal{F}_{t-1}$ the information about the process up to time $t - 1$.

A MEM for $x_t$ is specified as

$$x_t = \mu_t \varepsilon_t$$
The Base Model

Consider

- $x_t$, a non-negative univariate process,
- $\mathcal{F}_{t-1}$ the information about the process up to time $t - 1$.

A MEM for $x_t$ is specified as

$$x_t = \mu_t \epsilon_t$$

Conditional on $\mathcal{F}_{t-1}$:

$\mu_t$ is a nonnegative *predictable* process, depending on a vector of unknown parameters $\theta$,

$$\mu_t = \mu_t(\theta);$$
The Base Model

Consider

- $x_t$, a non-negative univariate process,
- $\mathcal{F}_{t-1}$ the information about the process up to time $t - 1$.

A MEM for $x_t$ is specified as

$$x_t = \mu_t \varepsilon_t$$

Conditional on $\mathcal{F}_{t-1}$: $\varepsilon_t$ is a conditionally stochastic i.i.d. process, with density having non-negative support, mean 1 and unknown variance $\sigma^2$,

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

- $x_t$, a non-negative univariate process,
- $\mathcal{F}_{t-1}$ the information about the process up to time $t - 1$.

A MEM for $x_t$ is specified as

$$x_t = \mu_t \varepsilon_t$$

As a consequence

$$E(x_t|\mathcal{F}_{t-1}) = \mu_t$$
$$V(x_t|\mathcal{F}_{t-1}) = \sigma^2 \mu_t^2.$$
The specification of $\mu_t$

- **Base (1, 1) specification for** $\mu_t$

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

- **Asymmetric (à la GJR) specification**: if $x_t$ can tap on info about $r_t$ (e.g. $hl_{t-1}$ can be associated with the observed sign of $r_{t-1}$):

  $\mu_t = \omega + \alpha x_{t-1} + \gamma x_{t-1}^{(-)} + \beta \mu_{t-1}$,

  where $x_{t}^{(-)} = x_t 1_{(r_t<0)}$.

- **Constant unconditional expectation** $E(x_t) = \frac{\omega}{1-\alpha-\beta-\gamma/2}$
A Gamma Assumption for $\varepsilon_t$

Flexible parameterization

$$\varepsilon_t \mid \mathcal{F}_{t-1} \sim \text{Gamma}(\phi, \phi),$$

with $E(\varepsilon_t \mid \mathcal{F}_{t-1}) = 1$ and $V(\varepsilon_t \mid \mathcal{F}_{t-1}) = 1/\phi$.

As a consequence, $x_t \mid \mathcal{F}_{t-1} \sim \text{Gamma}(\phi, \phi/\mu_t)$. 
A useful relationship is between the Gamma distribution and the Generalized Error Distribution (GED). We have:

\[ x_t|\mathcal{F}_{t-1} \sim \text{Gamma}(\phi, \phi/\mu_t) \iff x_t^\phi|\mathcal{F}_{t-1} \sim \text{Half - GED}(0, \mu_t^\phi, \phi). \]

The conditional densities of \( x_t \) and of \( x_t^\phi \) are related. In particular, \( \phi = 0.5 \)

\[ X_t = \mu_t \varepsilon_t \iff \sqrt{X_t} = \sqrt{\mu_t} \nu_t \]

where

\[ \nu_t|\mathcal{F}_{t-1} \sim \text{Half - Normal}(0, 1). \]

This will provide a trick to estimate a MEM with a standard GARCH package with normal innovations and Bollerslev–Wooldridge standard errors.
Consider a MEM for squared returns \( r_t^2 \)

\[
r_t^2 = h_t \varepsilon_t
\]

with \( h_t = E(r_t^2|\mathcal{F}_{t-1}) \) estimated by a GARCH routine choosing \( r_t \) as the dependent variable, setting \( E(r_t|\mathcal{F}_{t-1}) = 0 \) with normal errors for the returns.

Numerically the same results choosing \( |r_t| \) as the dependent variable, setting (nonsensically) \( E(|r_t||\mathcal{F}_{t-1}) = 0 \)

Hence, if \( hl_t \) is of interest, take \( \sqrt{hl_t} \) as the dependent variable, set its conditional mean to zero and normal errors: the GARCH results are the MEM estimation.

For GJR flavor recolor \( \sqrt{hl_t} \) with the sign of returns

\[
\sqrt{hl_t}(1 - 2 \mathbb{1}_{r_t<0})
\]
The poor person’s guide to MEM estimation cont. d

\[ \sqrt{h_t} \]

\[ \sqrt{h_t}(1 - 2 \mathbb{1}_{r_t < 0}) \]
The poor person’s guide to MEM estimation cont.d

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>z-Statistic</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.622314</td>
<td>0.080533</td>
<td>7.727400</td>
<td>0.0000</td>
</tr>
<tr>
<td>$hl_{t-1}$</td>
<td>0.195050</td>
<td>0.015805</td>
<td>12.34083</td>
<td>0.0000</td>
</tr>
<tr>
<td>$hl_{t-1} I_{r_{t-1} &lt; 0}$</td>
<td>0.052596</td>
<td>0.009957</td>
<td>5.282513</td>
<td>0.0000</td>
</tr>
<tr>
<td>$hl_{t-1}$</td>
<td>0.732919</td>
<td>0.019955</td>
<td>36.72779</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Listed as Variance Equation
Contribution of $x_t$ to the log–likelihood function $l_t$

$$l_t = \ln L_t = \phi \ln \phi - \ln \Gamma(\phi) + (\phi - 1) \ln x_t - \phi (\ln \mu_t + x_t / \mu_t).$$

Contribution of $x_t$ to the score $s_t = \begin{pmatrix} s_{t,\theta} \\ s_{t,\phi} \end{pmatrix}$ with components

$$s_{t,\theta} = \nabla_{\theta} l_t = \phi \left( x_t - \frac{\mu_t}{\mu_t^2} \right) \nabla_{\theta} \mu_t$$

$$s_{t,\phi} = \nabla_{\phi} l_t = \ln \phi + 1 - \psi(\phi) + \ln \left( \frac{x_t}{\mu_t} \right) - \frac{x_t}{\mu_t},$$

where $\psi(\phi) = \frac{\Gamma'(\phi)}{\Gamma(\phi)}$ is the *digamma* function and the operator $\nabla_{\lambda}$ denotes the derivative with respect to $\lambda$. 
Estimation – cont.d

Contribution of $x_t$ to the Hessian $H_t = \begin{pmatrix} H_{t,\theta\theta'} & H_{t,\theta\phi} \\ H_{t,\theta\phi}' & H_{t,\phi\phi} \end{pmatrix}$ with components

$$H_{t,\theta\theta'} = \nabla_{\theta\theta'} l_t = \phi \left( \frac{-2x_t + \mu_t}{\mu_t^3} \nabla_\theta \mu_t \nabla_{\theta'} \mu_t + \frac{x_t - \mu_t}{\mu_t^2} \nabla_\theta \theta' \mu_t \right)$$

$$H_{t,\theta\phi} = \nabla_{\theta\phi} l_t = \frac{x_t - \mu_t}{\mu_t^2} \nabla_\theta \mu_t$$

$$H_{t,\phi\phi} = \nabla_{\phi\phi} l_t = \frac{1}{\phi} - \psi'(\phi),$$

where $\psi'(\phi)$ is the trigamma function.
First order conditions for $\theta$ and $\phi$

\[
\frac{1}{T} \sum_{t=1}^{T} \frac{x_t - \mu_t}{\mu_t^2} \nabla \theta \mu_t = 0
\]

\[
\ln \phi + 1 - \psi(\phi) + \frac{1}{T} \sum_{t=1}^{T} \left[ \ln \left( \frac{x_t}{\mu_t} \right) - \frac{x_t}{\mu_t} \right] = 0
\]

- First–order conditions for $\theta$ do not depend on $\phi$, i.e. same point estimates for $\theta$ whatever value $\phi$ may take
- $\phi$ can be estimated after $\theta$.
- If $\mu_t = E(x_t|F_{t-1})$, the expected value of the score of $\theta$ evaluated at the true parameters is zero irrespective of the Gamma assumption on $\varepsilon_t|F_{t-1}$
- Estimator is QML
Asymptotic variance–covariance matrix

\[ V_\infty = \left( \begin{array}{ccc} 1 & 1 & 0 \\ \frac{1}{T} \sum_{t=1}^{T} \frac{1}{\mu_t^2} \nabla \theta \mu_t \nabla \theta' \mu_t & 0 & \psi'(\phi) - \frac{1}{\phi} \\ 0 & 0 & \psi'(\phi) - \frac{1}{\phi} \end{array} \right)^{-1} \]

- The variance of \( \widehat{\theta} \) is proportional to \( 1/\phi \).
- \( \widehat{\theta} \) and \( \widehat{\phi} \) asymptotically uncorrelated.
- With \( v_t = x_t/\mu_t - 1 \), simple MoM estimator

\[ \widehat{\phi^{-1}} = \frac{1}{T} \sum_{t=1}^{T} \hat{V}_t^2. \]

not affected by the presence of zero \( x_t \)'s.
The *sandwich* estimator gets rid of the dependence of the submatrix relative to $\theta$ on $\phi$ altogether

$$\hat{V}_\infty = \hat{H}_T^{-1} \hat{OPG}_T \hat{H}_T^{-1}$$

This is where the poor person’s way to estimate a MEM for a single equation via a GARCH for the square root of the variable of interest, needs Bollerslev-Wooldridge standard errors.
In spite of QML properties, pursue more flexible GMM without an explicit choice of the error term distribution, based on

\[ \varepsilon_t = \frac{x_t}{\mu_t} \]

Under model assumptions, \( \varepsilon_t - 1 \) is a conditionally homoskedastic martingale difference, with conditional expectation zero and conditional variance \( \sigma^2 \).

The efficient GMM estimator of \( \theta \), say \( \hat{\theta}_{GMM} \), solves the criterion equation

\[ \sum_{t=1}^{T} (\varepsilon_t - 1) a_t = 0, \quad \text{where} \quad a_t = \frac{1}{\mu_t} \nabla_{\mu_t} \theta \mu_t \]
GMM estimation – cont.d

- $\hat{\theta}_{GMM}$ has asymptotic variance matrix

$$\text{Avar}(\hat{\theta}_{GMM}) = \sigma^2 A^{-1},$$

where

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

- A consistent estimator of the asymptotic variance matrix is

$$\hat{\text{Avar}}(\hat{\theta}_{GMM}) = \tilde{\sigma}^2 \hat{A}^{-1},$$

with relevant objects $\hat{\varepsilon}_t$ and $\hat{a}_t$ evaluated at $\hat{\theta}_{GMM}$

$$\hat{\sigma}^2 = T^{-1} \sum_{t=1}^{T} (\hat{\varepsilon}_t - 1)^2 \quad \text{and} \quad \hat{A} = T^{-1} \sum_{t=1}^{T} \hat{a}_t \hat{a}'_t.$$
An important equivalence between GMM and QML

Let \( \varepsilon_t | F_{t-1} \sim \text{Gamma}(\phi, \phi) \) (so that \( E(\varepsilon_t | F_{t-1}) = 1 \) and \( V(\varepsilon_t | F_{t-1}) = \sigma^2 = 1/\phi \)). The log-likelihood function is

\[
I_T = \sum_{t=1}^{T} \left[ \phi \ln \phi - \ln \Gamma(\phi) + \phi \ln \varepsilon_t - \phi \varepsilon_t - \ln x_t \right].
\]

Maximization wrt \( \theta \) involves just (\( \phi \) is irrelevant)

\[
\sum_{t=1}^{T} (\ln \varepsilon_t - \varepsilon_t).
\]

The f.o.c. for \( \theta \) is equal to the GMM condition

\[
\sum_{t=1}^{T} \nabla_{\theta \mu_t} \frac{x_t - \mu_t}{\mu_t^2} = \sum_{t=1}^{T} (\varepsilon_t - 1) a_t = 0,
\]

Under correct specification of \( \mu_t \) the term has a zero expectation even when \( \varepsilon_t \) is not Gamma-distributed.
Univariate MEM à la carte

Univariate MEM

Base AMEM

\[
\omega + \left( \alpha + \gamma \cdot \mathbb{1}_{(r_{t-1} < 0)} \right) x_{t-1} + \beta_1 \mu_{t-1}
\]

Predetermined Variables

\[
\omega + \left( \alpha + \gamma \cdot \mathbb{1}_{(r_{t-1} < 0)} \right) x_{t-1} + \beta_1 \mu_{t-1} + \delta' z_{t-1}
\]

Stylized Facts

Component MEM

\[
x_t = \tau t \xi_t \epsilon_t
\]
Univariate extensions

1. Changing average level of volatility in MEMs
2. MEMs to mitigate measurement error effects in volatility dynamics
Some stylized facts: the typical behavior of a volatility series

The Fear Index: the VIX

Shaded areas indicate U.S. recessions. Source: Chicago Board Options Exchange fred.stlouisfed.org
Some stylized facts: the typical behavior of a volatility series

The S&P500 Realized Kernel Volatility

Zoom 1m 3m 6m YTD 1y All

From Oct 12, 2007 To Jan 28, 2022
Approaches to modeling a low frequency component

Need to modify assumption of a constant unconditional (long range) volatility: idea of a local average which is time–varying

1 Additive model as in the Two–Component GARCH (Engle and Lee, 1999): a permanent (identified by high persistence) and a transitory one

\[
h_t = q_t + \alpha (\epsilon_{t-1}^2 - q_{t-1}) + \beta (h_{t-1} - q_{t-1}) \\
q_t = \omega + \rho q_{t-1} + \phi (\epsilon_{t-1}^2 - h_{t-1})
\]

with \( \rho > \alpha + \beta \) for identification of the permanent component (extension to MEM available).

2 Multiplicative model: consider a combination of multiplicative components, one of which \( (\tau_t) \) corresponds to a slow moving average level of volatility.

3 Conrad and Schienle (2018) devise an LM test for such an omitted multiplicative component
Conditional models for volatility on the boxing ring

1. On your right side, true GARCH for returns:

\[ r_t = \sqrt{\tau_t h_t} \eta_t \quad E(r_t^2 | F_{t-1}) = \tau_t h_t \]

typically, \( \eta_t \sim N(0, 1) \) or Student’s t; \( r_t \) close–to–close log–returns; \( \tau_t \) is the low–frequency component, \( h_t \) is the high–frequency component;

2. On your left side, true MEM for volatility–type:

\[ x_t = \mu_t \varepsilon_t = \tau_t \xi_t \epsilon_t \quad E(x_t | F_{t-1}) = \mu_t = \tau_t \xi_t \]

where \( x_t = \sigma_t^2 \), or \( \sigma_t \), or \( \log(\sigma_t^2) \); \( \sigma_t^2 \) can be one of the many realized variance measures, daily range (or other market activity measures), \( \tau_t \) is the low–frequency component, \( \xi_t \) is the high–frequency component;
The Low-frequency Component $\tau_t$

Insightful review paper in the GARCH world by Amado, Silvennoinen and Teräsvirta (2019)

- Curve fitting approach - deterministic. Spline GARCH by Engle and Rangel (2008): goal to find macroeconomic determinants of volatility (ex post)
- Smooth Transition approach. Amado and Teräsvirta (2008): link $\tau_t$ to a logistic function
- Markov Switching approach. Dueker (1997), Haas et al. (2004): average level of volatility by regime gives $\tau_t$ as a step function
- GARCH–MIDAS approach. Derive $\tau_t$ as a filter of past observations of data available at different frequencies
Parallel treatment of the Low-frequency Component in MEMs

- Markov Switching and Smooth Transition MEMs are suggested in a IJoF paper with E. Otranto (2015) introducing the concept of Local Average Volatility.
- B-splines in a MEM are suggested by Brownlees and G. (2010).
- A common smooth factor extracted from a panel of realized volatilities is derived in Barigozzi et al. (2014).
- MEM–MIDAS suggested by Amendola et al.
Let’s take a Brian De Palma’s cut...
...showing the bottom line at the very start

S&P 500 – in solid blue line $\hat{\tau}_t$ from a MEM–MIDAS

Monthly aggregated realized volatility (solid red line), the predicted MEM-MIDAS aggregated realized volatility (dotted blue line). Annualized scale.
A different treatment of the low frequency component

S&P 500 – local average volatilities in a MS-AMEM(3)

The step function is the local average volatility calculated across three MS regimes. Annualized scale.
Let \( \{x_{i,t}\} \) refer to the \( i \)-th day \((i = 1, \ldots, N_t)\) of the period \( t \) (a week, a month or a quarter; \( t = 1, \ldots, T \)) with \( \mathcal{F}_{i,t} \) be the information set available at day \( i \) of period \( t \).

Reparameterize the base MEM as

\[
x_{i,t} = \mu_{i,t} \epsilon_{i,t} = \tau \xi_{i,t} \epsilon_{i,t},
\]

where: \( \tau \) is a constant; \( \xi_{i,t} \) is a quantity that, conditionally on \( \mathcal{F}_{i-1,t} \), evolves deterministically; \( \epsilon_{i,t} \) is an error term such that

\[
\epsilon_{i,t|\mathcal{F}_{i-1,t}} \sim iid \mathcal{D}(1, \sigma^2),
\]

\[
E(x_{i,t|\mathcal{F}_{i-1,t}}) = \tau \xi_{i,t} \quad Var(x_{i,t|\mathcal{F}_{i-1,t}}) = \sigma^2 \tau^2 \xi_{i,t}^2.
\]
Extension: Doubly Multiplicative Error Model

Recent joint work with A. Amendola, V. Candila and F. Cipollini

Specification for the conditional mean with a multiplicative component structure, with both factors time–varying.

\[ x_{i,t} = \tau_{i,t} \xi_{i,t} \varepsilon_{i,t}. \]

- \( \tau_{i,t} \) is the long run component: a slow–moving component determining the average level of the conditional mean at any given time. It may refer to a different frequency or not.

- \( \xi_{i,t} \) is the short run or fast–moving component: centered around one, with the role to dampen (when <) or to amplify \( \tau_{i,t} \) (when > 1).
Doubly Multiplicative Error Model: short run

The short run component can be expressed as a MEM, augmented by the contribution of a predetermined de–meaned (vector) variable $z$ within a DMEMX

$$\xi_{i,t} = (1 - \alpha_1 - \gamma_1/2 - \beta_1) + \alpha_1 x_{i-1,t}^{(\xi)} + \gamma_1 x_{i-1,t}^{(\xi^-)} + \beta_1 \xi_{i-1,t} + \delta'_1 z_{i-1,t}$$

where

$$x_{i,t}^{(\xi)} \equiv \frac{x_{i,t}}{\tau_{i,t}} \quad x_{i,t}^{(\xi^-)} \equiv x_{i,t}^{(\xi)} 1_{(r_{i,t}<0)}.$$ 

$x_{i,t}^{(\xi^-)}$ is a variable derived from $x_{i,t}^{(\xi)}$ which takes a non-zero value only if it corresponds to a negative return (for asymmetric effects).
Doubly Multiplicative Error Model: long run

Specifications for $\tau_{i,t}$

- A first possibility is to adapt a spline function
- A second possibility is to structure $\tau_{i,t}$ in a way similar to $\xi_{i,t}$, namely

$$
\tau_{i,t} = \omega(\tau) + \alpha_1(\tau) x_{i-1,t}^{(\tau)} + \gamma_1(\tau) x_{i-1,t}^{(\tau-)} + \beta_1(\tau) \tau_{i-1,t}
$$

where

$$
x_{i,t}^{(\tau)} \equiv \frac{x_{i,t}}{\xi_{i,t}} \quad x_{i,t}^{(\tau-)} \equiv x_{i,t}^{(\tau)} \mathbb{1}(r_{i,t} < 0).
$$

we call this CMEM
How to assemble a MEM–MIDAS

- In the case of a mixed–frequency framework

\[ x_{i,t} | \mathcal{F}_{i-1,t} = \tau_t \xi_{i,t} \varepsilon_{i,t} \]

\[ \varepsilon_{i,t} \sim \left( 1, \frac{1}{\phi} \right) \]

\[ \xi_{i,t} = (1 - \alpha - \beta - \gamma/2) + \left( \alpha + \gamma \cdot \mathbb{1}(r_{i-1,t} < 0) \right) \frac{x_{i-1,t}}{\tau_t} + \beta \xi_{i-1,t} \]

- MIDAS filter

\[ \tau_t = \exp \left\{ m + \vartheta \sum_{k=1}^{K} \delta_k(\omega) X_{t-k} \right\} \]

\[ \delta_k(\omega) = \frac{\left( \frac{k}{K} \right)^{\omega_1-1} \left( 1 - \frac{k}{K} \right)^{\omega_2-1}}{\sum_{j=1}^{K} \left( \frac{j}{K} \right)^{\omega_1-1} \left( 1 - \frac{j}{K} \right)^{\omega_2-1}} \]

- GMM inference works as before
Spline MEM

![Graph showing Spline MEM with data points for S&P 500, FTSE 100, NASDAQ, and Hang Seng from 2001 to 2013. The graph includes low-frequency component (Annualized %) for each year. The x-axis represents the years from 2001 to 2013, and the y-axis represents the low-frequency component percentage. The graph highlights fluctuations and trends in financial series over time.]
C–MEM

Low-frequency component (Annualized %)


S&P 500
FTSE 100
NASDAQ
Hang Seng
MEM–MIDAS

![Graph showing low-frequency component (Annualized %) over years from 2001 to 2013 for S&P 500, FTSE 100, NASDAQ, and Hang Seng.]
Multiplicative errors mitigate measurement errors effects

Joint work with F. Cipollini and E. Otranto

- MEMs provide an alternative to the treatment by Bollerslev, Patton and Quaedvlieg (2016) for measurement error in realized volatility dynamics
- Problem: realized variance measures integrated variance of a continuous time process with error
- When specifying dynamic models for $RV_t$ the estimated relationship less persistent than the “true” one (attenuation bias)
- Framework chosen: HAR enlarged to HARQ by including an interaction term between $RV_t$ and realized quarticity

$$rv_t = \omega + (\alpha_D + \alpha_Erq_{t-1}^{1/2})rv_{t-1} + \alpha_W\overline{rv}_{t-2(5)} + \alpha_M\overline{rv}_{t-6(22)} + ut$$
Our take: measurement errors are multiplicative

- Stylized facts: (sqrt-)quarticity is strongly correlated with RealVar (in our panel, median = 0.934)

- If we insert squared RealVar in lieu of the interaction term we have similar results (significant negative coefficient)

- When analyzing the nature of the measurement errors, they are heteroskedastic

\[
RV_t = IV_t + \eta_t \\
= IV_t + \sqrt{2\Delta}IQ^{1/2}_t z_t \\
\approx IV_t + \sqrt{2\Delta}IV_t z_t \\
= IV_t \cdot (1 + \sqrt{2\Delta}z_t) \\
= IV_t \cdot \varepsilon_t.
\]

hence multiplicative errors
Our strategy

If that is the case

\[
RV_t = \begin{cases} 
E(RV_t | I_{t-1}) + \eta_t, & \eta_t \text{ zero mean, heteroskedastic} \\
E(RV_t | I_{t-1}) \cdot \varepsilon_t, & \varepsilon_t \text{ unit mean, homoskedastic}.
\end{cases}
\]

- **Alternative explanation**: lagged variance has a curvature effect within HAR → nonlinear effect which reduces persistence

- High levels of lagged RealVar imply a faster absorption of news and a faster reversion to the mean

- **Fundamental Questions**: Is this HAR a well specified model (are we rather catching heteroskedasticity à la White?, cf also Corsi, Mittnik, Pigorsch\(^2\), 2008)

- Which mean to revert to? overall constant? or regime specific?
Comparison of robust AMEM Specifications of $\mu_t$

- **AMEM(/Q/2)**

\[
\mu_t = \omega + \beta_1 \mu_{t-1} + (\alpha_1 + \alpha E h_{t-1}) r\nu_{t-1} + \gamma_1 r\nu_{t-1}^{(-)}
\]

\[
\varepsilon_t | I_{t-1} \sim \text{Gamma}(a, 1/a)
\]

- **MS-AMEM(/Q/2)** [Gallo and Otranto, 2015]

\[
\mu_{t,s_t} = \omega_{s_t} + \beta_{s_t} \mu_{t-1,s_{t-1}} + (\alpha_{s_t} + \alpha_E h_{t-1}) r\nu_{t-1} + \gamma_{s_t} r\nu_{t-1}^{(-)}
\]

\[
\varepsilon_t | s_t, I_{t-1} \sim \text{Gamma}(a_{s_t}, 1/a_{s_t})
\]

where $s_t \in \{1, 2, 3\}$ and $P(s_t = j | s_{t-1} = i) = p_{ij}$.

- Different specifications depending on how $h_t$ is defined:
  - $h_t = \alpha_E \equiv 0 \rightarrow$ AMEM, MS-AMEM
  - $h_t = r q_t^{1/2} \rightarrow$ AMEMQ, MS-AMEMQ
  - $h_t = r\nu_t \rightarrow$ AMEM2, MS-AMEM2
Our conclusions

- Evidence of curvature within HAR class of models could be attributed to alternative explanations (higher variances induce a faster mean reversion)
- **but** HAR being misspecified, a multiplicative specification (AMEM) takes heteroskedasticity into account and does not find a strong evidence for the extra terms
- A further refinement with Markov switching regime specific mean and short term dynamics eliminates the evidence of a curvature, with a substantial gain in predictive terms (in–and out–of–sample)
- Simulating from AMEM’s **without curvature** provides estimated curvature effects in a HAR–type model.
- **No need to pay money for the quarticity series:** use MS-AMEM2 or AMEM2
The vector MEM

Extension to the multivariate case

Non-negative-valued processes taken together: several indicators of the same market activity OR same indicator (e.g. volatility) for different markets
Consider

$\mathbf{x}_t$, a non-negative univariate vector $(N \times 1)$ process,

$\mathcal{F}_{t-1}$ the information about the process up to time $t - 1$.

A MEM for $\mathbf{x}_t$ is specified as

$$ \mathbf{x}_t = \mu_t \odot \varepsilon_t $$

Conditional on $\mathcal{F}_{t-1}$:

- The components $\mu_{i,t}$ are predictable process, depending
  on a vector of unknown parameters $\theta$,

  $$ \mu_{i,t} = \mu_{i,t}(\theta); $$

- $\varepsilon_t$ is a conditionally stochastic i.i.d. process, with density
  having non-negative support, mean 1 and unknown variance $\Sigma^2$,

  $$ \varepsilon_t | \mathcal{F}_{t-1} \sim D(1, \Sigma^2). $$
vector MEM by Estimation Method

- Equation by Equation
- Copula Functions
- GMM
The vector Multiplicative Error Model

\[ x_t = \mu_t \odot \varepsilon_t = \text{diag}(\mu_t)\varepsilon_t. \]

Conditionally on \( \mathcal{F}_{t-1} \):

- \( \mu_t \) is a \( K \)-dimensional vector depending on a vector of parameters \( \theta \). Example:

\[
\mu_t = \omega + \alpha x_{t-1} + \gamma x_{t-1}^{(-)} + \beta \mu_{t-1}
\]

Equation by equation does not work if \( \beta \) is not diagonal

- \( \varepsilon_t \) is a iid multiplicative error term

\[
\varepsilon_t|\mathcal{F}_{t-1} \sim (1, \Sigma)
\]
The vector Multiplicative Error Model

From the definition:

\[ E(x_t | \mathcal{F}_{t-1}) = \mu_t \]
\[ V(x_t | \mathcal{F}_{t-1}) = \mu_t \mu'_t \odot \Sigma = \text{diag}(\mu_t) \Sigma \text{diag}(\mu_t) \]

▶ \( \theta \) is the parameter of main interest
▶ \( \Sigma \) is a nuisance parameter

For forecasting, considering a second lag in the specification:

\[ \mu_{t+\tau} = \omega^* + A_1 \mu_{t+\tau-1} + A_2 \mu_{t+\tau-2}, \]

can be solved recursively for any horizon \( \tau \).
Impulse Response Analysis

From

\[ h_l_t = \mu_t \odot \epsilon_t \] (2)

Interpret \( \mu_{t+\tau} = E (h_{l_t+\tau} | l_t, \epsilon_t = 1) \) and contrast it with \( \mu^{(i)}_{t+\tau} = E (h_{l_t+\tau} | l_t, \epsilon_t = 1 + s^{(i)}) \), for a generic vector of shocks \( s^{(i)} \).

The element–by–element division (\( \odot \)) of the two vectors

\[ \rho^{(i)}_{t,\tau} = (\mu^{(i)}_{t+\tau} \odot \mu_{t+\tau}) - 1 \quad \tau = 1, \ldots, K \] (3)

gives us the MEM impulse response function to a shock in a market.
Impulse Response to a shock in Hong Kong

Originating Market: HK - Starting Date: 10/22/97

Response vs. Horizon
Step 1:
Let us define
\[ u_t = x_t \odot \mu_t - 1 = \varepsilon_t - 1. \]
as an working residual. Hence
\[ E(u_t | F_{t-1}) = 0 \]
\[ V(u_t | F_{t-1}) = \Sigma \]
so that \( u_t \) is a martingale difference.
Step 2:
Let $G_t$ an instrument, i.e. a $(M, K)$-matrix

- depending deterministically on $\mathcal{F}_{t-1}$;
- (possibly) depending on a vector of nuisance parameters $\psi$, for the time being taken as fixed.

Then

$$E(G_t u_t | \mathcal{F}_{t-1}) = 0 = E(G_t u_t)$$

and $g_t = G_t u_t$ also is a martingale difference.

This provides $M$ moment conditions. If $M = p$, we have as many equations as the dimension of $\theta$
Step 3:
If $M = p$, we have the MM criterion

$$\frac{1}{T} \sum_{t=1}^{T} g_t = 0$$

where $g_t = G_t u_t$. 
Efficient GMM Inference

Main general results: (Wooldridge, 1994, th. 7.1, 7.2)
Under correct specification of the $\mu_t$ equation, the GMM estimator $\hat{\theta}_T$, obtained by solving the moment conditions for $\theta$, is consistent and asymptotically normal with asymptotic variance matrix

$$Avar(\hat{\theta}_T) = \frac{1}{T} S^{-1} V S^{-1'}$$

where

$$S = \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} E (\nabla_{\theta} g_t)$$

$$V = \lim_{T \to \infty} \frac{1}{T} V \left( \sum_{t=1}^{T} g_t \right)$$
Efficient GMM Inference

Step 4:
Being \( g_t = G_t u_t \) a martingale difference leads to a simple formulation for the efficient choice of the instrument \( G_t \)

\[
G^*_t = -E(\nabla_\theta u_t' | \mathcal{F}_{t-1}) V(u_t | \mathcal{F}_{t-1})^{-1}.
\]

Efficient is meant producing the ’smallest’ asymptotic variance matrix among the GMM estimators obtained solving the moment conditions.
Step 5:
Computing the efficient instrument $G_t^*$ for the vMEM and plugging it into the moment conditions we obtain

$$\frac{1}{T} \sum_{t=1}^{T} \nabla_{\theta} \mu_t' [\text{diag}(\mu_t) \Sigma \text{diag}(\mu_t)]^{-1} (x_t - \mu_t) = 0$$

together with a (relatively) simple expression of $\text{Avar}(\hat{\theta}_T)$. 
Efficient GMM Inference

Remarks:

- Identical inferences can be obtained by means of QML in a declination named Weighted Nonlinear Least Squares (WNLS) (Wooldridge, 1994)

- In the $K = 1$ case, the moment equation specializes as the 1-order condition of the univariate MEM under Gamma assumption of $\varepsilon_t$ (Engle and Gallo, 2006)

- Main difference of the vector case: it is impossible to remove $\Sigma$ from the moment equation. Hence, it is important to investigate its role in making inference about $\theta$
Inference on $\Sigma$

- The nuisance parameter $\Sigma$ is not fixed and has to be estimated. Are there consequences on inference for $\theta$?
- Omitting the details (rather technical!) the answer is... no.
- In practice, an inconsistent estimate of $\Sigma$ does not affect consistency of $\hat{\theta}_T$.
- Since $\Sigma = V(u_t|F_{t-1})$, a natural estimator for $\Sigma$ is

$$\hat{\Sigma}_T = \frac{1}{T} \sum_{t=1}^{T} u_t u_t'$$

where $u_t = x_t \odot \mu_t - 1$ is the working residual computed at current values of $\hat{\theta}_T$.

- **Remark:** this estimator is not compromised by zeros in the data.
Variable selection: a Lasso approach

- Put all coefficients of $\mu_t$ into a vector $\delta$.
- The unrestricted model often contains zero parameters: inefficient parameter estimates, and poor forecasting performance.
- The Adaptive Lasso selects the model and estimates the parameters simultaneously.
- Let $\hat{\delta}_j = 1/|\hat{\delta}_j(mle)|^\xi$ for some $\xi > 0$.

$$
\hat{\delta}(\lambda_T) = \arg\min_{\delta} \left\{ -\frac{1}{T} \ell(\tilde{\delta}) + \lambda_T \sum_{j=1}^d \hat{\omega}_j |\tilde{\delta}_j| \right\}. \quad (4)
$$

- $\lambda_T$ is selected with a cross validation approach.
- oracle property for Adaptive Lasso–vMEM: it is consistent in variable selection and performs as well as if the true underlying model were given in advance.
Network of interactions

The network of interactions across markets.
Open questions

- A low frequency component measures the secular movements of the volatility (local average concept)
- Statistically, it can be reproduced in a variety of ways:
  - Markov Switching has the appeal to allow for different dynamics and identify volatility regimes; possibility of a forcing variable in transition probabilities for interpretation
  - Smooth transition introduces the persistence in the component and possibility of a forcing variable for interpretation
  - Deterministic exploits the *fitting* capabilities
  - MIDAS is built on a forcing variable with more suitable lower frequency as the volatility component
Open questions

- Economic interpretability with transmission mechanisms from the real economy (try housing starts)
- Which monthly variable? credit spread, realized volatility
- Refinements on the MEM-MIDAS - insert double asymmetry in the MIDAS component
- Common component to different markets – what is left out? multivariate version?
- Different drivers – combined additively in $\tau_t$?
- Cascading components – combined multiplicatively
- Horse race with other $\tau_t$ specifications
Summing Up

- MEM as a flexible class of models to estimate conditional expectations of non-negative processes both univariate (with extra predetermined variables) and multivariate

- Doubly multiplicative model captures a wide range of features suggested by data structure

- Challenge: handle large panel of data/impose common component structure for a more parsimonious/more tractable specification
Estimation reasonably simple in a GMM framework
Needs a major econometric software to implement it
Thank You!