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.)

From Conditional Variance to Conditional Means

In financial econometrics, returns as the main object of analysis.

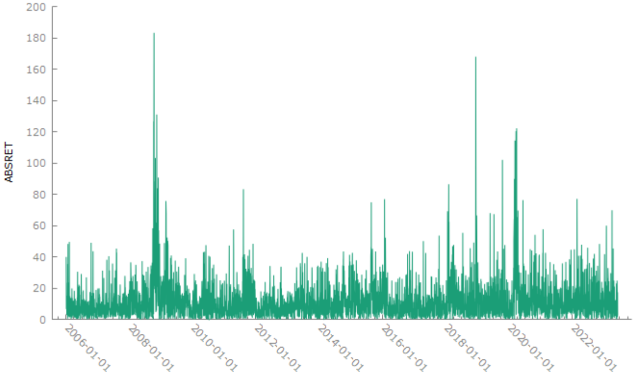
- Financial volatility has been extensively investigated for more than twenty-five years
- Risk-related motivations
- Conditional density evaluation for VaR, ES
- Strong empirical regularities about GARCH models
- Ultra-high frequency data have allowed for more detailed analysis of market activity
- Clustering spreads over to other financial time series

Modeling non-negative time series: GARCH as a MEM

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*

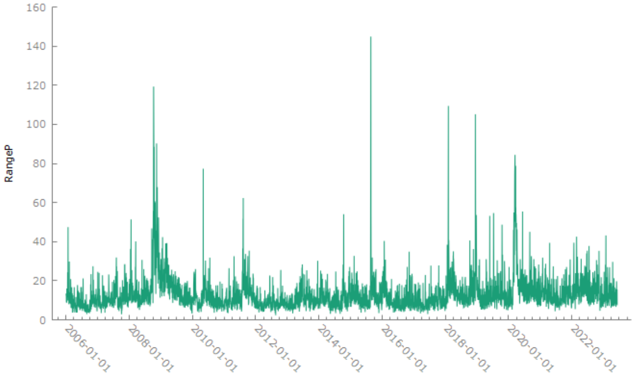
Abs Returns



Autocorrelation 0.29

Daily Range – Parkinson (1980)

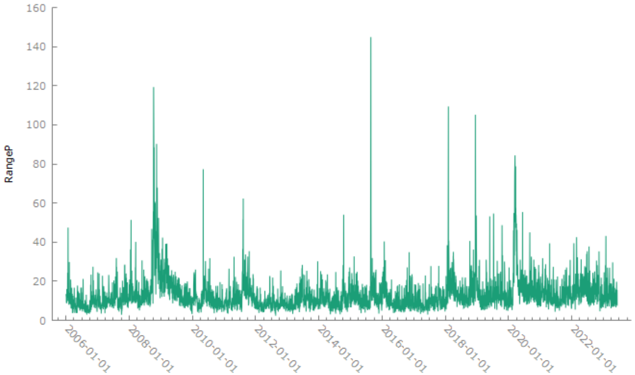
$$hl_t^P = \frac{1}{4\log(2)}(\log H_t - \log L_T)$$



Autocorrelation 0.59

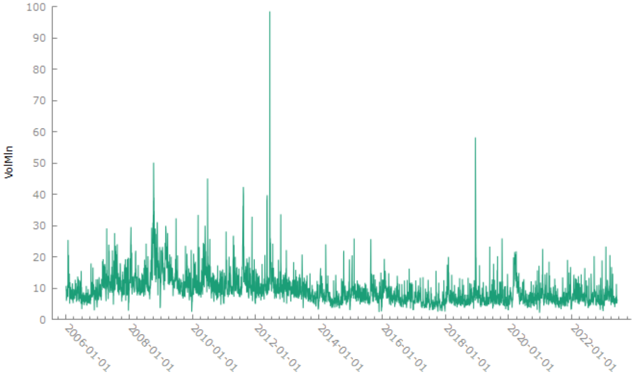
Daily Range – Garman and Klass (1980)

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



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

$$x_t = \mu_t \; \varepsilon_t$$

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

μ_t is a nonnegative *predictable* process, depending on a vector of unknown parameters $\boldsymbol{\theta}$,

$$\mu_t = \mu_t(\boldsymbol{\theta});$$

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

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

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

$$x_t = \mu_t \, \varepsilon_t$$

As a consequence

$$\begin{aligned} E(x_t|\mathcal{F}_{t-1}) &= \mu_t \\ V(x_t|\mathcal{F}_{t-1}) &= \sigma^2 \mu_t^2. \end{aligned}$$

The specification of μ_t

- *Base* $(1,1)$ specification for μ_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 $x_t^{(-)} = x_t \, \mathrm{I}_{(r_t<0)}$.

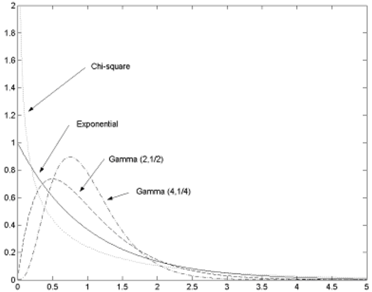
- Constant unconditional expectation $E(x_t) = \frac{\omega}{1-\alpha-\beta-\gamma/2}$

A Gamma Assumption for ε_t

Flexible parameterization

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

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



As a consequence, $x_t|\mathcal{F}_{t-1} \sim Gamma(\phi, \phi/\mu_t)$.

A useful relationship is between the Gamma distribution and the Generalized Error Distribution (GED). We have:

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

The conditional densities of x_t and of x_t^{ϕ} are related. In particular, $\phi = 0.5$

$$x_t = \mu_t \varepsilon_t \qquad \Leftrightarrow \qquad \sqrt{x_t} = \sqrt{\mu_t} \nu_t$$

where

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

This will provide a *trick* to estimate a MEM with a standard GARCH package with normal innovations and Bollerslev–Wooldridge standard errors.

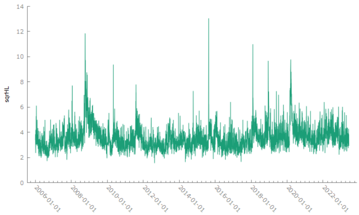
The poor person's guide to MEM estimation

- Consider a MEM for squared returns r_t^2

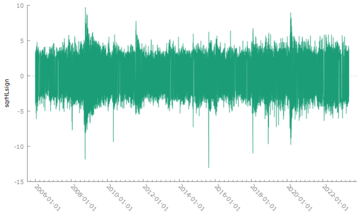
$$r_t^2 = h_t \varepsilon_t$$

  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 $\sqrt{hl_t}(1 - 2\,\mathrm{I}_{r_t<0})$

The poor person's guide to MEM estimation cont.d



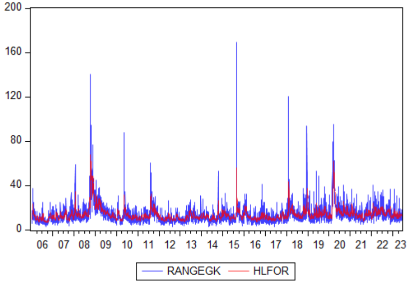
$$\sqrt{hl_t}$$



$$\sqrt{hl_t}(1 - 2\,\mathrm{I}_{r_t<0})$$

The poor person's guide to MEM estimation cont.d

Variable	Coefficient	Std. Error	z-Statistic	Prob.
	Listed as Variance Equation			
Constant	0.622314	0.080533	7.727400	0.0000
hl_{t-1}	0.195050	0.015805	12.34083	0.0000
$hl_{t-1}\, \mathrm{I}_{r_{t-1}<0}$	0.052596	0.009957	5.282513	0.0000
hl_{t-1}	0.732919	0.019955	36.72779	0.0000



Estimation

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 $\mathbf{s}_t = \begin{pmatrix} \mathbf{s}_{t,\boldsymbol{\theta}} \\ s_{t,\phi} \end{pmatrix}$ with components

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

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

where $\psi(\phi) = \frac{\Gamma'(\phi)}{\Gamma(\phi)}$ is the *digamma* function and the operator ∇_λ denotes the derivative with respect to λ.

Estimation – cont.d

Contribution of x_t to the Hessian $\mathbf{H}_t = \begin{pmatrix} \mathbf{H}_{t,\boldsymbol{\theta}\boldsymbol{\theta}'} & \mathbf{H}_{t,\boldsymbol{\theta}\phi} \\ \mathbf{H}'_{t,\boldsymbol{\theta}\phi} & \mathbf{H}_{t,\phi\phi} \end{pmatrix}$ with components

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

$$\mathbf{H}_{t,\boldsymbol{\theta}\phi} = \nabla_{\boldsymbol{\theta}\phi} l_t = \frac{x_t - \mu_t}{\mu_t^2} \nabla_{\boldsymbol{\theta}} \mu_t$$

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

where $\psi'(\phi)$ is the *trigamma* function.

First order conditions for θ and ϕ

$$\frac{1}{T}\sum_{t=1}^{T}\frac{x_t-\mu_t}{\mu_t^2}\nabla_{\boldsymbol{\theta}}\mu_t = 0$$

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

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

Asymptotic variance–covariance matrix

$$V_\infty = \begin{pmatrix} \phi \frac{1}{T} \sum_{t=1}^{T} \frac{1}{\mu_t^2} \nabla_{\boldsymbol{\theta}} \mu_t \nabla_{\boldsymbol{\theta}'} \mu_t & \mathbf{0} \\ \mathbf{0} & \psi'(\phi) - \frac{1}{\phi} \end{pmatrix}^{-1}$$

- The variance of $\widehat{\boldsymbol{\theta}}$ is proportional to $1/\phi$
- $\widehat{\boldsymbol{\theta}}$ 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.

Asymptotic variance–covariance matrix - cont.d

The *sandwich* estimator gets rid of the dependence of the submatrix relative to $\boldsymbol{\theta}$ on ϕ altogether

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

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.

Serious stuff: GMM estimation

1. In spite of QML properties, pursue more flexible GMM without an explicit choice of the error term distribution, based on
$$\varepsilon_t = \frac{x_t}{\mu_t}$$
2. Under model assumptions, $\varepsilon_t - 1$ is a conditionally homoskedastic martingale difference, with conditional expectation zero and conditional variance σ^2.
3. The *efficient* GMM estimator of $\boldsymbol{\theta}$, say $\widehat{\boldsymbol{\theta}}_{GMM}$, solves the criterion equation

$$\sum_{t=1}^{T}(\varepsilon_t - 1)\boldsymbol{a}_t = \mathbf{0}, \qquad \text{where } \boldsymbol{a}_t = \frac{1}{\mu_t}\nabla_{\boldsymbol{\theta}}\mu_t$$

GMM estimation – cont.d

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

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

where

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

- A consistent estimator of the asymptotic variance matrix is

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

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

$$\widehat{\sigma}^2 = T^{-1} \sum_{t=1}^{T} (\widehat{\varepsilon}_t - 1)^2 \qquad \text{and} \qquad \widehat{\boldsymbol{A}} = T^{-1} \sum_{t=1}^{T} \widehat{\boldsymbol{a}}_t \widehat{\boldsymbol{a}}_t'$$

An important equivalence between GMM and QML

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

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

Maximization wrt $\boldsymbol{\theta}$ involves just (ϕ is irrelevant)

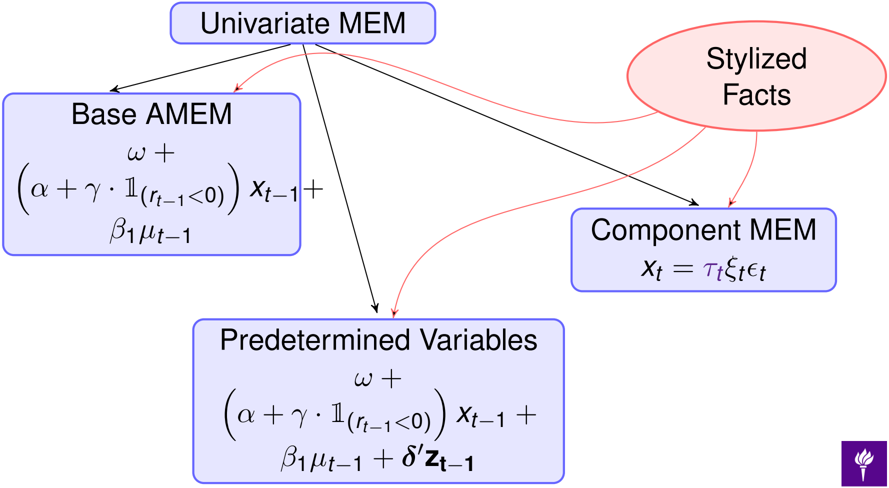
$$\sum_{t=1}^{T} (\ln \varepsilon_t - \varepsilon_t) .$$

The f.o.c. for $\boldsymbol{\theta}$ is equal to the GMM condition

$$\sum_{t=1}^{T} \nabla_{\boldsymbol{\theta}} \mu_t \frac{x_t - \mu_t}{\mu_t^2} = \sum_{t=1}^{T} (\varepsilon_t - 1)\boldsymbol{a}_t = \mathbf{0},$$

Under correct specification of μ_t the term has a zero expectation even when ε_t is not Gamma-distributed.

Univariate MEM *à la carte*



Univariate extensions

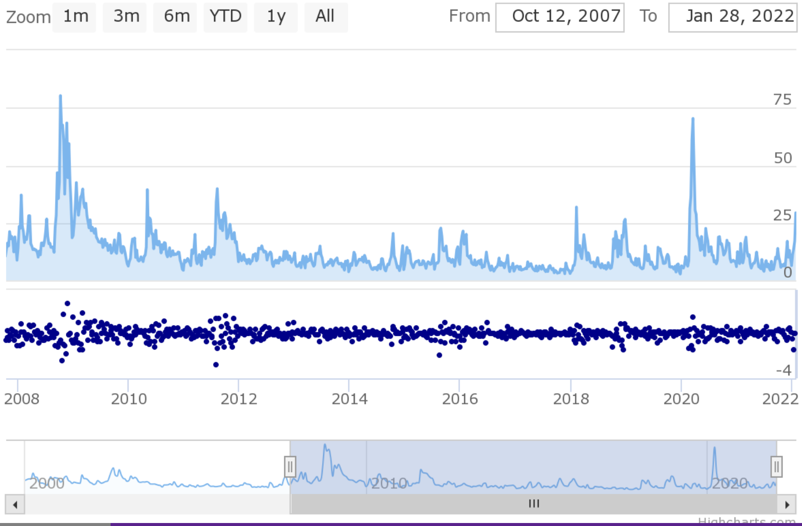
1. Changing average level of volatility in MEMs
2. MEMs to mitigate measurement error effects in volatility dynamics

Some stylized facts: the typical behavior of a volatility series

The Fear Index: the VIX

Some stylized facts: the typical behavior of a volatility series

The S&P500 Realized Kernel Volatility



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

$$\begin{aligned} h_t &= q_t + \alpha(\epsilon_{t-1}^2 - q_{t-1}) + \beta(h_{t-1} - q_{t-1}) \\ q_t &= \omega + \rho q_{t-1} + \phi(\epsilon_{t-1}^2 - h_{t-1}) \end{aligned}$$

   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 (τ_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

NYU

Conditional models for volatility on the boxing ring

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

$$r_t = \sqrt{\tau_t\, h_t}\,\eta_t \qquad E(r_t^2|\mathcal{F}_{t-1}) = \tau_t\, h_t$$

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

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

$$x_t = \mu_t\,\varepsilon_t = \tau_t\,\xi_t\,\epsilon_t \qquad E(x_t|\mathcal{F}_{t-1}) = \mu_t = \tau_t\,\xi_t$$

where $x_t = \sigma_t^2$, or σ_t, or $\log(\sigma_t^2)$; σ_t^2 can be one of the many realized variance measures, daily range (or other market activity measures), τ_t is the low–frequency component, ξ_t is the high–frequency component;

The Low-frequency Component τ_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 τ_t to a logistic function
- Markov Switching approach. Dueker (1997), Haas et al. (2004): average level of volatility by regime gives τ_t as a step function
- GARCH–MIDAS approach. Derive τ_t as a filter of past observations of data available at different frequencies

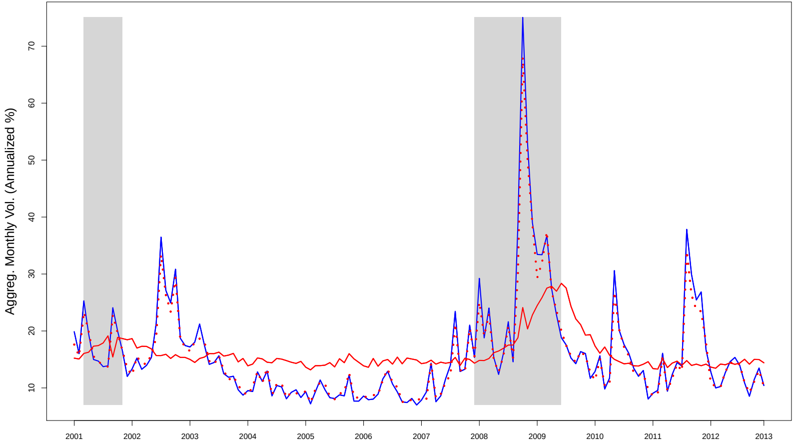
Parallel treatment of the Low-frequency Component in MEMs

- 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

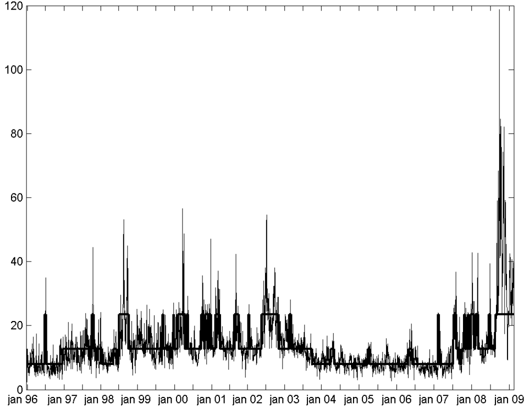
S&P 500 – in solid blue line $\hat{\tau}_t$ from a MEM–MIDAS



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.

Component Models

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: τ 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

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

$$E(x_{i,t}|\mathcal{F}_{i-1,t}) = \tau\xi_{i,t} \qquad Var(x_{i,t}|\mathcal{F}_{i-1,t}) = \sigma^2\tau^2\xi^2_{i,t}.$$

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 $\boldsymbol{z}$ within a DMEMX

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

where

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

$x^{(\xi-)}_{i,t}$ is a variable derived from $x^{(\xi)}_{i,t}$ 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

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

we call this CMEM

How to assemble a MEM–MIDAS

- In the case of a mixed–frequency framework

$$x_{i,t}|\mathcal{F}_{i-1,t} = \tau_t\, \xi_{i,t}\, \epsilon_{i,t} \qquad \varepsilon_{i,t} \overset{i.i.d}{\sim} \left(1, \frac{1}{\phi}\right)$$

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

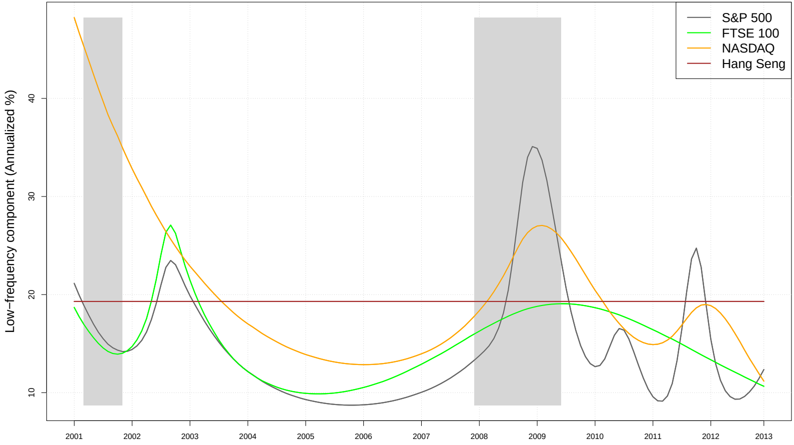
- MIDAS filter

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

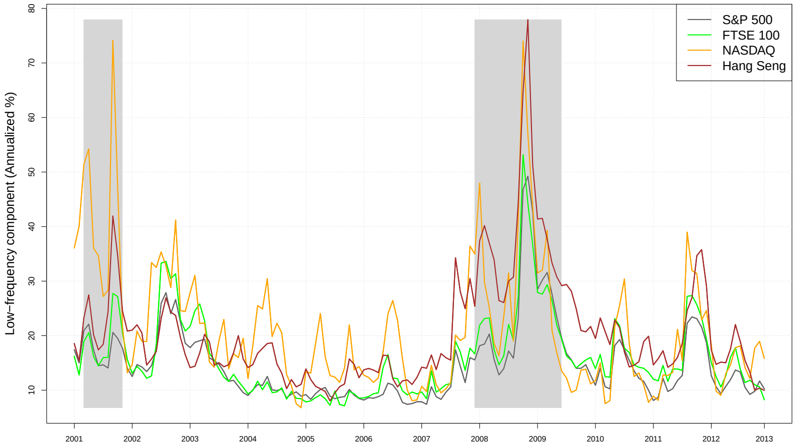
$$\delta_k(\omega) = \frac{(k/K)^{\omega_1-1}(1-k/K)^{\omega_2-1}}{\sum_{j=1}^{K}(j/K)^{\omega_1-1}(1-j/K)^{\omega_2-1}}$$

- GMM inference works as before

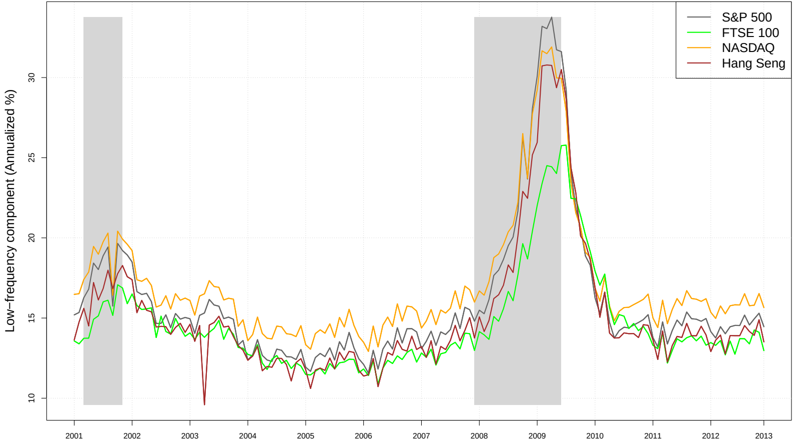
Spline MEM


S&P 500
FTSE 100
NASDAQ
Hang Seng
Low−frequency component (Annualized %)
10
20
30
40
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013

C–MEM


Low−frequency component (Annualized %)
80
70
60
50
40
30
20
10
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
S&P 500
FTSE 100
NASDAQ
Hang Seng

MEM–MIDAS


S&P 500
FTSE 100
NASDAQ
Hang Seng
Low−frequency component (Annualized %)
10
15
20
25
30
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013

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

$$rv_t = \omega + \underbrace{(\alpha_D + \alpha_E rq_{t-1}^{1/2})}_{\alpha_{1,t-1}} rv_{t-1} + \alpha_W \overline{rv}_{t-(2:5)} + \alpha_M \overline{rv}_{t-(6:22)} + u_t$$

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

$$
\begin{aligned}
RV_t &= IV_t + \eta_t \\
&= IV_t + \sqrt{2\Delta} IQ_t^{1/2} z_t \\
&\approx IV_t + \sqrt{2\Delta} \delta IV_t z_t \\
&= IV_t \cdot (1 + \sqrt{2\Delta} \delta z_t) \\
&= IV_t \cdot \varepsilon_t.
\end{aligned}
$$

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 $\rightarrow$ 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 μ_t

- AMEM(/Q/2)

$$\mu_t = \omega + \beta_1 \mu_{t-1} + (\alpha_1 + \alpha_E h_{t-1}) rv_{t-1} + \gamma_1 rv_{t-1}^{(-)}$$
$$\varepsilon_t | \mathcal{I}_{t-1} \sim Gamma(a, 1/a)$$

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

$$\mu_{t,s_t} = \omega_{s_t} + \beta_{s_t} \mu_{t-1,s_{t-1}} + (\alpha_{s_t} + \alpha_E h_{t-1}) rv_{t-1} + \gamma_{s_t} rv_{t-1}^{(-)}$$
$$\varepsilon_t | s_t, \mathcal{I}_{t-1} \sim Gamma(a_{s_t}, 1/a_{s_t})$$

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
  - $h_t = rq_t^{1/2} \rightarrow$ AMEMQ, MS-AMEMQ
  - $h_t = rv_t \rightarrow$ 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

The vector MEM

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

$$\mathbf{x}_t = \boldsymbol{\mu}_t \odot \boldsymbol{\varepsilon}_t$$

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

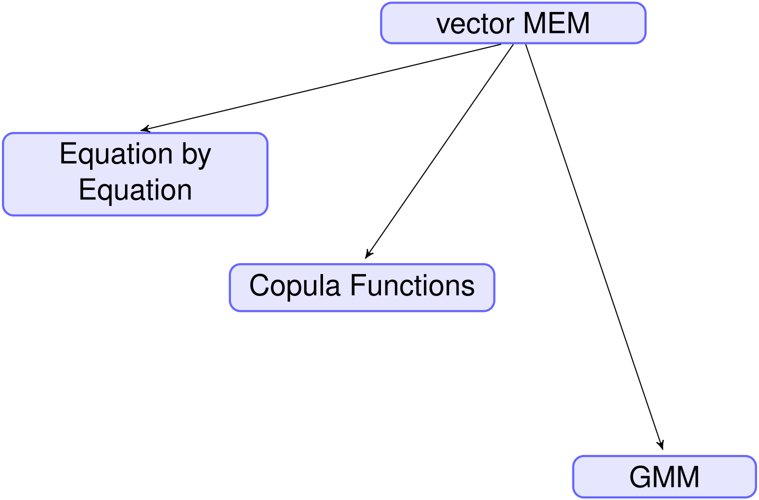
- the components $\mu_{i,t}$ are *predictable* process, depending on a vector of unknown parameters $\boldsymbol{\theta}$,

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

- $\boldsymbol{\varepsilon}_t$ is a *conditionally stochastic* i.i.d. process, with density having non–negative support, mean $\mathbb{1}$ and unknown variance $\boldsymbol{\Sigma}^2$,

$$\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1} \sim D(\mathbb{1}, \boldsymbol{\Sigma}^2).$$

vector MEM by Estimation Method



The vector Multiplicative Error Model

$$\mathbf{x}_t = \boldsymbol{\mu}_t \odot \boldsymbol{\varepsilon}_t = \text{diag}(\boldsymbol{\mu}_t)\boldsymbol{\varepsilon}_t.$$

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

- $\boldsymbol{\mu}_t$ is a $K-$ dimensional vector depending on a vector of parameters $\boldsymbol{\theta}$. Example:

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

  Equation by equation does not work if $\boldsymbol{\beta}$ is not diagonal

- $\boldsymbol{\varepsilon}_t$ is a iid multiplicative error term

$$\boldsymbol{\varepsilon}_t|\mathcal{F}_{t-1} \sim (\mathbb{1}, \boldsymbol{\Sigma})$$

The vector Multiplicative Error Model

From the definition:

$$E(\mathbf{x}_t|\mathcal{F}_{t-1}) = \boldsymbol{\mu}_t$$
$$V(\mathbf{x}_t|\mathcal{F}_{t-1}) = \boldsymbol{\mu}_t\boldsymbol{\mu}_t' \odot \boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\mu}_t)\boldsymbol{\Sigma}\,\text{diag}(\boldsymbol{\mu}_t)$$

- $\boldsymbol{\theta}$ is the parameter of main interest
- $\boldsymbol{\Sigma}$ is a nuisance parameter

For forecasting, considering a second lag in the specification:

$$\boldsymbol{\mu}_{t+\tau} = \boldsymbol{\omega}^* + \mathbf{A}_1\boldsymbol{\mu}_{t+\tau-1} + \mathbf{A}_2\boldsymbol{\mu}_{t+\tau-2},$$

can be solved recursively for any horizon τ.

Impulse Response Analysis

From

$$\mathbf{hl_t} = \boldsymbol{\mu}_\mathbf{t} \odot \boldsymbol{\epsilon}_\mathbf{t} \tag{2}$$

Interpret $\boldsymbol{\mu}_{t+\tau} = E\,(\mathbf{hl_{t+\tau}}|\mathbf{I_t}, \boldsymbol{\epsilon}_\mathbf{t} = \mathbf{1})$and contrast it with $\boldsymbol{\mu}^{(i)}_{t+\tau} = E\left(\mathbf{hl_{t+\tau}}|\mathbf{I_t}, \boldsymbol{\epsilon}_\mathbf{t} = \mathbf{1} + \mathbf{s^{(i)}}\right)$, for a generic vector of shocks $\mathbf{s}^{(i)}$.

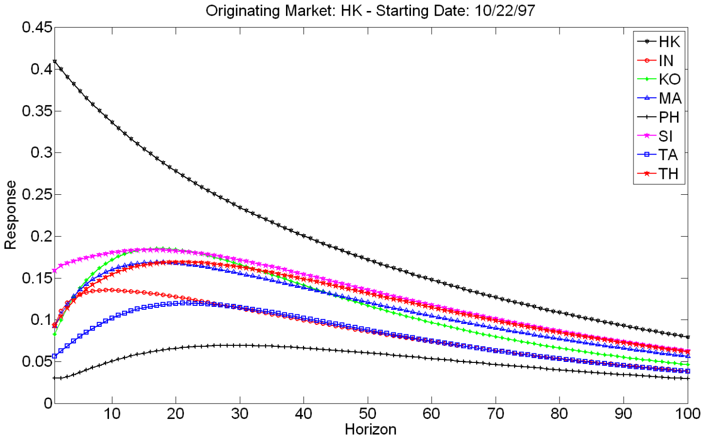
The element–by–element division ($\oslash$) of the two vectors

$$\rho^{(i)}_{t,\tau} = (\boldsymbol{\mu}^{(i)}_{t+\tau} \oslash \boldsymbol{\mu}_{t+\tau}) - \mathbf{1} \quad \tau = 1, \ldots, K \tag{3}$$

gives us the MEM impulse response function to a shock in a market.

NYU

Impulse Response to a shock in Hong Kong



Efficient GMM Inference

Step 1:
Let us define

$$\mathbf{u}_t = \mathbf{x}_t \oslash \boldsymbol{\mu}_t - \mathbb{1} = \varepsilon_t - \mathbb{1}.$$

as an working residual. Hence

$$E(\mathbf{u}_t|\mathcal{F}_{t-1}) = \mathbf{0}$$
$$V(\mathbf{u}_t|\mathcal{F}_{t-1}) = \boldsymbol{\Sigma}$$

so that $\mathbf{u}_t$ is a martingale difference.

Efficient GMM Inference

Step 2:

Let $\mathbf{G}_t$ an instrument, i.e. a (M, K)-matrix

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

Then

$$E(\mathbf{G}_t\mathbf{u}_t|\mathcal{F}_{t-1}) = \mathbf{0} = E(\mathbf{G}_t\mathbf{u}_t)$$

and $\mathbf{g}_t = \mathbf{G}_t\mathbf{u}_t$ also is a martingale difference.

This provides M moment conditions. If $M = p$, we have as many equations as the dimension of θ

Efficient GMM Inference

Step 3:

If $M = p$, we have the MM criterion

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{g}_t = \mathbf{0}$$

where $\boldsymbol{g}_t = \boldsymbol{G}_t \mathbf{u}_t$.

Efficient GMM Inference

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

$$\text{Avar}(\widehat{\theta}_T) = \frac{1}{T}\mathbf{S}^{-1}\mathbf{V}\mathbf{S}^{-1\prime},$$

where

$$\mathbf{S} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} E\left(\nabla_{\theta'}\mathbf{g}_t\right)$$

$$\mathbf{V} = \lim_{T\to\infty} \frac{1}{T} V\left(\sum_{t=1}^{T}\mathbf{g}_t\right)$$

Efficient GMM Inference

Step 4:
Being $\mathbf{g}_t = \mathbf{G}_t \mathbf{u}_t$ a martingale difference leads to a simple formulation for the efficient choice of the instrument $\mathbf{G}_t$

$$\mathbf{G}_t^* = -E(\nabla_\theta \mathbf{u}_t' | \mathcal{F}_{t-1}) V(\mathbf{u}_t | \mathcal{F}_{t-1})^{-1}.$$

Efficient is meant producing the 'smallest' asymptotic variance matrix among the GMM estimators obtained solving the moment conditions.

Efficient GMM Inference

Step 5:
Computing the efficient instrument $\mathbf{G}_t^*$ for the vMEM and plugging it into the moment conditions we obtain

$$\frac{1}{T}\sum_{t=1}^{T}\nabla_{\theta}\boldsymbol{\mu}_t'[\mathrm{diag}(\boldsymbol{\mu}_t)\boldsymbol{\Sigma}\,\mathrm{diag}(\boldsymbol{\mu}_t)]^{-1}(\mathbf{x}_t-\boldsymbol{\mu}_t)=\mathbf{0}$$

together with a (relatively) simple expression of $\mathrm{Avar}(\widehat{\theta}_T)$.

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 ε_t (Engle and Gallo, 2006)
- Main difference of the vector case: it is impossible to remove $\boldsymbol{\Sigma}$ from the moment equation. Hence, it is important to investigate its role in making inference about $\boldsymbol{\theta}$

Inference on Σ

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

$$\widehat{\boldsymbol{\Sigma}}_T = \frac{1}{T}\sum_{t=1}^{T}\mathbf{u}_t\mathbf{u}_t'$$

  where $\mathbf{u}_t = \mathbf{x}_t \oslash \boldsymbol{\mu}_t - \mathbb{1}$ is the working residual computed at current values of $\widehat{\boldsymbol{\theta}}_T$.
- Remark: this estimator is not compromised by zeros in the data.

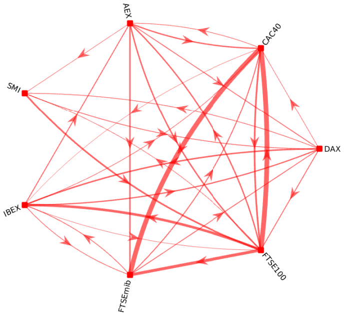
Variable selection: a Lasso approach

- Put all coefficients of μ_t into a vector δ.
- 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 $\hat{w}_j = 1/|\hat{\delta}_j(mle)|^{\xi}$ for some $\xi > 0$.

$$\hat{\delta}(\lambda_T) = \underset{\tilde{\delta}}{\text{argmin}} \left\{ -\frac{1}{T}\ell(\tilde{\delta}) + \lambda_T \sum_{j=1}^{d} \hat{w}_j |\tilde{\delta}_j| \right\}. \qquad (4)$$

- λ_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.
Left: 2010-2012 Right: 2013-2015.

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 τ_t?
- Cascading components – combined multiplicatively
- Horse race with other τ_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!