Variance Targeting for Heavy Tailed Time Series

June 2012

Conference "Advances in Econometrics", Yale SOM

Jonathan Hill*   Eric Renault**

*University of North Carolina at Chapel Hill

** Brown University

OUTLINE

1. Variance targeting in GARCH models
2. Sample variance with tail trimming
3. QMLE with (tail-trimmed) variance targeting = solution to infinite unconditional kurtosis problem
4. Trimming orthogonality conditions for GMM = solution to infinite conditional kurtosis problem
5. Monte Carlo study

1. Variance Targeting in GARCH Models

- *Example 1:* GARCH(1,1)

$$y_{t+1} = \left(h_{t+1}\right)^{1/2} \varepsilon_{t+1}$$

$$h_{t+1} = \omega + \alpha (y_{t+1})^2 + \beta h_t, \alpha > 0, \beta > 0,$$

$$\alpha + \beta < 1 \Rightarrow Var(y_t) = \frac{\omega}{1-\alpha-\beta} = \frac{\omega}{\eta} = \gamma$$

$$\rightarrow new\ parameterization:$$

$$\varphi = (\gamma, \xi), \xi = (\alpha, \eta)$$

Variance Targeting:

- Idea = direct estimation of unconditional variance $\gamma \rightarrow$ common practice:

(i) γ = estimated by sample variance: $\hat{\gamma}_T = \frac{1}{T}\sum_{t=1}^{T} y_t^2$

(ii) (Q)MLE = applied to remaining parameters after plugging in sample variance:

$$h_{t+1} = \omega + \alpha y_{t+1}^2 + \beta h_t = \gamma\eta + \alpha y_{t+1}^2 + (1-\alpha-\eta)h_t$$

$$\rightarrow \hat{\sigma}_{t+1,T}^2(\xi) = \hat{\gamma}_T\eta + \alpha y_{t+1}^2 + (1-\alpha-\eta)\hat{\sigma}_{t,T}^2(\xi),$$

$$\xi = (\alpha, \eta).$$

Cost-Benefit of variance targeting

- Benefits:

(i) Better finite sample performance (especially for estimation of β).

(ii) Robustness to misspecification of variance equation.

- Costs:

(i) Efficiency loss at least if QMLE = MLE

(ii) For asymptotic normality, requires a much more restrictive assumption than QMLE

$$Not\ only\ E\left(\varepsilon_{t+1}^{4}\right) < \infty\ (and\ E\left(h_{t+1}\right) = \gamma < \infty)$$

$$But\ also\ E\left(y_{t+1}^{4}\right) = E\left(h_{t+1}^{2}\varepsilon_{t+1}^{4}\right) < \infty \Leftrightarrow Var\left(h_{t+1}\right) < \infty$$

Does unconditional kurtosis exist?

- Existence of the unconditional fourth moment of stochastic process generating financial return data $\rightarrow$ maintained interest for researchers:
- He and Terasvirta (1999, ET)*"Fourth Moment Structure of the GARCH(p,q) Process"*

R1. Existence "would enable one to see how well the kurtosis and autocorrelation (of squared returns) implied by the estimated model match the estimates obtained directly from the data"

R2. Existence = far from being certain in case of volatility persistence

GARCH (1,1) case: (see He and Terasvirta for general GARCH(p,q))

$$y_{t+1} = \left(h_{t+1}\right)^{1/2} \varepsilon_{t+1}$$

$$E_t(\varepsilon_{t+1}) = 0, E_t(\varepsilon_{t+1})^2 = 1, E_t(\varepsilon_{t+1})^4 = \nu_4,$$

$$h_{t+1} = \omega + \alpha(y_{t+1})^2 + \beta h_t.$$

$$E(y_{t+1})^4 < \infty \Leftrightarrow \nu_4 \alpha^2 + 2\alpha\beta + \beta^2 < 1.$$

Conditionally normal case:

$$E(y_{t+1})^4 < \infty \Leftrightarrow 2\alpha^2 + (\alpha + \beta)^2 < 1.$$

Pareto Tail Index κ

- Basrak, Davis and Mikosch, SPA, 2002

"Regular variation of GARCH processes"

$$y_{t+1} = \left(h_{t+1}\right)^{1/2} \varepsilon_{t+1}$$

$$h_{t+1} = \omega + \alpha (y_{t+1})^2 + \beta h_t, \alpha > 0, \beta > 0,$$

$$P\left[\left|y_t\right| > c\right] \propto c^{-\kappa}, c \to \infty \Leftrightarrow E\left[\left(\alpha \varepsilon_t^2 + \beta\right)^{\kappa/2}\right] = 1$$

$$\alpha + \beta < 1 \Rightarrow \kappa > 2, E\left[\left|y_t\right|^p\right] < \infty \Leftrightarrow \kappa > p$$

$$\alpha + \beta < 1 \Rightarrow Var(y_t) = \frac{\omega}{1 - \alpha - \beta}$$

R3. Hill's tail index estimator for daily log-returns over the period 2001-2011:

90% confidence band:

SP 500 and NASDAQ : [2,3]

DAX: [2, 3.5]

NIKKEI : [1.8, 2.8]

Conclusion:

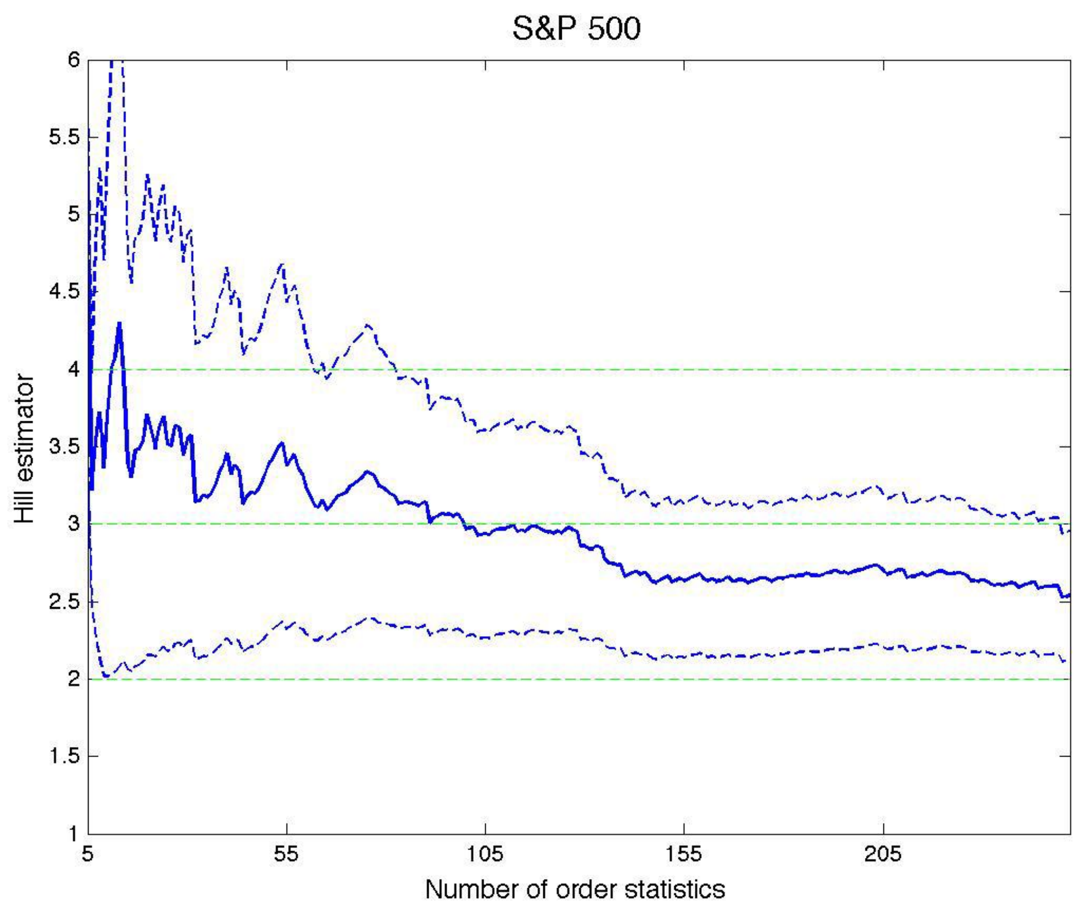
(i) Variance should be finite (not compelling for Nikkei!)

(ii) (Unconditional) moment of order 3 may exist

(iii) Hard to find a series that appears to have finite unconditional moment of order 4

Hill Plots for daily log-returns on SP500

(2513 days from Jan ,1st, 2001 to Jan 1st, 2011)



Multivariate examples:

- *Example 1:* DCC-GARCH(1,1):

Univariate GARCH for each asset + dynamic conditional correlations:

$$\varepsilon_{i,t} = \frac{y_{i,t}}{(h_{ii,t})^{1/2}} \rightarrow conditional\ \ correlation\ \ matrix:$$

$$Q_t = conditional \text{ variance matrix of } \varepsilon_t = (\varepsilon_{i,t})_{1 \le i \le n}$$

$$\rightarrow correlation \text{ targeting } in\ the\ DCC\ equation:$$

$$Q_t = (1 - \alpha - \beta)\overline{Q} + \alpha \varepsilon_t \varepsilon_t' + \beta Q_{t-1}$$

$$\rightarrow no \text{ additional } problem\ under$$

$$\text{maintained } assumption\ E\left(\varepsilon_{i,t}\right)^4 < \infty, \forall i.$$

Variance targeting and parsimony

- DCC example $\rightarrow$ Once the unconditional individual variances and correlation matrix is estimated : only individual GARCH + two more parameters α and β $\rightarrow$ saves N = n(n+1)/2 parameters.
- *Example 2:* VEC-GARCH(1,1)

$$y_t = H_t^{1/2}\varepsilon_t, h_t = vech(H_t), v_t = vech(\varepsilon_t \varepsilon_t')$$

$$h_t = c + A v_{t-1} + G h_{t-1},$$

$$\Rightarrow vech[Var(y_t)] = [Id_N - A - G]^{-1} c.$$

$\rightarrow$ Variance targeting still requires unconditional finite (matricial) kurtosis

2. Sample variance with tail trimming

- Univariate case:

$$\hat{\gamma}_T = \frac{1}{T}\sum_{t=1}^{T} y_t^2 \xrightarrow{T=\infty} \gamma^0 = Var(y_t) < \infty$$

$$BUT: \sqrt{T}\left(\hat{\gamma}_T - \gamma^0\right) \neq asymptotically\ normal\ if\ E(y_t^4) = \infty.$$

Key idea: Tail-Trimming of returns before computing sample variance

$$\hat{I}_{T,t}^{(y)} = 0 \Leftrightarrow \left|y_t\right|\ is\ one\ of\ the\ k_T^{(y)}\ \text{largest}$$

$$observations\ among\ \left|y_\tau\right|, \tau = 1,\ldots,T.$$

$$\hat{I}_{T,t}^{(y)} = 1\, otherwise, \hat{\gamma}_T^{(tr)} = \frac{1}{T}\sum_{t=1}^{T} y_t^2 \hat{I}_{T,t}^{(y)}$$

$$k_T^{(y)} \to \infty \text{ promotes Gaussian asymptotics}$$

$$k_T^{(y)}/T \to 0 \text{ ensures unbiasedness}: \hat{\gamma}_T^{(tr)} \to \gamma^0 = Var(y_t)$$

Asymptotic distribution of estimator of unconditional variance in GARCH:

1st case: Without trimming (with finite kurtosis):

Horvath, Kokoszka,Zitikis, JFEC,2006:

$$\hat{\gamma}_T = \frac{1}{T}\sum_{t=1}^{T} y_t^2 = \frac{1}{T}\sum_{t=1}^{T} h_t(\varepsilon_t^2 - 1) + \frac{1}{T}\sum_{t=1}^{T} h_t$$

$$= \gamma^0 + \frac{1-\beta^0}{\eta^0}\frac{1}{T}\sum_{t=1}^{T} h_t(\varepsilon_t^2 - 1) + o_P(1/\sqrt{T})$$

$$\Rightarrow v_T^2 = Var\left(T\hat{\gamma}_T\right) = Var\left[\sum_{t=1}^{T} y_t^2\right] \approx T\left(\frac{1-\beta^0}{\eta^0}\right)^2 E\left(h_t^2\right)Var\left(\varepsilon_t^2\right)$$

2nd case: With trimming:

$$v_T^2 = Var\left(T\hat{\gamma}_T^{(tr)}\right) = Var\left[\sum_{t=1}^{T} y_t^2 I\left(\left|y_t\right| < c_T^{(y)}\right)\right]$$

$$Lim_{T=\infty} \frac{v_T^2}{T} = \left(\frac{1-\beta^0}{\eta^0}\right)^2 E\left(h_t^2\right) Var\left(\varepsilon_t^2\right) if\ E\left(y_t^4\right) < \infty$$

$$Lim_{T=\infty} \frac{v_T^2}{T} = \infty\ if\ E\left(y_t^4\right) = \infty, BUT\ v_T^2 = o(T^2)$$

→Same asymptotic distribution if finite kurtosis

→ More involved in the general case because, by trimming, we lose the mds property:

Pareto tails:

$$P\left[y_t^2 > \left(c_T^{(y)}\right)^2\right] = \frac{k_T^{(y)}}{T} \approx d(c_T^{(y)})^{-\kappa}, 2 < \kappa \le 4$$

$$c_T^{(y)} \approx (d)^{-1/\kappa}\left[\frac{T}{k_T^{(y)}}\right]^{1/\kappa} \xrightarrow{T=\infty} \infty \; as \; \frac{k_T^{(y)}}{T} \to 0$$

Feller (1971) (regularly varying functions):

$$E\left[y_t^4 \hat{I}_{T,t}^{(y)}\right] = (c_T^{(y)})^{4-\kappa} L\left(c_T^{(y)}\right), L(x) = o(x^{\varepsilon}), \forall \varepsilon > 0$$

Corollary 1:

$$\kappa > 2 \text{ and } k_T^{(y)} \to \infty \Rightarrow E\left[y_t^4 \hat{I}_{T,t}^{(y)}\right] = o(T)$$

Corollary 2 (with geometric β-mixing):

Long term variance matrix = o(T)

$$v_T^2 = Var\left(\sum_{t=1}^{T} y_t^2 \hat{I}_{T,t}^{(y)}\right) = o(T^2), v_T = o(T).$$

$$BUT : E(y_t^4) = \infty \Rightarrow v_T / \sqrt{T} \to \infty$$

$$E(y_t^4) < \infty \Rightarrow \left(v_T / \sqrt{T}\right)^2 \to b\left[E(\varepsilon_t)^4 - 1\right]$$

$$b = \left(\frac{1-\beta}{1-\alpha-\beta}\right)^2 E(h_t^4)$$

R. Geometric β-mixing for y = implied by α+β<1 when ε has an absolutely continuous distribution

Asymptotic distribution of tail-trimmed estimators of unconditional variance:

Long-term variance matrix = o(T)

$\rightarrow$ Promotes asymptotic normality of tail trimmed sample variance:

$$\frac{T}{v_{T,t}}\left(\hat{\gamma}_T^{(tr)} - \gamma^0\right) \xrightarrow{d} \aleph(0,1)$$

$$\text{insofar as}: \frac{\mathrm{T}}{\mathrm{v}_{\mathrm{T,t}}}\left(E\left[y_t^2 \hat{I}_T^{(y)}\right] - \gamma^0\right) \rightarrow 0.$$

$$\textit{Always true}\ \text{if}: \sqrt{T}\left(E\left[y_t^2 \hat{I}_T^{(y)}\right] - \gamma^0\right) \rightarrow 0.$$

$\Rightarrow$ Lighter trimming for heavier tail (feasible from estimation of tail index)

1st improvement of tail-trimmed sample variance

- Peng (2001) : "Estimating the mean of a heavy tailed distribution", Statistics & Probability Letters

$\rightarrow$ Characterizes the systematic finite sample bias due to trimming:

$$\hat{\gamma}_T^{(tr)} = \frac{1}{T}\sum_{t=1}^{T} y_t^2 \hat{I}_{T,t}^{(y)} \text{ biased by :}$$

$$E\left[y_t^2\left(1-\hat{I}_{T,t}^{(y)}\right)\right] \approx \frac{\kappa}{\kappa-2}\frac{k_T^{(y)}}{T}\left(c_T^{(y)}\right)^2$$

$$P\left[\left|y_t\right| > c_T^{(y)}\right] = \frac{k_T^{(y)}}{T}$$

This bias can be estimated by:

$$\hat{R}_T^* = \frac{\hat{\kappa}}{\hat{\kappa}-2} \frac{k_T^{(y)}}{T} \left(y_{k_T^{(y)}}^{(a)} \right)^2$$

$\hat{\kappa} = Hill\ estimator\ based\ on\ k_T^{(y)}$ largest *absolute values*.

$$\hat{\kappa} = \left[\frac{1}{k_T^{(y)}} \sum_{i=1}^{k_T^{(y)}-1} Log\left(\frac{y_i^{(a)}}{y_{k_T^{(y)}}^{(a)}} \right) \right]^{-1}$$

$y_{k_T^{(y)}}^{(a)} = component\ k_T^{(y)} in\ order\ statistics\ of\ absolute\ values:$

$y_1^{(a)} \geq y_2^{(a)} \geq .... \geq y_T^{(a)}.$

$\rightarrow$ Improved estimator of conditional variance based on Peng (2001):

$$\hat{\gamma}_T^* = \hat{\gamma}_T^{(tr)} + \hat{R}_T^*$$

Pros and cons of Peng's improvement

- Pro 1:

Avoids a systematic finite sample under-estimation of unconditional variance

- Pro 2:

Provides asymptotic normality without the problematic maintained assumption:

$$\frac{T}{v_T}\left(E\left[y_t^2 \hat{I}_T^{(y)}\right] - \gamma^0\right) \rightarrow 0.$$

Con: The correction term is not asymptotically independent from the main term

$$\hat{\gamma}_T^* = \hat{\gamma}_T^{(tr)} + \hat{R}_T^*$$

$$\neq \textit{asymptotic distribution of } \hat{\gamma}_T^{(tr)}$$

$\rightarrow$ We propose a new estimator: by under-correction for the gap, we obtain an asymptotically normal estimator with the same distribution limit as the initial one:

$$\hat{R}_T^* = \frac{\hat{\kappa}}{\hat{\kappa}-2} \frac{\tilde{k}_T^{(y)}}{T} \left(y^{(a)}_{\tilde{k}_T^{(y)}} \right)^2$$

$$\tilde{k}_T^{(y)} / k_T^{(y)} \rightarrow \infty \Rightarrow y^{(a)}_{\tilde{k}_T^{(y)}} / y^{(a)}_{k_T^{(y)}} \xrightarrow{T=\infty} 0$$

$\hat{\kappa} = \textit{Hill estimator based on}$ even more $\breve{k}_T^{(y)}$ largest *absolute values.*

Price to pay for this simpler bias-corrected estimator:

$\rightarrow$Need to maintain the assumption:

$$\frac{\mathrm{T}}{\mathrm{v_T}} E\left[y_t^2\left(\hat{I}_T^{(y)} - \tilde{I}_T^{(y)}\right)\right] \rightarrow 0.$$

$\rightarrow$ we will work under this maintained assumption , or

$\rightarrow$ No-bias correction + the corresponding stronger assumption:

$$\frac{\mathrm{T}}{\mathrm{v_T}} \left(E\left[y_t^2 \hat{I}_T^{(y)}\right] - \gamma^0\right) \rightarrow 0.$$

3. QMLE with (tail-trimmed) variance targeting

$$y_{t+1} = \left(h_{t+1}\right)^{1/2} \varepsilon_{t+1}, h_{t+1} = \omega + \alpha (y_{t+1})^2 + \beta h_t .$$

$$\eta = 1 - \alpha - \beta, \gamma = \omega / \eta, \xi = (\alpha, \eta)', \varphi = (\gamma, \xi')'$$

$$\frac{T}{v_T} \Omega_T^{-1/2} \left(\hat{\varphi}_T - \varphi^0\right) \xrightarrow{d} \aleph(0, Id_3)$$

$$NW \text{ of } \Omega_T = 1$$

$$SW \text{ of } \Omega_T = -J^{-1} K$$

$$SE \text{ of } \Omega_T = \left(\frac{E(\varepsilon_t^4) - 1}{v_T^2 / T} \right) J^{-1} + J^{-1} K K' J^{-1}$$

$$J = E\left[\frac{1}{h_t^2}\frac{\partial h_t}{\partial \xi}\frac{\partial h_t}{\partial \xi'}\right], K = E\left[\frac{1}{h_t^2}\frac{\partial h_t}{\partial \xi}\frac{\partial h_t}{\partial \gamma}\right]$$

R1. Trimming has no impact on asymptotic distribution of QML-VT when finite kurtosis (Francq,Horvath,Zakoian,JFEC, 2011)

R2. In case of infinite kurtosis, the asymptotic variance of GARCH parameters (α,η) (SE block) is smaller:

$$\Omega_T \approx \begin{bmatrix} 1 \\ -J^{-1}K \end{bmatrix} \begin{bmatrix} 1 & -K'J^{-1} \end{bmatrix} = \Omega^*$$

R3. Price to pay for variance targeting = now in terms of rate of convergence

BUT Two directions in the parameter space with root-T rate of convergence:

$$\forall \lambda \in \Re^2, \zeta(\lambda) = \begin{bmatrix} \lambda' J^{-1} K \\ \lambda \end{bmatrix} \Rightarrow \zeta(\lambda)' \Omega^* \zeta(\lambda) = 0$$

$$\Rightarrow \frac{T}{v_T} \zeta(\lambda)' \left(\hat{\varphi}_T - \varphi^0 \right) \to 0$$

$$\Rightarrow \sqrt{T} \zeta(\lambda)' \left(\hat{\varphi}_T - \varphi^0 \right) \to \aleph\left(0, \left(E(\varepsilon_t^4) - 1\right) \lambda' J^{-1} \lambda\right)$$

$$\lambda = (1,0)' \Rightarrow \zeta(\lambda)' \varphi = \left(J^{-1} K\right)_1 \gamma + \alpha$$

$$\lambda = (1,1)' \Rightarrow \zeta(\lambda)' \varphi = \left[\left(J^{-1} K\right)_1 + \left(J^{-1} K\right)_2\right] \gamma + 1 - \beta$$

4. Trimming orthogonality conditions for GMM

QML score (with re-parameterization for variance targeting):

$$\left(e_t^2(\varphi)-1\right)s_t(\varphi)=\left(\frac{y_t^2}{h_t(\varphi)}-1\right)\frac{1}{h_t(\varphi)}\frac{\partial h_t(\varphi)}{\partial \xi}$$

→ May need tail-trimming if $E(\varepsilon_t^4)=\infty$
since then Gaussian QML (Hall and Yao, 2003) is not asymptotically normal and has a slow rate of convergence

R1. General results of Hill and Renault (2010, GMM with Tail Trimming) : shows that a well-suited tail trimming of the estimating equations will ensure asymptotic normality with a faster rate of convergence (albeit smaller than root T)

$\rightarrow$ Asymptotic distribution with standard formulas (but Jacobian and Variance matrices with trimmed estimating equations)$\rightarrow$ rate of convergence slower than root-T

R2. Tail trimming of estimating equations $\Rightarrow$ No reason anymore to consider just identified moment conditions (QML) $\Rightarrow$ more lags of the variance score $s(\varphi)$ for more estimating equations

General notations:

- Tail trimming of moment conditions

$$\hat{m}_{t,T}(\theta) = \left(\hat{m}_{i,t,T}(\theta)\right)_{1 \le i \le H}$$

$$\hat{m}_{i,t,T}(\theta) = m_{it}(\theta)\hat{I}_{i,t,T}(\theta), \hat{I}_{i,t,T}(\theta) \in \{0,1\}$$

$$\hat{I}_{i,t,T}(\theta) = 0 \Leftrightarrow \left|m_{it}(\theta)\right| \textit{ is one of the } k_{iT} \text{ largest}$$

$$\textit{observations among } \left|m_{i\tau}(\theta)\right|, \tau = 1,\ldots,T.$$

$$k_{iT} \to \infty, k_{iT} / T \to 0$$

$$(\neq \textit{ fixed quantile } \text{trimming}, k_{iT} = cT)$$

Trimming and local identification

$\rightarrow$Jacobian of moment conditions J(θ):

R1. Problem of non-existence of expected Jacobian matrix related to non-existence of moments :

$$m_t(\theta) = (y_t - \theta y_{t-1}) y_{t-1} \Rightarrow \frac{\partial}{\partial \theta} m_t(\theta) = -y_{t-1}^2$$

$$\Rightarrow J_T(\theta^0) \; and \; \mathrm{S}_T(\theta^0) \; may \; both \; diverge \; when \; T \rightarrow \infty$$

R2. Standard assumption of full-column rank with asymptotic non-degeneracy (after trimming).

Jacobian formula:

- Thanks to asymptotically vanishing trimming:

$$J_T(\theta^0) = \frac{\partial}{\partial \theta'} E[m^*_{t,T}(\theta)]_{\theta=\theta^0} \approx E\left[\left(\frac{\partial}{\partial \theta'} m_{i,t}(\theta)\right)_{\theta=\theta^0} I^*_{i,t,T}(\theta^0)\right]_{1\leq i\leq H}$$

- While the perverse term:

$$\frac{\partial}{\partial \theta'} E\left[m_{i,t}(\theta^0) I^*_{i,t,T}(\theta)\right]_{\theta=\theta^0}$$

Is asymptotically negligible (would not work with fixed quantile trimming)

Linearization of FOC:

$$J_T(\theta^0)'\Omega\sqrt{T}m_T^*(\theta^0)+$$

$$J_T(\theta^0)'\Omega J_T(\theta^0)\sqrt{T}(\hat{\theta}_T-\theta^0)=o_P(1)$$

R. In Cizek (2009) “Generalized Method of Trimmed moments”, W.P. Tilburg

→Fixed Quantile Trimming

→The “perverse” term is no longer negligible in the linearization

→Must deal with an asymmetric matrix J’ΩJ*instead of J’ΩJ (no perverse term in finite sample left Jacobian term)

Assuming:

$$J_T(\theta^0) = asymptotically\ \ full\ column\ rank$$

$$V_T(\theta) =$$

$$T[J_T^{'}(\theta)\Omega J_T(\theta)][J_T^{'}(\theta)\Omega S_T(\theta)\Omega J_T(\theta)]^{-1}[J_T^{'}(\theta)\Omega J_T(\theta)]$$

$$V_T^{1/2}(\theta^0)[\hat{\theta}_T - \theta^0] \xrightarrow{d} N(0, Id_p)$$

An "efficient" weighting matrix:

$$\Omega_T = [S_T(\theta^0)]^{-1}$$

$$T^{1/2}[J_T^{'}(\theta^0)S_T^{-1}(\theta^0)J_T(\theta^0)]^{1/2}(\hat{\theta}_T - \theta^0)$$

$$\xrightarrow{d} N(0, Id_p)$$

R1. Efficient weighting matrix = exists thanks to symmetric occurrence of Jacobian matrices because perverse terms are negligible (would not work with fixed quantile trimming, Cizek, 2009)

R2.Rate of convergence may depend upon the choice of trimming fractiles $\rightarrow$ questions the concept of "efficiency"

R3. We recover root-T asymptotic normality (with efficient asymptotic variance)if the matrices $\Sigma(T)$ and $J(T)$ have well-defined limits $\Rightarrow$ Tail trimming is always a safe practice .

R4. Asymptotic normality $\Rightarrow$ allows standard inference like J-test, Wald, with standard formulas in spite of non-standard rates.
(see Antoine and Renault (2009)).

R5. Fractile selection for sufficiently fast identification:

1st solution (Horowitz):

Compare trimmed GMM and naïve estimator with identity weighting matrix (averaged over subsamples)

2nd solution:

Iterative estimation of the tail indices

R. HR(2010) : trim estimating equations in a general environment

BUT here: We can focus the trimming on the source of the extreme

$\rightarrow$ Different trimming for the standardized innovation $e(\varphi)$ and the variance score $s(\varphi)$

$\rightarrow$ better control for artificially introduced bias due to trimming and even more importantly maintains (asymptotically) the martingale difference structure (no need of HAC estimators for GMM weighting matrix) by re-centering the tail-trimmed innovation.

Trimmed estimating equation:

$$\left(\hat{e}^{*}_{T,t}(\xi) - \frac{1}{T}\sum_{\tau=1}^{T}\hat{e}^{*}_{T,\tau}(\xi)\right)\hat{s}^{*}_{T,t-j}(\xi)$$

$$\hat{e}^{*}_{T,t}(\xi) = e_t(\xi)\hat{I}^{(e)}_{T,t}$$

$$\hat{s}^{*}_{T,t-j}(\xi) = s_{t-j}(\xi)\hat{I}^{(s)}_{T,t-j}$$

$\rightarrow$ Asymptotic theory of GMTTM (HR,2010) while taking into account that it is a two step GMM but no need of HAC estimation (while in HR 2010, perverse serial correlation due to trimming)

Additional simplification:

(Francq and Zakoian, 2004)

$$If\ \ s_{T,t-j}(\xi) = lagged\ value\ of$$

$$\frac{1}{h_t(\varphi)}\frac{\partial h_t(\varphi)}{\partial \xi} = uniformly\ square\ \text{integrable}$$

$$no\ need\ of\ \ \text{trimming when}\ Var(y_t) < \infty$$

$$\rightarrow GARCH\ \ polynomial\ in\ \text{denominator}$$

$$promotes\ \ finite\ \text{variance}$$

5. Monte Carlo study

4.1. 1st case: QML-GARCH without variance targeting (and global trimming of moment condition):

1000 samples of size T=1000

→Strong impact of a few trimmed observations

(breakdown point: Sakata and White (1998))

$$k_T = [T^{\lambda}],$$
$$\lambda \in \{.01,.02,.....,.99\}$$

•Tail trimming always delivers an approximately normal estimator

•In the cases of substantially heavy tails, standard GMM and QMLE strongly fails the KS normality test.

• Only a few tail observations need to be trimmed to ensure approximate normality

Example: GARCH with Pareto errors (κ=2.5 for errors$\rightarrow$ κ=1.5 for y):

•Select KS (test for normality) minimizing λ

$$k_T = [1000^{.35}] = 11$$

- Always safe to trim even when the variance is not infinite.
- Asymmetric trimming for asymmetrically distributed equations is always optimal, where less observations are trimmed from the heavier tail.
- Example: TARCH model :

$$(\lambda_1, \lambda_2) = (.26, .11) \Rightarrow k_{1T} = 6, k_{2T} = 2$$

- Two-step GMTTME with a QML plug-in dominates one-step GMTTME (and two-step GMTTME with a GMTTME plug-in

2^{nd} case: Variance Targeting

- Three distributions for standardized innovations:

N(0,1), Pareto with κ = 4.1 and 2.5

.Two GARCH(1,1) models:

(ω,α,β) = (1,0.3,0.4) and (1,0.3,0.6)

$\Rightarrow$6 different DGPs:

2 with infinite kurtosis for ε

4 with infinite kurtosis for y

$\rightarrow$ Focus on inference on β (Normality, MSE, Tests)

Main Monte-Carlo results

1. VT = improves upon QML when y has finite kurtosis but large β.
2. TTVT = needed to restore normality when y has infinite kurtosis (and no cost in terms of MSE)
3. G-TTVT = needed to restore normality when ε has infinite kurtosis (does not hurt if ε has finite kurtosis)
4. Peng’s bias correction matters (no significant need of under-correction in finite sample)

CONCLUSION

Variance targeting in GARCH-QML estimation:

(i) Cost in terms of efficiency as already documented by Francq, Horvath and Zakoian,2011, in the finite kurtosis case.

(ii) Advantages for finite sample performance and robustness to misspecification

(iii) Inference based on asymptotic normality = no longer valid if infinite kurtosis of asset returns.

(iv) Tail trimming = hedge against infinite kurtosis (unconditional and/or conditional)