The HEAVY GAS Skew-t Noncentral-F Model

for Realized Covariance Matrices

Michael Stollenwerk,

joint work with Taras Bodnar, Yarema Okhrin, Nestor Parolya.

Integrated Covariance - The Variable of Interest

Prices are assumed to be a semimartingale

$$\mathbf{p}_s = \boldsymbol{\mu}_s ds + \mathbf{A}_s d\mathbf{W}_s, \tag{1}$$

where $\mathbf{p}_s$ $(k \times 1)$ is the vector of log prices at time s, $\boldsymbol{\mu}_s$ is the instantaneous drift, $\boldsymbol{\Sigma}_s = \mathbf{A}_s\mathbf{A}_s'$ is the instantaneous covariance of returns, and $\mathbf{W}_s$ is a standard k-variate Wiener process.

We are interested in the *integrated covariance* over [day] t,

$$\mathbf{ICov_t} = \int_t \boldsymbol{\Sigma}_s ds. \tag{2}$$

Variables Used and Objective

We use

1. the simplest estimator of integrated covariance, the realized covariance matrix

$$\underset{(k\times k)}{\mathbf{RC}_t} = \sum_{j=1}^{m} \mathbf{r}_{j,t}\mathbf{r}'_{j,t}, \qquad (3)$$

   where $\mathbf{r}_{j,t}$ is the j'th intraday return vector on day t and

2. the vector of period t returns $\underset{(k\times 1)}{\mathbf{r}_t}$,

in a model with the aim of dynamically modelling and forecasting integrated covariance.

High-frEquency-bAsed VolatilitY (HEAVY) Models[1,2]

Heavy models are made up of the system

$$\begin{Bmatrix} \text{Cov}(\mathbf{r}_t|\mathcal{F}_{t-1}) \\ \mathbb{E}[\mathbf{RC}_t|\mathcal{F}_{t-1}] \end{Bmatrix}, \quad t = 2, \dots, T$$

under the assumption that $\mathbf{r}_t$ and $\mathbf{RC}_t$ are independent conditional on the joint dynamic parameter vector $\boldsymbol{\theta}_t$ and on $\mathcal{F}_{t-1}$, such that the period-t log-likelihood contribution obtains as

$$\mathcal{L}_t(\mathbf{r}_t, \mathbf{RC}_t|\boldsymbol{\theta}_t, \mathcal{F}_{t-1}; \boldsymbol{\beta}, \boldsymbol{\gamma}) = \log f_{\mathbf{r}}(\mathbf{r}_t|\boldsymbol{\theta}_t, \mathcal{F}_{t-1}; \boldsymbol{\beta}) + \log f_{\mathbf{RC}}(\mathbf{RC}_t|\boldsymbol{\theta}_t, \mathcal{F}_{t-1}; \boldsymbol{\gamma}), \tag{4}$$

where $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ represent the static parameters.

[1] Shephard and Sheppard (JAE, 2010)
[2] Noureldin, Shephard and Sheppard (JAE, 2012)

Generalized Autoregressive Score (GAS) Models[1]

A class of observation-driven time series models. The mechanism to update the dynamic parameters over time is the scaled score of the likelihood function, w.r.t. these parameters, $\mathbf{S}_t^{\boldsymbol{\theta}}$.

For example, if we assume that $\mathbf{RC}_t|\mathcal{F}_{t-1} \overset{iid}{\sim} \mathit{Wishart}(\boldsymbol{\Sigma}_t/\nu, \nu)$, then the updating equation for $\boldsymbol{\Sigma}_t$ could be

$$\boldsymbol{\Sigma}_t = \mathbf{A}\mathbf{A}' + \mathbf{B}\,\boldsymbol{\Sigma}_{t-1}\,\mathbf{B}' + \mathbf{C}\,\mathbf{S}_{t-1}^{\boldsymbol{\Sigma}}\,\mathbf{C}, \tag{5}$$

where **A-C** are full k by k matrices and $\mathbf{S}_t^{\boldsymbol{\Sigma}} = \mathbf{D}_t \nabla_t \mathbf{D}_t$, with $\nabla_t = \frac{\partial \log(pdf_{Wishart})}{\partial \boldsymbol{\Sigma}_t}$ and the (time varying) scale matrix $\mathbf{D}_t$.

[1] Creal, Koopman and Lucas (JAE, 2013)

Generalized Autoregressive Score (GAS) Models[1]

- The score defines a steepest ascent direction for improving the model's local fit in terms of the likelihood at time t given the current position of the parameters. This provides the natural direction for updating the parameter.
- The score depends on the complete density, and not only on the first- or second-order moments of the observations, which distinguishes the GAS framework from most of the other observation-driven approaches in the literature.
- By exploiting the full density structure, the GAS model introduces new transformations of the data that can be used to update the time-varying parameters.

[1] Creal, Koopman and Lucas (JAE, 2013)

Existing HEAVY GAS Models

Gorgi, Hansen, Janus and Koopman (JFEconometrics, 2019):

$$\mathbf{r}_t \sim \mathbf{N}(\mathbf{0}, \mathbf{\Lambda}\mathbf{\Sigma}_t\mathbf{\Lambda}), \tag{6}$$

$$\mathbf{RC}_t \sim \mathbf{W}(\mathbf{\Sigma}_t, \nu_1) \tag{7}$$

$$\mathbf{\Sigma}_t = \mathbf{A}\mathbf{A}' + \mathbf{B}\,\mathbf{\Sigma}_{t-1}\,\mathbf{B}' + \mathbf{C}\,\mathbf{S}^{\mathbf{\Sigma}}_{t-1}\,\mathbf{C}', \tag{8}$$

where $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are full k by k matrices, $\mathbf{\Lambda}$ is a diagonal matrix, $\mathbf{\Sigma}$ is a symmetric p.d. parameter matrix, ν_1 is a scalar parameter.

Opschoor, Janus, Lucas and Van Dijk (JBES, 2018):

$$\mathbf{r}_t \sim \mathbf{t}(\mathbf{0}, \mathbf{\Sigma}_t, \nu_0), \tag{9}$$

$$\mathbf{RC}_t \sim \textbf{matrix-F}(\mathbf{\Sigma}_t, \nu_1, \nu_2) \tag{10}$$

$$\mathbf{\Sigma}_t = \mathbf{A}\mathbf{A}' + b\mathbf{\Sigma}_{t-1} + c\mathbf{S}^{\mathbf{\Sigma}}_{t-1}, \tag{11}$$

where a and b are scalars and ν_0 and ν_2 are scalar parameters. The t-distribution accounts for the fat tails in financial returns. Realized covariance matrices also exhibit fat-tails (much probability mass on "large" covariance matrices), which the matrix-F can account for.

The HEAVY GAS skew-t noncentral-F Model

$$\mathbf{r}_t \sim \textbf{skew-t}(\mathbf{0}, \boldsymbol{\Lambda}\boldsymbol{\Sigma}_t\boldsymbol{\Lambda}, \boldsymbol{\alpha}_t, \nu_0), \tag{12}$$

$$\mathbf{RC}_t \sim \textbf{noncentral-F}(\boldsymbol{\Sigma}_t, \boldsymbol{\Omega}_t, \nu_1, \nu_2) \tag{13}$$

$$\boldsymbol{\Sigma}_t = \mathbf{A}\mathbf{A}' + \mathbf{B}\,\boldsymbol{\Sigma}_{t-1}\,\mathbf{B}' + \mathbf{C}\,\mathbf{S}^{\boldsymbol{\Sigma}}_{t-1}\,\mathbf{C}' \tag{14}$$

$$\boldsymbol{\Omega}_t = \mathbf{D}\mathbf{D}' + \mathbf{E}\,\boldsymbol{\Omega}_{t-1}\,\mathbf{E}' + \mathbf{F}\,\mathbf{S}^{\boldsymbol{\Omega}}_{t-1}\,\mathbf{F}' \tag{15}$$

$$\boldsymbol{\alpha}_t = \mathbf{g} + \mathbf{H}\,\boldsymbol{\alpha}_{t-1} + \mathbf{J}\,\mathbf{S}^{\boldsymbol{\alpha}}_{t-1}, \tag{16}$$

Nests Grigori et al., Opschoor et al. (with appropriate scaling of the score).

Multivariate Skew-t Distribution

Motivated by the stylized fact that large negative returns are more likely than large positive ones, i.e. skewness.

Probability density function:

$$f_{\mathbf{r}}(\mathbf{r}_t|\boldsymbol{\Sigma}_t,\boldsymbol{\alpha}_t,\nu_0;\mathcal{F}_{t-1}) = 2f_{T_{\nu_0}}(\mathbf{r}_t|\boldsymbol{\Sigma}_t;\mathcal{F}_{t-1})F_{T_{\nu_0+p}}\left(\frac{\boldsymbol{\alpha}_t{}^\top\mathbf{r}_t}{(\nu_0-2+\mathbf{r}_t{}^\top\boldsymbol{\Sigma}_t{}^{-1}\mathbf{r}_t)^{1/2}}\sqrt{\nu_0+p}\right), \quad (17)$$

where the $F_{T_{\nu_0}}(\cdot)$ denotes the c.d.f. of the standard univariate t-distribution with ν_0 degrees of freedom and $f_{T_{\nu_0}}(\cdot)$ is density of the (standardized) multivariate central t distribution with ν_0 degrees of freedom. The vector $\boldsymbol{\alpha}$ contains the skewness parameters.

$$\operatorname{Cov}(\mathbf{r}_t) = \frac{\nu_0}{\nu_0-2}\boldsymbol{\Sigma}_t - \frac{\nu_0}{\pi}\left(\frac{\Gamma\left(\frac{\nu_0-1}{2}\right)}{\Gamma\left(\frac{\nu_0}{2}\right)}\right)^2\frac{\boldsymbol{\Sigma}_t\boldsymbol{\alpha}_t\boldsymbol{\alpha}_t{}^\top\boldsymbol{\Sigma}_t}{1+\boldsymbol{\alpha}_t{}^\top\boldsymbol{\Sigma}_t\boldsymbol{\alpha}_t}, \quad \nu_0>2\,. \quad (18)$$

Matrix-Variate Noncentral F Distribution

Probability density function:

$$f_{RC}(RC_t|\boldsymbol{\Sigma}_t,\boldsymbol{\Omega}_t,\mathcal{F}_{t-1};\nu_1,\nu_2)$$
$$= \frac{1}{B_p\left(\frac{\nu_1}{2},\frac{\nu_2}{2}\right)}\left|\frac{\nu_1}{\nu_2-p-1}\boldsymbol{\Sigma}_t{}^{-1}\right|^{\nu_1/2}\frac{|RC_t|^{(\nu_1-p-1)/2}}{\left|\mathbf{I}+\frac{\nu_1}{\nu_2-p-1}\boldsymbol{\Sigma}_t{}^{-1}RC_t\right|^{(\nu_1+\nu_2)/2}}e^{\mathrm{tr}\left(-\frac{1}{2}\boldsymbol{\Sigma}_t{}^{-1}\boldsymbol{\Omega}_t\right)}$$
$$\times\, {}_1F_1\left(\frac{\nu_1+\nu_2}{2};\frac{\nu_1}{2};\frac{1}{2}\frac{\nu_1}{\nu_2-p-1}\boldsymbol{\Sigma}_t{}^{-1}\boldsymbol{\Omega}_t\boldsymbol{\Sigma}_t{}^{-1}RC_t[\mathbf{I}+\frac{\nu_1}{\nu_2-p-1}\boldsymbol{\Sigma}_t{}^{-1}RC_t]^{-1}\right) \tag{19}$$

where $B_p\left(a,b\right)$ denotes the multivariate Beta function and ${}_1F_1(a;b;\mathbf{C})$ is the matrix variate hypergeometric function.

$$\mathbf{E}[\mathbf{RC}_t]=\boldsymbol{\Sigma}_t+\frac{1}{\nu_1}\boldsymbol{\Omega}_t. \tag{20}$$

Matrix-Variate Noncentral F Distribution

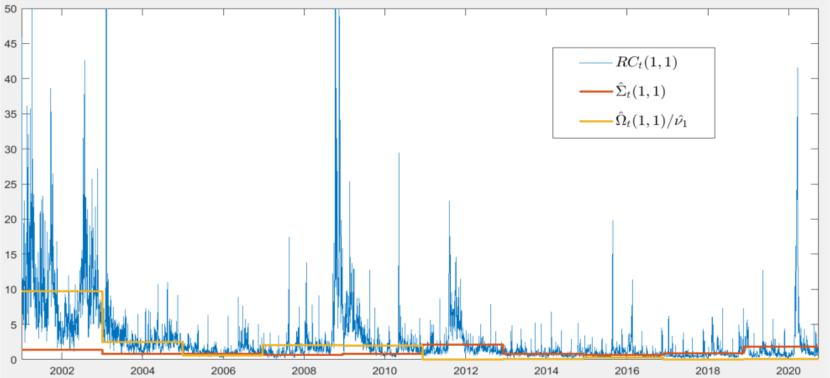

Figure 1: Static estimation of noncentral matrix-F distribution on subperiods of 500 datapoints each. $\hat{\mathbf{\Omega}}$ varies substatially over time, motivating our choice of the *noncentral* matrix-F. It seems that, the more volatile the period, the more weight rests on the $\mathbf{\Omega}/\nu_1$ part of $\mathbf{E}[RC_t]$ as opposed to the $\mathbf{\Sigma}$ part. (Recall $\mathbf{E}[\mathbf{RC}_t] = \mathbf{\Sigma}_t + \mathbf{\Omega}_t/\nu_1$.)

Matrix-Variate Noncentral F Distribution

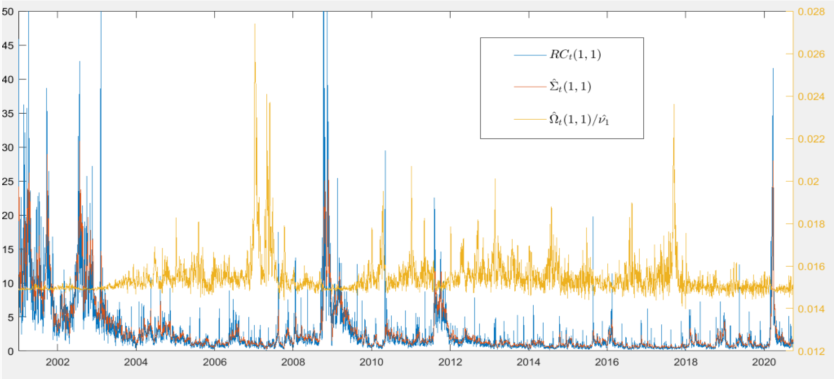

Figure 2: HEAVY GAS t noncentral-F model fit.

Problems and Plans

Problems:

- Parameter optimization is very unstable, highly dependent on starting values and ”precision“ of matrix-variate hypergeometric function evaluation.
- The matrix-variate hypergeometric function computing time explodes increasing ”precision“ and cross-sectional dimension.
- Differentiation of the matrix-variate hypergeometric function wrt matrix argument is not given in closed form. We use a complicated approximation for the score.

Plans:

- In-Sample fit comparison to ”restricted models“ using BIC.
- Forecasting experiment.
- Med -/High dimensional application.

Thank you for your attention!