Unrestricted, Restricted, and Regularized Models for Multivariate Volatility

Stanislav Anatolyev[1] Filip Staněk[2]

[1]CERGE-EI (Prague) & New Economic School (Moscow)

[2]CERGE-EI (Prague)

FMND 2021, online

Object of modeling

Time-varying $n \times n$ matrix S_t related to multivariate volatility for n assets

- **B**aba, **E**ngle, **K**raft, **K**roner (BEKK)

  (Engle and Kroner, ET 1995)

  and extensions (asymmetric BEKK, Caporin & McAleer, CSDA 2014)

  $\rightarrow S_t = var_{t-1}(r_t)$ where r_t is $n \times 1$ vector of returns

- **C**onditional **A**utoregressive **W**ishart model (CAW)

  (Golosnoy et al., JoE 2012)

  and extensions (threshold CAW, Anatolyev and Kobotaev, ER 2018)

  $\rightarrow S_t = E_{t-1}(RCov_t)$ where $RCov_t$ is $n \times n$ realized covariance matrix

CAW(p, q) volatility equation

$$S_t = CC^\top + \sum_{i=1}^{q} A^{(i)} RCov_{t-i} A^{(i)\top} + \sum_{i=1}^{p} B^{(i)} S_{t-i} B^{(i)\top}$$

Used in 3 (un)restricted forms:

- **f**ull CAW: $\varnothing$

  $\longrightarrow \#parameters = 1 + \frac{1}{2}n(n+1) + (p+q)n^2$

- **d**iagonal CAW: $\{A^{(i)}_{j,k:j\neq k} = 0, B^{(i)}_{j,k:j\neq k} = 0\}$

  $\longrightarrow \#parameters = 1 + \frac{1}{2}n(n+1) + (p+q)n$

- **s**calar CAW: $\{A^{(i)}_{j,k:j\neq k} = B^{(i)}_{j,k:j\neq k} = 0, A^{(i)}_{j,j} = A^{(i)}_{k,k}, B^{(i)}_{j,j} = B^{(i)}_{k,k}\}$

  $\longrightarrow \#parameters = 1 + \frac{1}{2}n(n+1) + (p+q)$

Degree of parameterization

Parsimony/prediction quality tradeoff

- fCAW is excessively parameterized leading to poor predictive performance
- sCAW and dCAW do not allow for flexible modeling of covariances

e.g.) fCAW(1,1), $n = 10$: fCAW/dCAW/sCAW has 256/76/58 parameters

Is there "optimal" parameterization?

We frame fCAW, dCAW and sCAW as special/limiting cases of **r**CAW

- apply ridge regularization to off-diagonal elements of $A^{(i)}$ and $B^{(i)}$ towards zero (Hoerl & Kennard, TM 1970)
- apply ridge regularization to diagonal elements of $A^{(i)}$ and $B^{(i)}$ towards common value (Anatolyev, JSCS 2020)

Frame fCAW, dCAW and sCAW as special cases of rCAW

Wishart loglikelihood

$$\underbrace{\left\{\widehat{C}, \widehat{A}, \widehat{B}, \widehat{v}\right\}}_{\widehat{\theta}} = \arg\max_{\theta} \sum_{t=1}^{T} \ell_t\left(RCov_t; v, \frac{S_t}{v}\right)$$

evolution of conditional mean

$$S_t = CC^\top + A\, RCov_{t-1} A^\top + B S_{t-1} B^\top$$

Frame fCAW, dCAW and sCAW as special cases of rCAW

Wishart loglikelihood L^2-penalized

$$\underbrace{\left\{\widehat{C},\widehat{A},\widehat{B},\widehat{v}\right\}}_{\widehat{\theta}} = \arg\max_{\theta} \sum_{t=1}^{T} \ell_t\left(RCov_t; v, \frac{S_t}{v}\right) - \lambda_f \tau_f\left(\theta\right) - \lambda_d \tau_d\left(\theta\right)$$

evolution of conditional mean

$$S_t = CC^\top + A\, RCov_{t-1} A^\top + B S_{t-1} B^\top$$

ridging off-diagonal elements towards zero:

$$\tau_f\left(\theta\right) = \sum_{j \neq k}^{n} A_{j,k}^2 + \sum_{j \neq k}^{n} B_{j,k}^2$$

ridging diagonal elements towards common value:

$$\tau_d\left(\theta\right) = \sum_{j=1}^{n} \left(A_{j,j} - \overline{A_{\cdot,\cdot}}\right)^2 + \sum_{j=1}^{n} \left(B_{j,j} - \overline{B_{\cdot,\cdot}}\right)^2$$

What we do next

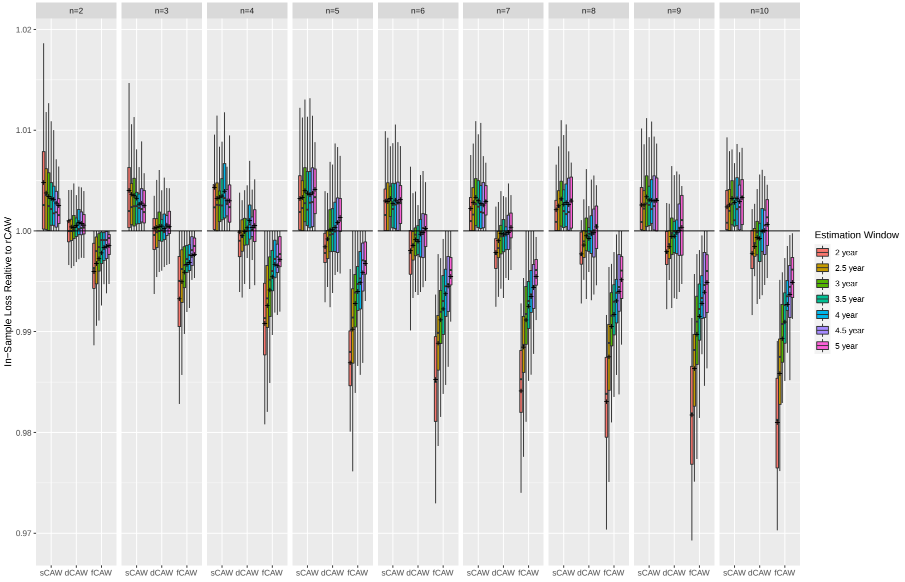
1. We investigate in-sample and forecasting performance of fCAW, dCAW and sCAW relative to ‘optimal’ rCAW
2. We investigate factors affecting forecasting performance:
   - which model is more ‘optimal’
   - dimensionality: number of stocks n
   - length of estimation window
   - distance from estimation window

Experimental setting

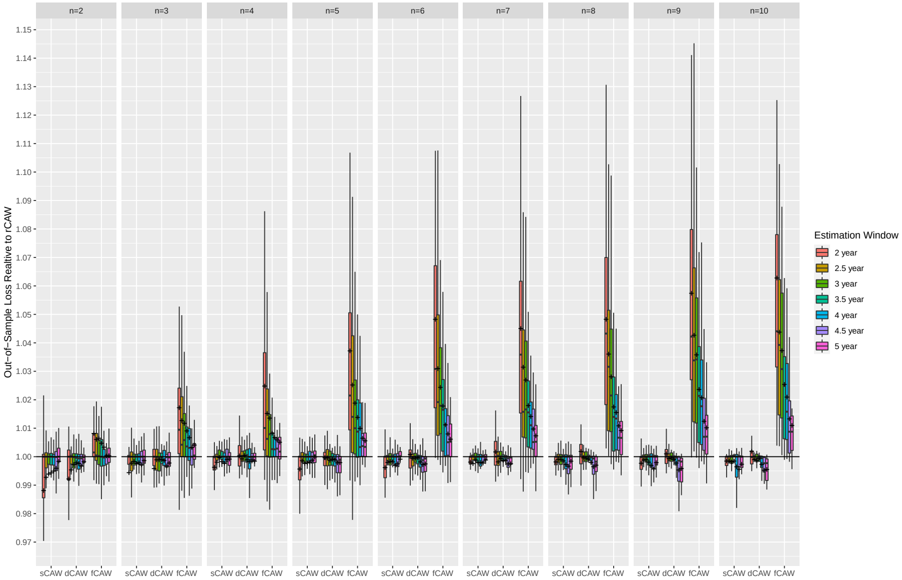
We adopt RCov data from Noureldin et al (JAE 2012) also used elsewhere

- Period: 2/2001–12/2009
- 10 stocks: BAC, JPM, IBM, MSFT, XOM, AA, AXP, DD, GE, KO
- 4 distinct random combinations of stocks of size n are selected
- Estimated models in rolling window:
  - sCAW(1,1)
  - dCAW(1,1)
  - fCAW(1,1)
  - rCAW(1,1) with hyper-parameters λ_f and λ_f selected on grid by evaluation on last 0.5 year of window
- 1-step ahead forecasting performance measured by Stein loss (coherent with Wishart likelihood)

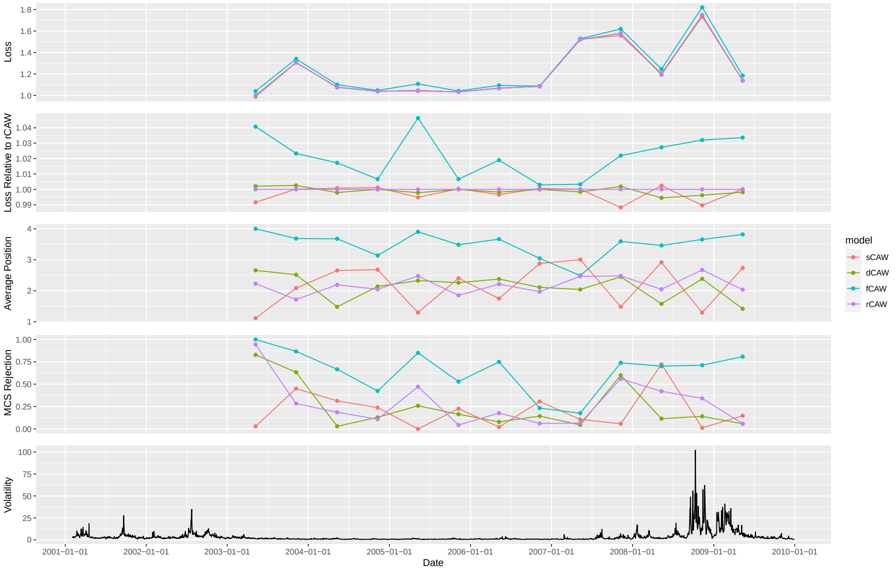
In-sample performance, Stein loss



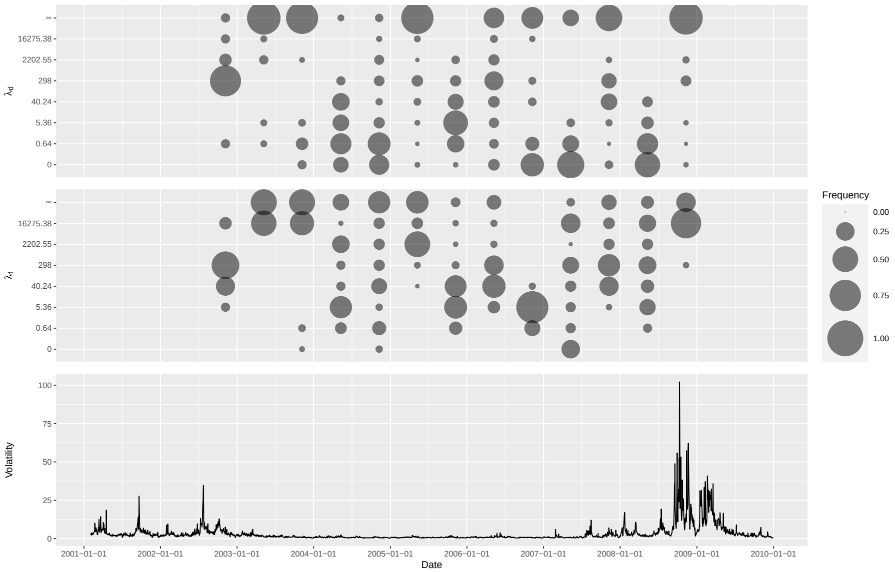
Out-of-sample performance, Stein loss



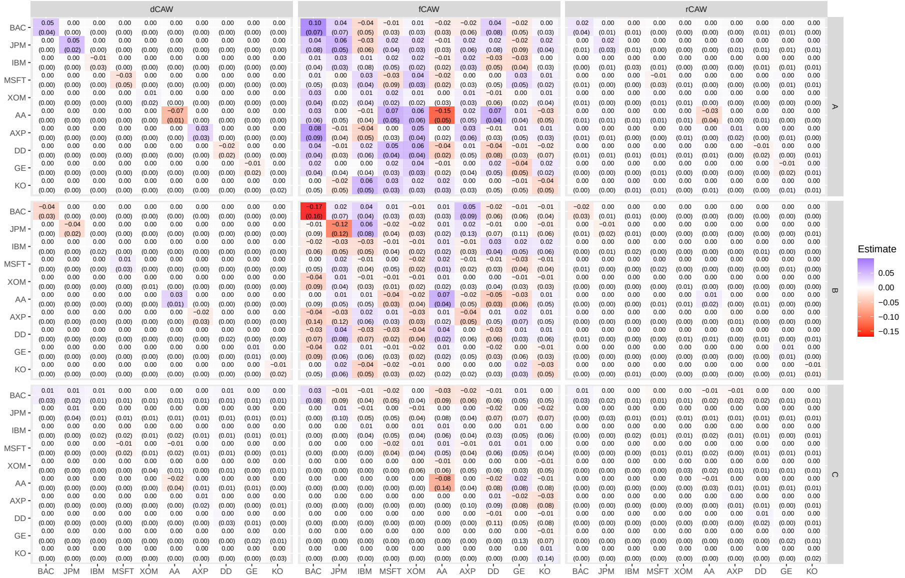
Out-of-sample performance: timing



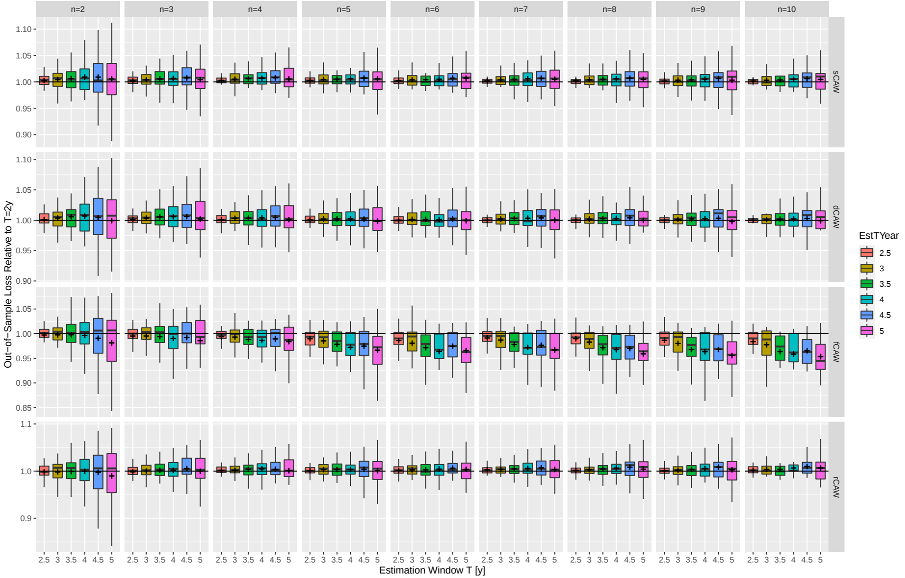
Selected λ_d and λ_f $(n \geq 5)$



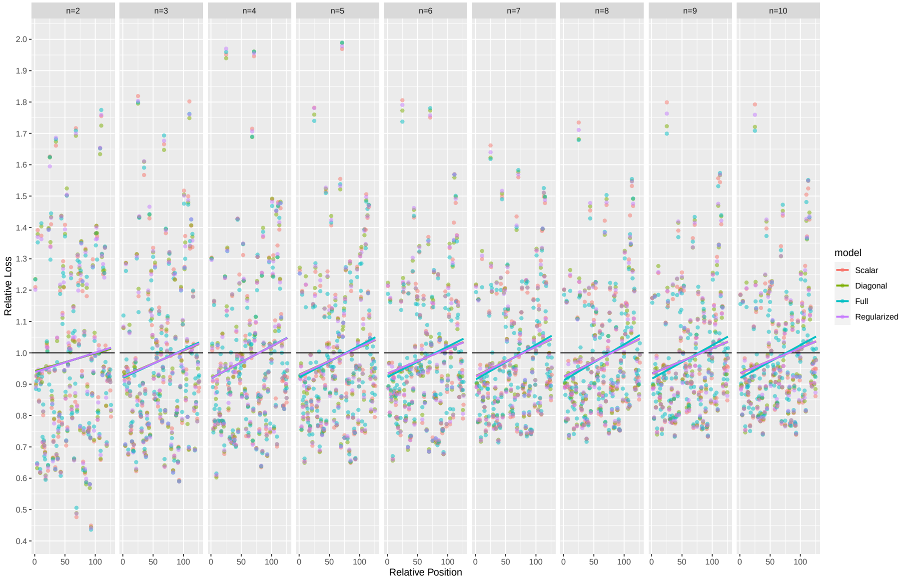
Estimated parameters relative to sCAW



Out-of-sample performance: estimation window



Out-of-sample performance: distance from estimation



Summary of findings

- fCAW is **systematically outperformed** by sCAW and dCAW
- Evidence on sCAW vs dCAW is **fuzzy**
- rCAW tends to choose **heavy penalization** close to sCAW
- Numerical differences between sCAW/dCAW/fCAW are **modest** on average ($\lesssim 2\%$) and **very noisy**
- Forecasting performance **deteriorates** with estimation window size
- Forecasting performance **strongly deteriorates** with distance from estimation window
- **Recommendation:** use sCAW/dCAW estimated in short rolling window