

Large Covariance Matrices

FAPESP Thematic Project Workshop
Machine Learning in Finance

September 2026

let's start in a Japanese way

what this lecture is about?

- ▶ how to estimate and forecast large covariance matrices (high dimensionality and positive definiteness)

what this lecture is not about?

- ▶ how to estimate and forecast volatility
- ▶ options-implied measures of covariation

main implication

- ▶ I'll assume you know how to measure/model/forecast volatility

going multivariate

why? portfolio allocation and risk management
call for covariance matrix!

modeling choices

- ▶ static vs dynamic specifications
- ▶ conditional vs realized covariances
- ▶ low vs high dimensionality

main issues how to limit number of parameters?
how to ensure positive definiteness?

main solutions

shrinkage estimators

- ▶ linear (Ledoit & Wolf, JMA'04, JPM'04)
- ▶ nonlinear (Ledoit & Wolf, RFS'17, AoS'20)

factor-based alternatives

- ▶ observed vs latent factors (Fan, Liao & Mincheva, AoS'11, JRSS'13)
- ▶ static vs dynamic models (Engle, Ledoit & Wolf, JBES'19)
- ▶ how to combine with shrinkage? (Ledoit & Wolf, JFinEcts'22)

nodewise regression

- (Meinshausen and Bühlmann, AoS'06)
- ▶ naïve vs residual (Callot et al, JBES'21; Caner et al, JoE'23)
- ▶ tracking error and weight constraints (Caner & Fan, JASA'26)

static covariance matrix

- ▶ sample covariance matrix is a natural estimator
 - if sample size is large enough (i.e, $T \gg N$)
 - then simple, consistent and, sometimes, efficient

however... not necessarily positive definite if N is large, and even infeasible if $N > T$

solutions

- ▶ shrinkage (linear vs nonlinear)
- ▶ thresholding (sparsity assumption)
- ▶ factor structure (observed vs latent)
- ▶ nodewise regression (naïve vs residual)

linear shrinkage

assumption $T \times N$ matrix X_T collects iid data, with mean zero and positive definite covariance matrix Σ_T

sample covariance matrix

$$S_T = \frac{1}{T} \sum_{t=1}^T X'_T X_T$$

► unbiased and, under Gaussianity, asymptotically efficient

though... ML equivalence is not necessarily desirable in finite samples (James & Stein, BSMSP'61)

James-Stein estimator

$(N \geq 3)$

linear combination between sample and target mean

- shrinking towards target reduces variance, even if introducing bias
- Ledoit & Wolf (JMA'04) extend approach to covariance matrices

linear shrinkage for covariance matrix

- ▶ quadratic loss function, leading to MSE risk

- ▶ **scaled** Frobenius norm

- ▶ MSE criterion

$$\|A\|_F = \sqrt{\text{tr}(A'A)/N}$$
$$\mathbb{E} \left[\|\hat{\Sigma}_T - \Sigma_T\|_F^2 \right]$$

- ▶ linear combination with identity matrix

$$\hat{\Sigma}_T = \lambda_S S_T + \lambda_I I_T$$

$$\min_{\lambda_S, \lambda_I} \mathbb{E} \left[\|\lambda_S S_T + \lambda_I I_T - \Sigma_T\|_F^2 \right]$$

solution

$$\varsigma_T := \text{tr}(\Sigma_T)/N$$

$$\Sigma_T^* = \frac{\mathbb{E} \left[\|S_T - \Sigma_T\|_F^2 \right]}{\mathbb{E} \left[\|S_T - \varsigma_T I_T\|_F^2 \right]} \varsigma_T I_T + \frac{\mathbb{E} \left[\|\Sigma_T - \varsigma_T I_T\|_F^2 \right]}{\mathbb{E} \left[\|S_T - \varsigma_T I_T\|_F^2 \right]} S_T$$

linear shrinkage of the covariance matrix

- ▶ quadratic loss function, leading to MSE risk

- ▶ scaled Frobenius norm

$$\|A\|_F = \sqrt{\text{tr}(A'A)/N}$$

- ▶ MSE criterion

$$\mathbb{E} \left[\|\hat{\Sigma}_T - \Sigma_T\|_F^2 \right]$$

- ▶ linear combination with identity matrix

$$\hat{\Sigma}_T = \lambda_S S_T + \lambda_I I_T$$

$$\min_{\lambda_S, \lambda_I} \mathbb{E} \left[\|\lambda_S S_T + \lambda_I I_T - \Sigma_T\|_F^2 \right]$$

solution

$$\varsigma_T := \text{tr}(\Sigma_T)/N$$

$$\Sigma_T^* = \lambda_T^* \varsigma_T I_T + (1 - \lambda_T^*) S_T$$

$$\text{given that } \mathbb{E} \left[\|S_T - \Sigma_T\|_F^2 \right] + \mathbb{E} \left[\|\Sigma_T - \varsigma_T I_T\|_F^2 \right] = \mathbb{E} \left[\|S_T - \varsigma_T I_T\|_F^2 \right]$$

linear shrinkage of the covariance matrix

- ▶ quadratic loss function, leading to MSE risk

- ▶ scaled Frobenius norm

- ▶ MSE criterion

$$\|A\|_F = \sqrt{\text{tr}(A'A)/N}$$
$$\mathbb{E} \left[\|\hat{\Sigma}_T - \Sigma_T\|_F^2 \right]$$

- ▶ linear combination with identity matrix

$$\hat{\Sigma}_T = \lambda_S S_T + \lambda_I I_T$$

$$\min_{\lambda_S, \lambda_I} \mathbb{E} \left[\|\lambda_S S_T + \lambda_I I_T - \Sigma_T\|_F^2 \right]$$

solution

$$\varsigma_T := \text{tr}(\Sigma_T)/N$$

$$\Sigma_T^* = \lambda_T^* \varsigma_T I_T + (1 - \lambda_T^*) S_T$$

- ▶ average variance as shrinking target $\varsigma_T I_T$

linear shrinkage of the covariance matrix

- ▶ quadratic loss function, leading to MSE risk

- ▶ **scaled** Frobenius norm

- ▶ MSE criterion

$$\|A\|_F = \sqrt{\text{tr}(A'A)/N}$$
$$\mathbb{E} \left[\|\hat{\Sigma}_T - \Sigma_T\|_F^2 \right]$$

- ▶ linear combination with identity matrix

$$\hat{\Sigma}_T = \lambda_S S_T + \lambda_I I_T$$

$$\min_{\lambda_S, \lambda_I} \mathbb{E} \left[\|\lambda_S S_T + \lambda_I I_T - \Sigma_T\|_F^2 \right]$$

solution

$$\varsigma_T := \text{tr}(\Sigma_T)/N$$

$$\Sigma_T^* = \lambda_T^* \varsigma_T I_T + (1 - \lambda_T^*) S_T$$

- ▶ shrinkage intensity λ_T^*

some remarks

- ▶ Σ_T^* is positive definitive even if $N > T$
(convex combination of the pd I_T and psd S_T)

though... Σ_T^* is infeasible due to dependence on Σ_T

- ▶ $\hat{\Sigma}_T^*$ is consistent to Σ_T^* , but not to Σ_T
(unless $N/T \rightarrow 0$, or $N/T \rightarrow c < \infty$ under sparsity)

alternative targets close as possible to true covariance matrix
with as few parameters as possible, ideally

- ▶ equicorrelation to reflect usually positive correlation across stocks
- ▶ CAPM-implied covariance matrix $\Phi_T = \sigma_M^2 \beta_T \beta_T' + \Delta_T$

nonlinear shrinkage

why does shrinkage work?

- ▶ it dampens the variation in eigenvalues by increasing/decreasing smallest/largest eigenvalues of the sample covariance matrix
- though...** common intensity λ_T^*

idea why not shrink more the noisiest entries of S_T ?

- ▶ number of shrinking intensities is $O(N^2)$
- ▶ no guarantees of positive definiteness

back to drawing board spectral decomposition $S_T = U_T \Delta_T U_T'$
 U_T orthogonal matrix with eigenvectors
 Δ_T diagonal matrix with eigenvalues

nonlinear shrinkage

why does shrinkage work?

- ▶ it dampens the variation in eigenvalues by increasing/decreasing smallest/largest eigenvalues of the sample covariance matrix
- though...** common intensity λ_T^*

idea why not shrink more the noisiest entries of S_T ?

- ▶ number of shrinking intensities is $O(N^2)$
- ▶ no guarantees of positive definiteness

back to drawing board spectral decomposition $S_T = U_T \Delta_T U_T'$
shrinkage $\Sigma_T^* = U_T \Delta_T^* U_T'$ ($\delta_{j,T}^* > 0$)
if linear, then $\delta_{j,T}^* = \lambda_T^* \varsigma_T + (1 - \lambda_T^*) \delta_{j,T}$

nonlinear shrinkage

why does shrinkage work?

- ▶ it dampens the variation in eigenvalues by increasing/decreasing smallest/largest eigenvalues of the sample covariance matrix
- though...** common intensity λ_T^*

idea why not shrink more the noisiest entries of S_T ?

- ▶ number of shrinking intensities is $O(N^2)$
- ▶ no guarantees of positive definiteness

back to drawing board spectral decomposition $S_T = U_T \Delta_T U_T'$
shrinkage $\Sigma_T^* = U_T \Delta_T^* U_T'$ ($\delta_{j,T}^* > 0$)
and if nonlinear?

nonlinear shrinkage

why does shrinkage work?

- ▶ it dampens the variation in eigenvalues by increasing/decreasing smallest/largest eigenvalues of the sample covariance matrix
- though...** common intensity λ_T^*

idea why not shrink more the noisiest entries of S_T ?

- ▶ number of shrinking intensities is $O(N^2)$
- ▶ no guarantees of positive definiteness

back to drawing board spectral decomposition $S_T = U_T \Delta_T U_T'$
shrinkage $\Sigma_T^* = U_T \Delta_T^* U_T'$ ($\delta_{j,T}^* > 0$)
if nonlinear? still equivariant to rotation!

how to do nonlinear shrinkage, then?

$$\min_{\Delta_T} \|U_T \Delta_T U_T' - \Sigma_T\|_F^2$$

- ▶ optimal solution in finite samples $\bar{\Delta}_T$ with $\bar{\delta}_{j,T} = u_{j,T}' \Sigma_T u_{j,T}$
 - interestingly** it combines sample and population information
 - though...** infeasible due to dependence on Σ_T
 - in addition, it depends on N parameters

solution define function that transforms sample eigenvalues in a dimension-free manner (no dependence on N)

restrictions ensure positivity of resulting eigenvalues
nonstochastic limit if dependent on T

$$\hat{\Sigma}_T = U_T \hat{\Delta}_T U_T', \quad \text{with } \hat{\delta}_{j,T} = \hat{\phi}_T(\delta_{j,T}) \text{ and } \hat{\phi}_T \rightarrow \phi$$

how to find optimal function ϕ ?

random matrix theory

- ▶ asymptotic behavior of sample covariance matrices, with a major focus on sample eigenvalues
- ▶ sample eigenvalues are asymptotically nonstochastic ($N/T \rightarrow c \neq 1$) (under tons of regularity conditions, including iidness)

$$\phi^*(x) = \begin{cases} \frac{1}{\pi(c-1)\mathcal{H}_{\bar{f}}} & \text{if } x = 0, \quad c > 1 \\ \frac{x}{[\pi c x f(x)]^2 + [1 - c - \pi c x \mathcal{H}_f(x)]^2} & \text{otherwise} \end{cases}$$

how to find optimal function ϕ ?

random matrix theory

- ▶ asymptotic behavior of sample covariance matrices, with a major focus on sample eigenvalues
- ▶ sample eigenvalues are asymptotically nonstochastic ($N/T \rightarrow c \neq 1$) (under tons of regularity conditions, including iidness)

$$\phi^*(x) = \begin{cases} \frac{1}{\pi(c-1)\mathcal{H}_{\bar{f}}} & \text{if } x = 0, \quad c > 1 \\ \frac{x}{[\pi cx \textcolor{red}{f}(x)]^2 + [1 - c - \pi cx \mathcal{H}_{\textcolor{red}{f}}(x)]^2} & \text{otherwise} \end{cases}$$

where $\textcolor{red}{f} \sim$ density of the nonstochastic limit $\textcolor{red}{F}$ of the sample eigenvalues' empirical distribution

how to find optimal function ϕ ?

random matrix theory

- ▶ asymptotic behavior of sample covariance matrices, with a major focus on sample eigenvalues
- ▶ sample eigenvalues are asymptotically nonstochastic ($N/T \rightarrow c \neq 1$) (under tons of regularity conditions, including iidness)

$$\phi^*(x) = \begin{cases} \frac{1}{\pi(c-1)\mathcal{H}_{\bar{f}}} & \text{if } x = 0, \quad c > 1 \\ \frac{x}{[\pi c x f(x)]^2 + [1 - c - \pi c x \mathcal{H}_f(x)]^2} & \text{otherwise} \end{cases}$$

where $\bar{f} \sim \text{density of } cF + (1 - c)\mathbf{1}_{(0, \infty)}$

how to find optimal function ϕ ?

random matrix theory

- ▶ asymptotic behavior of sample covariance matrices, with a major focus on sample eigenvalues
- ▶ sample eigenvalues are asymptotically nonstochastic ($N/T \rightarrow c \neq 1$) (under tons of regularity conditions, including iidness)

$$\phi^*(x) = \begin{cases} \frac{1}{\pi(c-1)\mathcal{H}_{\bar{f}}} & \text{if } x = 0, \quad c > 1 \\ \frac{x}{[\pi c x f(x)]^2 + [1 - c - \pi c x \mathcal{H}_f(x)]^2} & \text{otherwise} \end{cases}$$

where $\mathcal{H}_g \sim$ Hilbert transform of a real function g , which depends on Cauchy principal value

how to find optimal function ϕ ?

random matrix theory

- ▶ asymptotic behavior of sample covariance matrices, with a major focus on sample eigenvalues
- ▶ sample eigenvalues are asymptotically nonstochastic ($N/T \rightarrow c \neq 1$) (under tons of regularity conditions, including iidness)

$$\phi^*(x) = \begin{cases} \frac{1}{\pi(c-1)\mathcal{H}_{\bar{f}}} & \text{if } x = 0, \quad c > 1 \\ \frac{x}{[\pi c x f(x)]^2 + [1 - c - \pi c x \mathcal{H}_f(x)]^2} & \text{otherwise} \end{cases}$$

estimator $\hat{\phi}_T$ using $\hat{c}_T = N/T$, and kernel estimators for $(f, \bar{f}, \mathcal{H}_f, \mathcal{H}_{\bar{f}})$
global/linear vs local/nonlinear shrinkage

local nature of Hilbert transforms

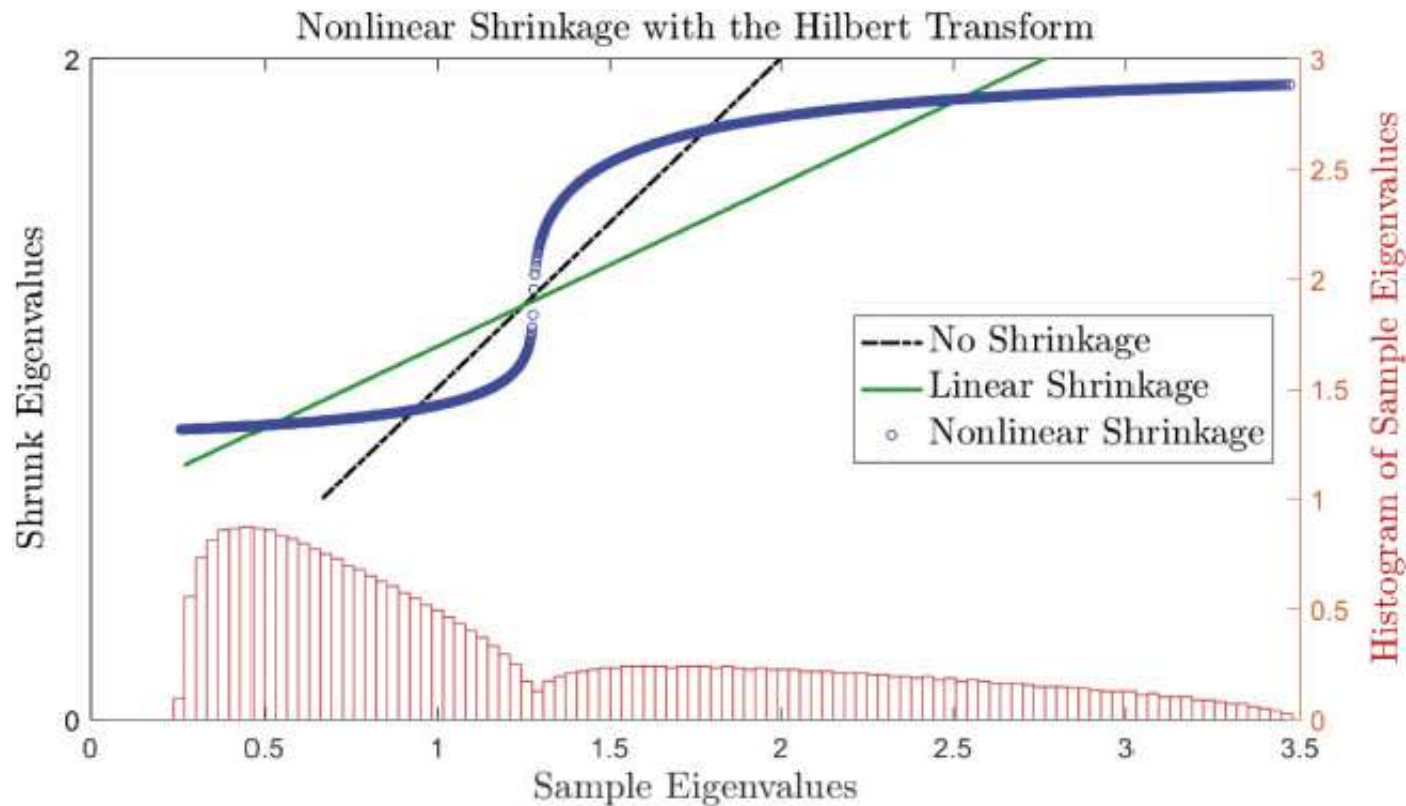


Figure 1 Local attraction effect. A total of 2500 population eigenvalues are equal to 0.8, and 1500 are equal to 2, so that $N=4000$. The sample size is $T=18,000$. At the bottom of the figure is a histogram displaying the location of the sample eigenvalues.

extension to dynamic specifications

multivariate GARCH curse of dimensionality severely limits N

- ▶ shrinkage methods are appropriate to static covariance matrix
- ▶ trick is to find specification with static component

example DCC specification with *correlation targeting*

pseudo **conditional** vs **unconditional** correlation matrix

$$Q_t = (1 - \alpha - \beta)C + \alpha s_t s_t' + \beta Q_{t-1} \quad \text{up for shrinkage!}$$

with $s_{jt} = r_{jt}/\sigma_{jt}$ denoting standardized returns

algorithm

(Engle, Ledoit & Wolf, RFS'19)

- (1) fit individual GARCH(1,1) models to standardize returns
- (2) estimate unconditional correlation matrix C using shrinkage
- (3) maximize composite likelihood to estimate α and β

extension to factors

factor structure interpretable dimensionality reduction

example CAPM-implied target covariance matrix

though... large N calls for multiple factors and, hence, for the estimation of their covariance matrix

► $\Sigma_t = B' \Sigma_f B + \Sigma_{u,t}$

algorithm

(De Nard, Ledoit & Wolf, JFinEcts'20)

- (1) estimate by OLS individual risk exposures to each factor B
- (2) estimate sample covariance matrix of the factors Σ_f
- (3) estimate idiosyncratic covariance matrix by DCC-NL $\Sigma_{u,t}$

homoskedastic factor structure

example under a time-varying beta specification, sometimes there remains no conditional heteroskedasticity

if static idiosyncratic covariance matrix Σ_u

then usual (non)linear shrinkage suffices

alternative thresholding estimators

- ▶ approximate factor structure implies a sparse idiosyncratic covariance matrix (Fan, Liao & Mincheva, JASA'11)
- ▶ principal orthogonal complement thresholding (POET) estimator under latent factor structure (Fan, Liao & Mincheva, JRSS'13)

nodewise regression

(Meinshausen & Bühlmann, AoS'06)

main idea estimate *precision* matrix by stacking matrix algebraic closed-form solutions for each row

► each row has main and off diagonal terms

main $\Gamma_{j,j} = (\Sigma_{j,j} - \Sigma_{j,-j} \Sigma_{-j,-j}^{-1} \Sigma_{-j,j})^{-1}$
where $\Sigma_{-j,-j} \sim$ matrix without j th row/column
 $\Sigma_{-j,j} \sim j$ th column of Σ without j th entry
 $\Sigma_{j,-j} \sim j$ th row of Σ without j th entry

off $\Gamma_{j,-j} = -\Gamma_{j,j} \Sigma_{j,-j} \Sigma_{-j,-j}^{-1}$

connection with linear regression (Callot et al, JBES'21)

if we run regression $\tilde{r}_{j,t} := r_{j,t} - \bar{r}_j = \tilde{r}'_{-j,t} \gamma_j + \eta_{j,t}$
then $\gamma_j = \Sigma_{-j,-j}^{-1} \Sigma_{-j,j}$, so that $\Gamma_{j,-j} = -\Gamma_{j,j} \gamma_j'$
 $\Gamma_{j,j} = (\Sigma_{j,j} - \Sigma_{j,-j} \gamma_j)^{-1}$

$$\Gamma = \text{diag} \begin{pmatrix} (\Sigma_{1,1} - \Sigma_{1,-1} \gamma_1)^{-1} \\ \vdots \\ (\Sigma_{N,N} - \Sigma_{N,-N} \gamma_N)^{-1} \end{pmatrix} \begin{pmatrix} 1 & -\gamma_{1,2} & \cdots & -\gamma_{1,N} \\ -\gamma_{2,1} & 1 & \cdots & -\gamma_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ -\gamma_{N,1} & -\gamma_{N,2} & \cdots & 1 \end{pmatrix}$$

- ▶ LASSO regularization to induce sparsity in **precision matrix**, which does not imply a sparse **covariance matrix** (conditional/partial vs unconditional/total correlation links)
- ▶ asymptotic theory requires only β -mixingness

idiosyncratic noise (Caner, Medeiros & Vasconcelos, JoE'23)

main idea estimate *precision* matrix of idiosyncratic component after adjusting excess returns for risk factor exposures

algorithm

- (1) estimate by OLS factor model $y_j = F'b_j + u_j$
- (2) estimate by LASSO precision matrix Γ_u using residuals

asymptotic theory

- ▶ weak dependence, as before
- ▶ stronger sparsity assumption due to growing number of factors
- ▶ control for reliance on residuals (rather than errors)

more on dynamic modeling

time-varying covariances

so far... brief description of DCC model

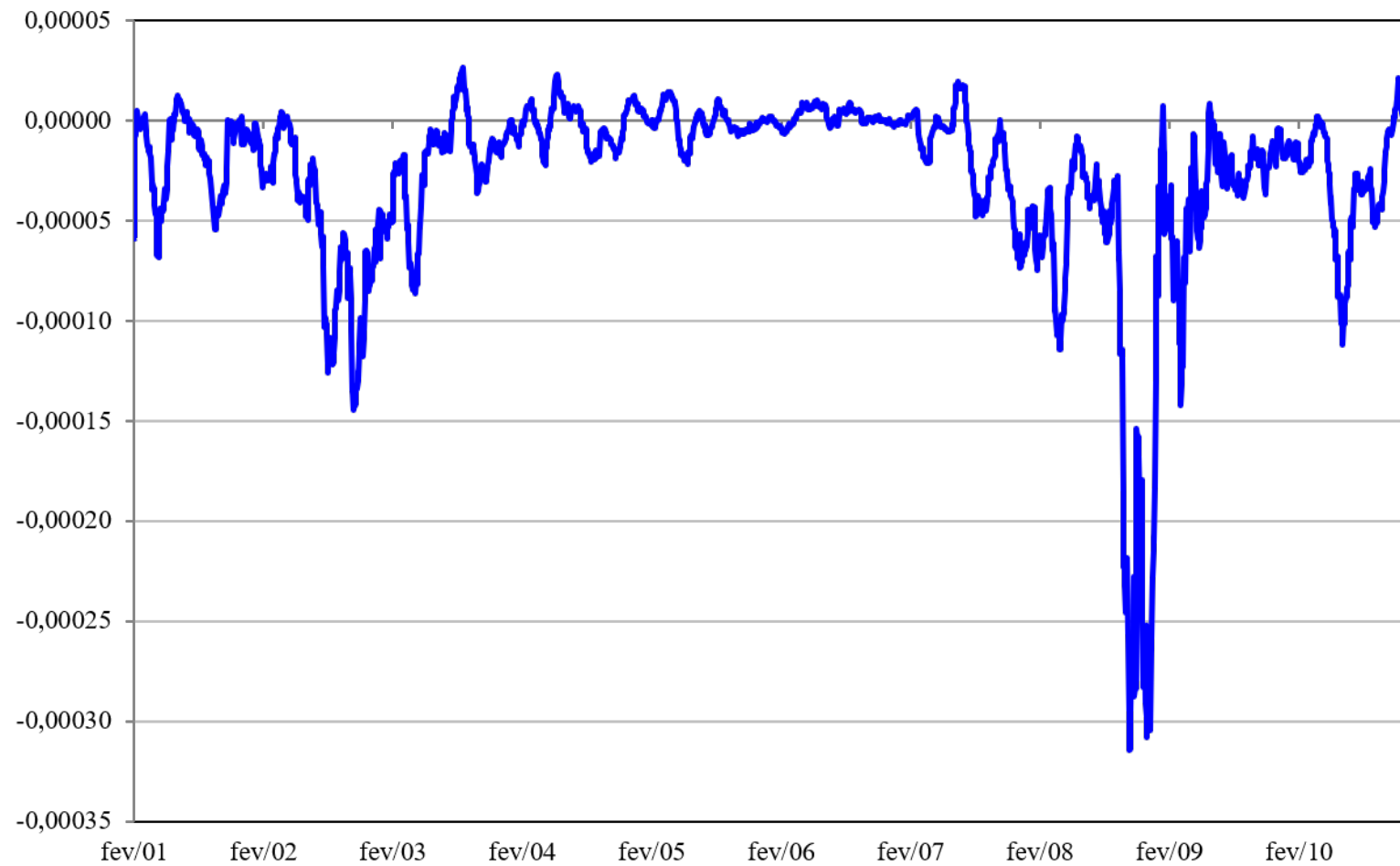
- ▶ curse of dimensionality in multivariate GARCH modeling
- ▶ positive definiteness guarantees

back to basics

- ▶ rolling covariances

$$\hat{\sigma}_{ij,t+1} = \frac{1}{L+1} \sum_{\ell=0}^L r_{i,t-\ell} r_{j,t-\ell}$$

rolling covariance, S&P500 vs 10-year T-note



time-varying covariances

so far... brief description of DCC model

- ▶ curse of dimensionality in multivariate GARCH modeling
- ▶ positive definiteness guarantees

back to basics

- ▶ rolling covariances

$$\hat{\sigma}_{ij,t+1} = \frac{1}{L+1} \sum_{\ell=0}^L r_{i,t-\ell} r_{j,t-\ell}$$

- ▶ easy extension to nonzero expected returns
- ▶ how to choose L in a robust manner?
- ▶ how sensible are uniform weights?

time-varying covariances

so far... brief description of DCC model

- ▶ curse of dimensionality in multivariate GARCH modeling
- ▶ positive definiteness guarantees

back to basics

- ▶ exponential smoothing

$$\hat{\sigma}_{ij,t+1} = (1 - \lambda)r_{i,t}r_{j,t} + \lambda\hat{\sigma}_{ij,t}$$

time-varying covariances

so far... brief description of DCC model

- ▶ curse of dimensionality in multivariate GARCH modeling
- ▶ positive definiteness guarantees

back to basics

- ▶ exponential smoothing

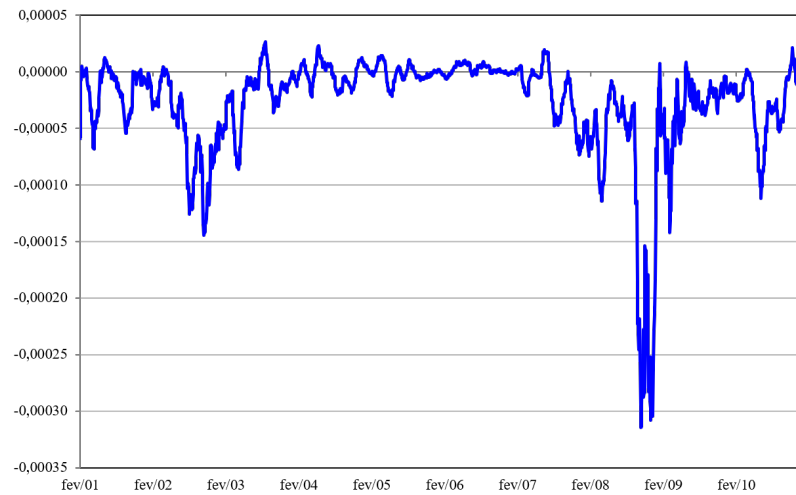
$$\hat{\sigma}_{ij,t+1} = (1 - \lambda)r_{i,t}r_{j,t} + \lambda\hat{\sigma}_{ij,t}$$

- ▶ easy extension to nonzero expected returns
- ▶ convex combination implies nonstationarity
- ▶ how to choose λ in a robust manner?

EWMA covariance, S&P500 vs 10-year T-note



rolling covariance vs EWMA



- ▶ perhaps a bit smoother, but not really much difference!

time-varying covariances

so far... brief description of DCC model

- ▶ curse of dimensionality in multivariate GARCH modeling
- ▶ positive definiteness guarantees

back to basics

- ▶ GARCH-type specification

$$\sigma_{ij,t+1} = \omega_{ij} + \alpha r_{i,t} + \beta \sigma_{ij,t}$$

- ▶ reverts to $\omega_{ij}/(1 - \alpha - \beta)$ in the long run
 - ▶ common (α, β) to ensure positivity of portfolio variance
- though...** it doesn't necessarily make sense!

time-varying correlations

- ▶ covariance is a confluence of correlation and variance,
 - ▶ time-varying feature could well stem only from variances
- ▶ correlations typically increase during financial turmoil
 - ▶ correlation dynamics is crucial to risk management

example exponential smoothing

- ▶ not particularly intuitive, though!

$$\hat{\rho}_{ij,t+1} = \frac{(1 - \lambda)r_{i,t}r_{j,t} + \lambda\hat{\sigma}_{ij,t}}{\sqrt{(1 - \lambda)r_{i,t}^2 + \lambda\hat{\sigma}_{i,t}^2}\sqrt{(1 - \lambda)r_{j,t}^2 + \lambda\hat{\sigma}_{j,t}^2}}$$

time-varying correlations

$$\Sigma_{t+1} = D_{t+1} \Upsilon_{t+1} D_{t+1}, \quad \text{with} \quad D_{t+1} = \text{diag}(\sigma_{1,t+1}, \dots, \sigma_{N,t+1})$$

$\Upsilon_{t+1} \sim \text{correlation matrix}$

example

$$\begin{aligned} \Sigma_{t+1} &= \begin{bmatrix} \sigma_{1,t+1}^2 & \sigma_{12,t+1} \\ \sigma_{12,t+1} & \sigma_{2,t+1}^2 \end{bmatrix} \\ &= \begin{bmatrix} \sigma_{1,t+1} & 0 \\ 0 & \sigma_{2,t+1} \end{bmatrix} \begin{bmatrix} 1 & \rho_{12,t+1} \\ \rho_{12,t+1} & 1 \end{bmatrix} \begin{bmatrix} \sigma_{1,t+1} & 0 \\ 0 & \sigma_{2,t+1} \end{bmatrix} \end{aligned}$$

modeling steps

- (1) estimate univariate GARCH model for each volatility $\hat{\sigma}_{i,t}$
- (2) standardize returns by their volatility $z_{i,t} = r_{i,t} / \hat{\sigma}_{i,t}$

$$\mathbb{E}(z_{i,t} z_{j,t}) = \text{corr}(r_{i,t} r_{j,t}) = \rho_{ij,t}$$

exponential smoothing for correlations

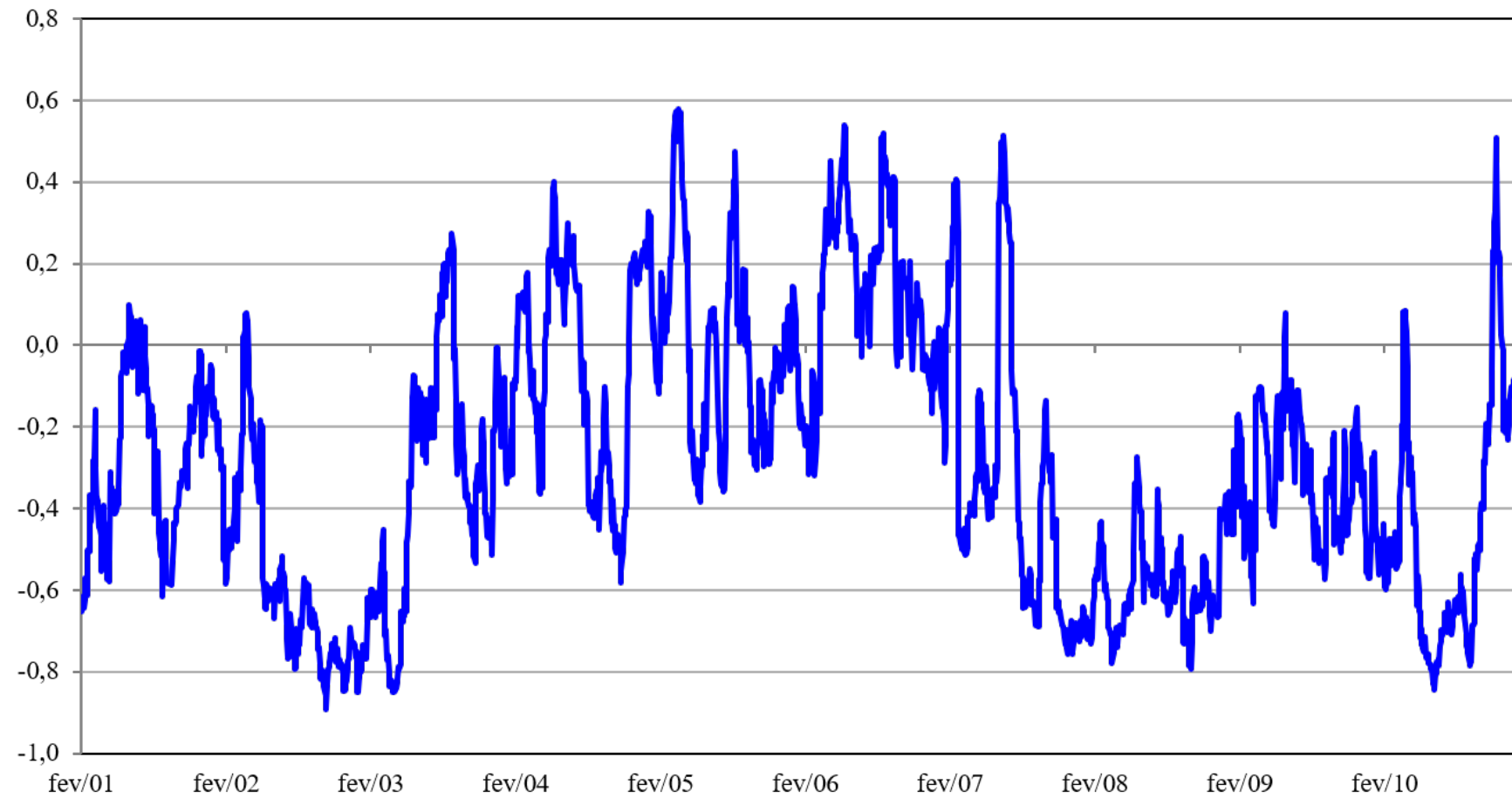
- ▶ updating by cross product of standardized returns

$$q_{ij,t+1} = (1 - \lambda) z_{i,t} z_{j,t} + \lambda q_{ij,t}$$

- ▶ apart from absence of mean reversion, there is no guarantee that magnitude lies in the unit interval normalize!

$$\hat{\rho}_{ij,t+1} = \frac{q_{ij,t+1}}{\sqrt{q_{ii,t+1} q_{jj,t+1}}}$$

EWMA correlation, S&P500 vs 10-year T-note



exponential smoothing for correlations

- ▶ updating by cross product of standardized returns

$$q_{ij,t+1} = (1 - \lambda) z_{i,t} z_{j,t} + \lambda q_{ij,t}$$

- ▶ apart from absence of mean reversion, there is no guarantee that magnitude lies in the unit interval normalize!

$$\hat{\rho}_{ij,t+1} = \frac{q_{ij,t+1}}{\sqrt{q_{ii,t+1} q_{jj,t+1}}}$$

natural next step

- ▶ impose mean reversion by considering GARCH extension

dynamic conditional correlation

DCC

- ▶ allowing correlations to revert to long-run correlation $\rho_{ij} = \mathbb{E}(z_{i,t}z_{j,t})$

$$\begin{aligned}q_{ij,t+1} &= \rho_{i,j} + \alpha (z_{i,t} z_{j,t} - \rho_{ij}) + \beta (q_{ij,t} - \rho_{ij}) \\&= \rho_{i,j}(1 - \alpha - \beta) + \alpha z_{i,t} z_{j,t} + \beta q_{ij,t}\end{aligned}$$

details

- ▶ plug in $\bar{\rho}_{i,j}$ to impose correlation targeting
- ▶ same persistence in correlation across assets
- ▶ normalize to obtain proper correlation measures

$$Q_{t+1} = \mathbb{E}(Z_t Z_t')(1 - \alpha - \beta) + \alpha Z_t' Z_t + \beta Q_t$$

- ▶ $Q_t \sim$ positive semi-definite as any weighted average of positive semi-definite and positive definite matrices

dynamic conditional correlation

DCC

- ▶ allowing correlations to revert to long-run correlation $\rho_{ij} = \mathbb{E}(z_{i,t}z_{j,t})$

$$\begin{aligned}q_{ij,t+1} &= \rho_{i,j} + \alpha (z_{i,t} z_{j,t} - \rho_{ij}) + \beta (q_{ij,t} - \rho_{ij}) \\&= \rho_{i,j}(1 - \alpha - \beta) + \alpha z_{i,t} z_{j,t} + \beta q_{ij,t}\end{aligned}$$

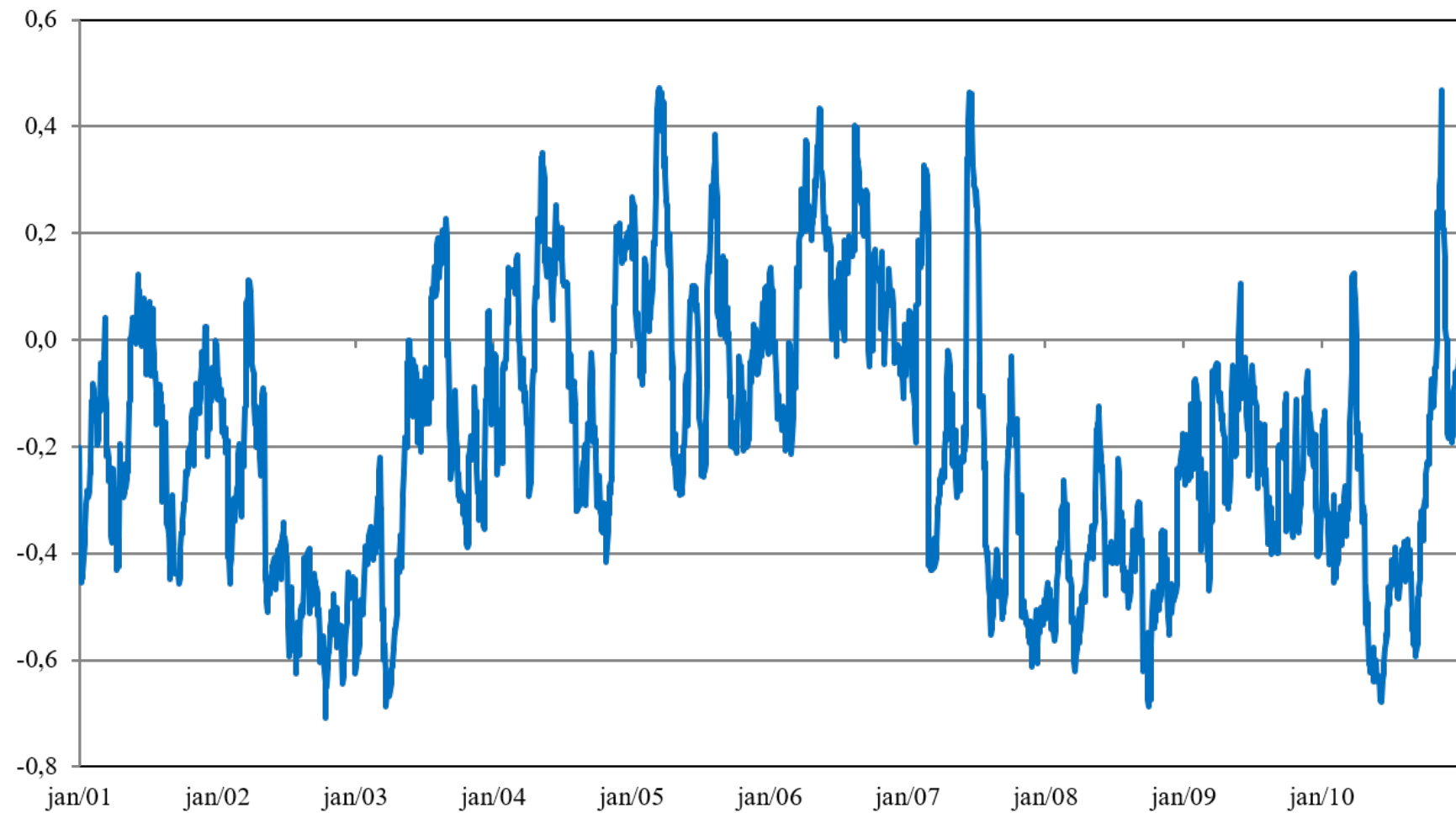
details

- ▶ plug in $\bar{\rho}_{i,j}$ to impose correlation targeting
- ▶ same persistence in correlation across assets
- ▶ normalize to obtain proper correlation measures

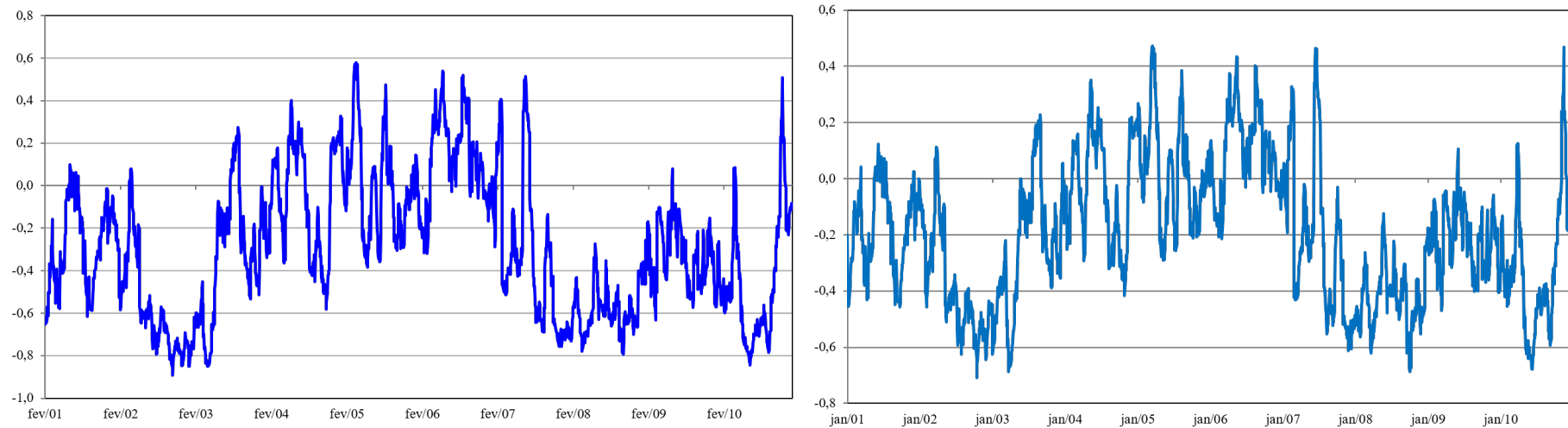
$$Q_{t+1} = \mathbb{E}(Z_t Z_t')(1 - \alpha - \beta) + \alpha Z_t' Z_t + \beta Q_t$$

- ▶ DCC $\sim$ very tractable model even for large portfolios given common persistence in correlation and targeting

GARCH correlation, S&P500 vs 10-year T-note



EWMA vs GARCH correlation



- ▶ as before, not really much difference!
(no surprise given that close-to-unit $\alpha + \beta$)

realized covariance matrix

realized measures of covariation

- ▶ straightforward extension to covariance estimation

$$\text{quadratic covariation} \approx \text{integrated covariance} + \text{cojump component} + \text{microstructure noise covariation}$$

examples

$$\text{rcov}_1 = \text{bpcov}_5 + \text{rcov}_5 - \text{bpcov}_5 + \text{rcov}_1 - \text{rcov}_5$$

details

- ▶ pre-averaging vs temporal aggregation vs refresh time
- ▶ pairwise approach, so very tractable for very large portfolios
- ▶ can we treat realized covariance matrix as observable?

modeling realized covariances

initial idea $y_{t+1} := \text{vech}(\hat{\Sigma}_{t+1}) \sim V(H)AR$ process

though... how to limit number of parameters?
how to ensure positive definiteness?

solution

- ▶ LASSO penalty to control for number of parameters that grows quadratically with N (Callot, Kock & Medeiros, JAE'17)

example S&P500 index constituents, with HAR structure

- ▶ each matrix $\hat{\Sigma}_t$ has $N(N+1)/2$ unique entries
- ▶ 125,250 equations with $4 \times 125,250$ variables each

modeling realized covariances

initial idea $y_{t+1} := \text{vech}(\Sigma_{t+1}) \sim V(H)AR$ process

though... how to limit number of parameters?
how to ensure positive definiteness?

solution

totally infeasible!

- ▶ LASSO penalty to control for number of parameters that grows quadratically with N (Callot, Kock & Medeiros, JAE'17)

example S&P500 index constituents, with HAR structure

- ▶ each matrix $\hat{\Sigma}_t$ has $N(N+1)/2$ unique entries
- ▶ 125,250 equations with $4 \times 125,250$ variables each

modeling realized covariances

initial idea $y_{t+1} := \text{vech}(\Sigma_{t+1}) \sim V(H)AR$ process

though... how to limit number of parameters?
how to ensure positive definiteness?

solution

- ▶ factor/sector structures + LASSO to mitigate drastically curse of dimensionality (Alves, Brito, Medeiros & Ribeiro, JFinEcts'24)

example S&P500 index constituents, with HAR structure

- ▶ factor covariance $\sim K(K+1)/2 \ll N(N+1)/2$
- ▶ residual covariance $\sim$ block diagonal by sector ($S < N$)

step-by-step modeling

- ▶ factor structure for excess returns, with time-varying loadings

$$\begin{aligned} \mathbf{r}_t^e &= \mathbf{B}_t' \mathbf{F}_t + \mathbf{e}_t, & \text{with } \mathbb{E}(\mathbf{e}_t | \mathbf{F}_t) &= 0 \\ \mathbf{F}_t &= \mathbf{W}' \mathbf{r}_t \end{aligned}$$

- ▶ factors are linear combinations of excess returns, with weights based on accounting and market information

$$\begin{aligned} \Sigma_{t+1|t} &= \mathbf{B}_{t+1|t}' \Sigma_{f,t+1|t} \mathbf{B}_{t+1|t} + \Sigma_{e,t+1|t} \\ \Sigma_{f,t+1|t} &= \mathbf{W}' \Sigma_{t+1|t} \mathbf{W} \end{aligned}$$

step-by-step modeling

- ▶ factor structure for excess returns, with time-varying loadings

$$r_t^e = B_t' F_t + e_t, \quad \text{with} \quad \mathbb{E}(e_t | F_t) = 0$$
$$F_t = W' r_t$$

- ▶ factors are linear combinations of excess returns, with weights based on accounting and market information

$$\Sigma_{t+1|t} = B_{t+1|t}' \Sigma_{f,t+1|t} B_{t+1|t} + \Sigma_{e,t+1|t}$$
$$\Sigma_{f,t+1|t} = W' \Sigma_{t+1|t} W$$

- ▶ factor covariances $\sim$ VHAR(1,5,22) by (ada)LASSO

step-by-step modeling

- ▶ factor structure for excess returns, with time-varying loadings

$$\begin{aligned} \mathbf{r}_t^e &= \mathbf{B}_t' \mathbf{F}_t + \mathbf{e}_t, & \text{with } \mathbb{E}(\mathbf{e}_t | \mathbf{F}_t) &= 0 \\ \mathbf{F}_t &= \mathbf{W}' \mathbf{r}_t \end{aligned}$$

- ▶ factors are linear combinations of excess returns, with weights based on accounting and market information

$$\begin{aligned} \Sigma_{t+1|t} &= \mathbf{B}_{t+1|t}' \Sigma_{f,t+1|t} \mathbf{B}_{t+1|t} + \Sigma_{e,t+1|t} \\ \Sigma_{f,t+1|t} &= \mathbf{W}' \Sigma_{t+1|t} \mathbf{W} \end{aligned}$$

- ▶ factor loadings $\sim \text{HAR}(1,5,22)$ by OLS

step-by-step modeling

- ▶ factor structure for excess returns, with time-varying loadings

$$\begin{aligned} r_t^e &= B_t' F_t + e_t, & \text{with } \mathbb{E}(e_t | F_t) &= 0 \\ F_t &= W' r_t \end{aligned}$$

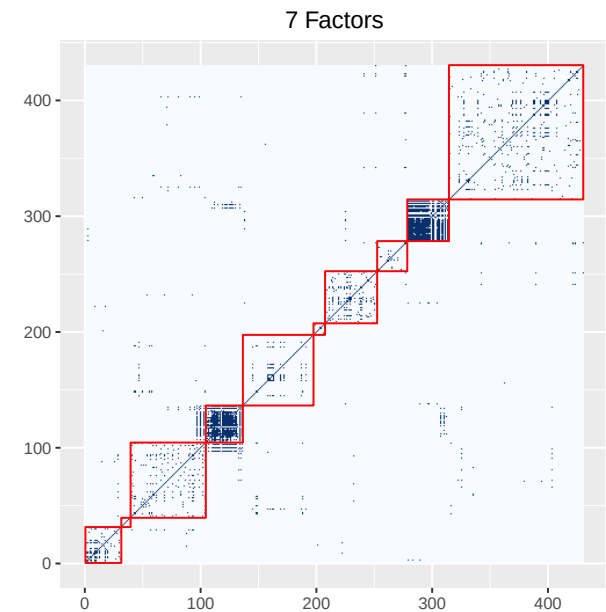
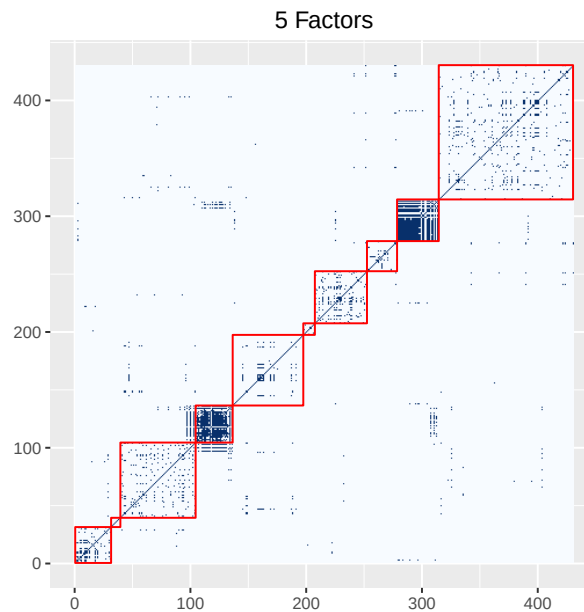
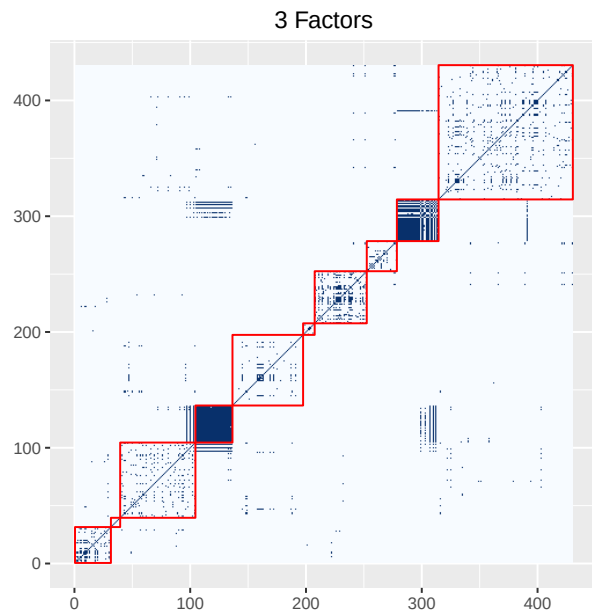
- ▶ factors are linear combinations of excess returns, with weights based on accounting and market information

$$\begin{aligned} \Sigma_{t+1|t} &= B_{t+1|t}' \Sigma_{f,t+1|t} B_{t+1|t} + \Sigma_{e,t+1|t} \\ \Sigma_{f,t+1|t} &= W' \Sigma_{t+1|t} W \end{aligned}$$

- ▶ residual covariances by sector **can** depend only on past variances

realized residual correlations

- ▶ S&P500 index constituents ($N = 430, S = 10$)
- ▶ January 2006 and December 2011 ($T = 1,495$ days)
- ▶ $W \sim$ (market, size, value, profitability, investment, asset growth, accruals)



forecasting performance

- ▶ rolling window of 1,000 observations
- ▶ 473 out-of-sample forecasts from $T_1 = 1,023$ to $T = 1,495$
- ▶ random walk vs VHAR with **LASSO** for factor covariance

$$\text{mean } \ell_2\text{-forecast error} = \frac{1}{T - T_1 + 1} \sum_{t=T_1}^T \left\| \text{vech}(\hat{\Sigma}_{f,t+1|t} - \Sigma_{f,t+1}) \right\|$$

factor structure	random walk	VHAR/RW	log(VHAR)/RW
market	0.40	0.96	0.92
+ size, value	0.44	0.98	0.90
+ profit, inv	0.51	0.95	0.89
+ growth, accruals	0.62	0.99	0.86

forecasting performance

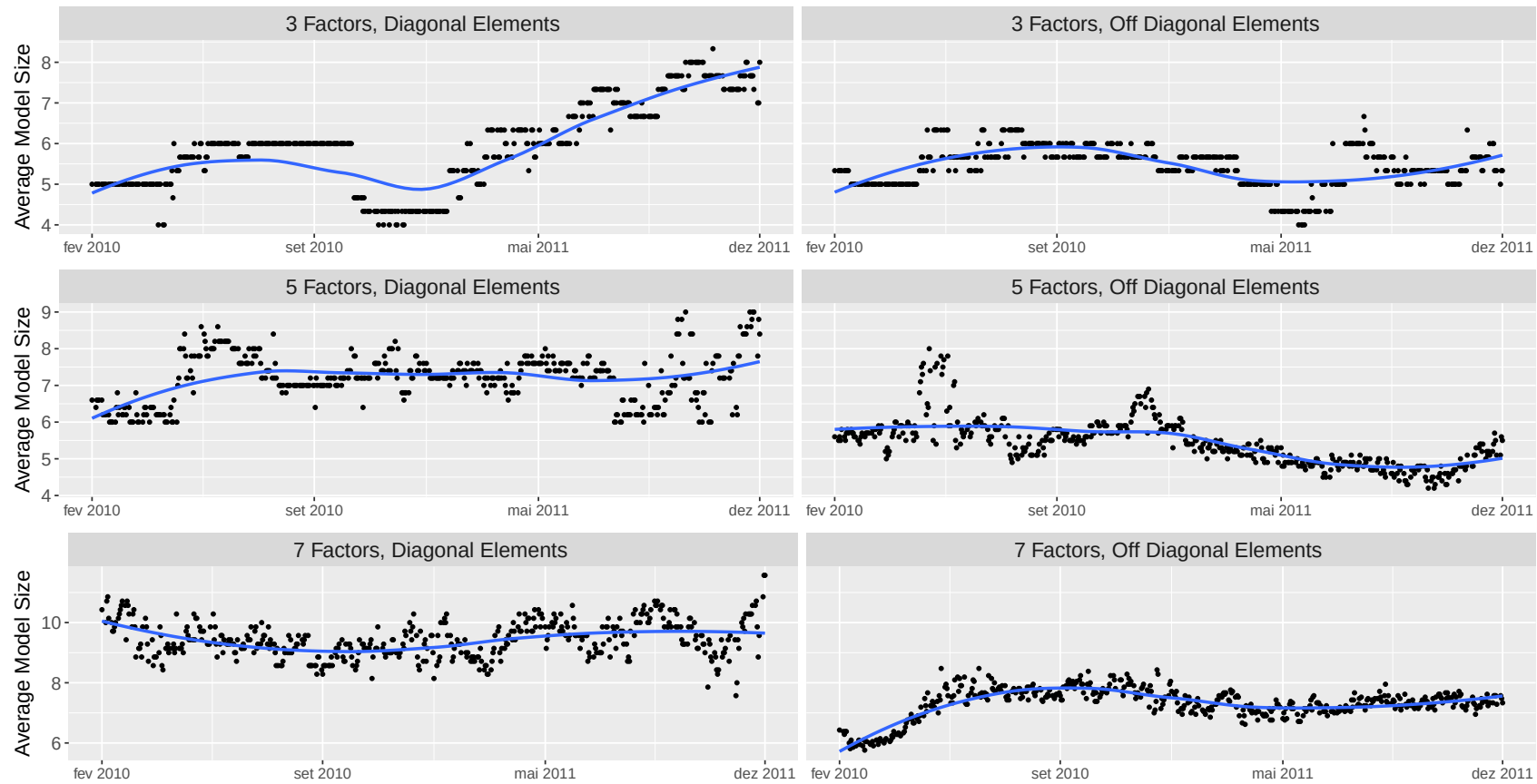
- ▶ rolling window of 1,000 observations
- ▶ 473 out-of-sample forecasts from $T_1 = 1,023$ to $T = 1,495$
- ▶ random walk vs VHAR with **adaLASSO** for factor covariance

$$\text{mean } \ell_2\text{-forecast error} = \frac{1}{T - T_1 + 1} \sum_{t=T_1}^T \left\| \text{vech}(\hat{\Sigma}_{f,t+1|t} - \Sigma_{f,t+1}) \right\|$$

factor structure	random walk	VHAR/RW	log(VHAR)/RW
market	0.40	0.96	0.92
+ size, value	0.44	0.97	0.90
+ profit, inv	0.51	0.95	0.89
+ growth, accruals	0.62	1.04	0.87

average number of predictors

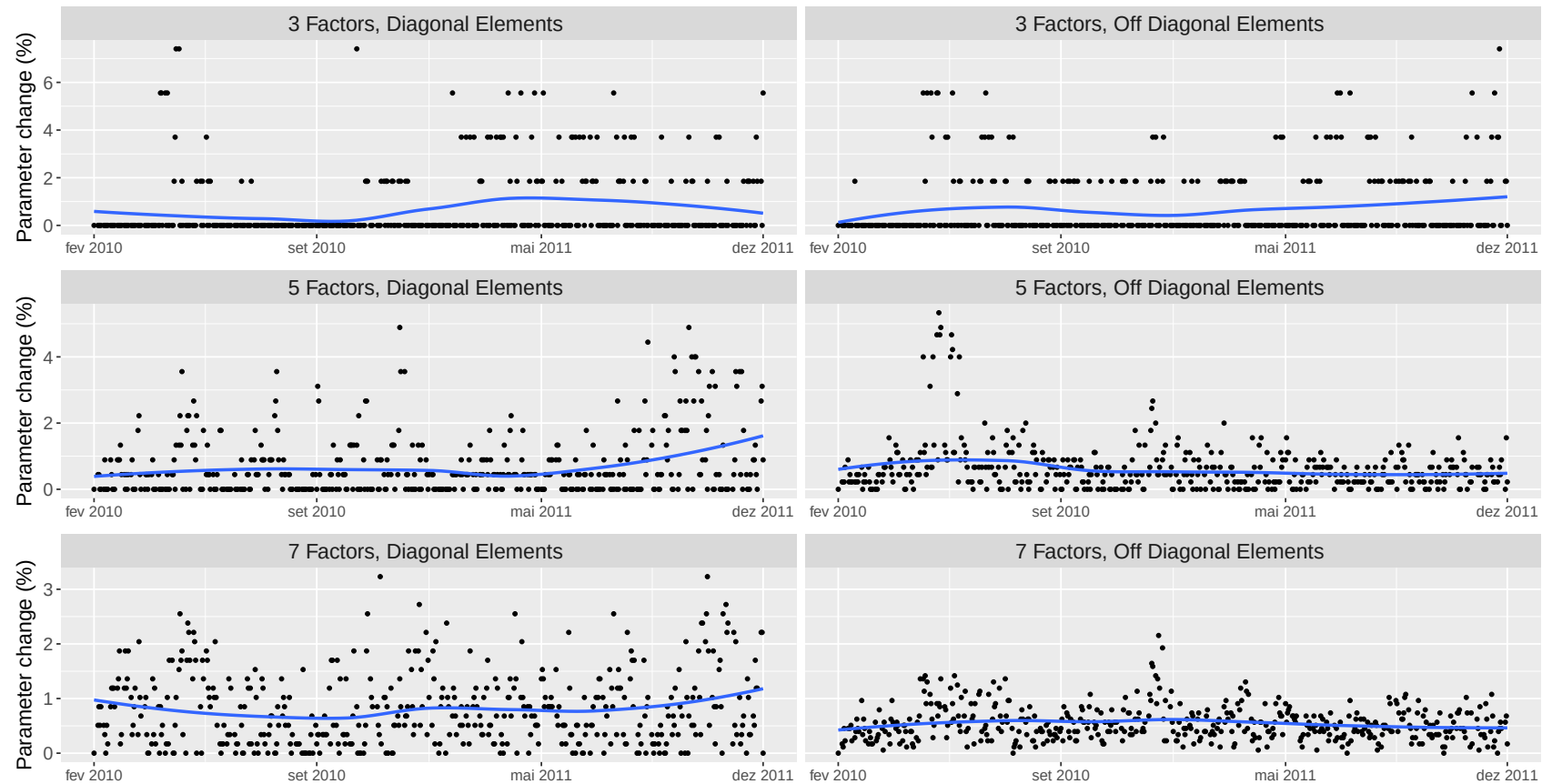
$\log(\text{VHAR})$



► substantial reduction in dimensionality

average change in predictors

$\log(\text{VHAR})$



► parameter estimates are stable over time

forecasting performance

- ▶ rolling window of 1,000 observations
- ▶ 473 out-of-sample forecasts from $T_1 = 1,023$ to $T = 1,495$
- ▶ random walk vs VHAR with **LASSO** vs DCC-NL

$$\text{mean } \ell_2\text{-forecast error} = \frac{1}{T - T_1 + 1} \sum_{t=T_1}^T \left\| \text{vech}(\hat{\Sigma}_{t+1|t} - \Sigma_{t+1}) \right\|$$

factor structure	random walk	DCC-NL/RW	log(VHAR)/RW
market	341.57	1.72	0.86
+ size, value	341.57	1.72	0.85
+ profit, inv	341.57	1.72	0.87
+ growth, accruals	341.57	1.72	0.86

forecasting performance

- ▶ rolling window of 1000 observations
- ▶ 473 out-of-sample forecasts from $T_1 = 1,023$ to $T = 1,495$
- ▶ random walk vs VHAR with [adaLASSO](#) vs DCC-NL

$$\text{mean } \ell_2\text{-forecast error} = \frac{1}{T - T_1 + 1} \sum_{t=T_1}^T \left\| \text{vech}(\hat{\Sigma}_{t+1|t} - \Sigma_{t+1}) \right\|$$

factor structure	random walk	DCC-NL/RW	log(VHAR)/RW
market	341.57	1.72	0.86
+ size, value	341.57	1.72	0.86
+ profit, inv	341.57	1.72	0.86
+ growth, accruals	341.57	1.72	0.86