Discussion of

"Forecasting and managing correlation risks"

by Bollerslev, Li, and Tang

Daniele Bianchi
Queen Mary, University of London

EFA 2024

Overview of the paper

Goal: Forecasting monthly pairwise realised correlation $\mathrm{RC}^{m}_{ij,t+1}$

$$\widehat{\mathrm{RC}}^{m}_{ij,t+1} = x'_{ij,t}\widehat{\boldsymbol{\theta}}$$

Overview of the paper

Goal: Forecasting monthly pairwise realised correlation $\mathrm{RC}^m_{ij,t+1}$

$$\widehat{\mathrm{RC}}^m_{ij,t+1} = x'_{ij,t}\widehat{\boldsymbol{\theta}}$$

Why? Correlations are key inputs for investment decisions:

↪ Hedging/diversification.

↪ Risk management.

Overview of the paper

Goal: Forecasting monthly pairwise realised correlation $\mathrm{RC}^m_{ij,t+1}$

$$\widehat{\mathrm{RC}}^m_{ij,t+1} = x'_{ij,t}\widehat{\boldsymbol{\theta}}$$

Why? Correlations are key inputs for investment decisions:

$\hookrightarrow$ Hedging/diversification.

$\hookrightarrow$ Risk management.

Issue(s): $x_{ij,t}$ can be large:

$\hookrightarrow$ Some variables may have little predictive ability.

$\hookrightarrow$ Risk of overfitting $\Longrightarrow$ sparsity vs shrinkage.

Overview of the paper

Assumption: only a subset of variables carry most of the predictive power:

$\hookrightarrow$ $\boldsymbol{\theta}$ is likely sparse.

Overview of the paper

Assumption: only a subset of variables carry most of the predictive power:

$\hookrightarrow$ $\boldsymbol{\theta}$ is likely sparse.

Variable selection via sparsity-inducing penalty (lasso):

$$\widehat{\boldsymbol{\theta}}_{\lambda} = \arg\min_{\boldsymbol{\theta}} \left(\underbrace{\sum_{t=1}^{T} \sum_{ij} \left(\mathrm{RC}^{m}_{ij,t+1} - x'_{ij,t}\boldsymbol{\theta} \right)^2}_{\text{loss function}} + \underbrace{\lambda \|\boldsymbol{\theta}\|_1}_{\text{penalty}} \right)$$

Overview of the paper

Assumption: only a subset of variables carry most of the predictive power:

$\hookrightarrow$ $\boldsymbol{\theta}$ is likely sparse.

Variable selection via sparsity-inducing penalty (lasso):

$$\widehat{\boldsymbol{\theta}}_\lambda = \arg\min_{\boldsymbol{\theta}} \left(\underbrace{\sum_{t=1}^{T} \sum_{ij} \left(\mathrm{RC}^m_{ij,t+1} - x'_{ij,t}\boldsymbol{\theta}\right)^2}_{\text{loss function}} + \underbrace{\lambda \|\boldsymbol{\theta}\|_1}_{\text{penalty}} \right)$$

Applications: $\widehat{\mathrm{RC}}^m_{ij,t+1}$ as an input of pairs trading, minimum-variance portfolios, etc.

My comments:

Comments/observations:

#1: Variable selection with correlated (!?) predictors.

#2: Asset quality and model performance.

#3: Testing performance differentials.

Comment #1: Lasso with correlated predictors

List of daily, weekly, and monthly predictors include:

↪ (1) Realised correlation and negative semicorrelation; (2) factor-driven realised correlation; (3) exponential realised correlation and negative semicorrelation; (4) sector-specific exp realised correlations and semicorrelations.

Comment #1: Lasso with correlated predictors

List of daily, weekly, and monthly predictors include:

↪ (1) Realised correlation and negative semicorrelation; (2) factor-driven realised correlation; (3) exponential realised correlation and negative semicorrelation; (4) sector-specific exp realised correlations and semicorrelations.

Predictors are all variations of pairwise correlation measures.

It is reasonable to believe these predictors carry similar information (potentially highly correlated!?);

↪ "Selection bias" of the lasso with correlated predictors.[1]

[1] Freijeiro-Gonzalez et al (2022). "A critical review of LASSO and its derivatives for variable selection under dependence among covariates." International Statistical Review.

Comment #1: Lasso with correlated predictors

Take a simple simulation example:

$\hookrightarrow$ DGP: $y = \theta_1 x_1 + \theta_2 x_2 + \text{noise}$

Comment #1: Lasso with correlated predictors

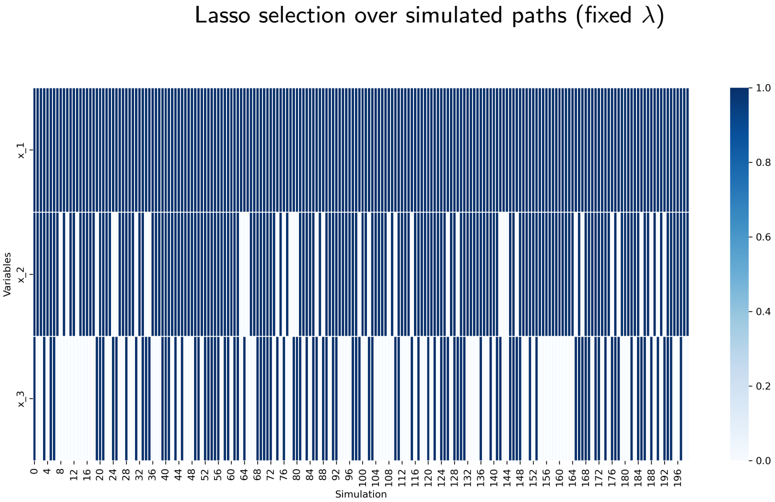
Take a simple simulation example:

↪ DGP: $y = \theta_1 x_1 + \theta_2 x_2 + \text{noise}$

↪ Model: $y = \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \text{noise}$

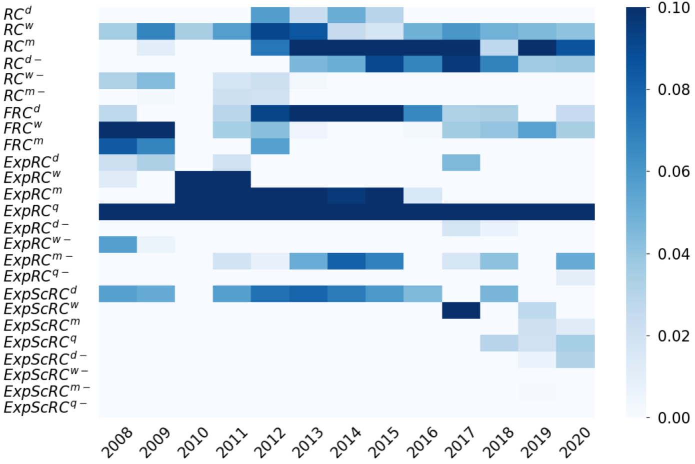
↪ x_1 strong, x_2 weak, x_3 irrelevant for y but correlated with x_2.

Comment #1: Lasso with correlated predictors



Comment #1: Lasso with correlated predictors

Something to consider for inference (Figure 3 in the version I have seen).



Comment #1: Lasso with correlated predictors

Do you really need sparsity? (4.2mln obs in training & 25 features)

↪ How does a simple OLS perform with these features?

↪ Are performances across methods really different?

Feature set	R^2_{OOS} relative to HAR Panel A: Equal-weighted				
	LASSO	Ridge	ENet	PCR	FNN
All 25 main features	10.16%	9.83%	10.14%	10.44%	10.12%
All 25 main features + 6 dummies (# of features = 31)	10.24%	9.96%	10.19%	9.61%	9.97%
All 25 main features + 150 feature × dummy combinations (# of features = 175)	10.35%	9.95%	10.31%	8.76%	9.88%

Comment #2: Asset quality and predictability

Value-weighted performance evaluation:

$$R^{2,VW}_{OOS} = 1 - \frac{\sum_{(ij,t)\in\mathcal{T}} \omega_{ij,t} \left(\mathrm{RC}^m_{ij,t} - \widehat{\mathrm{RC}}^m_{ij,t}\right)^2}{\sum_{(ij,t)\in\mathcal{T}} \omega_{ij,t} \left(\mathrm{RC}^m_{ij,t} - \widehat{\mathrm{RC}}^{m,\mathrm{HAR}}_{ij,t}\right)^2}$$

Comment #2: Asset quality and predictability

Value-weighted performance evaluation:

$$R^{2,VW}_{OOS} = 1 - \frac{\sum_{(ij,t)\in\mathcal{T}} \omega_{ij,t} \left(\mathrm{RC}^m_{ij,t} - \widehat{\mathrm{RC}}^m_{ij,t}\right)^2}{\sum_{(ij,t)\in\mathcal{T}} \omega_{ij,t} \left(\mathrm{RC}^m_{ij,t} - \widehat{\mathrm{RC}}^{m,\mathrm{HAR}}_{ij,t}\right)^2}$$

Goal: by weighting the forecast errors based on market cap, the role of small caps on R^2_{OOS} is downplayed.

Comment #2: Asset quality and predictability

Value-weighted performance evaluation:

$$R^{2,VW}_{OOS} = 1 - \frac{\sum_{(ij,t)\in\mathcal{T}} \omega_{ij,t} \left(\mathrm{RC}^m_{ij,t} - \widehat{\mathrm{RC}}^m_{ij,t}\right)^2}{\sum_{(ij,t)\in\mathcal{T}} \omega_{ij,t} \left(\mathrm{RC}^m_{ij,t} - \widehat{\mathrm{RC}}^{m,\mathrm{HAR}}_{ij,t}\right)^2}$$

Goal: by weighting the forecast errors based on market cap, the role of small caps on R^2_{OOS} is downplayed.

Yet, the model is estimated based on the same panel of assets:

↪ Slicing by mkt cap/illiq/vol may give further insights.

↪ Non-linear model to fit "lower-quality" assets?[2]

[2] Avramov et al. (2023) "Machine learning vs. economic restrictions: Evidence from stock return predictability." Management Science.

Comment #3: Testing performance differentials

Testing performance spreads via modified Diebold & Mariano (2002):

$$d_k = \frac{1}{N_k T} \sum_{ij,t} d_{ij,t} \mathbb{I}_{\{i=k \text{ or } j=k, i \neq j\}}, \qquad \text{with} \qquad d_{ij,t} = \left(\widehat{e}_{ij,t}^{(1)} \right)^2 - \left(\widehat{e}_{ij,t}^{(2)} \right)^2,$$

The DM statistics is defined as $t_{DM} = \overline{d} / \widehat{\sigma}_d$.

Comment #3: Testing performance differentials

Testing performance spreads via modified Diebold & Mariano (2002):

$$d_k = \frac{1}{N_k T} \sum_{ij,t} d_{ij,t} \mathbb{I}_{\{i=k \text{ or } j=k, i \neq j\}}, \qquad \text{with} \qquad d_{ij,t} = \left(\widehat{e}_{ij,t}^{(1)}\right)^2 - \left(\widehat{e}_{ij,t}^{(2)}\right)^2,$$

The DM statistics is defined as $t_{DM} = \overline{d}/\widehat{\sigma}_d$.

Intuitive, but:

↪ What are the asymptotic properties of t_{DM}?

↪ Consistent/efficient estimator of $\widehat{\sigma}_d$?

Comment #3: Testing performance differentials

Testing performance spreads via modified Diebold & Mariano (2002):

$$d_k = \frac{1}{N_k T} \sum_{ij,t} d_{ij,t} \mathbb{I}_{\{i=k \text{ or } j=k, i \neq j\}}, \qquad \text{with} \qquad d_{ij,t} = \left(\widehat{e}^{(1)}_{ij,t}\right)^2 - \left(\widehat{e}^{(2)}_{ij,t}\right)^2,$$

The DM statistics is defined as $t_{DM} = \overline{d}/\widehat{\sigma}_d$.

Intuitive, but:

- ↪ What are the asymptotic properties of t_{DM}?
- ↪ Consistent/efficient estimator of $\widehat{\sigma}_d$?

Qu et al. (2023) $\Longrightarrow$ forecast comparison in panel data (NW for $\widehat{\sigma}_d$).[3]

[3] Qu et al. (2023). "Comparing forecasting performance with panel data." International journal of forecasting.

Conclusion

Summary:

↪ Really interesting paper! I learnt a lot.

↪ Highly recommended reading.