Long range dependence and vector heterogeneous autoregressive models

- Basics properties of RVs
- LRD and HAR
- Factor augmented HAR and LSTM extension
- sparse VHAR models and extensions

Realized Volatility (RV)

- Volatility: usually we use (daily) log return

$$\log P_t - \log P_{t-1} \approx \frac{P_t - P_{t-1}}{P_{t-1}}$$

- Realized volatility (RV): use high frequency data for more precise measure of volatility (Andersen et al.;2003). It is the sum of intra-day squared returns

Realized Volatility at day t

$$RV_t = \sum_{j=0}^{M-1} r_{t,j}^2,$$

$$r_{t,j} = \log(P_{t-j/M}) - \log(P_{t-(j+1)/M}), \quad j = 1, \ldots, M.$$

RV of S&P 500 Index

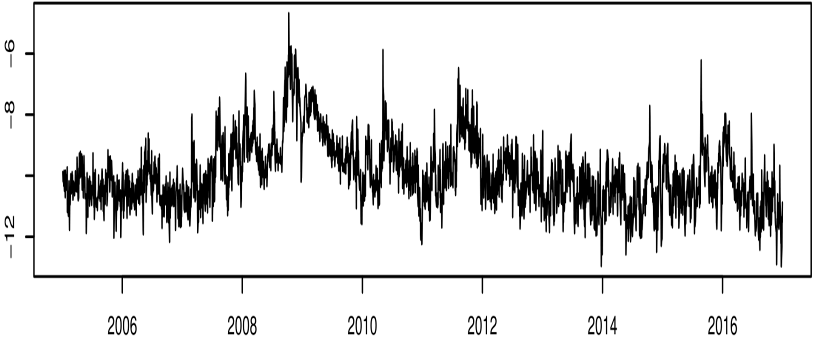


Figure: Time plot of log(Realized Volatilities) for S&P 500 Index from 2005 to 2016.

Long memory property of Realized Volatility

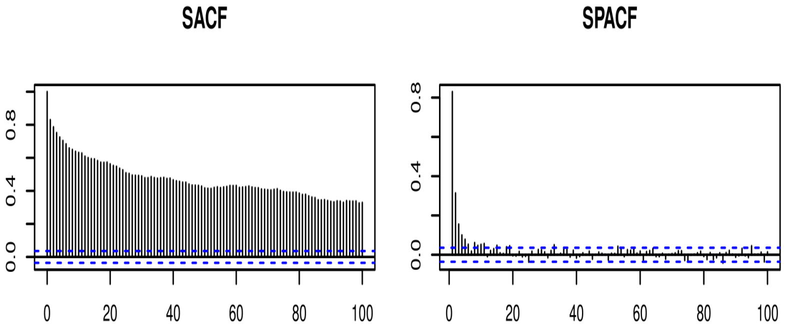


Figure: Sample ACF and PACF of log(RVs) for S&P 500 Index from 2005 to 2016.

Stylized facts on Realized Volatility

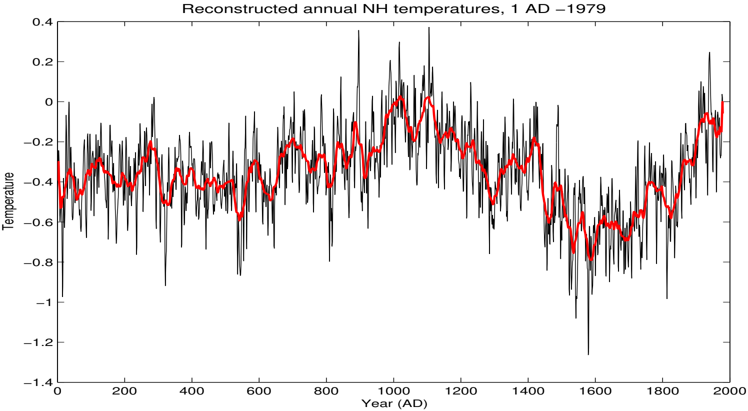
- Strong persistence in autocorrelations (or the long memory property).
- Time-varying conditional variance
- Non-Gaussianity
- Leverage effect
- Jumps, $\cdots$

We will see more on strong persistency in next slides. Such phenomenon is not only observed in finance. Recall that long memory is defined as the time series with very slow decaying of autocorrelations,

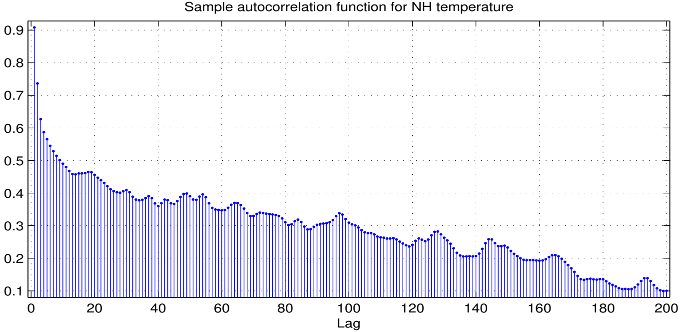
$$\gamma(h) \sim Ch^{2d-1}, \quad d \in (0, 1/2).$$

Northern Hemisphere temperature

Reconstructed data comes from Moberg et al. (2005), covering the period from 1 AD to 1979. Red curve represents 30 years moving average.



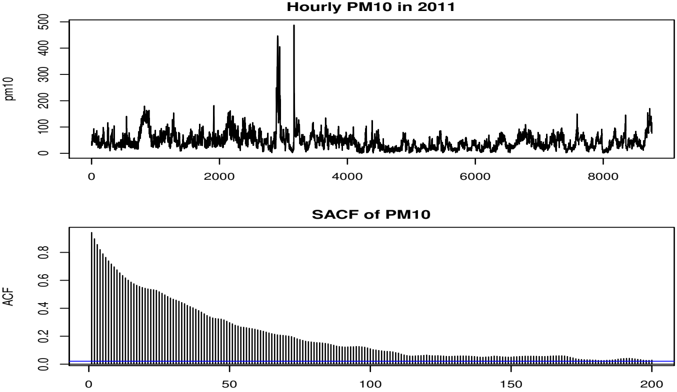
NH temperature data



Note a slow decay of the sample autocorrelation function!

Air pollution in Korea

Hourly observed PM10 (particular matter with diameter less than 10 $\mu m/m^3$) in 2011.



Note very slowly varying sample autocorrelations even for lag 100!

Long range dependence; Long memory

Long range dependence TS refers to a stationary TS with non ignorable long-term correlations.

Definition

A weakly stationary TS $\{X_t\}$ is called the long range dependent (LRD) if it has power-law decaying ACVF, that is,

$$\gamma_X(h) := \mathrm{Cov}(X_0, X_h) \sim Ch^{2d-1},$$

as $h \to \infty$ for $d \in (0, 1/2)$.

-
- Equivalently, in terms of $\rho_X(h) := \gamma_X(h)/\gamma_X(0)$,
$$\rho_X(h) \sim Ch^{2d-1}, \qquad \text{as } h \to \infty.$$
- $f \sim g$ means that two functions are ultimately equivalent in the sense that
$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1.$$

Long range dependence; Long memory

- Remark that d is called the LRD parameter, and

$$\boxed{d \in \left(0, \frac{1}{2}\right)}$$

for stationarity.

- If $d = 0$¡ then it is called short range dependence (SRD) implying that

$$\sum_{h=-\infty}^{\infty} |\gamma_X(h)| < \infty.$$

It covers popular ARMA(p, q) models. It can be shown for ARMA(p, q) seres that

$$|\rho(h)| < Cs^h = Ce^{h \log s}, \quad s \in (0, 1),$$

so that $\rho(h)$ decays exponentially fast as $h \to \infty$.

Modelling of LRD TS

- Recall Wold decomposition
  Every zero-mean weakly stationary TS $\{X_t\}$

$$X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} + \text{deterministic function}$$

- Since LRD is a stationary process, Wold decomposition implies that we can build linear process to model LRD, that is,

$$X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$$

- Q: How to incorporate long-term correlations?
  A: Need special structure on coefficients $\{\psi_j\}$

$I(d)$ process

Definition

The process $\{X_t\}_{t\in\mathbb{Z}}$ is called the fractionally integrated noise (I(d)) process if $\{X_t\}$ is a zero-mean stationary process satisfying

$$(1-B)^d X_t = \epsilon_t, \quad d \in (0, 1/2).$$

- The difference operator $\nabla^d = (1-B)^d$ is called the fractional differencing. Observe from the Taylor expansion that

$$(1-B)^d = 1 + d(-1)B + \frac{d(d-1)}{2!}(-1)^2 B^2 + \ldots.$$

By means of Binomial expansion,

$$(1-B)^d = \sum_{j=0}^{\infty} \binom{d}{j} (-1)^j B^j =: \sum_{j=0}^{\infty} \pi_j B^j,$$

$I(d)$ process

- Using Gamma function,

$$\Gamma(\alpha) := \int_0^\infty x^{\alpha-1} e^{-x} dx,$$

we can represent the coefficient as

$$\begin{cases} \pi_j & = \dfrac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} = \displaystyle\prod_{0<k\leq j} \frac{k-1-d}{k}, \quad j = 1, 2, \ldots, \\ \pi_0 & = 1 \end{cases}$$

- Therefore,

$$X_t = (1-B)^{-d}\epsilon_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j},$$

where

$$\psi_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} = \prod_{0<k\leq j} \left(\frac{k-1+d}{k}\right)$$

$I(d)$ process

- Using Stirling's formula

$$\Gamma(x) \approx \sqrt{2\pi} e^{-x+1} (x-1)^{x-\frac{1}{2}},$$

observe that as $j \to \infty$,

$$\psi_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} \approx \frac{\sqrt{2\pi} e^{-(j+d)+1} (j+d-1)^{j+d-\frac{1}{2}}}{\sqrt{2\pi} e^{-(j+1)+1} (j+1-1)^{j+1-\frac{1}{2}} \Gamma(d)}$$

$$= \frac{1}{\Gamma(d)} \frac{(j+d-1)^{j+d-\frac{1}{2}}}{j^{j+\frac{1}{2}}} e^{1-d} = \frac{1}{\Gamma(d)} \left(1+\frac{d-1}{j}\right)^{j+\frac{1}{2}} (j+d-1)^{d-1} e^{1-d}$$

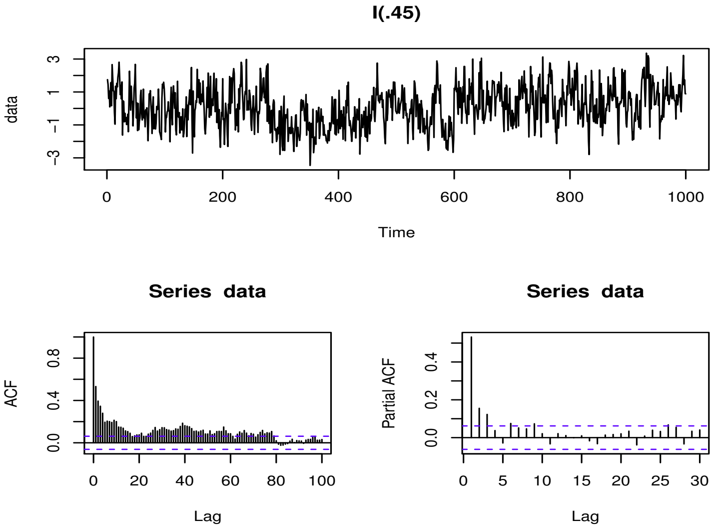
$$\approx \frac{1}{\Gamma(d)} e^{d-1} j^{d-1} \left(1+\frac{d-1}{j}\right)^{d-1} e^{1-d} = \frac{1}{\Gamma(d)} j^{d-1}$$

Thus, remark that the coefficient of $I(d)$ is given by

$$\psi_j \approx \frac{1}{\Gamma(d)} j^{d-1}$$

so that it decays very slowly, producing long-term correlations.

Simulated $I(.45)$



Properties if $I(d)$

- $\gamma(0) = \sigma^2\Gamma(1-2d)/\Gamma^2(1-d)$.
- The ACF is given by

$$\rho(h) = \frac{\Gamma(h+d)\Gamma(1-d)}{\Gamma(h-d+1)\Gamma(d)} = \prod_{0<k\leq h} \frac{k-1+d}{k-d}, \quad h = 1, 2, \ldots$$

- Using Stirling's formula, we obtain

$$\rho(h) \sim h^{2d-1}\frac{\Gamma(1-d)}{\Gamma(d)}$$

FARIMA(p, d, q) process

Definition

The process $\{X_t\}_{t\in\mathbb{Z}}$ is called the fractionally integrated ARMA (p, q) (FARIMA(p, q)) process if $\{X_t\}$ is stationary and satisfies

$$\phi(B)(1-B)^d X_t = \theta(B)\epsilon_t, \quad d \in (0, 1/2).$$

- $\nabla^d X_t$ is an ARMA(p, q) process.
- Furthermore, if $\theta(z) \neq 0$ for $|z| \leq 1$ so that it is invertible, then $Y_t = \phi(B)\theta^{-1}(B)X_t$ satisfies

$$\nabla^d Y_t = \epsilon_t$$
$$\phi(B)X_t = \theta(B)Y_t,$$

so we can also regard FARIMA as an ARMA(p, q) process driven by fractionally integrated noise.

MLE for FARIMA(p, d, q)

- Assume $\{\epsilon_1, ...\epsilon_n\}$ are Gaussian, then
$$(X_1, ..., X_n) \sim MVN(\mathbf{0}, \Sigma(\eta)),$$
where $\eta = (d, \sigma^2, \phi_1, ..., \phi_p, \theta_1, ..., \theta_q)$.
- MLE can be obtained as in ARMA(p, q) case by
$$\underset{\eta}{\operatorname{argmin}} \left\{ \log \left(\frac{1}{n} \sum_{j=1}^{n} \frac{\left(X_j - \widehat{X}_j\right)^2}{r_{j-1}} \right) + n^{-1} \sum_{j=1}^{n} \log r_{j-1} \right\},$$
where $r_{j-1} = \sigma^{-2} E(X_j - \widehat{X}_j)^2$, $j = 1, \ldots, n$ are the one step prediction errors. Innovations algorithm can be used for fast calculation of r_{j-1}.
- However, this method is quite slow because d is involved.
- For example, $\Sigma(\eta)$ is ill-conditioned when $d \approx \frac{1}{2}$

MLE for FARIMA(p, d, q)

- Numerically we can compute condition number

$$\frac{\lambda_{max}}{\lambda_{min}}$$

and larger condition number indicates that the inverse calculation is hard because it is close to singular.

d	.2	.3	. 4	.45
Condition Number	8.35	28.71	133.96	411.98

- Two popular "approximation" methods are
  i) Whittle's approximation of Gaussian MLE in spectral domain

$$\hat{\eta} = \underset{\eta}{\operatorname{argmin}} \left(\log \frac{1}{n} \sum_j \frac{I_n(\omega_j)}{f(\omega_j, \eta)} + n^{-1} \sum_j \log f(\omega_j, \eta) \right)$$

  ii) AR(∞) approximation based on the infinite past. As in ARMA case, replace $\widehat{X}_j$ by $\widetilde{X}_j$ and use $r_j \to 1$ as $n \to \infty$.

FARIMA(p, d, q) order selection

- Information criteria can be used in the model selection

$$
\begin{aligned}
(\mathrm{AIC}) &: -2\ell(\eta) + 2m \\
(\mathrm{BIC}) &: -2\ell(\eta) + 2\log n \\
(\mathrm{HQ}) &: -2\ell(\eta) + c\log(\log n), \quad c > 1
\end{aligned}
$$

- Diagnostics can be done by checking

$$
\hat{\epsilon}_j = \frac{X_j - \widehat{X}_j(\eta)}{\sqrt{r_{j-1}}} \approx \mathrm{WN}(0, \sigma^2)
$$

- Forecasting (Prediction) is exactly the same as ARMA. The BLP is given by minizing MSPE

$$
E(X_{n+h} - P_n X_{n+h})^2,
$$

where $P_n X_{n+h} = a_0 + a_1 X_n + \ldots + a_n X_1$.

Forecasting FARIMA(p, d, q)

- For large n, prediction based on infinite past gives

$$P_n X_{n+h} \approx \widetilde{X}_{n+h} = -\sum_{j=1}^{\infty} \pi_j \widetilde{X}_{n+h-j}$$

$$\widetilde{\sigma}_n^2(h) = E(X_{n+h} - \widetilde{X}_{n+h})^2 = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2,$$

  where $\sum_{j=0}^{\infty} \psi_j z^j = \theta(z)\phi^{-1}(1-z)^{-d}$ and $\sum_{j=0}^{\infty} \pi_j z^j = \phi(z)\theta^{-1}(z)(1-z)^d$.
- For FARIMA$(0, d, 0)$, $(1-B)^d X_t = \epsilon_t$, model observe that

$$\pi_j \sim c j^{-d-1}$$

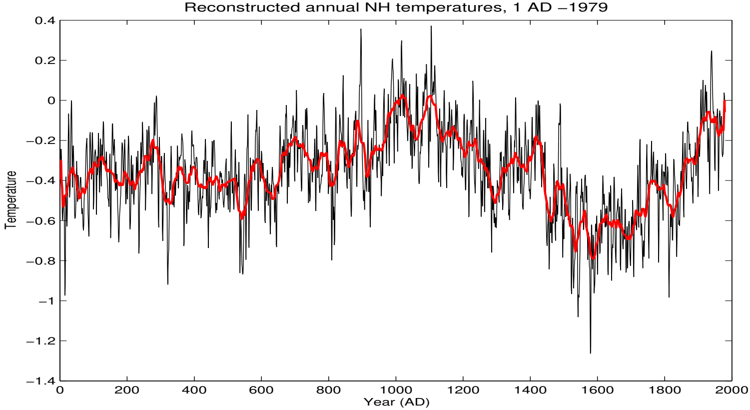
  implies that the influence of past observations decays as slow as a power function. In fact, finite past calculation gives

$$a_n = \frac{d}{n-d} \sim \frac{d}{n}$$

  so that it decays much slower than infinite past case.

Example: NH using `fracdiff`

Recall the NH temperature data reconstructed from 1 AD to 1979.



Example: NH using `fracdiff`

BIC gives the best model as FARIMA(2,d,2)

```
> temp_cen = temp - mean(temp)
> out = fracdiff(temp_cen, nar=2, nma=2)
> summary(out)

Call:
fracdiff(x = temp_cen, nar = 2, nma = 2)

Coefficients:
Estimate Std. Error z value Pr(>|z|)
d    0.44011    0.01731   25.42   <2e-16 ***
ar1  0.31866    0.03145   10.13   <2e-16 ***
ar2 -0.25286    0.02528  -10.00   <2e-16 ***
ma1 -1.50617    0.02672  -56.38   <2e-16 ***
ma2 -0.60350    0.02594  -23.27   <2e-16 ***
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1
sigma[eps] = 0.04843304
[d.tol = 0.0001221, M = 100, h = 3.354e-05]
Log likelihood:  3183 ==> AIC = -6353.015 [6 deg.freedom]
```

Example: NH using `fracdiff`

- The best model is written as

$$(1-.32B+.25B^2)(1-B)^{.44}(X_t+.3537) = (1+1.51B+.60B^2)\epsilon_t,$$

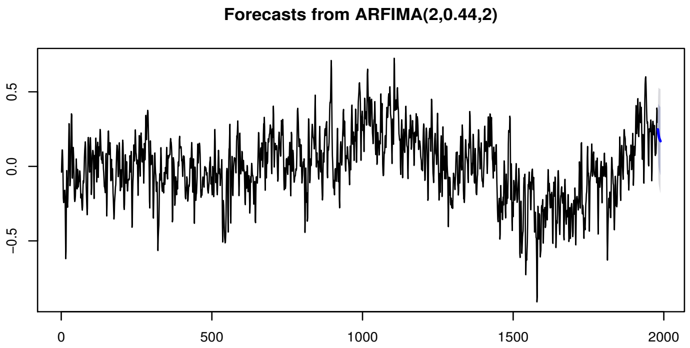
  where $\epsilon_t \sim \mathcal{N}(0, .048^2)$. Be careful that `fracdiff` uses slightly different notations for MA part.
- Forecasting is also easy using `forecast()` package.

```
> forecast(out, n.ahead=20)
Point Forecast        Lo 80      Hi 80        Lo 95      Hi 95
1980      0.2486070  0.18657254 0.3106414  0.153733492 0.3434805
1981      0.2338673  0.08027805 0.3874566 -0.001027194 0.4687618
1982      0.2193068  0.01981672 0.4187969 -0.085786958 0.5244006
1983      0.2061066 -0.00437106 0.4165842 -0.115791196 0.5280044
1984      0.1964311 -0.01761045 0.4104727 -0.130917211 0.5237794
1985      0.1893211 -0.02806719 0.4067095 -0.143145621 0.5217879
1986      0.1833164 -0.03772827 0.4043611 -0.154742271 0.5213751
1987      0.1778564 -0.04632732 0.4020401 -0.165003005 0.5207157
1988      0.1729220 -0.05371963 0.3995636 -0.173696453 0.5195404
1989      0.1685094 -0.06016942 0.3971882 -0.181224667 0.5182434
```

Example: NH using `fracdiff`

Plot gives the following:

```
> plot(forecast(out, n.ahead=20))
```



Modeling of RV

- (Genuine) Long memory model: fractionally integrated ARMA models (Andersen et al. 2003)
- Corsi (2004) proposed simple AR model for approximate long memory model:

$$X_t^{(d)} = c + \beta^{(d)} X_{t-1}^{(d)} + \beta^{(w)} X_{t-1}^{(w)} + \beta^{(m)} X_{t-1}^{(m)} + \omega_t, \quad (1)$$

where ω_t IID $\mathcal{N}(0, \sigma^2)$, $X_t^{(d)}$ is today's RV and

$$X_{t-1}^{(w)} = \frac{1}{5} \sum_{j=1}^{5} X_{t-j}^{(d)}, \quad X_{t-1}^{(m)} = \frac{1}{22} \sum_{j=1}^{22} X_{t-j}^{(d)} \quad (2)$$

- The second and third term represents weekly and monthly averages. This is a constrained AR(22) model.

HAR extensions

- Original HAR model is exted to HAR-GARCH for conditional heteroscedascity (Corsi et al.; 2008).
- Liu and Maheu (2008) further extended HAR-GARCH model with jump and asymmetric effects.
- We consider multivariate extension of HAR models
  - Factor augmentation
  - Sparse VHAR (sVHAR) model
  - sVHAR with non-convex penalties
  - sVHAR with GARCH erros

FAHAR

- Inspired by Bernanke et al. (2005), factor augmented HAR (FAHAR) model that incorporates factors obtained from multinational volatility.
- HAR model has some disadvangages; poor prediction of abrupt fluctuation and no coverage of RV non-linearity.
- We conjecture that abrupt fluctuations are partly due to the unstable and rapid swing of foreign stock markets; hence factors may adjust such effects and improve forecasting performance.
- Let $F_t^{(j)}$ be the j-th factor calculated from foreign RVs associated with the RV of interest. Then, FAHAR model is given by

$$X_t^{(d)} = c + \beta^{(d)} X_{t-1}^{(d)} + \beta^{(w)} X_{t-1}^{(w)} + \beta^{(m)} X_{t-1}^{(m)} + \sum_{j=1}^{r} \gamma_j F_{t-1}^{(j)} + \omega_t. \qquad (3)$$

FAHAR - PCA factor calculation

- Factor is calculated from PCA. Diagonalizing the sample correlation matrix $S = T^{-1} \sum_{t=1}^{T} X_t X_t' = UDU'$ where $D = \operatorname{diag}(d_1, \ldots, d_q)$ with $d_1 \geq d_2 \ldots \geq d_q$ and $U = (u_1, \ldots, u_q)$ is the orthogonal eigenvector matrix such that $UU' = U'U = I_q$. The principal factors are obtained by taking the r largest eigenvalues:

$$F_t = \frac{1}{q} \Lambda' Y_t, \quad \Lambda = \sqrt{q}(u_1, \ldots, u_r), \quad t = 1, \ldots, T.$$

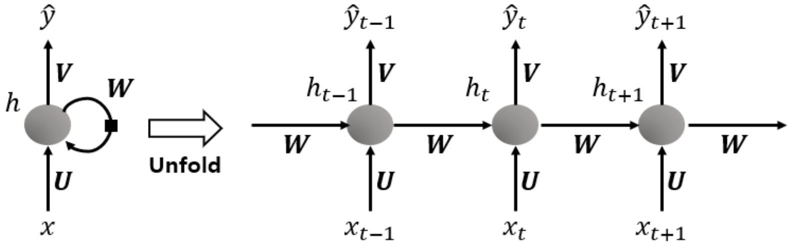
LSTM-FAHAR

We can also add nonlinearity by considering LSTM in deep learning. Namely,

- FAHAR-LSTM

$$X_{t+1} = f\left(1, X_t^{(d)}, X_t^{(w)}, X_t^{(m)}, F_t\right)$$

using long and short memory (LSTM) deep learning network.



Empirical result: FAHAR

- Daily volatility from the Oxford-Man Institute of Quantitative Finance (`http://realized.oxford-man.ox.ac.uk`).
- RV by aggregating 5-min within-day returns from January 4, 2010 to June 30, 2017 (post US housing crisis).
- 20 worldwide stock indices: KOSPI, S&P 500, FTSE, NIKKEI, DAX, Russell, All Ordinaries, DJIA, NASDAQ, CAC, Hang Seng, AEX, SIX Swiss, IBEX, S&P CNX Nifty 50, IPC Mexico, Brazil BOVESPA, S&P/TSX, EURO STOXX 50, and FTSE MIB.
- One-step-ahead out-of-sample forecasting is used for comparison. Mean squared prediction error (MSPE), mean quasi-likelihood function (MQLIKE), mean absolute relative error (MARE) and mean absolute scaled error (MASE) are used (Hyndman; 2006).

Empirical result: FAHAR

- Two-factors are used from the scree plot. But, the BIC criteria of Bai and Ng (2008) also suggests either 1 or 2 factors.

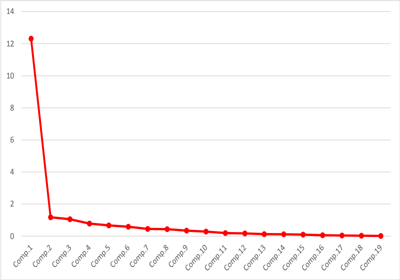


Figure: Scree plot for 19 RVs except the KOSPI. Elbow point 2 is selected for the number of factors.

Empirical result: FAHAR

	MSPE			MQLIKE		
	HAR	FAHAR	FAHAR-LSTM	HAR	FAHAR	FAHAR-LSTM
KOSPI	1.2881	1.3117	**1.2624**	0.0246	0.0264	**0.0243**
Nikkei	2.5884	**2.3194**	2.6121	0.0410	**0.0353**	0.0410
Hangseng	2.1070	**1.9437**	1.9675	0.0335	**0.0313**	0.0314
FTSE	0.8043	**0.7442**	0.8329	0.0234	**0.0222**	0.0253
DJIA	1.1702	**1.1335**	1.8923	0.0474	**0.0444**	0.0636
S&P	1.1394	**1.1053**	1.2620	0.0467	**0.0447**	0.0581
	MAPE			**MASE**		
	HAR	FAHAR	FAHAR-LSTM	HAR	FAHAR	FAHAR-LSTM
KOSPI	0.1727	**0.1588**	0.1759	0.8830	**0.8616**	0.8938
Nikkei	0.2681	**0.2174**	0.2577	1.0338	**0.8874**	1.0116
Hangseng	0.2102	**0.1834**	0.2003	0.8423	**0.7611**	0.8120
FTSE	0.1923	**0.1701**	0.1793	0.8760	**0.8116**	0.8541
DJIA	0.3176	**0.2888**	0.3639	1.1202	**1.0547**	1.3272
S&P	0.3054	**0.2766**	0.3494	1.0999	**1.0372**	1.2620

Table: One-step-ahead forecasting performance comparison for Asian and leading markets.

Empirical result: FAHAR

- In Asian markets such as Korea, Hong Kong, and Japan, the FAHAR or FAHAR-LSTM models provide better forecasting performance, meaning information from leading markets does indeed improve forecasting.
- The improvement from the LSTM network is only tangible for Korean markets; however, the overlaid plot shows that it has the power to follow sudden changes in financial markets.
- Hence, we may model multiple RVs in one large model. It leads to sparse VHAR models.

Sparse VHAR

- Vector heterogeneous autoregressive (VHAR) model is given by
$$Y_t^{(d)} = \Phi^{(d)} Y_{t-1}^{(d)} + \Phi^{(w)} Y_{t-1}^{(w)} + \Phi^{(m)} Y_{t-1}^{(m)} + \varepsilon_t, \qquad (4)$$
where $\varepsilon_t \sim WN(0, \Sigma), t = 23, \dots, T$ and $Y_{t-1}^{(w)} = 1/5 \sum_{j=1}^{5} Y_{t-j}^{(d)}$, $Y_{t-1}^{(m)} = 1/22 \sum_{j=1}^{22} Y_{t-j}^{(d)}$ are respectively the component-wise weekly and monthly average RVs.
- As dimension k grows, the number of VHAR parameters would increase quadratically, as $3k^2$ (without innovations Σ).
- Regularized estimation using (adaptive) lasso is introduced.

Sparse VHAR

- Rewrite VHAR as
$$\mathbb{Y} = \mathbb{A}\mathbb{X} + \mathbb{Z}, \tag{5}$$
where $\mathbb{Y} = (Y_{23}^{(d)}, \ldots, Y_T^{(d)})$, $\mathbb{A} = (\Phi^{(d)}, \Phi^{(w)}, \Phi^{(m)})$, and $\mathbb{X} = (X_{22}, \ldots, X_{T-1})$, with
$$X_t = \begin{pmatrix} Y_t^{(d)} \\ Y_t^{(w)} \\ Y_t^{(m)} \end{pmatrix}$$
and $\mathbb{Z} = (\varepsilon_{23}, \ldots, \varepsilon_T)$.
- Vectorizing gives further
$$\mathbf{Y} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{z}, \tag{6}$$
where response $\mathbf{Y} = \text{vec}(\mathbb{Y})$, the design matrix $\mathbf{X} = (\mathbb{X}' \otimes I_k)$, the parameter vector is $\boldsymbol{\alpha} = \text{vec}(\mathbb{A})$, and $\mathbf{z} = \text{vec}(\mathbb{Z})$.

Sparse VHAR: adaptive lasso

- The adaptive lasso estimation procedure

$$\widehat{\boldsymbol{\alpha}}^{AL} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \|(I_T \otimes \Sigma^{-1/2})\mathbf{Y} - (\mathbb{X}' \otimes \Sigma^{-1/2})\boldsymbol{\alpha}\|_2^2 + \lambda \|\mathbf{w}'\boldsymbol{\alpha}\|_1, \tag{7}$$

  where $\|\cdot\|_p$ is L_p-norm, and $\mathbf{w}$ is the adaptive weight vector.
  Initial estimators:

$$\widehat{\Sigma} = \frac{1}{T-22}\left(\mathbb{Y} - \widehat{\mathbb{A}}\mathbb{X}\right)\left(\mathbb{Y} - \widehat{\mathbb{A}}\mathbb{X}\right)', \tag{8}$$

  where $\widehat{\mathbb{A}}$ is obtained by OLS estimator

$$\widehat{\boldsymbol{\alpha}}^{OLS} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \|\mathbf{Y} - \mathbf{X}\boldsymbol{\alpha}\|_2^2 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{9}$$

- Note that it is differ from usual (adaptive) lasso estimation in the sense that Σ is incorporated.

Sparse VHAR: adaptive lasso

- Under mild assumptions, we can show the consistency of lasso estimator. It is basically a constrained VAR(22) model. See Proposition 1 of Baek and Kim (2020). For example,

$$\|\widehat{\boldsymbol{\alpha}}^{AL} - \boldsymbol{\alpha}\|_2 \leq 16\sqrt{\ell}\lambda_T/\alpha_R$$

- One of the key assumption is that the innovation covariance matrix $\Sigma = (\sigma_{ij})$ satisfies $\sigma_{ii} - \sum_{j \neq i} \sigma_{ij} > 0$ for $i = 1, \ldots, k$.
- Penalty parameter λ is chosen from either n-fold cross validation or BIC (Chen and Chen; 2008).
- To preserve time series feature, we applied the so-called block n-fold cross validation. For example 5-fold cross validation with sample size $T = 100$, divide the entire data into 5 subsamples, $(X_1, \ldots, X_{20})$, $(X_{21}, \ldots, X_{40})$,..., $(X_{81}, \ldots, X_{100})$, and calculate validation errors based on these subsamples.

Finite sample performance: sparse VHAR

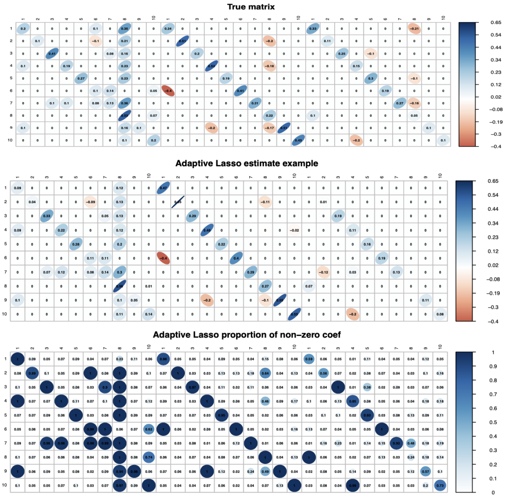


Figure: The true coefficient matrices (top), exemplary estimated coefficient matrices (middle), the proportion of selecting non-zero coefficients our of 1000 replications (bottom).

Empirical analysis: sparse VHAR

- The same 20 multinational RVs from Oxford-Man Institute of Quantitative Finance. But it ranges from January 4, 2010, to December 29, 2017. 1-step-ahead out-of-sample forecasting:

	VHAR			Univariate HAR			Sparse-VHAR		
	MSPE	MAPE	MASE	MSPE	MAPE	MASE	MSPE	MAPE	MASE
S&P 500	0.0875	4.1740	1.0175	0.0792	4.0007	0.9740	0.0796	3.9479	0.9593
FTSE 100	0.0449	3.0301	0.8227	0.0455	3.0757	0.8334	0.0412	2.9064	0.7855
Nikkei 225	0.0636	3.5591	0.8618	0.0662	3.6476	0.8846	0.0635	3.5182	0.8495
DAX	0.0697	4.0519	0.8458	0.0694	3.9568	0.8261	0.0667	3.9342	0.8169
Russel 2000	0.0841	4.4125	0.8541	0.0850	4.3819	0.8468	0.0798	4.3141	0.8307
All Ord	0.0703	3.9266	0.8291	0.0676	3.7402	0.7906	0.0663	3.7270	0.7839
DJIA	0.0851	4.0068	0.9625	0.0764	3.8718	0.9283	0.0765	3.8361	0.9181
Nasdaq 100	0.0866	4.2078	0.9508	0.0842	4.1255	0.9319	0.0830	4.0639	0.9138
CAC 40	0.0536	3.6093	0.8996	0.0539	3.5646	0.8886	0.0508	3.5581	0.8814
Hang Seng	0.0591	3.4614	0.7861	0.0606	3.4769	0.7944	0.0569	3.3437	0.7585
KOSPI	0.0460	2.9988	0.8632	0.0451	2.9574	0.8528	0.0462	3.0261	0.8728
AEX Index	0.0541	3.5145	0.9053	0.0554	3.5148	0.9046	0.0500	3.4053	0.8721
Swiss Market	0.0279	2.3907	0.8439	0.0283	2.4548	0.8653	0.0259	2.3114	0.8121
IBEX 35	0.0506	3.6476	0.8506	0.0518	3.6420	0.8546	0.0487	3.5677	0.8313
S&P CNX	0.0750	4.0657	0.8804	0.0741	3.9740	0.8607	0.0717	3.9374	0.8509
IPC Mexico	0.0661	3.9029	0.8449	0.0561	3.6218	0.7833	0.0606	3.6998	0.7964
Bovespa	0.0615	3.7333	0.8973	0.0647	3.7938	0.9206	0.0611	3.7128	0.8977
S&P TSX	0.0640	3.6792	0.8600	0.0645	3.6115	0.8453	0.0622	3.5760	0.8368
Euro STOXX	0.0731	3.9943	0.7876	0.0792	4.1627	0.8229	0.0692	3.9618	0.7765
FTSE MIB	0.0501	3.4884	0.8895	0.0488	3.4774	0.8883	0.0467	3.4130	0.8675
Average	0.0636	3.6927	0.8726	0.0628	3.6526	0.8649	0.0603	3.5880	0.8456

Table 6: Forecasting performance of the full VHAR, univariate HAR, and sparse VHAR models for 20 multinational RVs

Empirical analysis: sparse VHAR

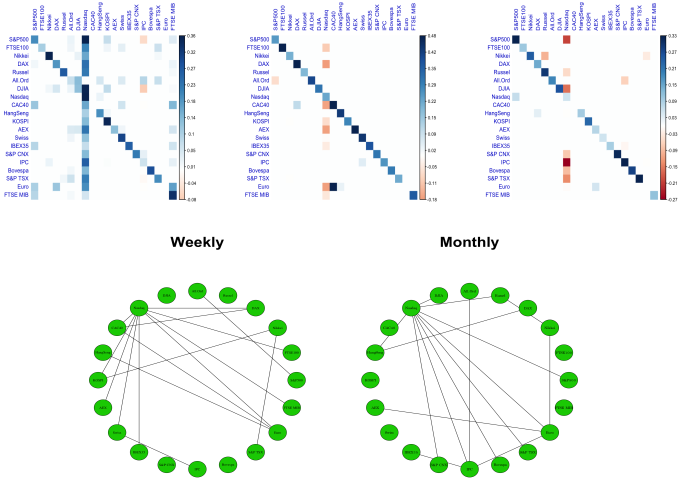


Figure: Coefficient matrix (top) and network diagram for weekly and monthly coefficient matrices. NASDAQ is the center of the graph.

Empirical analysis: sparse VHAR

- The volatility of the NASDAQ market has the strongest weekly and monthly influence on the other stock markets. This clearly indicates the leading nature of the NASDAQ market, and that researchers should closely monitor the dynamics of its volatility linkages.

Debiasing? Two stage procedure for adaptive lasso

To reduce bias in adaptive lasso estimation, two-stage is proposed. First, find non-zero coefficients from adaptive lasso. Then, apply constrained GLS to get the final estimates.

$$\hat{\boldsymbol{\alpha}}^{GLS} = R((\mathbf{X}R)'(I_{T-22} \otimes \hat{\Sigma}^{-1})(\mathbf{X}R))^{-1}(\mathbf{X}R)'(I_{T-22} \otimes \hat{\Sigma}^{-1})\mathbf{Y}, \tag{10}$$

where the sparsity constraint is $\boldsymbol{\alpha} = R\boldsymbol{\gamma}$. For example,

$$\alpha := \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ 0 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} =: R\boldsymbol{\gamma}.$$

Debiasing? SCAD

- Penalization method can be rewritten as

$$\hat{\alpha} = \underset{\alpha \in \mathbb{R}^q}{\text{argmin}} \frac{1}{N} \|\mathbf{Y} - \mathbf{X}\boldsymbol{\alpha}\|_2^2 + \sum_{j=1}^{q} p_\lambda(|\alpha_j|), \tag{11}$$

- Shin, Park and Baek (2022) consider non-convex penalties (which is known to reduce bias) and compared with two-stage procedure.
- For example, SCAD (Fan and Li (2001)) penalty is given by

$$\frac{d}{d\theta}\mathbf{p}_\lambda^{\mathbf{SCAD}}(\theta) = \lambda\left\{I(\theta \le \lambda) + \frac{(\gamma\lambda - \theta)_+}{(\gamma - 1)\lambda} I(\theta > \lambda)\right\} = \begin{cases} \lambda & , |\theta| \le \lambda, \\ \frac{\gamma\lambda - |\theta|}{\gamma - 1} & , \lambda < |\theta| < \gamma\lambda, \\ 0 & , \gamma\lambda \le |\theta|, \end{cases} \tag{12}$$

for some $\gamma > 2$ and $\theta > 0$.

Debiasing? MCP

- More generally, MCP (Breheny and Huang (2011)) can be considered with penalty function

$$\frac{d}{d\theta}\mathbf{p}_{\lambda}^{\mathbf{MCP}}(\theta) = \begin{cases} \mathrm{sign}(\theta)(\lambda - \frac{|\theta|}{\gamma}), & |\theta| \le \gamma\lambda, \\ 0, & |\theta| > \gamma\lambda, \end{cases} \quad (13)$$

where γ is usually set greater than 1.

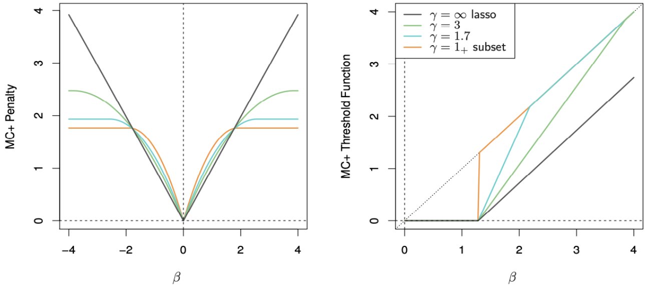


Figure: Hastie,Tibshirani and Wainwright (2015).

Finite sample performance: sVHAR with SCAD/MCP

- We are interested in the finite sample performance of penalization method. In particular, in terms of debiasing. Two stage vs SCAD/MCP?
- Simulation model: Gaussian VHAR model with $k = 10$ and coefficients

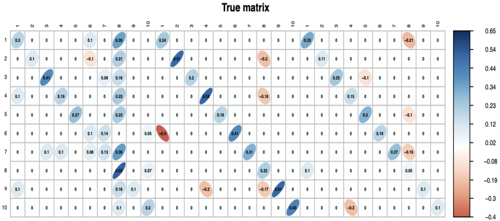


Figure: The true coefficient matrices.

Finite sample performance: DGP

- Spatial dependence is controlled by

$$\varepsilon_t := (\varepsilon_{t,1}, \ldots, \varepsilon_{t,10})' \sim N(0, \Sigma_z), \tag{14}$$

where

$$\Sigma_z = \begin{pmatrix} \delta^2 & \delta/12 & \delta/12 & \delta/16 & \delta/16 & \delta/20 & \delta/20 & \delta/24 & \delta/24 & \delta/28 \\ \delta/12 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \delta/12 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \delta/16 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \delta/16 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ \delta/20 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \delta/20 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \delta/24 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ \delta/24 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ \delta/28 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

with $\delta \in \{1, 2, 5, 10\}$.

- Performance measures:

$$\boldsymbol{RMSE} = \mathbb{E}\|\hat{\alpha} - \alpha\|_2, \quad \boldsymbol{ME} = \mathbb{E}\|1_{\{\hat{\alpha}=0\}} - 1_{\{\alpha=0\}}\|_2^2/(3k^2).$$

Simulation result : sparse VHAR

Table 2: Summary of performance measures for DGP2 with sample size 250.

		DGP2 / $\Sigma = I_k$			DGP2 / update Σ		
		eBIC			**eBIC**		
δ		**Adpative Lasso**	**SCAD**	**MCP**	**Adpative Lasso**	**SCAD**	**MCP**
1	**RMSE**	0.1084	0.0899	0.0969	0.1099	0.0910	0.0977
	ME	0.2044	0.1597	0.1784	0.2086	0.1630	0.1791
2	**RMSE**	0.1162	0.0872	0.0959	0.1114	0.0889	0.0974
	ME	0.2062	0.1561	0.1757	0.2085	0.1609	0.1777
5	**RMSE**	0.2674	0.0952	0.1088	0.1555	0.0925	0.1064
	ME	0.2107	0.1574	0.1729	0.2070	0.1580	0.1746
10	**RMSE**	0.5725	0.1717	0.1833	0.2836	0.1797	0.1847
	ME	0.2165	0.1683	0.1757	0.1623	0.1708	0.1767
		BLOCK-CV			**BLOCK-CV**		
δ		**Adaptive Lasso**	**SCAD**	**MCP**	**Adaptive Lasso**	**SCAD**	**MCP**
1	**RMSE**	0.0867	0.0823	0.0825	0.0934	0.0801	0.0802
	ME	0.1885	0.1854	0.1643	0.2104	0.1882	0.1661
2	**RMSE**	0.0968	0.0815	0.0821	0.0929	0.0792	0.0792
	ME	0.1802	0.1750	0.1604	0.2048	0.1848	0.1651
5	**RMSE**	0.1777	0.1095	0.1147	0.0970	0.0790	0.0799
	ME	0.1614	0.1631	0.1679	0.1935	0.1818	0.1670
10	**RMSE**	0.2889	0.1048	0.1087	0.1101	0.0995	0.1014
	ME	0.2111	0.1594	0.1667	0.1875	0.1944	0.1907

When the sample size is small, SCA/MCP tends to perform better than two-stage adaptive lasso.

Simulation result : sparse VHAR

Table 3: Summary of performance measures for DGP2 with sample size 500.

		DGP2 / $\Sigma = I_k$			DGP2 / update Σ		
		eBIC			eBIC		
δ		Adpative Lasso	SCAD	MCP	Adpative Lasso	SCAD	MCP
1	**RMSE**	0.0664	0.0865	0.0939	0.0659	0.0878	0.0949
	ME	0.1364	0.1520	0.1753	0.1372	0.1563	0.1775
2	**RMSE**	0.0674	0.0828	0.0924	0.0662	0.0853	0.0941
	ME	0.1366	0.1462	0.1717	0.1374	0.1541	0.1752
5	**RMSE**	0.1070	0.0880	0.1027	0.0790	0.0880	0.1034
	ME	0.1534	0.1480	0.1686	0.1481	0.1545	0.1733
10	**RMSE**	0.2042	0.1481	0.1580	0.1200	0.1030	0.1055
	ME	0.1691	0.1590	0.1683	0.1626	0.1534	0.1625
		BLOCK-CV			BLOCK-CV		
δ		Adaptive Lasso	SCAD	MCP	Adaptive Lasso	SCAD	MCP
1	**RMSE**	0.0576	0.076	0.0761	0.0580	0.0733	0.0735
	ME	0.1481	0.1634	0.1387	0.1550	0.1617	0.1362
2	**RMSE**	0.0622	0.0732	0.0736	0.0586	0.0721	0.0724
	ME	0.1372	0.1521	0.1353	0.1521	0.1623	0.1378
5	**RMSE**	0.1018	0.0908	0.0932	0.0655	0.0683	0.0688
	ME	0.1258	0.1451	0.1487	0.1494	0.1611	0.1449
10	**RMSE**	0.1433	0.1625	0.1661	0.0749	0.0814	0.0824
	ME	0.1350	0.1618	0.1679	0.1430	0.1709	0.1755

However, as sample size increases, two-stage adaptive lasso outperforms SCAD/MCP.

Simulation result : sparse VHAR

Table 4: Summary of performance measures for DGP2 with sample size 1000

		DGP2 / $\Sigma = I_k$			DGP2 / update Σ		
		eBIC			eBIC		
δ		Adpative Lasso	SCAD	MCP	Adpative Lasso	SCAD	MCP
1	RMSE	0.0524	0.0850	0.0931	0.0518	0.0851	0.0936
	ME	0.1023	0.1473	0.1736	0.1027	0.1478	0.1742
2	RMSE	0.0529	0.0801	0.0900	0.0518	0.0806	0.0905
	ME	0.1045	0.1399	0.1705	0.1066	0.1409	0.1708
5	RMSE	0.0712	0.0849	0.1011	0.0614	0.0845	0.1008
	ME	0.1252	0.1434	0.1669	0.1263	0.1442	0.1670
10	RMSE	0.1172	0.1258	0.1350	0.0842	0.1264	0.1363
	ME	0.1449	0.1466	0.1568	0.1485	0.1467	0.1581
		BLOCK-CV			BLOCK-CV		
δ		Adaptive Lasso	SCAD	MCP	Adaptive Lasso	SCAD	MCP
1	RMSE	0.0398	0.0718	0.0720	0.0392	0.0691	0.0693
	ME	0.1160	0.1470	0.1140	0.1199	0.1461	0.1150
2	RMSE	0.0428	0.0679	0.0682	0.0395	0.0674	0.0676
	ME	0.1070	0.1404	0.1148	0.1148	0.1427	0.1117
5	RMSE	0.0695	0.0735	0.0743	0.0462	0.0613	0.0619
	ME	0.0950	0.1363	0.1330	0.1063	0.1371	0.1224
10	RMSE	0.1004	0.1399	0.1415	0.0516	0.0716	0.0725
	ME	0.1120	0.1525	0.1600	0.0935	0.1662	0.1889

However, as sample size increases, two-stage adaptive lasso outperforms SCAD/MCP.

sparse VHAR: simulation study summary

- Updating Σ in estimation improve performance. (That is, use GLS type rather than OLS type)
- Block-CV performs slightly better than eBIC.
- SCAD/MCP reduce bias, in particular, when the sample size is small.
- For larger samples, two-stage adaptive lasso performs the best.

VHAR-MGARCH

- We further consider time varying dynamics of the financial return by adding DCC-GARCH structure.
- Assume
$$\varepsilon_t = H_t^{1/2} Z_t,$$
where $\{Z_t\}$ follows white noise with zero mean and unit variance.
- DCC-GARCH model assumes that
$$H_t = \mathrm{diag}(h_{11,t}^{1/2}, \ldots, {h_{kk,t}}^{1/2}) R_t \mathrm{diag}({h_{11,t}}^{1/2}, \ldots, {h_{kk,t}}^{1/2}),$$
where each component $\{h_{ii,t}\}$ follows a univariate GARCH(P_i, Q_i) process
$$h_{ii,t} = \alpha_{i0} + \sum_{q=1}^{Q_i} a_{iq} \varepsilon_{i,t-q}^2 + \sum_{p=1}^{P_i} b_{ip} h_{ii,t-p}. \tag{15}$$

sparse VHAR-MGARCH

- Conditional correlation matrix R_t is assumed to follow

$$R_t = \mathrm{diag}(q_{11,t}^{-1/2}, \ldots, q_{kk,t}^{-1/2}) Q_t \mathrm{diag}(q_{11,t}^{-1/2}, \ldots, q_{kk,t}^{-1/2})$$

  for a positive-definite matrix $Q_t = (q_{ij,t})$ given by the following relationship:

$$Q_t = (1 - \theta_1 - \theta_2)\overline{Q} + \theta_1 \eta_{t-1}\eta_{t-1}' + \theta_2 Q_{t-1}, \qquad (16)$$

  where $\eta_{i,t} = \varepsilon_{i,t}/\sqrt{h_{ii,t}}$ is the standardized innovation, $\overline{Q} = T^{-1}\sum_{t=1}^{T} \eta_t \eta_t'$ is the unconditional covariance matrix of η_t, and parameters θ_1, θ_2 satisfy $0 < \theta_1 + \theta_2 < 1$.

sparse VHAR-MGARCH: estimation

- The estimation of the sparse VHAR-MGARCH model is very similar to adaptive lasso except estimating H_t from DCC-GARCH.
- VHAR-GARCH is rewritten as

$$H_t^{-1/2} Y_t^{(d)} = H_t^{-1/2} (X'_{t-1} \otimes I_k) \boldsymbol{\alpha} + Z_t, \tag{17}$$

- The the adaptive lasso estimator of $\boldsymbol{\alpha}$ is given by

$$\widehat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \sum_{t=22}^{T} \| H_t^{-1/2} Y_t^{(d)} - H_t^{-1/2} (X_{t-1} \otimes I_k) \boldsymbol{\alpha} \|_2^2 + \lambda \sum_{j=1}^{3k^2} w_j |\alpha_j| \tag{18}$$

- Iterative procedure is required as follows.

sparse VHAR-MGARCH

- Step 0 (initiation)Use the ordinary least squares (OLS) estimator of the VHAR model to get resuduals

$$\widehat{\varepsilon}_t = Y_t^{(d)} - (X'_{t-1} \otimes I_k)\widehat{\boldsymbol{\alpha}} \tag{19}$$

- Step 1. fit DCC-GARCH model to obtain $\widehat{H}_t$, $t = 22, \ldots, T$.
- Step 2. Update the adaptive lasso estimator in (18) using plug-in $\widehat{H}_t$ in Step 1. The optimal penalty λ is chosen by applying block CV.
- Step 3. Iterate steps 1 to 2 until convergence. That is, update residuals (19) from the adaptive lasso estimates in Step 2, and then update the conditional covariance matrix H_t by refitting the DCC-GARCH model.

Spillover index for volatility changes

- VHAR-MGARCH model enables to calculate time varying spillpover index of Diebold and Yilmaz (2009) to see volatility changes.
- Spillover index indicates the impact of other variables (dimension) in h-step-ahead forecasting error decomposition.
- For

$$Y_t^{(d)} = \sum_{j=0}^{\infty} A_j \varepsilon_{t-j} = \sum_{j=0}^{\infty} A_j H_{t-j}^{1/2} Z_{t-j},$$

  where the coefficients are calculated from the recursion $A_j = \Phi_1 A_{j-1} + \Phi_2 A_{j-2} + \ldots + \Phi_{21} A_{j-21}$ with $A_0 = I_k$ and $A_j = 0$ if $j < 0$.
- Define the L-step ahead theta matrix as

$$\theta_{ij,t}(L) = \frac{\sum_{\ell=0}^{L-1} (e_i' A_\ell H_{t+L-\ell}^{1/2} e_j)^2}{\sum_{\ell=0}^{L-1} (e_i' A_\ell H_{t+L-\ell} A_\ell' e_i)}, \qquad (20)$$

  where e_i is the standard basis of $\mathbb{R}^k$ with 1 in the i-th coordinate and 0 otherwise.

Spillover index for VHAR-MGARCH

- Then, the spillovers at time t can be calculated as

$$S_t(L) = 100 \times \sum_{i,j=1, i \neq j}^{k} \theta_{ij,t}(L) / \sum_{i,j=1}^{k} \theta_{ij,t}(L).$$

- Higher spillover index implies that strong influence from other markets.
- Rolling window of size M to calculate the time-varying spillover index. Useful in detecting volatility changes.

Volatility changes in cryptocurrencies

- Interested in the volatility changes on the 20 largest cryptocurrencies based on market capital from January 1, 2018, to May 18, 2021. They are Bitcoin(BTC), Ethereum(ETH), Binance coin(BNB), XRP(XRP), Dogecoin(DOGE), Cardano(ADA), Bitcoin cash(BCH), Litecoin(LTC), Chainlink(LINK), Stella(XLM), Tron(TRX), Monero(XMR), Neo(NEO), Eos(EOS), IOTA(IOTA), Ethereum classic(ETC), Tezos(XTZ), Maker(MKR), Dash (DASH) and Waves(WAVES).
- Considered the volatility measure introduced in Parkinson (1980) of k-assets given by

$$\nu_{i,t}^{d} := (\log p_{i,t}^{H} - \log p_{i,t}^{L})^2/(4 * \log 2),$$

where $p_{i,t}^{H}$ and $p_{i,t}^{L}$ are the daily highest and lowest prices for the i-th asset at time t, respectively.

Volatility changes in cryptocurrencies

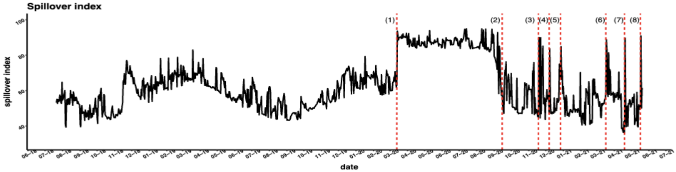


Figure 1: Time varying spillover index of 20 cryptocurrencies.

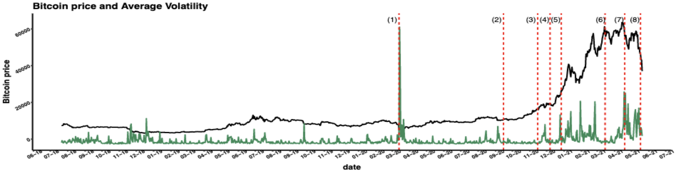


Figure 2: The price of Bitcoin(black) overlaied by the average volatility*10^6(green). Vertical red lines are as follows:

(1) 2020/03/13 : Price of Bitcoin dropped about 43 percent. This is one of the biggest one-day plunges in the history of Bitcoin.
(2) 2020/09/15 : Spillover index returned to its previous level.
(3) 2020/11/15, (4) 2020/12/08, (5) 2020/12/27 : Price of Bitcoin soared replacing its all -time high price frequently. Spikes correspond to the dates of the drop in the price of Bitcoin.
(6) 2021/03/15 : Price of Bitcoin declined by nearly six percent.
(7) 2021/04/17 : Price of Bitcoin declined by nearly eight percent.
(8) 2020/05/17 : Price of Bitcoin dropped to the lowest level in recent three months.

Volatility changes in cryptocurrencies

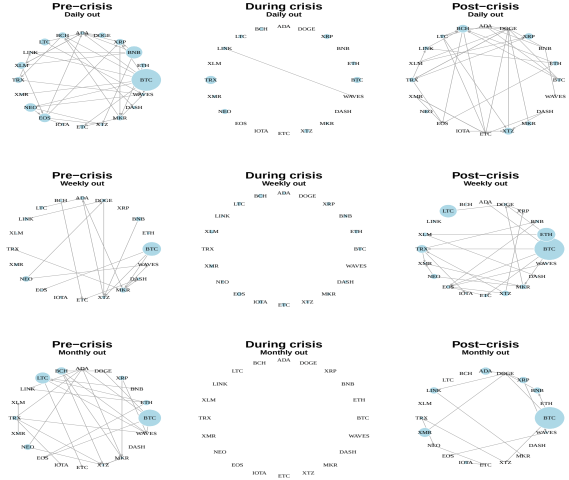


Figure: Network diagram of cryptocurrencies on three regimes with outward centrality measure.

Volatility changes in cryptocurrencies

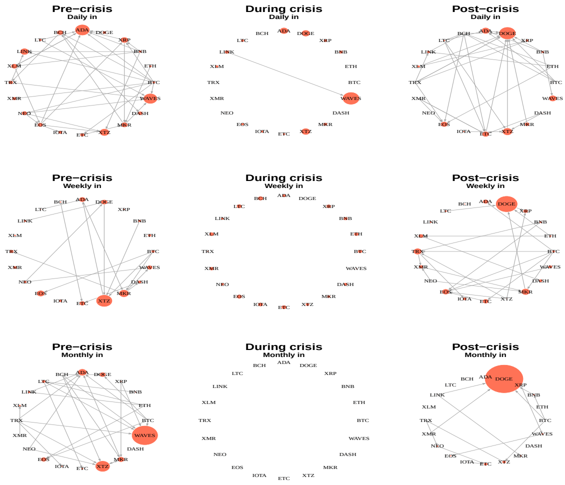


Figure: Network diagram of cryptocurrencies on three regimes with inward centrality measure.

Volatility changes in cryptocurrencies: summary

- VHAR-MGARCH indicates that cryptocurrency volatility can be classified as pre-crisis, during the crisis, and post-crisis regimes.
- Before the market crash on March 13, 2020, the spillover index showed a somewhat irregular but long-term trend; however, after the remarkable plunge due to COVID-19, it remained high for around six months. Then, volatility showed frequent, unexpected sharp spikes.
- Connectivity changes were also detected for each regime. During the crisis period, almost no connection between coins. This implies that the market behaves independently during a crisis. Bitcoin played a main role before the crisis. However, the post-crisis diagram shows that Bitcoin remained influential in the mid- and long-term, but not in short term.
- We conjecture that deep impact/shock may lead the market scattered while still volatile.
- The rise of Dogecoin after the market crash.

References

- Corsi, F. (2004), A simple long memory model of realized volatility, Tech. rep., University of Southern Switzerland.
- Corsi, F., Mittnik, S., Pigorsch, C. and Pigorsch, U. (2008), The volatility of realized volatility, Econometric Reviews 27(1-3), 46–78.
- Baek, C and Park, M (2020), Sparse vector heterogeneous autoregressive modeling for realized volatility, Journal of the Korean Statistical Society.
- Bernanke, B. S., J. Boivin, and P. Eliasz. 2005. "Measuring the Effects of Monetary Policy: a Factor-augmented Vector Autoregressive (FAVAR) Approach." The Quarterly Journal of Economics 120 (1): 387–422. doi:10.1162/ 0033553053327452.
- Diebold, F. X., and K. Yilmaz. 2009. "Measuring Financial Asset Return and Volatility Spillovers, with Application to Global Equity Markets." The Economic Journal 119 (534): 158–171. https://doi.org/10.1111/j.1468-0297.2008.02208.x.
- Kim, D. and Baek, C (2019), 'Factor augmented HAR model improves realized volatility forecasting: Evidence from Asian market', Applied Economics Letters.
- Lee, S and Baek, C (2022). Volatility changes in cryptocurrencies: evidence from sparse VHAR-MGARCH model, Applied Economics Letters, DOI: 10.1080/13504851.2022.2064417.
- Shin, J, Park, M and Baek, C (2022). Sparse vector heterogeneous autoregressive model with nonconvex penalties, Communications for Statistical Applications and Methods 29 (1), 53-64