Sparse-Group Network Autoregressive Model for Cryptocurrencies

Ying Chen

Wolfgang Karl Härdle

Simon Trimborn

Jiejie Zhang

Department of Statistics and Applied Probability
National University of Singapore
Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. – Center for Applied Statistics
Humboldt–Universität zu Berlin

Bitcoin and Blockchain

Bitcoin is the most popular and liquid cryptocurrency supported by a digital payment system that allows transactions happen between users directly.

- $\sim 650,000$ active accounts on 15 May 2017
- $\sim 350,000$ transactions on 15 May 2017

Bitcoin transactions are executed globally.

- Regional impact: Kristoufek (2015) documented Bitcoin transactions in China has an influence on the US market.
- Size impact: In early 2017 Bitcoin trading price plunged after Chinese authorities regulated the margins.

Global Bitcoin market

- Is there any dynamic dependence of transactions on its own past values?
- How is the transaction size of one region affected by other regions?
- Which region is dominating the dynamic influences in the global market? Which region has the least influence?

Continents and Transaction Sizes

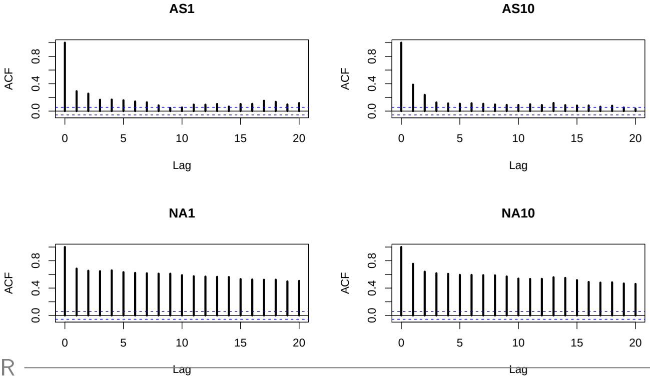
We group the bitcoin transactions according to continent and transaction size and investigate the dynamic dependence of the transaction flows in the bitcoin network.

Region	Transactions in x to y-percentile of total transaction sizes			
	Group 1	Group 2	...	Group 10
AF,AS,EU,NA,OC and SA	1% to 10%	10% to 20%	...	90% to 100%

Six continents: AF: Africa; AS: Asia; EU: Europe; NA: North America; OC: Oceania; SA: South America.

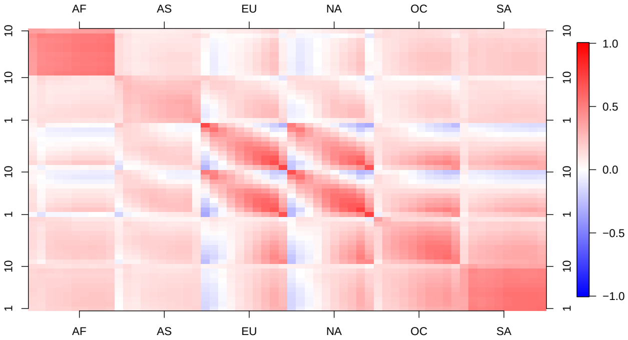
Serial dependency

Figure 1: Autocorrelation functions of the daily transactions in the smallest and largest sized groups in Asia (AS1 and AS10) and North America (NS1 and NS10).



Lead-lag cross dependence

Figure 2: Lag 1 cross-correlations among the sixty groups. On the diagonal line displays the serial dependence in each continent. Off the diagonal shows the lead-lag cross dependence between continents.



Network Vector Autoregression

The lead-lag cross dependence in (social) network can be described by the Network Vector Autoregression model proposed by Zhu, Pan, Li, Liu and Wang (2017):

$$Y_{i,t} = \beta_0 + \beta_1 Y_{i,(t-1)} + \beta_2 \frac{1}{N} \sum_{j=1}^{N} a_{ij} Y_{j,(t-1)} + Z_i^\top \gamma + \epsilon_{i,t}$$

- $Y_{i,t} \in \mathbb{R}^N$ represents the attributes of user/node $\#i$ at time t.
- β_1 reflects serial dependence, and β_2 controls the strength of network's dynamic connectivity. They are constant for all the users $i = 1, \cdots, N$.
- Adjacency factors a_{ij} measures the connectivity and are assumed to be known and binary. Moreover $a_{ii} = 0$.
- Covariates Z denotes user's individual features e.g. the gender of the user.

Sparsity in network

- The adjacency matrix $A = \{a_{ij}\}_{i,j=1}^{N}$ is unknown in bitcoin transactions.
- There are only a few number of nodes/groups dominating the transactions in the network.
- Detecting them can help to understand the essential dynamic dependence structure in the global bitcoin markets.
- The lead-lag cross dependence and thus the adjacency factors can be sparse both elementwise (i.e. previous transactions of user/group j have no impact on another user/group i) and groupwise (i.e. previous transactions of user/group j has no influence on any other users/groups).

Sparsity: LASSO and Group LASSO

- Regression Shrinkage and Selection via the Lasso (Tibshirani 1996):

$$\frac{1}{2}||Y - X\beta||^2 + \lambda|\beta|$$

  where $||\cdot||$ is the Euclidean norm.

- Yuan and Lin (2006) extended to the case of grouped variables by penalizing groups of variables:

$$\frac{1}{2}||Y - \sum_{j=1}^{N} X^{(j)}\beta^{(j)}||^2 + \lambda \sum_{j=1}^{N} ||\beta^{(j)}||$$

  where $X^{(j)}$ is the submatrix of predictors denoted by X with columns corresponding to the predictors in group j, $\beta^{(j)}$ is the coefficient vector of group j.

Simon et al. (2013) developed a model with two penalization parameters and algorithm which penalizes groups and the parameters in the groups separately.

Objectives

We propose a Sparse Group Network AutoRegressive (SG-NAR) model to describe the dynamic dependence in network with unknown yet sparse adjacency matrix:

$$Y_{i,t} = \beta_0 + \beta_1 Y_{i,(t-1)} + \sum_{j=1}^{N} a_{ij} Y_{j,(t-1)} + \mathbf{Z}_i^T \gamma + \epsilon_{i,t} \tag{1}$$

where a_{ij} is unknown and not necessarily binary. It introduces flexibility on the existence and level of connectivity in network.

- We derive the SG-NAR estimation using least squares.
- We investigate finite sample performance with simulations
- We implement the SG-NAR model to the bitcoin transactions and study the connectivity in the global market.

Outline

Bitcoin blockchain

We consider Bitcoin Blockchain data from 02 February 2012 until 01 August 2015 (1247 days).

- Datasource: Blockchain.info.
- Sample frequency: The transactions are accumulated to daily frequency based on the raw data sampled every 10 minutes, given the low liquidity of the network.
- Attributes: transaction size, account ID, transaction parties (participating accounts), timestamp of transactions and the IP address from the origin of the transaction.

IP address and grouping

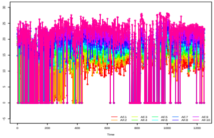
Identify the continent of each transaction (origin) using MaxMind Inc.

Group transactions to six continents and ten sizes: Africa (AF), Asia (AS), Europe (EU), North America (NA), Oceania (OC), South America (SA). Each continent is further split into 10 groups according to the evenly distributed quantiles of transaction sizes.

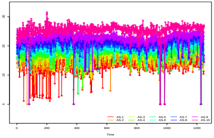
Region	Transactions in x to y-percentile of total transaction sizes			
	Group 1	Group 2	...	Group 10
{AF,AS,EU,NA,OC, and SA}	1% to 10%	10% to 20%	...	90% to 100%

Transactions

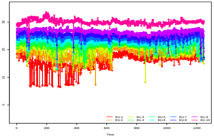
Figure 3: Time series of daily log transaction size over the 1247 days in the 6 continents.



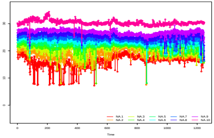
(a) AF



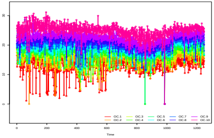
(b) AS



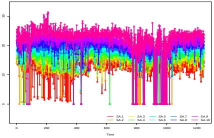
(c) EU



(d) NA



(e) OC



(f) SA

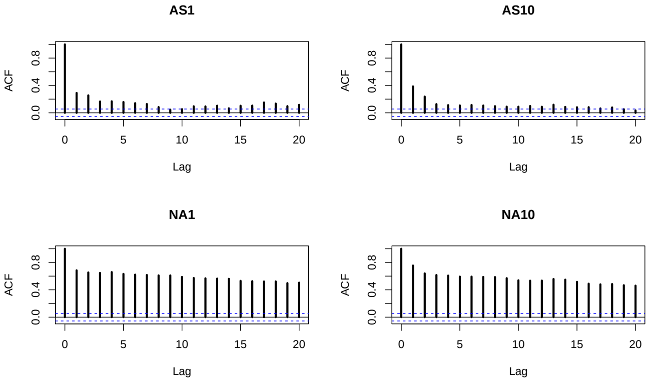
Descriptive statistics

	AF	AS	EU	NA	OC	SA
mean	147.25	190.78	226.92	229.72	189.70	172.64
sd	67.52	20.41	9.21	9.00	17.09	39.43
skewness	-1.52	-4.89	-1.38	-1.80	-1.79	-3.36
kurtosis	3.75	42.77	6.94	10.58	15.78	14.95
min	0.00	0.00	171.08	154.25	0.00	0.00
max	222.39	228.82	246.76	249.44	234.54	225.25

Table 1: Descriptive statistics of the accumulated log transactions of the 6 regions Africa (AF), Asia (AS), Europe (EU), North America (NA), Oceania (OC), South America (SA)

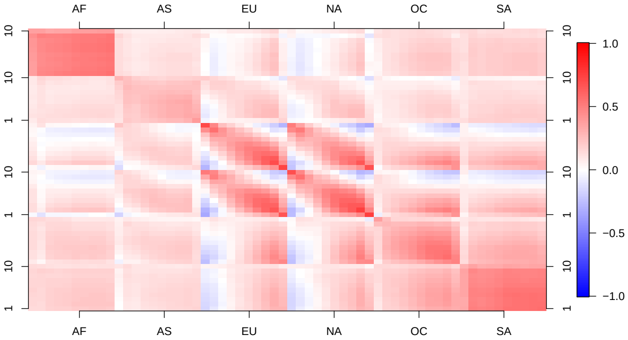
Serial dependency

Figure 4: Autocorrelation Functions of the daily transaction amount in Asia and Europe



Cross-correlation

Figure 5: Lag 1 cross-correlations between the size of the transactions - ordered in 10 groups - in the 6 regions: Africa (AF), Asia (AS), Europe (EU), North America (NA), Oceania (OC), South America (SA)



Sparse Group Network AutoRegression

SG-NAR:

$$Y_{i,t} = \beta_0 + \beta_1 Y_{i,(t-1)} + \sum_{j=1}^{N} a_{ij} Y_{j,(t-1)} + \mathbf{Z}_i^T \gamma + \epsilon_{i,t} \qquad (2)$$

where $Y_{i,t}$ denotes the transactions of Node (Group) i at time t. Serial dependence is controlled by β_1. Adjacency matrix $A = (a_{ij})_{1 \le i,\ j \le N}$ reflects the conenctivity (Node j's lag on Node i). Note a_{ij} shows both connectivity (zero or nonzero) and strength (magnitude). Moreover $a_{ii} = 0$. ϵ_{ti} is white noise s.t. $E(\epsilon_{i,t}) = 0$, $E(\epsilon_{i,s}\epsilon_{i,\tau}) = 0$, $Var(\epsilon_{i,t}) = \sigma_i^2$, $1 \le i \le N$ and $1 \le t,\ s,\ \tau \le T$.

Objective function under sparsity

W.L.O.G., the SG-NAR can be represented in matrix form:

$$\mathbf{Y}_t = \{\beta_1 I_N + A\}\mathbf{Y}_{t-1} + \boldsymbol{\epsilon}_t \tag{3}$$

which gives

$$\min_{(\beta_{\mathbf{1}}, A)^\top} \sum_{t=2}^{T} \frac{1}{2} ||\mathbf{Y}_t - (\beta_1 I_N + A)\mathbf{Y}_{t-1}||^2 + \sum_{i=1}^{N} \lambda_1 (||A_{.i}||) + \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_2 |a_{ij}| \tag{4}$$

where λ_1 controls the group sparsity and λ_2 handles the elementary sparsity.

Estimator under sparsity

We adopt a two-step algorithm to iteratively estimate β_1 and A (Simon et al. 2013). The estimator satisfies

$$\hat{\beta}_1 = \operatorname*{argmin}_{\beta_1} L(r_{-A_{.i}}; A_{.i}) \tag{5}$$

$$\hat{A}_{.i} = \operatorname*{argmin}_{A_{.i}} L(r_{-A_{.j},-\beta_1}; A_{.j}, \beta_1) + \sum_{i=1}^{N} \lambda_1(||A_{.i}||) + \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_2 |a_{ij}| \tag{6}$$

General estimator under sparsity

Let $\theta_1 = \beta_1$, $\theta_k = A_{\cdot i}$, $k = 2, \cdots, N+1$. We denote (5) and (7) generally as

$$\hat{\theta}_k = \underset{\theta}{\text{argmin}}\, L(r_{-\theta_k}; \theta_k) + \sum_{i=1}^{N} \lambda_1(||A_{\cdot i}||) + \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_2 |a_{ij}| \quad (7)$$

where $r_{-\theta_k}$ is the partial residual of $\mathbf{Y}_t$, $1 \le t \le T$ after subtracting all fits other than group θ_k

$$r_{-\theta_k} = \sum_{t=2}^{T} ||\mathbf{Y}_t - \sum_{l \neq k} (\theta_1, \cdots, \theta_{k-1}, \theta_{k+1}, \cdots, \theta_{N+1}) \mathbf{Y}_{t-1}||^2 \quad (8)$$

and

$$L(r_{-\theta_k}; \theta_k) = \frac{1}{2N} ||r_{-\theta_k} - \sum_{t=2}^{T} (\mathbf{0}, \cdots, \theta_k, \cdots, \mathbf{0}) \mathbf{Y}_{t-1}||^2 \quad (9)$$

is the unpenalized loss function. $\lambda_1 = \lambda_2 = 0$ if $k = 1$.

Derivation of the optimization problem

By Taylor's expansion, we can derive that

$$L(r_{-\theta_k};\theta_k) \leq L(r_{-\theta_k};\theta_0) + (\theta-\theta_0)^\top \nabla L(r_{-\theta_k};\theta_0) + \frac{1}{2t}||\theta-\theta_0||^2 \quad (10)$$

where t is sufficiently small that the quadratic term dominates the Hessian of the loss function.

Minimizing the penalized loss function is equivalent to minimizing

$$\begin{aligned} M(\theta_k) =& L(r_{-\theta_k};\theta_0) + (\theta-\theta_0)^\top \nabla L(r_{-\theta_k};\theta_0) + \frac{1}{2t}||\theta-\theta_0||^2 \\ &+ \sum_{i=1}^{N} \lambda_1(||A_{\cdot i}||) + \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_2|a_{ij}| \end{aligned} \quad (11)$$

We establish the algorithm to solve (11) based on Simon et al. (2013).

Algorithm

1. Initiate $\beta_1 = 0,\ A = 0,\ l = 1,\ t = 1$
2. Derive until $A^{(l)} - A^{(l-1)} < \epsilon_1$ or $\beta_1^{(l)} - \beta_1^{(l-1)} < \epsilon_1$
3. $\theta_1 = \beta_1, \theta_k = A_{.i}, i = 1, \dots, N, k = 2, \dots, N + 1$
   - 3.1 Iterate through each $k = 1, \dots, N + 1$
     - 3.1.1 $m = 1$
     - 3.1.2 Derive until $\theta_k^{(m)} - \theta_k^{(m-1)} < \epsilon_2$
     - 3.1.3 Derive $S = \text{sign}(\theta_k - t*?\nabla L(r_{\theta_k}; \theta_k))*?(|\theta_k| - \lambda)_+$
     - 3.1.4 Derive $U = (1 - t\lambda/||S||)_+ S$
     - 3.1.5 Update $\theta_k = U_l + l/(l + 3)(U_{l+1} - U_l)$
     - 3.1.6 $k = k + 1,\ t = 0.1 * t$
   - 3.2 $l = l + 1$

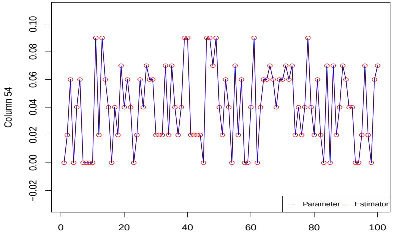
Simulation

We investigate the finite sample performance (accuracy and sparsity) of the SG-NAR estimation along with simulated super-sparse (low connectivity at intensity of $d = 2\%$) and sparse (middle connectivity at $d = 10\%$) networks.

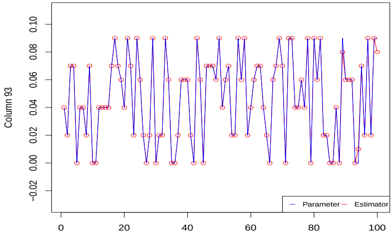
- $N = 100$ nodes (users or groups). Sample periods $T = 10$.
- Attributes $(Y_{1t}, \cdots, Y_{Nt})$, $1 < t \leq T$, are with serial dependence $\beta_1 = 0.5$ on its own past values and cross dependence A on others' past values. $(Y_{11}, \cdots, Y_{N1})$ are independently from $N(0, 0.05)$.
- The nonzero entries of the adjacency matrix A (correspond to the active users) are randomly sampled from $(0, 0.02, 0.04, 0.06, 0.07, 0.09)$.
- Generation is repeated 100 times.

Scenario Low Connectivity: Accuracy

Figure 6: The detected active nodes (nonzeros). They are Column 54 (a) and 93 (b) in A and $\hat{A}$). $d = 2\%$. $N = 100$. $T = 10$.



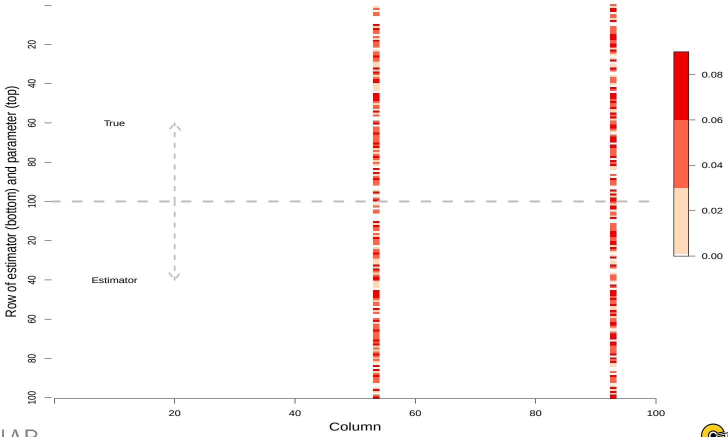
(a)



(b)

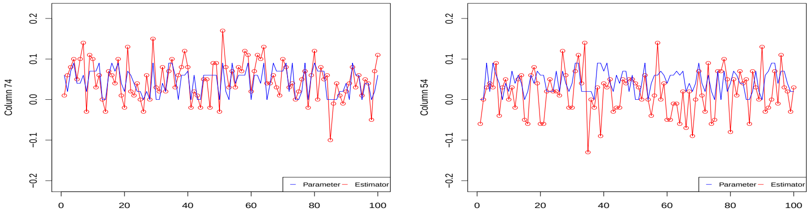
Scenario LC: accuracy

Figure 7: $\beta_1 = 0.5$, $\hat{\beta}_1 =$. Adjacency matrix A (top) and its estimator $\hat{A}$ (bottom). $d = 2\%$. $N = 100$. $T = 10$



Scenario Mid Connectivity: Accuracy

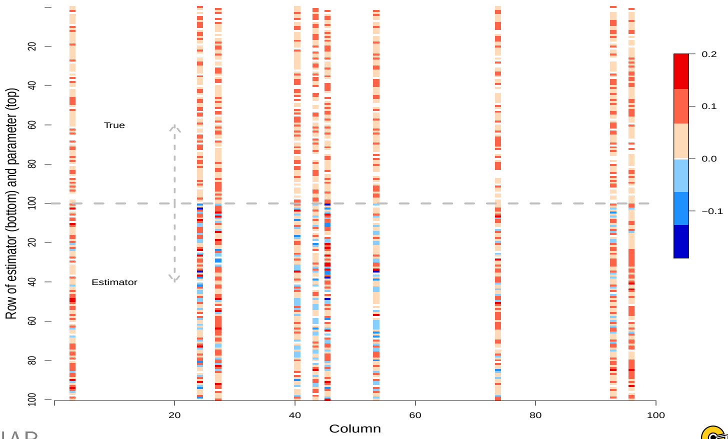
Figure 8: The detected active nodes (nonzero columns) in the scenario. Two are reported: Column 74 (a) with the smallest RMSE 0.31 and Column 28 (b) with the largest RMSE 2.79 among the 10 active users. $d = 10\%$. $N = 100$. $T = 10$.



(a) (b)

Scenario MC: Accuracy

Figure 9: $\beta_1 = 0.5$, $\hat{\beta}_1 = 0.44$. Adjacency matrix A (top) and its estimator $\hat{A}$ (bottom). $d = 10\%$. $N = 100$. $T = 10$



Sparsity: Elementwise

Table 2: Elementwise summary of the estimators β_1 and $\hat{A}$. $d\% = 2\%$, 10%. $N = 100$. $T = 10$.

Parameter	$d\% = 2\%$				$d\% = 10\%$			
	Estimator	Size	Percentage(%)	RMSE(%)	Estimator	Size	Percentage(%)	RMSE(%)
$\beta_1 = 0.5$	$\hat{\beta}_1 = 0.5$	-	-	-	$\hat{\beta}_1 = 0.44$	-	0	6.0
$a_{ij} = 0$	$\hat{a}_{ij} = 0$	9,734	100	0	$\hat{a}_{ij} = 0$	8,927	98.4	6.0
	$\hat{a}_{ij} \neq 0$	1	0	1.0	$\hat{a}_{ij} \neq 0$	146	1.6	5.0
$a_{ij} \neq 0$	$\hat{a}_{ij} = 0$	0	0	-	$\hat{a}_{ij} = 0$	33	4.0	5.0
	$\hat{a}_{ij} \neq 0$	165	100	0.1	$\hat{a}_{ij} \neq 0$	794	96.0	5.0

Sparsity: Groupwise

Table 3: Groupwise summary of the estimator $\hat{A}$. $d\% = 2\%, 10\%$. $N = 100$. $T = 10$.

Parameter	Estimator	$d\% = 2\%$		$d\% = 10\%$	
		Size	Percentage(%)	Size	Percentage(%)
$a_{.j} = 0$	$\hat{a}_{.j} = 0$	98	100	90	100
	$\hat{a}_{.j} \neq 0$	0	0	0	0
$a_{.j} \neq 0$	$\hat{a}_{.j} = 0$	0	0	0	0
	$\hat{a}_{.j} \neq 0$	2	100	10	100

Bitcoin Transaction Analysis

We implement the proposed Sparse Group NAR model on the transaction size in the Bitcoin network. The transactions are grouped by continents and transaction quantiles.

We answer the following questions:

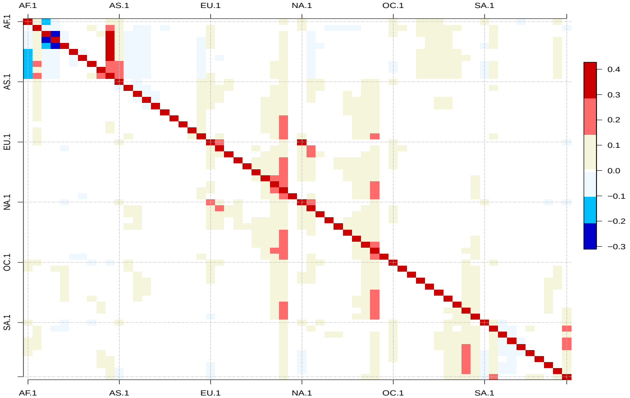
- How is the transactions size of a certain region (denoted as G) at time t affected by G's lagged transaction size at $t-1$?
- Which region/regions affect the transaction size of G? How is the strength of connectivity?
- Which region/regions influence most of the rest regions? Which region/regions influence little on the other regions?

Grouping

Region	Transactions in upper x to y-percentile of total transaction sizes			
	Group 1	Group 2	...	Group 10
AF,AS,EU,NA,OC, and SA	1% to 10%	10% to 20%	...	90% to 100%

Real data: $\hat{A}$ and $\hat{\beta}_1$

Figure 10: Estimator of $\hat{A}$ and $\hat{\beta}_1$ of the accumulated Bitcoin transaction data. $N = 60$, $T = 1247$, $\lambda_1 = 0.0001$, $\lambda_2 = 0.00026$.



Real Data Analysis

- The transaction size of all the groups: AF.1 to SA.10 are positively influenced by their own lag 1, with a ratio of 0.30.
- Besides the own lag 1, the second largest transaction sizes of Europe (EU.9) influence AS.7–AS.10, EU.4–EU.10, NA.5–NA.10 and OC.8–OC.10.
- Besides the own lag 1, the second largest transaction sizes of North America (NA.9) influence AS.10, EU.8-EU.10, NA.8–NA.10 ans OC.5–OC.10.
- The smallest transaction size of AS.8-AS.10
- Africa (AF) are mostly influenced by their own groups (AF.10, AF.1) and by AS.8–AS.10. Africa show negative influence, indicating a diminishing transaction volume.

Conclusion

- We propose a practical and flexible Sparse Group Network AutoRegressive (SG-NAR) Model to investigate the connectivity and detect the essential active users in the network, with the adjacency matrix A being unknown.
- We develop the least square estimator under group-elementaty sparsity.
- Simulation study shows SG-NAR is accurate in identifying the active nodes and estimating the adjacency matrix accurately.
- We apply the SG-NAR model in the Bitcoin transactions. For the first time, we investigate the dynamic dependence in Bitcoin transactions among different continents.

Sparse-Group Network Autoregressive Model for Cryptocurrencies

Ying Chen
Wolfgang Karl Härdle
Simon Trimborn
Jiejie Zhang
Department of Statistics and Applied Probability
National University of Singapore
Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. – Center for Applied Statistics
Humboldt–Universität zu Berlin
http://lvb.wiwi.hu-berlin.de
http://www.stat.nus.edu.sg

Bibliography I

- Jianqing Fan, Yang Feng and Yichao Wu (2009)
  Network Exploration via the Adaptive LASSO and SCAD penalties
  The Annals of Applied Statistics

- Ming Yuan and Yi Lin (2006)
  Model selection and estimation in regression with grouped variables
  Journal of the Royal Statistical Society: Series B

Back

Bibliography I

William B. Nicholson, David S. Mattesony and Jacob Bienz (2016)
VARX-L: Structured Regularization for Large Vector Autoregressions with Exogenous Variables
Discussion paper

Song Song, Peter J. Bickel (2011)
Large Vector Auto Regressions
Discussion paper

Back

Bibliography II

Robert Tibshirani (1996)
Regression shrinkage and selection via the lasso
Journal of the Royal Statistical Society. Series B

William B. Nicholson, David S. Mattesony and Jacob Bienz (2016)
VARX-L: Structured Regularization for Large Vector Autoregressions with Exogenous Variables
Discussion paper

Back

Bibliography III

David Hallac, Jure Leskovec and Stephen Boyd (2015)
Network Lasso: Clustering and Optimization in Large Graphs
Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining

Gareth M. James, Courtney Paulson and Paat Rusmevichientong (2012)
The Constrained Lasso
Discussion paper

Back

Bibliography IV

- Xuening Zhu, Rui Pan, Guodong Li, Yuewen Liu and Hansheng Wang (2016)
  Network Vector Autoregression
  Discussion paper

- Xuening Zhu, weining Wang, Hansheng Wang and Wolfgang K. Härdle (2016)
  Network Quantile Autoregression
  Discussion paper

Back