Imperial College London

Learning the causal graph structure of financial markets

Théophile Griveau-Billion & Ben Calderhead

tg1314@imperial.co.uk

Abstract

We apply the causal graph process cyclical coordinate descent (CGP-CCD) algorithm on European and American stocks to learn the causal graph structure of these markets. The CGP-CCD is able to compute the adjacency matrix and infer the degree of sparsity using only a group of time series as input. With the automatic sparsity level selection, the algorithm provides a novel view of the underlying causal structure of the financial market. We present the results obtained for the S&P500 and a group of 467 European stocks.

Motivation

We want to learn the causal graph structure of the financial market with the CGP-CCD algorithm. However, before studying this structure in detail we need to assess the quality of the sparse directed graph. In this poster we present the results from applying the CGP-CCD algorithm on financial datasets and assess the quality of the obtained graph by studying the impact of the financial crisis on the learned structure.

Causal Graph Process

We follow [3, 2] to model the causal graph process as an autoregressive process with the coefficients of each lag assumed to follow a polynomial function of the adjacency matrix. Let us denote by A the adjacency matrix, N the number of nodes, K the number of time steps, M the number of lags, C the vector of polynomial coefficients. We consider a directed graph where the coefficient $A_{i,j}$ of the adjacency matrix corresponds to an edge going from node j to i and vice-versa. We define the graph signal at time k by $x(k) = (x_0(k) \dots x_{N-1}(k))^T \in \mathbb{R}$. The causal graph process follows at k:

$$\begin{aligned} x(k) &= w(k) + \sum_{l=1}^{M} P_l(A) x(k-l) \,, \\ &= w(k) + \sum_{l=1}^{M} \left(\sum_{j=0}^{l} c_{l,j} A^j \right) x(k-l) \,, \\ &= w(k) + A x(k-1) + \dots + \left(c_{M,0} I + \dots + c_{M,M} A^M \right) x(k-M) \,. \end{aligned}$$

Where $w(k)$ corresponds to a Gaussian noise. Without lost of generality, we fixed the lag-one coefficients $(c_{1,0}, c_{1,1}) = (0, 1)$ in order to easily retrieve the adjacency matrix A. We can retrieve the CGP coefficients (A, C) by solving the optimisation problem proposed by [2]:

$$(A, C) = \min_{A,C} \frac{1}{2} \sum_{k=M}^{K-1} \left\| x(k) - \sum_{l=1}^{M} P_l(A) x(k-l) \right\|_2^2 + \lambda_1 \|A\|_1 + \lambda_1^c \|C\|_1 + \lambda_2^c \|C\|_2^2 \,, \qquad (1)$$

CGP-CCD

In this section we detail the CGP-CCD algorithm proposed by [1] to obtain the adjacency matrix A from a potentially large group of time series. The CGP-CCD uses a block coordinate descent to obtain first, the adjacency matrix A, then the polynomial coefficients C. For the first step they consider the optimisation problem 1 as a function of the matrix coefficients $R_l = P_l(A)$ and cycle over the different lags for the descent. Since there is a LASSO regularisation term on the adjacency matrix to impose sparsity they consider the case of lag one separately from the rest. For lags $l > 1$ instead of relying on the gradient of a matrix the CGP-CCD algorithm uses a matrix update formula for each R_l consisting of one matrix-matrix multiplication. Hence, the CCD update for the matrix coefficient R_l is:

$$R_l = \left(\sum_{k=M}^{K-1} S_k^l x(k-l)^T \right) \left(\sum_{k=M}^{K-1} x(k-l) x(k-l)^T \right)^{-1} .$$

For the first lag $l = 1$ the CGP-CCD uses a cyclical coordinate descent algorithm on the columns of the matrix to impose sparsity on each node separately. Indeed, each column of the adjacency matrix corresponds to the edges directed from that node to the others it influences. Let us denote by A^j the column j of A, A^{-j} the matrix A without the column j, x^j the element j of the vector x and x^{-j} the vector x without the coefficient j, the CCD update for the vector A^j is:

$$A^j = \frac{S\left(\sum_{k=M}^{K-1} \left(S_k - A^{-j} x^{-j}(k-1) \right) x^j(k-1), \lambda_1 \right)}{\sum_{k=M}^{K-1} (x^j(k-1))^2} .$$

The second step applies a CCD on the coefficients of C. Since this poster only studies the adjacency matrix we let the reader refer to [1] for this step. While the use of a LASSO coefficient is efficient in enforcing sparsity, the selection of its coefficient is non-trivial. Cross-validation is the main method of selection but does not work in the case of time series where the order is important such as causal processes. [1] proposed two new metrics to automatically select the LASSO coefficient in the case of CGP. The first one is:

$$err = \sum_{j=1}^{N} \frac{1}{\sum_{i=1}^{N} \mathbb{I}_{\{A_j \neq 0\}}} \frac{1}{K-M} \sum_{k=M}^{K} \left\| x(k) \mathbb{I}_{\{A_j \neq 0\}} - A x_j(k-1) \right\|_2^2 ,$$

Where $\mathbb{I}_{\{A_j \neq 0\}}$ is a binary vector to select the non-zeros elements of A_j. The second error metric, err^d, uses the degree of the matrix $\sum_{i=1}^{N} |A_{i,j}|$ instead of the number of edges $\sum_{i=1}^{N} \mathbb{I}_{\{A_j \neq 0\}}$. [1] proved the efficiency of these two metrics in simulated environments of CGP following stochastic-block-models.

Datasets

We apply the CGP-CCD algorithm on stocks from the US and EU markets. We select the 371 stocks from the S&P500 that have a history of prices from $01/01/2000$ to $01/01/2018$. For the Europe we select the 467 most liquid stocks from different European markets with data between $01/01/2000$ to $01/01/2018$. Both datasets consist of the last quoted price of the day. When the quote is missing for a day we carry over the previous price. We run the CGP-CCD algorithm on the daily log-returns for both datasets separately. In both cases the algorithm uses time windows of 4 years, $K = 1040$, and a weekly time lag, $M = 5$. We implement the algorithm to run on GPU with the PyTorch library.

Results

In order to assess the sparse directed adjacency matrix we look at the impact of a known market change. More specifically, we compute the adjacency matrix before the financial crisis, using data between $06/11/2003$ and $01/01/2008$, and after the crisis with the 4 years time window $03/11/2009$ to $01/01/2014$. Since we know that the financial crisis impacted the structure of the market, the causal structure of the graph should change after the crisis. Indeed we can observe in both markets that the adjacency matrix is denser before the crisis and that the importance of the financial sector decreases after the crisis. Figures 1 for the US and 3 for the EU show the adjacency matrix before and after the crisis with the stocks grouped by financial sectors using the BCIS (Bloomberg Industry Classification Standard) sector classification. For a better visualisation we do not show the weight of the edge but instead a black square for every non-zero edge.

S&P500

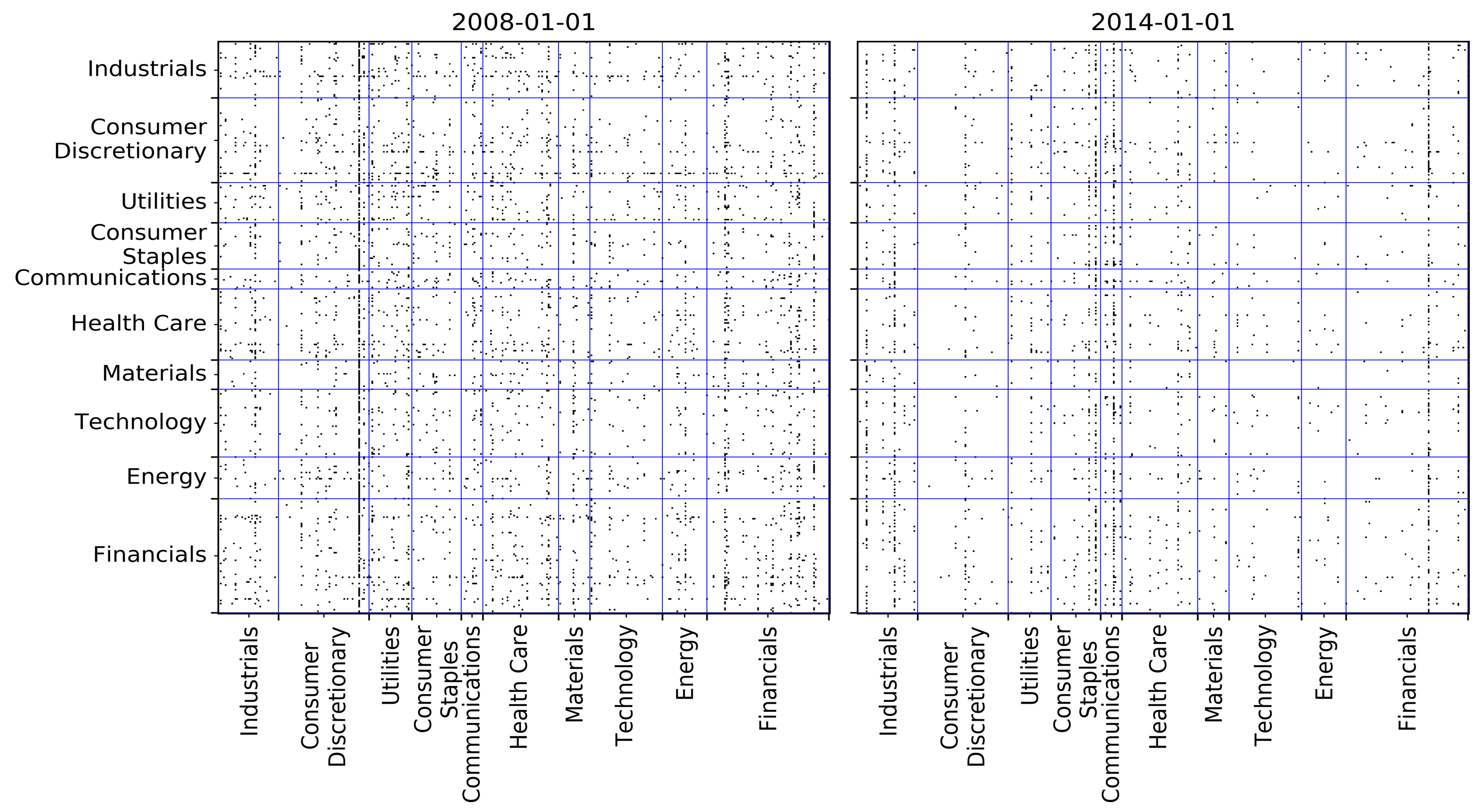

Figure 1: Adjacency matrix of 371 stocks from the S&P500 computed using data between $06/11/2003$ and $01/01/2008$ on the left, and $03/11/2009$ to $01/01/2014$ on the right. The stocks are grouped by sector.

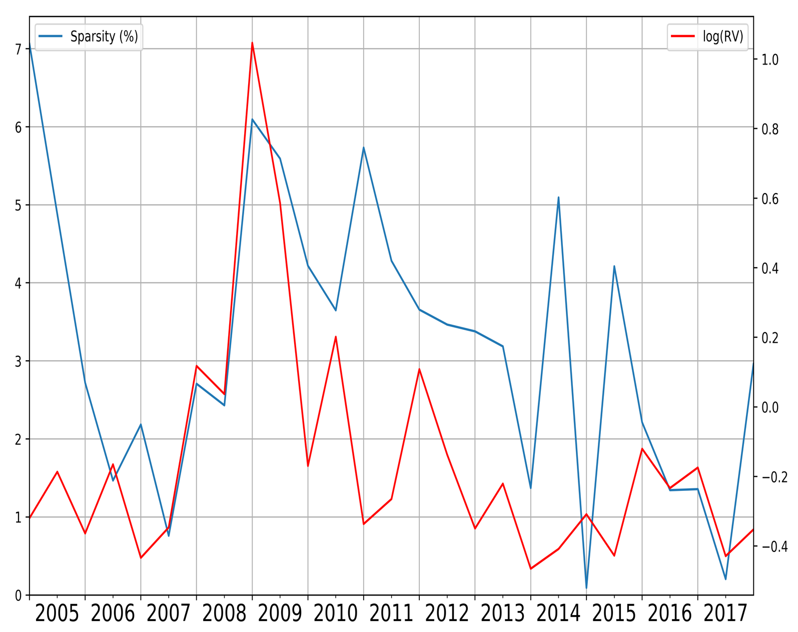

Figure 2: For 371 stocks from the S&P500, evolution of the sparsity level of the adjacency matrix, left y-axis, and the $log(RV)$ on the right y-axis.

Since the CGP-CCD algorithm automatically selects the sparsity level, in addition to studying the different connection we can also study the evolution of the sparsity level. Let us define the sparsity level as the percentage of non-zeros edges, hence 100.0 corresponds to a fully connected graph. We estimate the variance of the market by the average of the log-Realised-Variance, $log(RV)$, of each stock. We computed the $log(RV)$ with exponential weighting on the daily log-returns, using a coefficient of 0.99 and time window of 40 days. Figure 2 shows the evolution of the sparsity level of the adjacency matrix and of the $log(RV)$ of the market through time. We computed the adjacency matrix every 6 months and the corresponding $log(RV)$ at the last date of that time window.

European Stocks

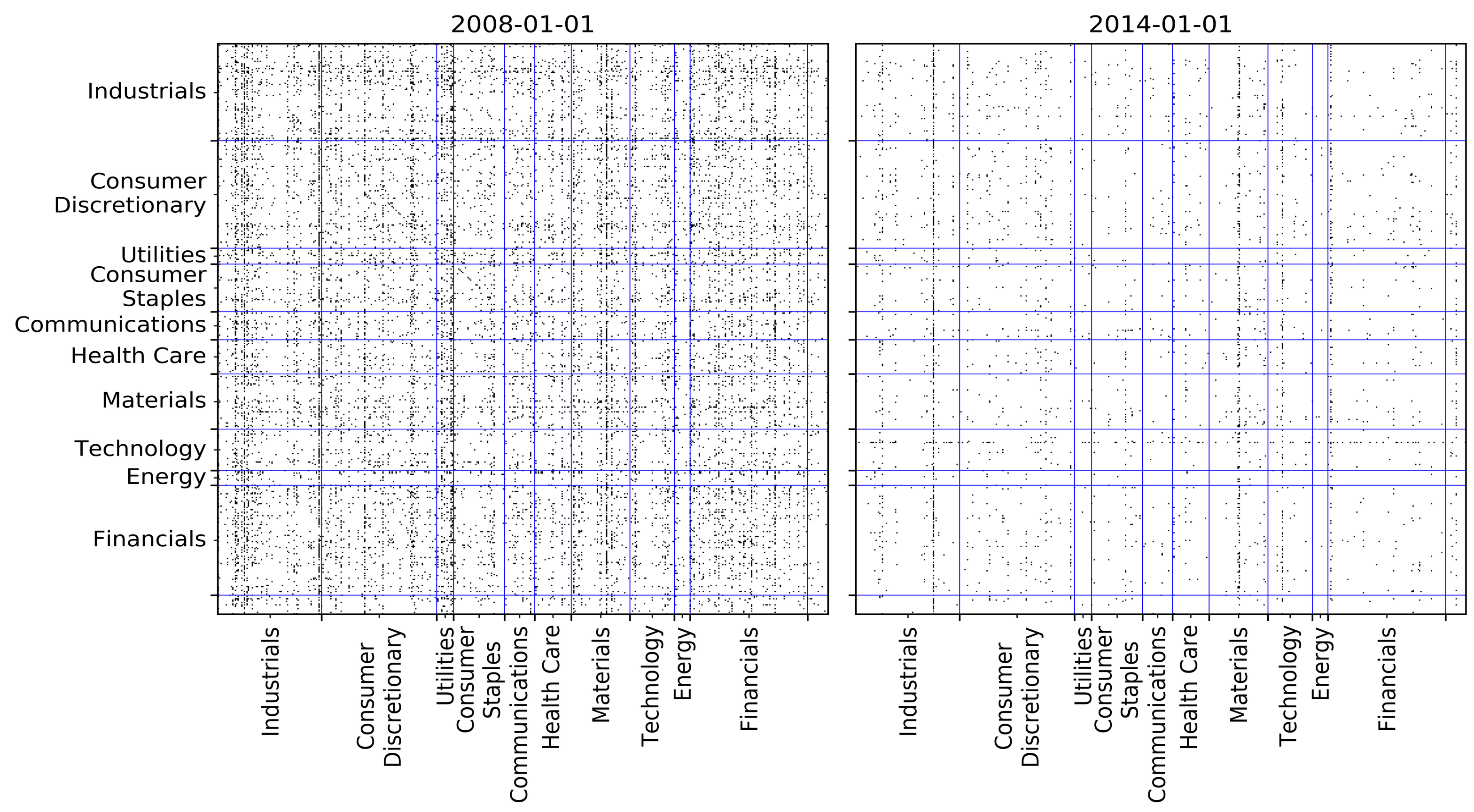

Figure 3: Adjacency matrix of 467 European stocks computed using data between $06/11/2003$ and $01/01/2008$ on the left, and $03/11/2009$ to $01/01/2014$ on the right. The stocks are grouped by sector.

References

[1] Théophile Griveau-Billion and Ben Calderhead. Efficient structure learning with automatic sparsity selection for causal graph processes. *arXiv.org*, June 2019.

[2] Jonathan Mei and Jos e M F Moura. Signal processing on graphs: causal modeling of unstructured data. *IEEE Transactions on Signal Processing*, 65(8):2077–2092, 2017.

[3] Jonathan Mei and Jose M F Moura. Signal processing on graphs: Estimating the structure of a graph. In *ICASSP 2015 - 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5495–5499. IEEE, 2015.