Bayesian Glasso model for stock return series analysis

Jian Wang

Financial math Ph.D. candidate

Florida state university *jwang@math.fsu.edu*

April 6, 2015

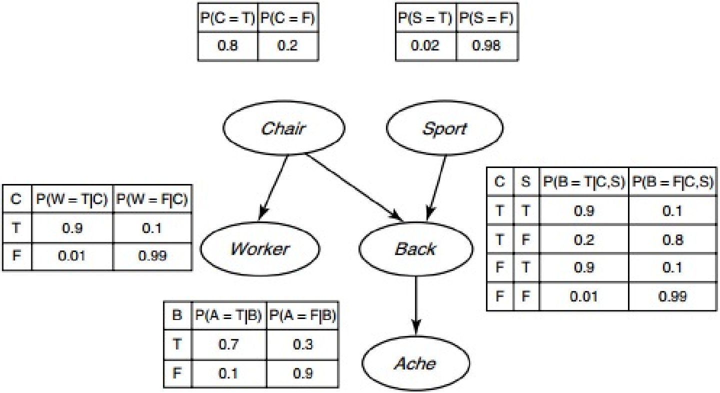
Overview

Introduction

- Our final goal: Use efficient Bayesian network to predict the stock return series
- Lasso model and Glasso model for variable selection
- Bayesian network structure learning based on the previous results of Glasso model
- Model validation with famous bioinfo datasets
- Pick one stock in the network and use linear regression,logistic regression and support vector machine to predict the future price
- Parallel computing and high frequency data

Bayesian network

Bayesian network is a probabilistic graphical model (a type of statistical model) that represents a set of random variables and their conditional dependencies via a directed acyclic graph (DAG)
The following chart is an example of BN:



P(C = T)	P(C = F)
0.8	0.2

P(S = T)	P(S = F)
0.02	0.98

C	P(W = T|C)	P(W = F|C)
T	0.9	0.1
F	0.01	0.99

C	S	P(B = T|C,S)	P(B = F|C,S)
T	T	0.9	0.1
T	F	0.2	0.8
F	T	0.9	0.1
F	F	0.01	0.99

B	P(A = T|B)	P(A = F|B)
T	0.7	0.3
F	0.1	0.9

Bayesian network

Three main research areas for Bayesian network

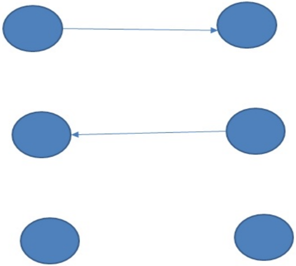
1. Structure Learning
2. Parameter Learning
3. Bayesian inference

Bayesian network

Our main concern is on the structure learning process since it is exponentially increasing complicated and is the most challenging part in Bayesian network research area,here We aimed to use score based method to conduct the structure learning process.

BIC score: $BIC = -2 * ln\hat{L} + K * ln(n)$, where:

$\hat{L} =$ maximum likelihood estimator, k is the number of free parameters and n is the number of data points in x



Lasso and Glasso

Linear regression:$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \epsilon_i, \epsilon_i$ i.i.d $\backsim N(0, \sigma^2)$

Linear regression

$$\hat{\beta}^{ls} = argmin_{\beta}\{\sum_{i=1}^{p} (y_i - \hat{y}_i)^2\} \tag{1}$$

Ridge regression

$$\hat{\beta}^{ridge} = argmin_{\beta}\{\sum_{i=1}^{p} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} \beta_j^2\} \tag{2}$$

Lasso regression

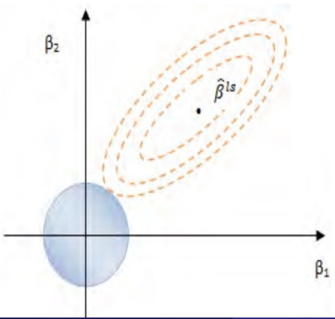
$$\hat{\beta}^{lasso} = argmin_{\beta}\{\sum_{i=1}^{p} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} \beta_j\} \tag{3}$$

Comparison of L1 and L2 Penalized Model

Ridge regression

$\hat{\beta}^{ridge} = argmin_{\beta}\{\sum_{i=1}^{p}(y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} \beta_j^2\}$

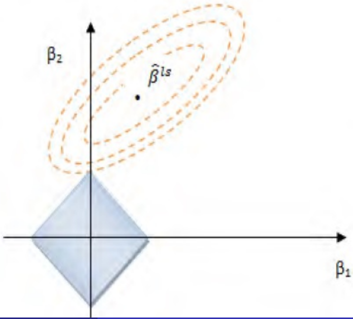
Coefficients:



Lasso regression

$\hat{\beta}^{lasso} = argmin_{\beta}\{\sum_{i=1}^{p}(y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} \beta_j\}$

Coefficients:

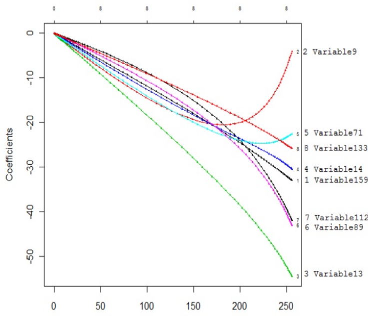


Comparison of L1 and L2 Penalized Model

Ridge regression

$\hat{\beta}^{ridge} = argmin_{\beta}\{\sum_{i=1}^{p}(y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} \beta_j^2\}$

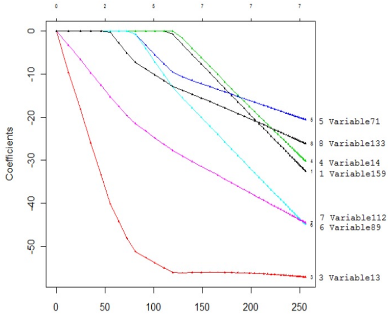
Path::



Lasso regression

$\hat{\beta}^{lasso} = argmin_{\beta}\{\sum_{i=1}^{p}(y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{p} \beta_j\}$

Path::



Lasso and Glasso

Reference:

- **The original paper:**
  Tibshirani, R. (1996). Regression shrinkage and selection via the lasso.
- **The elast angle regression(LAR) algorithm for solving the Lasso:**
  Efron, B., Johnstone, I., Hastie, T. and Tibshirani, R. (2004). Least angle regression,
- **Details and comparison:**
  Hastie, T., Tibshirani, R. and Jerome, F. The elements of statistical learning

R packages

- LASSO in R: glmnet, lasso2, lars
- Relaxed LASSO in R: relaxo

Lasso and Glasso

Reference:

- **The original paper:**
  Tibshirani, R. (1996). Regression shrinkage and selection via the lasso.
- **The elast angle regression(LAR) algorithm for solving the Lasso:**
  Efron, B., Johnstone, I., Hastie, T. and Tibshirani, R. (2004). Least angle regression,
- **Details and comparison:**
  Hastie, T., Tibshirani, R. and Jerome, F. The elements of statistical learning

R packages

- LASSO in R: glmnet, lasso2, lars
- Relaxed LASSO in R: relaxo

Lasso and Glasso

Glasso:

Algorithm for learning the structure in an undirected Gaussian graphical model via using the L_1 regularization to control the zeros in the inverse covariance matrix

Suppose we have N multivariate normal observations of dimension p , with mean μ and covariacne Σ. Let $\Theta = \Sigma^{-1}$ and S be the empirical covariance matrix, the problem is to maximize the log- likelihood

$$logdet\Theta - tr(S\Theta)) - \lambda||\Theta||_1$$

Algorithm

Many algorithms for this problem, The following might be the oldest and simple one by Meinshausen and Buhlmann(2006)

- Estimate a sparse graphical model by fitting a lasso model to each variable, using others as predictors
- Set Σ_{ij}^{-1} to be non zero, if either the estimated coefficient of variable

Lasso and Glasso

Glasso:

Algorithm for learning the structure in an undirected Gaussian graphical model via using the L_1 regularization to control the zeros in the inverse covariance matrix
Suppose we have N multivariate normal observations of dimension p , with mean μ and covariacne Σ. Let $\Theta = \Sigma^{-1}$ and S be the empirical covariance matrix, the problem is to maximize the log- likelihood

$$logdet\Theta - tr(S\Theta)) - \lambda||\Theta||_1$$

Algorithm

Many algorithms for this problem, The following might be the oldest and simple one by Meinshausen and Buhlmann(2006)

- Estimate a sparse graphical model by fitting a lasso model to each variable, using others as predictors
- Set Σ_{ij}^{-1} to be non zero, if either the estimated coefficient of variable

Lasso and Glasso

Reference

- Nicolai Meinshausen and Peter Buhlmann(2006),HIGH DIMENSIONAL GRAPHS AND VARIABLE SELECTION WITH THE LASSO
- Jerome Friedman,Trevor Hastie and Robert Tibshiraniz(2007), Sparse inverse covariance estimation with the graphical lasso
- Rahul Mazumder and Trevor Hastie,(2012),The graphical lasso: New insights and alternatives

R package

- Glasso: by Jerome Friedman, Trevor Hastie and Rob Tibshirani, from Standford, oldest one however updated till now(version 1.8)
- Huge by Tuo Zhao, Han Liu, Kathryn Roeder, John Lafferty, Larry Wasserman from John hopkins and CMU, developed by C instead of fortran, faster!

Lasso and Glasso

Reference

- Nicolai Meinshausen and Peter Buhlmann(2006),HIGH DIMENSIONAL GRAPHS AND VARIABLE SELECTION WITH THE LASSO
- Jerome Friedman,Trevor Hastie and Robert Tibshiraniz(2007), Sparse inverse covariance estimation with the graphical lasso
- Rahul Mazumder and Trevor Hastie,(2012),The graphical lasso: New insights and alternatives

R package

- Glasso: by Jerome Friedman, Trevor Hastie and Rob Tibshirani, from Standford, oldest one however updated till now(version 1.8)
- Huge by Tuo Zhao, Han Liu, Kathryn Roeder, John Lafferty, Larry Wasserman from John hopkins and CMU, developed by C instead of fortran, faster!

Bayesian-Glasso model

For the high dimensional problem, it is not very easy to built the Bayesian network due to its exponentially increasing complexity.
Our idea is to first use the Glasso model to conduct the model selection and then use Bayesian network structure learning process to define the network structure.

Algorithm

- Use Glasso algorithm to find the edges among variables
- Use greedy search methods to change the direction only on those existed edges
- Choose the direction which has the lowest BIC score
- Finish when all the edges are reached or attain the maximum iteration numbers

Bayesian-Glasso model

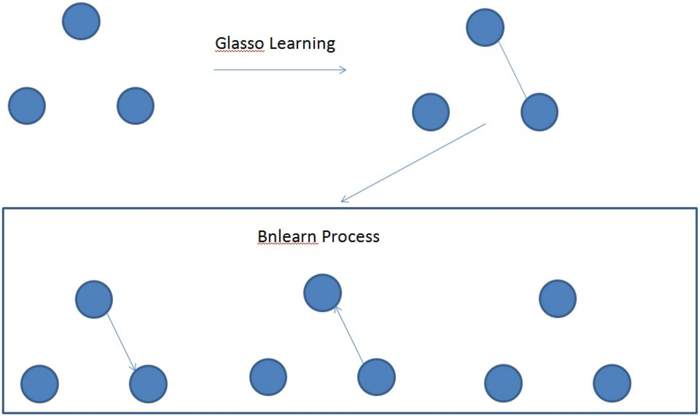
For the high dimensional problem, it is not very easy to built the Bayesian network due to its exponentially increasing complexity.
Our idea is to first use the Glasso model to conduct the model selection and then use Bayesian network structure learning process to define the network structure.

Algorithm

- Use Glasso algorithm to find the edges among variables
- Use greedy search methods to change the direction only on those existed edges
- Choose the direction which has the lowest BIC score
- Finish when all the edges are reached or attain the maximum iteration numbers

Bayesian-Glasso model

One easy example(only contain three nodes) to illustrate the learning process as follows:



Bayesian-Glasso model

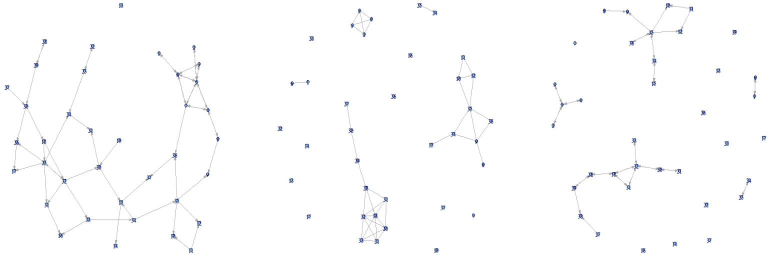
Famous data set test

- **ALARM**, with 37 nodes, 46 arcs and p=509 parameters
- **HEPAR II**, with 70 nodes, 123 arcs and p = 1453 parameters
- **ANDRES**, with 223 nodes, 338arcs and p = 1157 parameters
- **PIGS** with 441 nodes, 592 arcs and 5618 parameters

Bayesian-Glasso model

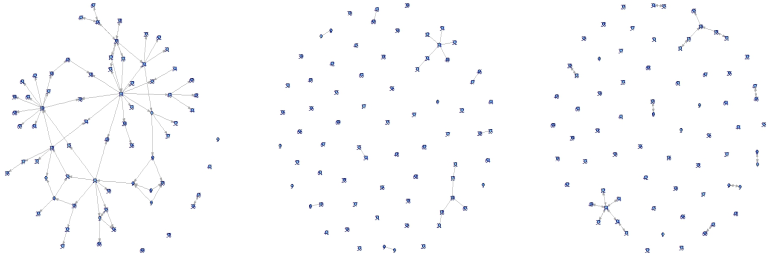
ALARM

The left is the result of bayesnetwork model, the middle is the result of glasso model, the right is the result of Bayesian-Glasso model



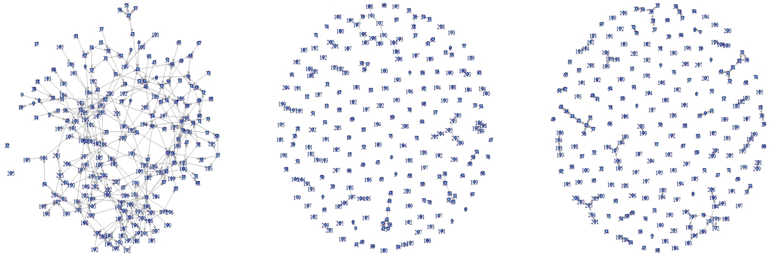
HEPARII

The left is the result of bayesnetwork model, the middle is the result of glasso model, the right is the result of Bayesian-Glasso model



ANDRES

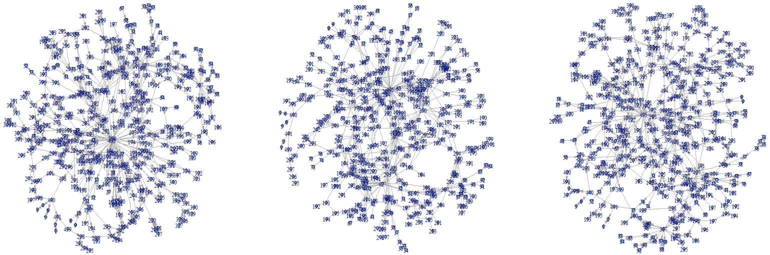
The left is the result of bayesnetwork model, the middle is the result of glasso model, the right is the result of Bayesian-Glasso model



Bayesian-Glasso model

PIGS

The left is the result of bayesnetwork model, the middle is the result of glasso model, the right is the result of Bayesian-Glasso model



Bayesian-Glasso model

SCORE

Table: BIC scores for datasets

Dataset	bn.hc	bn.glasso
ALARM	-54428	-69037
HEPARII	-163823	-167191
ANDES	-468395	-535242
PIGS	-1675890	-1684347

Benefits

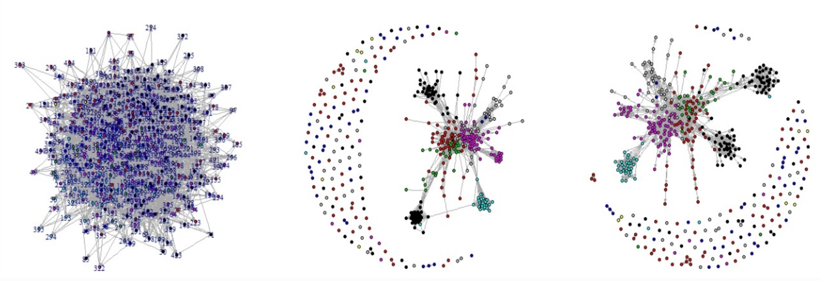
- The score function were all improved
- The number of edges decreased

Stock data analysis

Data set

We chose the daily stock data set on the huge package, it contains the S&P 500 stocks from the date January 1, 2003 to January 1, 2008. This gave us 1258 samples for the 452 stocks(reduced some stocks that are not during the entire period), Besides each stock was categorized into one industry and there are totally 10 industries according to the Global Industry Classification Standard(GICS),Use log return series of the stock for analysis

Bayesian-Glasso learning results

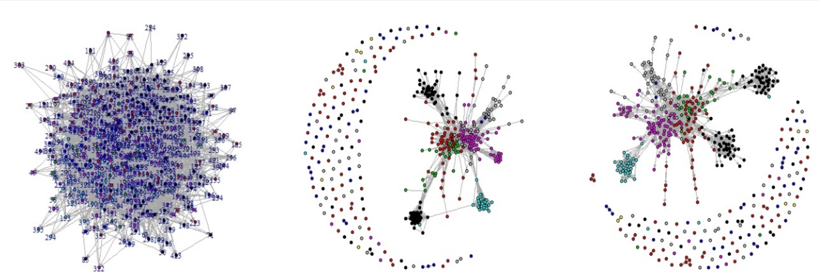


Stock data analysis

Data set

We chose the daily stock data set on the huge package, it contains the S&P 500 stocks from the date January 1, 2003 to January 1, 2008. This gave us 1258 samples for the 452 stocks(reduced some stocks that are not during the entire period), Besides each stock was categorized into one industry and there are totally 10 industries according to the Global Industry Classification Standard(GICS),Use log return series of the stock for analysis

Bayesian-Glasso learning results



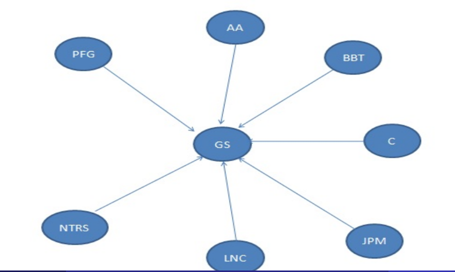
Score

Table: BIC scores for Stockdata

Dataset	bn.hc	bn.glasso
Stockdata	1543594	1510058

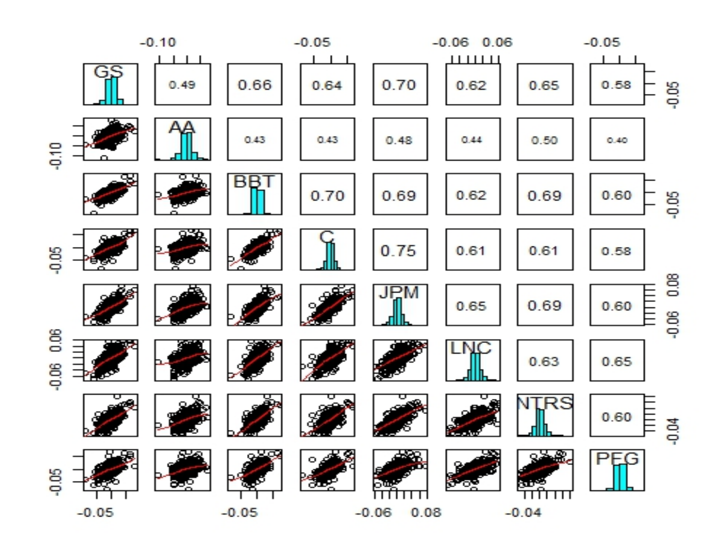
Forecast

Now random chose 1006 days without replacement as the training set and the remaining 251 days as the testing set. We chose stock 186(GS),which had 7 parents, 15 (AA), 53 (BBT), 91(c),229(jpm),249(LNC),302(NTRS), 339(PFG)



Forecast

Correlation and linear regression comparison:



Forecast

We classified the stock returns as increasing, unchanged and decreasing with the threshold 0.5% and -0.5%, to test the accuracy rate of the model.

Among all the 251 days, there are 158 pairs between predicting values and true values that matched each other.

The accuracy rate for the model was around 63%

Future work

- Compare with the time series model, such as Garch
    - Stock return series is badly described by Gaussian(fat tailed and volatility cluster),most time series model assumed the normal or t distribution
    - Only consider the stock itself, do not have some comprehensive-view
- Parallel computing and high frequency data
    - Multiple clusters, mpi and hadoop in software R
    - How to deal with unsynchronized high frequency data