A thesis on Tail Risk interconnectedness using LassoNet

Project-II (CE47006) report submitted to

Indian Institute of Technology Kharagpur in

partial fulfilment for the award of the degree of

Bachelor of Technology

in

Civil Engineering

by

Tuhin Subhra De

(19CE36007)

Under the supervision of

Professor Sujoy Bhattacharya

Vinod Gupta School of Management

Indian Institute of Technology Kharagpur

Spring Semester 2023

April 27, 2023

DECLARATION

I certify that

(a) The work contained in this report has been done by me under the guidance of my supervisor.

(b) The work has not been submitted to any other Institute for any degree or diploma.

(c) I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.

(d) Whenever I have used materials (data, theoretical analysis, figures, and text) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the references. Further, I have taken permission from the copyright owners of the sources, whenever necessary.

Date: April 27, 2023 (Tuhin Subhra De)
Place: Kharagpur (19CE36007)

CERTIFICATE

This is to certify that the project report entitled “**A thesis on Tail Risk interconnectedness using LassoNet**” submitted by **Tuhin Subhra De** (Roll No. 19CE36007) to Indian Institute of Technology Kharagpur towards partial fulfilment of requirements for the award of degree of Bachelor of Technology in Civil Engineering is a record of bona fide work carried out by him under my supervision and guidance during Spring Semester, 2023.

Date: April 27, 2023
Place: Kharagpur

Professor Sujoy Bhattacharya
Vinod Gupta School of Management
Indian Institute of Technology Kharagpur
Kharagpur - 721302, India

Table of Contents

Abstract

Name of the student: **Tuhin Subhra De** Roll No: **19CE36007**

Degree for which submitted: **Bachelor of Technology in Civil Engineering**

Department: **Vinod Gupta School of Management**

Thesis title: **A thesis on Tail Risk interconnectedness using LassoNet**

Thesis supervisor: **Professor Sujoy Bhattacharya**

Month and year of thesis submission: **April 27, 2023**

When evaluating the tail risk of financial institutions, tail risk interconnectivity is a significant aspect and a risk indicator that should be measured. In times of financial crisis, the bad luck is portrayed as a tail risk that might wipe out the entire financial system.. Many of the previous works have shown potential non-linearity in tail risk contagion. Recent advancement of machine learning, especially neural networks have played an important role in data analysis and prediction. They can capture linear as well as non-linear patterns within data and can take the form of any estimating function. However, they are criticized due to their black-box nature. We throw a light into the unexplainable nature of neural networks and use a pre-developed interpretable neural network model called LassoNet for estimating tail risk interconnectedness among public banks of Japan, Taiwan, and Malaysia. We also show that LassoNet stands a robust and accurate-fit estimating model for extracting variable importance and adjustability in regularization as compared to other machine learning models for this scenario.

---

1. Introduction

The spread of bad luck, or "tail risks," among financial institutions during financial crises puts the entire financial system at risk. Systemic risk is created by the unique characteristics of both enterprises and their connections to other organisations. Specifically, a single breakdown could trigger a cascading failure of the financial system. The liquidity risk of failed firms cascading through the money market, the counterparties' credit risk, or simply the panic reflected in price fluctuations are possible causes of the spillover effects across institutions. Therefore, it is inappropriate for regulators or policymakers to evaluate the risk of the company in a singular way. Interconnectedness is acknowledged as a significant quantitative risk indicator. Investors are more concerned with unexpected and unprecedented losses as compared to gains. Hence, we are more focused in enlightening the left fat tail of the returns.

Inspired by some recent research in the domain, we decided to propose a much more robust and accurate fit model for estimating the tail risk within the financial institutions. We investigate a recently developed work which created LassoNet: A neural network with feature sparsity. With development in the field of Machine learning Neural Networks have boomed for adapting to any form of transformation function. However, they are criticized heavily for their black-box behaviour or unexplainable nature in getting correct results. We scrutinize this behaviour and demonstrate the interpretable characteristics for our estimation. Because LassoNet employs a modified objective function with restrictions and directly incorporates feature selection and parameter learning, it may capture significant input variables.

2. Literature Review

Researchers have up several workable solutions to quantify the interconnection of financial institutions and their effect on system risk, which may generally be split into two categories: economic modelling and statistical measurement.
From the standpoint of economic modelling, Acemoglu, Ozdaglar, and Tahbaz-Salehi (2015) [1] offer a framework to demonstrate how the close links between financial institutions aid in the spread of contagion if the size of the negative shock reaches a certain threshold. Yi, Zishuang, and Wang (2018) [2] parse the financial network in the statistical sector using variance decomposition, allowing Generalised Vector Autoregression to determine each node's contribution. The suggested CoVaR (Tobias and Brunnermeier 2016) [3] is based on the idea of Value at Risk and seeks to characterise the tail risk through pairwise linear quantile regression. Linh HN et.al. (2020) [4] used LASSO (Least Absolute Shrinkage and Selection Operator) to capture the important features for assessment of interconnectedness. They have directly incorporated the linear coefficients of the regressor as feature importance.

The existing approaches to estimating conditional tail risk have certain potential drawbacks, as pointed out by Härdle, Wang, and Lining (2016). First off, although tail risk spillover effects are somewhat nonlinear, most of the suggested solutions are based on the linear assumption. Second, a multi-dimensional setup should be used when choosing risk factors. Thirdly, if a flexible and nonlinear model is used, it would be challenging to assess the contribution of particular risk variables. Having an indicator that distils nonlinear and interacting risk components is useful for regulators. Therefore, it necessitates a data-driven model that is not only adaptable enough but also capable of offering simple interpretability. Keeping these under consideration Long Y. et al (2022) [5] moved towards using gradient boosting regression for estimating tail risk. They used the decrease in impurity of the input variables as metric to quantify spillover effects. It took care of potential non-linearity and had no issue of formation of any single-index model for calculating partial derivatives.

3. Problem Statement and overview

Value at Risk (VaR) is a measurement of the likelihood of a loss over a certain time period. Regulators and risk managers in the financial sector frequently use it to evaluate the magnitude and probability of possible loss in their assets.
VAR_q^i can be defined as

$$P\left(X_i \leq VAR_q^i\right) = q$$

where X_i is the log return of a financial institute i and q is the quantile level. The next step involves estimating the VAR of institutes with lagged Macro Variable:

$$X^i = \alpha^i + \gamma M$$

Where M is lagged macro variable. Also, according to Tobias et.al (2016) [3] the linear model adopted is:

$$X^i = \alpha^i + \beta W^{-i}$$

Where W^{-i} is the influence from another institute other than i. Therefore, it is intuitive to interpret the β as the sensitivity of the log return of institute i to the influences from institutes other than i.

3.1 Problem Statement

We start with a generalised parametric model in order to investigate the tail risk contagion problem in a flexible parametric environment.

$$X^j = F\left(W^{-j}\right)$$

F is a generalized function and a linear format of the above equation. So, given a set of financial institutes ***(S)***, Their daily returns ***(X)***, Macro state variables ***(M)***. For some institute **j ϵ S**, we need to estimate an abstract number **(D)** which will quantify the tail risk interconnectedness with institute **i ϵ S** such that **i ≠ j**. To do this, we form a feature vector $\boldsymbol{W_{-j}} \equiv (\boldsymbol{X_i}, \boldsymbol{M})$ representing the influences from institutes other than ***j*** and macro state variables.

Next the generalized function $\boldsymbol{F}$ is developed to predict the returns of $\boldsymbol{j}$ by letting $\boldsymbol{X_j} = \boldsymbol{F}(\boldsymbol{W_{-j}})$ and then $\boldsymbol{D}$ is determined from the internal parameters of F after it is optimized to predict $\boldsymbol{X_j}$ subjected to minimizing the quantile loss function given by:

$$QL_q^{(j)} = \sum_{t=1}^{T} L_q\left(X_j^{(t)} - F\left(W_{-j}^{(t)}\right)\right)$$

where, $L_q(u) = u\big(q - I(u < 0)\big)$. Here some institute j ∈ S

T = Total number of training samples

I is an indicator function, and the quantile loss is particular to quantile q. If the supplied condition is met, $I(\ )$ returns 1, else it returns 0. By varying the weight of the samples, which is managed by the quantile parameter q, the quantile loss function narrows down on the data that are of interest. In this study, q is set at 0.05 to explore the tail risk.

3.2 Modelling Overview

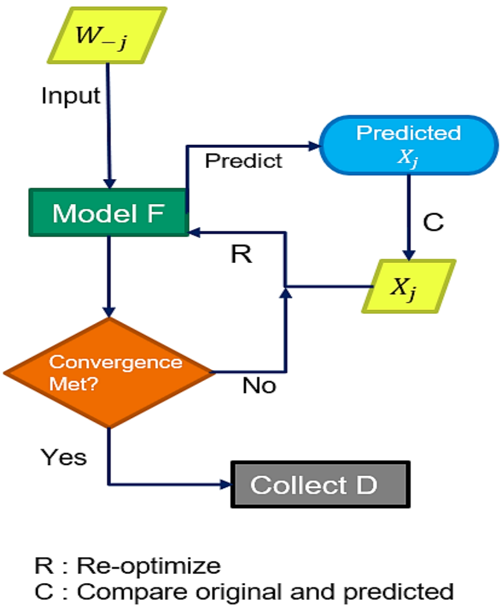

Figure 1: Model Overflow

We follow this structure for finding the optimum value of $\boldsymbol{F}$. $\boldsymbol{W_{-j}}$ acts as an input for $\boldsymbol{F}$, it the predicts the returns and compared to original value using the loss function described above. We then run another iteration for optimizing the model.

In every iteration, before comparing with the ground truth the convergence criterion is checked. If it is met then we stop optimizing and extract D from the model else we keep on optimizing it till meeting of the criterion. The convergence criterion is discussed in detail in the further sections.

3.3 Dataset Information

Reviewing through most of the works done in past, involvement of Chinese and US banks has been in larger amounts than any other financial institutions. Some were also based on multi dollar ventures and cryptocurrencies. We decided to investigate the Asia-Pacific region banks. Banks especially in countries like Japan, Malaysia, and Taiwan play an important role in regulating the Asian as well as the global market. Hence the intuition of tail risk interconnectedness among intranational banks is prevalent. Data of daily returns of the public banks of the three countries viz. Japan, Malaysia, and Taiwan are obtained over a period of 10 years starting from April 1st, 2013 to January 31st, 2023. Most of them were publicly available from Yahoo Finance.

Following Tobias and Brunnermeier (2016) [3] we obtain the Macro State variables for each of the countries. The variables significantly affect the country's economic state. These are:

- The change in three-month treasury bond yield
- The difference between the yields on 3-month and 10-year Treasury bonds
- The difference between the 10-year Treasury bond yield and AAA corporate bond yield
- The difference between the 3-month Treasury bond yield and the 3-month interbank offered rate
- The daily return of the country's corresponding stock index
- The stock index's real estate daily return
- The conditional variance of a country's stock index returns determined using the GARCH (1,1) model represents market volatility.

Most of the bond related data are taken from investing.com and Bloomberg. The banks associated with the corresponding countries are:

Japan	Taiwan	Malaysia
Bank of Japan	Taiwan Cooperative Financial Holding Co. Ltd	CIMB group holdings
Mitsubishi UFJ Financial Group	Chang Hwa Commercial Bank, Ltd	Public Bank
Sumitomo Mitsui Financial Group	Shanghai Commercial & Savings Bank	RHB Bank
Mizuho Financial Group	Mega International Investment Trust Co. Ltd	Hong Leong Bank
Sumitomo Mitsui Trust Holdings	Taichung Commercial Bank Co., Ltd	AMMB Holdings
Resonac Holdings Corporation	King's Town Bank	Affin Bank
Aozora Bank, Ltd	Union Bank of Taiwan	
SBI Shinsei Bank, Limited	Taishin Financial Holding Co., Ltd	
The Chiba Bank, Ltd	CTBC Financial Holding Co., Ltd	
Hokuhoku Financial Group, Inc		

Table 1: List of banks from countries

4. Methodology

Keeping in mind about the potential non-linearity involved in tail contagion and black box nature of non-linear models we decide to move forward with a hybrid model. The model is a combination of linear and non-linear operators and is called LassoNet.

4.1 LassoNet

4.1.1 Architecture

As stated by Lemhadri I et.al. (2021) [6], in linear models, Lasso (or "L1-regularized") regression, which is popular in data science, zero-weights the most redundant or irrelevant features. The Lasso, however, only works with linear models. Here LassoNet is a global feature-selected neural network architecture. This approach achieves feature sparsity by including a skip (residual) layer and restricting a feature's participation in any concealed layer to just while its skip-layer representation is active. This method directly integrates feature selection with parameter learning, unlike other approaches to feature selection for neural networks that use modified objective functions with constraints. It thus provides a complete regularisation path of solutions with a variety of feature sparsity.

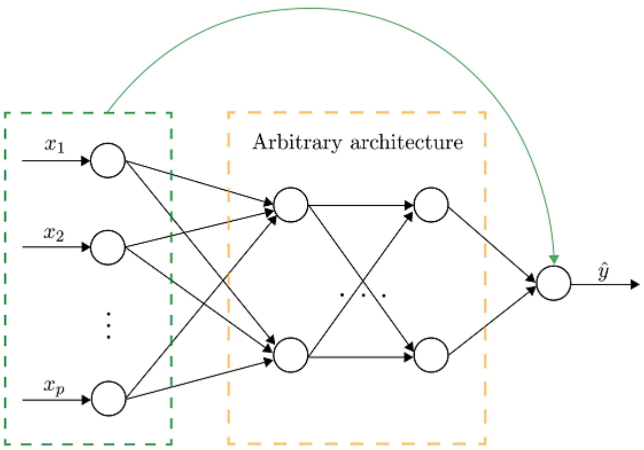

*Figure 2:(Source: Lemhadri et.al 2021 [6])LassoNet architecture. LassoNet's design comprises of a single residual link (shown in green) and an undefined feed-forward neural network (represented in **black**). Together, the first hidden layer and the residual layer go through a hierarchical soft-thresholding optimizer.*

4.1.2 Representation

Hence the model $\boldsymbol{F}$ can be formulated as a class of residual feed-forward neural networks. It is well known that residual networks are simpler to train. According to Raghu et al. (2017) [7] and Lin and Jegelka (2018) [8,] they also serve as universal approximators to a variety of function classes. So, it can be represented by:

$$\boldsymbol{F} = \{f \equiv f_{\theta,w}: x \to \theta^T x + g_w(x)\}$$

Where:

- g_w denotes a feed-forward network with weights w (fully connected)
- $w \in R^{d \times k}$ denotes the weights in the first hidden layer
- k denotes the number of units in the first hidden layer
- $\theta \in R^d$ denotes the weights in the residual layer
- d denotes the data dimension

4.1.3 Components

The LassoNet majorly contains two components that it is dealing with.

- $L(\theta, W)$ the loss function associated with the network. In our case it is a quantile loss function with q = 0.05
- $S_\lambda(x) = sgn\,(x)\ * max(|x| - \lambda, 0)$, is the soft-thresholding operator and λ is the L1 regularization intensity

$$\operatorname{sgn} x := \begin{cases} -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ 1 & \text{if } x > 0. \end{cases}$$

The method of LassoNet contains these important steps that are followed while training the network and finding important features.

1. The original empirical risk minimization is replaced with a **penalty** that promotes feature sparsity. By altering the severity of the penalty, the formulation converts the combinatorial search into a continuous search.

2. A mathematically beautiful use of a **proximal** method allows for a straightforward and effective implementation on top of back-propagation with just minor changes to the original design. This method is described in detail in algorithm 2.

4.1.4 Training

The objective function which the network follows is:

$$\boldsymbol{Minimize} \underset{(\theta, w)}{\longrightarrow} \quad L(\theta,\ w) + \lambda\|\theta\|_1$$
$$\boldsymbol{Subject\ to} \quad |W_{jk}| \le M|\theta_j|, k = 1, 2, 3 \ldots K$$

As discussed previously that LassoNet works on feature sparsity. The constraint $|W_{jk}| \le M|\theta_j|$ says that if the weight of a particular neuron on residual network θ_j is zero then the neuron becomes inactive in the whole network $W_{jk} = 0$. The hierarchical coefficient M in this situation regulates the balance between the linear and nonlinear components. Without extensive understanding of the job or area, it can be challenging to determine the hierarchy coefficient. By considering the hierarchy coefficient as a hyper-parameter, LassoNet avoids this issue. It makes use of a naïve search to thoroughly assess the candidates for the given hyper-parameters' correctness using a validation dataset. This process can be carried out concurrently.

The basic algorithm to train LassoNet is

Algorithm 1: Training LassoNet

---

1. **Input:** training dataset $X \in R^{nxd}$, training labels Y, feed-forward neural network $g_w(\cdot)$, number of epochs B, hierarchy multiplier M, path multiplier ϵ, learning rate α
2. Initialize and train the feed-forward network on the loss L (θ, W)
3. Initialize the penalty, $\lambda = \lambda_0$, and the number of active features, k = d
4.
```
while k>0 do
   Update λ ← (1 + ϵ)λ
      for b ∈ {1 ... B} do
            grad_w = ∇_W(Loss_Function) //[via back propagation]
            grad_θ = ∇_θ(Loss_Function)
            Update θ ← θ – α*grad_θ and W ← W – α*grad_w
            Update (θ, W^(1)) ← Hier-Prox(θ, W^(1), αλ, M)
      end for
   Update k to be non-zero tensor of θ //features still in use
end while
```
5. where `Hier–Prox` is defined in Algorithm 2

LassoNet is an open-source python package made available by the users. It is available at lassonet.ml. The default loss function is programmed to be mean squared error. Hence, we changed it to quantile loss by making a few changes in the source code

The Hierarchical Proximal algorithm takes care of the constraint $|W_{jk}| \leq M|\theta_j|$. The construction of the proximal operator is based on the determination of comparable optimality requirements that completely describe the global solution of the non-convex minimization problem. It turns out that the inner loop can be broken down into different aspects. The algorithm is formulated by:

Algorithm 2: Hierarchical Proximal Operator

Procedure Hier-Prox $(\theta, W^{(1)}; \lambda, M)$

  for j ∈ {1, 2, . . . d} **do**

    Sort $W_j^{(1)}$ such that $\left|W_{(j,1)}^{(1)}\right| \geq \cdots \geq \left|W_{(j,K)}^{(1)}\right|$

   for m ∈ {0, . . . K} **do**

     Compute $w_m = \frac{M}{1+mM^2} * S_\lambda\left(\left|\theta_j\right| + M * \sum_{i=0}^{m}\left|W_{(j,1)}^{(1)}\right|\right)$

   end for

   $\tilde{m} = the\ first$ m ∈ $\{0,...K\} such\ that\ \left(\left|W_{(j,m+1)}^{(1)}\right| \leq w_m \leq \left|W_{(j,m)}^{(1)}\right|\right)$

   $\breve{\theta} \leftarrow \frac{1}{M} * sign(\theta_j) * w_{\tilde{m}}$

   $\widetilde{W_{(j)}^{(1)}} = sign\left(W_{(j,1)}^{(1)}\right) * \min\left(w_{\tilde{m}}, W_{(j)}^{(1)}\right)$

  end for

  return $\left(\breve{\theta}, \breve{W}^{(1)}\right)$

End Procedure

//Conventions d denotes the number of features; K denotes the size of the first hidden layer

Starting the training of LassoNet using Algorithm 1 which uses Algorithm 2 within itself, the values of the weights of the residual layer keeps getting small with increase in the value of λ. Once they are negligible enough to not pass through the soft-thresholding operator they are set to zero. In this way the features are filtered through the path of dense to sparsity. This phenomenon is later used to estimate the tail risk contagion metric and is described in detail in section 4.3

4.2 Why LassoNet?

Feature importance extraction has been a challenging task in machine learning and data science. Tail risk interconnectedness is solely based on how the returns of a particular institute is affected by the other institute's returns. This gives rise to the challenge of interpretation of the models used for this purpose. Neural Networks being a non-linear model itself is criticized for its black-box nature. However, with models like LassoNet we can easily select the important features responsible for prediction of results.

Parameters	GBR	Lasso Regression	Other ML models	LassoNet
Linearity	✓	✓	✓	✓
Non-linearity	✓	✗	✓	✓
Variable Importance (VI)	✓	✓	✗	✓
Adjustability Of VI	✗	✓	✗	✓
Is it OK with no Hyperparameter tuning	✗	✓	✗	✓

Table 2: Comparison of LassoNet with other relevant models. GBR - gradient boosting regression, Lasso Regression signifies L1 norm of weights applied to simple linear regression. Other ML models - all the commonly used machine learning models.

Yunshen Long et.al. (2020) [5] tackled non-linearity using a gradient boosting approach. They used Gradient Boosting Regressor (GBR) for predicting the returns of each Chinese bank using the other bank's return and macro state variables. GBR has property of fetching relative importance of input features. The total of the reductions in impurity measures attributable to the splits connected to one input variable in each regression tree in the GBR model determines its relative relevance. The relative relevance of each input variable is then noted and totalled across all the GBR's decision trees. The more improvement this input variable makes to the decision tree's fitting, the more the impurity measure will reduce as a result of a particular input variable's split.

However, the GBR approach does not have the degree of freedom to adjust the amount of regularization or variable importance. Moreover, it requires finding best model using hyperparameter tuning which may be time-consuming depending upon the sample space of possible hyperparameters to be tried out with. LassoNet on the other hand does not requires hyperparameter tuning as it is a parametric model and the parameters are optimized during the training itself. (Raghu et al., 2017 [7]; Lin and Jegelka, 2018 [8]) also stated that residual networks act as universal approximators to any function as they are hybrid in nature (combination of both linear and non-linear architecture). The weights of the networks are randomly initialized and do not expect the input data to follow any kind of probability distribution.

4.3 Estimation of tail risk interconnectedness

We use both θ (weights of residual connection) and W (weights of fully-connected network) for evaluation metrics. θ can be directly interpreted as feature importance Linh et. Al (2020) [4]. However, it inculcates the linear dependence part of tail risk connectedness. To facilitate the non-linear dependence, we calculate the absolute partial derivatives of the output with respect to the inputs Härdle et. Al (2016) [9] and Yannis D et. Al (1995) [10]. The use of chain rule is handy in this case to back-propagate through the entire network. Hence the metric can be defined by:

$$G_{ij} = |\theta_{ij}| + |\frac{\partial y_{ij}}{\partial W_{-j}^{(i)}}|$$

For some financial institute j with respect to i

The explanatory power of the input variables W_{-j} regarding the variance of the answer j, however, is not disclosed by G_{ij}. R-squared, abbreviated as R_q^j, is used to assess how much of the variation in the answer can be accounted for by these input variables, W_{-j}. In our setting of quantiles, Davino et al. When inserting input variables into the trained model, R^2 is frequently used to gauge the model's correctness. As our focus is on the tail characteristics, the quantile format of R^2 is given by:

$$R_q^j = 1 - \frac{QL_q^j}{TL_q^j} \qquad \text{Where;} \quad TL_q^{(j)} = \sum_t^T L_q\left(X_j^{(t)} - VAR_q^j\right)$$

VAR = Value at Risk at q quantile, for j^{th} financial institute. QL_q^j is the quantile loss defined previously in Problem statement section

Hence, our final estimate D_{ij} (tail risk from j to i) stands:

$$D_{ij} = 0.5 * R_q^j(|\theta_{ij}| + |\frac{\partial y_{ij}}{\partial W_{-j}^{(i)}}|)$$

A factor of 0.5 is introduced to take the mean of the quantities that were added up in G_{ij}. In this case since there are multiple banks involved hence to represent the tail risk contagion a matrix is formed. Hence, the matrix let us say Z can be represented by:

$$Z = \begin{bmatrix} 0 & D^{2|1} & \dots & D^{N|1} \\ D^{1|2} & 0 & D^{3|2} & D^{N|2} \\ \vdots & D^{2|3} & \ddots & \vdots \\ D^{1|N} & D^{2|N} & \dots & 0 \end{bmatrix} \quad \text{where, N : Number of Institutes}$$

The matrix's rows show the effect that each institute has externally, while its columns show how each institute is impacted by other institutes. Another way to visualise the created contagion network is as a directed graph with weighted edges and vertices. Financial institutions are represented in the graph as vertices, and their contributions to the tail risk of the other financial institutions are shown as directed and weighted edges.

The procedure to obtain D_{ij} from LassoNet is stated below:

Algorithm 3: Obtaining Z from LassoNet

Required $W_{-j}^{(i)}$, X_j for all banks
Z = []
for bank_i in list(banks) **do**
    D(row) = []
    for bank_j in list(banks) and not equal bank_i **do**
        Initialize LassoNet model for bank_j
        Result = training_history(LassoNet($W_{-j}^{(i)}$, X_j))
        θ_{ij} = θ such that maximum λ and $\theta \neq 0$ in Result
        Calculate partial derivatives and R^2
        Fill D_{ij} in D(row)
    end for
    Fill D(row) in Z
end for

The number of different LassoNet models used are equal to the number of distinct banks present. Each bank's particular model involves the tail risk interconnectedness based on the input features. Hence it is more likely that Z is not a symmetric matrix, meaning $D^{i|j} \neq D^{j|i}$. Additionally, this suggests that institute-to-institute spillover effects do not necessarily have to be symmetric. Since a bank cannot provide any contribution to itself for tail risk, we set the diagonal elements of Z to be 0.

---

5. Results

There are three countries involved viz. Japan, Taiwan, and Malaysia. We present the results according to the order of countries stated. The Z matrix is developed for all three of them. We worked out the tail risk estimation using both gradient boosting regressor as followed by Yushen Long et. al (2022)[5] and LassoNet. The average value of loss for the models are also present for reference. The top five highest tail risk contagion values are marked in bold and green colour in shade.

For Japan:

Bank name	Abbreviation	Ticker Code
Bank of Japan	BOJ	8301.T
Mitsubishi UFJ Financial Group	MUFG	8306.T
Sumitomo Mitsui Financial Group	SMFG	8316.T
Mizuho Financial Group	MFG	8411.T
Sumitomo Mitsui Trust Holdings	SMTH	8309.T
Resonac Holdings Corporation	RHC	4004.T
Aozora Bank, Ltd	AB	8304.T
SBI Shinsei Bank, Limited	SBI	8303.T
The Chiba Bank, Ltd	CB	8331.T
Hokuhoku Financial Group, Inc	HFG	8377.T

Table 3: Banks in Japan, their abbreviations and Ticker codes

Average value of quantile loss at q = 0.05 by Gradient Boosting Regressor = **4.23**

Average value of quantile loss at q = 0.05 by LassoNet = **3.86**

	BOJ	MUFG	SMFG	MFG	SMTH	RHC	AB	SBI	SB	HFG
BOJ	0.000	0.027	0.028	0.031	0.031	0.031	0.031	0.032	0.030	0.031
MUFG	0.032	0.000	0.029	0.031	0.031	0.029	0.031	0.030	0.031	0.031
SMFG	0.029	0.030	0.000	0.028	0.028	0.028	0.028	0.027	0.028	0.028
MFG	**0.064**	0.015	0.010	0.000	0.015	0.034	0.017	0.015	**0.047**	0.013
SMTH	0.031	0.029	0.025	0.031	0.000	0.031	0.031	0.031	0.028	0.031
RHC	0.032	0.030	0.025	0.033	0.032	0.000	0.032	0.031	0.026	0.032
AB	0.015	0.002	**0.042**	0.028	0.033	0.031	0.000	**0.038**	0.007	**0.034**
SBI	0.032	0.035	0.029	0.032	0.031	0.032	0.031	0.000	0.031	0.030
CB	0.032	0.035	0.032	0.028	0.028	0.027	0.028	0.028	0.000	0.028
HFG	0.031	0.034	0.029	0.030	0.029	0.030	0.031	0.030	0.031	0.000

Table 4: The tail risk contribution between banks of Japan in an adjacent matrix. Each column represents the tail risk contribution received from the banks.

For Taiwan:

Bank name	Abbreviation	Ticker Code
Taiwan Cooperative Financial Holding Co. Ltd	TCFH	5880.TW
Chang Hwa Commercial Bank, Ltd	CHCF	2801.TW
Shanghai Commercial & Savings Bank	SCSB	5876.TW
Mega International Investment Trust Co. Ltd	MIIT	00921.TW
Taichung Commercial Bank Co., Ltd	TCB	2812.TW
King's Town Bank	KTB	2809.TW
Union Bank of Taiwan	UBT	2838A.TW
Taishin Financial Holding Co., Ltd	TFH	2887.TW
CTBC Financial Holding Co., Ltd	CTBC	2891.TW

Table 5: Banks in Taiwan, their abbreviations and Ticker codes

Average value of quantile loss at q = 0.05 by Gradient Boosting Regressor = **5.12**

Average value of quantile loss at q = 0.05 by LassoNet = **4.83**

	TCFH	CHFC	SCSB	MIIT	TCB	KTB	UBT	TFH	CTBC
TCFH	0.000	0.008	0.017	0.001	0.011	0.025	0.005	0.014	0.014
CHCF	0.018	0.000	0.018	0.018	0.020	0.020	0.030	0.016	0.021
SCSB	0.019	0.024	0.000	0.021	0.022	0.021	**0.035**	0.016	0.023
MIIT	0.002	0.001	0.002	0.000	0.011	0.002	0.016	0.003	0.011
TCB	0.021	0.026	0.022	0.017	0.000	0.023	**0.029**	0.019	0.024
KTB	0.020	0.026	0.023	0.018	0.024	0.000	**0.032**	0.018	0.025
UBT	0.010	0.012	0.010	0.015	0.012	0.007	0.000	0.012	0.012
TFH	0.009	0.011	0.008	0.024	0.012	0.008	**0.035**	0.000	0.011
CTBC	0.005	0.019	0.013	0.020	0.018	0.017	**0.036**	0.013	0.000

Table 6: The tail risk contribution between banks of Taiwan in an adjacent matrix. Each column represents the tail risk contribution received from the banks.

For Malaysia:

Bank name	Abbreviation	Ticker Code
CIMB group holdings	CIMB	1023.KL
Public bank	CHCB	1295.KL
RHB Bank	RHB	1066.KL
Hong Leong Bank	HLB	5819.KL
AMMB Holdings	AMMB	1015.KL
Affin Bank	AFB	5185.KL

Table 7: Banks of Malaysia, their names, abbreviations, and Ticker codes

Average value of quantile loss at q = 0.05 by Gradient Boosting Regressor = **4.12**

Average value of quantile loss at q = 0.05 by LassoNet = **3.63**

	CIMB	CHCB	RHB	HLB	AMMB	AFB
CIMB	0.000	0.013	0.017	0.034	0.024	0.014
CHCB	0.015	0.000	**0.066**	0.023	**0.106**	**0.060**
RHB	0.042	0.033	0.000	0.042	0.044	0.043
HLB	0.038	0.035	0.044	0.000	**0.056**	0.042
AMMB	0.041	0.037	0.046	0.040	0.000	0.046
AFB	0.038	0.033	0.046	0.035	**0.062**	0.000

Table 8: The tail risk contribution between banks of Malaysia in an adjacent matrix. Each column represents the tail risk contribution received from the banks

6. Conclusion and Future work

In this work, using LassoNet we present a novel approach towards measuring tail risk associated amongst various financial institutions. LassoNet is parametric neural network model with hybrid architecture accounting for linear as well as non-linear dependencies. The useful property of neural networks being able to assume and take form of any function is key motivation behind this approach. To quantify tail risk interconnectedness, we propose a convenient metric which is a combination of linear and non-linear feature importance. Regulators should pay close attention to institutions that are highly integrated with other financial institutions in terms of tail risk.

We observe from the results that majority of the banks in Japan receive and emit average magnitude of tail risk from each other. In case of Taiwan this distribution is skewed towards some specific banks and is not uniform. Malaysian banks on the other hand have partially skewed distribution of tail risk interconnectedness. Half of the banks receive and emit considerable amount of risk whereas the other half emit high risk. The emission and reception are also dependent of various other economic factors of the banks like Market Capitalization, Total Assets, and liabilities etc.

In future, better models with greater interpretability and more accurate fit can be used to obtain better results. Various other economic measures and metrics can also be integrated with the tail risk contagion estimation allowing greater degrees of flexibility and exploring certain more angles of risks associated.

7. References

[1] Acemoglu Daron, Asuman Ozdaglar, Alireza Tahbaz-Salehi, “Systemic Risk and Stability in Financial Networks,” *American Economic Association,* vol. 105, no. 2, p. 44, 2015.

[2] Shuyue Yi, Zishuang Xu, Gang-Jin Wang, “Volatility connectedness in the cryptocurrency market: Is Bitcoin a dominant cryptocurrency?,” *International Review of Financial Analysis,* vol. 60, no. 1, pp. 98-114, 2018.

[3] Tobias Adrian, Markus K. Brunnermeier, “CoVaR,” *American Economic Association,* vol. 7, no. 1, pp. 1705-1741, 2016.

[4] Linh Hoang Nguyen, Thanaset Chevapatrakul, Kai Yao, “Investigating tail-risk dependence in the cryptocurrency markets: A LASSO quantile regression approach,” *Journal of Empirical Finance,* vol. 58, no. 1, pp. 333-335, 2020.

[5] Yunshen Long, LinQing Zeng, Jing Wang, Xingchen Long, Liang Wu, “A gradient boosting approach to estimating tail risk Interconnectedness,” *Applied Economics,* vol. 52, no. 8, pp. 862-879, 2022.

[6] Ismael Lemhadri, Feng Ruan, Louis Abraham, Robert Tibshirani, “LassoNet: A Neural Network with Feature Sparsity,” *Journal of Machine Learning Research,* vol. 22, no. 127, pp. 1-29, 2021.

[7] Maithra Raghu, Ben Poole, Jon Kleinberg, Surya Ganguli, Jascha Sohl-Dickstein, “On the expressive power of deep neural networks,” in *International conference on machine*, 2017.

[8] Hongzhou Lin, Stefanie Jegelka, “Resnet with one-neuron hidden layers is a universal approximator,” in *Advances in neural information processing systems*, 2018.

[9] Wolfgang Karl Härdle, Weining Wang, Lining Yu, “TENET: Tail-Event driven NETwork risk,” *Journal of Econometrics,* vol. 192, no. 2, pp. 499-513, 2016.

[10] Yannis Dimopoulos, Paul Bourret, Sovan Lek, “Use of some sensitivity criteria for choosing networks with good generalization ability,” *Neural Processing Letters,* vol. 2, pp. 1-4, 1995.