FINTECH RISK MANAGEMENT

www.fintech-ho2020.eu

Slides I: Credit risk in Peer to peer lending

Prepared by Branka Hadji-Misheva, Arianna Agosto,
Paolo Pagnottoni, Anca Toma
Approved by Tomaso Aste and Paolo Giudici

January 30, 2019

Introduction to the FIN-TECH project (Pavia, February 1st, 2019)

Project Overview (Paolo Giudici and Anca Toma, WP1 and WP5)
Credit risk in peer to peer lending (Tomaso Aste and Alessandro Spelta, WP2)
Similarity patterns, Correlation network and credit scoring models (Giudici and Hadji-Misheva, 2018. Hadji-Misheva, Spelta, 2019)
Deep Learning based credit scoring models (Aste and Turiel, 2018)
Market risk in financial robot advisory (Wolfgang Härdle and Paolo Pagnottoni, WP3)
Case study: connectedness models for cryptocurrency exchange prices
Operational risk in blockchain payments (Dominique Guegan and Paola Cerchiello, WP4)
Case studies: cyber risk and ICOs fraud detection
Training by coding (Jochen Papenbrock and Branka Hadji Misheva, WP6)
Validation of Fintech risk management models (Dave Remue and Arianna Agosto, WP7)
Discussion (Claudia Tarantola)

Introduction to the FIN-TECH project (Pavia, February 1st, 2019)

Project Overview (Paolo Giudici and Anca Toma, WP1 and WP5)

Financial Technologies - I

- The Financial Stability Board (2017) defines Financial Technology (FinTech) as “technologically enabled financial innovation that could result in new business models, applications, processes, or products with an associated material effect on financial markets and institutions and on the provision of financial services”
- Peer to peer lending, robot advisory asset management and crypto payments are examples of Financial Technologies, enabled by big data analytics, artificial intelligence and blockchain technologies.

Financial Technologies - II

- Fintech services are competitive, and can increase financial inclusion, but may bring disadvantages: credit risks, market risks, cyber risks and fraud risks. All amplified by systemic risks, due to the high interconnectdness of fintech platforms, which increases contagion.
- Fintech risk management becomes a central point of interest for regulators and supervisors, to protect consumers and preserve financial stability.

FINTECH - HO2020: Motivation I

"Across the board, we are working to strike the right balance between risks and opportunities, so that Europe can benefit fully from new technologies in the financial services sector."

Valdis Dombrovskis, Vice President of the European Commission

FINTECH - HO2020: Motivation II

- There is a strong need to improve the competitiveness of the European fintech sector, introducing a framework for a common risk management approach across all countries, that can supervise fintech companies without stifling their economic potential.
- A framework that can help both fintech and supervisors: on one hand, Fintech firms that want to grow and scale-up across Europe need advanced regulatory technology (RegTech) solutions; on the other hand, the supervisory bodies' ability to monitor innovative financial products proposed by fintechs is limited, and advanced supervisory technology (SupTech) solutions are required.

FINTECH - HO2020: Aim

- Development of a European fintech risk management framework that encourages innovations while protecting their users.
- A framework that can close the gap between technical and regulatory expertise, providing risk management procedures common to Regtech and Suptech, and uniform across countries.

FINTECH - HO2020: Methodology

Deliverables organised in a platform, from four types of events:

- Three workshops, where research and developments in fintech risk management are discussed. Each workshop is built using previous research, organised in repositories.
- Twenty nine training sessions, of 72 hours each, for the supervisors in each of the 29 involved countries. Each session uses a set of common slides.
    - The training slides are based on the research discussed during the workshops, and are split in three parts: i) credit risk ih P2P lending; ii) market risk in robot advisory; iii) operational risk in blockchain payments. Each part contains three risk management case studies, fully reproducible using the data and the code repositories, plus the background theory.

FINTECH - HO2020: Methodology continued

- Six coding sessions carried out at a centralized European level. These sessions allow participants to implement themselves the risk management models explained in the training sessions.
- A validation session, in which all deliverables of the project (research papers, training slides, codes) will be evaluated by the risk management functions of a group of European banks, and the results will be shared during the final workshop.

FINTECH - HO2020: Project Network

- Fintechs and fintech hubs, who have detailed understanding of business models based on financial technologies;
- International regulators and advisors, who have detailed understanding of the regulations and risks that concern financial technologies;
- National supervisors, who have detailed understanding on how to apply regulations to the specific national contexts;
- Universities and research centers, which have detailed understanding of the risk management models that can be applied to financial technologies.

FINTECH - HO2020: Project network

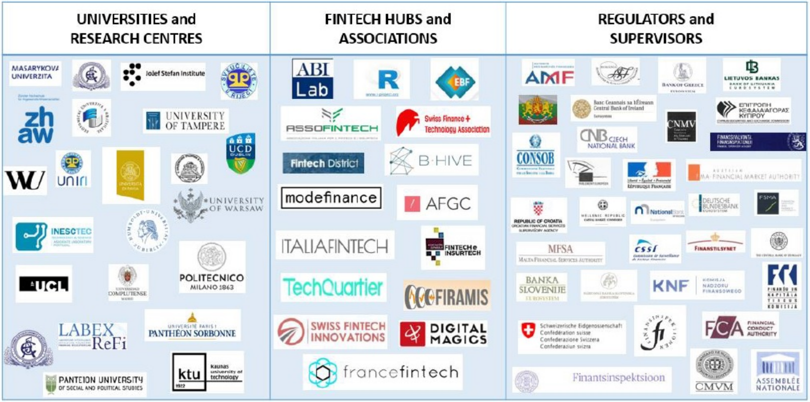

UNIVERSITIES and RESEARCH CENTRES
FINTECH HUBS and ASSOCIATIONS
REGULATORS and SUPERVISORS
MASARYKOVA UNIVERZITA
Jožef Stefan Institute
zh aw
UNIVERSITY OF TAMPERE
UCD DUBLIN
W/
unirı
UNIVERSITY OF WARSAW
INESCTEC
POLITECNICO MILANO 1863
UCL
UNIVERSIDAD COMPLUTENSE MADRID
LABEX ReFi
UNIVERSITÉ PARIS 1 PANTHÉON SORBONNE
PANTEION UNIVERSITY OF SOCIAL AND POLITICAL STUDIES
ktu kaunas university of technology
ABI Lab
R
EBF
ASSOFINTECH
Swiss Finance + Technology Association
Fintech District
B-HIVE
modefinance
AFGC
ITALIAFINTECH
FINTECHe INSURTECH
TechQuartier
COFIRAMIS
SWISS FINTECH INNOVATIONS
DIGITAL MAGICS
francefintech
AMF
BANK OF GREECE EUROSYSTEM
LIETUVOS BANKAS BANK OF LITHUANIA EUROSYSTEM
Banc Ceannais na hÉireann Central Bank of Ireland Eurosystem
CNMV
ΕΠΙΤΡΟΠΗ ΚΕΦΑΛΑΙΑΓΟΡΑΣ ΚΥΠΡΟΥ
CONSOB
CNB CZECH NATIONAL BANK
FINANSSIVALVONTA FINANSINSPEKTIONEN
AUSTRIAN FMA FINANCIAL MARKET AUTHORITY
Liberté • Égalité • Fraternité RÉPUBLIQUE FRANÇAISE
REPUBLIC OF CROATIA CROATIAN FINANCIAL SERVICES SUPERVISORY AGENCY
HELLENIC REPUBLIC CAPITAL MARKET COMMISSION
National Bank
DEUTSCHE BUNDESBANK EUROSYSTEM
FSMA
MFSA MALTA FINANCIAL SERVICES AUTHORITY
cssf
FINANSTILSYNET
BANKA SLOVENIJE EUROSYSTEM
KNF KOMISJA NADZORU FINANSOWEGO
FKTK
Schweizerische Eidgenossenschaft Confédération suisse Confederazione Svizzera Confederaziun svizra
fi
FCA FINANCIAL CONDUCT AUTHORITY
Finantsinspektsioon
CMVM
ASSEMBLÉE NATIONALE

Figure 1: The FINTECH-HO2020 Consortium

Credit risk in peer to peer lending (Tomaso Aste and Alessandro Spelta, WP2)

P2P Platforms

- Among FinTech applications that rely on big data analytics, innovative ones are those based on **peer-to-peer** (P2P) financial transactions, such as peer to peer lending, crowdfunding and invoice trading.
- The concept peer-to-peer captures the **interaction** between units, which eliminates the need for a central intermediary
- **Advantages** of P2P Lending Platforms:
  - Improved financial inclusion
  - Higher rates of return compared to bank deposits
  - Lower fees
  - High speed of service
  - Customized user experience

Banks vs P2P Platforms: Risk concerns and data availability

- Both classic banks and P2P platforms rely on **credit scoring models** for estimating credit risk but the incentive for model accuracy is different:
  - Banks assumes the risk so they interested in having the most accurate possible model
  - In a P2P lending platform, the risk is fully borne by the lender
- P2P Platform often do not have access to borrowers' data usually employed by banks
- P2P Platforms operate as **social networks**:
  - Data from such activity can be leveraged for improving credit risk accuracy
  - If we lack P2P direct interaction data we can exploit similarity patters between borrower feature

Similarity patterns, Correlation network and credit scoring models (Giudici and Hadji-Misheva, 2018. Hadji-Misheva, Spelta, 2019)

Aim of Research

- Analyze the **predictive performance** of scoring models employed by P2P platforms.
- Built **similarity network** from the available platform data.
- Extract **topological information** for describing the relationships between players' local interactions and the global network structure
- Investigate whether **pattern of similarities** between borrowers features can **improve loan default predictions**.

Similarity Network

- In the distance network nodes represent borrowing companies and the **edges** the similarities between adjacent nodes.
- There exist different metrics to build up **distances** between objects:
  - Correlation
  - Cosine
- There exist different algorithm to extract a **sparse representation** of the fully connected distance matrix revealing the strongest pattern of similarities
  - Minimum Spanning Tree
  - Maximum Planar Graph
  - F-test

Similarity Networks

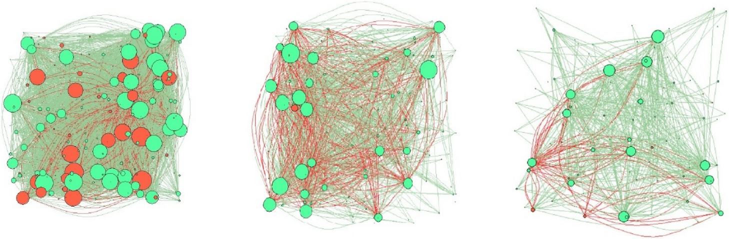


Figure 2: Left: Correlation network based on the activity indicator. Number of nodes= 386; Middle: Correlation network based on the solvency indicator. Number of nodes= 288; Right: Correlation network based on the return on equity ratio. Number of nodes= 226

Topological Coefficients

- We extract various type of information from such networks
- **Nodes Importance**
  - How many partners a Company has (degree)?
  - How strong are the weights embedded in such connections (strength)?
  - How crucial a node is in letting information to spreads over the network (betweennes)
- **Community Structure**
  - Dense sub-graphs sharing some common characteristics
  - For each node we have the identifier of the community the node belongs
- Such information are used to complement balance sheet information of each Company.
- We compare different credit scoring models (logistic, knn, svm, random forest...) using non parametric measures (roc, accuracy, precision).

ROC comparison

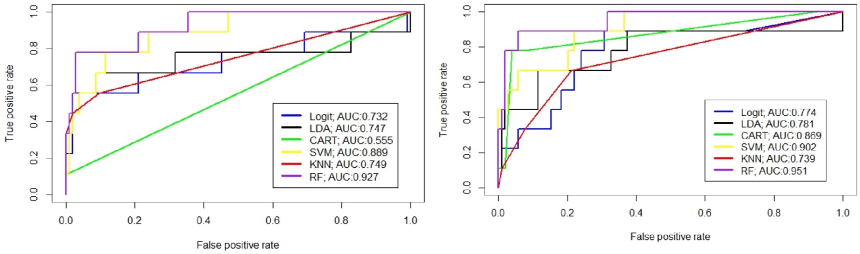
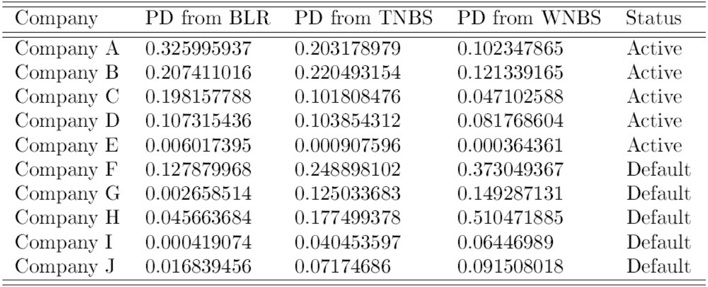


Figure 3: Comparison predictive utility of models with and without network information

Predictive Modeling



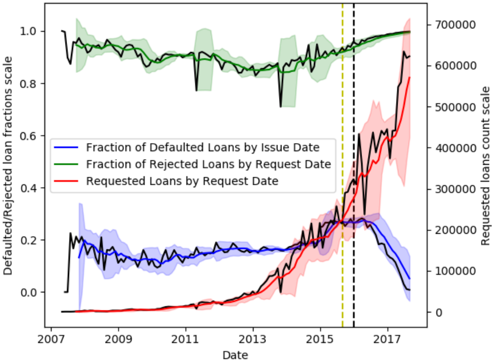
Company	PD from BLR	PD from TNBS	PD from WNBS	Status
Company A	0.325995937	0.203178979	0.102347865	Active
Company B	0.207411016	0.220493154	0.121339165	Active
Company C	0.198157788	0.101808476	0.047102588	Active
Company D	0.107315436	0.103854312	0.081768604	Active
Company E	0.006017395	0.000907596	0.000364361	Active
Company F	0.127879968	0.248898102	0.373049367	Default
Company G	0.002658514	0.125033683	0.149287131	Default
Company H	0.045663684	0.177499378	0.510471885	Default
Company I	0.000419074	0.040453597	0.06446989	Default
Company J	0.016839456	0.07174686	0.091508018	Default

Table 1: Comparison of PD Estimates across different models. BLR indicated the baseline regression model; TNBS the network based model, with all types; WNBS the network based model, with only Type C edges.

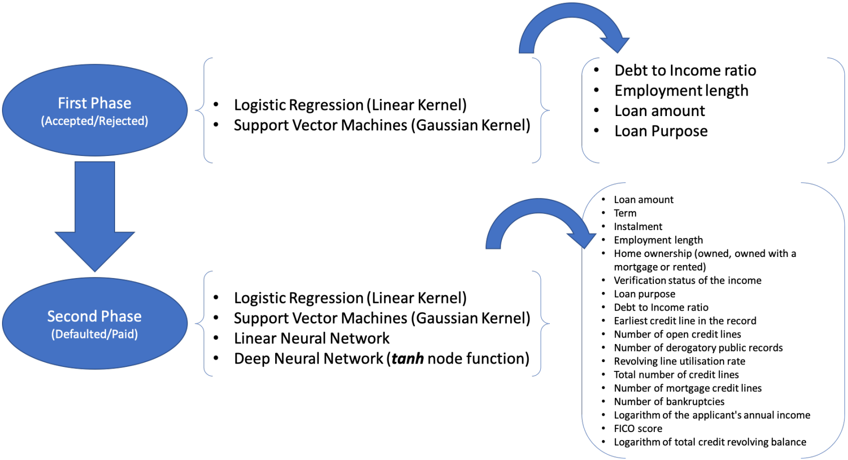
Deep Learning based credit scoring models (Aste and Turiel, 2018)

Aim of Research

- Automation of loan screening and acceptance through Machine Learning.
- Accurate prediction of default risk through Big Data analytics and Machine Learning techniques.
- P2P lending data investigated to understand limitations of default predictability.



Method: Two-Phase Model



First Phase
(Accepted/Rejected)

- Logistic Regression (Linear Kernel)
- Support Vector Machines (Gaussian Kernel)

- Debt to Income ratio
- Employment length
- Loan amount
- Loan Purpose

Second Phase
(Defaulted/Paid)

- Logistic Regression (Linear Kernel)
- Support Vector Machines (Gaussian Kernel)
- Linear Neural Network
- Deep Neural Network (***tanh*** node function)

- Loan amount
- Term
- Instalment
- Employment length
- Home ownership (owned, owned with a mortgage or rented)
- Verification status of the income
- Loan purpose
- Debt to Income ratio
- Earliest credit line in the record
- Number of open credit lines
- Number of derogatory public records
- Revolving line utilisation rate
- Total number of credit lines
- Number of mortgage credit lines
- Number of bankruptcies
- Logarithm of the applicant's annual income
- FICO score
- Logarithm of total credit revolving balance

Figure 4: Representative diagram outlining the two phases of the model with machine learning methods applied and features considered for each phase

Neural Network Visualisation

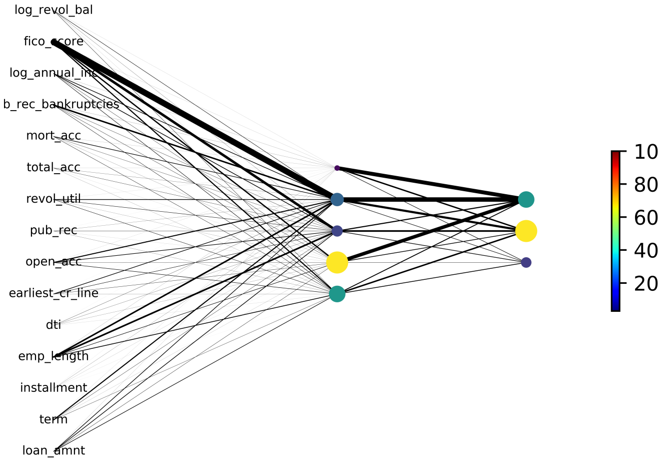
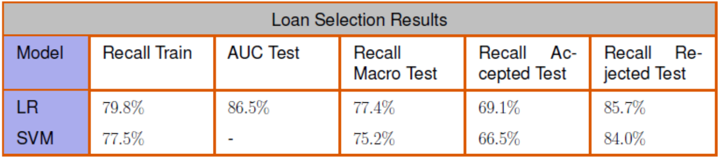


Figure 5: Neural network representation with node size and colour representing total outgoing weight and edge width proportional to the weight.

Loan Selection

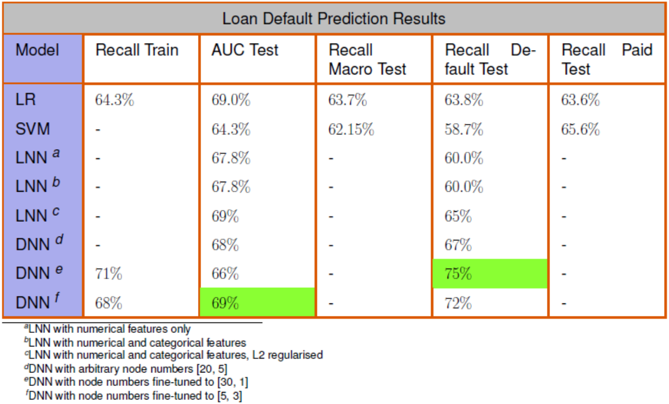


Loan Selection Results					
Model	Recall Train	AUC Test	Recall Macro Test	Recall Accepted Test	Recall Rejected Test
LR	79.8%	86.5%	77.4%	69.1%	85.7%
SVM	77.5%	-	75.2%	66.5%	84.0%

Table 2: Results for the ML algorithms applied to the 1^{st} model phase.

- Simple Logistic Regression model replicates analyst rejections with recall above 85%.
- Target feature class imbalance in training set affects class scores. Would benefit from more training data.
- Replicability of screening leads to more complex models applied to default prediction.

Loan Default Prediction



Loan Default Prediction Results					
Model	Recall Train	AUC Test	Recall Macro Test	Recall Default Test	Recall Paid Test
LR	64.3%	69.0%	63.7%	63.8%	63.6%
SVM	-	64.3%	62.15%	58.7%	65.6%
LNN [a]	-	67.8%	-	60.0%	-
LNN [b]	-	67.8%	-	60.0%	-
LNN [c]	-	69%	-	65%	-
DNN [d]	-	68%	-	67%	-
DNN [e]	71%	66%	-	75%	-
DNN [f]	68%	69%	-	72%	-

[a] LNN with numerical features only
[b] LNN with numerical and categorical features
[c] LNN with numerical and categorical features, L2 regularised
[d] DNN with arbitrary node numbers [20, 5]
[e] DNN with node numbers fine-tuned to [30, 1]
[f] DNN with node numbers fine-tuned to [5, 3]

Table 3: Results for the ML algorithms applied to the 2^{nd} model phase.

- Increasing complexity of the model captures complex phenomenon of default, with NN outperforming in recall.

Market risk in financial robot advisory (Wolfgang Härdle and Paolo Pagnottoni, WP3)

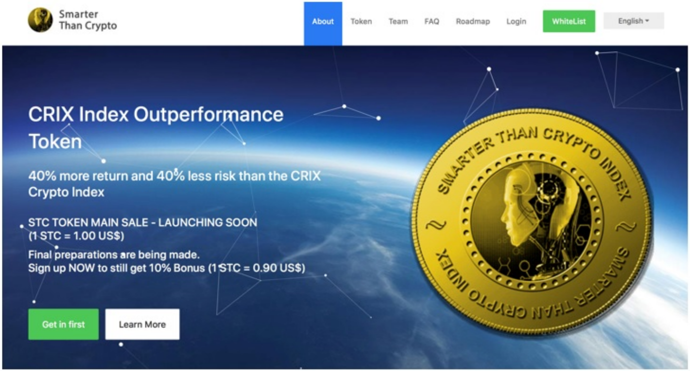
Content of WP3

- Methods
  - AI in finance, application to robo advisory, main risk concerns
  - Cluster analysis, distance models and community detection
  - Volatility and connectedness models, VAR and VECM models
- Risk management (case-studies)
  - Market risk and contagion models in financial markets
  - Market risk and contagion models in crypto markets
  - Asset allocation and compliance risk management

Data for case studies and coding sessions

- Data on price and trading volume of cryptocurrencies
- Start date: 2013/12/27
- Daily data
- 3792 coins (as on 2019/01/21)

CRIX (Simon Trimborn, Wolfgang Härdle)



CRIX Index Outperformance Token

40% more return and 40% less risk than the CRIX Crypto Index

STC TOKEN MAIN SALE - LAUNCHING SOON
(1 STC = 1.00 US$)

Final preparations are being made.
Sign up NOW to still get 10% Bonus (1 STC = 0.90 US$)

▸ CRIX Code ▸ thecrix.de

CRIX (Simon Trimborn, Wolfgang Härdle)

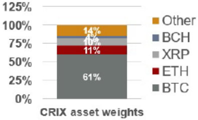
- a market cap weighted index
- Dominance of BTC...
- reallocation: 3M evaluation of k = constituents

$$Division = \frac{\sum_i MV_{i0}}{1000}$$

CRIX Methodology

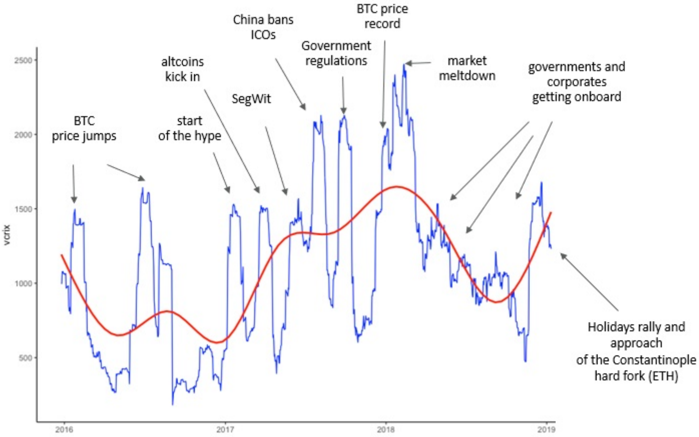
- $INDEX_t^{Laspeyers} = \frac{\sum_i P_{it} Q_{i0}}{\sum_i P_{i0} Q_{i0}}$
- $INDEX_t^{CRIX} = \frac{\sum_i MV_{it}}{Division}$
- Only price changes cause a change in index development.
- Divisor: changes in coin volume does not affect index price

2018/10/31 Underlying Crypto Assets and Weights



30 underlying cryptocurrencies constituent realloc 3 M Constituent ranking every M

CRIX (Simon Trimborn, Wolfgang Härdle)



▸ VCRIX Code ▸ thecrix.de

VCRIX

- log-returns of CRIX from 12.2015 to 01.2019 (T = 1626), (RV=realised volatility, in case of VCRIX a 30-day rolling volatility)
- VCRIX $= \frac{RV^d_{t+1d}}{\text{Divisor}}$
  $RV^d_{t+1d} = \alpha + \beta^d RV^d_t + \beta^w RV^w_t + \beta^m RV^m_t + \omega_{t+1d}$
  $RV^w_t = 1/7 \left(RV^d_t + RV^d_{t-1} + \ldots + RV^d_{t-6d} \right)$
  $RV^m_t = 1/30 \left(RV^d_t + RV^d_{t-1d} + \ldots + RV^d_{t-29d} \right)$
- VCRIX1 = 1000
- Divisor adjusts to changes in constituents

Covariate-assisted Spectral Clustering in Dynamic Networks: An Application to CCs Market

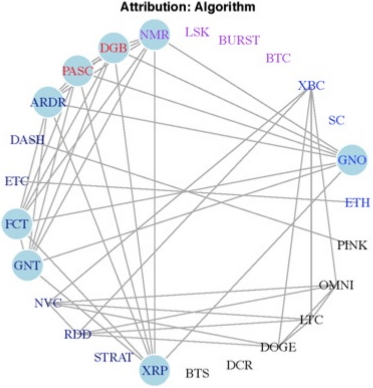


Figure 6: Node Features (Attribution Network Structure). Node size - eigenvector centrality of a CC

Covariate-assisted Spectral Clustering in CCs Market

- Return network structure from Adaptive LASSO
  - Find connection between returns of top 200 cryptos
  - $Ret_{eth} = \beta_1 Ret_{btc} + \beta_2 Ret_{xrp} + \beta_3 Ret_{qtum} + \dots$
  - Result 24 cryptos

Covariate-assisted Spectral Clustering in CCs Market

- Return network structure from Adaptive LASSO
  - Find connection between returns of top 200 cryptos
  - $Ret_{eth} = \beta_1 Ret_{btc} + \beta_2 Ret_{xrp} + \beta_3 Ret_{qtum} + \dots$
  - Result 24 cryptos

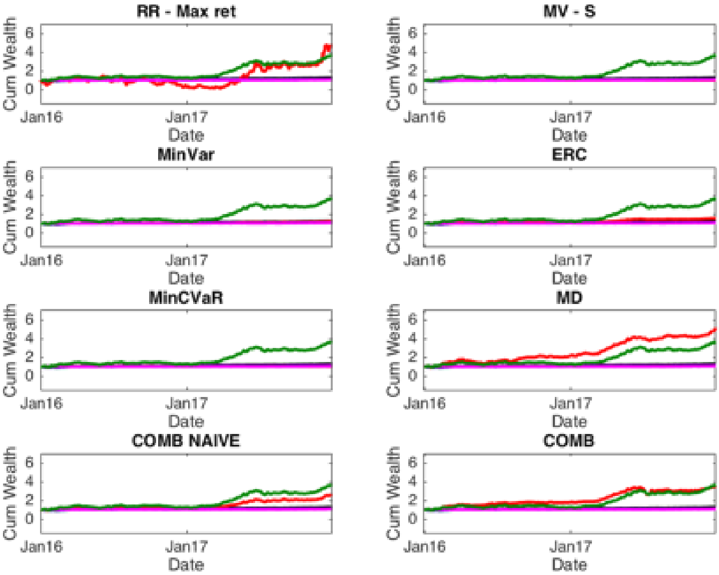
Dynamic Stochastic Block Model

$$A_t(i,j) = \begin{cases} \text{Bernoulli } \{P_t(i,j)\}, & \text{if } i < j \\ 0, & \text{if } i = j \\ A_t(j,i), & \text{if } i > j \end{cases}$$

$$\mathscr{A}_t \overset{\mathrm{def}}{=} \mathrm{E}(A_t|Z_t) = Z_t B_t Z_t^\top$$

- Adjacency matrix based on return information A_t
- Connection between i and j $P_t(i,j) = P\{A_t(i,j) = 1\}$
- Clustering Matrix: $Z_t \in \{0,1\}^{N \times K}$
- Block Probability Matrix: $B_t \in \mathscr{M}^{K \times K}$
  $B_t(k,k') = P_t(i,j), \forall k,k' = \{1, \cdots, K\}$

Asset allocation with cryptos



Asset allocation with cryptos

Model	Reference	Abbreviation
Equally weighted	DeMiguel et al. (2009)	EW
Risk-return-oriented strategies		
Mean – Var – max Sharpe	Jagannathan and Ma (2003)	MV – S
Return-oriented strategies		
Risk – Return – max return	Markowitz (1952)	RR – max ret
Risk-oriented strategies		
Mean – Var – min var	Merton (1980)	MinVar
Equal Risk Contribution ERC	Roncalli et al. (2010)	ERC
Mean – CVaR – min risk	Rockafellar and Uryasev (2000)	MinCVaR
Maximum Diversification	Rudin and Morgan(2006)	MD
Combination of models		
Naïve combination	Schanbacher (2015)	COMB NAÏVE
Combination bootstrap	Schanbacher (2014)	COMB

Research ideas and ongoing projects

- LSTM for trading and portfolio allocation
- Modelling Systemic Risk using Neural Network Quantile Regression
- Ensemble machine learning in portfolio allocation with cryptocurrencies

Case study: connectedness models for cryptocurrency exchange prices

Research questions

Main research questions:

- How much are Bitcoin exchanges interconnected? And which are the exchanges showing high/low degree of interconnectedness among each other?
- Which are the price setter exchanges and which are the followers?
- To answer the questions we analyze the spillovers proposed by Diebold and Yilmaz (2012, 2014), providing an extension of their methodology

Modelling strategy

- Granger Representation Theorem (Engle and Granger, 1987)

Vector Error Correction Model (VECM)

$$\Delta p_t = \alpha\beta' p_{t-1} + \sum_{i=1}^{k-1} \zeta_i \Delta p_{t-i} + \varepsilon_t \tag{1}$$

- $\Delta p_t = (\Delta p_t^1, \Delta p_t^2, ..., \Delta p_t^n)'$
- α : $(n \times h)$ adjustment coefficients matrix
- β : $(n \times h)$ cointegrating matrix
- ζ_i : $(n \times n)$ parameter matrices
- k : autoregressive order
- h : cointegrating rank
- ε_t : zero-mean white noise process having variance-covariance matrix Σ_ε

Spillover measures

- using $\theta_{ij}^{g}(H)$ to denote the KPSS H-step forecast error variance decompositions, with $H = 1, \cdots, n$, we have:

Variance shares

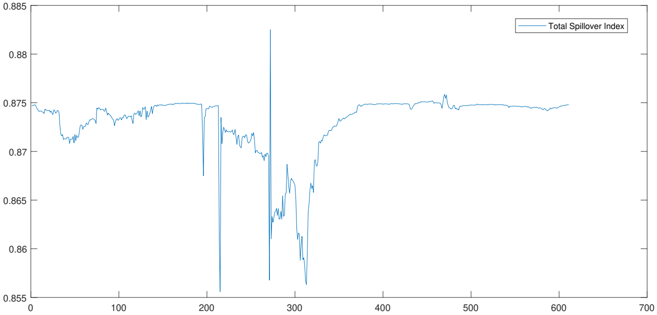
$$\theta_{ij}^{g}(H) = \frac{\sigma_{jj}^{-1} \sum_{h=0}^{H-1} (e_i' \Psi_h \Sigma_\varepsilon e_j)^2}{\sum_{h=0}^{H-1} (e_i' \Psi_h \Sigma_\varepsilon \Psi_h' e_j)} \tag{2}$$

- σ_{jj} is the standard deviation of the innovation for equation j and e_i represents the selection vector with one as element i and zeros elsewhere.
- Note that we extend the methodology from Diebold and Yilmaz (2012) - from a Vector AutoRegressive (VAR) framework to a VECM framework - as well as the analysis of Giudici and Abu-Hashish (2018)

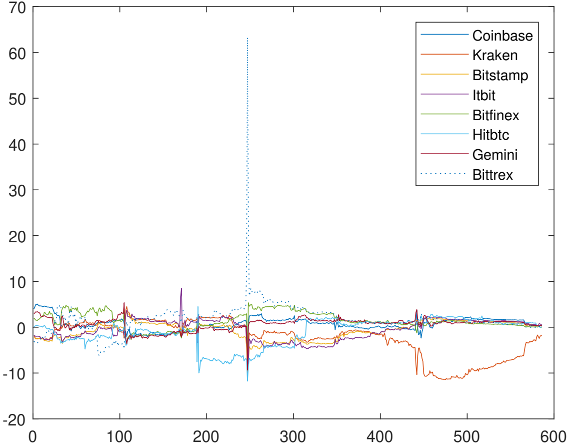
Data

- 8 price series (USD) belonging to selected Bitcoin exchanges
  - Bitfinex
  - Coinbase
  - Bitstamp
  - Kraken
  - Hitbtc
  - Gemini
  - Itbit
  - Bittrex
- Market exchanges make up at least 60% of BTC daily trading volume during the sample period
- Time period analyzed: 18 May 2016 - 30 April 2018
- sampling interval: daily

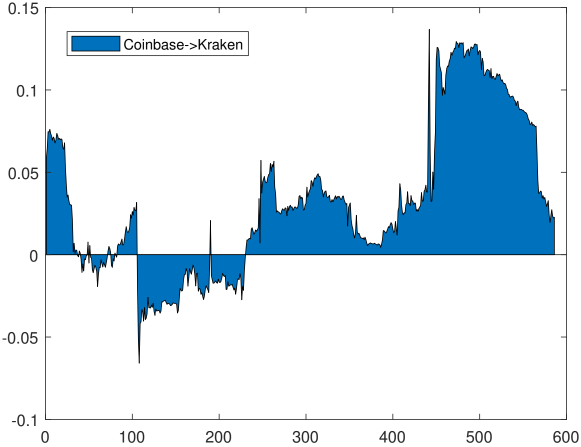
Total Spillover Index



Net Spillover Index



Pairwise Spillover Indexes



Operational risk in blockchain payments (Dominique Guegan and Paola Cerchiello, WP4)

Blockchain Technology

- Securized technology
  - Peer-to-peer technology
  - Consensus protocol
  - Cryptography
- All blockchains are not equivalent
  - What is the objective for using the blockchain technology?
  - Creation of crypto-assets
  - Payments
  - Transfers of goods
- Risks associated to the blockchains

Risks associated to the blockchain environment

- 51% attack
- Errors in codes (smart contracts)
- Hacking of the platforms (not directly link to the technology blockchain)
- Lost of private keys
- Theft of private keys

Frauds link to the financial system

- Payments
    - Fraudulous exchange payments
    - The money laundering (ML), terrorist financing (TF) : 0.4% of statements of suspicion in 2017, (TracFin , December 2018). On Coinhouse: 50 cases of money laundering have been identified during the less 18 months.
    - Evasion sanctions (circumventing exchanges and capital controls)
    - Erroneous transactions and transactions never executed
- Economics
    - Impact on the monetary policy and financial system
    - Stability on the financial system
    - High risk investment opportunities (pump and dump)

Scam to crypto exchanges

- Main fraud in 2018: between 500 M to 1B.
- sellers show off juicy returns to private investors
- They propose to you to give a good return on a small amount, then you receive it
- Then, you send more money, and you receive again a good return
- When finally you send a very high amount of money, the sellers disappear: they close their account and you cannot find them.
- To get back the money is impossible: persons have to file a complaint

Initial Coins Offering or Initial Tokens Offering

- Since early 2016, a new way of raising funds has rapidly emerged as a major issue for FinTech founders and financial regulators
- A new method
  - to raise funds through the offer and sale by a group of developers or a company to a crowd (i.e. investors or contributors) of ad hoc crypto-assets (also coined as "tokens") specifically created and issued on a distributed ledger,
  - sometimes preceded by an early sale of the crypto-assets called "pre-sale",
  - for the purpose of launching a business or of developing ad hoc governance of projects based,
  - in exchange for pre-existing 'mainstream' crypto-assets, such as Bitcoin and Ether among others, or even fiat currencies.
    Perceived by several entrepreneurs as a less burdensome way of fundraising, at least 25 billion dollars have been raised between March 2016 and August 2018 through ICOs only.

- Perceived by several entrepreneurs as a less burdensome way of fundraising, at least 25 billion dollars have been raised between March 2016 and August 2018 through ICOs only (coinschedule.com).

Case studies: cyber risk and ICOs fraud detection

Frauds and ICOs

- Importance of white paper
- New regulation (France, US, ASIA)
- Frauds concern the ICO which have no valuable project
- In 2016 – 2017 specific behavior
- In 2018 and in the future, due to the regulations which arise, frauds will diminish.
- Investigation on the ICOs which work, identifying the empty shells.
- New phenomena: DAICO

Risks and regulation

- Crypto-assets cannot be regulated
- Regulation of the payments platforms
- Information on the ICO: in France possibility to have a Label by AMF (optional)
- Information on the frauds link to the use of cryptocurrencies
- Banks and account in cryptocurrencies
- Uniform regulation between the different countries
- New fiscal legislation in France for tokens emitters and tokens acquirers.

Case studies for risk management

- Open blockchains
    - Security of blockchains to avoid frauds: study of 51% attack investigating the protocols: definition of an economic indicator – ranking.
    - For ML/TF (whose volume is negligible in crypto-assets compare to the whole financial system), study of the volume exchanges considering the dynamic sequence of the cryptographic keys.
    - Speculative phenomena: studies of the bubbles, pump and dump events, strategies of investment based on crypto assets which are largely risky.
    - Importance of the second market: future of the tokens issued by ICOs.
    - Frauds on ICOs: the empty shells
- Close blockchains
    - Creation of commodities back digital assets
    - Central banks and monopolistic new market
- Creation and sharing of Database
- Development of new approaches for measuring the risks associated to the crypto-assets link with the blockchain

Case study I: Fraud detection in ICOs

- Initial Coin offerings are a new yet uncovered mean to raise funds through tokens: a conjunction of **crowdfunding** and **blockchain**.
- ICOs are a relatively new phenomenon but have quickly become a dominant topic of discussion within the fintech community.
- Few numbers (based on Coinschedule.com)
  - around **6** bi USD raised in 2017 by **456** ICOs
  - around **21.7** bi USD raised till the end of 2018 by **1076** ICOs
- The risky counter part is the presence of criminal activity.
- Financial market authorities are very prudent and some countries ban straightaway all ICOs from their jurisdiction.

Methodology - Response Variable

The analyzed status of an ICO is made up of 3 classes, intended as follows:

- **Success**: the ICO collects the predefined cap within the time horizon of the campaign;
- **Failure**: the ICO does not collect the predefined cap within the time horizon of the campaign;
- **Scam**: the ICO is discovered to be a fraudulent activity during the campaign and described as such by all the platforms we use for data gathering (namely ICObench and Telegram).

Methodology - Explanatory variables

Table 4: Employed Covariates

class0	f=failed, sc=scam su=success
class1	0=success, 1=scam
class2	0=failed, 1= success
w_site	Website (dummy)
tm	Telegram (dummy)
w_paper	White paper (dummy)
usd	presale price in USD
tw	Twitter (dummy)
fb	Facebook (dummy)
ln	Linkedin (dummy)
yt	Youtube (dummy)
gith	Github (dummy)
slack	Slack (dummy)
reddit	Reddit (dummy)
btalk	Bitcointalk (dummy)
mm	Medium (dummy)
nr_team	Number of Team members
adv	Existence of advisors (dummy)
nr_adv	Number of advisors
project	Official name of the ICO
nr_tm	Number of users in Telegram
tot_token	Number of Total Tokens
Pos_Bing	Standardized number of positive words for BL list
Neg_Bing	Standardized number of negative words for BL list
Sent_Bing	Standardized sentiment for BL list
Pos_NRC	Standardized number of positive words for NRC list
Neg_NRC	Standardized number of negative words for NRC list
Sent_NRC	Standardized sentiment for NRC list

Results - I

Table 5: Results from Logistic regression on Success/Failure

	Dependent variable:
	class2
tw	2.63$^{\cdot}$
	(1.49)
w_paper	1.51*
	(0.65)
Sent_NRC	2.36***
	(0.61)
Nr_adv	0.53***
	(0.15)
Nr_team	0.30**
	(0.10)
Constant	-4.40
	(1.64)
Observations	196
Residual Deviance	71.14
Akaike Inf. Crit.	83.14
Note:	*p<0.1; **p<0.05; ***p<0.01

Results – II

Table 6: Results from multilogit regression: failure and scam compared to success

	Dependent variable:	
	f	sc
	(1)	(2)
Oweb_dum	0.363	-1.731^{*}
	(0.859)	(1.042)
tw	-3.046^{**}	-2.768^{**}
	(1.310)	(1.350)
adv_dum	-1.679^{***}	-0.943
	(0.607)	(0.855)
Paper_du	-2.060^{***}	-0.737
	(0.722)	(0.954)
Sent_NRC_sc	-2.934^{***}	-1.585^{**}
	(0.785)	(0.790)
Constant	1.732	1.685
	(1.365)	(1.459)
Akaike Inf. Crit.	161.230	161.230
Note:	*p<0.1; **p<0.05; ***p<0.01	

Case study II: Cyber risk prioritisation

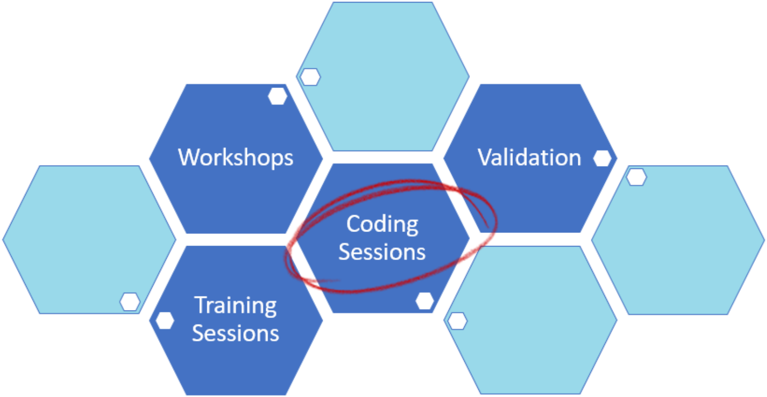
- Cyber risks can be defined as: operational risks emerging from the use of ICT, that compromises the confidentiality, availability, or the integrity of data or services (IMF, 2018).
- Data on cyber risk is scarce: there is no common standard to record them, and companies have no incentives to report them For example, among around 4,000 annual reports for U.S. firms published in 2017, only 7 percent included a reference to cyber-risk.
- There have been very few quantitative analyses of cyber risk. We extend IMF (2018) in two main directions: i) modelling data available only at an ordinal scale; ii) capturing interdependence between event types by means of contagion models, to improve predictive performance.

Preliminary Results - criticality index (Facchinetti et al. (2018))

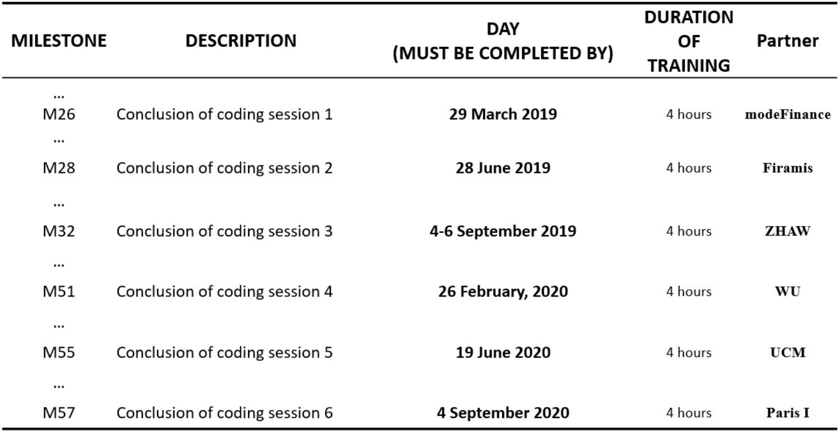
	$\hat{I}$ (SE)			
Attack technique	Cybercrime	Hacktivism	Espion./Sab.	Inf.Warfare
0-day	0.600 (0.126)	1.000 (0.000)	1.000 (0.000)	1.000 (0.000)
Account Cracking	0.188 (0.061)	0.281 (0.088)	1.000 (0.000)	-
DDoS	0.370 (0.078)	0.188 (0.121)	-	1.000 (0.000)
Malware	0.291 (0.024)	0.600 (0.126)	0.971 (0.023)	0.938 (0.058)
Multiple Thr./APT	0.409 (0.082)	0.500 (0.000)	0.952 (0.038)	0.950 (0.047)
Phishing/Soc.Eng.	0.096 (0.035)	-	1.000 (0.000)	0.875 (0.108)
Phone Hacking	-	-	1.000 (0.000)	1.000 (0.000)
SQLi	0.500 (0.000)	0.500 (0.000)	-	-
Unknown	0.162 (0.026)	0.352 (0.081)	0.969 (0.043)	1.000 (0.000)
Vulnerabilities	0.280 (0.051)	0.325 (0.075)	1.000 (0.000)	1.000 (0.000)
Geometric mean	**0.239**	**0.342**	**0.973**	**0.952**

Training by coding (Jochen Papenbrock and Branka Hadji Misheva, WP6)

Overview of Coding Sessions - I



Overview of Coding Sessions - II



MILESTONE	DESCRIPTION	DAY (MUST BE COMPLETED BY)	DURATION OF TRAINING	Partner
...				
M26	Conclusion of coding session 1	**29 March 2019**	4 hours	**modeFinance**
...				
M28	Conclusion of coding session 2	**28 June 2019**	4 hours	**Firamis**
...				
M32	Conclusion of coding session 3	**4-6 September 2019**	4 hours	**ZHAW**
...				
M51	Conclusion of coding session 4	**26 February, 2020**	4 hours	**WU**
...				
M55	Conclusion of coding session 5	**19 June 2020**	4 hours	**UCM**
...				
M57	Conclusion of coding session 6	**4 September 2020**	4 hours	**Paris I**

Table 7: Schedule of the Coding Sessions

Functionalities included in the coding lab

The coding sessions will allow participants to experiment and test the proposed fintech risk management tools by means of an open source reproducible implementation.

- Repository for coding session material (syllabus, scripts, datasets)
- Code interaction with notebooks
- Cloud server environment, located and hosted in Europe

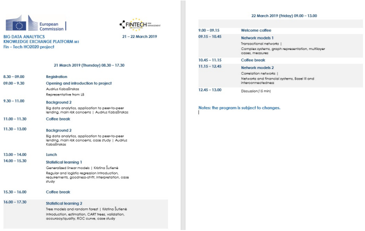
Functionalities included in the dissemination part of the infrastructure

- Upload and exchange of publications, slides, codes, testdata
- Download and feedback functionality
- Publishing process
- External communication channels: web site and social media
- Event participation repository
- Feedback repository
- Evaluation lab
- Communication channels
- Participation repository
- Event managing support

FinTech-HO2020 - Logo

Dissemination: Examples

Agenda of the Training Sessions



European Commission

FINTECH RISK MANAGEMENT

BIG DATA ANALYTICS
KNOWLEDGE EXCHANGE PLATFORM M1
Fin – Tech HO2020 project

21 – 22 March 2019

21 March 2019 (Thursday) 08.30 – 17.30

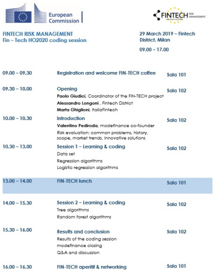
8.30 – 09.00	Registration
09.00 – 9.30	Opening and introduction to project Audrius Kabašinskas Representative from LB
9.30 – 11.00	Background 2 Big data analytics, application to peer-to-peer lending, main risk concerns | Audrius Kabašinskas
11.00 – 11.30	Coffee break
11.30 – 13.00	Background 2 Big data analytics, application to peer-to-peer lending, main risk concerns, case study | Audrius Kabašinskas
13.00 – 14.00	Lunch
14.00 – 15.30	Statistical learning 1 Generalized linear models | Kristina Šutienė Regular and logistic regression introduction, requirements, goodness-of-fit, interpretation, case study
15.30 – 16.00	Coffee break
16.00 – 17.30	Statistical learning 2 Tree models and random forest | Kristina Šutienė Introduction, estimation, CART trees, validation, accuracy/quality, ROC curve, case study

22 March 2019 (Friday) 09.00 – 13.00

9.00 – 09.15	Welcome coffee
09.15 – 10.45	Network models 1 Transactional networks | Complex systems, graph representation, multilayer cases, measures
10.45 – 11.15	Coffee break
11.15 – 12.45	Network models 2 Correlation networks | Networks and financial systems, Basel III and interconnectedness
12.45 – 13.00	Discussion(15 min)

Notes: the program is subject to changes.

Agenda of the Coding Sessions



European Commission

FINTECH RISK MANAGEMENT

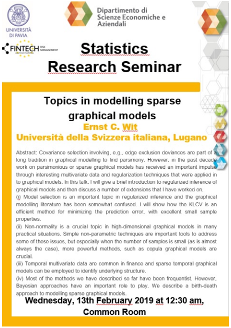
FINTECH RISK MANAGEMENT
Fin – Tech HO2020 coding session

29 March 2019 – Fintech District, Milan

09.00 – 17.00

09.00 – 09.30	Registration and welcome FIN-TECH coffee	Sala 101
09.30 – 10.00	Opening **Paolo Giudici**, Coordinator of the FIN-TECH project **Alessandro Longoni** , Fintech District **Marta Ghiglioni**, ItaliaFintech	Sala 102
10.00 – 10.30	Introduction **Valentino Pediroda**, modefinance co-founder Risk evaluation: common problems, history, scope, market trends, innovative solutions	Sala 102
10.30 – 13.00	Session 1 – Learning & coding Data set Regression algorithms Logistic regression algorithms	Sala 102
13.00 – 14.00	FIN-TECH lunch	Sala 101
14.00 – 15.30	Session 2 – Learning & coding Tree algorithms Random forest algorithms	Sala 102
15.30 – 16.00	Results and conclusion Results of the coding session modefinance closing Q&A and discussion	Sala 102
16.00 – 16.30	FIN-TECH aperitif & networking	Sala 101

Seminars



UNIVERSITÀ DI PAVIA

Dipartimento di Scienze Economiche e Aziendali

FINTECH RISK MANAGEMENT

Statistics Research Seminar

Topics in modelling sparse graphical models

Ernst C. Wit

Università della Svizzera italiana, Lugano

Abstract: Covariance selection involving, e.g., edge exclusion deviances are part of a long tradition in graphical modelling to find parsimony. However, in the past decade work on parsimonious or sparse graphical models has received an important impulse through interesting multivariate data and regularization techniques that were applied in to graphical models. In this talk, I will give a brief introduction to regularized inference of graphical models and then discuss a number of extensions that I have worked on.

(i) Model selection is an important topic in regularized inference and the graphical modelling literature has been somewhat confused. I will show how the KLCV is an efficient method for minimizing the prediction error, with excellent small sample properties.

(ii) Non-normality is a crucial topic in high-dimensional graphical models in many practical situations. Simple non-parametric techniques are important tools to address some of these issues, but especially when the number of samples is small (as is almost always the case), more powerful methods, such as copula graphical models are crucial.

(iii) Temporal multivariate data are common in finance and sparse temporal graphical models can be employed to identify underlying structure.

(iv) Most of the methods we have described so far have been frequentist. However, Bayesian approaches have an important role to play. We describe a birth-death approach to modelling sparse graphical models.

Wednesday, 13th February 2019 at 12:30 am, Common Room

Dissemination: Journals

- Frontiers in Artificial Intelligence

frontiers
in Artificial Intelligence
Artificial Intelligence in Finance

- Digital Finance

Digital Finance
Smart Data Analytics, Investment Innovation, and Financial Technology
Springer

Registration and Feedback Forms

- Please make sure to always use the registration and feedback forms for any Project event
- ▸ Registartion Form ▸ Evaluation Form

Validation of Fintech risk management models (Dave Remue and Arianna Agosto, WP7)

Dissemination and Validation - I

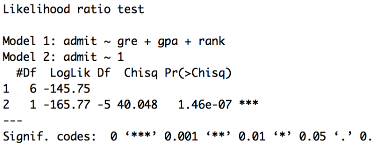
- Both dissemination and validation are enacted and made transparent within the project platform of WP6, which includes a communication infrastructure, a dedicated project website, and a Slack social network channel, aimed at engaging all stakeholders, existing and potential.
- The platform is continuously fed with feedback participation and evaluation to all projects' events: workshops, training sessions, training with coding sessions.

Dissemination and Validation - II

- The feedback from the project network (partners and Regulators) is only the first step of evaluation: a key part of WP7 concerns the validation of the fintech risk management models built in the project by European banks.
- Through the European Bank Federation (EBF) in collaboration with ABI Lab, the developed models and the methods for their validation are proposed to the risk management functions of European Banks.
- In this framework it is crucial that model validation is performed in line with the risk management regulations valid for the European banks, adapted to the fintech context. The next slides introduce some technical guidelines.

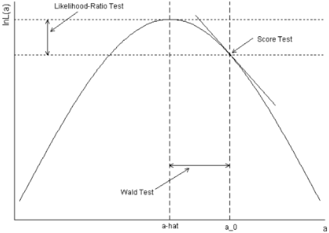
Model evaluation

- The traditional way to choose a model is statistical testing: a sequence of pairwise model comparisons, on the basis of a test statistics whose distribution is known.



```
Likelihood ratio test

Model 1: admit ~ gre + gpa + rank
Model 2: admit ~ 1
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   6 -145.75
2   1 -165.77 -5 40.048   1.46e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.
```



- A complete ordering of models can be obtained using scoring likelihood-based methods such as AIC (Akaike Information Criterion) or BIC (Bayesian Information Criterion):

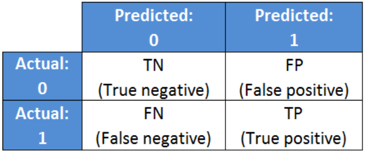
$$AIC = -2logL(\hat{\theta}; x_1......, x_n) + 2q$$
$$BIC = -2logL(\hat{\theta}; x_1......, x_n) + qlog(n)$$

Predictive accuracy criteria

- Neither statistical tests nor scoring methods are generally applicable to machine learning models, which do not necessarily have an underlying probabilistic model.
- A different approach measures predictive accuracy assuming that one part of the data is not observed, but should be predicted.
- The predicted values can be compared with the observed ones, to measure predictive accuracy.
- The problem with this approach is that accuracy measures differ depending on the type of reponse to be predicted.

Predictive accuracy for binary variables: the confusion matrix

- A confusion matrix contains information about actual and predicted classifications.



	Predicted: 0	Predicted: 1
Actual: 0	TN (True negative)	FP (False positive)
Actual: 1	FN (False negative)	TP (True positive)

- Accuracy $= \frac{TP+TN}{TP+FP+TN+FN}$
- Sensitivity $= \frac{TP}{TP+FN}$
- Specificity $= \frac{TN}{TN+FP}$

Predictive accuracy for binary variables: the ROC curve - I

- A widely used predictive accuracy measure for binary responses is the Receiver Operating Characteristics (ROC curve).
- The ROC curve displays the relationship between the sensitivity (on the y-axis) and the complement of the specificity (1- specificity, on the x-axis), across a series of predetermined cut-off points.
- The Area Under the ROC curve (AUROC) is the area under the ROC curve, and summaries predictive accuracy into a single statistics.
- **The ideal curve coincides** with the y-axis between 0 and 1, and the AUROC, in this case, is equal to 1.

Predictive accuracy: the binary variables: the ROC curve - II

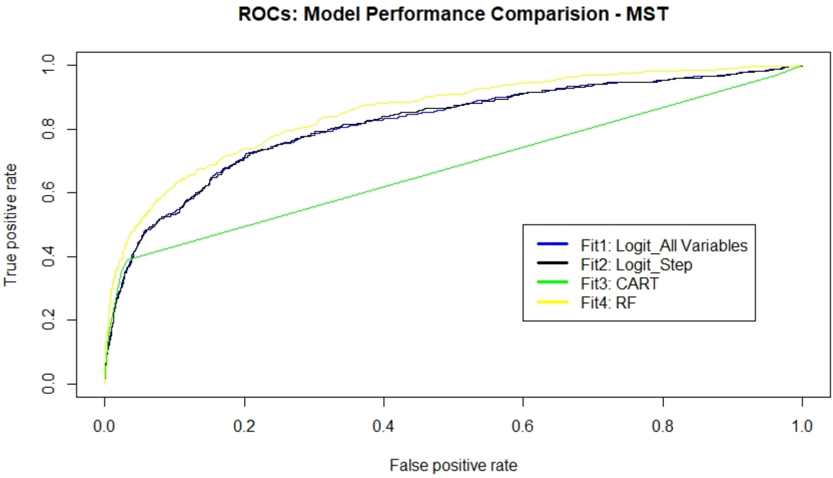
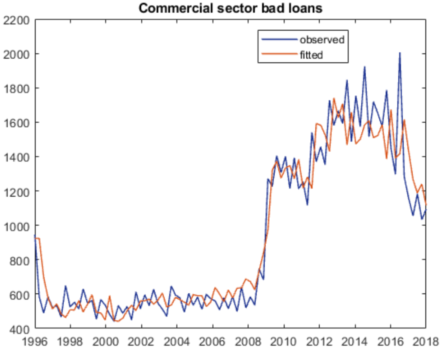


Figure 7: Model performance comparison through the ROC curve

Predictive accuracy: continuous case



- When the response variable is continuous, a widely used measure to assess how far the predicted values are from the observed ones is the Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{x}_i - x_i)^2}{n}}$$

x_i actual values; $\hat{x}_i$ fitted values

Discussion (Claudia Tarantola)

Summary

- The actual financial landscape is characterised by a heavy use of technology. We cannot deal with the current financial framework without considering the crucial rule of FinTech.
- The combination of financial services with modern, innovative technologies leads to the creation of new business models characterised by specific risk profiles.
- Different types of risk were considered: Credit and systemic risk, Market risk and Operational risk.
- Standard methods for risk evaluation are no more sufficient, hence the necessity to develop new models to evaluate and deal with them.
- Idea for model construction and evaluation were discussed
- Issues regarding implementation and dissemination of the proposed methodologies were also discussed.

Discussion

- Both Spelta and Härdle present network models. How are these model constructed? Are these models based on probabilistic or a deterministic approach?
- Aste and Spelta: Credit risk models are concerned with differentiating between vulnerable and safe institutions. The data at hand are usually unbalanced with a small number of zeros (low default rate). It is well known that measures based on ROC curves do not perform well in case of unbalanced data. How do you suggest to proceed?
- Härdle and Pagnottoni: What is the advantage of your model compared to alternative ones already existing in the Literature? What is the computational time? Have you tested your model with other cryptocurrencies data?

- Guegan and Cerchiello: Cyber risk evaluation is a very interesting and actual topic, but data availability is a hard problem. Can you tell us something more about the sources and the collection of the data? Is there a standard way to classify the gravity of a cyber attack? In which way you extended IMF(2018) to take into account cardinality of the data?
- Agosto and Remue: AUROC and RMSE are measures that can be easily implemented but suffer from some drawbacks, for example in case of unbalanced data sets or in presence of outliers. Could these issues represent a problem when dealing with risk management data? Can you tell us something about Bayesian model evaluation?
- Hadji-Misheva and Papenbrock: Are you planning to implement ad hoc packages in R and share them between all R users?