Payment network and bank stress nearcasting

by *Jacopo Di Simone*, Sébastien Kraenzlin, Christoph Meyer, Thomas Nellen,
Alfred Sutter, *Paolo Vanini*, *Ermin Zvizdić*

Central bankers go data driven: applications of AI and ML for policy and prudential supervision

De Nederlandsche Bank
Amsterdam 12-13 May 2022

SCHWEIZERISCHE NATIONALBANK
BANQUE NATIONALE SUISSE
BANCA NAZIONALE SVIZZERA
BANCA NAZIUNALA SVIZRA
SWISS NATIONAL BANK

swissQuant
Mastering Complexity

Introduction

Basic idea: *Stress events* change behaviour of counterparties and/or change behaviour of the stressed bank itself – mirrored in payment data – **network indicators (features)**

Stress event (label): defined by pronounced changes **or** comparatively high levels in default risk on the basis of 5y CDS spread data

Contribution 1: Unsupervised ML (heterogeneous outlier ensemble) applied to capture the dynamics of payment topology of RTGS PS and its participants. Resulting ***outlier score*** is based on a comprehensive set of network indicators

Contribution 2: Weakly supervised ML is used to “nearcast” a ***stress event likelihood*** of a particular bank [based on ***outlier score*** residuals as features and ***stress events*** as label] (stacked ML approach for binary labels) based on interpretable concepts [& data background]

[**Contribution 3:** Estimated **supervised learning model**s should be **transfer**rable to nonCDS banks (using a common CDS bank model or bank subset model based on «similarity»)]

Literature review

- **Bank run nearcasting:** Sabetti and Heijmans (2021), Rainone (2020), Triepels et al. (2018)

→ *We focus on individual banks too but have a broader concept of bank vulnerability in mind that is likely more in line with modern bank runs (as witnessed during the GFC such as eg repo runs)*

→ *We set up a likely more robust and versatile ML pipeline and allow for supervised learning…*

→ *…and that is accessible to interpretation*

- **ML methods used in PS context**: «Timmermans et al. (2017)», Triepels et al. (2018), Heijmans and Zhou (2019), Sabetti and Heijmans (2021), Castro et al. (2021)

→ *We provide a new ML pipeline in the context of PS - stacked learning based on heterogeneous outlier ensemble for binary labels*

- **RTGS monitoring (alert / outliers indicators) in the context of CPMI-IOSCO's PFMI:** Berndsen and Heijmans (2020): Near-real-time monitoring in RTGS systems: A traffic light approach; Heijmans and Wendt (2020): Measuring the Impact of a Failing Participant in Payment Systems

→ *Unsupervised ML provides outlier indicator for the topology of RTGS PS & participants*

Data

*Note: An increase of 12bp in the 5-year CDS spread corresponds to a 1 percentage point increase of the implied default probability (assuming a 40% recovery ratio)

- **Stress event label:** Pronounced changes and comparatively elevated levels in default risk
  - Markit 5y CDS spread data 2005-2020 (from Bloomberg & Reuters) at daily frequency
  - Stress event (for $bank_i$) = 1 if **{**CDS_i > 120**}** AND **{**(daily ΔCDS_i > 12) OR (*period-specific* z-score of CDS_i-vs-$MeanCDS_{j \neq i}$ > 1.96)**}** = TRUE, else = 0*
- **SIC data:** Selection of banks (accounts) with continuous activity & priced default risk 2005-2020 (CDS bank versus nonCDS bank): 18 domestic & foreign CDS banks
  - Participant # ~ 350 – 18 selected banks: # = 54% & CHF = 76%
  - Daily transaction # 1-3 million & CHF 140-180 billion
  - ***Interbank payments:*** # = 4%, CHF = 90%
- **Network indicators (features):** Comprehensive set of 13 (nodal/account) to 15 (overall-system) network indicators: SIC transaction-level data of *interbank* payments from 2005-2020:
  - Aggregated to three intraday periods (overnight, morning and afternoon) and daily frequencies
  - Intraday dynamics important in stress periods (Bech and Garratt, 2012; Benos et al., 2014)

Methodology

Data – Stress event label and payment network indicators

- CDS spread based **binary label**** named *stress event*
- Comprehensive set of **payment network indicators***. *Nodal* centrality indicators are
  - *unweighted*
  - *reciprocally weighted*
  - *iteratively weighted*
- All network indicators are based on transaction level data
  - daily and three intraday periods (overnight, morning and afternoon)

*SIC account level and system level

Data – Feature engineering

- Data = time series data (e.g. trend, seasonality, cyclicality):
  - Ensure stationarity
  - Input features as dev. from normal dynamics
- ARIMA model considering weekday, change-of-month, change-of-year
- Generate input features quarterly, using 3-year moving window → account for long-term shifts in dynamics
- **ARIMA residuals as input features**

*SIC account level and system level

Unsupervised machine learning – outlier detection

- **Heterogeneous outlier ensemble** → output = **outlier scores**:
  - **K-fold cross-validation, observation** & **feature subsampling**
  - Base learner types:
    - **kNN**: anomaly = very distant
    - **GMM**: anomaly = low likelihood region
    - **RAE**: anomaly = large reconstruction error
    - **iFOR**: anomaly = quickly isolated

*SIC account level and system level

Weakly supervised machine learning – stacked learning

- Supervision for selected banks** considering stress events as labels
- Model: **Binary lasso (L1) logistic regression**
  - L1 penalty: implicit feature selection through zero weights
- Output = **stress-event likelihood**
- Interpretability: pipeline set up to measure contribution of input features to stress event likelihood

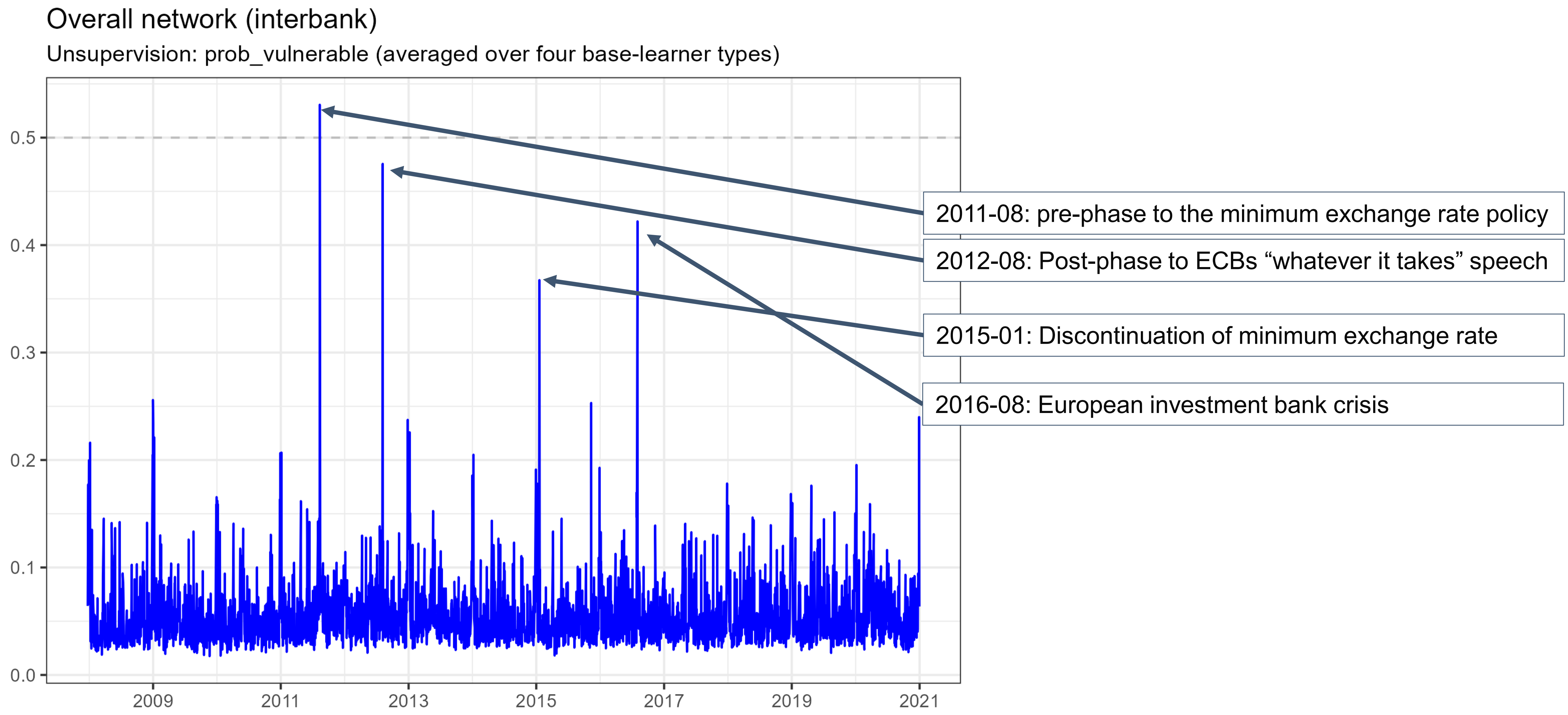
**Economic unit level

Evaluation

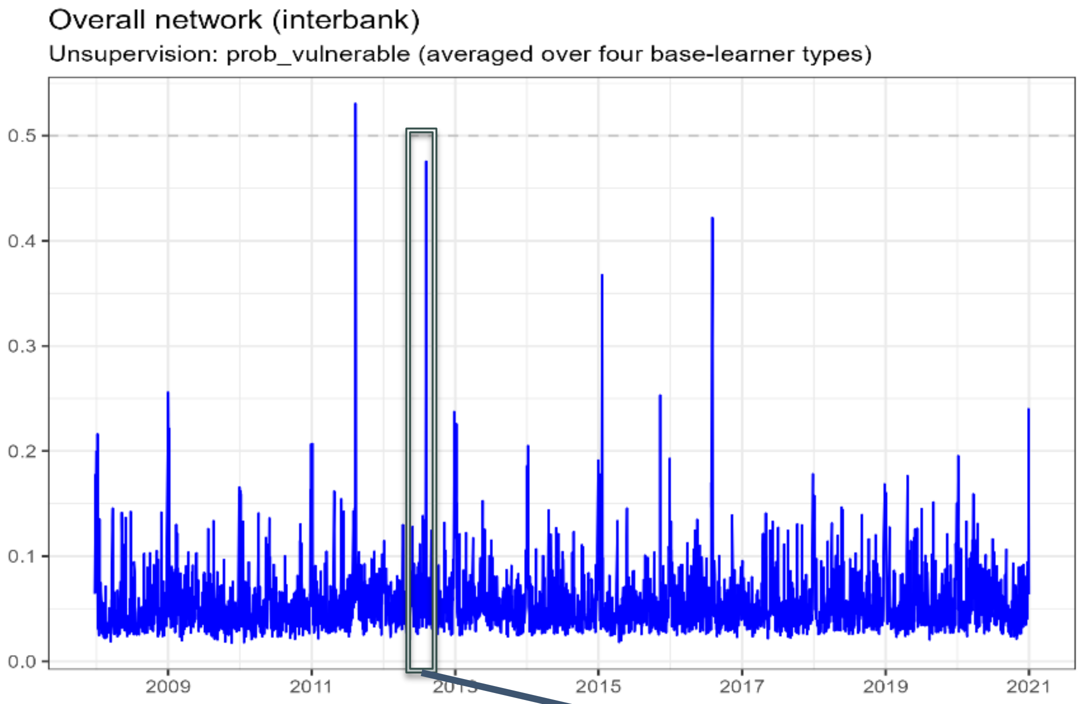
- Metric: **AUC / ROC curve**
  - K-fold cross-validation to i) compute “out-of-sample” outlier scores from each base learner and to
  - ii) estimate “out-of-sample” predictive performance
  - Threshold for binary alerts is selected to maximize the TPR at a fixed FPR
- Metric: **Feature importance** using feature permutation
- Qualitative analysis and visualization

* Possible to aggregate SIC accounts of an economic unit (here not applied); ** Turnover-weighted average based on all SIC accounts of the economic unit

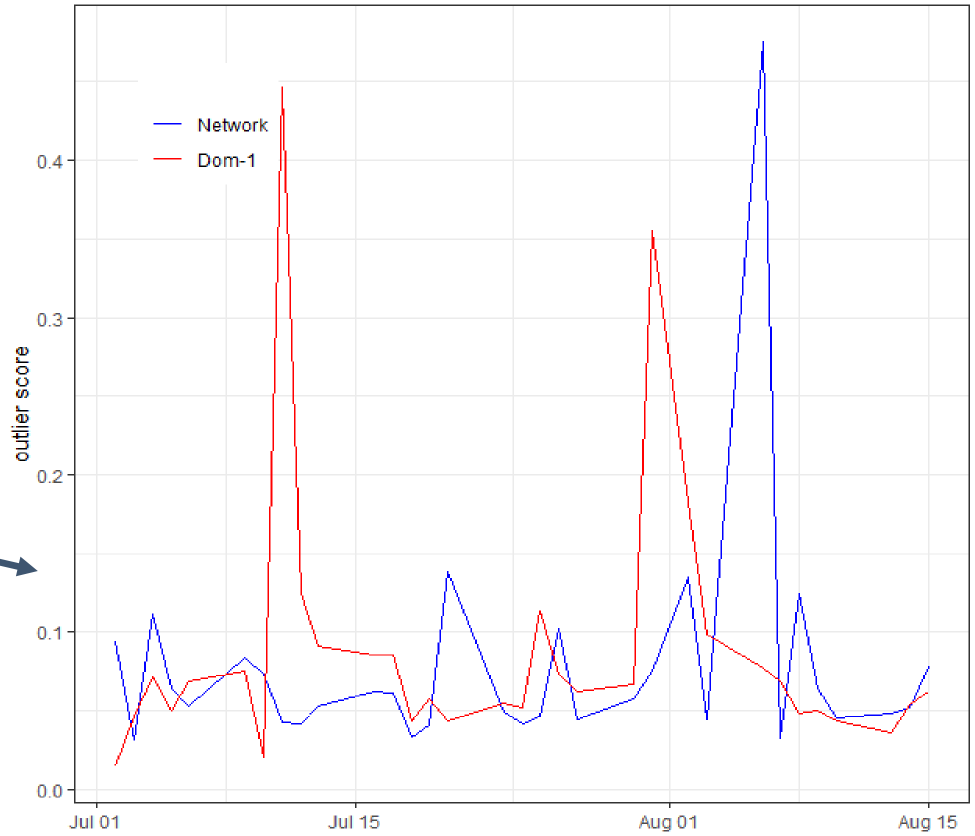
Results – Outlier score at SIC network level…



… differs from outlier score at bank level



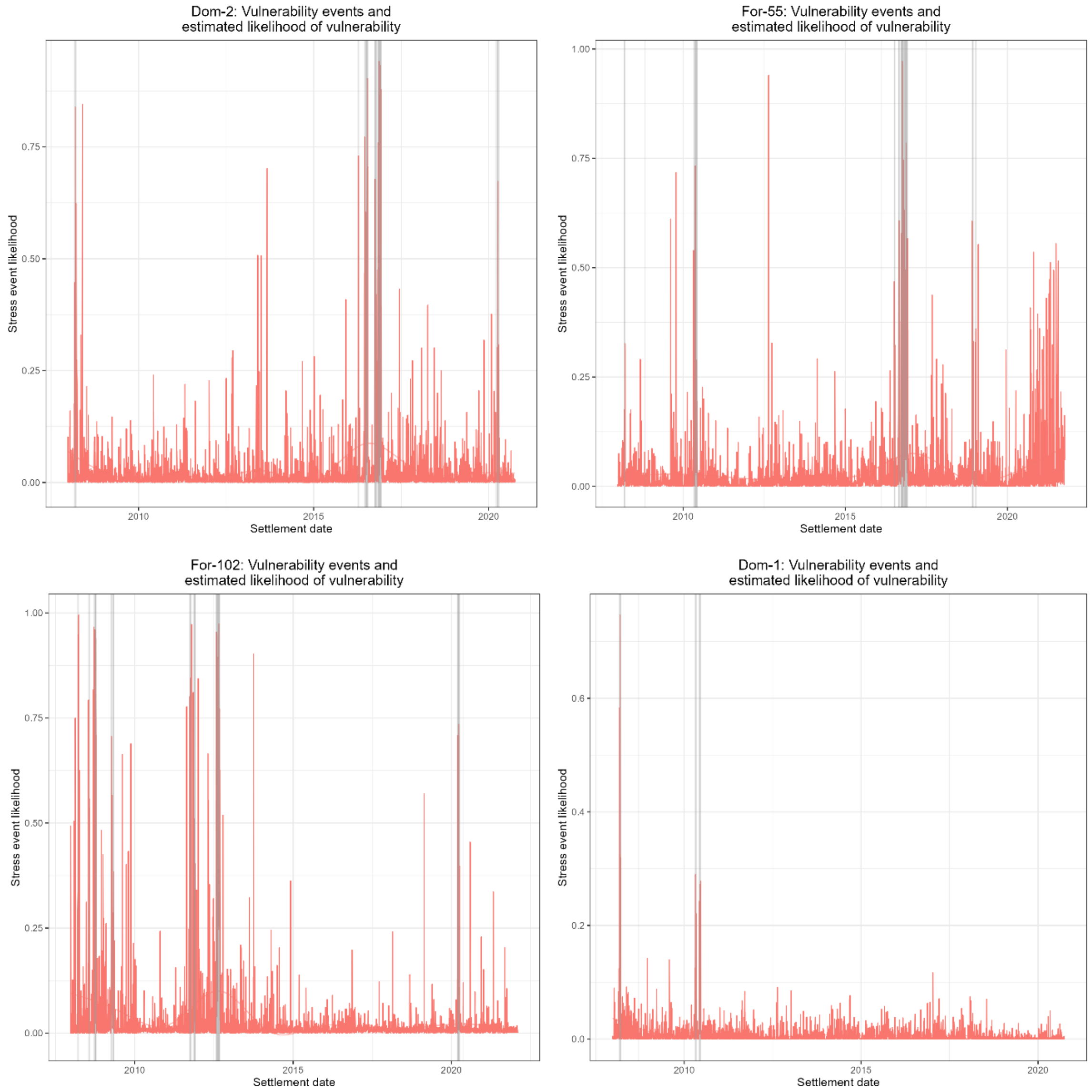
To get a comprehensive picture, it seems advisable to monitor outlier scores at both network *and* bank level.



High individual bank OS seem to coincide with periods of financial stress of the whole system but also of particular banks.

SNB BNS

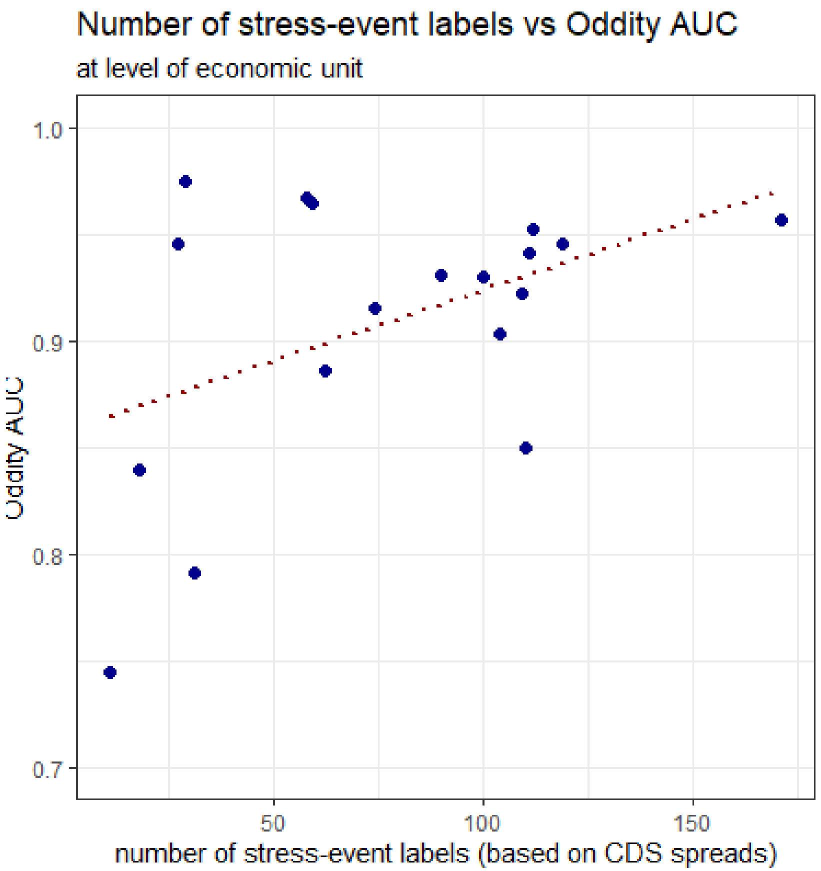
Results – Stress event likelihood for selected banks



- (Generally reasonable occurrence of stress events in line with conventional wisdom)
- All banks suffered from vulnerability during the GFC, but *unevenly*
- Not all banks suffered from the European investment bank crisis 2016 or from the outbreak of the Covid-19 pandemic
- Overall, high stress event likelihoods (red lines) cluster with stress events (grey)
- Perfection would be wrong and imperfection rises the question of performance:

Results – Stress event likelihood – AUC values

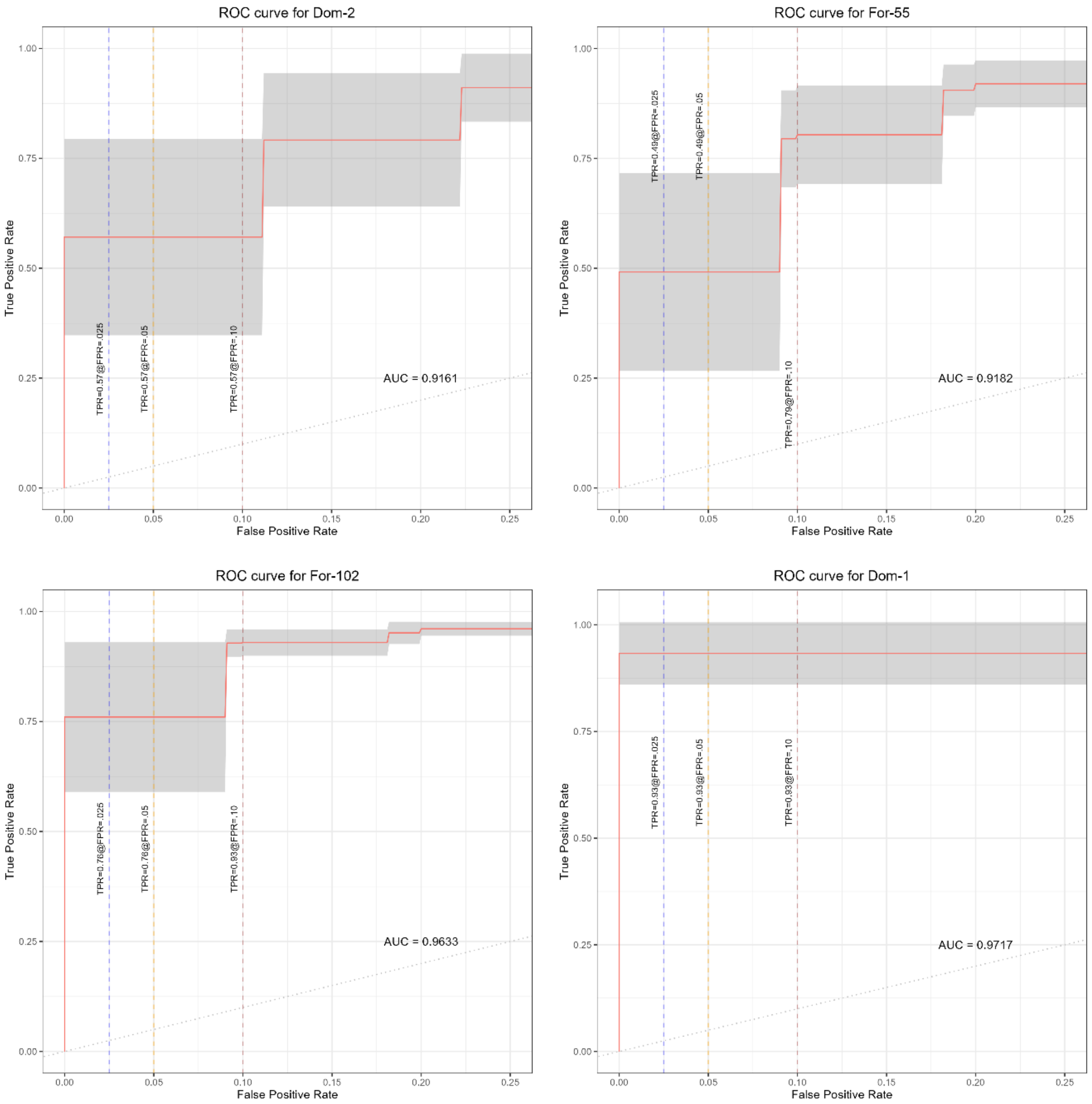
Bank	Stress events	AUC Base	AUC SupSt
Dom-1	29	0.739	**0.972**
For-57	58	0.804	**0.967**
For-111	59	0.517	**0.964**
For-102	112	0.766	**0.963**
For-39	171	0.588	**0.957**
For-163	27	0.448	**0.945**
For-11	119	0.732	**0.945**
For-252	111	0.569	**0.941**
For-264	100	0.781	**0.930**
For-55	109	0.684	**0.918**
Dom-2	90	0.679	**0.916**
For-188	74	0.627	**0.916**
For-131	104	0.486	**0.903**
For-27	62	0.591	**0.886**
For-127	110	0.837	**0.850**
For-68	18	0.387	**0.839**
For-90	31	0.649	**0.791**
For-13	11	0.695	**0.744**



– Accuracy confirmed in terms of AUC (13>AUC 0.9)
– Both domestic and foreign banks show promising results

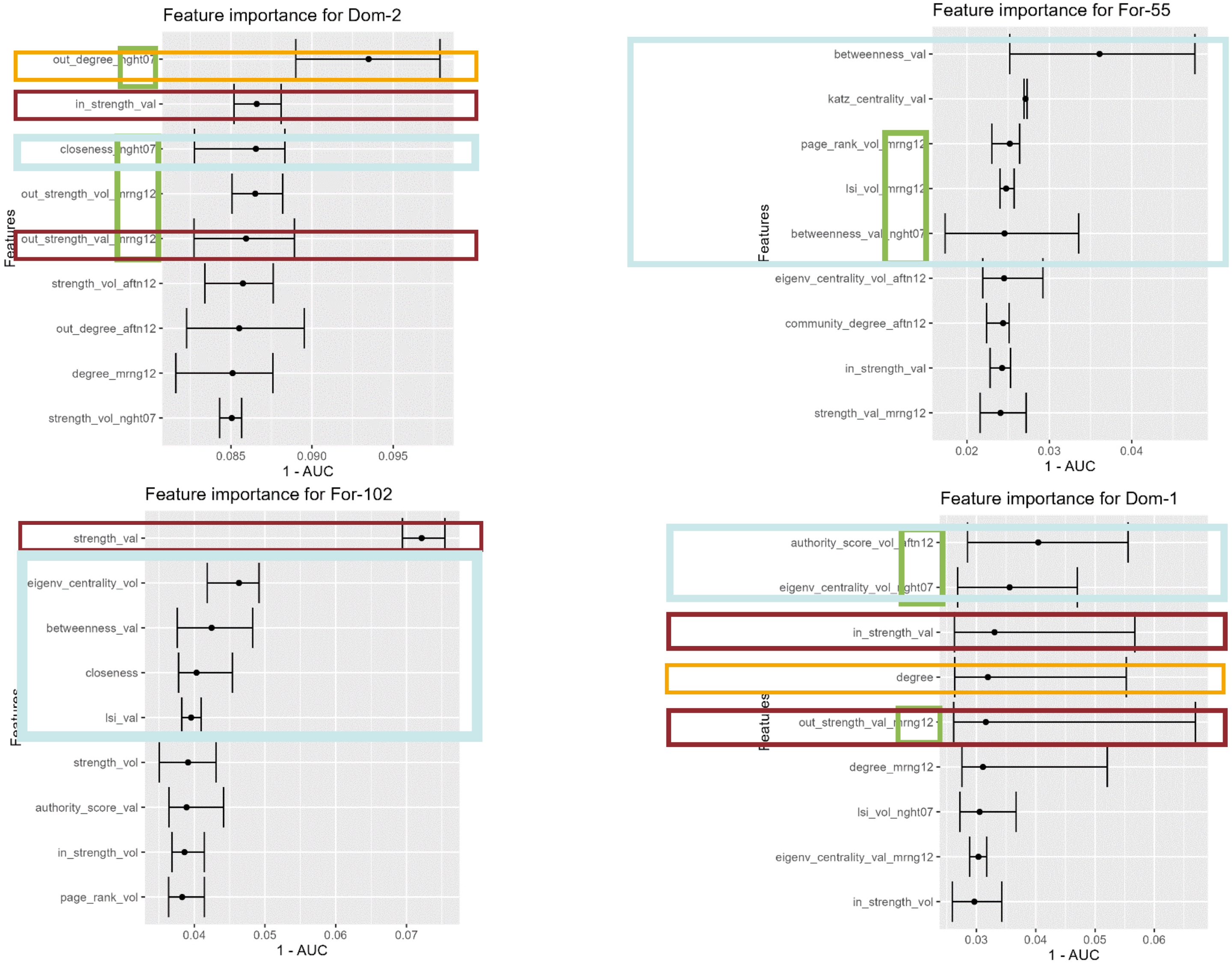
– AUC lowers with a lower # of stress events
– (How many stress event should we have? → How should we define stress events?)

Results – Stress event likelihood – ROC curves



- High steepness of the ROC curve in lower FPR regions → high detection accuracy comes at a low cost of mislabeling non-vulnerable events
- If FPR of 10% acceptable, a generally high detection accuracy is achieved
- For some banks the model performs outstandingly well, e.g. for Dom-1 it identifies 92% of vulnerability events at almost no false alarms

Results – Interpretability – Feature importance (top 10)



- Indirect measure of feature importance: the reduction of the AUC when omitting a feature (Permutation Feature Importance)
- **Per bank** / (per stress event) / [increase feature granularity]
- Each bank shows a different set of important features → banks have an **"individual fingerprint"**
- **"Common feature importance":** Value-based turnover, connectedness and centrality KPIs tend to contribute strongly to accurate nearcasting. Changes in intraday network dynamics contribute to stress detection too.

SNB BNS

Results – Interpretation – Common & individual feature importance

Performance of the benchmark model for selected four banks

Bank	AUC		
	Base	SupSt	Benchmark
Dom-2	0.679	0.916	0.818
For-55	0.684	0.918	0.805
For-102	0.766	0.963	0.809
Dom-1	0.739	0.972	0.801

- **Base:** AUC based on outlier scores
- **SupSt:** Weak supervised learning with single bank
- **Benchmark:** Weak supervised learning with all banks (but the evaluated one)

«Common» model approach as benchmark

- Markup from the Base model (nonsupervised AUC) *«is due to»* common and individual factors
- Roughly 50% may be attributed to common and individual feature importance with Base as the reference
- [Promising and troubling starting point for the third application – model transfer]

[Supervised model transfer – estimate stress event likelihood for nonCDS banks]

- **Idea:** Supervised model transfer to nonCDS banks = estimate stress event likelihood for nonCDS banks (based on supervised regression weights of CDS banks using features of the nonCDS bank)
- **Implementation:** Idiosyncratic footprint suggests using «common» model (benchmark) or a model based on a subset of «similar» banks
- **Evaluation:** non-considered CDS banks (eg Lehman Brothers,…) and nonCDS banks (eg cantonal banks with & without cantonal guarantee, insolvent banks)

Where to go from here?

- **Label – definition of stress event (robustness):**
  - 120bp/12bp → 60bp/24pb / share prices
- **Features:**
  - Use more information: 80+ additional transaction attributes available / move away from the pure network feature approach and use additional features
- **Methodology:**
  - *ARIMA model:* improve ARIMA or use other method to achieve stationarity / 3y training of ARIMA model – reduce loss of data
  - *ML pipeline:* evaluate ensemble (for instance, iFOR is most relevant) / benchmarking to understand drivers and inhibiting factors
  - *Operational usage:* improve interpretability
- **Extensions:**
  - Move from nearcasting to forecasting? CDS spreads or implied default probability as label?
- …

Conclusion

- **Policy relevance:**
  - **Application 1:** RTGS and participant monitoring by means of ***outlier scores*****:**
    - *Comprehensive* monitoring requires both levels, the network and individual participants
  - **Application 2:** Supervisory monitoring by means of the ***stress event likelihood***
    - *Complementing* low-frequency regulatory or manipulated market data
      - Very accurate stress event nearcasting for domestic and foreign banks
    - *Qualifying* CDS spread changes
      - Interpretable approach
  - [**Application 3:** Supervisory monitoring for ***nonCDS banks***
    - *Substituting* missing market data by means of transferring supervised learning models to estimate the stress event likelihood for nonCDS banks]
- **Improvable and adaptable model** (features, label, methodology, interpretation, extensions) **as a first step towards a model that can be put into operation and that covers all three apps**

Thank you for your attention!