Bank of Italy Workshop on *Big Data & Machine Learning Applications for Central Banks* - 21-22 October 2019

Anomaly Detection in RTGS Systems

Performance Comparisons Between Shallow and Deep Neural Networks

L. Arciero, G. Bruno, S. Marchetti, J. Marcucci - Bank of Italy

INTRODUCTION AND MOTIVATION

TARGET2-BdI is Trans-European Automated Real-Time Gross Settlement (RTGS) system, owned by the Eurosystem. Among its objectives, stands the minimization of systemic risk in the financial system (no credit risk).
To the date, resilience of the platform to (liquidity) stress test scenarios was appreciated. Yet, the platform interlinks all participants into a dense network: single liquidity failures may impact the smooth functioning of the system and have severe financial stability repercussions. This motivates monitoring activities with the aim to proactive timely detect anomalies - illiquidity circumstances, bank runs - and to assess build-up risk, and prevent its materialization (also in light of the end of QE).
Traditional approaches for measurement of systemic risk were previously proposed. However, they suffer from data availability issues, undermining their effective real-time employment.

TARGET2-BdI DATA

We focus on payments exchanged in the Italian component of TARGET2 (so called TARGET2-BdI):

- 583 working days with 15-minutes frequency from Jan-2017 to April-2019: $T^{\text{tot}} = 24,192$ observations
- $\mathbf{X}^t = \{x^t_{ij} : i, j = 1, \ldots, N\}$, $N(N-1)$ payments flows between the largest $N = 20$ banks. At time t:

$$\mathbf{X}^t = \begin{bmatrix} 0 & x^t_{1,2} & \cdots & x^t_{1,N} \\ x^t_{2,1} & 0 & \cdots & x^t_{1,N} \\ \vdots & \vdots & \vdots & \vdots \\ x^t_{N,1} & x^t_{N,2} & \cdots & 0 \end{bmatrix}, \quad t = 1, \ldots, T^{\text{tot}}$$

- Cumulative payments flows show high-frequency periodicity: short interval of times are expected to provide several representations of each period's phase.

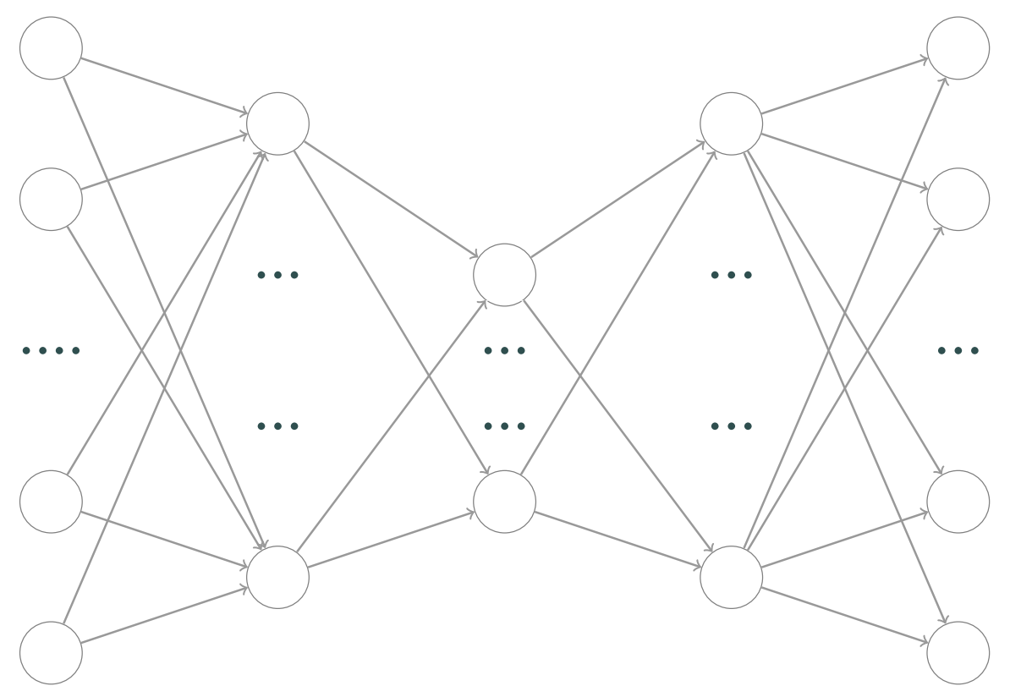
METHODS

Neural Networks (NN) are used to model complex and dynamic relationships in data via successive layers of representation.
Each NN is composed of a stacking of *layers*, usually an Input Layer, H Hidden Layers ($H \geq 1$) and an Output Layer. Each layer is composed in turn by processing units called *neurons*.
The way neurons from a layer aggregate input information defines a NN's architecture.
The most basic type of layer is called *Dense*. Neurons from Dense layer h aggregate input information into a linear combination (with weights $\mathbf{W}_h$ and biases $\mathbf{b}_h$), apply a (possibly non-linear) *activation function* f_h and pass it on, until the Output layer is reached:

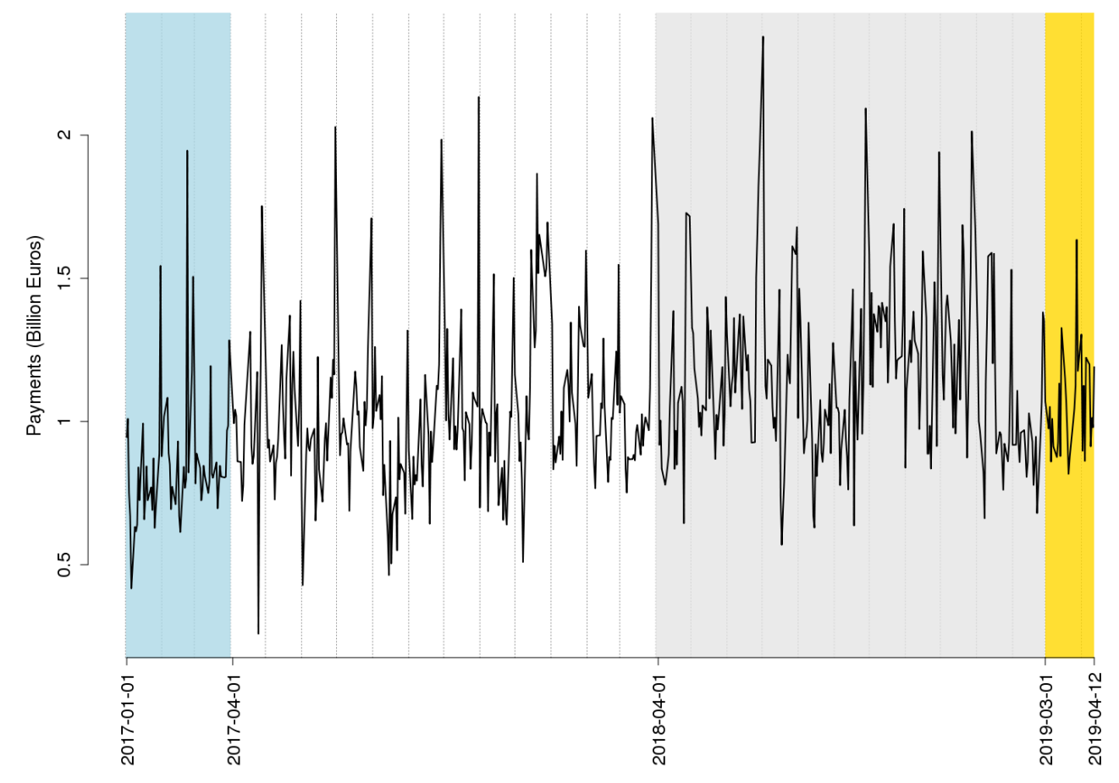
$$\mathbf{Z}^{(h)} = f_l\left(\mathbf{W}_h \mathbf{Z}^{(h-1)} + \mathbf{b}_h\right) \qquad h = 0, \ldots, H$$

Autoencoder

Autoencoders (AE [2]) constitute a class of NN, characterized by a **fully symmetric structure**. In their simplest form, they are equivalent to PCA. In our application, we consider a plain Dense Autoencoder structure with $H = 3$; the innermost layer provides a compressed knowledge representation.
The anomaly detection task with AE is unsupervised: no feedback is provided in training data.
↪ Reconstruction of data points by an *encoding* step to a latent dimension, followed by *decoding*: target variable coincides the input. Our application extends previous work of [3], where a single shallow AE was considered.



Training and Tuning



Tuning: (light blue) Choice of the number of neurons in Hidden layers and activation function ($f_h = f, \forall h$) on a Validation set of 3 months;
Training: (white and gray) 12 to 24 long months window; Scaling of input data was considered according to the matrix representation of data, with respect to each bank's outflows, inflows and overall transactions: three sets of input data were considered
Test: (yellow) Deviance-based anomaly detection with Reconstruction Error (RE):

$$RE(\mathbf{x}^t) = \|\mathbf{x}^t - \hat{\mathbf{x}}^t\|_2^2$$

Anomaly whenever $RE(\mathbf{x}^t) > \mu^{val} + \alpha\sigma^{val}$.

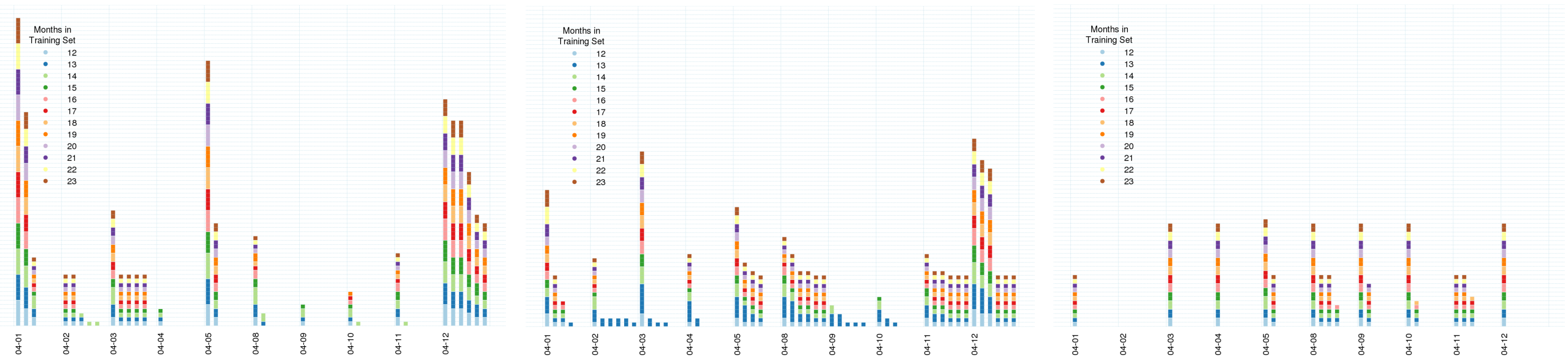
RESULTS

Sensitivity Analysis

- Overall, performance of the models proved robust to different length of the training interval. This finding is supported by analysis of data in the frequency domain. However, shorter windows yield to slight over-reporting of anomalies only if threshold parameter α is small ($\alpha \leq 1$);
- Detection of anomalies is sensitive to changes of α for small values only, otherwise few differences may be appreciated (see results reported below for an example).

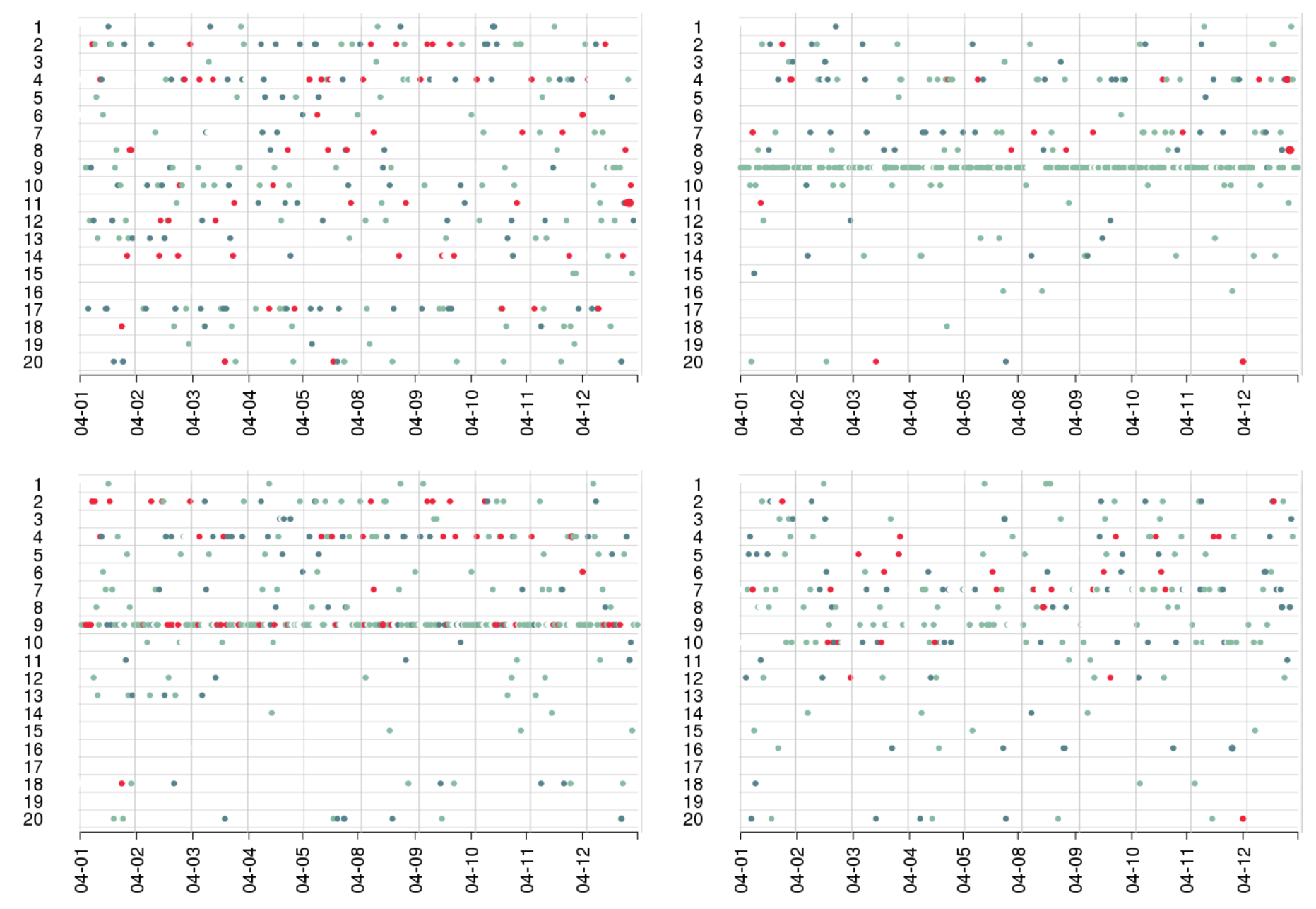
Real World Case-study

We considered two critical weeks from 2019, when a bank suffering a major outage that prevented it from submitting payments to TARGET2-BdI for several hours in a day jeopardized stability of the whole Italian transactions system.
Remarkably, our AEs were able to detect occurrence of anomalous data points. Figure below shows results for different scaling approach and length of the training interval.



Left-to-Right: Daily anomalies detected in the Test Set by the AE trained on Out-, In- and Overall-scaled data in April 2019. Vertical segments depict the number of daily detected anomalies. Length of the training interval is color-coded. Each day can have up to six vertical segments referring to values of threshold parameter $\alpha \in \{0.5, 1.0, 1.5, 2.0, 2.5, 3.0\}$.

Contributions to the general RE by each bank's cumulative outflows and inflows are hereby depicted. As a result, most singularities appeared as related to payments settled by banks on behalf of their customers. Those constitute a frequently executed kind of payments via TARGET2-BdI, for pairs of banks which exchange customer payments quite rarely. The AEs succeeded in detecting singularities caused by the bank suffering.



Simulated Case-Study

In order to be able to evaluate the accuracy of our AEs to detect anomalies, several supervised scenarios were considered, with simulated anomalies. While guaranteeing a balance between those latter and original values, we considered:

1. Gradually increasing deviations in the payments (outflows/inflows/all) of groups of banks over a working week period; anomalies were either *mild* or *strong*
2. Abrupt deviations, on a single day
3. Extreme anomalies on a single bank's payments (outflows/inflows/all), with varying %

We hereby report the F1-Score obtained for the last scenario, with $\alpha = 2$.

	F1 Score		
Extreme anomalies (%)	Sys	Out	In
40%	99.5%	94.8%	96.6%
50%	99.7%	95.8%	97.3%
60%	99.7%	96.7%	98.0%

Overall, the AEs provided us with reliable and timely detection of anomalies. Interestingly, training of the models on different scalings of input data led to an increased ability to recognize perturbations of the expected pattern of payments, while informing us on the nature of such deviations. As an example, consider the scenario with abrupt mild deviations on a single bank's outflows: AE trained on data scaled with respect to each bank's outflows would not detect significant deviations from the expected pattern, while general and inflow-based scalings would yield detection of, respectively, $14.29\%(+20.08\%)$ and $40.48\%(+41.69)\%$ anomalous observations with $\alpha = 2$.

CONCLUSIONS AND FORTHCOMING RESEARCH

- Both idiosyncratic and system-level anomalies were detected in the RTGS system;
- Application of our method to real world case-studies proved effective in timely detecting anomalous payment flows among pairs of banks. Future research will focus on further testing the Autoencoder's ability to timely detect anomalous payment flows and prevent systemic risk, with targeted applications for critical years;
- Empirical applications suggest potential usage of AEs for wholesale payment fraud detection: AEs were able to recognize "unusual or uncharacteristic payment patterns (e.g., in terms of timing, value, volume or location)" as required by the Committee on Payments and Market Infrastructures [1];
- Simulated scenarios were mainly useful in two accounts: 1) Autoencoders were proved to represent a robust method for timely and reliable assessment for anomaly detection on TARGET2-BdI, 2) They provided us with insights on the behavior of different scaling approaches for input data. However, the scenarios would not account for spillover effects: further research may also focus on the propagation of anomalies through the network of payments flows;
- Models were implemented in R, extending code previously available to the *deep* framework. Also, Python-Tensorflow implementation of the Autoencoders achieved significant reduction in the computational time (mainly training/validation phase);
- Finally, forthcoming research will account for additional types of Autoencoders, including those embedding time dependencies.

References

[1] Bank for International Settlement Committee on Payments and Market Infrastructures. Reducing the risk of wholesale payments fraud related to endpoint security. CPMI Report 178, Basel, CH, 2018.

[2] Geoffrey E. Hinton and Ruslan R. Salakhutdinov. Reducing the Dimensionality of Data with Neural Networks. *Science*, 313(5786):504–507, 2006.

[3] Ron Triepels, Hennie Daniels, and Ronald Heijmans. Anomaly detection in Real-Time Gross Settlement Systems. In *ICEIS (1)*, pages 433–441, 2017.