DNB Occasional Studies
Volume 25-3

Federated Modelling

A new framework and an application to system-wide stress testing

DeNederlandscheBank
EUROSYSTEEM

Authors

Sébastien Gallet and Julja Prodani

Views expressed are those of the authors and do not necessarily reflect official positions of De Nederlandsche Bank.

Acknowledgements

We would like to thank Lennart Dekker, Iman van Lelyveld, Robert Vermeulen, David-Jan Jansen and other DNB colleagues for their valuable comments and feedback.

Contents

Abstract

This paper builds on existing literature on federated learning to introduce an innovative framework, which we call federated modelling. Federated modelling enables collaborative modelling by a group of participants while bypassing the need for disclosing participants' underlying private data, which are restricted due to legal or institutional requirements. While the uses of this framework can be numerous, the paper presents a proof of concept for a system-wide, granular financial stress test that enables effective cooperation among central banks without the need to disclose the underlying private data and models of the participating central banks or their reporting entities (banks and insurers). Our findings confirm that by leveraging machine learning techniques and using readily available computational tools, the framework allows participants to contribute to the development of shared models whose results are comparable to those using full granular data centralization. This has profound implications for regulatory cooperation and financial stability monitoring across jurisdictions.

1. Introduction

An interconnected and complex financial system requires cross-jurisdictional and -sectoral collaboration from the part of central banks, supervisory authorities, financial regulators, private sector financial institutions, and other institutions of the international financial architecture. To cater to an interconnected world and financial system, central banks, supervisory authorities, financial regulators, and other institutions of the international financial architecture need to develop and use system-wide models to comprehensively and accurately represent the financial sector and measure financial risks. Privacy and other legal requirements oftentimes inhibit these institutions from sharing not only the data that these requirements are meant to protect, but also models embedding this data. The result is that collaboration across different institutions is often times not possible or inefficient.

Legal constraints on data sharing pose a challenge to international collaboration. Confidentiality is one of the key Fundamental Principles of Official Statistics (UNSD, 2014; BIS, 2023). This means that individual data, whether it refers to natural or legal persons, is to be strictly confidential. In Europe, the General Data Protection Regulation (GDPR) governs the processing of personal data and restricts the sharing of any data that could directly or indirectly identify individuals. Regulation 2018/1725 applies similar rules to EU institutions like the ECB. For central banks, the exchange of statistical data with *external users* is tightly regulated to protect confidentiality and reflect institutional mandates. For example, Regulation 2533/98 stipulates that confidential statistical data can only be shared within the European System of Central Banks (ESCB). This poses a restriction to the data that can be shared outside of the ESCB. In many cases, specific data exchange agreements or memorandums of understanding – such as the Letter of Agreement between the European Central Bank (ECB) and European Insurance and Occupational Pensions Authority (EIOPA) or between European Supervisory Agencies (ESAs) and the European Systemic Risk Board (ESRB) – define the scope and strict conditions under which statistical data can be shared among these entities (EIOPA, 2024; BIS, 2023). When it comes to making data *publicly available*, authorities need to ensure that the information is sufficiently aggregated so as to not be traceable to individual reporting entities.

One of the areas affected by the safeguards around data sharing is that of financial stability assessments and in particular system-wide stress testing. The restrictions imposed when sharing data with *external users* mean that stress tests that rely on granular data can be either sector-specific – for the sector for which the respective regulator or supervisor is responsible (such as the EBA EU-wide stress tests on banks) or country-specific – for different financial sectors in the country where a particular supervisor is based (such as a stress test conducted by de Nederlandsche Bank on Dutch banks and insurers). Given the limitations on data sharing, a truly system-wide stress test – such as a stress test that would rely on granular data from bank and non-bank sectors in the EU or more internationally – has until now not been feasible. Yet, such a stress test is necessary to capture the complexity of today's interconnected financial landscape. Initiatives such as the European Central Bank's Financial Stability Committee Working Group on Stress Testing aim to develop unified frameworks capable of modelling contagion and stress propagation between banks and non-banks at the institutional level (Budnik et al. 2023). This initiative enabled the extension of the ECB's system-wide stress testing framework to include the insurance sector so as to more thoroughly assess risks to financial stability, but it had to rely on country-level aggregated data for insurers (Sydow et al., 2024).

The concept of federated learning (FL) offers a solution to the tension between the need to collaborate across jurisdictions and institutions and the need to preserve the privacy of confidential data and models. FL works by training a "global machine-learning model" across multiple participating institutions, with each participant training the same common model on their own private data and

only sharing model parameter updates with a central server that averages these updates to improve the "global model" (Oualid et al., 2023). This concept has been applied in different disciplines, such as healthcare and finance, but as of yet not to stress testing (Fernandez et al., 2024). We draw inspiration from advances in FL to propose a new framework, which we call federated modelling (FM). The difference between FL and FM is that while the former centralizes the *parameters* of the *same model* trained differently by the different data of each participant, the latter centralizes a *sample of local – participant specific - model outputs* securely aggregated to create learning points for the "global model". In this way, the FM framework preserves the privacy of the local data and models while building on their ability to model participant-specific behaviors. The result is the ability to co-develop publicly available "global models" while bypassing the need for sharing private data and models and therefore preserving compliance with legislative and organizational requirements.

The FM framework solves many of the challenges traditionally posed to system-wide stress testing, given that it i. uses decentralized calibration data that does not need to be shared, ii. uses private models that do not need to be shared, and iii. only needs sub-sets of outputs to be shared that are effectively untraceable back to the participants or their reporting entities. The FM framework relies on participants' decentralised data and private models, allowing cross-sector and cross-country granular data to be used while preserving privacy and avoiding the high costs usually associated with bottom-up stress testing (ECB, 2017; BoE, 2024). In addition, the FM framework is by design able to reflect a multitude of cross-sectors and borders transmission channels that are missing in more traditional stress tests that focus on specific sectors and/or jurisdictions (Sydow et al. 2024). Lastly, the iterative simulations foreseen by the FM framework endogenize variables that are traditionally only introduced as exogenous shocks, meaning that it captures feedback loops that are often excluded in more traditional approaches (EBA, 2025 EU-wide stress test).

By bridging cutting-edge machine learning techniques with institutional constraints, this paper aims to establish federated modeling as a practical, scalable solution for modern financial stability assessments. Our intended audience includes central banks, supervisory authorities, financial regulators, and other international institutions like the IMF and BIS.

The paper first presents the theory behind this framework (section 2), the application to a system-wide liquidity stress test (section 3), the results of the stress test application (section 4), and lastly a discussion of the relevance of FM for the central banking community in terms of financial stability monitoring and policy setting (section 5).

2. The Federated Modelling framework: the theory

FM is a four-step framework that enables collaborative modelling while preserving the privacy of participants' data and models (Figure 1). Below we present these components theoretically. For a deep-dive into a use-case for central banks, please refer to the system-wide stress test application in section 3.

A prerequisite for FM is model decomposition, i.e. the existence of two types of models: the private models of the participants and a public system-level model. We call the private models of the participants "participant-level models". These models are private and local in that they are developed by, visible, and used only by the specific participants that built them. These models are behavioral models that are different for each participant; they describe a participant's responses to system-wide input variables that are the same for each participant. Such input variables can be prices, interest rate changes, etc. The only requirement imposed on the participant-level models is therefore that all these models use the same input (system-wide variables) and produce the same output types (such as quantities or flows). In this way, the outputs can be aggregated or averaged across all participants. In contrast to participant-level models, a system-level model is one publicly available global model that describes the market. The system-level model uses the aggregation of outputs from participant-level models and estimates the market response.

The FM framework consists of four steps (Figure 1).

1. Participants feed the system-wide input variables (e.g. interest rate changes) into participant-level models, which use a participant's private data (e.g. detailed balance sheet information) to generate outputs that are participant-specific (e.g. loss in asset value of its securities). This results in a participant-level datapoint cloud of outputs[1].

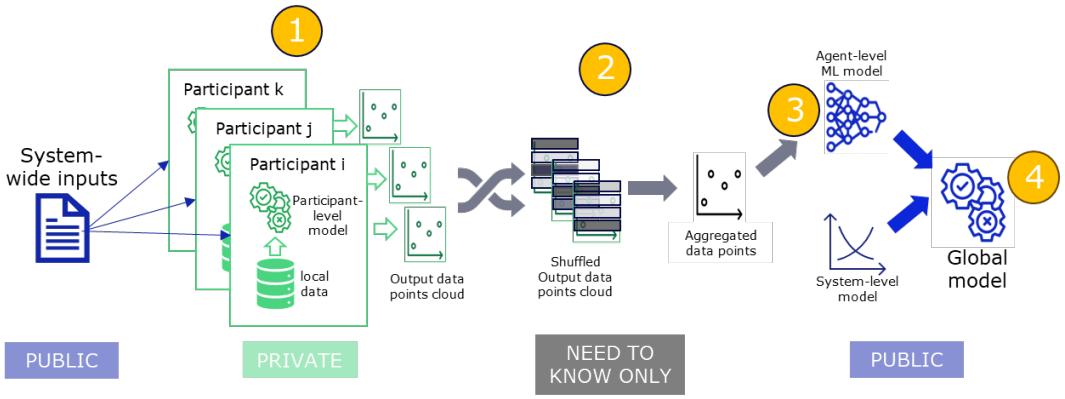
2. Participants submit the output datapoint clouds to the host using a security protocol that ensures individual privacy. This protocol ensures that only the processed data – the sum or average – that is not directly traceable to a participant is visible to the host and will serve as training data for the system-level model. The host here could be either one of the participants or an external counterparty organizing the exercise (e.g. one of the participating central banks, international institutions, financial institutions, etc).

3. Using machine learning, the host creates an approximate function from the combined system-wide input scenarios used in step 1 and the processed output datapoint clouds received in step 2. This function describes how system-wide inputs affect the behaviour of the market participants, i.e. the behaviour of all participants together. We call this the participant-level machine learning (ML) model, which can be thought of as an approximation of all participant-level models used in step 1. As an example, the approximate function could show the impact of an interest rate shock on participants' bond holdings and more specifically how much of their holdings participants would try to sell off.

4. The ML model is then used together with the system-level model. The system-level model models the reaction of the system, e.g. the whole financial market, to the behaviour of the participants. As an example, it would estimate the impact that the collective sell-off behaviour of the participants in step 3 would have on government bond prices. Equilibrium is found through iteration[2], meaning that second and higher round effects are a design feature of the framework. In this way, the behavior of participants impacts the market, and then the new market equilibrium feeds again as a

[1] The system-wide input variables can be in a public datapoint cloud.
[2] Convergence to an equilibrium depends on the model's properties and not on the federated modelling process itself.

system-wide input scenario that in turn affects the participants' behavior and so on until an equilibrium is reached. The combination of the participant-level ML model and the system-level model makes up the Global model.



Note: The numbering in this figure corresponds to the steps of the FM framework as outlined above.

Figure 1: The federated modeling framework

One of the steps of the FM framework consists in i. each participant *sharing its own output* with another participant and then ii. *aggregating their outputs* so that they are not traceable to the participants and their reporting entities (step 2 in section 2). The *aggregated output data* is by definition not confidential and traceable and does not therefore need extra security protection. The *participant-specific output data* shared between participants is also deemed to not be traceable to the participants and their reporting entities. However, as it cannot be proven that this is always the case, we propose a data-protection technique that consists in sharing only parts (i.e. slices) of output, meaning that only processed, i.e. shuffled, output data that cannot be traced back to reporting entities is shared between participants. In this section, we analyze the types of security threats that could face this data sharing and explain our proposed data protection technique.

We identify four types of security threats posed to the sharing of each participant's output with another participant, as envisaged in step 2 of section 2. Drawing on literature, we identify the below four types of security threats (Liu et al., 2022).

- *Direct raw data theft*: This category consists of data theft via e-mails or on private disks.
- *Reverse engineering*: This category considers honest-but-curious participants, which follow the protocol correctly but due to curiosity want to learn more than they should from the data they have access to. To that end, they may try to use the participant-specific output data exchanged during the process to reconstruct the participant-specific private calibration data of one or all other participants. This participant is still aiming to develop the most accurate model possible and is not actively trying to disrupt the framework (Lu et al., 2022).
- *Poisoning attacks*: These attacks consist of a participant injecting fake data or altering the model in order to compromise the framework, i.e. to get wrong outputs for all participants (Tolpegin et al., 2020).
- *Byzantine attacks*: This is a variation of poisoning attacks, whereby a malicious participant injects false data in order to get protected information.

This is different from reverse engineering in that the participant is attempting to extract protected information even if it results in an unusable global model.

Of the four categories of security threats identified above, we focus on how to tackle the threat of *reverse engineering* by an honest-but-curious participant. While the *direct raw data theft* is the most obvious risk, it is out of scope of this analysis given that it is traditionally fully managed by IT and operational risk management given that the treatment would be similar to an external threat. The *poisoning attacks* and *byzantine attacks* are unlikely to happen in a central banking context, given that participation in this exercise would be voluntary and that participants would be duly identified and would represent the credibility of the participating NCA itself. Therefore, the *reverse engineering* by an honest-but-curious participant represents the worst-case scenario considered in this paper.

The Secure Multi-Party Computation (SMPC) data-protection technique can address the risk of *reverse engineering* by an honest-but-curious participant. To counter the risk of reverse engineering by an honest-but-curious participant, multiple techniques can be used. Differential Privacy protects individual data by adding carefully calibrated noise, while still preserving the overall statistical properties of the dataset (Dinur & Nissim, 2003). Homomorphic encryption allows computations on encrypted data without decryption, revealing the correct result only after final decryption (Paillier, 2005) . The Secure Multi-Party Computation (Merino & Cabrero-Holgueras, 2023) technique is a protocol that allows multiple parties – in our case participants - to jointly aggregate data – in our case the output datapoint cloud - while keeping individual data private. SMPC is the technique we propose to protect the data aggregation foreseen in step 2 of the FM framework given that it offers strong protection, is simple to implement, and it does not add noise to data. While we consider the security resulting from SMPC sufficient for a real central banking exercise in an honest-but-curious setting, it always remains an option to combine different protection methods to enhance security when deemed necessary.

The concept of the SMPC is first illustrated through an intuitive example in Box 1 and then applied to a system-wide stress test in section 3.2. While the data aggregation undertaken in the illustrative example is a simple sum of the participants' output data, it is possible to compute other functions like averages, weighted averages, products or weighted products. Given that SMPC protects private participant-level information without allowing the data to be traced back to participants for data quality checks, adding automatic plausibility checks to SMPC would enhance data quality for sensitive uses like stress testing. Ready-to-use open-

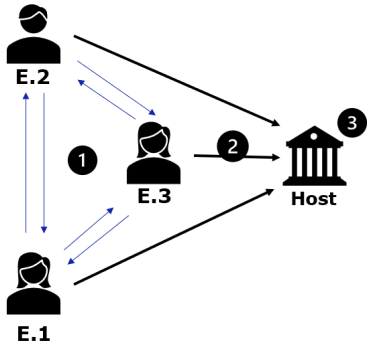
source code is freely available online and can be directly applied in a pilot exercise without additional development (i.e. package PySyft in Python).[3]

Box 1:Illustration of the SMPC technique[4]

Let say that three employees (E.1 , E.2 and E.3) – participants – want to calculate their average salary without revealing their own salaries. The SMPC protocol would then be:

1. Each employee splits their salary into three random parts. Then, it randomly sends two of those three parts to one other participant, keeping one part for themselves.

2. Every participant sums the received parts along with their own remaining portion and forwards the total to a host (which could be a participant).



3. The host adds up the contributions received by all participants and calculates the average salary by dividing the total contributions by the number of participants. The added randomness cancels out in the sum, leaving the final result accurate and keeping the individual salaries hidden. The total sum is public, free of noise, and does not reveal any single participant's input.

The below section applies the FM framework, including SMPC, to a system-wide fire-sale stress test.

[3] Four production-ready SMPC open-source packages are MP-SPDZ, FRESCO, PySyft, and TF Encrypted. Each of these packages offers secure computation capabilities tailored to use cases ranging from cryptographic protocols to privacy-preserving machine learning.
[4] For this example and more generally, to prevent potential inference of the contribution of individual participants, thought should be given to the number and size of participants.

3. The Federated Modelling framework: application to a system-wide stress-test

In this section we apply the Federated Modelling framework to one of the main challenges facing central banks today: building truly system-wide stress tests. A system-wide stress test using federated modelling could assess how the financial system - including different types of financial institutions and their interlinkages - would perform under severe scenarios, without requiring cross-border or cross- jurisdictional confidential data sharing from the part of (inter)national authorities or extra reporting from supervised institutions.

Our specific case study is a fire-sale stress test based on a model proposed by the Federal Reserve (Cetorelli et al., 2023). This model was used by the IMF during the Financial Sector Assessment Program (FSAP) for the Netherlands in 2024 (IMF, 2024).

3.1 The original stress test model

The model proposed by Cetorelli et al. (2023) estimates how financial institutions adjust their portfolios by buying or selling assets in response to price shifts. The model consists of two key components. The first component is a participant-level model, which in this case is a Sell-off Model. It determines i. which asset categories are sold and ii. how much of these categories is sold by each financial entity in order to maintain the entities' initial leverage ratio following an initial system-wide price drop. The second component is a system-level model, which in this case is a Market Model. It adjusts market prices based on overall volumes sold by the participants and market depths for each asset category.

By iterating the results of both components of this model – the participant model component and the system model component – the model assesses shock amplification within the financial system. As the system model does not use sensitive data, but rather only price elasticities of the market – i.e. the relation between the price shock and selloff volumes for different categories of assets – , we categorize it as public and freely shared by all participants. The rest of this section focusses on the participant-level model.

To simplify the participant-level model for the purposes of this proof of concept case while preserving its core logic, three main assumptions are made. The assumptions can always be adjusted and refined later on as needed, without changing the core framework. The three assumptions are: i. a drop in asset prices leads to losses that directly reduce balance sheet equity, while debt remains constant; ii. the leverage ratio is restored to its initial level in each iteration by selling assets as needed; and iii. assets are sold in order to keep the initial portfolio allocation.

The model requires assets to be grouped into pre-defined categories[5]. Once chosen, the asset categories are the same for all participants. For this proof of concept, we used Monthly Security Reporting (MSR). MSR is a national dataset that provides detailed denomination on securities held by banks, insurers, and pension funds[6]. We define 15 asset categories, which we consider as a good balance between reflecting participants' portfolios while keeping the computation easily manageable[7].

[5] While more categories enhance realism, they also increase computational load. Therefore, a balance must be found.
[6] MSR provides ESA2010 denomination.
[7] Based on ESA 2010 categorization, the list of asset categories is: Cash; F_31_O_EUR_S_13; F_33_K_EUR_S_12; F_33_K_EUR_S_unknown; F_33_K_USD_S_12; F_33_O_EUR_S_13; F_33_O_USD_S_13; F_511_C_USD_S_11; F_511_J_USD_S_11; F_511_K_USD_S_12; F_521_K_EUR_S_12; F_522_K_EUR_S_12; F_522_unknown_EUR_S_12; F_522_unknown_USD_S_12; others

Data on leverage ratios comes from COREP for banks and Solvency II reporting for insurers.

3.2 Applying the federated modeling framework to the original stress test model

In this section we conduct the fire-sale stress test proposed by Cetorelli et al. 2023 using the FM framework. For this stress test, we consider three participants, namely a supervisor of banks, a supervisor of insurers, and an external host. The FM framework is most useful for collaboration across participants that are established in different jurisdictions – e.g. EU central banks – and/or across participants that are responsible for different financial sectors – e.g. the ECB and EIOPA. In order to test the framework, however, we need data that is already available to DNB. To that end, we have used data that DNB has in its capacity as a supervisor of both banks and insurers. The steps of the framework would remain unchanged for different types of participants.

Below are the four steps of the FM framework outlined in Section 2, applied to the system-wide fire-sale stress test. Steps 1 to 3 are visualized in Figure 2, while step 4 is visualized in Figure 3.

Protocol to run the Federated Modeling-System Wide Stress Test

Initialisation

The host shares with the participants the code for the participant-level "sell off" model. It must be plugged to local data and can be modified partially or fully by the participants.

The host shares with the participants the code for the system-level "market" model. This makes the framework transparent, as it allows the participants to re-run both models and reproduce the final results at the end of the protocol.

The host provides participants with common input datapoint clouds (system-wide inputs) of variations in prices, covering the space of plausible price variations.

1 – Local calculation

Participants generate output datapoint clouds -sell-off volumes- based on the sell-off model, calibrated on each participant's private data.

2 - Aggregation

Participants randomly split each output data point cloud into two shares and send one to the other participant. Each participant adds the kept to the received output datapoint cloud. Each participant sends its resulting output datapoint cloud to the host. The host aggregates them and creates the global input-output datapoint clouds, i.e. a table mapping the system-level price variations to the aggregated participant sell-off behaviour.

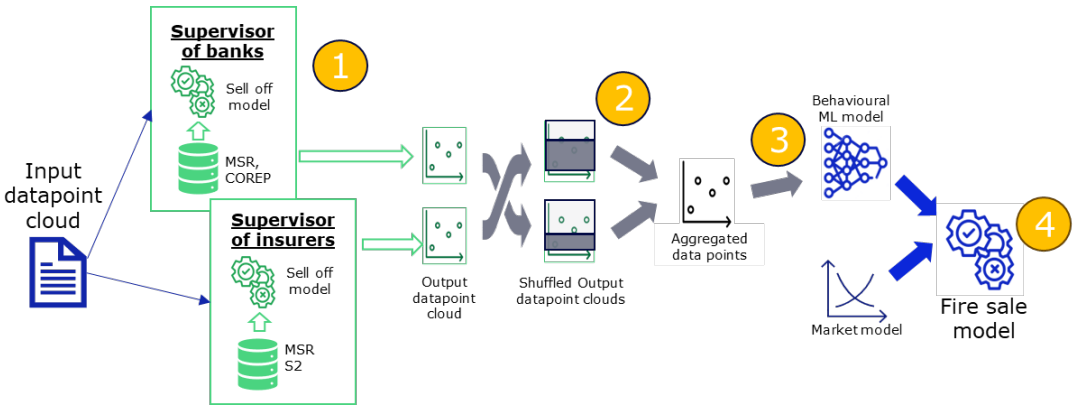
3 – ML approximation

The host creates a proxy-model by using machine learning trained on part of the global input-output datapoint cloud and tested on the remaining part of the global input-output datapoint cloud.

4 – Global modelling

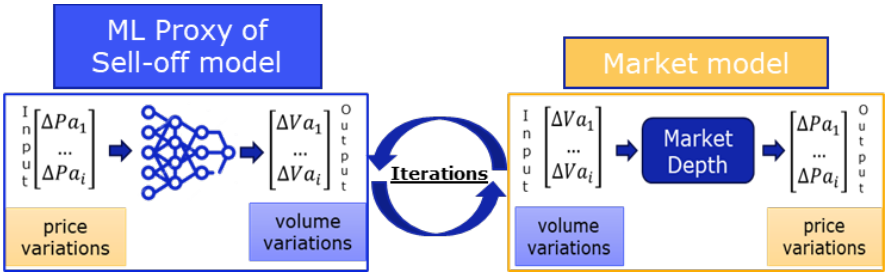
The host shares the ML model with the participants. Each participant, including the host, can run both the ML model (proxy of the participant-level sell-off model) and the Market model (the system-wide model) simultaneously to proxy the endogenous amplification of initial exogenous price shocks by all the participants collectively (Figure 3).

Figure 2: Fire-sale model proxied by the FM framework



Note: The numbering in this figure corresponds to the steps of the FM framework as outlined above.

Figure 3: Endogenous amplification of initial price shocks through multiple iterations of both models



The next section addresses the question of the quality and feasibility of training the ML model. The main challenge of the FM framework is to determine to what extent and at what level of complexity it is possible to train a generic ML model that can not only replicate known input-output points but also accurately interpolate new, unseen inputs with a low error compared to the output produced by the sum of all private individual models. This capability will be tested and demonstrated through the concrete example in the next section.

4. Results of the stress test application

This section focuses on the quality and feasibility of training a machine learning model to replicate the combined output of all participant-level models, i.e. step 3 defined in Section 3.2. For the stress test model defined in section 3, the inputs are vectors of price changes for the 14 asset categories. We generate 5,000 randomly distributed vectors within a range of -1% to -90% to ensure coverage and materiality. Each participant runs its private sell-off model based on these inputs and calibrated on private data, producing 5,000 outputs. The output vectors — showing sell-off amounts across 14 asset categories — are aggregated, completing the SMPC step. A machine learning model is then trained on one part of the aggregated data and is tested for accuracy on the other part of the aggregated data.

We first assess how well the ML model replicates the private selloff models (section 4.1) and then how, once used in combination with the market model to proxy the original fire-sale model, it proxies the price evolution produced by the original fire-sale model (section 4.2).

4.1 Quality of the ML model

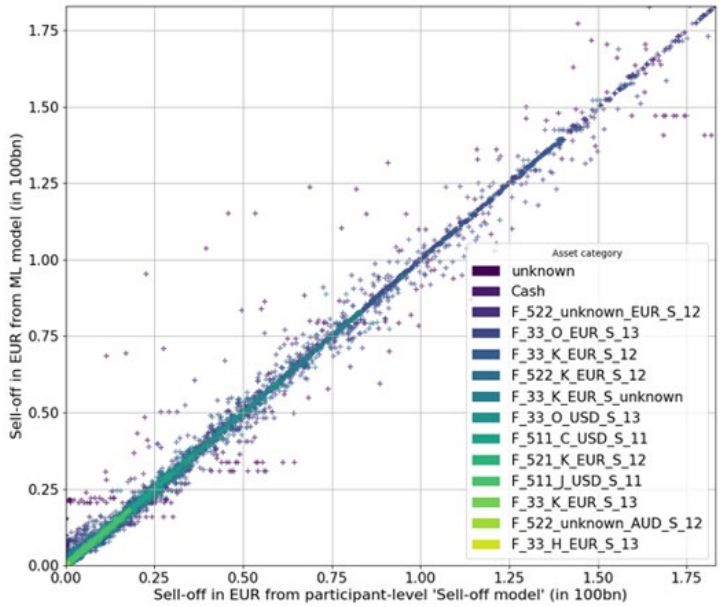
While dozens of ML model types are available, for this proof of concept we use the XGBoost model. There are dozens of Machine Learning (ML) models available, ranging from simple regression models to advanced models like neural networks and deep learning. The choice of machine learning model is important. It must be well suited to the use case—complex enough to capture nonlinear behavior in multidimensional spaces, but with as few degrees of freedom[8] as possible to avoid requiring an excessive number of training points. For this proof of concept, we choose the XGBoost model given that it is well-suited for high-dimensional data - here 14 asset categories for input and output - and balances strong performance with moderate training and computational requirements[9].

To assess the quality of the ML model, we compare its performance to that of the private sell-off models of the participants. As per standard practice when using data to train and then assess a model, the set of input-output learning points is first randomly split into two sets. Most of it is used as training data (80% of total datapoints, consisting of 4000 training points), and the rest is used to evaluate the performance of the model (20% of total datapoints, consisting of 1000 test points).

The performance of the model is evaluated by comparing the outputs, i.e. the sell-off amounts, that the trained ML model produces to the outputs produced by the original private sell-off models of each participant. Any deviation between the two outputs is considered an error. Errors are visualized in Figure 4 as deviations from the diagonal. In case the ML model perfectly replicates the private selloff models, all the testing points as represented by dots in Figure 4 should be on the diagonal, reflecting the fact that sell-off amounts produced by the private models (x-axis) are equal to the sell-off amounts produced by the ML model (y-axis).

Figure 4: The quality of the ML model in replicating the private sell-off models of the participants

[8] A degree of freedom refers to a free parameter that can be adjusted during calibration to improve the model's fit to data. The more degrees of freedom a model has, the more flexible it becomes. This comes at the cost of more complex and time-consuming calibration.
[9] From the computational point of view, we used 'Scikit-Learn' python package for XGBoost as ML model. Hyperparameters of the model were optimized with 'Optuna'. Several other open source configurations exist that have not been tested.



The performance of the ML model is good, suggesting that it could be a good substitute to the private sell-off models. The ML model shows good accuracy in predicting 14 selloff amounts based on 14 price variations, with most test points resulting in relatively small errors. The average error weighed across asset categories is -0.021%. The standard deviation of the error weighed across asset categories is 0.89%, meaning that in 68.27% of cases the error will be smaller than 1%. While this performance was achieved using only 4,000 training points, we found that increasing the number of points to 8,000 did not significantly improve replication quality. This suggests that adding more data is not the most effective approach. To enhance accuracy and flexibility, alternative machine learning models should be considered. More adaptable methods - such as neural networks - could offer better performance, albeit at the price of much larger datasets and greater computational resources.

The next section shows how the ML model performs when run jointly with the market model to produce a system-wide fire-sale stress test.

4.2 Quality of the FM framework- jointly run ML model and market model

To understand how accurate the use of the ML model is in combination with the market model to mimic the original fire-sale model, we compare the outputs of these two approaches after several iterations. In the first round, we apply a random vector of price shocks to both approaches, i.e. to the original fire-sale model using centralized data from banks and insurers (as defined in section 3.1) and to the trained ML model (XGBoost, as explained in section 4.1). The output vector of asset selloffs produced by both approaches is separately fed into the same market model. The market model then updates asset prices, and the process repeats over multiple rounds[10]. The evolution of the asset prices produced by both approaches is shown in Figure 5.

[10] This is the same process as that explained in section 3.2.

a) Original fire-sale model **b) FM fire-sale model using ML**

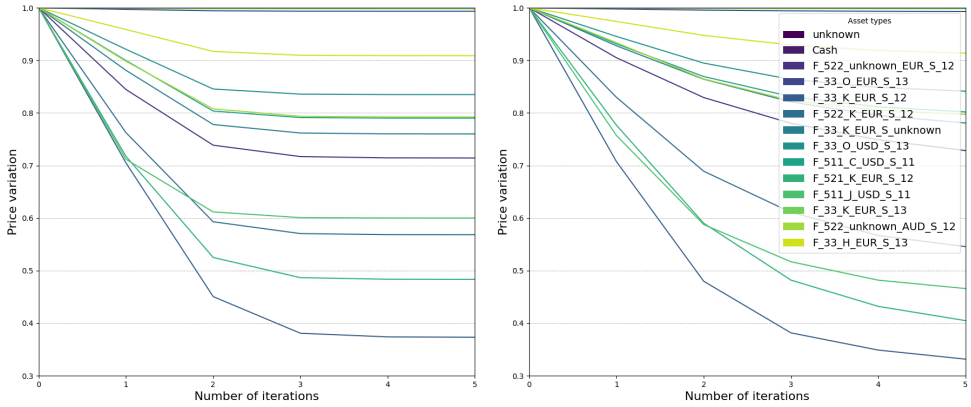
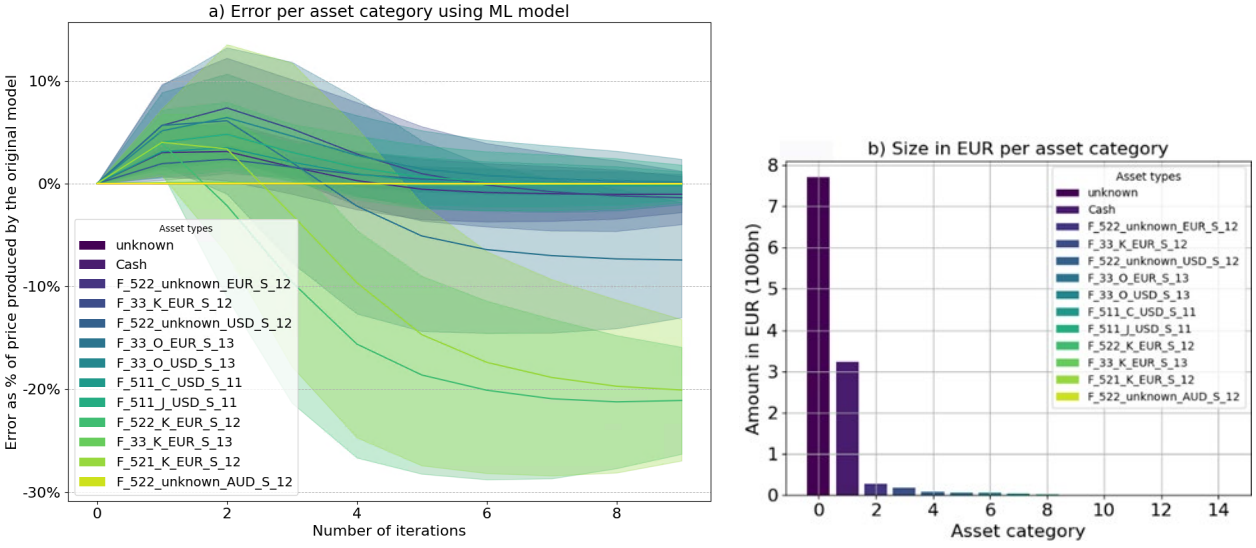


Figure 5: Assets price evolution following an initial random shock

The convergence of the error to zero for the largest asset categories validates the proof of concept and demonstrates that federated modeling is a viable and scalable approach for multi-sector, cross-border stress testing. To assess the model's error, we first simulate thousands of scenarios with random price shocks and analyze the resulting error distribution—defined as the relative difference in price evolution over several iterations between the original centralized fire-sale model and the FM global fire-sale model using the ML approximation. Figure 6 shows that the XGBoost model introduces biased errors in particular for smaller asset categories. In contrast, larger asset categories - such as the "unknown" type - are replicated more accurately, with errors below 10% of the modelled price change[11]. The smallest asset categories are represented by lighter colors, while the largest asset categories are represented by darker colors in Figures 6 a and b. Importantly, over successive iterations the error term converges to zero for the largest asset categories.

Figure 6: Price evolution – Original Fire-Sale Model vs. ML Proxy



[11] For the "unknown" asset category, we used the weighted average of all other asset categories.

In this proof of concept, access to centralized data allowed for direct comparison of the FM global fire-sale model to the original centralized global fire-sale model. In a real-world setting, however, the original global model built from private data would not exist. Instead, only the training data would be known. By splitting the data into training and testing – as explained in section 4.1 - the test data can serve as "new" input to assess the quality of the ML proxy model as shown in Figure 4. However, reconstructing model error across multiple iterations, as done in Figure 6, would not be possible.

The comparison between the two models' dynamics reveals the potential of the FM framework and highlights several takeaways. First, accuracy declines over successive iterations due to the natural accumulation of approximation errors and then improves and stabilizes as both models' prices converge after several iterations. In addition, despite its simplicity and training time of only a few minutes, the ML model already offers a reasonably accurate approximation. This makes it suitable for many analytical and stress-testing exercises, especially in areas in which no model currently exists and even indicative estimates are valuable. Lastly, replication accuracy is flexible and can be improved depending on model design choices and available computational resources.

5. Conclusion

This paper introduces Federated Modelling (FM) as a secure, scalable, and transparent solution to the conflict between data privacy and the need for cross-jurisdictional and cross-sectoral collaboration. FM bridges the gap between data privacy and the collective need for system-wide analysis, connecting bottom-up private models that rely on confidential data with a top-down holistic approach. In this way, data is used locally, models are defined locally, and outputs are analyzed globally in order to develop global insights.

The proof of concept of a system-wide fire-sale stress test model establishes FM as a new, viable, privacy-preserving and cost-efficient approach to system-wide stress testing. The proof of concept shows how collaboration across multiple participants – central banks or supervisory authorities – could happen without needing to share granular confidential data. This is a practical alternative to centralized data collections that are resource-intensive and often hindered by legal, operational, and institutional barriers.

The proof of concept shows promise, achieving reasonably good results with minimal computational and manpower resources. The results confirm that all necessary computations can be executed in just a few minutes on regular hardware using open-source Python tools. For more complex use cases involving higher-dimensional or less regular models, more adaptable machine learning techniques — such as neural networks —combined with more efficient function approximation methods can yield a broad spectrum of versatile and effective solutions.

The security features of the model call for high data quality standards and automated checks from the part of the participants. The decentralized nature of the data and models foreseen in the FM framework, together with the security protocol applied in this framework, mean that errors made by each participant are not traceable to them. Therefore, there is an inherent risk that either the system-wide outputs could be perceived as reliable when they are not and/or that when the outputs are considered unreliable it is not possible to identify the root cause of the problem. As a result, it is crucial that participants are required to implement internal automated data validation checks. A more thorough process could also be established to further ensure the reliability of the data points that each participant sends to the host, potentially including requirements related to local models and evidence that those requirements are met.

The benefits of the FM framework extend beyond the central banking community. The framework can serve any organization aiming to build shared, publicly-usable models from private granular data. Creating a shared library of trusted and accessible proxy models between regulators, academia, and the private sector could speed up interdisciplinary work — such as on climate or geoeconomic risks — and facilitate international cooperation.

A real-world pilot would pave the way for FM to shape the next generation of system-wide stress tests. With the technology already available and the computational cost proven manageable, a system-wide stress test pilot could be feasible in the short term. The pilot would establish FM as the framework to be used for the next generation of stress testing and broader financial stability assessments.

6. References

Bank of England. 2024. System-Wide Exploratory Scenario Exercise: Final Report. 29 November. Available at: https://www.bankofengland.co.uk/-/media/boe/files/financial-stability/system-wide-exploratory-scenario/boe-swes-final-report.pdf (Accessed: 22/03/2025).

BIS. "Data sharing practices". *Bank for International Settlements*, 17 Mar. 2023, https://www.bis.org/ifc/data_sharing_practices.pdf

Budnik, et al. 2023. Advancements in Stress-Testing Methodologies for Financial Stability Applications. ECB Occasional Paper No. 348.

Byrd, David, and Antigoni Polychroniadou. 2020. Differentially Private Secure Multi-Party Computation for Federated Learning in Financial Applications. arXiv:2010.05867v1, October 12.

Cetorelli, Nicola, Mattia Landoni, and Lina Lu. 2023. Non-Bank Financial Institutions and Banks' Fire-Sale Vulnerabilities. FRB of New York Staff Report No. 1057.

Cifuentes, Rodrigo, Ferrucci, Gianluigi, and Hyun Song Shin. 2005. "Liquidity Risk and Contagion." Journal of the European Economic Association 3(2–3): 556–566.

Cont, Rama, and Eric Schaanning. 2017. "Fire Sales, Indirect Contagion and Systemic Stress Testing. "

Duarte, Fernando, and João A. C. Santos. 2021. "Fire-Sale Spillovers and Systemic Risk." The Journal of Finance 76(3): 1251–1294.

Dinur and Nissim. 2003. "Revealing information while preserving privacy." Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems: 202–210.

European Central Bank. "Setting Standards for Granular Data." *European Central Bank*, 28 Mar. 2017, www.ecb.europa.eu/press/key/date/2017/html/sp170328.en.html.

Edward, Franklin R. 1999. "Hedge Funds and the Collapse of Long-Term Capital Management." Journal of Economic Perspectives 13(2): 189–210.

EIOPA. "Data Exchange Agreements." *European Insurance and Occupational Pensions Authority*, 2024, www.eiopa.europa.eu/about/accountability-and-transparency/data-exchange-agreements_en.

Eisenberg, Larry, and Thomas H. Noe. 2001. "Systemic Risk in Financial Systems." Management Science 47(2): 236–249.

Fernandez, J.D., Baima, R.L., Barbereau, T., Rieger, A. (2024). Opportunities and Applications of Federated Learning in the Financial Services Industry. In: Fridgen, G., Guggenberger, T., Sedlmeir, J., Urbach, N. (eds) Decentralization Technologies. Financial Innovation and Technology. Springer, Cham.

Fukker et al. 2022. "Contagion from Market Price Impact: A Price-at-Risk Perspective. " ECB Working Paper No. 2692.

International Monetary Fund (IMF). 2024. "Kingdom of the Netherlands–The Netherlands: Financial System Stability Assessment."

Lee, C. M., Fernández, J. D., Potenciano Menci, S., Rieger, A., & Fridgen, G. 2023. "Federated Learning for Credit Risk Assessment." Hawaii International Conference on System Sciences.

Liu, P., Xu, X. & Wang, W. 2022. "Threats, attacks and defenses to federated learning: issues, taxonomy and perspectives. " Cybersecurity 5, 4.

Lu, S., Li, R., Chen, X. et al. 2022."Defense against local model poisoning attacks to byzantine-robust federated learning. " Frontiers of Computer Science 16, 166337.

McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. y. 2017. "Communication-Efficient Learning of Deep Networks from Decentralized Data." Artificial Intelligence and Statistics: 1273–1282.

Merino, LH., Cabrero-Holgueras, J. 2023. "Secure Multi-Party Computation. " In: Mulder, V., Mermoud, A., Lenders, V., Tellenbach, B. (eds) "Trends in Data Protection and Encryption Technologies. " Springer, Cham.

Oualid, Ahmed, Yassine Maleh, and Laila Moumoun. 2023. "Federated Learning Techniques Applied to Credit Risk Management: A Systematic Literature Review. " EDPACS 68(1): 42–56.

Paillier, P. 2005. "Paillier Encryption and Signature Schemes. " In: van Tilborg, H.C.A. (eds) "Encyclopedia of Cryptography and Security. " Springer, Boston, MA.

Sydow et al. 2024. "Banks and Non-Banks Stressed: Liquidity Shocks and the Mitigating Role of Insurance Companies. " ECB Working Paper Series No. 3000.

Tolpegin, V., Truex, S., Gursoy, M.E., Liu, L. 2020. Data Poisoning Attacks Against Federated Learning Systems. In: Chen, L., Li, N., Liang, K., Schneider, S. (eds) Computer Security – ESORICS 2020. ESORICS 2020. Lecture Notes in Computer Science(), vol 12308. Springer.

UNSD. "Fundamental Principles of Official Statistics." *United Nations Statistics Division*, 29 Jan 2024, Fundamental Principles of Official Statistics