Economics and Organization of Logistics
9 (3), 2024, 103–138

DOI: 10.22630/EIOL.2024.9.3.22

Aleksandra Helena Pasieczna-Dixit✉**, Dariusz Starkowski**
Pomeranian University in Starogard Gdansk

Management and Identification of model risk sources: Examples using systemic risk measures

Zarządzanie i identyfikacja źródeł ryzyka modelu na przykładzie miar ryzyka systemowego

Abstract. As finance practitioners increase their reliance on computational models and data, the risk associated with erroneous models and improper model usage increases, with consequences for individual companies, regulatory bodies, and the larger economy. To help control model risk, we identify five categories of model risk sources: 1) dataset issues; 2) data processing related issues; 3) model construction related issues, 4) model implementation related issues; 5) model interpretation related issues. We justify this classification with examples in each category that study the impact of these sources on the estimates of systemic risk measures. The aim of the article is to draw attention to the need to create a framework for risk management in a logistics company based on a bank, which would include quantitative models. The article also presents general strategies that can be implemented in such a risk management framework in a logistics enterprise using the example of a bank.

Key words: Model Risk, Model Risk Management, Systemic Risk Measures, Monte Carlo, bank, Banking, Logistics company

Synopsis. W miarę jak praktycy finansowi coraz bardziej polegają na modelach obliczeniowych i danych, wzrasta poziom ryzyka związanego z błędnymi modelami i niewłaściwym wykorzystaniem modeli, co ma konsekwencje dla poszczególnych firm, organów regulacyjnych i całej gospodarki. Aby pomóc w kontrolowaniu ryzyka modelu, identyfikujemy pięć kategorii źródeł ryzyka modelu: 1) problemy ze zbiorem danych; 2) problemy związane z przetwarzaniem danych; 3) kwestie związane z konstrukcją modelu; 4) kwestie związane z implementacją modelu; 5) kwestie związane z interpretacją modelu. Uzasadniamy tę klasyfikację przykładami w każdej kategorii, które badają wpływ tych źródeł na oszacowania miar ryzyka systemowego. Celem artykułu jest zwrócenie uwagi na potrzebę stworzenia

✉**Aleksandra Helena Pasieczna-Dixit** – Pomeranian University in Starogard Gdansk, Institute of Security and Aministration; e-mail: wwwpasiecznapl@gmail.com; https://orcid.org/0000-0001-6867-3584

Dariusz Starkowski – Pomeranian University in Starogard Gdansk, Institute of Mining Economics and Logistics; e-mail: dariusz.starkowski@twojestudia.pl; https://orcid.org/0000-0001-9232-7673

ram zarządzania ryzykiem w przedsiębiorstwie logistycznym na podstawie banku, które obejmowałyby modele ilościowe. W artykule przedstawiono również ogólne strategie, które można wdrożyć w takich ramach zarządzania ryzykiem w przedsiębiorstwie logistycznych na przykładzie banku.

Słowa kluczowe: ryzyko modelu, zarządzenie ryzykiem modelu, miary ryzyka systemowego, Monte Carlo, bank, bankowość, przedsiębiorstwo logistyczne instytucja finansowa

Kody JEL: D81, C63, O1/P4

Introduction

Quantitative finance is usually assumed to have started with the doctoral thesis "The theory of speculation" written by Louis Bachelier [Bachelier 1900, Cesa 2017]. Since then, mathematical and computational models have been used for a wide variety of applications in finance, including pricing [Black and Scholes 1973] and evaluating risk profiles [Merton 1974]. Due to the digitalization of modern businesses, mathematical and computational models have been incorporated within banking logistics as well, such as for reducing the costs associated with cash inventory optimizations [Baker et al. 2012], cash demand forecast [Cedolin et al. 2024] and customer flow optimization [Madadi et al. 2013].With much of the finance world relying on models, it is not surprising that, at some point, many of these models may cause various financial crises, whereas others indicate the reliance on these models for causing the crises [Weatherall 2013].

Given the widespread use of models, it is thus important that the practitioners understand the risks associated with their use. Improper model risk management can be damaging for investors, companies, and, under some circumstances, propagate these negative effects to the larger economy. The term model risk here refers to the risk associated with the general development and usage of models and is a subject of active research. Models are built on observed data and have implicit and explicit assumptions about the workings of the world.

Since they approximate various phenomena, their value in guiding practitioners can only be as good as the data used to build the model, the assumptions that entered the model, the implementation of the model, and the interpretation of the results. According to Derman [Derman 1996], model risk is a consequence of general model construction and uncertainty in the field of finance, a view shared by some researchers [Crouhy et al. 1998]. Everything related to a model is thus a part of the model risk, including data contamination, incorrect implementations, poorly approximated solutions, software or hardware bugs, and the practitioners themselves. With this view, it is difficult to isolate the individual sources of risk for risk control.

The goal of any organization is, among other things, to manage efficient logistics processes. Its implementation requires the implementation of a risk management system so that it is possible to estimate significant risk groups accompanying the implementation of the logistics processes, develop integrated risk management strategies, and develop risk analysis tools in logistics processes. Risk management is the planned and deliberate analysis, control and control of risk positions. Planning means a systematic, not random analysis, while purposefulness means conscious tracking of the positions, opportunities and risks of a specific institution derived from the company's goals. Effective risk management in logistics requires the identification and analysis of potential threats, planning and implementation of the risk minimization measures, plus regular monitoring of their effectiveness and adaptation to changing the business conditions.

Planning means a systematic rather than a random analysis, while purposefulness means conscious tracking of the positions, opportunities and risks of a specific institution derived from the company's goals. Effective risk management in logistics requires the identification and analysis of potential threats, planning and implementation of risk minimization measures, and regular monitoring of their effectiveness and adaptation to changing business conditions.

Some authors study model risk with more focused points of view. For example, few research works define model risk as the inaccuracy arising from estimation errors and the use of incorrect models [Boucher et al. 2014, Glasserman and Xu 2014, Hendricks 1996]. Few other authors consider model risk induced by the data-fitting approach used for statistical modelling, namely, the choice of tests for the data and estimation of the model parameters [Sibbertsen et al. 2008]. For readers interested in the quantification of model risk, we indicate the active research occurring in this domain [Danielsson et al. 2016, Banulescu-Radu et al. 2020, Pasieczna 2021]. As opposed to quantifying the model risk, the aim of this work is to identify and categorize the model risk sources.

This paper highlights the need for better model risk management, and endeavors to push for proper documentation, implementation, and usage of quantitative models, especially for those relying on large amounts of data. While listing the model risk sources, this work relies on examples based on market-data based models that quantify systemic risk (SR). Specifically, we highlight potential issues that occur with the computational tools used for estimating stress in a financial network. SR is the risk of collapse of the entire financial system. As opposed to the risk on an individual bank, the risk is spread (and realized) across the entire financial network. A systemic collapse tends to have triggers at individual banks (e.g., bankruptcy of a very important bank), and can have consequences that leak into the larger economy (e.g., a national level recession). Systemic risk measures (SRMs) aim to quantify the level of stress of the system and recognize the major contributors, and so SRMs are a crucial tool for regulators and banking institutions. Since systemic events impact the larger economy, it is highly important that they accurately measure that for which they were proposed. Underestimating SR might result in more risk-taking behaviors by banks, leading to an increase in the overall financial stress and an increased risk of a systemic collapse with long-term consequences for the real economy. On the other hand, overestimating SR can cause regulators to apply excessive penalties on banks, leading to a decrease in economic stimulus and a consequent slow-down in economic growth.

The importance of banking logistics cannot be understated in the context of SR, since failures in appropriate modeling of the various processes in liquidity and risk management can result in a shock that negatively impacts the broader financial system. For example, delays or disruptions in replenishment of cash at ATMs due to external factors (e.g., natural disasters, war) can lead to customers causing bank runs, resulting in a liquidity crisis and an eventual propagation of the contagion into the larger financial system.

Typically, there are two main model families (not a strict classification) used when measuring SR: network models and market-data based models. Network or graph models rely on balance sheet data to model the financial network, and use interbank connections (such as interbank loans) to propagate shocks [Cifuentes et al. 2014, Hurd 2016]. Market-data based models, the focus in this work, rely on market data, such as stock prices, to estimate the correlations that indicate bank co-movements due to interbank dependencies as well as exposures to common assets. The latter family has certain advantages, the main being that market data is accessible to the public and considered more informative (transparent) for investors. Additionally, market data is available at different frequencies (monthly, weekly, daily, intra-day), as compared to balance sheet data, typically available only once a quarter. Higher frequency data can be very insightful to analyze extremely fast SR contagion events.

While this work applies to a wide variety of SRMs and other computational models, we focus on two candidate SRMs to highlight potential sources of model risk. Both SRMs are related to the too-connected-to-fail (TCTF) aspect of SR. The first SRM is the marginal expected shortfall (MES), which looks at how a particular bank reacts when the market under-performs [Acharya et al. 2010, Brownlees and Engle 2012]. It can be interpreted as the marginal contribution of a bank to market falls [Idier et al. 2013], and thus attempts to quantify the SR using the inter-connectedness of the financial network. If the MES of the bank is estimated with an external market index (as opposed to the index it belongs to), then it can represent the performance (sensitivity) of the bank to bad market days, and thus looks at the company's exposure to SR. The second SRM studied in this work is the Delta Conditional Value at Risk (ΔCoVaR), which looks at how the market risk changes to bank crashes [Adrian and Brunnermeier 2008, Castro and Ferrari 2014]. Like the MES, the ΔCoVaR also considers the aspect of inter-connectedness in quantifying the SR of the financial system. The chosen SRMs differ only in direction: one examines the reaction of banks to market falls (banks' exposure to SR), the other examines the reaction of the market to bank crashes (financial network's exposure to individual banks).

We expect both SRMs to act independently of each other, except when the reference market index and the bank are highly correlated to each other. Typically, if an appropriate reference market index is chosen, such situations do not arise, and these SRMs capture slightly different effects from one another.

Despite their advantages, market-data based SRMs have certain limitations. Firstly, inter-connectedness is inferred from price returns correlations. It is difficult to conclude whether an observed correlation is due to an interbank dependency or an exposure to a common factor. Secondly, market-data based metrics give little to no importance to the precise cause of systemic triggers, and typically simulate shocks as drops in stock value of a bank or drops in a reference market index. Thirdly, they have high model risk [Danielsson et al. 2016], with even identification of systemically important banks with SRMs being prone to errors due to estimation risk alone [Danielsson et al. 2016]. The last point makes their use limited in the context of regulation, since if the model risk is high, the model outcomes are less dependable, and it becomes difficult for regulators to judge the systemic importances of banks. Given the high model risk of SRMs, they are good candidates to illustrate the need for better model risk management, which is the aim of this paper. The risks associated with improper management of model risk can be devastating for individual banks, which are specific logistics companies, and the effects can translate into the real economy.

The rest of the article has the following structure. Section 2 – Theoretical Concepts and Tools contains the theoretical concepts and tools used, namely, classification of the model risk sources, brief definitions of the SRMs, description of the data used, and the implementation details of the algorithm used to estimate the SRMs. Section 3 – Research results, being the main contribution of this article, expands upon the classification of the model risk sources with technical examples pertinent to the estimation of the chosen SRMs. When possible, we provide suggestions that might be used to reduce the model risk associated with certain sources. Section 4 discusses the implications of the work in the context of SRMs, along with schemes to manage the model risk. Section 5 concludes the article.

The aim of this work is primarily to identify and classify the sources of model risk, along with the intention to create a dialogue between model developers and risk managers so that model risk can be properly estimated by the logistics company. To substantiate this classification, examples are provided of the impact that these sources have on the estimates of logistic and systemic risk measures used by the banks and regulators to quantify the magnitude of stress in the financial system. The examples used cover a specific application, although it can be used in other branches of the economy, including logistics.

Materials and Methods

Theoretical Concepts and Tools

Identification of Model Risk Sources

Multiple researchers have attempted to identify model risk sources in different financial models. For example, Derman [Derman 1996] identifies the following seven sources of model risk:

- inapplicability of modelling,
- incorrect model,
- correct model, incorrect solution,
- correct model, inappropriate use,
- poorly approximated solution,
- software and hardware bugs,
- unstable data.

In another example, Kato et al. [2000]) state that in pricing models, the sources of model risk include:

- use of incorrect assumptions,
- errors in the estimations of parameters,
- errors resulting from discretization,
- errors in market data.

For risk measurement models, they identify the difference between assumed and actual distribution, and errors in the logical framework of the model. In a paper by Management Solutions [Lamas et al. 2014], the authors identified three categories of model risk sources:

- data deficiencies in terms of availability and quality,

– estimation uncertainty or model error,
– model misuse.

While these classifications differ from one practitioner to another, model risk sources tend to lie along three main dimensions:
– data,
– the computational model,
– the interpretation (more generally, the usage) of the model outcomes.

Given that data is becoming increasingly central to financial models, we find it appropriate to divide the data axis into two categories – (1) issues related to dataset collection and description, (2) issues related to data-processing, where the raw data is treated or transformed before being used in computational models. Additionally, finance practitioners continue to build and rely on computational models, and so we divide the model risk sources within the computational model axis in two categories – (1) model construction related sources, i.e., issues related to the model development at the abstract level, and (2) model implementation related sources, i.e., issues dealing with the practicalities of software and hardware. With this, we propose the following classification of model risk sources.

1. Dataset Issues: These sources refer to risk sources that practitioners are exposed to during dataset collection and description processes.
2. Data Processing Issues: These are risk sources that appear during the transformation of raw data to make the datasets more 'model friendly.'
3. Model Construction Related Issues: These sources of model risk specifically consider issues associated with the development of the model algorithm, and not with the practicalities of the model implementation.
4. Model Implementation Related Issues: These model risk sources refer to problems linked to the practical implementation of the model, including software and hardware limitations.
5. Model Interpretation Related Issues: This category of model risk sources deals with issues linked to the usage of model outcomes for decision-making processes.

While this proposal is based on literature review and personal experience, it can be used to initiate discussions for better model risk management. In Section 3, we provide a detailed explanation of these categories along with examples in each category that help justify our proposed list. Though the examples lie within the context of market-data based SRMs, the list applies to all types of models (e.g., forecasting, optimization models, inventory management models risk models, pricing models, portfolio optimization models) that use data and computational tools. This proposed classification aims to create dialogue between the modelers and risk managers, with the intent of identifying major model risk sources within different processes and (hopefully) mitigating model risk.

Systemic Risk Measures

To highlight the model-risk sources present in SRMs, we selected two widely used market-data based SRMs, which we briefly describe here.

Both SRMs do not consider the cause of the shock that causes a collapse, and simply consider that the bank or the market has had an event which caused its market value to fall. The event at a bank could be due to realization of the systematic risk (broader trend of market behavior), or unsystematic risk (specific to the company, e.g., inventory problems, cash mismanagement, local bank runs). The event at the market could be due to macro-announcements (e.g., pandemic lockdowns, war, trade sanctions, supply chain restrictions). While we focused on SRMs, the analysis of the model risk can be applied to most computational models, including forecasting and risk management.

i. Marginal Expected Shortfall

The MES of a bank is defined as the average performance of the bank when a reference market index is in its left tail [Acharya et al. 2010, Brownlees and Engle 2012, Idier et al. 2013]. Thus, it looks at the exposure of the bank to market falls. Mathematically:

$$MES_{i,t}(\alpha) = \mathbb{E}\left[R_{i,t} \middle| R_{m,t} \leq VaR_{m,t}(\alpha) \right]. \tag{1}$$

Here, $R_{i,t}$ and $R_{m,t}$ represent the price returns of the bank i and the market m at time t. $\mathbb{E}[x]$ is the expectation value of x. $VaR_{m,t}(\alpha)$ is the Value-at-Risk (VaR) of the market at confidence level α, and is defined as the maximum possible loss, whose probability is within a pre-defined confidence level over a predefined time horizon [Hendricks 1996, Holton 2003, Pasieczna 2019]. In a Monte Carlo (MC) setup like ours, the MES of a bank is the average simulated returns of the bank over the MC iterations satisfying the condition that the market's simulated returns are below the simulated market VaR.

ii. Delta Conditional Value at Risk

The ΔCoVaR was proposed to study the contribution of a bank to overall SR [Adrian and Brunnermeier 2008, Castro and Ferrari 2014]. As the name suggests, it looks at the change in the CoVaR, defined as the market VaR conditional on an event at the bank. Mathematically:

$$\Delta CoVaR_{i,t}(\alpha) = CoVaR_t^{\left[m \middle| R_{i,t} = VaR_{i,t}(\alpha)\right]}(\alpha) - CoVaR_t^{\left[m \middle| R_{i,t} = VaR_{i,t}(0.5)\right]}(\alpha). \tag{2}$$

Here, $CoVaR_t^{\left[m \middle| R_{i,t} = VaR_{i,t}(\beta)\right]}(\alpha)$ is defined as the VaR of the market m at confidence level α when the returns of the bank i are at their VaR (confidence level β). $\beta = 0.5$ implies median returns at the bank ('normal' functioning). Thus, the ΔCoVaR SRM looks at how the market risk (defined through the market VaR) changes when the bank crashes. While it is possible to estimate the ΔCoVaR using a quantile regression [Bianchi and Sorrentino 2020], we use a slightly different approach where we compute the market VaR over a certain range (defined as 1% = ±0.5% of the total MC iterations) of simulated market returns around the required quantile (i.e., $\beta \pm 0.005$) of the bank's simulated returns.

Data Description

Our chosen list of banks was a subset of the banks considered systemically important and directly supervised by the European Central Bank (ECB). ECB regularly updates and publishes the full list online along with criteria for significance [European Central Bank 2021]. These banks function in countries which have adopted the Euro as their official currency, or whose currencies are pegged to the Euro (as is the case of Bulgaria).

We had collected the data in August 2021, when the ECB supervised 114 banks, of which, we kept 47 banks for study. The reasons for rejecting the remaining institutions were unavailability of data in Bloomberg, unlisted (or pending listing) on the market exchanges, a private company, or having been acquired by a bank already included. The complete list of banks along with their grounds for significance is provided in Table 1.

Price and outstanding shares data were obtained using the Bloomberg terminal [Bloomberg L.P. 2021]. Whenever possible, the data for each bank began on the first of January 2000, implying a history of slightly over 20 years until August 2021. Since the list of banks is static, based on the supervised list in August 2021, we expect survivorship bias in the dataset, and we discuss its implications in an example (Section 3.A.ii).

Computational Methods

This section describes the computational methods used to estimate the MES and ΔCoVaR. It consists of two subsections, the first dealing with the self-built market index used as the reference market index in the SRMs, and the second describing the Monte Carlo process to compute the SRMs.

i. Self-built market index

Here, a self-built market index was used as a reference market index for the SRMs. This index is constructed from the 47 banks described in Table 1, using the following steps:

- For every day within the simulation period, compute the market capitalization for each bank as the product of its last known outstanding shares and price.
- Add the market capitalization over all banks present on that day.
- Normalize this aggregated market value by dividing it using a divisor.

$$\text{index}_d = \frac{\sum_{i \in \text{comp}_d} \text{mcap}_i(d)}{\text{divisor}_d}. \tag{3}$$

Here, index_d is the value of the index, copm_d is the composition of the index, $\text{mcap}_i(d)$ is the market capitalization of bank i, and divisor_d is the divisor value on day d. The missing price and outstanding shares values were forward filled (last known data used) for the market index, as it represents the amount of wealth generated by these banks.

The divisor was reconstructed each time a new bank enters the market and used from the following day. This was necessary to ensure that the market index only takes past data into account. The first day was skipped for every bank in the index. Additionally, the starting value of the divisor was chosen such that the index begins with a value of 100 points. The expression for updating the divisor is:

$$\text{divisor}_{d+1} = \text{divisor}_d = \frac{\sum_{i \in \text{comp}_d} \text{mcap}_i(d)}{\sum_{i \in \text{comp}_{d-1}} \text{mcap}_i(d)}; \tag{4}$$

$$\text{divisor}_{d=2} = \frac{1}{100} \sum_{i \in \text{comp}_{d=1}} \text{mcap}_i(d=1). \tag{5}$$

There are two advantages to this index – (1) the start date of the market index is the first simulation date of the banks (01/01/2000), and (2) the index fully describes the movements of the banks and contains no external information. However, as mentioned earlier, we expect the presence of survivorship bias in the dataset, thereby impacting the perceived index performance (see Section 3.A.ii for its impact).

ii. Computational algorithm for SRMs

We use a Monte Carlo (MC) approach to estimate the two SRMs. MC techniques are used in simulating problems with uncertainty, and so are quite apt for risk analysis [Savvides 1994, Glasserman et al. 2000, Pasieczna 2019], including SRMs [Lehar 2005, Glasserman 2005, Minderhoud 2006, Siller 2013, Koike and Hofert 2020]. Since our aim is to identify model risk sources, as opposed to the estimation of SRMs or their model risk, we opt for a simplistic MC scheme to emphasize the risk associated with the mismanagement of model risk sources. Except in Section 3.C.i, price returns of the bank or the market are defined as logarithmic returns. The algorithm consists of two steps: 1) estimation of the distribution of price returns using the past data, and 2) generating future returns using the estimated distribution, from which we estimate the SRMs:

1. Estimation of price returns distribution: For each day of the simulation period, we use a rolling window (length 250 trading days, approximately 1 year) to estimate the price returns distribution:
   – We estimate the mean, standard deviation of the returns of all banks and the market index. The correlation values between the returns of each bank and the market index are also estimated.
   – An additional degrees of freedom parameter (proxy for the tailedness) is estimated for the returns of all banks and the market index when the Student's-*t* distribution is used.
2. Estimation of SRMs with the estimated distribution:
   – Using a Gaussian or Student's-t distribution with the parameters estimated, we generate 20,000 future price returns (except in Section 3.D.ii) for each bank and the market index using a bivariate process [Brownlees and Engle 2012]:

$$r^n_{m,d+1} = \mu_{m,d} + \sigma_{m,d}\epsilon^n_{m,d}, \tag{6}$$

$$r^n_{m,d+1} = \mu_{i,d} + \sigma_{i,d}\rho_{im,d}\epsilon^n_{m,d} + \sigma_{i,d}\sqrt{1-\rho^2_{im,d}}\ \epsilon^n_{i,d}. \tag{7}$$

– Indices i, m, n and d refer to the bank, the market index, the MC iteration, and the simulation day, respectively. $r^n_{i/m,d+1}$ are the simulated returns for the next simulation day, $d+1$. $\mu_{i/m,d}$, $\sigma_{i/m,d}$ and $\epsilon^n_{i/m,d}$ are the estimated means, the estimated standard deviations and drawn random numbers from a standardized distribution (Gaussian or Student's-*t*) respectively.$\rho_{im,d}$ is the correlation between the returns of the bank and market index.

– MES estimation (confidence level α): Estimate the VaR of the market as the α^{th} quantile from the simulated market index values. The MES of a bank is the average of the simulated returns of a bank when the simulated market returns were below this VaR.

– ΔCoVaR estimation (confidence level α): For a given bank, consider all simulations of the market index returns, where the simulated returns of the bank were between the VaR at $\alpha + \varepsilon$ and $\alpha - \varepsilon$ quantiles. The VaR at α confidence level of these selected market returns is the stressed CoVaR. The unstressed CoVaR is similarly computed as the VaR at confidence level α of the selected market returns, where the bank's simulated returns were between the $0.5 + \varepsilon$ and $0.5 - \varepsilon$ quantiles. The difference between the stressed and unstressed CoVaRs is the ΔCoVaR. ε was set to 0.005 (1% of all MC iterations are selected), and the impact of changing this value is discussed in Section 3.C.ii.

Here, the missing data points were not treated (Section 3.B.i discusses the impact of the treatment type), and the distribution parameters were estimated by dropping these points. For statistical stability purposes, we required a minimum of 80% of available data within the rolling window for computing the mean, standard deviation and tailedness parameters, and a minimum of 60% for the correlation parameter. If this was not satisfied for a particular bank on a given day, no SRM was estimated for that bank on that day. For the Student's-t distribution, we had to impose a minimum value of 5 for the degrees of freedom parameter. The impact of this choice is discussed in Section 3.D.i.

Research Results and Discussion

Model Risk Sources: Examples using systemic risk measures.

This work is based on case studies that show the impact of different sources of model risk on the results obtained. The proposal to classify the sources of risk of the model was presented on the basis of our own research and a literature review.

The case studies focus on the quantification of two measures of systemic risk - Marginal Exceeded Shortfall (FEM) and Delta Conditional Value at Risk (ΔCoVaR), whose values are estimated using the Monte Carlo simulation method. The data used for the analysis includes market data on banks monitored by the Logistics company on the example of European Central Bank.

Dataset Issues

Data collection and description can be huge sources of model risk, depending on how the data is used. Both issues can be due to technological limitations (poor measurement tools), the nature of rare events, or human error (e.g., negligence). Poor quality data, for example, can make it extremely hard to extract the necessary information for practical use. Models that are built on bad data tend to give unreliable results, referred to in data science fields as 'Garbage In – Garbage Out' [Kilkenny and Robinson 2018]. Incomplete or missing data can adversely impact models that rely on inferring statistical properties of certain variables from the data. Just like quality and the completeness of data, documentation and description of the dataset is important. If the datasets are incorrectly described, then models built on those datasets will be exposed to high model risk. Irrelevant data might be used during model construction, acting as another potential source of model risk.

Data collection can also be subject to biases, such as survivorship bias where the dataset reflects only the entities that 'survived' until the time of the data collection process

[Gilbert and Strugnell 2010]. For example, if we collect historical market prices of the S&P-500 companies based on today's composition, then our dataset does not contain the companies removed from the index.

This leads to under-representation of companies that performed worse, and models built on this dataset are exposed to this bias. Financial data has been known to be subject to other types of biases as well [Bryant et al. 2019, Zhong and Hamilton 2023].

Since computational models are built on datasets, then dataset issues are a major source of model risk. Managing model risk associated with this source requires a huge effort on the part of banks, mainly on the side of data collection (either directly from markets, or from data providers), maintenance (includes monitoring and when possible, correcting, for any potential bias) and documentation of the datasets. This is a non-trivial problem, and many banks tend to have dedicated data teams for leading this effort. Data management is a rapidly evolving domain, and risk managers need to consider and account for the ever-growing need of researchers and practitioners for more data for better computational models [Dicuonzo et al. 2019].

To further underline the need for better data management, we study two specific examples – (1) poor data quality due to missing data, and (2) survivorship bias. In the first example, we show the deviations of the MES at 95% confidence levels for three banks as a function of missing data. In the second example, we show how the reference market index with survivorship bias has a slightly optimistic outlook when compared to another reference market index without this bias.

i. Poor data quality

This example highlights the impact of missing data on the MES estimates (95% confidence levels), using three banks:

– Crédit Agricole S.A.
– Deutsche Bank AG, a
– Banco Santander S.A.

First we estimated the MES using all available data, then 5% of all price data was randomly removed before recomputing the MES, and, finally, 10% of the total data was randomly removed for the third MES computation. Figure 1 shows the histograms of the deviations (in relative units with respect to the MES computed with all data) of the MES at 95% confidence levels for the three banks when 5% (orange) or 10% (green) of the data was removed.

We observe that the green histograms that have 10% missing data have more spread (mean: –0.26%, std: 5.09%, min: –22.34%, max: 13.00%) than the orange histograms which have 5% missing data (mean: 0.61%, std: 3.22%, min: –19.14%, max: 15.74%). This indicates that the model risk (defined as variability of the estimates) increases with the percentage of missing data. As price data goes missing, estimation of the statistical properties of the returns becomes unreliable, leading to an increase in the MES deviations. This has important consequences for SRMs. For example, if one were to choose the most significant banks from a larger dataset, then with just 5% of missing data (selected 5,259 data points out of 5,536) we observed variability of about 3%. With 10% of the missing data, the observed variability reaches about 5%. The perceived relative risk contributions of the banks can vary quite significantly, leading to an unreliable ranking of the banks'

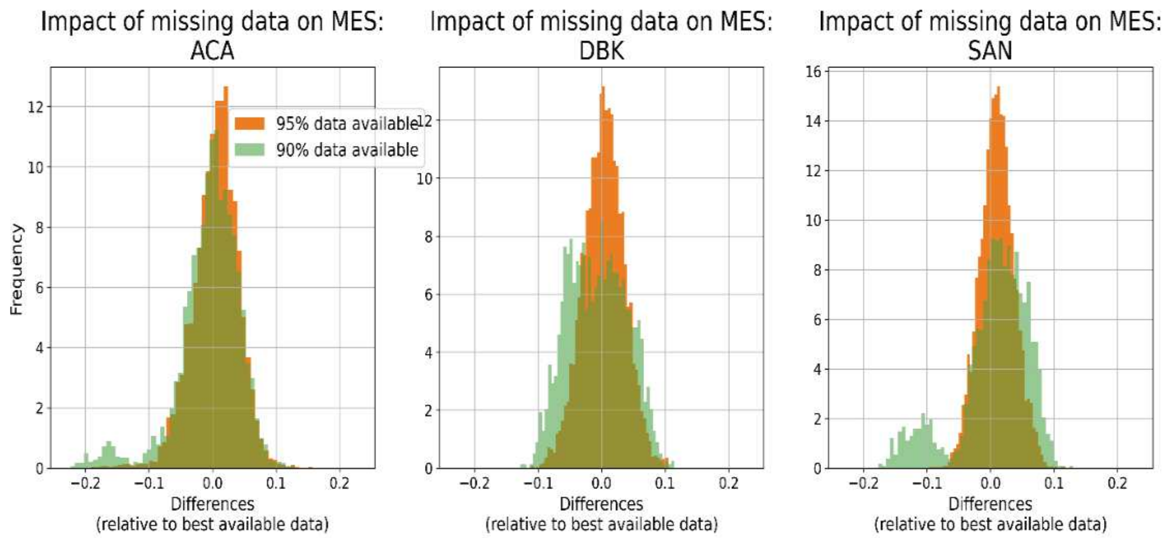

Figure 1. Relative differences of MES at a 95% confidence level from the MES computed with the best available data

Rysunek 1. Względne różnice MES obliczonego na podstawie najlepszych dostępnych danych z MES obliczonym na podstawie brakujących danych

Source: banks chosen: ACA (Crédit Agricole S.A.), DBK (Deutsche Bank AG) and SAN (Banco Santander, SA)
Źródło: wybrane banki: ACA (Crédit Agricole S.A.), DBK (Deutsche Bank AG) i SAN (Banco Santander, SA)

systemic importances. Usage of the results by regulatory bodies without accounting for the associated model risk can lead to the regulators applying inappropriate penalties on less risky banks or being lenient on more risky banks.

i. Survivorship bias

This second example related to dataset issues deals with the issue of survivorship bias and discusses the implications, in the context of the perceived performance of the market indices. The dataset used in this work consists of market prices and the outstanding shares amount of 47 banks, which form a subset of systemically important banks supervised by the European Central Bank (ECB).

As mentioned earlier, the dataset reflects the banks supervised in August 2021, and does not consider the temporal change of this list. Hence, we expect survivorship bias to be present. To understand why, let us consider a bank that was supervised by the ECB prior to the 2015–2016 stock market sell off, but not supervised (became less significant) after the crash. We would not have downloaded price data about this bank. A reference market index built without this bank would not see the downward performance of the bank, the consequence of which is that this index would outperform another index that considers the temporal evolution appropriately. This perceived outperformance is an artefact of how the dataset was built and is the outcome of the survivorship bias present.

In the left subplot of Figure 2, we compare the temporal evolution of the self-built market index (Section 2.D.i) that contains the survivorship bias and the Euro Stoxx Banks index, or SX7E (QONTIGO, 2011) without this bias. The data for the SX7E index was obtained from https://www.investing.com/ and begins on 28 December 2012, which implies a limited history for comparison of the two indices. Over the common history, we

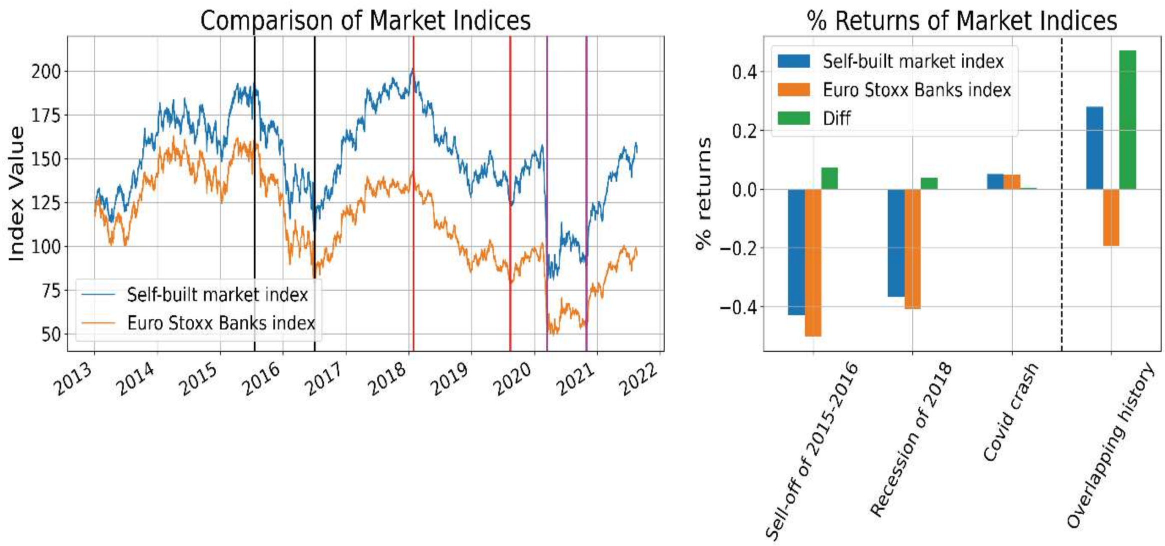

Figure 2. Comparison of the self-built market index that contains survivorship bias and the Euro Stoxx Banks index that does not contain survivorship bias

Rysunek 2. Porównanie samodzielnie zbudowanego indeksu rynkowego, zawierającego błąd przeżywalności, z indeksem Euro Stoxx Banks

Source: own work based on market index with Euro Stoxx Banks index

Źródło: opracowanie własne na podstawie indeksu rynkowego z indeksem Euro Stoxx Banks

observe similar behaviors for the two indices, with correlations between the respective log-returns reaching 97.76%. The self-built index seems to be a good proxy of the market index and has more available data.

Nonetheless, the self-built index outperforms the SX7E index, hinting at a non-negligible impact of the survivorship bias. To further analyze this issue, we computed the cumulative returns of both indices over 4 periods:

– sell-off of 2015–2016 [Irwin 2015, Li et al. 2015],
– recession of 2018 [DW 2019],
– COVID crash [Haldar and Sethi 2021],
– the entire overlapping history.

The results are presented in the bar plot on the right subplot of Figure 2. The difference bar (in green) corresponds to the perceived outperformance of the self-built market index over the SX7E index. We observe that the self-built index consistently beats the SX7E, with the difference being the smallest (but still positive) during the Covid crash period that is much closer to the data collection time.

Despite the remarkable similarity over the common history, we expect the impact of survivorship bias to be more pronounced before the start of the SX7E index since the index compositions would have differed much more. There are consequences on the evaluation of the SRMs as well, since the market index appears less risky, leading to an under-estimation of the network stress. This is particularly important in the context of the development of new SRMs, and the back-testing of the previous crises. Having survivorship bias might cause the SRM to help ‘avert’ a particular crisis, simply because the data consists of banks that have already averted (survived) the crisis.

Data Processing Related Issues

Before being used in models, some type of data processing might be done to make the data more ‘model friendly.’ This step introduces more potential sources of model risk, and practitioners need to be able to deal with them. Some examples of data processing steps include dealing with missing datapoints, correcting or flagging erroneous datapoints, re-arranging data into a more appropriate form, and performing small transformations on the dataset. If datasets are small, then manual verification, even if slightly cumbersome, of the data processing steps is possible, and the risk associated with this stage can be minimized. However, as datasets become larger, performing tasks manually become intractable, with many steps tending to become automated, leading to an increase of model risk.

To better understand this stage, consider the problem of detecting and flagging large price jumps in a stock’s daily time-series. If the jump exists for a day, it might most likely indicate that there is an error in the data, less likely reflect an extreme event, and flagging this jump as an error would require confirmation and verification. If the jumps are uncommon, a human might be able to verify each jump case by case and flag them manually. However, if the jumps are too many, the problem becomes intractable for manual verification, and practitioners turn to automation with some statistical metric (such as percentage of datapoints flagged as errors). Automation then must deal with having either too many datapoints flagged as errors (missing out true tail events) or having too many erroneous datapoints.

When the data processing stage becomes automated, more sources of model risk are introduced, typically in the form of computer bugs. In an ideal case, these bugs are caught and managed when the data processing steps are applied on many tests. However, bugs can pass all internal controls, and their impacts become known only when the transformed data is used for model building or during an audit. A well-known example is the rounding error bug of the Vancouver Stock Exchange Index [Nievergelt 2000], where small errors in rounding accumulated to a very large tracking error (around 50%) of the published index value with respect to the actual index value, causing the index to appear to be falling instead of growing. Using the published index value in computational models would indicate poor market performance (despite an increase in the value of daily transactions).

Model risk management at this stage typically depends on whether the data processing is manual or automatic. If it is manual, rigorous checks by different data analysts need to be implemented with very stringent controls. If automation is used, then good software practices would additionally need to be implemented. As is the case with datasets, computer programs need to be documented, tested, and released. Monitoring will additionally be required to ensure that the data processing programs remain robust to newer, and untested, circumstances of the real world. Implementing these practices comes at a cost to banks, both in terms of money and time, but the risk associated with bad implementation is typically much higher. To highlight the impact of the data processing stage, one example is provided below, where we compare the MES at a 95% confidence level for one bank computed with three ways of dealing with missing price data.

i. Dealing with missing price data

In this case study, three options are considered to deal with missing price data:
- drop the points,
- forward-fill the prices (use the last available information) before estimating the price return statistics,
- back-fill the prices (use the next available information in the past) before estimating the price return statistics.

All three approaches have problems, which manifest in the estimation of the SRMs. If we drop the points, then depending on the number of missing datapoints, we might end up with insufficient data, or unreliable statistics. If we forward-fill the prices, then the price returns will be zero for all the missing points, which might cause the variance of the returns to be lower. Finally, if the missing time-series data is back-filled, then the models are provided information before it should be available, leading to leaked information from the future (breaks in causality). The third approach should ideally never be used for SRMs, since it is possible to gain information about a market crash before the actual crisis. It is included only for teaching purposes.

We estimated the MES at a 95% confidence level for HSBC Bank Malta p.l.c. using these three options. Figure 3 contains: (left) the prices (in red), along with a rolling 250-day count (in black) of available price data, and (right) the MES curves for the three approaches of dealing with the missing data. The plots are zoomed in on the period from July 2008 to January 2010. We see that there are more holes between August 2008 and October 2008, the period containing the September 2008 crash, and the MES curves

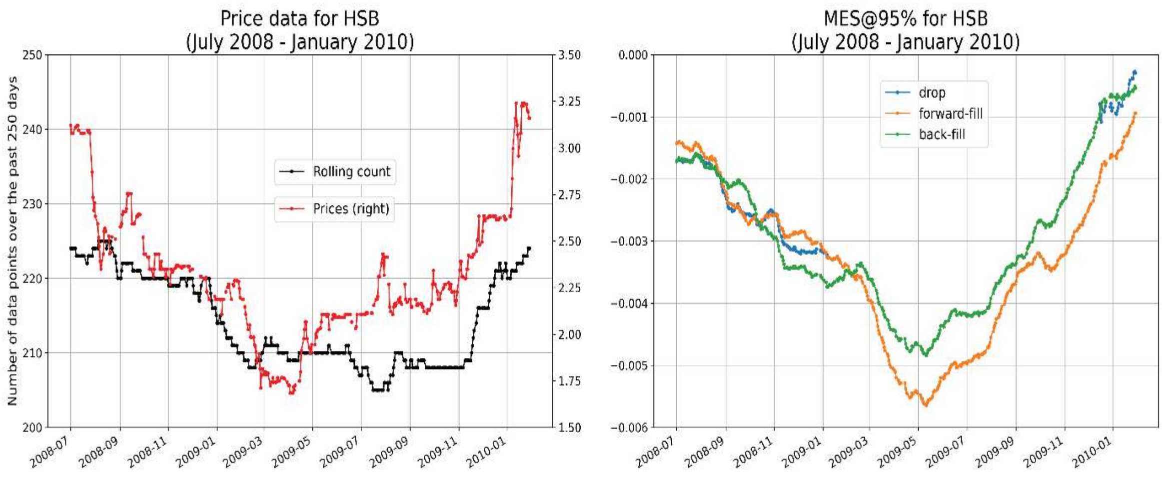

Figure 3. Impact of dropping, forward-filling, or back-filling missing datapoints on the MES at a 95% confidence level. Left subplot: Prices (red) and rolling count of available points (black). Right subplot: MES computed by dropping (blue), forward-filling (orange) or back-filling (green) missing data

Rysunek 3. Wpływ usuwania, uzupełniania do przodu, lub uzupełniania wstecznego brakujących danych dla MES przy poziomie ufności 95%. Po lewej: Cena (czerwony) i liczba dostępnych punktów danych (czarny). Na prawo: MES obliczone poprzez usunięcie (niebieski), uzupełnienie do przodu (pomarańczowy), lub uzupełnienie wsteczne (zielony) brakujących danych

Source: chosen bank HSB (HSBC Bank Malta p.l.c.)

Źródło: na przykładzie banku HSB (HSBC Bank Malta p.l.c.)

indicate the perceived exposure of the bank to the market fall. The three curves are quite different in this period, with the MES computed using the back-fill option (green curve) indicating the largest risk. However, after January 2009, the MES with the forward-fill option (orange curve) indicates the largest risk. The MES curve computed with the drop option (blue curve) typically follows one or the other. The interpretation is as follows. Since the back-fill option (green) has information about the price drop in the past, the perceived risk is higher before (and during) the fall. Similarly, once the price starts picking up, the information is available before, and the risk drops. With the forward-fill option (orange), there is a delay in the information availability due to the smaller perceived variance of the returns, and so the curve lags. The MES computed with the drop option (blue) reacts to the market prices as they are available, and so oscillates between the two.

If we ignore the back-fill option on the argument of causality, we might be tempted to conclude that the drop-option is the better option over the forward-fill option, but that is not always clear, and the better option depends upon the use case. If we are estimating a market index, then using the last available price might be the best option, since the market index reflects the amount of wealth created. However, if we are trying to estimate the statistics of the price returns, then dropping too many datapoints might lead to statistics that vary more, whereas forward-filling prices leads to standard deviation estimates that are smaller. Depending on the model requirements, the data needs to be processed correctly at this stage.

Model Construction Related Issues

During the abstract model construction phase, just before implementation, practitioners are exposed to many sources of model risk. The construction stage that we discuss in this section refers to the development of the algorithm (dealing with abstract aspects as opposed to the model implementation dealing with practical aspects) as part of the modeling process. This stage can equivalently be thought of as the model conception stage, where decisions regarding the data usage, algorithmic steps and expected model outcomes are made, without regard to the practicalities of software and hardware choices.

During the abstract development of the models, practitioners might inadvertently incorporate some biases within the models. These biases might be within the data itself (especially pertinent for machine learning algorithms, where the bias might be hidden in the training set), or in the algorithmic steps. For example, Google's Sentiment Analyzer gave negative or positive sentiment values depending on the sociocultural information mentioned in the input [Thomson 2017], which can be considered offensive and discriminatory against certain groups. These can have huge consequences when the models exposed to such biases are used to help decision-making.

Another major source of model risk during this phase lies in the assumptions made by the practitioners. These assumptions might be well accepted, such as the efficiency of markets, (all information is incorporated in market prices), or less universally accepted, such as static seasonality assumptions (ignoring the dynamic or shifting nature of the cycles) related to the distribution choice for an asset's returns. Depending on how appropriate the assumptions are in the real world, the model risk can be considered high or low. For example, during periods of high volatility, financial markets tend to become correlated [Junior & Franca, 2012]. A model that assumes the relative independence of the markets may not accurately capture the dynamics during stressful periods, and thus develop a higher model risk. Approximations might fail for mathematical limitations as well, where certain phenomena are simply not captured.

Model inapplicability, i.e., the application of a model to a problem out of its development context, is yet another source of model risk for practitioners. For example, the Black-Scholes-Merton option pricing model was developed for European-style options (the right to exercise on the settlement date), and not for American-style options (the right to exercise at any time before the settlement date). If we apply the model on American-style options, then the model might fail to describe certain options exercised earlier.

Uncertainty in the model parameters, sometimes referred to as estimation risk [Klein and Bawa 1976, Lewellen and Shanken 2000], also contributes to model risk. Quantifying this risk typically involves quantifying the uncertainty in the model's outputs due to the uncertainty in the model's parameters. Parameter estimations can come from completely different approaches, such as using bootstrapping techniques [Christoffersen and Gonçalves 2004], or by building different variants of the models [Danielsson, et al., 201]. For example, consider a model that uses a historical demand of ATM cash to predict the minimum required cash levels to be maintained in future. A change in a parameter, such as the historical window length, causes variation in the estimation of the distribution of the cash demands, and consequently, the model outputs might change. Managing this source of model risk requires stringent model development practices and documentation of every choice that enters the model. Practitioners need to be able to envisage and find ways to circumvent biases in the datasets and the model development process. At the company level, this would typically involve setting up a model risk management team, whose sole responsibility is to ensure proper model development practices. Model risk management is a relatively new field, with much active research still being done [Garro 2020]. As financial institutions become more dependent on computational models, and use larger datasets, the need for model risk management becomes more important.

To highlight how important model risk management is at this stage, we study two examples – (1) the choice of price returns, and (2) the choice of a parameter value in the estimation of ΔCoVaR. In the first example, we study the impact of logarithmic or relative returns on the MES at a 95% confidence level for three banks. In the second example, we look at the ε parameter used in our implementation of estimating the ΔCoVaR at the 95% confidence level for two banks. This parameter controls the number of MC iterations around a required quantile (VaR or median) of the bank, over which the VaR of the market is estimated.

i. Choice of price returns

Many works in finance typically use logarithmic (natural base) returns and relative returns. Below are their definitions, with $p(t)$ the price, $\eta(t)$ the (natural) logarithmic return, and $r_r(t)$ the relative return at time t:

$$r_r(t) = \frac{p(t) - p(t-1)}{p(t-1)} \Leftrightarrow p(t) = p(t-1)\left[1 + r_r(t)\right] \tag{8}$$

$$r_l(t) = \log\left[\frac{p(t)}{p(t-1)}\right] \Leftrightarrow p(t) = p(t-1)\exp\left[r_l(t)\right] \tag{9}$$

$$r_r(t) = \exp\left[r_l(t)\right] - 1 \Leftrightarrow r_l(t)\log\left[1 + r_r(t)\right] \tag{10}$$

If we assume that prices are non-negative, with some exceptions [Fernandez-Perez et al. 2023], relative returns are always greater than or equal to –1. Due to their very nature, logarithmic returns will satisfy this requirement, since by Eq. (10):

$$\lim_{r_l(t)\to-\infty} r_r(t) = -1 \tag{11}$$

Thus, by using logarithmic returns, 'real' prices are effectively lower bound at 0. Additionally, we have:

$$\log\left[1 + x\right] \approx x \Rightarrow r_l(t) \approx r_r(t) \tag{12}$$

Thus, relative returns are the first order approximation of the logarithmic returns under the Taylor's expansion, and so using logarithmic returns appears to make good mathematical sense for negative price changes. However, logarithmic returns do not always behave well with respect to positive price changes. Consider the following log-returns:

(–3, –2, –1, 0, 1, 2, 3)

By Eq. (10), these translate to the following relative returns:

(–0.9502, –0.8647, –0.6321, 0, 1.7183, 6.3891, 19.0855)

This implies that if we use symmetric distributions for logarithmic returns with a given probability of losing 95%, then we end up with the same probability of gaining 1909%. These effects are naturally more pronounced in heavy-tailed distributions like the Student's-t distribution. Additionally, relative returns are more intuitive.

In Figure 4, we compare the MES at the 95% confidence levels during the period from January 2007 to December 2008 for three banks: Bank of Valletta plc, BPER Banca S.p.A. and Banca Popolare di Sondrio, Società Cooperativa per Azioni. The MC algorithm under the Gaussian approximation was used with either logarithmic returns (blue) or relative returns (orange). The curves are plot in relative returns units for comparison.

We observe a small non-zero impact on the MES curves due to the choice of type of price returns. For some points, the risk perceived by logarithmic returns is higher (BPSO, between July 2007 and July 2008), whereas at some other points, the risk perceived by relative returns is higher (BPE, April 2007 to July 2007).

From these curves, it is difficult to say which type of returns have a higher perceived risk. While it makes sense to use logarithmic returns for SRMs (since SRMs typically focus on negative price changes), there may be situations when relative returns are more appropriate (say if we are interested in looking at positive price changes only). Model developers need to analyze the effects of their choice depending on their use cases.

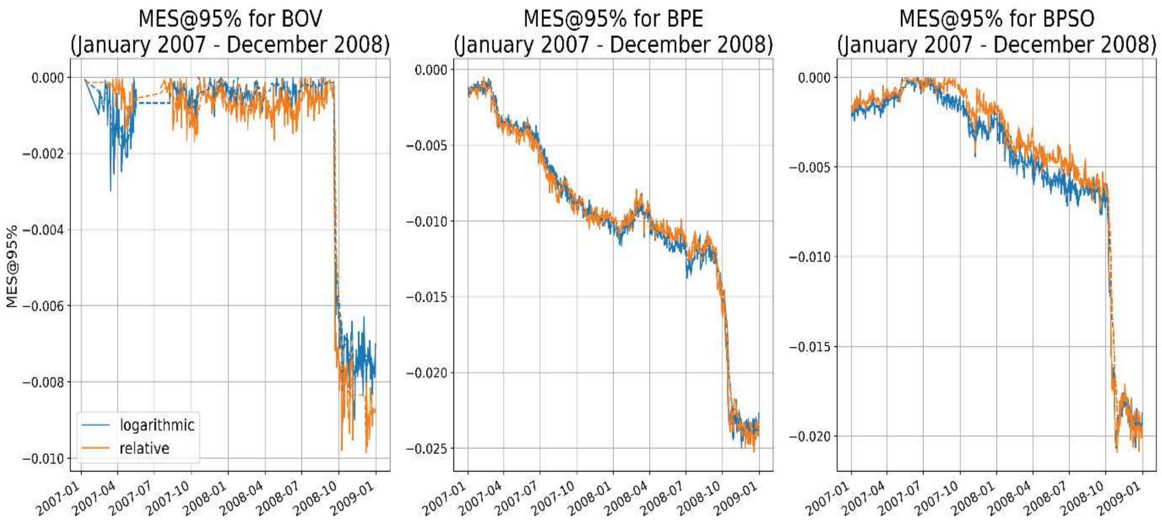

Figure 4. Impact of choice of type of price returns (logarithmic vs. relative) on the MES at 95% confidence levels for three banks. Chosen banks: BOV (Bank of Valletta plc), BPE (BPER Banca S.p.A.), BPSO (Banca Popolare di Sondrio, and Società Cooperativa per Azioni)

Rysunek 4. Wpływ wyboru rodzaju stopy zwrotu (logarytmicznej lub względnej) na MES przy poziomie ufności 95% dla trzech banków. Wybrane banki: BOV (Bank of Valletta plc), BPE (BPER Banca S.p.A.), BPSO (Banca Popolare di Sondrio, Società Cooperativa per Azioni)

Source: chosen banks: BOV (Bank of Valletta plc), BPE (BPER Banca S.p.A.), and BPSO (Banca Popolare di Sondrio, Società Cooperativa per Azioni).

Źródło: na przykładzie wybranych banków : BOV (Bank of Valletta plc), BPE (BPER Banca S.p.A.), BPSO (Banca Popolare di Sondrio, Società Cooperativa per Azioni).

ii. Choice of ε in ΔCoVaR

This second example considers the model risk due to the estimation risk of a parameter used in our implementation of ΔCoVaR (Section 2.D.ii). The parameter is ε, which defines the range around a desired quantile of a bank's simulated price returns, over which the market VaR (CoVaR) is estimated. If the parameter value is larger, a greater number of MC iterations are selected for estimating the CoVaR, bringing a higher stability (lower uncertainty) to the value. However, it also implies iterations that are not necessarily close in value to that of the desired quantile have been selected.

Figure 5 contains a visual representation of the selection process to better understand the algorithm. The blue line is the cumulative distribution function of the bank, and the blue diamond represents the VaR at the desired confidence level set to 95% (black dashed-dotted line at 5%) for this illustration. Three values of ε are studied here: 0.0025 (orange straight), 0.005 (green dashed) and 0.01 (red dotted). The ranges thus are 0.005 (100 iterations out of 20,000), 0.01 (200 iterations out of 20,000) and 0.02 (400 iterations out of 20,000) around the desired quantile. The CoVaR is then computed as the VaR of the market returns over iterations where the bank's returns are between the selected price returns. Note that ε ensures a symmetric range along the probability distribution (y) axis but does not guarantee symmetry along the price returns (x) axis. A larger (smaller) range implies that more (less) iterations are selected increasing (decreasing) the stability of the final CoVaR, but the required range of the bank's returns is much larger (smaller)

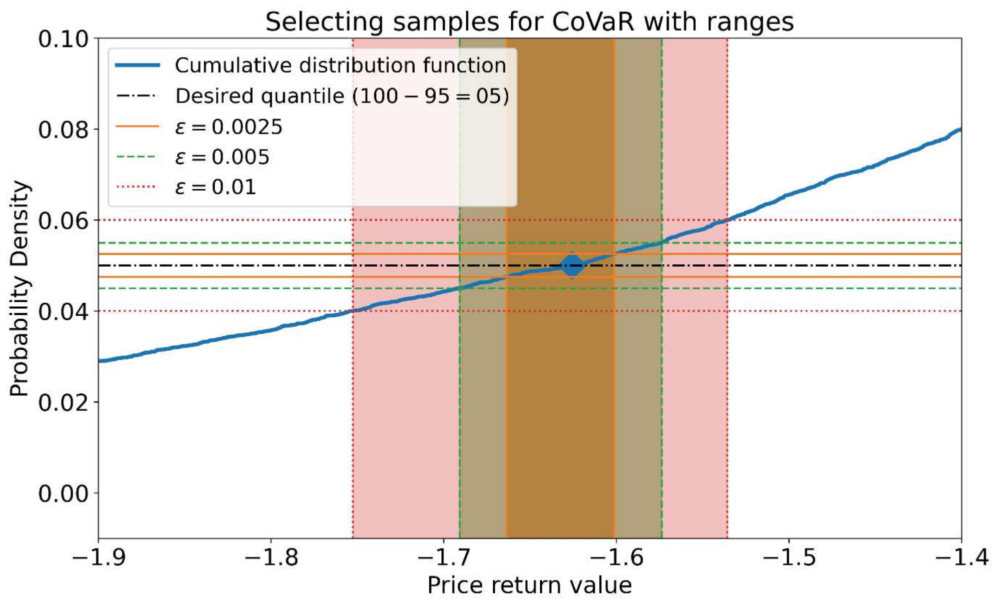

Figure 5. Visual representation of how samples are chosen for CoVaR. Given a distribution function (blue) of a bank, and a desired quantile (black dashed-dotted line), we select samples that lie between the desired quantile plus ε and the desired quantile minus ε. MC iterations where the bank's price returns lie between these limits (x-axis) are selected. The CoVaR is estimated as the average of the market returns of these selected iterations.

Rysunek 5. Wizualna reprezentacja sposobu wyboru próbek dla CoVaR. Biorąc pod uwagę funkcję rozkładu (niebieski) i pożądany kwantyl (czarna linia przerywana), wybieramy próbki, które leżą pomiędzy pożądanym kwantylem plus ε i pożądanym kwantylem minus ε. Wybierane są iteracje MC, w których zwroty cenowe banku mieszczą się w tych przedziałach (oś x). CoVaR szacuje się jako średnią zwrotów rynkowych z wybranych iteracji.

Source: chosen banks: BOV (Bank of Valletta plc), BPE (BPER Banca S.p.A.), and BPSO (Banca Popolare di Sondrio, Società Cooperativa per Azioni).

Źródło: na przykładzie wybranych banków: BOV (Bank of Valletta plc), BPE (BPER Banca S.p.A.), BPSO (Banca Popolare di Sondrio, Società Cooperativa per Azioni).

and contains iterations much further away from (much nearer to) the required quantile value, represented as a blue diamond.

Thus, the value of ε impacts ΔCoVaR estimates in two ways: numerical stability (due to the number of iterations chosen), and relevance of the selected points. In Figure 6, we highlight the numerical stability aspect of the choice. To do so, we first computed the ΔCoVaR using the three ε values for two banks: BNP Paribas S.A. (BNP) and AXA Bank Belgium SA (CS). Then a five-day exponentially weighted moving average was computed to obtain a 'trend.' Deviations from this trend are provided in the figure. As the value of ε becomes smaller (larger), fewer (more) iterations are selected, decreasing (increasing) the numerical stability, and causing larger (smaller) fluctuations from the trend. The green histogram ($\varepsilon = 0.01$) has the narrowest distribution, representing the smallest fluctuations (highest stability) with respect to the others, whereas the blue histogram ($\varepsilon = 0.0025$) has the widest distribution, representing the largest fluctuations (lowest stability). The orange histogram ($\varepsilon = 0.05$) lies in between. As can be seen, choosing a particular value of a parameter has a certain model risk associated with it. The model risk highlighted here

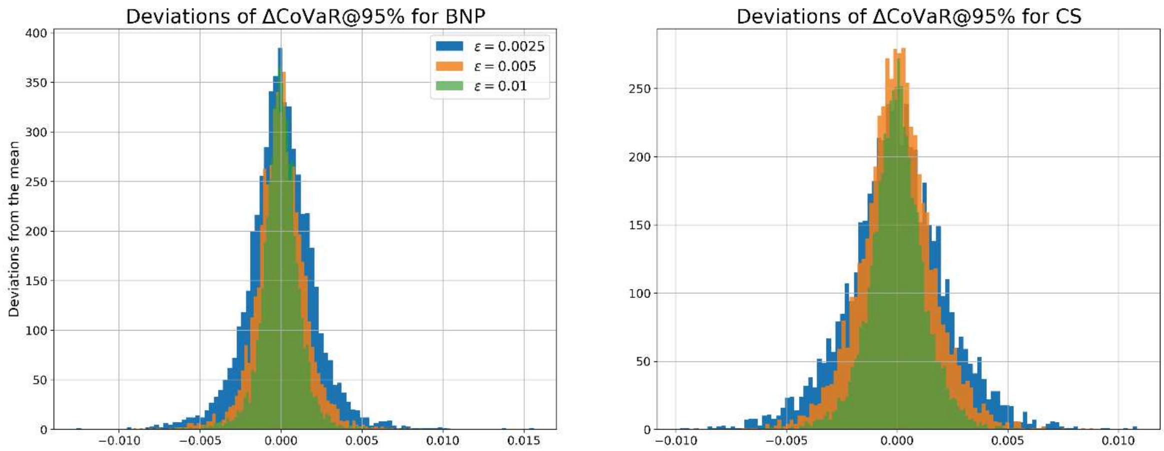

Figure 6. Histogram of the fluctuations of the ΔCoVaR at the 95% confidence levels around a 5-day exponentially weighted mean for different values of ε. Chosen banks: BNP (BNP Paribas S.A.) and CS (AXA Bank Belgium SA)

Rysunek 6. Histogram wahań ΔCoVaR przy 95% poziomie ufności wokół 5-dniowej wykładniczej średniej ważonej dla różnych wartości ε. Wybrane banki: BNP (BNP Paribas S.A.), CS (AXA Bank Belgium SA)

Source: chosen banks: BNP (BNP Paribas S.A.) and CS (AXA Bank Belgium SA).
Źródło: na przykładzie wybranych banków : BNP (BNP Paribas S.A.), CS (AXA Bank Belgium SA).

is linked to the model risk of the MC process, since one way to increase the stability is to increase (by a large amount) the number of MC iterations, while choosing a value of ε that is as small as possible.

Model Implementation Related Issues

The model risk sources considered here refer to the practical aspects of the modelling process. In the context of computational modelling, these sources usually appear in the form of computer bugs. Bugs can be introduced due to multiple factors: human error (bad programming style, misinterpretation of algorithms, etc.), faulty assumptions about the data and algorithm (data type mismatch, duplicate data, assumptions about edge cases), or software and hardware limitations (use of ‘alpha’ or ‘beta’ code, error mishandling, security restrictions, insufficient available memory, etc.).

Bug-free programs are extremely difficult to implement [Simmonds, 2018] and practitioners need to understand and anticipate the potential problems in code usage. As more ready-made computational and statistical packages are developed for users’ convenience, the model risk that users become exposed to increases. This can happen if the documentation of the implementation is incomplete/incomprehensible, or if the user does not spend time understanding the documentation. Black-box implementations, where the implementation is typically opaque to the user, tend to amplify these issues.

Computational complexity also increases model risk. Complexity refers to the ability of programmers to comprehend and debug a computer program. As a program becomes more complex, the chances of having bugs in the code increases. While the complexity of an implementation depends on the choice of language (high level language programs might be easier to debug than assembly level language programs), bugs are found everywhere, even in common spreadsheet software. For example, in 2012, JP Morgan, a prominent investment bank, lost over 6 billion dollars due to a small bug in their Excel implementation [Pollack 2013, EuSpRIG 2013].

During this stage, practitioners make implementation choices (as opposed to decisions in methodology) based on certain preliminary tests. These decisions are made to make the problem tractable, deal with bugs, have desirable outcomes in edge cases, or limit improbable scenarios (realistic scenarios). When certain choices are inappropriate in situations not considered in the tests, the model risk increases.

In addition, computer programs have hardware limitations, which, if not managed appropriately, can increase model risk. For example, memory issues might occur while performing large computations on a computer, which might lead to faulty results. Computers work in binary, requiring floating-point representations [Goldberg 1991] which might cause results to vary slightly (practically undetectable) for small runs of the program, but aggregate over larger runs.

Managing risk at this stage requires good computer engineering practices, such as Cleanroom engineering [Cobb and Mills 1990]. Cleanroom engineering develops the software under statistical quality control by: (a) specifying statistical usage, (b) defining an incremental pipeline for software construction that permits statistical testing, and (c) separating development and testing (only testers compile and execute the software being developed). While this setup might seem too extreme for banks, mitigating the risk associated with computer bugs might pay out the costs of implementing a similar setup.

We study two examples to underline the importance of model risk management at this stage – (1) the choice of the minimum value of the tailedness-parameter in the Student's-t distribution, and (2) the number of MC iterations. The first example looks at the MES at a 95% confidence level for three banks when the $v_{\min}$ parameter is altered. The parameter controls how heavy the Student's-t distribution is allowed to be, and thus has an impact on the tail events, which causes the MES estimates to vary. The second example analyzes the impact of the number of MC iterations on the variation and computational effort of the MES at a 95% confidence level for ten banks.

i. Choice of $v_{\min}$ in Student's-*t* distribution

While there are different distributions available to describe the observed price returns [(McDonald, 1996], as a starting point, practitioners use the normal distribution, since it is symmetric and not heavy- or light-tailed. To introduce some heavy-tailed behavior, practitioners sometimes use the Student's-*t* distribution, which introduces one extra parameter v (called the degrees of freedom parameter) that controls the heaviness of the tails. As v approaches infinity, the distribution approaches the normal distribution: the smaller the v, the heavier the tails.

In our implementation of the SRMs (Section 2.D.ii), this parameter was estimated on historical data, and thus represents the observed heaviness of the tails. Initial tests indicated that when relatively small values of v were used in the MC process to generate price returns, we obtained tail scenarios that were unrealistic. On further analysis, we found that since the Student's-t distribution is symmetric, it could generate price returns that were extremely large or small. For practical purposes, we had to set $\nu_{\min} = 5$. This implementation choice was done on a relatively small set of tests, and here we demonstrate the choice by reducing and increasing $\nu_{\min}$ by a factor of 2. We also include the simulation with the normal distribution, i.e., $\nu = \nu_{\min} = \nu_{\max} = \infty$.

In Figure 7, we look at how the MES at the 95% confidence levels changes across four values of $\nu_{\min}$: 2.5 (reduced by a factor of 2), 5 (choice), 10 (increased by a factor of 2), and ∞ (normal distribution). The upper subplots look at the MES curves of three banks: AIB Group plc, Eurobank Ergasias Services and Holdings S.A. and Sberbank Europe AG.

The bottom subplots show the ν curves for these banks so that we can understand the heavy-tailed nature (smaller values imply heavier tails). Since we expect the Student's-t distribution to be heavy-tailed, the risk (as seen by MES here) is expected (and observed) to be larger with this distribution than when the normal distribution (red curve,

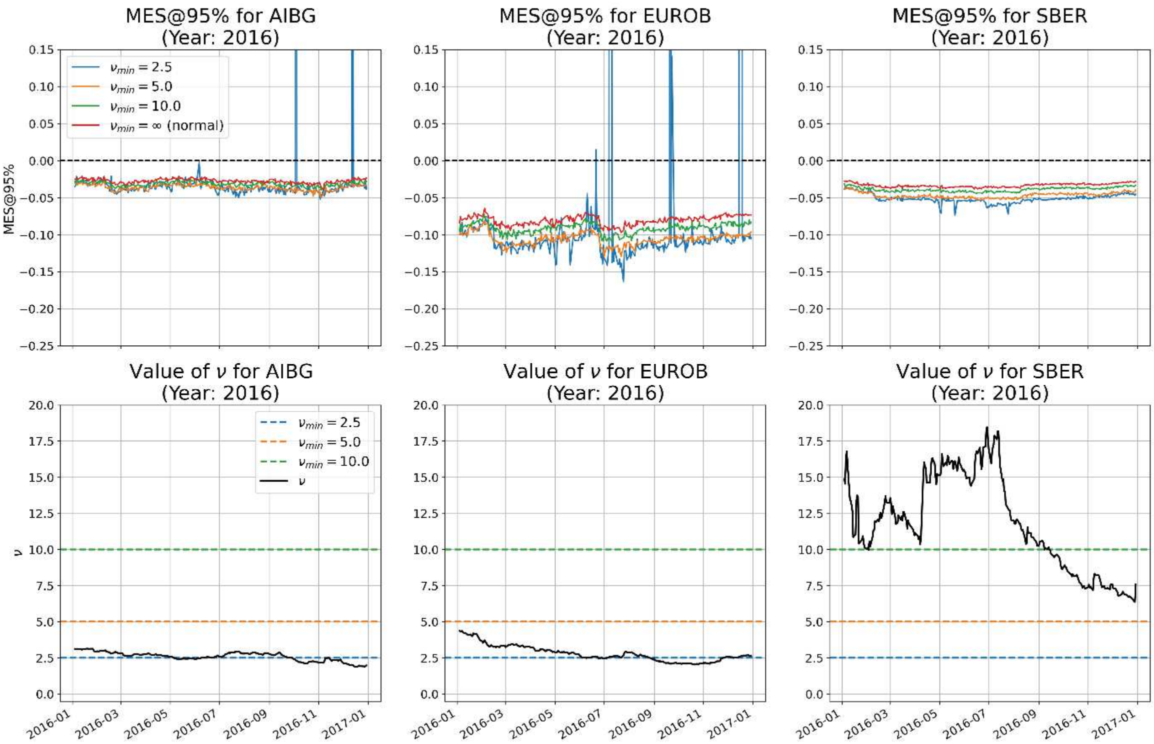

Figure 7. Impact of the choice of $\nu_{\min}$ on the MES at the 95% confidence levels (top subplots) for the year 2016. No filtering of minimum risk levels for the MES was done to highlight the sometimes non-physical heavy-tailedness nature of the Student's-t distribution. Provided in the three bottom subplots are the ν values of the banks.

Rysunek 7. Wpływ wyboru $\nu_{\min}$ na MES przy 95% poziomie ufności (górne wykresy) dla roku 2016. Nie przeprowadzono żadnego filtrowania minimalnych poziomów ryzyka dla MES, aby podkreślić czasami nierealistyczne właściwości rozkładu t-studenta. Na dolnych wykresach przedstawiono wartości ν banków.

Source: chosen banks: AIBG (AIB Group plc), EUROB (Eurobank Ergasias Services and Holdings S.A.), and SBER (Sberbank Europe AG).

Źródło: na przykładzie wybranych banków: AIBG (AIB Group plc), EUROB (Eurobank Ergasias Services and Holdings S.A.), SBER (Sberbank Europe AG).

$\nu_{min} = \nu_{min} = \infty$) is used. As ν_{min} is set to smaller values, the tails become heavier, and we expect and observe the MES curves to move down (losses and thus risk metrics are represented as negative price returns). Indeed, the green curve ($\nu_{min} = 10$) is above the yellow curve ($\nu_{min} = 5$), which is above the blue curve ($\nu_{min} = 2.5$). However, this is true only as long as the estimated ν is below the ν_{min}. In the case of Sberbank Europe AG (upper and lower rightmost subplots), the ν parameter is always above 5, and so the yellow and blue curves are remarkably similar.

Nonetheless, there are some jumps in the blue curve that make the choice of $\nu_{min} = 2.5$ questionable. In the case of the first two banks (AIB Group plc and Eurobank Ergasias Services and Holdings S.A.), we see exceptionally large jumps that make the MES levels positive. While it is possible for MES to take positive values if the bank and market are not positively correlated, the fact that these jumps happen only for a few instances in an otherwise 'properly behaving curve' indicate that these are probably artefacts of the implementation. Indeed, when $\nu_{min} = 5$, these jumps are not visible, and hence we chose the value of 5 as the minimum value. The jumps are present on the negative side as well, albeit much smaller since we used logarithmic returns (see the first example in the previous section).

This example demonstrates how programmers might make implementation choices. The best way to mitigate the risk with these choices is to perform tests on a relatively large set of examples to understand the impact of these choices. In this exercise, we were 'lucky' since the jumps are visible, but there might be cases where the impact of computational artefacts is not visible in any of the tests, and practitioners need to be able to debug bugs that appear when the software is released and applied in real-world scenarios.

ii. Impact of number of MC iterations

The second example looks at another implementation decision made due to practical constraints. Specifically, we study the choice of the number of MC iterations, which impacts: (a) the fluctuations of the MES at a 95% confidence level, and (b) the computational time. In all the other examples provided in this work, 20,000 MC iterations were used to estimate the MES or the ΔCoVaR. Preliminary tests guided this choice to optimize the tradeoff between the MC uncertainty and computational time.

In Figure 8, we show how practitioners might make the decision. The MES at the 95% confidence level was computed for ten banks (names provided in the caption of the figure), and for five values for the number of MC iterations (5,000, 10,000, 20,000, 40,000, and 80,000). In addition, we estimated the computational time in seconds for these ten banks. We expect the computational time to increase with the number of iterations. To estimate the fluctuations of the MES, we simply looked at deviations from a 5-day exponentially moving average. These deviations from the trend are expected to be higher if the number of iterations is small. The subplots show: (left) the computational effort measured in seconds as a function of the number of iterations along with a quadratic fit, and (right) the mean absolute deviation from the trend as a function of the number of iterations for the ten chosen banks.

As expected, the computational effort increases rapidly with the number of iterations, with an appropriate quadratic fit. If the quadratic term begins to dominate, then doubling the number of iterations would quadruple the computational time. We also confirm our

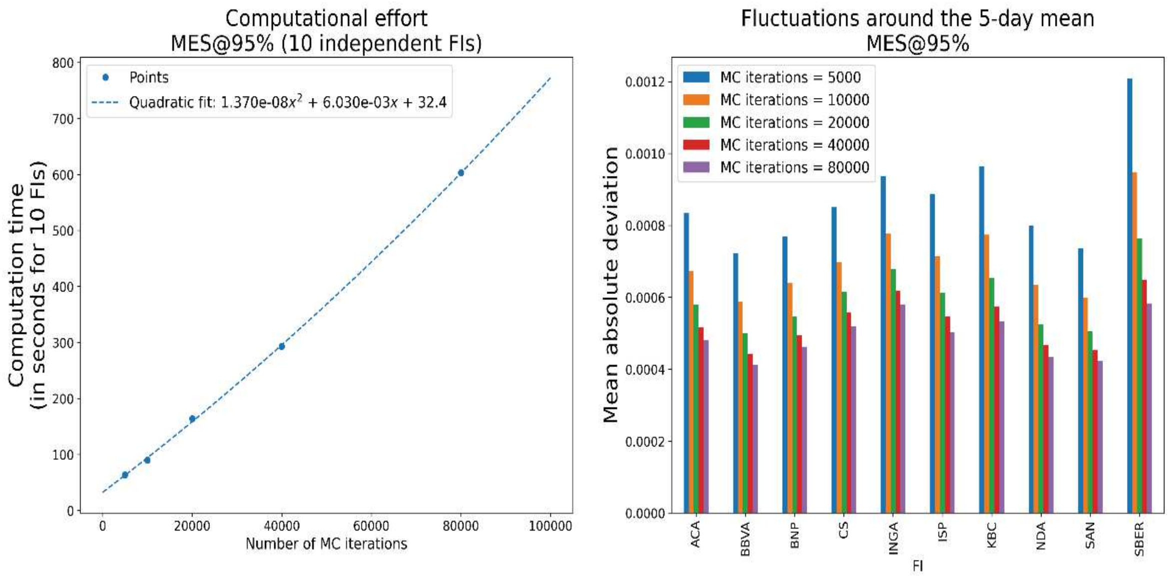

Figure 8. Left: Computational effort (in seconds for 10 banks) as a function of the number of MC iterations. Right: Fluctuations of the MES at the 95% confidence levels as a function of the number of MC iterations for the ten chosen banks around the 5-day exponentially weighted mean.

Rysunek 8. Po lewej: Nakład obliczeniowy (w sekundach, dla 10 banków) jako funkcja liczby iteracji MC. Po prawej: Wahania MES (przy 95% poziomie ufności jako funkcja liczby iteracji MC dla dziesięciu wybranych banków) wokół 5-dniowej średniej ważonej wykładniczo.

Source: chosen banks: ACA (Crédit Agricole S.A.), BBVA (Banco Bilbao Vizcaya Argentaria, S.A.), BNP (BNP Paribas S.A.), CS (AXA Bank Belgium SA), INGA (ING Groep N.V.), ISP (Intesa Sanpaolo S.p.A.), KBC (KBC Group NV), NDA (Nordea Bank Abp), SAN (Banco Santander, S.A.), and SBER (Sberbank Europe AG).

Źródło: na przykładzie wybranych banków: ACA (Crédit Agricole S.A.), BBVA (Banco Bilbao Vizcaya Argentaria, S.A.), BNP (BNP Paribas S.A.), CS (AXA Bank Belgium SA), INGA (ING Groep N.V.), ISP (Intesa Sanpaolo S.p.A.), KBC (KBC Group NV), NDA (Nordea Bank Abp), SAN (Banco Santander, S.A.), SBER (Sberbank Europe AG).

hypothesis that the deviations reduce as the number of iterations increases. However, the added benefit of minimizing the fluctuations is reduced as well, since as the number of iterations increases, the reduction of the deviations is slower than a linear growth. Thus, practitioners are required to make a choice that is a tradeoff between acceptable MC uncertainty and acceptable computational time.

This choice is one dictated by available hardware. If a faster computer were available, practitioners might be able to push for more MC iterations, whereas practitioners with slower computers would be forced to further reduce the number of iterations. There is an inherent model risk of the MC process, and this risk propagates through to the SRM estimates due to the choice and available hardware. General approaches (more of a desire) to reduce this particular risk include usage of:

- optimized models, so that more computational time is available for the actual iterations than for the overhead;
- appropriate models that focus on the variable of interest than anything else;
- low-variance models that target the model risk of MC process itself. However much the risk is reduced, implementation decisions will always carry some risk with them.

Model Interpretation Related Issues

This source of model risk refers to the risk associated with the usage of model outcomes for decision-making. Accordingly, the term ‘model interpretation’ here refers to the analysis and application of the model outcomes. Since models are simplified explanations (approximations) of real-world phenomena, they are limited in scope and applicability. Practitioners use models to understand a particular problem, but the models do not capture the entire picture. Misinterpretation thus can happen due to practitioners using models as ‘answer machines’ [Wagner et al. 2010], or when the results are taken out of context. In the context of SRMs, understanding and managing this risk is quite important. Typically, regulatory bodies might use SRMs with the intention of monitoring the overall systemic stress and controlling the contributions of individual banks. If the overall systemic stress is underestimated, then this can lead to an unobserved buildup of stress, which might culminate in an eventual collapse of the financial system and adversely impact the larger economy. An overestimation of the overall systemic stress can also adversely affect the financial system since regulators might stifle economic growth by restricting perceived risky ventures.

While aiming to control the contributions of individual banks to overall systemic stress, the model risk of SRMs can lead to suboptimal regulatory action.

Typically, systemically important banks are required to allocate a pre-determined part of their capital to offset systemic risk. For example, consider the bucketing scheme used by the Financial Stability Board (FSB) in conjunction with the Basel Committee to determine the global systemically important banks (G-SIBs). Depending on the eventual bucket into which a bank is placed, it must have from 1% (for bucket 1) to 3.5% (for bucket 5, which is empty when the study is performed) of its Common Equity Tier 1 capital as a buffer for its systemic risk contribution [Bank For International Settlements 2018]. If a bank is wrongly allocated to a lower bucket (perceived less risky by regulators), then (a) it will be allowed (unjustly) to take more risky investments without a buffer to reduce the systemic stress, and (b) in the event of a systemic shock, it may not have enough capital to deal with the shock. On the other hand, if the bank is wrongly allocated to a higher bucket (perceived riskier), then the bank will be blocked from taking on ventures (the result might be that economic growth is stifled) with reduced available liquidity within the financial network.

Even though the two SRMs studied here are linked to the TCTF aspect of SR, they approach the problem slightly differently, and thus may not be directly compatible if used independently. For example, if a bank has the highest contribution according to ΔCoVaR, it is not guaranteed that the same bank will have the highest contribution according to MES. This problem is further amplified for regulatory bodies, who must analyze and weigh various aspects of SR (e.g., too-big-to-fail or TBTF, TCTF, impact on non-financial sectors). Given the types of SRMs in the finance literature, and their associated model risk, identification of systemically important banks and the management of overall systemic stress is a daunting task for regulators. When using model outcomes, human intervention might be needed, and such an option should be provided in any model-based framework. Indeed, in the bucketing scheme by the FSB, the G-SIBs may be placed in buckets other than what the model dictates, referred to as exercise of supervisory judgement in their framework.

We provide one example below that attempts to highlight the risks of model misinterpretation within the context of SR. Specifically, we show that model outcomes using one SRM can lead to incompatibilities with outcomes from another SRM.

In this example, we rank all banks based on three criteria:

– size (specifically, the market capitalization without free-float corrections),
– MES,
– ΔCoVaR.

The size-based ranking acts as an indicator of the TBTF aspect of SR, whereas the MES- and ΔCoVaR-based ranking function as indicators of the TCTF aspect of SR. MES quantifies the impact of the market on a bank, and ΔCoVaR quantifies the impact of a bank on the market. The size-based ranking is done using the market capitalization of the banks (without correcting for the free-float factor) and has minimal computational requirements (just a product of the market price and outstanding shares amount). The other rankings are achieved using the outputs of the MC algorithm described earlier under the Gaussian approximation for the returns.

In Figure 9, we show the comparison of the ranking across the three indicators for five banks (out of 47) during the year 2020, where these banks were always among the top five according to size. We immediately observe that the rankings are not compatible across the three indicators. For example, the bank Sberbank Europe AG has the first rank according to size but lies between the 35th and 40th ranks according to MES, and between the 10th and 40th ranks according to ΔCoVaR. Even the relative ranking among

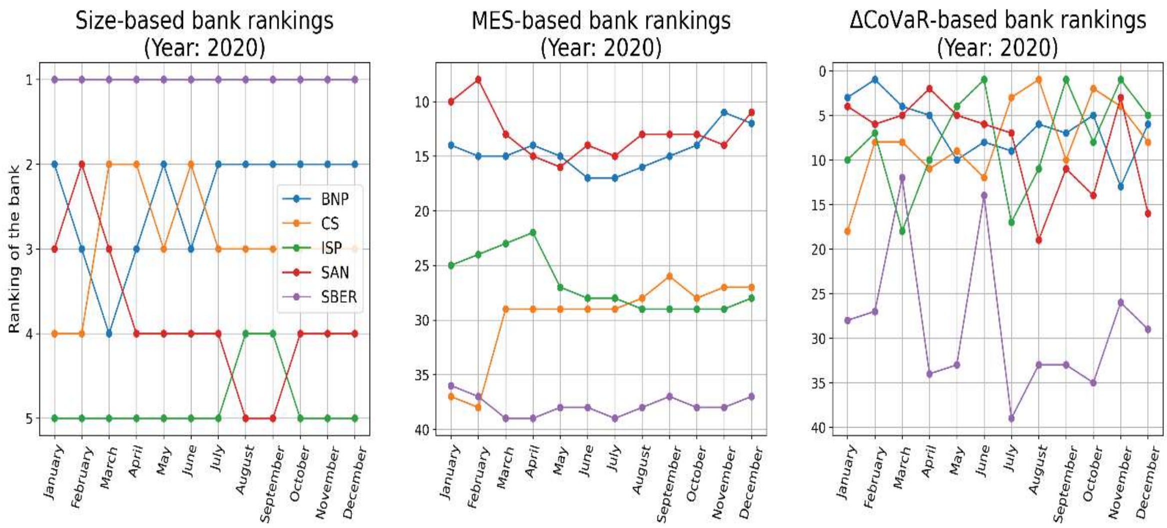

Figure 9. Ranking of five banks according to the size (market-capitalization), MES and ΔCoVaR.
Rysunek 9. Ranking pięciu banków według wielkości (kapitalizacji rynkowej), MES i ΔCoVaR.

Source: chosen banks are those with largest size in the year 2020: BNP (BNP Paribas S.A.), CS (AXA Bank Belgium SA), ISP (Intesa Sanpaolo S.p.A.), SAN (Banco Santander, S.A.), and SBER (Sberbank Europe AG).
Źródło: na przykładzie wybranych banków o największej wielkości w 2020 roku: BNP (BNP Paribas S.A.), CS (AXA Bank Belgium SA), ISP (Intesa Sanpaolo S.p.A.), SAN (Banco Santander, S.A.), SBER (Sberbank Europe AG).

these banks is not consistent. This has important implications in regulatory and banking contexts, since relying on one SRM is not enough to really determine the systemically significantly important banks. Indeed, one might require many indicators to fully understand the contribution of a bank to the overall SR. While the usage of more indicators leads to a potential increase in model risk, missing out potential SR contributions can have negative implications for the financial system.

One approach to deal with multiple indicators might be to combine class effects. For example, we might wish to average the MES and ΔCoVaR outcomes to become a net TCTF contribution. Multiple aspects of SR, such as the TBTF and TCTF contributions, can then be added together before the ranking or clustering is done. This type of approach is used to identify the G-SIBs, where five aspects of SR are considered with equal weights: size, cross-jurisdictional activity, interconnectedness, substitutability/financial institution infrastructure and complexity. Each of these aspects is a (weighted) sum of multiple indicators so that regulators can obtain an overall picture of a bank's contribution to SR. Despite this, the framework allows for the exercise of supervisory judgement (human intervention) so that banks can be manually placed into risk buckets as deemed necessary, since a sum of indicators is still a model and might be insufficient to explain SR contributions.

Discussion

The examples highlighted in Section 3 indicate that practitioners are exposed to model risk at every stage of the modelling process – from data collection to model deployment and interpretation. Before attempting to provide some general strategies to help mitigate model risk, we will discuss some consequences of model failure at the level of individual companies and regulatory bodies.

Model failures can be devastating for individual companies and have negative effects on the financial network. For example, the failure of Long-Term Capital Management (LTCM) is often attributed to its poor risk management in terms of model risk [Kolman, 1999, Kato et al. 2000, Powell 2023]. Some have suggested that simply increasing the amount of data for their VaR models from the past five years to the past eleven years would have captured earlier stock market falls, and perhaps have reduced the impact of the crash [Ferguson 2008]. The consequences of LTCM's fall involved a major bailout with many major banks to prevent a systemic contagion event [Lowenstein 2001]. It seems that the benefits of a proper risk management framework that includes models and the modelling process far outweigh the risks associated with a failure of the models.

In the context of SR regulation, regulatory model failures can impact the entire financial network and the larger economy. Regulators typically rely on models and require banks to allocate a part of their capital to offset 'known' individual and systemic risks. As banks reduce these risks, they are incentivized to reinvest their freed-up capital to take up risk elsewhere. If there is a sector (or asset class) that is unregulated or considered less risky, banks will probably gravitate towards it, leading to a build-up of risk there [Acharya et al. 2011]. Depending on how the models work, this build-up might pass under the regulatory radar, creating a vulnerable spot for a systemic failure. Furthermore, banks are tempted to 'game the rules,' sometimes referred to as 'regulatory arbitrage'.

Broadly speaking, banks would like to structure their activities in a way that minimizes regulatory constraints while allowing them to take as much risk as possible [Nouy 2017]. In a model-based regulatory framework, banks might 'optimize' their reporting so that they can take more risks than the regulations allowed [Behn, et al. 2022]. To avoid unintended consequences, regulatory bodies thus must require risk management frameworks to include models, and use manual intervention when models are lacking.

Given the consequences associated with model failure, it makes sense to envisage a risk management framework that actively includes model risk. Furthermore, risk is perceived differently by various professions [Harkins 2013], and so there is a need for harmonization and open communication between the data management teams, computer programmers, analysts, risk managers and decision makers. At a team level, strategies would need to be implemented that aim to control their individual contributions to model risk. Some general strategies that might help with model risk management are provided below:

1. Managing dataset issues:
   - proper documentation of collected data, including source, time, data variables and description of the data variables;
   - summary statistics for missing datapoints;
   - highlighting potential biases within the dataset documentation.
2. Managing data processing related issues:
   - proper documentation of the transformations of the data, especially expected data input and expected transformed output;
   - highlighting the purpose of the data processing, and potential use cases, preferably with examples;
   - open communication about the automation tools used, including specific choices in code and test cases on which the code was evaluated.
3. Managing model construction related issues:
   - proper definition of the abstract model using algorithms and flowcharts that help highlight how the data is used and manipulated to gain the required outputs;
   - employment of strict model development practices that envisage model failures due to data issues, software and hardware issues and the inapplicability of the model;
   - brainstorming with other experts with the aim of 'breaking' the model to find potential limitations that may have been missed by the modeler;
   - highlighting as many assumptions and limitations of the model as possible;
   - indicating the choices used in the model development along with justification and, if possible, the impact of alternate choices.
4. Managing model implementation related issues:
   - proper documentation of the code so that other programmers can understand and alter it if needed;
   - inclusion and documentation of as many error catching scenarios as possible;
   - intensive testing of the code on all conceivable data, software, and hardware issues;
   - highlighting all implementation choices that were made based on the testing phase;
   - careful deployment of the code with sufficient monitoring in production before the model is used in decision-making.

5. Managing model interpretation related issues:
   - avoidance of complete reliance on model outputs for decisions;
   - open debate of the usage and limitations of data and computational models in the decision-making process;
   - push for regular revisits of the models to understand model limitations and drive better model development;
   - inclusion of a human judgement option to handle model failure.

These strategies target specific model risk sources, and every risk management framework will need to adapt them, depending on their needs. For example, regulatory bodies might wish to emphasize the risk associated with model interpretation, since they would have to justify increasing (or decreasing) the capital requirements of individual banks. Banks, on the other hand, might focus on the risk associated with data management, model construction and implementation, to optimize their daily operations, from cash management to investing, while following the regulatory guidelines.

Large banks, on the other hand, might focus on the risk associated with model construction and implementation, so that their activities take no more risk than is deemed necessary (and permitted). Hedge funds might target the risk associated with data and model aspects so that they can discover market inefficiencies for arbitrage.

To manage model risk, practitioners need to recognize that it will always be present. Models are approximations to observed phenomena and cannot be treated as the absolute truth. In quoting Box again [Box 1979], there is no need to ask the question 'Is the model true?' if 'truth' is the 'whole truth' as the answer must be 'No.' The only question of interest is 'Is the model illuminating and useful?'

Conclusions

As practitioners continue building their reliance on computational models, the risk associated with model failure increases. Model risk in the context of SR has more negative consequences for the financial network and the larger economy. In this work, we provided a classification of the model risk sources and gave specific examples in each category to highlight how model risk might impact the estimation of SRMs. We additionally provided certain general strategies that might be employed at banks and regulatory bodies to mitigate the risks associated with these sources.

Given that model risk is always present, risk management teams must adapt their strategies to include modelling. This may be daunting due to the omnipresence of model risk – from the dataset collection to model usage. However, we argue that the benefits associated with controlling model risk outweigh the costs. As it is with models, there may be no one-size-fits-all framework to deal with model risk. Risk managers, data management teams, computer programmers, analysts and decision makers would have to work together to identify and control the sources of model risk. No matter what the scope of applicability of models is, practitioners will require transparent communication about models and their limitations.

Bibliography

Acharya V.V., Cooley T.F., Richardson M.P., Walter I. 2011: Market failures and regulatory failures: Lessons from past and present financial crises, ADBI Working Paper Series 264.

Acharya V.V., Pedersen L.H., Philippon T., Richardson M.P., 2010: Measuring Systemic Risk. AFA 2011 Denver Meetings Paper, University of Chicago, Booth School of Business, Chicago.

Adrian T., Brunnermeier M.K., 2008: CoVaR. Federal Reserve Bank of New York, Federal Reserve Bank of New York, New York.

Bachelier L., 1900: Theorie de la Speculation, Annales Scientifiques Ecole Normale Sperieure 3(17), 21–86.

Baker T., Vaidyanathan J., Ashley N., 2012: A Data-Driven Inventory Control Policy for Cash Logistics Operations: An Exploratory Case Study Application at a Financial Institution, Decision Sciences 44(1), 205.

Banulescu-Radu D., Hurlin C., Leymarie J., Scaillet O., 2020: Backtesting Marginal Expected Shortfall and Related Systemic Risk Measures, Management Science 67(9), 5301.

Bank For International Settlements, 2018: Basel Committee on Banking Supervision. Global systemically important banks: revised assessment methodology and the higher loss absorbency requirement, [electronic source] https://www.bis.org/bcbs/publ/d445.htm [accessed: 10.09.2024].

Behn M., Haselmann R., Vig V. (2022). The limits of model-based regulation. The Journal of Finance 77, 1635–1684.

Bianchi M.L., Sorrentino A.M., 2020: Measuring covar: An empirical comparison, Computational Economics 55, 511–528.

Black F., Scholes M., 1973: The pricing of options and corporate liabilities. Journal of political economy 81, 637–654.

Bloomberg L.P., 2021: Bloomberg Terminal, [electronic source] https://www.bloomberg.com/professional/solution/bloomberg-terminal/ [accessed: 29.08.2024].

Boucher C.M., Danielsson J., Kouontchou P.S., Maillet B.B., 2014: Risk models-at-risk, Journal of Banking & Finance 44, 72–92.

Box G.E., 1979: Robustness in the strategy of scientific model building, [in:] R.L. Launer, G.N. Wilkinson (eds), Robustness in statistics. Research Triangle, Elsevier, North Carolina, 201–236.

Brownlees C.T., Engle R., 2012: Volatility, correlation and tails for systemic risk measurement, SSRN, 1611229.

Bryant R., Cintas C., Wambugu I., Kinai A., Diriye A., Weldemariam K., 2019: Evaluation of bias in sensitive personal information used to train financial models, 2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP), IEEE, Ottawa, 1–55.

Castro C., Ferrari S., 2014: Measuring and testing for the systemically important financial institutions, Journal of Empirical Finance 25, 1–14.

Cedolin M., Orhan D., Genevois M., 2024: Statistical and Artificial Intelligence Based Forecasting Approaches for Cash Demand Problem of Automated Teller Machines, Journal of Engineering and Smart Systems 12(1), 21.

Cesa M., 2017: A brief history of quantitative finance. Probability, Uncertainty and Quantitative Risk, 2, 1–16.

Christoffersen P., Gonçalves S., 2004: Estimation Risk in Financial Risk Management, CIRANO, Montreal.

Cifuentes A., Hawas F.J., Chavarria E., 2014: A Numerical Simulation Approach to Study Systemic Risk in Banking Systems, SSRN, 2539020.

Cobb R.H., Mills H.D., 1990: Engineering software under statistical quality control. IEEE software 7, 45–54.

Crouhy M., Galai D., Mark R., 1998: Model risk, Journal of Financial Engineering 7, 267–288.

Danielsson J., James K.R., Valenzuela M., Zer I., 2016: Can we prove a bank guilty of creating systemic risk? A minority report, Journal of Money, Credit and Banking 48, 795–812.

Danielsson J., James K.R., Valenzuela M., Zer I., 2016: Model risk of risk models, Journal of Financial Stability 23, 79–91.

Derman E., 1996: Model Risk: What are the assumptions made in using models to value securities and what are the consequent risks? Risk-London-Risk Magazine Limited 9, 34–38.

Dicuonzo G., Galeone G., Zappimbulso E., Dell'Atti V., et al., 2019: Risk management 4.0: The role of big data analytics in the bank sector, International Journal of Economics and Financial Issues 9, 40–47.

DW, 2019: The worst year for stocks since financial crisis, [electronic source] https://www.dw.com/en/2018-the-worst-year-for-stocks-since-financial-crisis/a-46915652 [accessed: 24.08.2024].

European Central Bank, ECB, ESCB and the Eurosystem, [electronic source] https://www.ecb.europa.eu/ecb/orga/escb/html/index.en.html [accessed: 14.09.2024].

EuSpRIG, 2013: The European Spreadsheet Risk Interest Group, [electronic source] https://eusprig.org/research-info/horror-stories/ [accessed: 15.09.2024].

Ferguson N., 2008: The ascent of money: A financial history of the world, Penguin, London.

Fernandez-Perez A., Fuertes A.-M., Miffre J., 2023: The Negative Pricing of the May 2020 WTI Contract, The Energy Journal 44, 119–142.

Garro M., 2020: The evolution of model risk management processes, Journal of Risk Management in Financial Institutions 13, 16–23.

Gilbert E., Strugnell D., 2010: Does survivorship bias really matter? An empirical investigation into its effects on the mean reversion of share returns on the JSE (1984–2007), Investment Analysts Journal 39, 31–42.

Glasserman P., 2005: Measuring marginal risk contributions in credit portfolios, FDIC Center for Financial Research Working Paper, 5–12.

Glasserman P., Xu X., 2014: Robust risk measurement and model risk, Quantitative Finance 14, 29–58.

Glasserman P., Heidelberger P., Shahabuddin P., 2000: Efficient Monte Carlo methods for value-at-risk, IBM Thomas J. Watson Research Division, Yorktown Heights, NY.

Goldberg D., 1991: What every computer scientist should know about floating-point arithmetic, ACM computing surveys (CSUR) 23, 5–48.

Haldar A., Sethi N., 2021: The news effect of COVID-19 on global financial market volatility, Buletin Ekonomi Moneter dan Perbankan 24, 33–58.

Harkins M., 2013: The Misperception of Risk, [in:] Managing Risk and Information Security: Protect to Enable, Apress, Berekley, 15–25.

Hendricks D., 1996: Evaluation of value-at-risk models using historical data, Economic Policy Review 2, 40–49.

Holton G.A., 2003: Value-at-Risk Theory and Practice, Academic Press, San Diego.

Hurd T.R., 2016: Contagion!: Systemic Risk in Financial Networks, Springer, Cham.

Idier J., Lamé G., Mésonnier J.-S., 2013: How useful is the marginal expected shortfall for the measurement of systemic exposure? A practical assessment, ECB Working Paper, European Central Bank (ECB), Frankfurt a. M., 1546.

Irwin N., 2015: Why the Stock Market Is So Turbulent, New York Times, [electronic source] https://www.nytimes.com/2015/08/25/upshot/why-global-financial-markets-are-going-crazy.html [accessed: 10.08.2024].

Junior L.S., Franca I.D., 2012: Correlation of financial markets in times of crisis, Physica A: Statistical Mechanics and its Applications 391, 187–208.

Kato T., Yoshiba T. 2000: Model risk and its control, Institute for Monetary and Economic Studies, Bank of Japan.

Kilkenny M.F., Robinson K.M., 2018: Data quality: "Garbage in – garbage out", Health Information Management Journal 47, 103–105.

Klein R.W., Bawa V.S., 1976: The effect of estimation risk on optimal portfolio choice, Journal of Financial Economics 3, 215–231.

Koike T., Hofert M., 2020: Markov chain monte carlo methods for estimating systemic risk allocations, Risks 8, 6–12.

Kolman J., 1999: LTCM speaks, Derivatives Strategy 4, 12–17.

Lamas L., Gallejones C.F., Moral R.G., Cascales J.G., Calvo J., 2014: Model Risk Management: Quantitative and qualitative aspects. Management Solutions.

Lehar A., 2005: Measuring systemic risk: A risk management approach, Journal of Banking & Finance 29, 2577–2603.

Lewellen J., Shanken J., 2000: Estimation Risk, Market Efficiency, and the Predictability of Returns, NBER Working Papers 7699.

Li S., Chang A., Dave P., 2015: Stock market suffers worst one-day drop since 2008, Los Angeles Times, [electronic source], https://www.latimes.com/business/la-fi-0822-financial-markets-20150821-story.html [accessed: 20.09.2024].

Lowenstein R., 2001: When genius failed: The rise and fall of Long-Term Capital Management, Random House Publishing Group, New York.

Madadi N., Roudsari A.H., Wong K.Y., Galankashi M.R., 2013: Modeling and Simulation of a Bank Queuing System, 2013 Fifth International Conference on Computational Intelligence, Modelling and Simulation, Seoul, Korea (South).

McDonald J.B., 1996: 14 Probability distributions for financial models, Handbook of Statistics 14, 427–461.

Merton R.C., 1974: On the pricing of corporate debt: The risk structure of interest rates, The Journal of finance 29, 449–470.

Minderhoud K., 2006: Systemic risk in the Dutch financial sector, De Economist 154, 177–195.

Nievergelt Y., 2000: Rounding errors to knock your stocks off, Mathematics Magazine 3,1, 47–48.

Nouy D., 2017: Gaming the rules or ruling the game? How to deal with regulatory arbitrage. Shadow Banking, 53.

Pasieczna A.H., 2019: Monte Carlo Simulation Approach to Calculate Value at Risk: Application to WIG20 and MWIG40, Financial Sciences. Nauki o Finansach, Sciendo 24, 61–75.

Pasieczna A.H., 2021: Model Risk of VaR and ES Using Monte Carlo: Study on Financial Institutions from Paris and Frankfurt Stock Exchanges, [in:] Contemporary Trends and Challenges in Finance, Springer, Cham, 75–85.

Pollack L., 2013: A tempest in a spreadsheet, Financial Times, [electronic source] https://www.ft.com/content/80bc5046-6ed2-30ef-b31a-bcbeac55d8bb [accessed: 15.08.2024].

Powell R., 2023: Six lessons from the collapse of LTCM, Timeline, [electronic source] https://www.timeline.co/blog/six-lessons-from-the-collapse-of-ltcm [accessed: 24.08.2024].
QONTIGO, 2011: EURO STOXX Banks, [electronic source] https://qontigo.com/index/sx7e/ [accessed: 05.09.2024].
Savvides S., 1994: Risk analysis in investment appraisal, Project Appraisal 9, 3–18.
Sibbertsen P., Stahl G., Luedtke C., 2008: Measuring model risk, Journal of Risk Model Validation 2, 65–82.
Siller T., 2013: Measuring marginal risk contributions in credit portfolios, Quantitative Finance 13, 1915–1923.
Simmonds D.M., 2018: Complexity and the engineering of bug-free software, Proceedings of the International Conference on Frontiers in Education: Computer Science and Computer Engineering (FECS), 4–100.
Thomson A., 2017: Google's Sentiment Analyzer Thinks Being Gay Is Bad, Motherboard, Tech by Vice, [electronic source] https://www.vice.com/en/article/j5jmj8/google-artificial-intelligence-bias [accessed: 05.09.2024].
Wagner W., Fisher E., Pascual P., 2010: Misunderstanding models in environmental and public health regulation, NYU Envtl. LJ 18, 293.
Weatherall J., 2013: It is not the maths that causes the crisis, Financial Times, [electronic source] https://www.ft.com/content/30fc4ece-760b-11e2-9891-00144feabdc0 [accessed: 05.09.2024].
Zhong H., Hamilton M., 2023: Exploring gender and race biases in the NFT market, Finance Research Letters 53, 103651.

APPENDIX

Table 1. List of banks in the study
Tabela 1. Lista banków uwzględnionych w badaniu

Country	Bank Name	Grounds for significance
Austria	Addiko Bank AG	Significant cross-border activities
	BAWAG Group AG	Size (total assets EUR 30–50 bn)
	Erste Group Bank AG	Size (total assets EUR 150–300 bn)
	Raiffeisen Bank International AG	Size (total assets EUR 150–300 bn)
	Sberbank Europe AG	Significant cross-border activities
Belgium	AXA Bank Belgium SA	Article 6(5)(b) of Regulation (EU) No 1024/2013
	KBC Group NV	Size (total assets EUR 150-300 bn)
Bulgaria	DSK Bank AD	Among the three largest credit institutions in the Member State
Cyprus	Bank of Cyprus Holdings Public Limited Company	Total assets above 20% of GDP
	Hellenic Bank Public Company Limited	Total assets above 20% of GDP
Estonia	AS SEB Pank	Total assets above 20% of GDP
	Swedbank AS	Total assets above 20% of GDP
Finland	Nordea Bank Abp	Size (total assets EUR 500–1,000 bn)
France	BNP Paribas S.A.	Size (total assets above EUR 1,000 bn)
	Crédit Agricole	Size (total assets above EUR 1,000 bn)
	Société Générale	Size (total assets above EUR 1,000 bn)
Germany	Aareal Bank AG	Size (total assets EUR 30–50 bn)
	COMMERZBANK Aktiengesellschaft	Size (total assets EUR 300–500 bn)
	Deutsche Bank AG	Size (total assets above EUR 1,000 bn)
	Deutsche Pfandbriefbank AG	Size (total assets EUR 50–75 bn)
Greece	ALPHA SERVICES AND HOLDINGS S.A.	Size (total assets EUR 50–75 bn)
	Eurobank Ergasias Services and Holdings S.A.	Size (total assets EUR 50–75 bn)
	National Bank of Greece S.A.	Size (total assets EUR 50–75 bn)
	Piraeus Financial Holdings S.A.	Size (total assets EUR 50–75 bn)
Ireland	AIB Group plc	Size (total assets EUR 75–100 bn)
	Bank of Ireland Group plc	Size (total assets EUR 100–150 bn)

cont. table 1
cd. tab. 1

Country	Bank Name	Grounds for significance
Italy	BANCA MONTE DEI PASCHI DI SIENA S.p.A.	Size (total assets EUR 100–150 bn)
	BPER Banca S.p.A.	Size (total assets EUR 75–100 bn)
	Banca Carige S.p.A. – Cassa di Risparmio di Genova e Imperia	Article 6(5)(b) of Regulation (EU) No 1024/2013
	Banca Popolare di Sondrio, Società cooperativa per azioni	Size (total assets EUR 30–50 bn)
	Banco BPM S.p.A.	Size (total assets EUR 150–300 bn)
	Intesa Sanpaolo S.p.A.	Size (total assets EUR 500–1,000 bn)
	Mediobanca – Banca di Credito Finanziario S.p.A.	Size (total assets EUR 75–100 bn)
	UniCredit S.p.A.	Size (total assets EUR 500–1,000 bn)
Lithuania	Akcinë bendrovë Ðiauliř bankas	Among the three largest credit institutions in the Member State
Malta	Bank of Valletta plc	Total assets above 20% of GDP
	HSBC Bank Malta p.l.c.	Total assets above 20% of GDP
Portugal	Banco Comercial Português, SA	Size (total assets EUR 75–100 bn)
Slovenia	Nova Ljubljanska Banka d.d. Ljubljana	Total assets above 20% of GDP
Netherlands	ABN AMRO Bank N.V.	Size (total assets EUR 300–500 bn)
	ING Groep N.V.	Size (total assets EUR 500–1,000 bn)
Spain	Banco Bilbao Vizcaya Argentaria, S.A.	Size (total assets EUR 500–1,000 bn)
	Banco Santander, S.A.	Size (total assets above EUR 1,000 bn)
	Banco de Sabadell, S.A.	Size (total assets EUR 150–300 bn)
	Bankinter, S.A.	Size (total assets EUR 75–100 bn)
	CaixaBank, S.A.	Size (total assets EUR 300–500 bn)
	Unicaja Banco, S.A.	Size (total assets EUR 50–75 bn)

Source: bank data provided by the [European Central Bank].
Źródło: dane uzyskane z banku [European Central Bank].