EUROPEAN CENTRAL BANK

EUROSYSTEM

Working Paper Series

Andreas Beyer, Arndt-Gerrit Kund

Severity over quantity. Drivers of supervisory capital add-ons in internal ratings-based models

No 3268

Abstract

Banks use their internal models to estimate capital requirements in a risk-sensitive way, subject to a set of rules laid down in banking regulation. However, these models are not flawless as the usage of models suffers from imperfections, such as oversimplifications or wrong assumptions. As a result, risks may be underestimated. This is particularly troublesome, where models are used to assess risks to banks' solvency. In this paper we address an important gap in the literature with regard to such model risk. We find that a small set of high severity deficiencies in models is responsible for the majority of the counterfactual RWA burden imposed by supervisory capital add-ons. We trace the underlying non-compliances to a subset of CRR articles that mostly govern the handling of IRB-relevant data by banks. Our results help improve the supervision of IRB-banks by proposing more efficient use of scarce supervisory resources.

JEL Classification: G21, G28, G29

Keywords: Microprudential Regulation, Margin of Conservatism
Banking Supervision, Internal Models, Model Risk

Non-technical summary

Statistical models are widely used in the financial sector to analyse balance sheets, prepare accounting and reporting statements, but also to calculate capital requirements for banks.

While these models allow to quantify and manage certain risks, they can give rise to other, novel risks, at the same time. One example is model risk, that is the potential losses that an institution may incur due to a decision that was made on the erroneous output of an internal model. Such losses could arise from mistakes in the development and implementation of quantitative models, but also their application outside their intended scope.

Against this backdrop, numerous lines of defence are employed to prevent model risk to a reasonable extent. Internally, banks validate their models through the internal validation function, while externally auditors and supervisors scrutinize new and existent models alike.

Despite these efforts, model risk materialises yet continues to be under-reported. Even if this risk is reported, its true impact is likely to be concealed by the fact that its realisation rather impacts banks' profitability, and only in the utmost instances their viability. The conjunction of a higher prevalence of deficiencies in models on the one hand and their comparatively material impacts regarding risk-weighted assets (RWA) and hence solvency ratios on the other hand has so far not received sufficiently attention in the academic literature. We fill this gap by using hand-collected data from Internal Model Inspections (IMIs) carried out by the European Central Bank (ECB) and/or the National Competent Authorities on its behalf. Our data set spans banks under the supervision of the Single Supervisory Mechanism between 2014 and 2020. During this period, we identify 267 IMIs, which have culminated in a so called "limitation", i.e. the supervisor imposing higher RWA until a non-compliance in an internal model is remediated. The aggregated additional capital that banks have to hold in response thereto amounts to a double-digit billion Euro figure, which greatly exceeds the equivalent impact of supervisory sanctions. The stark contrast in order of magnitude illustrates the importance of this topic.

We find that the number of deficiencies uncovered in the course of IMIs is statistically not a significant driver of the RWA impact of limitations. Instead, the severity of these deficiencies almost exclusively explains those limitations and hence the implied capital add-ons. We can link these high severity deficiencies to non-compliances with sections of the bank regulation that concern primarily the handling of data for internal models, implying that poor data governance leads to poor model performance.

Our policy implications are twofold. First, our results show that banking supervision could become more risk-based as evidenced by high-severity findings being a main driver of limitations. Second, we show possible room for improvement with respect to banking supervisors' risk-sensitivity, when calibrating limitations. These findings are robust to possible reverse causality and distortions from outliers.

1 Introduction

Statistical models have become de facto ubiquitous in today's financial system. The recent transition to the expected credit loss accounting within the IFRS 9 accounting framework has yet increased this prevalence of internal models in banking. In light of this ever growing importance, it is highly desirable that the applied models are "[...] consistent, reliable and valid [...]" (see Rogers (2021) and CRE36 of Basel III). While models allow the quantification of risks and hence make them manageable, they are no panacea. Derman points out as early as in 1996 that merely applying statistical models does not exclusively help managing risks. Instead, their usage also introduces a novel kind of risk: model risk. Generally speaking, model risk refers to the potential loss an institution may incur due to a decision that was made on the erroneous output of an internal model (cf. CRD Art. 3(1)(11)). More specifically, such losses could arise from the development and implementation (e.g. incorrect computer codes applied or wrong calibration of model parameters), but also from the application of models outside their intended scope (Barrieu and Scandolo, 2015). Against this background and context numerous lines of defense are employed in order to prevent model risk to a reasonable extent. Internally, banks validate their models through the internal validation function (IVF), while externally auditors and supervisors scrutinize new and existent models alike.

Despite theses efforts, the materialization of model risk prevails. While it may be perceived as a theoretical issue in the public's eye, this view is possibly biased by the fact that banks under-report the materialization of model risk. Even if this risk is reported, its true impact is likely to be concealed by the fact that its realization rather impacts banks' profitability, and only in the utmost instances their viability (Aggarwal et al., 2015). Not surprisingly, a literature or web search that is based on public information for losses stemming from erroneous internal rating-based (IRB) models only yields very few results with e.g. the work of Aggarwal et al. (2015) being such an exception. However, from supervisory data we know that the actual number is larger by orders of magnitude. Nevertheless, understanding the conjunction of a higher prevalence of deficiencies in models on the one hand and their comparatively material impacts with regard to risk-weighted assets (RWA) and hence solvency ratios on the other hand has so far not received sufficiently attention in the academic literature. We fill this gap by using hand-collected data from Internal Model Inspections (IMIs) carried out by the European Central Bank (ECB) and/or the National Competent Authorities (NCAs) on its behalf. Our data set spans banks under the supervision of the Single Supervisory Mechanism (SSM) between 2014 and 2020. During this period, we identify 267 IMIs, which have culminated in a so called "limitation" (i.e. the supervisor imposing higher RWA until a non-compliance/deficiency in an internal model is remediated). The aggregated additional Common Equity Tier 1 (CET1) capital that banks have to hold in response thereto amounts to a double-digit billion Euro figure, whereas over the same time, the equivalent impact of supervisory sanctions imposed by the ECB is only a double-digit million Euro figure.[1] This stark contrast in order of magnitude illustrates the importance of the topic at hand.

In our empirical analysis we provide evidence that the number of deficiencies (henceforth "findings") identified by supervisors in the course of IMIs is – in statistical terms – not *per se* a significant driver of the RWA impact of limitations. Instead, it is the severity of the findings that almost exclusively explains those limitations and hence the implied capital add-ons. We can link these high severity findings and non-compliances to particular sections of the Capital Requirements Regulation (CRR). More specifically, we identify non-compliance with data governance requirements as a main channel through which model risk propagates. Eventually, the reliability of a banks' model output is constrained by the quality of its inputs, where biased, erroneous, and/or incomplete data are compromising the statistical capabilities of the model. Against this backdrop, our findings are crucial, as they help to gauge the potential severity of such supervisory findings. That is, they help banks to identify key areas to test and validate their models, and auditors as well as supervisors to challenge them, respectively. The policy implications thereof are twofold. First, our results show in the context of supervisory simplification that banking supervision could become more risk-based without compromising resilience by focusing on non-compliances with data governance that lead to high severity findings and are hence the main driver of limitations. Second, we show possible room for improvement with regard to banking supervisors' risk-sensitivity, when calibrating limitations. These findings are robust to possible reverse causality and distortions from outliers as well as a possible structural break from a change in inspectors' supervisory approach under the Targeted Review of Internal Models (TRIM).[2]

Our results are important for numerous stakeholders: Firstly, for banking supervisors as they might prioritise scarce supervisory resources to work on the particular subset of deficiencies that we identify in our analysis as drivers of model risk. At the same time a targeted model inspection that is based on such early warning signals may help to facilitate a risk-based approach in the follow-up of imposed limitations after the findings of deficiencies. In doing so and recognizing that certain limitations entail a more severe risk than others, a standard "one size fits all" approach can be substituted in favour of a more tailor-made approach and thus contribute to more efficient banking supervision. Secondly, investors may benefit from more targeted supervision because banks possess private information on the quality of

[1] Since its inception until the end of our observation period, i.e. 2020Q4, the ECB had imposed sanctions totalling EUR 44.3 mn.

[2] At the beginning of 2016, the ECB launched in close cooperation with national competent authorities (NCAs) that are part of European banking supervision, a multi-year effort to assess Pillar I internal models in significant institutions that they jointly supervise. This Targeted Review of Internal Models (TRIM) investigated if models meet regulatory requirements and their results are reliable and comparable. Beyond that, it harmonised supervisory practises regarding internal models within the SSM, see European Central Bank (2021).

their models, which can have profound implications on capital ratios and hence investors' decisions whether or not to invest, and thereby exercise a disciplining effect for under-capitalized banks.

This paper is organized as follows: Section (2) provides the institutional background on the supervision of internal models by the ECB and reviews the current literature. Given the novelty of our hand collected data set, we present it in more detail in Section (3), before deriving our research questions in Section (4). We present our results in Section (5), followed by robustness tests in Section (6). Section (7) concludes.

2 Institutional Background and Literature Review

2.1 Literature review on internal models

A fundamental shift in banking regulation occurred with the adoption and implementation of the Basel II regulation in 2007. Since then, banks are allowed to model inter alia the credit risk of their portfolio with IRB models. Subject to compliance with numerous regulatory conditions banks thus do not have to abide by the "one size fits all" approach within the so-called "Standardised Approach" (STA), see Basel Committee on Banking Supervision (2013). The underlying rationale is that as banks have individually a better view of the idiosyncratic risks in their portfolio they may foster systemic stability by measuring risks more accurately through dedicated models at the institutional level. Despite these liberties, banks are not unconstrained in their modelling choices. Instead, features such as a minimum probability of default (PD), or more recently the introduction of an output floor are intended to prevent a "race to the bottom" (Basel Committee on Banking Supervision, 2013).

The continuous evolution and implementation of numerous safeguards envisages to ensure that there is "no free lunch" from applying IRB models. Indeed, Derman (1996) was one of the first to point out that the introduction of models to measure risk carries a risk of its own: model risk. In this line, Aggarwal et al. (2015) argue that any model will eventually be subject to model risk, as it has to either make simplifications, given the inherent complexity of the modelled relationships; or – in the absence of historical observations for very high percentiles – appropriate assumptions have to be made. Similarly, Cont (2006) points to the risk of parametrizing the model unintentionally wrong and thereby introducing another form of model risk. After all, the critique of Lucas (1976) holds in this context as well: The historical links on which a model was calibrated, might fail "to hold true" at the time of the model's application due to agents' shifts in expectations and therefore a shift in the parameters of the estimated model. As a consequence, the predictions of the estimated model become unreliable. Given these unfavourable properties, numerous attempts have been made by regulators around the globe to reduce model risk.

For example, the FED issued in 2011 the "Supervisory Guidance on Model Risk Management", thereby setting minimum standards for, inter alia, the model development and validation. Similarly, the ECB has published its "Guide to Internal Models", which gives granular guidance on topics such as the appropriate counting of defaults. Given that the majority of IRB models is calibrated against the number of observed defaults, this apparently minuscule provision has substantial real world implications. A notable tension lies in the granularity of these provisions: if on the one hand there are too many degrees of freedom, banks may opportunistically manipulate their estimates as suggested by the works of Mariathasan and Merrouche (2014) and Behn et al. (2022). If on the other hand provisions are too strict and banks would price financial instruments unanimously, the resulting lack of divergent views may constitute a financial stability risk through "group think" (Daníelsson et al., 2004). Colliard and Georg (2020) discuss a combination of both cases, where regulation becomes so complex that even a faithful bank fails to apply it correctly, demonstrating the inherent tension on the *right* level of regulation. In this light, BCBS (2013) calls for banking regulation to be generally (i) simple, (ii) comparable, and (iii) risk sensitive.

In particular, the call for models to be simple has been repeated numerous times. Hakenes and Schnabel (2014) argue for example that complexity in models can be exploited by sophisticated agents. Mariathasan and Merrouche (2014) give credence to this suspicion, by showing empirically that banks tend to engage in regulatory arbitrage, where regulatory complexity is high. Barucci and Milani (2018) pin point this concern in particular to the Advanced-IRB approach, where banks are not constrained to only model the PD. Plosser and Santos (2018) demonstrate that this observation does not apply exclusively to European banks, but also holds for U.S. banks, by assessing the riskiness of the same syndicated loan across different banks. Other instances of issues in banks' internal modelling approaches are documented by Vallascas and Hagendorff (2013) who show from a market-point of view a disconnect between the implied portfolio risk and the risk inferred from IRB models. Similarly, Behn et al. (2022) show that the charged interest rate and default rate for an IRB portfolio go up, despite the estimated PD going down according to the bank's internal model. Haldane and Madouros (2012) argue along these lines that not only models should become simpler, but rather bank regulation altogether. More specifically, they call for focusing on simpler metrics such as the Leverage Ratio. Barrieu and Scandolo (2015) stress in this context that the model choice itself is a substantial driver of the final risk measure. This view is also shared by Leitner and Yilmaz (2019) who show that in theory banks may only disclose one of many possible models. The chosen model produces risk-weights which are optimal, relative to its discovery costs and at the same time do not generate a negative signal for the supervisor to scrutinize the model more carefully. Nevertheless, with hindsight, results of many of these papers might be taken with some scepticism, as the data they are based on has since been superseded by yet a new regulation, that is Basel III. At the same time, there is also literature pointing to the benefits of IRB models beyond their increased risk-sensitivity. Exemplary, Bruno et al. (2023) show that more intensive usage of IRB models, and in particular the advanced IRB approach, reduces the opacity of banks due to better data quality and disclosure thereof. Against this background, we set out to investigate the status quo on internal models, and look in particular for drivers of model risk.

2.2 Supervisory background on internal models and findings

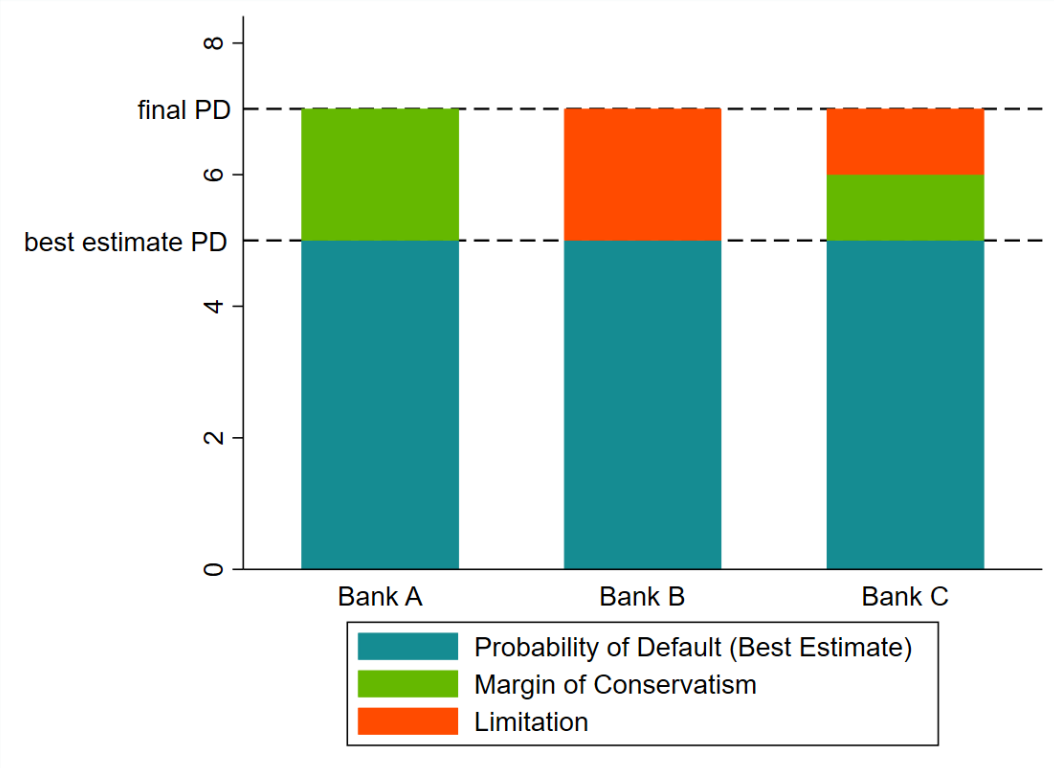
Given that the topic under review in this paper represents rather a niche within the existing academic literature, this section will provide some institutional background on IRB models in banks under the supervision of the ECB.

Internal Model Inspections (IMIs) are usually triggered and requested by banks for two reasons: (i) either a bank seeks initial approval of a newly developed model or (ii) an existent model has undergone material changes as defined in EBA/RTS/2026/05. Following such a request for an IMI, the applying bank will be notified when the investigation will take place, and be asked to provide first information for a kick-off meeting that precedes a more intense on-site phase, where model experts review the model on the bank's premises. This period concludes the inspection phase, after which an Assessment Report (AR) is drafted. It lists all deficiencies, ranked by severity on a scale from F1 (low) to F4 (very high), and is checked for consistency. In the absence of comments, respectively incorporation thereof, the findings of the report become a legally binding decision by the ECB through approval by the Supervisory Board of the SSM. At this point in time, the inspected bank is obliged to remediate the listed deficiencies within a given period of time that is proportionate to the severity of the finding and the complexity to remediate it. Where high severity findings suggest an underestimation of the actual risk, a limitation may additionally be imposed to address the risk of the identified underestimation and hence to preserve the level playing field. For example, the limitation could legally oblige the bank to increase its final PD estimates by a given factor to prevent it from gaining an advantage by calculating unduly low RWAs. For our analysis, we look in particular at the RWA equivalent of these limitations which represent a gap between the model-based estimation of the PD and its hypothetically *true* value. It is worthwhile mentioning that the notion of a "true" PD can be debated in the sense of e.g. Aggarwal et al. (2015) who question this concept, arguing that differences between reality and a model's output are more reminiscent of an institution's risk appetite. Similarly, Leitner and Yilmaz (2019) suggest that the bank never reveals this *true* model, but rather opportunistically chooses one that performs well for the given portfolio.[3] Nevertheless, this concept allows us to derive a credible counterfactual RWA equivalent of the difference between the "true" final PD and the best-estimate PD in Figure (1) below as a variable of interest for exploring the drivers of model deficiencies.

While this makes for a reasonable measurement of a model's limitation, it is worth noting that imposing a limitation is not the rule to an inspection. Instead, there are numerous banks that prefer to err on the side of conservatism and faithfully apply Margins of Conservatism (MoC) in order to account for potential inaccuracies in their model in the first place. In fact, comparable supervisory data by Casellina et al. (2026) suggest that there is a sizeable share of models that is without such risk underestimations. To illustrate the impact where they nevertheless do occur, consider the hypothetical examples of Figure (1). Let us assume that all three banks A, B and C are asked to estimate independently from each other the PD of an identical obligor, based on the same information set and without any constraints. Bank A computes the best estimate of the PD and supplements it with a MoC to account for uncertainty in the estimation. This approach is broken down in three different categories, related to data and methodological deficiencies, structural breaks in the data (e.g. changes in underwriting standards, policies, risk appetite, etc.), and general estimation errors. We will jointly refer to them as merely the MoC, which when added to the best estimate PD forms the final PD in our example. Bank B arrives at the same best estimate PD, yet neglects possible deficiencies of the model, and would hence report the final PD as being equal of the best estimate PD, thereby underestimating the remaining ex post risks that together comprise the *true* entire risk and hence the *true* final PD. In order to prevent this infraction on the level playing field, the inspection team would consequentially step in, and impose a limitation, such that the *true* final PD is reached. Mixtures of this case as illustrated for Bank C can of course occur.

[3] In the applied economic and econometric literature there is often a confusion with respect to the concept of "the true model". True models exist in the context of controlled experiments such as statistical simulations. In a modelling context based on realised observed empirical data, however, all models are "wrong", see e.g. Box (1976). In the context of this paper we therefore rather interpret the "true" final PD as the best statistical approximation of its unknown Data Generating Process, see e.g. Hendry (1995).

Figure 1: Examples of differences in estimating the final Probability of Default across stylized banks, in %.



3 Our Data Set

Given the novelty of our hand collected non-public supervisory data set, we will describe it in more detail in this section. In doing so, two dimensions are highlighted in particular: (i) the data collection and management process that preceded our data analysis, and (ii) an illustration of the final data set.

Our analysis starts from the set of IRB banks supervised by the SSM since its inception in 2014. In order to create a balanced panel, we only look at banks that still continue to exist after the cut-off date in the fourth quarter of 2020. We deliberately choose 2020 as a cut-off to account for three effects: First, to avoid the presence of data in our sample that is based on IMI's which cannot be fully evaluated as their remediation phase has not been concluded, yet. Second, the introduction of the European Banking Authority's (EBA) Guideline on PD estimation, LGD estimation and the treatment of defaulted exposures (EBA/GL/2017/16) effective as of January 2021 and therefore possibly constituting a structural break. Third, numerous regulatory and governmental interventions were initiated and implemented during the COVID-19 pandemic, possibly biasings our results. An example thereof could be the suspension to count days past due, thereby effectively preventing obligors from defaulting - a key input in each rating model. For the banks meeting our sampling conditions, we proceed to manually review the so called "Supervisory Board Notes". Each of them constitutes an official ECB Decision, which are the final outcome of a corresponding IMI. We collect the RWA-equivalent of the imposed limitations – where available, i.e. not all IMIs conclude with a limitation – from the aforementioned Notes. Notice that we limit the scope of our sample to IRB models for calculating the credit risk RWA, in order to keep our identification strategy as clean as possible and to reduce the effect of different sets of rules for modelling different risk categories. In doing so, we leverage on the unique properties of our supervisory data set that allows us to put a quasi counterfactual price-tag on model deficiencies. This metric, the additional RWA impact from a limitation standardized by total assets in order to control for larger limitations in banks with larger portfolios, will serve as the dependent variable in all subsequent analyses. One important feature of this analysis is that we isolate from the *overall* RWA impact induced by the model change the *individual* RWA resulting from the imposed *supervisory* limitation. As such, our dependent variable has a strict lower bound at zero, as a limitation cannot have a negative RWA impact. This feature would not hold, if we were to look just at the aggregate *overall* RWA impact of any model change, which could even be negative due to capital relief from model refinements or merely updating the time-series for periods with less actual defaults.

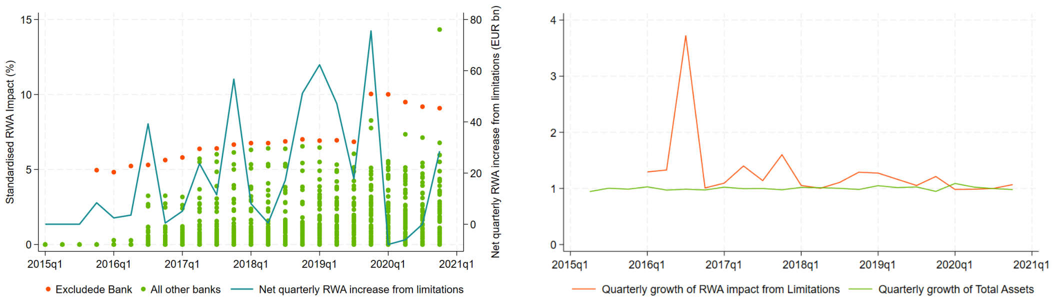
We then proceed to the "Assessment Reports" (ARs), which contain a summary of all reported model deficiencies ("findings"), and, yet more importantly, also provide a reference to any particular non-compliance *vis-à-vis* the regulatory framework, i.e.

CRR/CRD. This information along the attached severity of the finding is available in structured form through the confidential supervisory "Agora" data base. We use this information in all subsequent models as explanatory variables, which are complemented by additional idiosyncratic control variables. A full list of those controls, including comprehensive descriptive statistics can be obtained from Table (1) in the Appendix. To ensure comparability we have chosen our set of explanatory variables to broadly correspond with those employed in other academic studies, such as e.g. Basel Committee on Banking Supervision (2023) and Mariathasan and Merrouche (2014). They find that models are responsible for up to 80 % and 77 % of the overall capital requirements, respectively. The slightly higher number in our sample (i.e. 88 %) is possibly driven by legally mandated higher minimum coverage ratios in Germany (92 %), which accounts for roughly one in four banks in our sample. Our figures are consistent with the 50 % minimum coverage ratio to be eligible for using an IRB approach (see ECB Guide to Internal Models General Topics para. 27). As we constrain our analyses on credit risk models, some articles of the CRR will never be infringed in first place, while others are more frequently mentioned. We address this heterogeneity in their distribution by introducing clusters in line with the sections of the CRR. As an illustrative example, articles 138 to 141 would be counted as one section on "Use of the ECAI credit assessments for the determination of risk weights". Lastly, because a bank can have multiple IMIs in a given year, we aggregate the aforementioned information at the bank $\times$ quarter level as represented by the green dots in Figure (2a), effectively looking at the stock of open findings. As we track a bank through the panel, the reported variables are hence the respective levels of the variable at point t in time, for bank i, unless stated otherwise.

For our analysis we neither winsorize nor truncate our data. Instead, we drop severe outliers as they exclusively relate to one particular bank in our sample. This bank dominates the time-series throughout the panel we analyse, as illustrated by the orange dots in Figure (2a). As a result, our final sample consists of 50 IRB banks which we track through 24 quarters. We demonstrate that this choice does not drive our results by running a winsorized model in Section (6.1) below which is dedicated to our robustness tests. Beyond justifying the outlier treatment, Figure (2a) illustrates a few key properties of our dependent variable: First, one can observe a shift in the standardised RWA impact towards the end of 2016. We conjecture that this relates to the conclusion of the first TRIM missions. The impact of the TRIM missions is assessed in more detail in a robustness test in Section (6). Second, one can observe that the standardised impact of limitations grows over time. This trend is due to two factors. As the remediation of existing findings takes a substantial amount of time, there is an accumulation effect as more new findings are raised than old findings are closed. This observation is illustrated by the net quarterly flow in the blue line. In addition, due to supervisory learning since the inception of the SSM and hence a more consistent interpretation of the regulation over time, imposed limitations grew on average faster than total assets, see Figure (2b). Both effects jointly imply that our numerator grows faster as our denominator, hence explaining the positive trend.

Figure 2: Development of the RWA Impact of the imposed limitations.

(a) Impact of the standardised limitation (b) Limitation and Total Assets growth rate



4 Research Questions

Given the importance of statistical and empirical models in crucial operations of today's financial system, we set out to shed more light on the caveats that are associated with their usage. More specifically, we investigate factors that contribute to our proxy variable for model risk. In doing so, two opposing views exist in the supervisory policy space: One that argues that many small deficiencies are concerning because they compound, whereas another sees few but severe deficiencies as the culprit of poor model performance.

Research Question 1 *Is the quantity or the severity of findings the actual driver of the imposed limitation?*

Additionally, we investigate whether early warning signals exist that could alert internal and external stakeholders alike about possible issues with any model under scrutiny. As we can extract the exact non-compliance with the CRR-regulation from our data set, we envisage to pin-point individual pockets of risk that call for greater attention.

Research Question 2 *Can we identify early warning signals that would point to possible drivers of model risk?*

In order to address these questions we apply a standard panel ordinary least squares (OLS) model with fixed effects. In addition to the hand-collected metrics mentioned in Section (3), we complement our regression model with additional metrics. More specifically, we control for different bank business models by calculating the revenue diversification metric (ROID) as shown in Equation (1):

$$\text{ROID}_{i,t} = 1 - \left| \frac{\text{Net Interest Income}_{i,t} - \text{Net Fee and Commission Income}_{i,t}}{\text{Net Interest Income}_{i,t} + \text{Net Fee and Commission Income}_{i,t}} \right| \quad (1)$$

In this notation, t refers to the time, while i represents the individual bank. Additionally, we proxy the riskiness of a bank. To this end, we measure the capital in excess of the minimum Common Equity Tier 1 (CET1) requirements as capital headroom (henceforth Headroom), and its cost of risk (COR), that is:

$$COR_{i,t} = \frac{Impairments_{i,t}}{RWA_{i,t}} \tag{2}$$

where *Impairments* refers to the quarterly flow of loan loss provisions, and RWA is the stock of total risk-weighted assets. In addition, we control for the profitability of a bank by employing Return on Assets (ROA), which is net income divided by total assets. We chose ROA over Return on Equity, in order to prevent any systematic biases that may be induced by structurally higher leverage ratios for some banks in the sample. Furthermore, we control for the size of the bank with the natural logarithm of total assets, in order to account for the inherent skewness from banks that enter the sample due to either their relative importance in smaller Euro Area member states and/or being significantly connected with entities abroad.

We regress the aforementioned variables on the ratio of the limitation standardized by total assets (henceforth labelled "Limitation"), in order to account for banks with large IRB models that would hence, ceteris paribus, naturally also incur larger limitations. It is important to note that we do not standardise by RWA for reasons of reverse causality, given that we investigate the possible underestimation of the RWA in the first place. Our full specification is thus:

$$\begin{aligned} Limitation_{i,t} = {} & \beta_1 F1_{i,t-1} + \beta_2 F2_{i,t-1} + \beta_3 F3_{i,t-1} + \beta_4 F4_{i,t-1} + \beta_5 ROA_{i,t-1} \\ & + \beta_6 Size_{i,t-1} + \beta_7 Headroom_{i,t-1} + \beta_8 COR_{i,t-1} + \beta_9 ROID_{i,t-1} \\ & + \alpha_i + \gamma_t + \epsilon_{i,t} \end{aligned} \tag{3}$$

We lag our regressors by one period, in order to address potential concerns with regard to simultaneity as we track the cumulating stocks of limitations and findings alike. Furthermore, we employ bank (α) and time-fixed effects (γ) to account for additional unmeasured heterogeneity, while $\epsilon_{i,t}$ is the error term. All standard errors are clustered on the bank level due to the number of periods being small, compared to the number of banks, possibly biasing our estimators due to a lack of clusters as described by Abadie et al. (2022). Additionally, policies that may induce cross-sectional correlation in the time dimension, e.g. policies effecting all banks, have been excluded to the extent possible (cf. discussion on EBA/GL/2017/16 and COVID interventions in Section 3).

5 Results

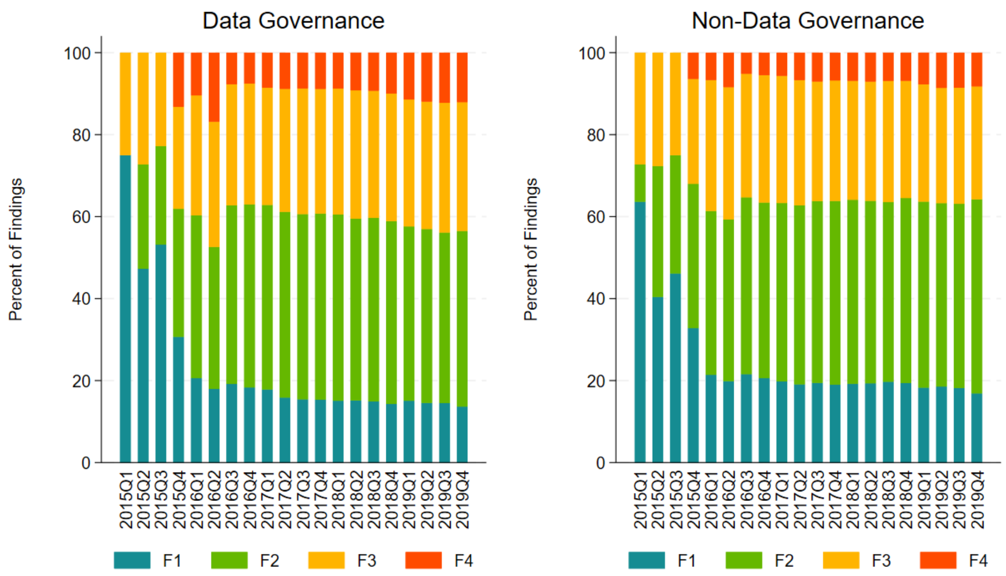
Following our first Research Questions above, we begin our analysis by verifying whether there exists a link between the number of findings – clustered by their severity – and the magnitude of the limitation standardized by total assets. We find in the first column of Table (2) that only in the instance of the highest severity, i.e. "F4", the number of findings is a driver for the severity of the resulting limitation. As this effect is statistically and economically significant we proceed with our analysis and introduce additional regressors to improve the robustness of our model.

Table (2) about here

We conjecture that this observation could also be due to the fact that lower severity findings do not matter in isolation but only in conjunction with higher severity findings. Accordingly, we interact the counts of low severity findings, i.e. F1 and F2, with the high severities F3 and F4, respectively. We find in the second column of Table (2) that this interaction does not change our results. Instead, F4 remain in statistical terms the exclusive driver of the severity of a limitation. Having established this link, we further extend our model to include bank-level controls. More specifically, it is conceivable that more resilient banks behave differently from riskier banks, which may also be reflected by the modelling approach of these banks. Recall that Blum (2008) shows a stronger incentive for weakly capitalized banks to move towards the IRB approach. Similarly, due to more available resources, larger banks may e.g. be in a more favourable position to model risks along higher quality standards, be it through more available staff or data to calibrate the model. Against this background, we show in column three of Table (2) that our established link between F4 findings and the impact of limitations remains robust at the 95 % confidence level. At the same time, better capitalized banks appear to receive larger limitations as judged by the coefficient of Headroom. A similar story is told from the other end, where riskier banks as judged by COR appear to receive smaller limitations. This observation could be interpreted in two ways: First, in line with Casellina et al. (2026), portfolios with more defaults are better calibrated because there are more observations to calibrate the model with. This would counter-intuitively imply that riskier banks have on average better models. Second, drawing on Agarwal et al. (2024), this observation could results from noise in supervisory decision due to human discretion. From a methodological perspective this might not necessarily be harmful, as it allows to incorporate soft information into the decision. In fact, a case can be made that such discretion is necessary, in order to prevent banks from opportunistic behaviour to gamble a possible strictly mechanistic and hence predictable scoring-mechanism. In any case, we scrutinize this observation in more detail in Section (6.2) of the robustness tests, as another way of framing it could arise from reverse causality: banks are only better capitalized because they under-estimate the *true* risk, and hence need bigger limitations to correct for this.

Having established the transmission channel that findings of the highest severity are drivers of the impact of the limitation, we proceed to investigate our second research question, that is: is there a possible early warning signal for high severity limitations? We investigate this question using our established empirical model, which we extend by another dimension that is unique to our data: from the findings, we can track the actual non-compliance with the corresponding article of the CRR that led to the finding. Following the clustering technique described in Section (3), we introduce a dummy for each section of the CRR. In untabulated results, we include these dummies in our regression, and find that some of them are particularly meaningful for our model. More specifically, there is a subset of sections in the CRR, which relates to data governance that leads on average to more high severity findings, i.e. F3 or F4, as illustrated in Figure (3) below.

Figure 3: Visualisation of subset of high severity drivers.



This graphical evidence is further supported by a test for equality of the two groups, illustrated in Table (3). Therefrom, the assumption of both groups being equal is strongly rejected, with an excess of low severity findings in non-data governance related deficiencies, and vice versa for high severity findings.

Table (3) about here

Following this evidence, we proceed to include a dummy in our regression that is one, whenever a finding is associated with an article of the data governance group, and zero otherwise. Including this dummy in the specification of our model as shown in column four of Table (2) suggests that the data governance group is both, a statistically and economically meaningful driver of the severity of the limitation.

Taken together, we establish a link between high severity findings and the magnitude of the imposed limitation. We show that this link is robust, when controlling for bank characteristics such as riskiness, size, profitability, and business model. We can use this observation to extend our model and show that a possible early warning signal for high severity limitations exists: findings associated with non-compliances with regard to data governance should get the undivided attention of supervisors and bank managers alike. Indeed, a case could be made that if the input data are subpar, so will be the prediction derived therefrom. While we dispel concerns with regard to simultaneity by employing lagged values of our independent variables, the subsequent section will further demonstrate the robustness of our findings.

6 Robustness Tests

6.1 Winsorization

As explained in the data section of this paper, we exclude one bank from our sample as it continuously dominates all other observations in the sample (cf. Figure (2)), and hence biases the results. In this subsection, we investigate the counterfactual of not excluding this particular bank, but instead winsorizing our observations at the 1^{st} and 99^{th} percentile. The results thereof are reported in Table (4).

Table (4) about here

For comparison, we display the original values of our regression adjacent to the winsorized results in Columns (2) to (7) of Table (4). We find that our main observation holds: findings of the highest severity drive the standardized RWA impact of the imposed limitations. All control variables demonstrate comparable signs and magnitudes as well as statistical significance.

6.2 Endogeneity of bank characteristics

Following our research question, we econometrically investigated if the severity or quantity of model deficiencies matters concerning the model risk, that is the risk underestimation by the model. Throughout our analysis we generate evidence, suggesting that high severity findings matter more than numerous low severity findings. In fact, we can pinpoint the non-compliant handling of data, as a major channel that leads to high severity findings. While we thereby answer our research question, we want to address a possible endogeneity issue in this robustness test. Specifically, the relevance of high severity findings we observe could be pre-determined by bank characteristics. For example, better capitalised banks receiving harsher limitations than their less robust peers. Against this backdrop, we interact the F4s we observe with key bank characteristics, such as profitability as measured by RoA, in order to investigate this possibility.

Table (5) about here

Comparing to our base model in the first column, we find that neither RoA in the second, nor COR in the third column, explain as a bank characteristic the prevalence of F4 findings. We conclude from this observation that F4s are not predetermined by bank characteristics, but instead idiosyncratic to the bank $\times$ joint supervisory team (JST) combination. That is to say that decisions on the severity are done on a case by case basis by the inspection team and subsequently the JST. Note that this does not mean a judgement error, as in different JSTs assessing the same information differently, but instead that all idiosyncrasies of a case are considered and that the outcome is not predetermined by the type of bank that is investigated.

6.3 Time Effects

As discussed in Section 3, we cluster the standard errors of our analysis at the bank-level and control for time with fixed effects. However, in this section we want to further explore the robustness of these assumptions. To this end, we use the Targeted Review of Internal Models (TRIM) exercise as a control. It was designed in order to ensure a level playing field by investigating the largest IRB models in the banks under the direct supervision of the SSM. Given the effort required, the corresponding model inspections were not done exclusively by ECB and NCA staff, but also drawing on external consultants. Arguably, doing so could constitute a structural break, which we can leverage to challenge the assumption of our data to be uncorrelated in the time dimension. To do so, we drop in a first step the time-fixed effect and instead introduce a dummy that takes the value of one during the time of the TRIM missions. In a second step, we relax our assumption on the temporal correlation and introduce double clustering at the bank and time level. Lastly, we employ the most restrictive specification of our model, including a dummy for the TRIM period as well as double clustering of standard errors. Ultimately, we drop the TRIM dummy, in favour of a mere time dummy, which should – if TRIM were the driving factor of the temporal correlation – only be significant during the TRIM quarters. The results of this analysis can be obtained from Table (6) below.

Table (6) about here

We find in column two that the dummy for the TRIM period is statistically highly significant. However, its introduction does not change our original interpretation: the main driver of the limitation, i.e. our proxy for model risk, remains the number of F4 findings. All other transmission channels, particularly on the data compliance remain in tact. Similarly, F4s remain the main transmission channel in column three, where we double cluster by bank and time. While all other variables become insignificant, we caution against over-interpreting this observation given the small number of clusters (i.e. T $<$ N) in line concerns raised by Abadie et al. (2022). Lastly, column four depicts the baseline model amended by the TRIM dummy and double clustering, showing - again - that severity of the findings matters more than their quantity, and that non-compliances on the data treatment are a strong driver thereof. In unreported results, we have introduced a random time dummy that determines arbitrary pre- and post-periods of the TRIM exercise. These dummies are only significant for the period, where the actual TRIM exercise was conducted, attesting to the robustness of our results.

7 Conclusions

Modern finance has grown increasingly complex and thereby risky. In order to manage these risks, a vast number of statistical techniques are applied, which we broadly refer to as "models". While these models may allow to quantify and thus arguably manage risks, they are not flawless in and by themselves. Mistakes in models due to e.g. misspecification or wrong parametrisation may lead to erroneous decisions because of misleading model output. Those models thereby create a new risk of their own, that is model risk.

In our study, we set out to get a better understanding of model risk by analyzing a confidential data set of supervisory inspections of IRB models which are employed by European banks for measuring Pillar 1 credit risk. We narrow down our analysis to this particular set of risk, in order to have an as clean as possible identification of the actual drivers of this new category of risk, i.e. model risk.

In doing so, we investigate two research questions. The first relates to two opposing schools of thought as to what constitutes a "bad" model: Is it either a model with numerous deficiencies, or rather a model with limited but substantial weaknesses? Our analysis shows that the number of issues identified in a model is – statistically speaking – *per se* not a significant driver of model risk as approximated in this study by the counterfactual penalty that the supervisor has to impose in order to maintain a level playing field. Instead, it is the severity of the associated findings that drive the limitation.

Having established a link between high-severity findings and the underestimation of a risk model, we look closer at the possible transmission channel in our second research question. Identifying the mechanics behind this observation could allow to construct early warning signals for these issues and thereby make banking supervision more risk-based. Indeed, we find a subset of articles from the CRR, which are related to data governance and are more often associated with higher severities and should hence be focal points for supervisors to check. At the same time, this should not create a false sense of security. Even with this knowledge there is no notion of a universally true model. Instead "all models are wrong" in the sense that they can only predict outcomes in a statistical sense.

That being said, our results are useful for bankers and supervisors alike. First, they allow to prioritize scarce resources on both sides in order to be particularly mindful of key topics that are risk drivers within the field of risk modelling. Second, and in line with this thought of prioritization, we generate evidence suggesting that supervisors should not occupy themselves with checking every minuscule detail, but instead focus on the material deficiencies in a model. This may be a valuable observation at a time of discussing simplification without compromising resilience. This recommendation is robust to possible reverse causality concerns as well as outlier treatments in the data we analyze.

Future research should try to reinstate our findings in different jurisdictions and different legal settings in order to narrow down the list of early warning signals that point to deficiencies in models that are used to ultimately calculate solvency ratios.

References

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. M. (2022). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics*, 138(1):1–35.

Agarwal, S., Morais, B. C., Seru, A., and Shue, K. (2024). Noisy Experts? Discretion in Regulation. Working Paper 32344, National Bureau of Economic Research.

Aggarwal, A., Beck, M. B., Cann, M., Ford, T., Georgescu, D., Morjaria, N., Smith, A., Taylor, Y., Tsanakas, A., Witts, L., and Ye, I. (2015). Model Risk - Daring to open up the Black Box. *British Actuarial Journal*, 21:229–296.

Barrieu, P. and Scandolo, G. (2015). Assessing financial model risk. *European Journal of Operational Research*, pages 546 – 556.

Barucci, E. and Milani, C. (2018). Do European Banks manipulate Risk Weights? *International Review of Financial Analysis*, 59:47–57.

Basel Committee on Banking Supervision (2013). The regulatory framework: balancing risk sensitivity, simplicity and comparability.

Basel Committee on Banking Supervision (2023). Basel III Monitoring Report.

Behn, M., Haselmann, R., and Vig, V. (2022). The Limits of Model-Based Regulation. *Journal of Finance*, 77(3):1635–1684.

Blum, J. M. (2008). Why 'Basel II' may need a leverage ratio restriction. *Journal of Banking & Finance*, 32(8):1699–1707.

Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association*, 71(356):791–799.

Bruno, B., Marino, I., and Giacomo, N. (2023). Internal ratings and bank opacity: Evidence from analysts' forecasts. *Journal of Financial Intermediation*, 56:101062.

Casellina, S., Chionsini, G., Kopp, R. M., and Riabi, M. (2026). Systematic Backtesting of Probability of Default Models with Regulatory Data. Staff Paper 24, European Banking Authority.

Colliard, J.-E. and Georg, C.-P. (2020). Measuring Regulatory Complexity. CEPR Discussion Papers 14377, C.E.P.R. Discussion Papers.

Cont, R. (2006). Model uncertainty and its impact on the pricing of derivative instruments. *Mathematical Finance*, 16.

Daníelsson, J., Shin, H. S., and Zigrand, J.-P. (2004). The Impact of Risk Regulation on Price Dynamics. *Journal of Banking & Finance*, 28(5):1069–1087.

Derman, E. (1996). Model Risk in Quantitative Strategies Research Notes from Goldman Sachs.

European Banking Authority (2017). Guidelines on PD and LGD estimation.

European Central Bank (2021). Targeted Review of Internal Models - Project Report.

European Central Bank (2024). ECB guide to internal models.

Hakenes, H. and Schnabel, I. (2014). Regulatory Capture by Sophistication. *ERN: Econometric Studies of Government Regulation of Financial Markets (Topic)*.

Haldane, A. and Madouros, V. (2012). The Dog and the Frisbee. *Proceedings - Economic Policy Symposium - Jackson Hole*, pages 109–159.

Hendry, D. F. (1995). *Dynamic Econometrics*. Oxford University Press.

Leitner, Y. and Yilmaz, B. (2019). Regulating a model. *Journal of Financial Economics*, 131(2):251–268.

Lucas, R. E. (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy*, 1:19–46.

Mariathasan, M. and Merrouche, O. (2014). The manipulation of basel risk-weights. *Journal of Financial Intermediation*, 23(3):300–321.

Plosser, M. C. and Santos, J. A. C. (2018). Banks' Incentives and Inconsistent Risk Models. *Review of Financial Studies*, 31(6):2080–2112.

Rogers, C. (2021). Basel III and global cooperation: where do we go from here? – Keynote speech at the The Kangaroo Group virtual debate.

Vallascas, F. and Hagendorff, J. (2013). The Risk Sensitivity of Capital Requirements: Evidence from an International Sample of Large Banks. *Review of Finance*, 17(6):1947–1988.

8 Appendix

Table 1: Descriptive Statistics of used Variables

	Obs.	*Min.*	$Q_{0.25}$	*Mean*	*Median*	$Q_{0.75}$	*Max.*	σ
Limitation (%)	1,180	0.0000	0.0000	0.0099	0.0020	0.0140	0.1433	0.0152
F1 (#)	1,200	0.0000	0.0000	15.6175	8.0000	24.0000	121.0000	19.7479
F2 (#)	1,200	0.0000	0.0000	39.7808	22.0000	60.5000	272.0000	48.2814
F3 (#)	1,200	0.0000	0.0000	24.3858	10.0000	32.0000	274.0000	38.1216
F4 (#)	1,200	0.0000	0.0000	6.5717	2.0000	8.0000	84.0000	12.0921
ROA (%)	1,180	-0.0507	0.0041	0.0092	0.0087	0.0135	0.0497	0.0094
Size (ln)	1,180	22.5844	24.8145	25.7818	25.7247	26.9714	28.5260	1.3344
Headroom (%)	1,200	0.0200	0.0873	0.1158	0.1037	0.1290	0.3262	0.0437
COR (%)	1,180	-0.0395	0.0003	0.0031	0.0013	0.0039	0.663	0.0060
ROID ($\epsilon[-1;1]$)	1,180	-0.9925	0.3399	0.5088	0.5384	0.7472	0.9995	0.3369
Data governance ($\epsilon[0;1]$)	1,200	0.0000	0.0000	0.7217	1.0000	1.0000	1.0000	0.4484
TRIM ($\epsilon[0;1]$)	1,200	0.0000	0.0000	0.3333	0.0000	1.0000	1.0000	0.4716

Note: The table above gives an overview of the used variables and their descriptive statistics. The scale of the variable is given in parenthesis behind the name. Size is the natural logarithm of total assets. ROID is calculated in line with Equation (1) and COR according to Equation (2). Data governance refers to being non-compliant with at least one provision of the CRR that governs the quality and handling of data used to calibrate internal models. TRIM is a dummy that is one for all IMIs after the conclusion of the TRIM exercise.

Table 2: Regression on the EUR impact of a limitation over total assets

	(1)	(2)	(3)	(4)
F1 (#)	0.0002	0.0002	0.0002	0.0001
F2 (#)	0.0000	0.0000	0.0000	-0.0000
F3 (#)	-0.0000	-0.0000	-0.0000	-0.0000
F4 (#)	0.0004*	0.0005*	0.0005*	0.0005*
low severity × F3 (#)		-0.0000	0.0000	0.0000
low severity × F4 (#)		-0.0000	-0.0000	-0.0000
ROA (%)			-0.0542	-0.0603
Size (ln)			-0.0040	-0.0031
Headroom (%)			0.1574**	0.1283*
COR (%)			-0.2471*	-0.2428*
ROID ($\epsilon[0;1]$)			0.0133	0.0082
Data governance ($\epsilon[0;1]$)				0.0063**
Bank-fixed effects	Yes	Yes	Yes	Yes
Time-fixed effects	Yes	Yes	Yes	Yes
N	1,136	1,136	1,130	1,130
R^2	0.2952	0.3085	0.3666	0.3897

Note: The table above demonstrates the step-wise extension of our model, which establishes in the first column a significant link between the severity of a limitation and the associated findings. As shown in the second column, this link is not severed by the coinciding of many lower severity findings. From there we proceed to include bank-level controls to prove the resilience of this transmission channel, even when accounting for idiosyncrasies. Drawing from further analysis, we demonstrate in the fourth specification that findings concerning data governance are a statistically significant driver of the severity of the limitation. Significance is denoted at the 5 % (*), 1 % (**), and 0.1 % (***) level.

Table 3: Test for equality using the mean.

	Average Findings		
	Data Governance	Other	p-value
F1	25.66	26.52	0.0000
F2	36.79	38.88	0.0000
F3	29.17	28.91	0.0000
F4	8.35	5.66	0.0000

Note: The table above displays the percentage of findings for the two groups by severity. We test for the equality of the mean of both groups and find that they are different at a statistically highly significant level.

Table 4: Regression on the EUR impact of a limitation over total assets

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
F1 (#)	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
F2 (#)	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
F3 (#)	-0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
F4 (#)	0.0005*	0.0005*	0.0005*	0.0005*	0.0005*	0.0005*	0.0004*
low severity × F3 (#)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
low severity × F4 (#)	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
ROA (%)	-0.0603		-0.0387	-0.0276	-0.0193	-0.0384	
Size (ln)	-0.0031	-0.0064		-0.0054	-0.0061	-0.0063	
Headroom (%)	0.1283*	0.1252**	0.1252**		0.1254**	0.1248**	
COR (%)	-0.2428*	-0.1723	-0.1695	-0.1745		-0.1688	
ROID ($\epsilon[0;1]$)	0.0082	0.0110	0.0110	0.0105	0.0111		
Data governance ($\epsilon[0;1]$)	0.0063**	0.0069*	0.0069*	0.0067*	0.0069*	0.0069*	0.0064**
win. ROA (%)		-0.0609					-0.2163
win. Size (ln)			-0.0061				-0.0090
win. Headroom (%)				0.1451**			0.1596*
win. COR (%)					-0.1943		-0.1627
win. ROID (%)						0.0120	0.0254*
Bank-fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time-fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	1,130	1,153	1,153	1,153	1,153	1,153	1,153
R^2	0.3987	0.3806	0.3802	0.3832	0.3797	0.3809	0.3896

Note: The table above depicts the original regression in Column (1) and contrasts it to the individually winsorized models in Column (2) to (6) with the fully winsorized model depicted in Column (7). We winsorize at the 1% level, though our results are also robust to the 5% level. We generally show that our variable of interest, the number of findings with severity F4, remains statistically significant in all specifications, as does the transmission channel of non-compliances in the data treatment. Significance is denoted at the 5 % (*), 1 % (**), and 0.1 % (***) level.

Table 5: Regression on the EUR impact of a limitation over total assets

	(1)	(2)	(3)
F1 (#)	0.0001	0.0001	0.0001
F2 (#)	-0.0000	-0.0000	-0.0000
F3 (#)	-0.0000	-0.0000	-0.0000
F4 (#)	0.0005*	0.0005*	0.0005*
low severity $\times$ F3 (#)	0.0000	0.0000	0.0000
low severity $\times$ F4 (#)	-0.0000	-0.0000	-0.0000
ROA (%)	-0.0603	-0.0679	0.0571
Size (ln)	-0.0031	-0.0032	-0.0030
Headroom (%)	0.1283*	0.1289*	0.1275*
COR (%)	-0.2428*	-0.2419*	-0.2809*
ROID ($\epsilon[0;1]$)	0.0082	0.0082	0.0082
Data governance ($\epsilon[0;1]$)	0.0062**	0.0063**	0.0064**
F4 $\times$ RoA		0.0024	
F4 $\times$ COR			0.0135
Bank-fixed effects	Yes	Yes	Yes
Time-fixed effects	Yes	Yes	Yes
N	1,130	1,130	1,130
R^2	0.3897	0.3898	0.3913

Note: We investigate in this robustness test a potential endogeneity issue, where bank characteristics pre-determine the severity of findings, and hence the imposed limitation. To this end, we interact our variable of interest, F4 findings, with bank characteristics. We find that these interaction terms are insignificant, implying that bank characteristics are unlikely to explain the number of F4 findings. The results of interacting F4s with other bank characteristics were inconclusive, i.e. the interaction terms not being significant, yet simultaneously depleting other variables of explanatory power. Significance is denoted at the 5 % (*), 1 % (**), and 0.1 % (***) level.

Table 6: Regression on the EUR impact of a limitation over total assets

	(1)	(2)	(3)	(4)
F1 (#)	0.0001	0.0001	0.0001	0.0001
F2 (#)	-0.0000	-0.0000	-0.0001	0.0000
F3 (#)	-0.0000	0.0000	0.0000	0.0000
F4 (#)	0.0005*	0.0004*	0.0004*	0.0004*
low severity × F3 (#)	0.0000	0.0000	0.0000	0.0000
low severity × F4 (#)	-0.0000	-0.0000	-0.0000	-0.0000
ROA (%)	-0.0603	-0.0122	0.0521	-0.0122
Size (ln)	-0.0031	-0.0072	-0.0104	-0.0072
Headroom (%)	0.1283*	0.1174*	0.1010	0.1174*
COR (%)	-0.2428*	-0.2162	-0.2007	-0.2162
ROID ($\epsilon[0;1]$)	0.0082	0.0061	0.0043	0.0061
Data governance ($\epsilon[0;1]$)	0.0062**	0.0061**	0.0045	0.0061*
TRIM ($\epsilon[0;1]$)		0.0050**		0.0050*
Bank-fixed effects	Yes	Yes	Yes	Yes
Time-fixed effects	Yes	No	Yes	No
Bank cluster	Yes	Yes	Yes	Yes
Time cluster	No	No	Yes	Yes
N	1,136	1,136	1,130	1,130
R^2	0.2952	0.3085	0.3666	0.3897

Note: The table above illustrates further robustness checks concerning the time dimension of our data. Column (1) depicts our baseline model as a reference point. We investigate the impact of TRIM missions in Column (2), showing that the coefficient is statistically significant, yet does not alter our initial results, i.e. the severity of findings matters more than their quantity, with data governance non-compliances being a possible transmission channel. Column (3) introduces double-clustering of the standard errors on the bank and time level, with F4s still being the dominant force in our model. While our variable for data non-compliances becomes insignificant we caution an over-interpretation of the result, given the low number of clusters. Ultimately, Column (4) combines both robustness tests, i.e. TRIM and double-clustering, confirming previous results. Significance is denoted at the 5 % (*), 1 % (**), and 0.1 % (***) level.

Acknowledgements

We thank Tancrède Polge for his outstanding analytical data support. We are grateful to Andrea Enria, Frank Elderson, Klaus Düllmann, Robert Lauter, Mario Quagliariello, Tobias Beck, Daniel Foos, participants of the 3rd Workshop on Banks and Financial Markets at the Paris School of Business, 13th EBA Policy Research Workshop, BCBS Research Group and to participants and our discussant Elizaveta Sizova at the 2024 ECB Banking Supervision Research Conference and an anonymous referee for their valuable comments.
All remaining errors are our own.

Andreas Beyer
European Central Bank, Frankfurt am Main, Germany; email: andreas.beyer@ecb.europa.eu

Arndt-Gerrit Kund
European Stability Mechanism, Luxembourg, Luxembourg; email: a.kund@esm.europa.eu

Postal address 60640 Frankfurt am Main, Germany
Telephone +49 69 1344 0
Website www.ecb.europa.eu

PDF ISBN 978-92-899-7996-2 ISSN 1725-2806 doi:10.2866/8796354 QB-01-26-196-EN-N