J. Finan. Intermediation 53 (2023) 101015

Contents lists available at ScienceDirect

Journal of Financial Intermediation

journal homepage: www.elsevier.com/locate/jfi

The disciplining effect of supervisory scrutiny in the EU-wide stress test[☆,☆☆]

Christoffer Kok[a,1], Carola Müller[b,c,*,1], Steven Ongena[d,e,f,g,h,1], Cosimo Pancaro[a,1]

[a] European Central Bank, 60640 Frankfurt am Main, Germany
[b] Center for Latin American Monetary Studies, Durango 54, 06700 Ciudad de Mexico, Mexico
[c] Halle Institute for Economic Research, Kleine Märkerstrasse 8, 06108 Halle (Saale), Germany
[d] University of Zurich, Rämistrasse 71, 8006 Zürich, Switzerland
[e] Swiss Finance Institute, Walchestrasse 9, 8006 Zürich, Switzerland
[f] KU Leuven, Oude Markt 13, 3000 Leuven, Belgium
[g] CEPR, 33 Great Sutton Street, London, United Kingdom
[h] NTNU Business School, Postboks 8900, NO-7491, Trondheim, Torgarden, Norway

ARTICLE INFO

JEL classification:
G11
G21
G28

Keywords:
Stress testing
Credit risk
Internal models
Banking supervision
Banking regulation

ABSTRACT

Relying on confidential supervisory data related to the 2016 EU-wide stress test, this paper presents novel empirical evidence that supervisory scrutiny associated to stress testing has a disciplining effect on bank risk. We find that banks that participated in the 2016 EU-wide stress test subsequently reduced their credit risk relative to banks that were not part of this exercise. Relying on new metrics for supervisory scrutiny that measure the quantity, potential impact, and duration of interactions between banks and supervisors during the stress test, we find that the disciplining effect is stronger for banks subject to more intrusive supervisory scrutiny during the exercise. We also find that a strong risk management culture is a prerequisite for the supervisory scrutiny to be effective. Finally, we show that a similar disciplining effect is not exerted neither by higher capital charges nor by more transparency and related market discipline induced by the stress test.

1. Introduction

Since the financial crisis, stress tests have become an important supervisory and financial stability tool and have been used for different goals. During and in the immediate aftermath of the financial crisis, stress tests were used mainly as crisis solution tools aiming at identifying capital shortfalls in the banking sector and enhancing market discipline through the publication of consistent and granular data on a bank-by-bank level. In more recent years, stress tests have rather served the purpose of crisis prevention, thus aiming to identify vulnerabilities in the financial system and to assess the resilience of the banking sector and individual banks to adverse macro-financial shocks, thereby informing supervisory evaluations and contributing towards macroprudential policy discussions. Against this background, several studies document that stress tests contribute to financial stability by promoting risk reduction in the banking sector (Acharya et al., 2018; Cortés et al., 2020; Calem et al., 2020).

In this paper, we contribute by exploring different channels through which stress tests can have a disciplining effect on the banking sector. We are able to measure the effect of each of them. More specifically, we study how stress tests can lead to a reduction of risk by exploiting bank-level variation of stress test intensity along these channels. The set-up of the European stress tests thereby offers a testing ground particularly suited to explore the role of supervisory scrutiny. European stress tests often involve interactions between banks and supervisors on banks' risk management as well as confidential communications about best stress-testing practices and techniques. As such stress tests require that many resources have to be invested in these activities both on the side of the supervisors as well as of the supervised firms. However, mostly due to the confidentiality of supervisory actions, we know little about the effectiveness of these efforts. Do risk management capabilities built up for compliance purposes spill over into bank real outcomes? Has supervisory scrutiny an effect on bank risk? We address these questions by studying what we call the *supervisory scrutiny channel*.

Apart from the supervisory scrutiny channel, we distinguish two additional channels through which stress tests can impact bank risk. First, stress tests are often used to calibrate additional capital requirements.

☆ All errors are our own. This paper should not be reported as representing the views of the European Central Bank or any affiliated institution.

☆☆ Funding: Steven Ongena acknowledges financial support from ERC ADG 2016 - GA 740272 lending.

* Corresponding author.
E-mail addresses: christoffer.kok@ecb.europa.eu (C. Kok), cmuller@cemla.org (C. Müller), steven.ongena@bf.uzh.ch (S. Ongena), cosimo.pancaro@ecb.europa.eu (C. Pancaro).

[1] All authors have contributed equally to all aspects of this work.

https://doi.org/10.1016/j.jfi.2022.101015
Received 17 September 2021; Received in revised form 8 November 2022; Accepted 13 November 2022
Available online 24 November 2022
1042-9573/© 2022 Elsevier Inc. All rights reserved.

As capital is an important determinant of banks' risk choices (Berger and Bouwman, 2013), increases in capital can have a disciplining effect. We refer to this as the *capital channel*. Second, the disclosure of stress test results contains potentially new information about banks' vulnerability and, thus, plays an important role in improving market discipline and restoring confidence in the banking system. Indeed, markets react to the publication of stress test results (Georgescu et al., 2017; Fernandes et al., 2020; Ponte Marques et al., 2022) and other transparency exercises which effectively lowered risk for investors (Klein et al., 2021). We refer to this as the *market discipline channel*.

Our results suggest that stress tests are not merely a *check-the-box* regulatory constraint but rather able to affect bank risk. We provide evidence that stress testing can have a disciplining effect on bank risk via the supervisory scrutiny channel. Moreover, our results confirm that supervisory stress tests interact with banks' risk management. We show that a strong risk management culture is a prerequisite for the supervisory scrutiny to be effective. On the other hand, we cannot find a comparable disciplining effect stemming from the capital and the market discipline channels. Our results intrinsically depend on the design of stress tests in Europe which we exploit to disentangle between the different channels and their impact.

We focus on the EU-wide Stress Test conducted in 2016 by the European Banking Authority and the European Central Bank. Our analysis has two main steps. First, we corroborate existing evidence from U.S. stress tests showing reduced bank risk after testing. We do so by using difference-in-differences approach which compares the change in bank risk between 2015 and 2017 around the 2016 EU-wide stress test exercise of banks that were tested and banks that were not tested. Since tested banks were not selected randomly from a homogeneous population as it would occur in a perfect experimental setting, but instead were chosen according to their status of being systemically important,[2] we control in our analysis for the selection based on observables that define systemic importance and implement a series of robustness tests.[3] We provide further evidence indicating that tested banks achieve the reduction in credit risk by rebalancing their loan portfolios towards less risky borrowers, especially by cutting unsecured lending to small and medium sized firms. Similar findings were obtained in Acharya et al. (2018) and Cortés et al. (2020) although the latter as well as several other studies show that effects on aggregate loan supply might be neutralized by non-tested banks (Berrospide and Edge, 2019; Connolly, 2021).

Then, we study which one of the channels leads to the estimated reduction in credit risk. To do so, we exploit the variation between stress-tested banks in how intense the exercise was for each bank in terms of supervisory scrutiny, market discipline, and ensuing capital guidance. In order to test these possible channels, we extend our baseline difference-in-differences setting including a triple interaction term which captures the intensity of the treatment, i.e., the intensive margin of being stress tested. To avoid the selection problem mentioned above, we estimate the channels only within the set of stress-tested banks.

The European design offers a good testing ground to highlight the effect of supervisory scrutiny in contrast to other channels, most notably from the effects stemming from stress-test induced capital measures. By controlling for changes in capital requirements that are related to the stress test results,[4] we are able to disentangle their effects from those caused by the tighter scrutiny due to the stress test. We can further distinguish supervisory from market scrutiny by using a difference in the publication strategy which groups stress-tested banks in two samples, one with more and one with less transparent stress test results. The use of confidential data on documented interactions between supervisors and tested institutions allows us to construct direct metrics of the intensity of scrutiny exerted by supervisors on banks' stress testing projections and models during the exercise. These interactions arise because the EU-wide stress test exercises follow a constrained bottom-up approach which foresees a thorough Quality Assurance (QA) process carried-out by the European Central Bank. In this context, banks use their own internal models to generate projections conditional on a common macro-financial scenario and subject to a pre-set methodology. Meanwhile, banks' projections are challenged by the competent supervisory authorities to ensure their prudence and credibility typically by applying top-down models and other challenger tools. During this process, which lasts several months, if a bank's internal model-based projections materially deviate from the supervisor's projections, a process to discuss and possibly revise them is launched.

In this context, we use the data collected during the QA process to construct three measures of the intensity of the interactions that took place between supervisors and banks. The three measures are: the quantity of interactions, the potential impact of the interactions on the stress test results, and the duration of the interactions. It should be noted that supervisors do not choose how often to interact. Instead, an automated comparison of supervisory models outputs to the banks' models outputs generates interaction issues that have to be followed upon by the supervisors.[5]

We find that banks that had more subjects to interact upon with supervisors reduced their credit risk more than banks with fewer matters. Furthermore, banks that endured more than one repetition of revisions exhibit lower credit risk than peers going through shorter periods of discussions. Furthermore, we test the existence of a connection between supervisory scrutiny and banks' risk management culture. For this, we use two proxies of risk management culture: The Risk Management Index from Ellul and Yerramilli (2013) and a proxy of safety-oriented risk culture based on Fiordelisi and Ricci (2014).[6] Both proxies try to capture how important risk management and cautiousness are in a banks' culture or banks' governance. We find that supervisory scrutiny

[2] Significant institutions are defined as those SSM banks that (i) have more than EUR 30 bil. in total assets, (ii) are of economic importance for a specific country or the EU economy, (iii) have more than EUR 5 bil. in total assets and cross-border exposures above 20 percentage points than their total assets, or (iv) have requested or received funding from the ESM or EFSF (ECB, 2019). Less significant institutions, from which we sample the control group, are SSM banks that do not fulfil any of the significance criteria to be qualified as significant institutions. Less significant institutions are not under the direct supervision of the ECB. They are directly supervised by the National Competent Authorities under the oversight of the ECB which ensures the consistency of the regulatory framework and supervisory practices applied to these banks.

[3] First we re-estimate our baseline specification by gradually excluding the smallest banks in the control group and the largest banks in the treatment group allowing the size of the banks included in the two groups to progressively converge. Second, we employ the bias-corrected matching estimator of Abadie and Imbens (2011) and we exploit two matching strategies in line with those used in Gropp et al. (2019) which allow to balance the possible differences between the treatment and the control groups.

[4] In contrast to the U.S., the results of the EU-wide stress test serve only as one of the inputs used by the ECB to inform the calibration of the bank-specific capital guidance. Therefore, there is no mechanic relation between stress tests and capital requirements making it easier to disentangle the capital and the scrutiny channel. Further, we have the confidential information of bank-specific capital guidance.

[5] We provide a detailed description of the process to show that the intensity of supervisory scrutiny is not driven by ex-ante bank risk. We further provide some evidence that these scrutiny measures are not correlated to bank risk characteristics.

[6] The Risk Management Index measures the strength and independence of risk management functions in a bank, especially the standing of the Chief Risk Officer and the activity of the Risk Committee. Safety-oriented risk culture is measured as the share of risk-avoidant language banks use in their annual reports viewing the language chosen as an expression of their corporate culture.

is more effective in banks with strong risk management culture. All in all, these findings provide novel evidence that the tighter and more intrusive supervisory scrutiny associated to the EU-wide stress-test interacts with banks' risk management culture and has the potential to induce lower bank risk.

Regarding the capital channel, we fail to establish that banks that received higher stress-test-informed capital requirements reduced credit risk more than their less heavily levied peers. This finding contrasts with the evidence from the U.S. where the capital channel seems highly relevant (Acharya et al., 2018). This result, however, reflects the fact that the capital channel is by design less central in the European exercise that we study in this work. The results of the 2016 EU-wide stress test were only used to inform supervisory decisions about capital guidance. Therefore, our analysis does not rule out that increases in capital requirements could lower credit risk in different settings. Instead, the less deterministic relationship between capital requirements and stress test results makes the European stress test an ideal testing ground to study the otherwise more hardly distinguishable supervisory scrutiny channel.

Finally, to test the market discipline channel, we use a variation in the degree of disclosure of published stress test results that parts tested banks into two groups: (i) Banks that were tested under the mandate of the EBA for which detailed granular bank-by-bank results were published and (ii) banks that were tested under ECB mandate for which only aggregate results were published that did not allow to extract bank-specific information. Against this background, we expect market discipline to be stronger for banks on which investors receive more information. However, we cannot find evidence that banks whose results were published and disclosed at a granular level reduced credit risk more than those banks whose individual results were not published. Yet, Ahnert et al. (2020) and Fernandes et al. (2020) show that investors of tested banks react to the publications of stress test results, especially when news are negative. Motivated by these findings, we further test whether risk decreases more for those banks for which relatively severe results were published using several measures of what could be considered a negative outcome. For this test, we focus on EBA-banks only to avoid confounding the effect of transparency with other differences. But again, we cannot establish a significant differential effect.[7] In line with our results, Flannery et al. (2017) also find no significant change in bank portfolios towards riskier assets although stock returns and CDS spreads react to stress test results. Hence, we cannot find evidence that the market disciplining channel plays a significant role in inducing banks to reduce credit risk in our setting. Overall, our evidence is consistent with a strong role of the supervisory scrutiny channel in the context of the European stress test.

As noted earlier, the evidence that we present in this work depends on the specific design of the 2016 EU-wide stress test and therefore the implications drawn from it need to be contextualized. Because we are only considering a single stress test, the results cannot readily be applied to other stress tests given the different objectives and features of exercises across jurisdictions and time. We thus highlight the 2016 exercise as one that allows us to study the effect of supervisory scrutiny. In this light, the tests of capital guidance and market discipline provide robustness to the identification of the effect of supervisory scrutiny since we can show that none of these alternative channels is mixed with the effect of intense supervision. However, these tests do not provide evidence that capital guidance or market discipline could not have an important impact on bank credit risk in a different context.

Furthermore, the set-up and data of our analysis have limitations. We conduct the analysis on the bank level using fixed effects at the country-time level to control for credit demand. Hence, we cannot completely rule out that the results are affected by changes in credit demand although it is unlikely that these vary systematically in accordance with our measures of supervisory scrutiny. In addition, we cannot fully dismiss the possibility that events other than the stress test in 2016 are affecting our results.

Our study contributes to the literature in several ways. First, we contribute to the growing literature evaluating the effect of stress testing on bank risk. Acharya et al. (2018) find that stress tested banks in the U.S. increase spreads on loans and decrease credit supply especially in riskier market segments using syndicated loan data. Cortés et al. (2020) find very similar results for small business loans. Calem et al. (2020) provide evidence that stress tests led to a reduction in credit supply in the mortgage market which is arguably a riskier part of banks' loan portfolios. On the other hand, Flannery et al. (2017) find no evidence for risk shifting in banks' portfolios after stress test publications. All of these studies rely on U.S. data which we complement by providing evidence on the effects of stress testing on bank risk in Europe. More importantly, none of the works above investigate the channels through which stress tests can have these effects on banks.

In contrast to the U.S. stress test, the European stress tests do not automatically trigger higher capital requirements. Unlike other papers, this allows us to relate a risk reduction effect to specific stress test channels. Only Acharya et al. (2018) mention that their finding is driven by the channel of higher capital although they do not explicitly distinguish between a general effect attributable to the stress test per se and a stress-test induced capital effect.[8] Also Pierret and Steri (2020) emphasize the importance of controlling for capital requirements when measuring the effect of supervision. They find that participating in a stress test decreases banks' risk-taking when controlling for the stress-test induced increase in capital. However, contrary to our work, they do not have a direct measure of supervisory scrutiny in the stress test and do not distinguish between supervisory and market scrutiny. Our precise measures allow us to separate the supervisory scrutiny channel from the capital channel and market discipline channel. We show that stress tests can have effects on bank behaviour that are distinct from those of an elaborately calibrated capital requirement and persist even in absence of capital measures or market vigilance. Furthermore, we show that not external control but the quality of internal control mechanisms in the form of strong risk management culture are essential to ensure the effectiveness of supervisory scrutiny in stress tests. We are the first paper to show a connection between supervisory stress testing and banks' internal risk management functions.

With these results, we also contribute to the literature on the effectiveness of banking supervision and the interplay between the Basel Pillars, i.e., between capital adequacy, supervisory review, and market discipline. Our findings complement several papers that provide micro evidence on the existence of a significant link between supervision and bank risk by exploiting the variation in the intensity of supervisory scrutiny. For example, Buch and DeLong (2008) show that banks shift risks away from countries with strong supervision. Kandrac and Schlusche (2021) exploit an exogenous reduction in bank supervision, measured by the presence of supervisors' offices, to prove a causal effect of supervisory resources on financial institutions' willingness to take risk. Hirtle et al. (2020) using a matched sample approach, find that top-ranked banks that receive more supervisory attention, measured by the hours worked at supervised banks, hold less risky loan portfolios and are less volatile and less sensitive to industry downturns, but do not have slower growth or profitability. Rezende and Wu (2014) find that more frequent inspections increase profitability by decreasing loan losses and delinquencies suggesting that supervisors limit the risks that banks are exposed to and, consequently, limit banks'

[7] An extended analysis of the market discipline channel and how it interacts with the capital channel can be found in Konietschke et al. (2022).

[8] Indeed, they view stress tests as "essentially forward-looking capital requirements". Accordingly, the arguments they provide for the risk management hypothesis rely mostly on an increase in capital initiated by the stress test.

losses on risky assets. Bonfim et al. (2022) find that an inspected bank becomes less likely to refinance zombie firms, immediately spurring their default. Passalacqua et al. (2020) demonstrate positive effects of supervisory inspections on loan portfolio composition of Italian banks. All these studies unanimously advocate for a disciplining effect of supervisory scrutiny. However, none of the works above, study the effects of the supervisory scrutiny carried out in connection with stress tests as instead is done in our paper. Additionally, none of these studies have information about the interactions between supervisors and banks. By focussing on interactions discussing credit risk positions, we are able to link the content of supervisory scrutiny and the bank outcome. Furthermore, to the best of our knowledge, we are the first to contribute evidence of the importance of a strong risk management culture for the effectiveness of supervisory scrutiny to work.

Finally, we contribute to the literature linking internal risk management practices and bank supervision. So far, most attention has been focused on the potential drawbacks of allowing internal models for regulatory purposes. Critiques argue that the use of internal models might give banks too much leeway for regulatory arbitrage. Evidence has been collected on the strategic usage of internal risk models under the Internal Ratings Based approach for the calculation of regulatory capital requirements (Mariathasan and Merrouche, 2014; Behn et al., 2022; Begley et al., 2017; Plosser and Santos, 2018). With respect to the use of internal models in stress testing, Niepmann and Stebunovs (2018) point out that banks misuse the bottom-up design of the EU-wide exercise to strategically adjust their models to improve their loan loss projections. These views reflect a common idea, formalized in Leitner and Yilmaz (2019), that banks optimize one model for regulatory purposes while using another model for their risk management processes and decision making. In contrast, our results indicate that outcomes from compliance processes do seem to impact banks' decisions when they have a strong risk management culture. The rationale of allowing banks to use internal models is to exploit their superior knowledge about their own risks, to create incentives for investing in risk management and the establishment of best practices. Our work can be seen as providing evidence that in the stress testing context relying on bank internal models and thus on a bottom-up approach might not necessarily be detrimental as far as banks' results are subject to an intensive Quality Assurance process as in the EU-wide stress test.

The paper is structured as follows. Section 2 illustrates the possible mechanisms through which the scrutiny exerted by supervisors during the QA process might affect bank risk. Section 3 gives an overview of the institutional setting of the EU-wide stress test in 2016. Section 4 describes the data sources and the final sample we use in the analysis. In Section 5, we describe the estimation methodology, the variables we employ, and the metrics we construct to measure the intensity of the supervisory scrutiny. In Section 6, we show our baseline result on the impact of being stress tested on bank risk. Section 7 contains the analysis on the supervisory scrutiny channel, which can explain how stress testing can affect bank risk. It also reports the results for the capital and the market discipline channels. Finally, we conclude our arguments in Section 8.

2. Hypotheses about supervisory scrutiny and bank risk

In this section, we briefly describe the mechanisms through which the scrutiny exerted by supervisors during the stress test Quality Assurance process might affect bank risk. We also illustrate how this supervisory scrutiny channel is connected to specific features of the EU-wide stress test.

Recent literature provides several plausible explanations for the disciplining effect of banking supervision. First, supervision might improve risk management and bank governance practices. Second, it might produce relevant information about risks and malpractices that lead to corrective actions.

Supervision is based on interactions between supervisors and supervised entities in the form of information exchange (reporting), communications and meetings, as well as on-site inspections and off-site monitoring. The assessment of risk management practices and governance structures is part of the agenda of supervisors. Hence, higher scrutiny exerted by supervisors may impact bank governance. For example, Passalacqua et al. (2020) show that banks are more likely to change board members and hire more workers in supervision and control units after supervisory inspections. Hirtle et al. (2020) point out that these interactions might soften principle-agent problems between risk managers and risk takers within banks. Enhanced supervision might strengthen incentives favouring more conservative risk attitudes aligned with supervisory views and lead to a reduction in risk-taking. Supervisory requests for information may also cause banks to invest in data and technology systems that then enable them to manage their business more efficiently and prudently over the long run. Furthermore, as supervisors oversee many banks, they may transmit knowledge of best practices in the industry when they set expectations and provide feedback to banks about their risk management practices leading to an overall improvement of these practices. Finally, increased supervision might reduce misconduct risk and contribute to a different risk culture (Chaly et al., 2017).

The more intrusive supervision gets, the more weight supervisors can acquire as bank stakeholders representing the public interest in financial stability. Clearly, such abstract changes in power structures cannot be easily observed. Evidence suggests that the mere act of supervision – without the researcher's further knowledge about the content of this supervision – appears to be effective. Several studies document a disciplining effect of more intense supervision. They measure the intensity of supervision by the mere presence of supervisors' offices (Gopalan et al., 2021; Kandrac and Schlusche, 2021) or their hours worked at a supervised banks (Eisenbach et al., 2016; Hirtle et al., 2020).

Furthermore, enhanced supervisory scrutiny can produce new information by detecting unrecognized or unattended risks and misconduct that can lead to corrective actions. Supervision often demands an exchange of information and entails substantial reporting requirements. Corrective actions may be taken voluntarily, upon supervisory recommendation or in response to sanctions. In any case, they should result in a more prudent management of the unveiled risks or cessation of malpractice. Several studies corroborate a disciplining effect of targeted supervisory scrutiny (Ivanov and Wang, 2019; Bonfim et al., 2022; Delis and Staikouras, 2011).

Several aspects of the EU-wide stress tests as conducted within the SSM imply a tighter supervisory scrutiny. One aim of supervisory stress tests is to improve risk management practices.[9] Their mandatory use for regulatory purposes requires banks to invest resources in the development of stress testing techniques, especially in case of bottom-up stress tests where banks themselves have to estimate their own models to generate their projections. In fact, one objective of European regulators to use the constrained bottom-up approach is to foster risk management. As we discuss in detail in the next section, EU-wide stress tests induce interactions on banks' stress testing models and projections between supervisors and supervised banks which might strengthen banks' incentives to improve their risk management strategies. Furthermore, EU-wide stress tests allow the generation and collection of a high amount of new quantitative information. For example, in the 2016 EU-wide stress test banks had to fill-in 35 templates.[10] This additional

[9] Bottom-up stress tests in Europe are an important tool to strengthen banks' risk management (Enria, 2019; Guindos, 2019). Further, former Fed Governor Tarullo stressed in a speech the importance of combining the quantitative and qualitative assessment which includes scrutiny of risk management in the annual Comprehensive Capital Analysis and Review (CCAR) (Tarullo, 2016).

[10] Only a portion of this information is published by the EBA. Still, published information amounts to about 16.000 data points per bank (EBA, 2016b).

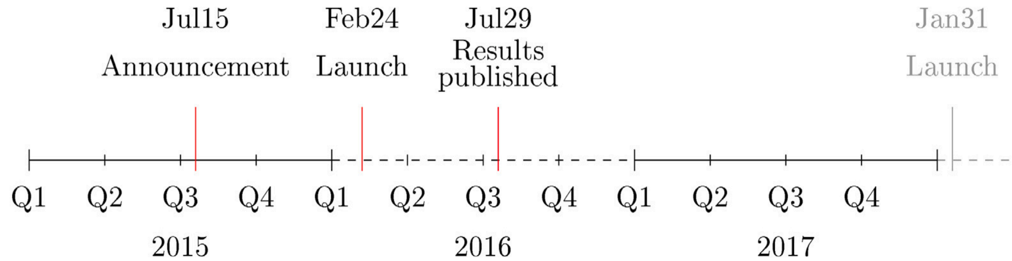


Fig. 1. Timeline of the EU-wide 2016 Stress Test. *Notes*: Solid line segments show quarters in the pre-period (2015Q1–2015Q4) and in the post-period (2017Q1–2017Q4). Dashed line segments show quarters which are excluded (2016Q1–2016Q4 and 2018Q1 onwards).

information facilitates the identification of bank vulnerabilities and, thus, the implementation of possible follow-up actions by the supervisors. Banks themselves might also benefit from insights gained during the stress test implementation to carry out more prudent risk strategies.

These arguments underline that an enhanced supervisory scrutiny is indeed associated to the implementation of EU-wide stress tests for SSM banks. Hence, our hypothesis is that stress-tested banks that were under tighter supervisory scrutiny due to the stress test would show lower risk after the stress test exercise.

3. The 2016 EU-wide stress test

The EU-wide stress test is a complex exercise involving several stakeholders. It is initiated and coordinated by the European Banking Authority (EBA) in cooperation with the European Systemic Risk Board (ESRB), the ECB and national competent authorities in line with the EBA regulation.[11]

It is conducted following a constrained bottom-up approach. Under this approach, banks generate stress test projections using their own models, relying on a common predefined macro-financial scenario[12] and subject to a pre-set methodology. Against this background, banks have to fill-in and submit a number of pre-defined templates prepared by the European Banking Authority in cooperation with the National Competent Authorities. These are structured along risk categories and accounting items. Banks are required to fill-in these templates with for example their credit risk, net interest income and market risk projections over the stress test horizon. Finally, the templates track the impact of these projections under the two common scenarios on bank capital ratios. The final bank level results of the stress test are often summarized by the bank Core Equity Tier 1 (CET1) ratio at the end of the stress test horizon and by the capital depletion under the two scenarios.

Furthermore, in this exercise, the ECB quality assures the stress test results of the banks under its direct supervision. During the QA process the ECB, as a competent authority, reviews and challenges banks' projections and models to ensure their plausibility. The ECB first assesses the compliance of banks' submissions with the constraints imposed by the EBA methodology. Second, it assesses the credibility of banks' submissions by comparing them with the projections produced by the ECB top-down models and with the projections submitted by peer banks. The QA is a thorough process lasting several months over three cycles which, within the ECB, benefits from the contribution of various teams composed of financial stability economists, horizontal supervisors and the direct supervisors of the Joint Supervisory Teams (JSTs). At the end of the first QA cycle, banks receive reports providing them with detailed assessments of their submissions and informing them of any material deviations, called QA flags,[13] between their own projections and the ECB challenger views and are asked to "comply or explain". This implies that in the presence of material deviations, banks are asked to provide quantitative and qualitative evidence on their modelling and supporting their own projections. In the last QA cycle, if the deviations persist and banks' explanations are not deemed sufficient, banks are asked to "comply" with the supervisory challenger view. Overall, the QA process involves extensive interactions between different counterparties, a substantial amount of resources and implies a tight and relatively intrusive supervisory scrutiny.[14]

The 2016 EU-wide stress test, which is the exercise taken into consideration by the analysis conducted in this paper, was first announced in July 2015 and was then officially launched by the EBA on February 24th 2016 with the publication of the common macroeconomic scenarios and methodology. The QA process was conducted between banks' first submission which took place in April 2016 and end-July 2016. The 2016 EU-wide stress test officially ended with the announcement of the results on July 29th 2016 (EBA, 2016a). The sequence of the events is illustrated in Fig. 1.

Overall, 93 SSM Significant Institutions (SI) participated in this exercise. Among these, 37 banks were part of the EBA stress test sample which was overall composed of 51 European banks and included banks that accounted for a share of over 70% of bank assets in Europe (EBA, 2016c). Additional 56 banks, which were also SSM significant institutions but were below the EBA threshold for asset size, also participated in the stress-test as part of their Supervisory Review and Evaluation Process (SREP). A key difference between the EBA and SREP sample is that only the stress-test results of the banks, which were part of the EBA sample, were published at a high level of granularity while the results of the SREP banks were generally not disclosed at individual bank level.[15]

Overall, the 2016 EU-wide stress-test results showed that SSM banks improved their resilience to adverse macroeconomic developments with respect to the 2014 when the previous EU-wide stress-test had been carried out. More specifically, the 2016 results showed that, under the adverse scenario, the 37 SSM banks in the EBA sample would experience on average a CET1 ratio depletion of 3.9 percentage points resulting in a final CET1 ratio of 9.1.

The 2016 EU-wide stress test did not contain a pass and fail CET1 ratio threshold, however, its results fed into the 2016 Supervisory Review and Evaluation Process (SREP) decisions (ECB, 2016a,b). The 2016 SREP process consisted for the first time of two parts: Pillar 2

[11] Regulation No 1093/2010 of the European Parliament and of the Council of 24 November 2010 establishing a European Supervisory Authority (European Banking Authority).

[12] The ECB provides the macroeconomic baseline scenario and contributes to the design of the adverse macroeconomic scenario in cooperation with the ESRB.

[13] The process of generating QA flags is automated and is conducted after the implementation of a comprehensive set of data quality checks. The QA flags are first reviewed and assessed by the ECB stress test teams and only those which are deemed meaningful are effectively shared with the banks.

[14] See Mirza and Zochowski (2017) for further details on the functioning of the ECB QA process.

[15] Only some of these banks decided to voluntarily disclose their stress test results at a very low level of granularity.

capital requirements and Pillar 2 capital guidance.[16] In this context, the fall in the CET1 ratio a bank faced between its starting point at end of 2015 and 2018 under the adverse stress test scenario, was one of the input factors for the calibration of Pillar 2 guidance. However, in defining Pillar 2 guidance, the ECB used also other information, e.g. the specific risk profile of the individual institution and possible measures taken by the bank to mitigate risk sensitivities, such as relevant asset sales, after the stress-test cut-off date. Banks' qualitative performance in the stress test (e.g. the overall quality of the submitted stress test data and possible delays in the submission of the stress test data) is taken into consideration in the determination of the Pillar 2 requirements, especially in the element of risk governance.

4. Data sources and sample

In our analysis we exploit two data sources. We use quarterly bank-level data from the confidential Supervisory Banking (SUBA) database[17] available at the European Central Bank for the period between 2015Q1 and 2017Q4. This database comprises information on balance sheet as well as profit and loss accounting items and regulatory capital ratios. Furthermore, we rely on an ECB proprietary data set which contains the data submitted by the banks for the 2016 EU-wide stress test and also provides information on the interactions which took place between the ECB and the stress tested banks during the related QA process. In addition, we use hand-collected information from banks' annual reports for the analysis of risk management culture.

Our treatment group is composed of banks which took part in the 2016 stress test. We have available data for 93 stress tested banks. In terms of geographies, we constrain our analysis to exposures in the European Union. We use SSM banks that did not participate in the stress test as control group. These banks are Less Significant Institutions (LSI). The sample of LSIs for which we have accounting data from the SUBA dataset is limited due to restricted reporting requirements. Indeed, we can construct covariates relying on balance sheet items only for 175 out of 369 banks for which the ECB has some supervisory data available. Furthermore, covariates relying on profit and loss accounts can only be defined for 81 out of 369 banks.

In line with EBA's decision to exclude Greek banks from the stress test exercise due to the precarious situation of the Greek economy at that time, we exclude these banks from the control group. Further we drop banks that were in resolution, took part in a merger, and those that are part of the banking groups which were stress tested within or outside of the SSM (subsidiaries or branches). Hence, we only consider banks at the highest level of consolidation.

For our baseline analysis, we strongly balance the sample according to the availability of all covariates at the consolidated bank level. This reduces our sample to 63 banks in the treatment group, of which 31 are EBA banks and 32 are SREP banks, and 69 banks in our control group totalling to 924 bank-quarter observations.

Table 1 presents descriptive statistics for the covariates of the banks in our sample in the pre-treatment period. We present the statistics separately for banks in the treatment group and banks in the control group. The stress test treatment is not selected randomly. Instead the selection is based on observables, namely the status of being stress tested and thus being a SSM SI. Hence, banks in the treatment group are substantially larger than the banks in the control group. Treated banks have on average 287 bn EUR in total assets (corresponding to a logarithmic value of 25.4) while control banks have on average only 8.8 bn EUR (corresponding to a logarithmic value of 22.2) as shown in the first rows of Table 1. Column (3) of Table 1 shows that this difference in size is statistically significant at a 1% level. It further reveals that stress tested banks are less reliant on retail business, have a significantly lower share of liquid assets relative to total assets, and lower loan loss provisions relative to total loans than the banks in the control group. Further, judging from the difference in means tests reported in column (3) of Table 1, stress tested banks seem relatively similar to the banks in the control group in terms of some indicators as the return on equity, the cost to income ratio, the interest income ratio as well as the voluntary capital ratio. Unsurprisingly, the regulatory capital ratios also do not significantly differ between the two groups, since all banks are subject to the same minimum requirements and the amounts of additional capital buffers for SIs were still relatively small in our period of observation.

Imbens and Wooldridge (2009) point out that the p-values of the difference in means tests can be misleading in large samples and suggest to normalize differences with variances. As a rule of thumb they mention that estimations are able to balance covariates if normalized differences lie within a range of 25 percentage points around zero. Therefore, we also report normalized differences in column (4) of Table 1. According to this rule, our estimations can handle the aforementioned differences for all covariates but for the liquidity ratio and, as expected, the logarithm of total assets. The normalized differences of both variables are outside the range suggested by Imbens and Wooldridge (2009). The normalized difference of the former is equal to −0.404 while the one of the latter is equal to 1.584.

Being aware that the material difference in the size of the banks between the treatment group and the control group could bias our estimates and curtail the comparability of the two groups, we carry out a set of robustness checks regarding both the estimators as well as the sample which are reported in the Online Appendix A1. However, it is also worth noticing that the cross-country standard variation of total assets within the treatment group is quite high. The smallest stress tested bank has a balance sheet size in terms of total assets of 4.2 bn EUR while the median treated bank has a balance sheet size in terms of total assets of 94 bn EUR which is below the largest control bank that has balance sheet size in terms of total assets of 111 bn EUR.

5. Estimation strategies

In this paper we first investigate if banks' participation in the 2016 EU-wide stress test has an attenuating effect on banks' credit risk in subsequent quarters. Then we study if this effect on banks' risk is at least partly due to the supervisory scrutiny prompted by the stress test QA process, or other features of the stress test. Hereafter, we illustrate the empirical strategies adopted to carry-out these analyses.

5.1. Baseline estimation

To investigate whether banks that were stress tested showed a significantly lower credit risk after the stress test than banks that were not stress tested, we rely on a difference-in-differences approach where we use the stress test as a treatment. Accordingly we estimate the following equation:

$$Risk_{i,t} = Post_t \times Tested_i + Bank_i + Country_i \times Time_t + Controls_{i,t-1} + \epsilon_{i,t}. \quad (1)$$

where the dependent variable $Risk_{i,t}$ is the measure of risk for bank i in period t. Our main yardstick for risk is the risk-weighted asset density for credit risk exposures. $Post_t$ is a dummy variable which takes a value equal to 1 in the 4 quarters of 2017 and 0 in the 4 quarters of 2015. In other words, a symmetric window around the event is used, meaning that the four quarters of 2016 during which the stress test was performed are omitted. $Tested_i$ is a dummy variable which takes a value equal to 1 if a bank participated in the 2016 stress test and 0

[16] Pillar 2 requirements are binding and breaches can have direct legal consequences for banks. Pillar 2 guidance is not directly binding and a failure to meet Pillar 2 guidance does not automatically trigger legal action. Nonetheless, the ECB expects banks to meet Pillar 2 guidance. If a bank does not meet its Pillar 2 guidance, supervisors will carefully consider the reasons and circumstances and may define fine-tuned supervisory measures.

[17] The SUBA database contains COREP and FINREP data collected under SSM mandate.

Table 1
Summary statistics for and differences in means between the treatment group and the control group.

		(1) Mean	(2) Std	(3) Diff (T-C)	(4) NormDiff (T-C)	(5) Min	(6) Max
Log(Total Assets)	T	25.358	1.501	3.181***	1.584[a]	22.159	28.305
	C	22.177	1.334			19.090	25.432
Regulatory Capital Ratio	T	0.081	0.037	−0.003	−0.065	0.045	0.297
	C	0.084	0.020			0.045	0.130
Voluntary Capital Ratio	T	0.087	0.046	−0.002	−0.030	−0.021	0.255
	C	0.090	0.066			−0.063	0.463
Retail Ratio	T	1.178	0.231	−0.062**	−0.176	0.592	1.595
	C	1.239	0.264			0.456	1.782
Liquidity Ratio	T	0.054	0.060	−0.064***	−0.404[a]	0.001	0.377
	C	0.119	0.149			0.000	0.747
Loan Loss Provisions Ratio	T	0.001	0.015	−0.019**	−0.173	−0.071	0.089
	C	0.019	0.106			−0.090	1.341
Cost-Income Ratio	T	0.653	0.702	−0.135	−0.060	0.068	9.189
	C	0.787	2.126			0.159	30.835
Return on Equity	T	0.020	0.020	−0.001	−0.013	−0.091	0.078
	C	0.020	0.036			−0.121	0.092
Interest Income Ratio	T	0.722	1.349	0.034	0.016	0.065	18.778
	C	0.688	1.589			−0.002	21.895

Notes: The table shows summary statistics of the covariates separately for banks in the treatment group (T) and the control group (C). Column (1) shows the mean, column (2) the standard deviation, column (5) the minimum value, and column (6) the maximum value. Columns (3) and (4) show difference in means tests. Column (3) show the difference in means. Stars indicate significance according to the p-value of a two-sided test for differences in means: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Column (4) shows normalized differences as in Imbens and Wooldridge (2009), i.e., difference in means is normalized with the sum of variances.
[a]Indicates that the normalized difference is outside of the range ±0.25 (which serves as a rule of thumb).

otherwise. $Controls_{i,t-1}$ is a vector of bank-specific controls. $Bank_i$ are bank fixed effects and $Country_i \times Time_t$ is an interaction term between country and time-fixed effects, which is included in the regressions to control for loan demand effects. As banks face considerable demand-driven differences across European countries and at the local level, we use the home country of a bank, i.e., the location of its headquarters, reflecting the fact that banks still earn a considerable share of profits in the country of origin (ECB, 2017).

A number of control variables are included to account for bank-specific characteristics which could affect bank risk.[18] We lag these control variables by one quarter to reduce possible endogeneity concerns. They comprise the *Regulatory Capital Ratio*, which allows disentangling the effects of supervisory scrutiny and capital requirements,[19] and the*Voluntary Capital Ratio*, which is the capital held by banks in addition to the amount required by the regulation and the supervisors. Furthermore, other control variables include the ratio of loan loss provisions over total loans (*Loan Loss Provisions Ratio*) to account for asset quality, the *Cost-Income-Ratio* to measure management capability, the *Return on Equity* as a yardstick for earnings, the share of cash and other liquid assets over total assets (*Liquidity Ratio*) to capture bank liquidity risk, the *Retail Ratio* and the ratio of interest income over total asset (*Interest Income Ratio*) as proxies for banks' business models. Finally, bank size is controlled by using the logarithm of banks' total assets (*Log(Assets)*), as this variable is key in determining the selection for the treatment and control groups. Given the inclusion of these controls, we assume that there are no further unobservable time-varying differences between the treatment and control group banks for our analysis to be valid. To answer our question, we are particularly interested in the significance and sign of the estimated coefficient of the interaction term of $Tested_i \times Post_t$.

In order to assess the effects of the stress test on bank risk it would have been ideal if the stress test, i.e., the treatment, had been distributed randomly among a homogeneous group of banks to identify the causal link between the treatment and the changes in banks' risk behaviour after the exercise. Clearly, this was not the case: whether a bank took part in the 2016 EU-wide stress test was determined by its status of being a SI under the ECB direct supervision. Indeed, all banks in our treatment group are SI while for the control group we have to rely on a sample of large LSI. This implies that we cannot claim that our treatment is randomly assigned. Instead, it is assigned based on observables.

Since we know the criteria used for selecting the significant institutions, we can control for the selection based on observables. Matching estimators could also be used to estimate a causal treatment effect (Rosenbaum and Rubin, 1983). However, these estimators cannot account for unobservable differences between treatment and control group that might still influence the outcome variable. Therefore, we use the difference-in-differences approach which also allows us to exploit the panel structure of the data by including bank-fixed effects. Thereby we can eliminate structural time-invariant differences between the two groups. Nevertheless, we also provide results based on Abadie and Imbens (2011) matching estimator as a robustness check.

5.2. The supervisory scrutiny channel

After estimating the baseline model reported in Eq. (1) to assess whether there is an external margin in being stress tested, we continue to investigate the internal margin of being stress tested by defining various measures of intensity of the treatment. More specifically, we focus on exploring the supervisory scrutiny channel. The supervisory scrutiny and interactions between the ECB and the banks, which take place during the stress test, provide information about the variation in the intensity of the QA process across the banks in the sample. This can be exploited as a measure of the intensity of the treatment. Against this background, the following regression is estimated:

$$\begin{aligned} Risk_{i,t} = \; & Post_t \times Tested_i + Post_t \times Tested_i \times High\ ST\ Intensity_i \\ & + Bank_i + Country_i \times Time_t + Controls_{i,t-1} + \epsilon_{i,t} \end{aligned} \quad (2)$$

[18] Table A16 in the Online Appendix provides detailed definitions of the variables.

[19] Capital requirements comprise the sum of Pillar 1 capital ratios as implemented by CRD III and IV, Pillar 2 capital requirements and guidance, as well as macro- and micro-prudential capital buffers. Details are described in Table A16 in the Online Appendix.

where

$$ST\,Intensity = \{QA\ Quantity, QA\ Potential\ Impact, QA\ Duration\}\ (Supervisory\ Scrutiny).$$

We use three different measures of the intensity of supervisory scrutiny. These measures are built relying on ECB proprietary information documenting the flags that were raised during the QA process. We use only flags that were raised due to the comparison of bank submissions to ECB challenger models regarding credit risk. Further, we only regard flags that were communicated to the bank such that interaction between supervisors and banks took place.

The first yardstick, $QA\ Quantity_i$, is the logarithm of the number of credit risk flags which were raised and communicated to the banks during the QA process.[20] This measure is a proxy of the amount of interactions, which took place between the supervisors and the banks during the QA. The second measure, $QA\ Potential\ Impact_i$, is the sum of the potential impact that the QA credit risk flags communicated to the banks could have in terms of CET1 depletion.[21] This yardstick provides a measure of the possible effect of the QA on the final stress test results. Generally, the flags with higher potential impact might entail more discussions and receive more attention. The third measure, $QA\ Duration_i$, is an indicator ranging from 1 to 3 depending on the number of QA cycles during which a bank was communicated a credit risk flag. This indicator reflects the length of the interactions between the ECB and the banks and could be likened to a measure like hours worked per bank as in Hirtle et al. (2020). Two of these three measures are continuous and one is ordinal. Here, we use for our analysis the dummies $High\ QA\ dim_i$, which are equal to 1 for values above the median QA treatment of the respective category (and equal to 0 if below the median). This approach eases the interpretation of the triple interaction term and makes the various results comparable. We add the results where all three measures are treated as continuous $QA\ dim_i$ in the robustness.

Against this background, when assessing the estimates of Eq. (2) we are particularly interested in the significance and sign of the estimated coefficient of the following triple interaction term: $Post_t \times Tested_i \times High\ ST\ Intensity_i$.

5.3. The capital and market discipline channel

In order to test the alternative channels, we follow the same approach described in the previous subsection, but we replace *High ST Intensity* with alternative measures of treatment intensity. For the sake of comparability, we keep using dummy variables for the distinct channels splitting at the median if we are considering continuous measures and estimate Eq. (2) with

$$\begin{aligned} ST\ Intensity &= \{P2G, Effective\ P2G, Voluntary\ Capital\}\ (Capital) \\ ST\ Intensity &= \{Transparency, Severe\ Result\}\ (Market\ Discipline)\,. \end{aligned} \tag{3}$$

To measure the capital channel, we use the information of the Pillar 2 Capital Guidance (P2G) issued the first quarter of 2017 which is when the stress test results where taken into account. As further robustness, we take the ratio of P2G to available voluntary capital as *Effective P2G* to account for banks with high buffers above the regulatory minimum for whom additional P2G might be less relevant compared to banks close to the minimum. We further use the *Voluntary Capital* buffer itself to test for differences between banks with high and low capitalization.

[20] We use log-levels due to the high non-normality displayed by the distribution of the number of flags by banks according to Shapiro–Wilk test.

[21] During the QA process, the deviation between the ECB and banks' projections is calculated automatically for each flag in terms of CET 1 ratio.

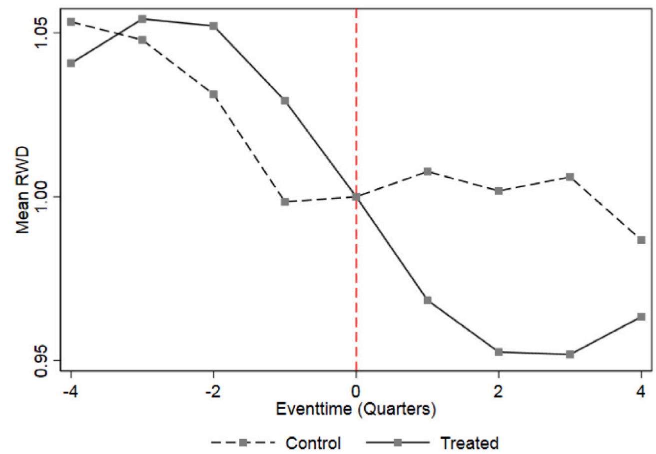


Fig. 2. Time trends of RWD around the 2016 stress test by treatment. *Notes*: The figure shows average RWD of the treatment and control group for each quarter of the pre- and post-period normalized with the average RWD of the respective group during the stress tests quarters which are excluded, i.e., four quarters of 2016 are summarized to eventtime 0. Hence, eventtime −1 corresponds to 2015Q4, eventtime 1 to 2017Q1, and so on. Banks in the treatment group participated in the 2016 stress test. Banks in the control group did not.

To measure the market discipline channel, we construct the *High Transparency* dummy which is equal to one for EBA banks whose individual stress test results were published at granular level and zero for SREP banks whose results were only published at aggregate level. As we assume the market discipline channel to be stronger not only for banks with more disclosed information but especially for banks with negative information, we further use three measures of severe stress test results within the sample of EBA banks, namely (i) banks with the lowest final CET1 ratios in the adverse scenario, (ii) banks with the highest capital depletion (difference between initial and final capital in the adverse scenario), and (iii) banks with the highest depletion relative to their initial CET1 ratios.

6. The effect of being stress-tested on bank risk

In this section, we discuss the results of our difference-in-differences analysis examining whether stress-tested banks change their credit risk level after the stress test relative to banks that were not part of the stress test. In particular, we report the results for the estimates of Eq. (1).

A necessary identifying assumption for this setting to be valid is that the change in outcomes, i.e., the trend in credit risk developments, in the period before the stress test is comparable between the control and treatment group. If the outcome variable is on a comparable trend before the stress test but diverges between the two groups after the stress test, we can attribute this divergence to the execution of the stress test. Fig. 2 illustrates the trend of average risk-weighted density for the treatment and control group around the 2016 stress test. The level of *RWD* was normalized to one for the stress test period in 2016 for both groups. Hence, the figure shows the level of *RWD* in the four quarters before the stress test (*Post ST16* = 0) in 2015 and in the four quarters after the stress test (*Post ST16* = 1) in 2017 for both groups relative to their average 2016 *RWD* level. The Fig. 2 corroborates the findings in Table 2.[22]

[22] For our analysis to be valid, the trends of the dependent variable for the treatment and the control group in the pre-period need to be parallel. Looking at the average slopes in the pre- and post-period as reported in Table 2, we see a significant change in the slope of the variable of interest for the control group banks between the pre- and post-period. As we cannot observe the counterfactual, in the graph we focus on the pre-period and not on the

Table 2
Summary statistics of the dependent variable RWD.

	(1) *Levels*	(2)	(3) Diff.	(4) *First differences*	(5)	(6) Diff.
	Control	Tested	(T-C)	Control	Tested	(T-C)
Pre ST16	0.484 (0.161)	0.437 (0.225)	−0.047** [0.018]	−0.009 (0.033)	−0.002 (0.066)	0.006 [0.247]
Post ST16	0.472 (0.167)	0.398 (0.187)	−0.074*** [0.000]	−0.001 (0.033)	−0.003 (0.026)	−0.002 [0.361]
Diff. (Post-Pre)	−0.012 [0.427]	−0.039* [0.055]	−0.027* [0.073]	0.008** [0.010]	−0.001 [0.897]	−0.009 [0.211]

Notes: Columns (1), (2), (4), and (5) show means and standard deviations in parentheses of RWD for the control group and treatment group before the 2016 stress test (Pre ST16) and after (Post ST16). The bottom row shows the difference in means between the pre and post stress test period and in parentheses the *p*-value of a t-test for differences in means. Columns (3) and (6) show the difference in means between the two groups within the pre or post stress test period and in parentheses the *p*-value of a t-test for differences in means. The bottom row in col. (3) and (6) show the difference in differences and in parentheses the *p*-value of a t-test. Stars indicate significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Columns (4) to (6) of Table 2 show means and differences in means for the quarter-on-quarter change in *RWD*. The first two rows of columns (4) and (5) document that *RWD* was on average decreasing in both groups and during both time periods. Column (6) shows that differences between control and tested banks in the slope of *RWD* in the pre-period are not significantly different from zero. We take this as evidence that the parallel trend assumption is valid.

Furthermore, columns (1) and (2) of Table 2 show the average *RWD* of the treatment and control group in the pre- and post-test periods. The last row indicates that while both groups exhibit lower *RWD* on average in the period after the stress test compared to the average *RWD* before, this difference is only significantly different from zero for the group of stress tested banks. Column (3) further documents that the average *RWD* of stress tested banks is significantly lower than the average *RWD* of control banks in the pre-period as well as in the post-period. Our analysis accounts for this difference in levels by effectively demeaning the outcome variable through the introduction of bank fixed effects. Finally, the bottom row of column (3) shows the unconditional difference-in-differences effect. We find preliminary evidence for our hypothesis that the stress test exercise impacted banks' risk. The coefficient shows that banks that took part in the stress test on average reduced their risk-weighted density subsequently by 2.7 percentage points more than banks that did not participate in the test.

In Table 3, we report the estimated results of Eq. (1). The results reported in column (1) show that the estimated coefficient on the interaction term of our interest is negative and significant once we include bank and time fixed effects. Interestingly, the magnitude of the coefficient is the same as in the univariate analysis shown in Table 2. However, as the treatment is not assigned randomly to banks, we have reasons to believe that this estimation is biased. Hence, in column (2) we expand our specification by including bank size in the form of *Log(Assets)* as a control variable being aware that this is the main variable that drives the selection into the treatment group. We find that size is a relevant determinant of *RWD* levels as its estimated coefficient is significant at the 1% level and the explanatory power of our estimation increases with respect to within-bank variation. Further, the coefficient is negative corroborating the existing evidence that larger banks might pose more systemic risk, but are inclined to take less individual risk (Laeven et al., 2016). Thereby this effect has the same direction as the effect of being stress-tested. Thus, by conditioning on size, the probability that we have to reject our hypothesis that tested banks reduce credit risk after the stress test with respect to non-stress tested banks decreases. In column (3), we add further control variables that might influence bank risk. While the signs on all estimated coefficients are broadly in line with expectations, only two of them are significant. More specifically it results that banks with higher voluntary capital ratios as well as banks with a higher share of liquid assets to total assets show lower RWDs. Additionally, we include an interaction term between fixed effects for the country of banks' headquarters and a dummy for each time period to capture demand conditions which vary at country level in addition to the pan-European macroeconomic developments which are captured by the time fixed effects. This takes into account that even the large international European banks still hold a majority of their credit risk exposures in their country of origin. Admittedly, we are not able to control for demand factors, which influence bank risk and might vary at a more local level, nor for cross-country exposures of these large international banks. Notwithstanding, we consider this as our preferred specification to explain changes in RWD given the data we have available. The results show that even when including this additional interaction term the estimated coefficient of interest remains negative and significant. In particular the results show that the reduction in RWD of tested banks after the stress test was on average 4.2 percentage points lower than the reduction of not-tested banks. This effect is economically material as it amounts to a change in RWD of about 20 percent of the standard deviation of RWD of the tested banks.

A general possible concern for any analysis on bank risk is related to the use of an appropriate yardstick for risk. RWD is based on information reported for regulatory purposes and might not fully represent bank risk. First, credit risk is only a part of overall bank risk. Second, banks might have incentives to underreport risk and manipulate risk weights for regulatory purposes. Indeed, there is evidence of strategic usage of internal risk models under the Internal Ratings Based approach for the calculation of regulatory capital requirements (Behn et al., 2022; Plosser and Santos, 2018; Mariathasan and Merrouche, 2014; Begley et al., 2017). And lastly, reported credit risk exposures might still miss credit risk exposures outside of the reporting framework. However, Baule and Tallau (2021) show that RWD of large European banks is sensitive to asset portfolio risk and captures quite adequately asset volatility, especially in normal times.

In order to address some of these shortcomings, we employ alternative measures related to banks' default probability. In column (4), we show that participating in the 2016 stress test had a negative significant effect on stress test banks' Expected Default Frequencies (EDFs). EDFs are a measure of the probability of default of a bank within the next year provided by Moody's Analytics. In column (5), we show that the z-score, i.e., the distance to default, of tested banks increases relative to that of non-tested banks. We built the z-score of the banks in sample relying on balance sheet data provided by SNL Financials. These data are available only on a yearly basis so we estimate Eq. (1) by averaging all covariates in the pre-test and post-test period. We provide additional robustness tests concerning the assumptions, variable definitions, and

post-period. Anything happening in the post can potentially be caused by the treatment. Therefore, we run regressions in which we control for differences in levels of RWD, differences between banks and other confounding factors.

Table 3
Baseline result of stress test participation.

Dependent:	(1) RWD	(2) RWD	(3) RWD	(4) Moody's EDF	(5) z-score (yearly)
Post ST16 × Treated	−0.027*	−0.035**	−0.042**	−1.275*	0.674**
	(0.015)	(0.015)	(0.019)	(0.640)	(0.270)
L.Log(Assets)		−0.119***	−0.145***	−1.558	0.047
		(0.036)	(0.039)	(1.234)	(1.253)
L.Regulatory Capital			−0.150	−4.958	4.156
			(0.191)	(5.966)	(7.034)
L.Voluntary Capital			−0.254*	−9.696*	3.739
			(0.144)	(5.387)	(3.909)
L.Retail			0.013	1.014	1.421
			(0.059)	(2.461)	(1.645)
L.Liquidity			−0.175**	2.312	0.251
			(0.078)	(3.315)	(1.980)
L.LLP			0.039	6.893	−0.125
			(0.105)	(6.609)	(0.621)
L.CIR			0.001	2.162***	0.092
			(0.003)	(0.549)	(0.100)
L.RoE			0.166	−16.247*	
			(0.207)	(8.122)	
L.Interest Income			−0.001	−2.618***	−0.114
			(0.004)	(0.665)	(0.115)
Bank FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	No	No
Country × Time FE	No	No	Yes	Yes	Yes
Observations	924	924	924	299	212
R2	0.923	0.927	0.942	0.915	0.955
R2-within	0.016	0.069	0.122	0.102	0.116
Tested Banks	63	63	63	32	51
Control Banks	69	69	69	14	55
Mean *Dependent*	0.462	0.462	0.462	1.82	2.073
(SD *Dependent*)	(0.195)	(0.195)	(0.195)	(3.074)	(2.802)

Notes: Clustered standard errors at the bank-level in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Difference-in-differences estimation with a 4-quarters before and after the 2016 stress test with a strongly balanced sample. *Post ST16* is a dummy for 2017Q1–2017Q4. *Tested* is a dummy for stress-tested banks. Bank-level control variables are lagged by one quarter. Time fixed effects are dummies for each quarter. Country time fixed effects indicate the country of each bank's headquarters. In columns (4) and (5) *RWD* is replaced as the dependent variable by a different measure indicated in the column heads. In Column (4) Expected Default Frequencies (EDFs) are provided by Moody's Analytics which measure the probability of default within the next year. In Column (5) we rely on yearly data from SNL Financials. Therefore, we collapse the time dimension in the covariates by averaging over the pre-period and post-period quarters. The dependent variable is *z-score* defined as the difference between Return-on-Assets (ROA) and total capital ratio, both calculated as 3-year rolling averages, relative to the standard deviation of ROA, calculated with all available data until the current period. *Return on Equity* is omitted here as a control variable due to collinearity.

methods of this analysis in the Online Appendix A1. Among these, we show that our results are not biased due to the non-random selection of tested banks nor the differences in size between treatment and control group by using a bias-corrected matching estimator.

These results demonstrate that stress tests contribute to the resilience of tested banks. While this corresponds to the findings for the U.S. in Acharya et al. (2018) and Cortés et al. (2020), these studies also show that tested banks lower credit risk by reducing credit supply to riskier borrowers (see also Acharya et al. (2018), Berrospide and Edge (2019), and Connolly (2021)). Further, Konietschke et al. (2022) provide similar evidence of a contraction in credit and shift towards less risky lending based on the 2016 and 2018 EU-wide stress tests. To assess the overall effect of stress tests, it should be considered that higher resilience might come at the cost of less bank lending to firms. In particular, small and medium sized firms (SMEs) are often the first group affected by cuts in credit (Cortés et al., 2020). As Doerr (2021) shows this can imply negative consequences for innovation on a regional and sectoral level. While a complete analysis of real effects is beyond the scope of this paper, we provide additional results showing that tested banks significantly reduce unsecured lending to SMEs as well as reallocate their lending portfolio away from riskier industries towards more conservative ones. Overall, this evidence is in line with the findings in the literature. Details can be found in Online Appendix A1.

7. How stress testing can affect bank risk

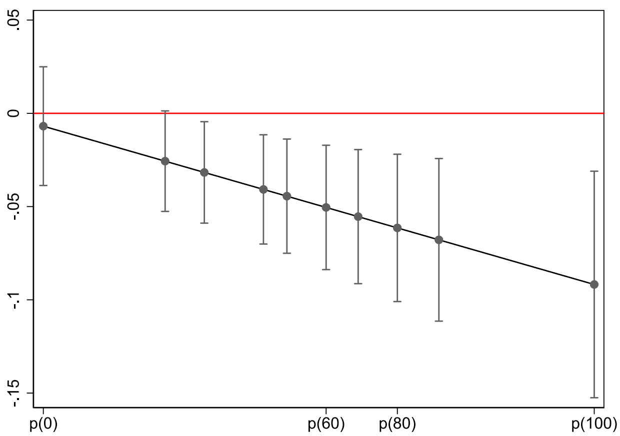
Following on the evidence provided in the previous section that stress testing caused a significant difference in bank risk between banks that participated in the 2016 EU-wide stress test and those that were not stress tested, we now investigate how stress testing might affect bank risk. We accomplish this by exploiting variation in the extent to which banks were exposed to different features of the stress test.

7.1. The supervisory scrutiny channel

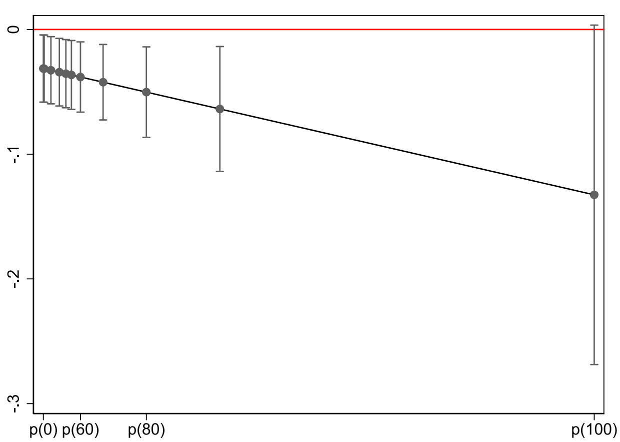
Stress tests are an intense supervisory exercise that last for several months. As illustrated in Section 3 during the stress test QA supervisors review banks' projections and models to ensure their prudency and credibility. In case of material deviations between the projections of the ECB and those of the banks, a dialogue with the banks, which can potentially lead to revisions of the banks' projections, is initiated. The process involves an exchange of views on banks' stress testing strategies and on their risk management practices and generates a vast amount of information on banks' risk profiles. In the following, we first study how supervisory scrutiny is linked to the disciplining effect outlined above and then how it is linked to banks' risk management culture.

As illustrated in Section 5.2, we estimate Eq. (2) using the three aforementioned measures of supervisory scrutiny. The main results are reported in Table 4 and the robustness using the continuous measures of scrutiny are reported in Table A8.

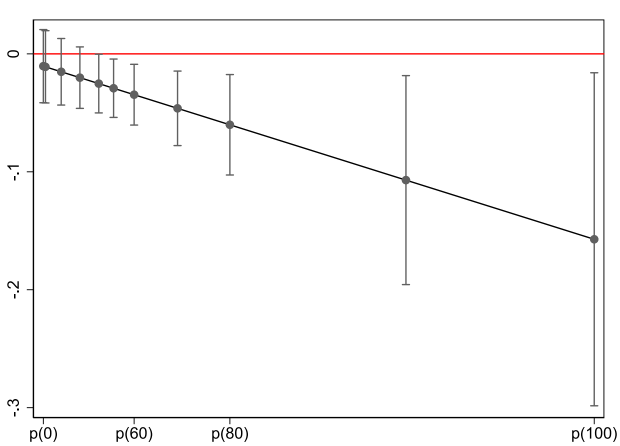
The results in column (1) of Table 4 show that stress tested banks that were exposed to *High QA Quantity*, i.e., received an above median number of flags during the QA process, significantly reduced their risk-weight density after the stress test relative to stress tested banks that received a number of QA flags lower than the median. We estimate that banks decrease RWD by 5.6 percentage points more if they belong to



(a) QA Quantity



(b) QA Potential Impact - Unwinsorized (baseline).



(c) QA Potential Impact - Winsorized at 5%.

Fig. 3. Marginal effects along percentiles of QA measures. *Notes*: The figure shows the marginal effect and 95% confidence interval of being stress-tested while receiving a defined amount of Quality Assurance along the distribution of two different measures of QA. Marginal effects are calculated for the minimum, the 10th, 20th, 30th, etc. percentile to the maximum. Upper Figure (a) shows marginal effects along the distribution of *QA Quantity* used in the estimation of Column (1) of Table 4. The lower two Figures (b) and (c) show marginal effects along the distribution of *QA Potential Impact*. The middle Figure (b) shows marginal effects corresponding to the unwinsorized measure used in the estimation of Column (2) of Table 4. It reveals an outlier in the distribution of *QA Potential Impact*. Hence, the lower Figure (c) shows marginal effects after winsorizing *QA Potential Impact* at 5%.

Table 4
The supervisory scrutiny channel.

	(1) High QA Quantity	(2) High QA Potential impact	(3) High QA Duration
Post ST16 × Tested	−0.014	−0.031*	−0.008
	(0.016)	(0.016)	(0.024)
Post ST16 × Tested × High QA	−0.056***	−0.023	−0.041*
	(0.020)	(0.024)	(0.022)
L.Log(Assets)	−0.151***	−0.144***	−0.144***
	(0.039)	(0.038)	(0.039)
L.Regulatory Capital Ratio	−0.107	−0.162	−0.126
	(0.181)	(0.184)	(0.182)
L.Voluntary Capital Ratio	−0.263*	−0.247*	−0.247*
	(0.135)	(0.143)	(0.142)
L.Retail Ratio	0.025	0.012	0.009
	(0.059)	(0.057)	(0.058)
L.Liquidity Ratio	−0.173**	−0.181**	−0.184**
	(0.079)	(0.079)	(0.079)
L.Loan Loss Provisions Ratio	0.024	0.041	0.043
	(0.106)	(0.106)	(0.105)
L.Cost-Income Ratio	0.002	0.001	0.001
	(0.003)	(0.003)	(0.003)
L.Return on Equity	0.145	0.129	0.191
	(0.189)	(0.211)	(0.196)
L.Interest Income Ratio	−0.002	−0.001	−0.001
	(0.004)	(0.004)	(0.004)
Bank FE	Yes	Yes	Yes
Country × Time FE	Yes	Yes	Yes
Observations	924	924	924
Banks	132	132	132
R2within	0.155	0.126	0.129

Notes: Clustered standard errors at the bank-level in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Difference-in-differences estimation with a 4-quarters before and after the 2016 stress test with a strongly balanced sample. *Post ST16* is a dummy for 2017Q1–2017Q4. *Tested* is a dummy for stress-tested banks. Bank-level control variables are lagged by one quarter. Time fixed effects are dummies for each quarter. Country time fixed effects indicate the country of each bank's headquarters. In Column (1) *High QA* is a dummy indicating above median QA Quantity defined as the logarithm of the number of flags communicated to banks with respect to credit risk. In Column (2) *High QA* is a dummy indicating above median QA Potential Impact defined as the sum of potential impact on CET1 in the adverse scenario of flags communicated to the banks with respect to credit risk. In Column (3) *High QA* is a dummy indicating above median QA Duration defined as two or more cycles.

the High QA group relative to the other stress tested banks. This impact amounts to a differential effect of about 13 percent of their pre-stress test RWD. We carry out this same exercise relying on the continuous measure of *QA quantity*, i.e., the logarithm of the number of credit risk QA flags communicated to the banks during the QA process. The result displayed in column (1) of Table A8 shows that banks reduce RWD by 2.7 percentage points if they receive 1 percentage point more of credit risk QA flags. This 1 percentage point increase corresponds roughly to a quintile in the distribution of the QA quantity measure. The upper panel (a) of Fig. 3 depicts the marginal effects along different percentiles of the distribution of *QA Quantity*. As expected, as intensity gets stronger the effect is stronger remaining significantly different from zero.

The results displayed in column (2) of Table 4 and in column (2) of Table A8 show that the potential impact, in terms of capital depletion, of the QA credit risk flags on the final stress test results (measured relying respectively on the dichotomous and continuous measures) does not seem to matter for the risk reduction in the aftermath of the exercise since both estimated coefficients for the triple interaction term are not significant.[23] This measure should capture the case of a bank having a very intense QA process due to the potentially very severe impact on the stress test final outcome of the received QA flags. A closer

[23] We also used the realized impact and not only the potential impact of the credit risk QA in terms of CET1 depletion to capture the effectiveness of the QA procedure. However, we did not find any significant results.

Table 5
The supervisory scrutiny channel in the subsample of stress-tested banks.

	(1)	(2)	(3)	(4)
Supervisory scrutiny				
Post ST16 × High QA Quantity	−0.040**			−0.040*
	(0.019)			(0.023)
Post ST16 × High QA Pot. Impact		−0.014		0.004
		(0.019)		(0.023)
Post ST16 × High QA Duration			0.029	−0.020
			(0.025)	(0.022)
Bank Controls	Yes	Yes	Yes	Yes
Bank FE, Country × Time FE	Yes	Yes	Yes	Yes
Observations	413	413	413	413
Tested	59	59	59	59
R2-within	0.208	0.182	0.184	0.21

Notes: Clustered standard errors at the bank-level in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Difference-in-differences estimation with a 4-quarters before and after the 2016 stress test with a strongly balanced sample. We use only stress-tested banks in the sample. Due to four singleton groups the sample drops from 63 to 59 banks. *Post ST16* is a dummy for 2017Q1–2017Q4. Bank-level control variables are lagged by one quarter and comprise *Log(Assets), Voluntary Capital Ratio, Regulatory Capital Ratio, Liquidity Ratio, Retail Ratio, LLP Ratio, CIR, RoE*, and *Interest Income Ratio*. Time fixed effects are dummies for each quarter. Country time fixed effects indicate the country of each bank's headquarters. In Columns (1) to (3) we test each one of the QA intensity measures separately. In Column (4) we include all three QA intensity measures in triple interaction terms.

look at the data, however, reveals that this insignificance is driven by an outlier. The middle panel (b) in Fig. 3 shows negative marginal effects that are significantly different from zero for all percentiles of the distribution of *QA Potential Impact* except for the maximum. We cannot disclose the nature of this outlier due to confidentiality restrictions, but when winsorizing at 5%, Fig. 3 shows in the lower panel (c) that banks with a higher materiality of communicated QA flags reduce their risk significantly more than banks with less material QA flags.

In column (3) of Table 4 we further find evidence that a longer duration had a negative significant differential effect. We find that one more round of discussions between regulators and banks results in an additional 2.5 percentage point drop in RWD compared to banks that had no further flags to discuss. Qualitatively similar results are show in column (3) of Table A8.

To further address the concerns addressed in section A1 about the comparability of the control and treatment group, we estimate the scrutiny channel in a subsample that does not include non-tested banks. Hence, we compare only outcomes of stress-tested banks depending on the intensity of scrutiny that they experienced during the exercise. The results in Table 5 show in column (1) that the effect of supervisory scrutiny also prevails within the sample of stress-tested banks. Banks that received more than the median number of flags reduce their RWD by 4 percentage points more than banks that received less than the median. Column (2) and (3) show that the effect for scrutiny is neither significant in the sample with stress-tested banks when measured by *High QA Potential Impact* nor *High QA Duration*.

Additional concerns might be that our measures of supervisory scrutiny are endogenous or capture an effect that stems from another omitted variable. Indeed it might be that institutions with higher credit risk or those more important to supervisors, e.g. larger banks, receive more scrutiny. We provide robustness analyses that in general reject such interpretations in the Online Appendix A3.

In Section 2, we argue that supervisory scrutiny might work through a shift in banks' risk management culture towards risk attitudes which are more aligned with regulatory views. We therefore use two proxies of risk management culture which have been used in the literature to test if there is a connection between the effectiveness of supervisory scrutiny and risk culture.

The first proxy is the *Risk Management Index* (RMI) from Ellul and Yerramilli (2013) which measures the strength and independence of the risk management function in a bank. It uses information on the relative importance of the Chief Risk Officer (CRO) and the supervisory board's risk committee within a bank that can usually be found in banks' annual reports. For example, we collect data on the salary of the CRO relative to the Chief Executive Officer or the number of meetings held by the risk committee to build this index. Ellul and Yerramilli (2013) show that stronger risk management (higher RMI) is associated with lower bank risk.

The second proxy is *Safety-oriented Culture* based on a linguistic analysis of banks' annual reports. Song and Thakor (2019) emphasize how safety-oriented bank culture can reduce agency conflicts and a competition-induced excessive growth focus within banks. We follow (Fiordelisi and Ricci, 2014) in the construction of corporate culture variables for banks. The idea is that corporate culture is expressed in the language of banks' publications. According to the competing values framework (Quinn and Rohrbaugh, 1983; Cameron et al., 2014), specific words of the English language are associated with control-oriented or collaboration-oriented cultures in opposition to competition-oriented or creation-oriented cultures. We scan reports for these control- and collaborate-oriented words relative to the total word count and take this ratio as the proxy. We provide details on how we replicate these methodologies in the Online Appendix A4.

To assess the meaningfulness of these proxies, we test if they behave as expected which indeed they do. We find that stronger risk management culture and more safety-oriented culture are negatively correlated with bank risk in our sample (see Table A15 in Online Appendix A4). This is consistent with the theoretical considerations in Song and Thakor (2019) and mirrors the findings in Ellul and Yerramilli (2013).

For both measures, we analyse banks' annual reports of 2015 (pre stress test) and 2017 (post stress test) giving us only two observations per bank. This precludes us from using any estimation design that depends on within-bank time variation of risk management culture (cf. low within-standard deviation shown in column 7 of Table A14 in Online Appendix A4).[24] Culture by definition changes only slowly. We therefore cannot observe if supervisory scrutiny in the stress test improves risk management culture immediately. Consequently, we choose a research design that uses the cross-sectional variation in bank culture that nevertheless allows us to draw a connection between supervisory

[24] We have to note that we could not find annual reports of 2015 and 2017 for all banks in our baseline sample, that some did not contain the necessary information to construct RMI, and some where not in the English language. As a result, we compute the RMI (Safety-oriented Culture) for 42 (49) out of 63 stress-tested banks and 53 (55) out of 69 control group banks.

Table 6
The dependence of the supervisory scrutiny channel on risk management culture.

Sample:	(1)	(2)	(3)	(4)	(5)	(6)
	High RMI			Low RMI		
Post ST16 × Tested	−0.004 (0.018)	−0.022 (0.023)	0.004 (0.023)	−0.005 (0.026)	−0.018 (0.023)	−0.023 (0.033)
Post ST16 × High QA Quantity	−0.094** (0.038)			−0.020 (0.018)		
Post ST16 × High QA Potential Impact		−0.065 (0.046)			0.005 (0.016)	
Post ST16 × High QA Duration			−0.078** (0.034)			0.008 (0.018)
Bank Controls	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE, Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	96	96	96	94	94	94
R2-within	0.328	0.288	0.289	0.423	0.408	0.408
Sample:	High safety-oriented culture			Low safety-oriented culture		
Post ST16 × Tested	−0.008 (0.022)	−0.021 (0.024)	0.023 (0.021)	0.008 (0.028)	0.006 (0.028)	0.003 (0.034)
Post ST16 × High QA Quantity	−0.118** (0.051)			−0.042 (0.027)		
Post ST16 × High QA Potential Impact		−0.048 (0.047)			−0.047* (0.027)	
Post ST16 × High QA Duration			−0.086* (0.047)			−0.031 (0.030)
Bank Controls	Yes	Yes	Yes	Yes	Yes	Yes
Bank FE, Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	98	98	98	110	110	110
R2-within	0.394	0.259	0.27	0.345	0.355	0.336

Notes: Clustered standard errors at the bank-level in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The table shows the estimation of the supervisory scrutiny channel for three measures of scrutiny specified as in Table 4 in four subsamples. We use yearly averages of the control variables and RWD. In the upper panel, we distinguish banks with strong risk management (Columns (1) to (3)) and banks with weak risk management (columns (4) to (6)) by splitting the sample along the median *Risk Management Index* of banks in 2015. In the lower panel, we distinguish banks with strong safety-oriented culture (columns (1) to (3)) and banks with weak safety-oriented culture by splitting the sample along the median *Safety-oriented Culture* of banks in 2015.

scrutiny and bank risk. Specifically, we split the sample along the medians of both proxies into banks with strong risk management cultures and those with weak risk management cultures and repeat the estimation of the supervisory scrutiny channel as in Eq. (2). The results are reported in Table 6.

The estimated results reveal that the supervisory scrutiny channel operates predominantly within banks with strong risk management cultures. The estimates are consistently higher in columns (1) to (3) where we test the strong risk management culture banks than in columns (4) to (6) where we test weak risk management culture banks. Further, we only find a significant effect of *High QA Intensity* and *High QA Duration* in the sample of strong RMI and strong *Safety-oriented Culture* banks (cf. columns (1) and (3)). We interpret this as evidence that strong risk management culture and supervisory scrutiny interact and that a safety-oriented culture is a prerequisite for supervisory scrutiny to be transmitted to real bank outcomes. It is important to note though that supervisory scrutiny could be crucial to improve risk management over a longer horizon.

Overall, these findings indicate that there is a significant value added in the interactions which take place between the ECB and banks during the QA process of the stress test. Indeed, the fact that banks are asked to explain and/or adjust their modelling strategies and projections in presence of material deviations between their projections and the ECB ones seems to be a relevant factor in influencing banks' risk attitude in the aftermath of the stress test. Interestingly, supervisory scrutiny proofs more effective in banks with strong risk management cultures. This result corroborates the idea that the QA process itself and the intrusion of supervisors into banks' sphere is the channel through which banks are disciplined.

7.2. The capital and the market discipline channels

In this section, as a further analysis, we examine if other drivers in addition to the supervisory scrutiny channel can potentially explain our baseline result. In particular we consider the market discipline channel and the capital channel. The former refers to the discipline potentially imposed by markets on banks as a consequence of the granular disclosure of the stress test results. The latter instead refers to the possible effects on bank risk related to additional capital requirements or capital distribution limits associated to the stress test results as bank capital is an important determinant of banks' risk choices.[25]

To test the hypothesis that banks could react to market discipline exerted as a result of stress tests, we associate high stress test intensity with a high level of transparency. The market discipline channel is related to the transparency's enhancement led by the publication of granular stress test results at the end of the exercise. This enhanced disclosure allows market investors to better price bank risk by providing additional information about banks' possible vulnerabilities and thus reducing information asymmetries. Accordingly, various studies show that the disclosure of stress test results has an effect on banks' stock returns and CDS spreads (Petrella and Resti, 2013; Morgan et al., 2014; Flannery et al., 2017; Georgescu et al., 2017; Lazzari et al., 2017; Ahnert et al., 2020; Fernandes et al., 2020).

[25] Acharya et al. (2018) detail at least four channels through which stress-test related capital measures might affect bank risk: (i) mechanical connection through risk-weighted capital requirements, (ii) moral hazard channel, (iii) charter value channel, and (iv) reach-for-yield channel.

Table 7
The market discipline channel.

Sample:	(1) Baseline	(2) Tested	(3) EBA	(4) EBA	(5) EBA
Post ST16 × Tested	−0.031* (0.018)				
Post ST16 × Tested × High Transparency	−0.026 (0.029)	−0.014 (0.025)			
Post ST16 × Tested × Low CET1(Adverse)			−0.012 (0.024)		
Post ST16 × Tested × High Depletion				−0.015 (0.018)	
Post ST16 × Tested × High Depletion/CET1(Initial)					0.015 (0.018)
Bank Controls	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes
Country × Time FE	Yes	Yes	Yes	Yes	Yes
Observations	924	413	210	210	210
R2-within	0.126	0.181	0.179	0.18	0.18

Notes: Clustered standard errors at the bank-level in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Difference-in-differences estimation with a 4-quarters before and after the 2016 stress test with a strongly balanced sample. *Post ST16* is a dummy for 2017Q1–2017Q4. *Tested* is a dummy for stress-tested banks. Bank-level control variables are lagged by one quarter and comprise *Log(Assets), Voluntary Capital Ratio, Regulatory Capital Ratio, Liquidity Ratio, Retail Ratio, LLP Ratio, CIR, RoE*, and *Interest Income Ratio*. Country-time fixed effects indicate the country of each bank's headquarters and each quarter. In Column (1) *High Transparency* is a dummy indicating whether a treated bank is part of the EBA sample and therefore its stress test results were published on a granular level. In Column (2) the sample includes only stress-tested banks. In Columns (3)-(5) the samples include only stress-tested banks under the EBA mandate, i.e., those with *High Transparency*. In Column (3) *Low CET1(Adverse)* is a dummy equal to one for EBA banks with below median CET1 ratio projections in the adverse scenario. In Column (4) *High Depletion* is a dummy equal to one for EBA banks with above median capital depletion in the adverse scenario, i.e., a high difference between initial CET1 ratio and final CET1 ratio projection. In column (5) *Depletion/CET1(Initial)* is a dummy equal to one for EBA banks with above median capital depletion relative to their starting CET1 ratio.

We exploit the dichotomous distinction in the way stress test results were published. As mentioned in Section 3, bank-specific results are only published for the banks which were part of the EBA sample while only aggregate results of the SREP sample were published that did not allow to extract bank-specific information. Hence, we should expect that the disciplining effect driven by the market discipline channel is stronger for EBA banks. We estimate Eq. (2) replacing *High ST Intensity* with the dummy *High Transparency* which is equal to 1 if a bank was part of the EBA stress test sample while it is equal to 0 if it was stress tested as part of the SREP sample as described in Section 5.2.

The results reported in column (1) of Table 7 show that we cannot find a significant difference in terms of credit risk in the aftermath of the stress test between EBA and SREP banks. The insignificance persists when estimating within the sample of stress-tested banks, see column (2). In columns (3) to (5), we test whether market discipline might only be exerted on banks for which a comparably severe result was published. We, however, cannot estimate a significant difference between positive and negative news in the sample of EBA banks using three summary metrics of what market participants usually consider the "result" of the stress test: projected CET1 ratio at the end of the adverse scenario, the CET1 depletion between initial CET1 ratio and the final projection, and this depletion relative to initial CET1 ratios.[26]

Thus, we cannot find evidence that the disciplining effect stemming from the participation in the stress test, which we find in our baseline estimation, is driven by the increased transparency due to the publication of stress test results. This finding contrasts with the evidence in Klein et al. (2021) which shows that enhanced mandated transparency of banks' ABS loan portfolios has improved the quality of the loan pool and thus benefited investors. However, the two transparency exercises differ in important ways regarding the periodicity and complexity of reporting. While stress tests are conducted only biannually and are based on a very complex and partly undisclosed methodology which makes them harder to interpret, information in the ABS reporting initiative must be disclosed on a regular basis and consists of objective characteristics. This may in part explain why stress tests do not lead to a comparable lasting impact. Furthermore, if market discipline ultimately aims at changing bank strategy by replacing bank management, it might take time to become effective (Ferreira et al., 2021). We might therefore not be able to detect the market discipline operating with a gap that goes beyond our sample period.[27] Another limitation of our analysis in this respect could be its main focus on credit risk. While our supervisory scrutiny measures directly refer to credit risk as does our dependent variable, the banks' stress test results published by the EBA provide an overview of all banks' risks. How market participants value individual parts of this information, we cannot tell. Nevertheless, taking these limitations into account we do not find any evidence that market discipline contributed to the estimated reduction in credit risk after the 2016 EU-wide stress test.

In order to test the capital channel, we associate high stress test intensity with a high impact of the stress test on banks' capitalization. We explicitly examine if the change in capital guidance related to the stress test results drives our main finding. Several studies using U.S. data show that banks decrease risk-taking after stress tests due to the associated increases in capital requirements and limits to capital distribution plans (Acharya et al., 2018; Pierret and Steri, 2020). We measure the impact on capitalization by looking at the capital requirements that resulted from the stress test. As pointed out in Section 3, stress test results of the 2016 exercise did not map directly into supervisory capital measures. They were used among other information for the Pillar 2 Guidance (P2G) that came into effect in 2017q1 as clarified in EBA (2016d). The Pillar 2 capital guidance (P2G) does not constitute a binding minimum capital requirement but determines an "adequate level of capital to be maintained in order to have sufficient capital as a buffer to withstand stressed situations". Supervisors "expect

[26] We also tested the *High Transparency* dummy in the sample of banks with negative news as well as quadruple interaction terms in the full and tested-only samples without finding any contradictory evidence. We chose the above for being the easiest to interpret.

[27] However, we tested if the effect of market discipline becomes stronger towards the end of our sample period but we could not find any trend that would support a lagged impact (the results are not reported but available upon request).

Table 8
The capital channel.

Sample:	(1)	(2)	(3)	(4)
	Tested	Full sample		
Post ST16 × Tested		−0.049**	−0.048**	−0.048**
		(0.021)	(0.023)	(0.024)
Post ST16 × Tested × High P2G	0.026	0.018		
	(0.023)	(0.023)		
Post ST16 × Tested × High Effective P2G			0.015	
			(0.017)	
Post ST16 × Low Voluntary Capital				−0.037*
				(0.022)
Post ST16 × Tested × Low Voluntary Capital				0.016
				(0.030)
L.Log(Assets)	0.121	−0.146***	−0.144***	−0.118***
	(0.165)	(0.040)	(0.038)	(0.037)
L.Regulatory Capital Ratio	−0.102	−0.149	−0.183	0.111
	(0.348)	(0.188)	(0.186)	(0.171)
L.Voluntary Capital Ratio	−0.065	−0.246*	−0.248*	
	(0.371)	(0.140)	(0.141)	
Bank Controls	Yes	Yes	Yes	(Yes)
Bank FE	Yes	Yes	Yes	Yes
Country × Time FE	Yes	Yes	Yes	Yes
Observations	413	924	924	924
Banks	63	132	132	132
R2-within	0.124	0.124	0.123	0.137

Notes: Clustered standard errors at the bank-level in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Difference-in-differences estimation with a 4-quarters before and after the 2016 stress test with a strongly balanced sample. *Post ST16* is a dummy for 2017Q1–2017Q4. *Tested* is a dummy for stress-tested banks. Bank-level control variables are lagged by one quarter and comprise *Log(Assets), Voluntary Capital Ratio, Regulatory Capital Ratio, Liquidity Ratio, Retail Ratio, LLP Ratio, CIR, RoE*, and *Interest Income Ratio*. To avoid collinearity *Voluntary Capital Ratio* is excluded from the list of covariates in Column (4). Countryxtime fixed effects indicate the country of each bank's headquarters and each quarter. In Columns (1) and (2) *High P2G* is a dummy indicating above median change in Pillar 2 capital guidance in 2017Q1 when the guidance was informed by the stress test results. In Column (1) we estimate within stress-tested banks. In Column (3) *High Effective P2G* is dummy indicating above median change in Pillar 2 capital guidance in 2017Q1 relative to CET1 capital buffers. In Column (4) *Low Voluntary Capital* is a dummy indicating on average below median voluntary capital buffers in the quarters before the stress test.

banks to comply with'' the P2G (ECB, 2016b). Hence, similar to capital requirements the P2G creates incentives to lower risk weighted assets in order to comply with the supervisory expectations.

To investigate the existence of the capital channel we extend our baseline equation (1) by introducing a triple interaction term between $Post_t \times Tested_i$ and a dummy indicating a high impact on bank capitalization. We use supervisory data to construct the dummy variable *High P2G* that indicates above median P2G in 2017q1. Furthermore, in line with Pierret and Steri (2020), we also take into account how much P2G would effectively affect banks' capital structure by weighting the P2G with available CET1 capital buffers before constructing the dummy *High Effective P2G*. To account for the possibility that P2G might not correlate strongly with stress test results, we further test whether we find a stronger effect on bank risk at banks that entered into the stress test with lower capital buffers. For this, we take the average *Voluntary Capital Ratio* of banks in the four quarters before the stress test and divide banks in two groups at the median. Since we expect a stronger effect for banks with low capital buffers, we substitute *High Intensity* in Eq. (2) here with a dummy *Low Voluntary Capital* indicating below median capital buffers before the stress test.

In column (1) of Table 8 we report the result when estimating within the sample of tested banks. In columns (2) to (4), we estimate within the full sample including the non-tested banks. As the estimated coefficient of the interaction terms are not significant in columns (1) and (2), we cannot find evidence supporting the hypothesis that the reduction in credit risk in the aftermath of the stress test found by our baseline analysis was driven by larger changes in the P2G. Nevertheless, it could be that the relevant driver of the credit risk reduction in the aftermath of the stress test is not the change in the P2G but rather banks' ability to comply with the additional requirements. Then banks whose capital ratio is closer to the requirements (including the P2G), i.e., banks which hold smaller voluntary capital buffers at the start of the stress test or those where P2G is higher relative to voluntary buffers, might have stronger incentives to reduce risk in order to reduce the probability of breaching the regulatory requirements. As the coefficient on the triple interaction in column (3) is also insignificant, we cannot find evidence that higher effective P2G are relevant for our findings. While we find that banks with lower capitalization reduced RWD after the 2016 EU-wide stress test, we cannot find a significant difference between tested and non-tested banks. The former result is in line with the general finding that lower voluntary capital buffers significantly reduce RWD (cf. Table 3).

To sum up, we cannot find evidence underlining the capital channel that could explain our baseline result that stress tested banks on average reduce their RWD by more than non-tested banks. Therewith, we cannot confirm the findings of other studies regarding the predominance of the capital channel in the U.S. stress testing framework. This could simply reflect the fact that the European stress test design does not focus on the evaluation of banks' capital plans and the 2016 EU-wide stress did not entail an mechanic link between the stress test results and the P2G. Our results might also be affected by the choice of measure of risk that only reflects one way for banks to adjust to higher capital requirements, i.e., by reducing the average credit risk weight.

8. Conclusions

We assess the effect of the 2016 EU-wide stress test coordinated by the European Banking Authority and conducted by the European Central Bank on banks' credit risk in the aftermath of the exercise using confidential supervisory data. To identify this effect, we rely on a difference-in-differences approach examining the change in risk weighted densities (RWD) between stress tested banks and non-stress tested banks between the four quarters of 2015 leading up to the stress test and the four quarters of 2017 after the stress test. We find that stress tested banks reduced their RWD by 4.2 percentage points more than not stress tested banks. This is a rather sizeable change caused by the participation in the stress test exercise as it amounts to about 20 percent of the standard deviation of RWD. We further show that our findings are not driven by the difference in size between the treated and untreated banks. We therefore conclude that the 2016 EU-wide stress test had a disciplining effect on banks which corroborates similar evidence for the U.S. stress tests. It should be noted, however, that this might lead tested banks to cut lending to small and medium-sized firms.

Furthermore, we test the hypothesis that the increased supervisory scrutiny carried out during the QA process of the 2016 EU-wide stress test is the driver that leads banks to ease their risk in the aftermath of the exercise. To this end we again rely on a difference in differences approach and we exploit ECB proprietary data on the interactions that took place between the ECB and the banks during the QA process to construct measures of the internal margin of being stress tested. More specifically, in our regression analysis we build and use three yardsticks which measure the intensity, potential impact, and duration of these interactions. We find that banks that had more interactions or interacted with the ECB over a longer period of time reduced RWD more than the other banks. This result proves that the enhanced supervisory scrutiny that is entailed by the stress test QA process has a disciplining effect on banks' risk. We further find that this disciplining effect depends on banks' risk management culture. Indeed, supervisory scrutiny is more effective for banks with strong and independent risk management functions and those with a strongly safety-oriented risk culture. This shows that regulatory stress tests interact with banks' risk management culture and can therefore play an important part in strengthening these. We cannot find evidence that our results can be explained by an increase in market discipline connected to the publication of stress test results nor proof that the risk reduction that we find is driven by an increase of capital guidance as a result of the bank-specific assessment.

With respect to the discussion on different stress test designs, our results highlight some merit in the use of a constrained bottom-up approach. Indeed, our work provides evidence that stress tests conducted applying a robust Quality Assurance of banks' bottom-up projections and models by competent authorities, which ensures the credibility and reliability of the results, may have beneficial disciplining effects on stress tested banks' risk. On the other side, though, it has to be noted that this merit is not costless. One of the stress tests' primarily objectives is to correctly assess banks' risk profiles. Our findings do not provide information on how well this objective is met. However, the possible strategic underreporting of banks' vulnerabilities under a bottom-up approach could undermine the reliability of the stress test outcomes from this perspective. Pursuing a more top-down approach could possibly reduce these costs.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgements

We thank an anonymous referee, Rym Ayadi (discussant), Thorsten Beck, Christa Bouwman (editor), Enzo Cerletti, Piotr Danisewicz (discussant), Reint Gropp, Anna Kovner, David Marqués-Ibáñez, Juan Francisco Martínez, Thomas Mosk, Matías Ossandon Busch, Diane Pierret (discussant), Kasper Roszbach, Alessandro Scopelliti, and participants at the ECB Macroprudential Stress-testing Conference as well as the 2020 FED Stress Testing Research Conference for their helpful comments.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jfi.2022.101015.

References

Abadie, A., Imbens, G.W., 2011. Bias-corrected matching estimators for average treatment effects. J. Bus. Econom. Statist. 29 (1), 1–11.
Acharya, V.V., Berger, A.N., Roman, R.A., 2018. Lending implications of US bank stress tests: Costs or benefits? J. Financial Intermediation 34, 58–90.
Ahnert, L., Vogt, P., Vonhoff, V., Weigert, F., 2020. Regulatory stress testing and bank performance. Eur. Financial Manag. 26 (5), 1449–1488.
Baule, R., Tallau, C., 2021. The risk sensitivity of Basel risk weights and loan loss provisions: evidence from European banks. Eur. J. Finance 27 (18), 1855–1886.
Begley, T.A., Purnanandam, A., Zheng, K., 2017. The strategic underreporting of bank risk. Rev. Financ. Stud. 30 (10), 3376–3415.
Behn, M., Haselmann, R., Vig, V., 2022. The limits of model-based regulation. J. Finance 77 (3), 1635–1684.
Berger, A.N., Bouwman, C.H., 2013. How does capital affect bank performance during financial crises? J. Financ. Econ. 109 (1), 146–176.
Berrospide, J.M., Edge, R.M., 2019. The Effects of Bank Capital Buffers on Bank Lending and Firm Activity: What Can We Learn from Five Years of Stress-Test Results? Working Paper.
Bonfim, D., Cerqueiro, G., Degryse, H., Ongena, S., 2022. On-site inspecting zombie lending. Manage. Sci. (in press).
Buch, C.M., DeLong, G., 2008. Do weak supervisory systems encourage bank risk-taking? J. Financial Stab. 4 (1), 23–39.
Calem, P., Correa, R., Lee, S.J., 2020. Prudential policies and their impact on credit in the United States. J. Financial Intermediation 42, 100826.
Cameron, K.S., Quinn, R.E., DeGraff, J., Thakor, A.V., 2014. Competing Values Leadership. Edward Elgar Publishing.
Chaly, S., Hennessy, J., Menand, L., Stiroh, K., Tracy, J., 2017. Misconduct Risk, Culture, and Supervision. Working paper, Federal Reserve Bank of New York.
Connolly, M., 2021. The Real Effects of Stress Testing. Working Paper.
Cortés, K., Demyanyk, Y., Li, L., Loutskina, E., Strahan, P.E., 2020. Stress tests and small business lending. J. Financ. Econ. 136 (1), 260–279.
Delis, M.D., Staikouras, P.K., 2011. Supervisory effectiveness and bank risk. Rev. Finance 15 (3), 511–543.
Doerr, S., 2021. Stress tests, entrepreneurship, and innovation. Rev. Finance 25 (5), 1609–1637.
EBA, 2016a. EU-Wide Stress Testing 2016. European Banking Authority, https://www.eba.europa.eu/risk-analysis-and-data/eu-wide-stress-testing/2016. (Accessed 01 July 2019).
EBA, 2016b. EBA Publishes 2016 EU-Wide Stress Test Results, 2016-07-29. European Banking Authority, https://eba.europa.eu/eba-publishes-2016-eu-wide-stress-test-results. (Accessed 01 July 2019).
EBA, 2016c. 2016 EU-Wide Stress Test: Frequently Asked Questions. (24 February 2016), European Banking Authority.
EBA, 2016d. Information Update on the 2016 EU-Wide Stress Test - Using the 2016 EU-Wide Stress Test Results in the SREP Process. (1 July 2016), European Banking Authority.
ECB, 2016a. Frequently Asked Questions on the 2016 EU-Wide Stress Test. European Central Bank Banking Supervision, https://www.bankingsupervision.europa.eu/about/ssmexplained/html/stresstestFAQ.en.html. (Accessed 01 July 2019).
ECB, 2016b. The Supervisory Review and Evaluation Process: What's New?. European Central Bank Banking Supervision, Supervision Newsletter. https://www.bankingsupervision.europa.eu/press/publications/newsletter/2016/html/nl161116.en.html. (Accessed 01 July 2019).
ECB, 2017. Report on Financial Structures. Technical Report, (October 2017), European Central Bank.
ECB, 2019. What Makes a Bank Significant? European Central Bank Banking Supervision, https://www.bankingsupervision.europa.eu/banking/list/criteria/html/index.en.html. (Accessed 01 July 2019).
Eisenbach, T.M., Lucca, D.O., Townsend, R.M., 2016. The Economics of Bank Supervision. Working Paper 22201, National Bureau of Economic Research.

Ellul, A., Yerramilli, V., 2013. Stronger risk controls, lower risk: Evidence from U.S. bank holding companies. J. Finance 68 (5), 1757–1803.
Enria, A., 2019. The future of stress testing - some further thoughts, speech by andrea enria, chair of the supervisory board of the ECB, 8th annual research workshop the future of stress tests in the banking sector - approaches, governance and methodologies, Paris 2019-11-27. https://www.bankingsupervision.europa.eu/press/speeches/date/2019/html/ssm.sp191127~2f9bdabff9.en.html.
Fernandes, M., Igan, D., Pinheiro, M., 2020. March madness in wall street:(what) does the market learn from stress tests? J. Bank. Financ. 112, 105250.
Ferreira, D., Kershaw, D., Kirchmaier, T., Schuster, E., 2021. Management insulation and bank failures. J. Financial Intermediation 47, 100909.
Fiordelisi, F., Ricci, O., 2014. Corporate culture and CEO turnover. J. Corp. Finance 28, 66–82.
Flannery, M., Hirtle, B., Kovner, A., 2017. Evaluating the information in the Federal Reserve stress tests. J. Financial Intermediation 29, 1–18.
Georgescu, O.M., Gross, M., Kapp, D., Kok, C., 2017. Do Stress Tests Matter? Evidence from the 2014 and 2016 Stress Tests. Working Paper Series 2054, European Central Bank.
Gopalan, Y., Kalda, A., Manela, A., 2021. Hub-and-spoke regulation and bank leverage. Rev. Finance 25 (5), 1499–1545.
Gropp, R., Mosk, T., Ongena, S., Wix, C., 2019. Banks response to higher capital requirements: Evidence from a quasi-natural experiment. Rev. Financ. Stud. 32 (1), 266–299.
Guindos, L.d., 2019. The evolution of stress-testing in Europe, keynote speech by Luis de Guindos, Vice-President of the ECB, at the annual US-EU symposium organised by the program on international financial systems, Frankfurt, 2019-09-04. https://www.ecb.europa.eu/press/key/date/2019/html/ecb.sp1909042~4c8236275b.en.html. (Accessed 20 March 2020).
Hirtle, B., Kovner, A., Plosser, M., 2020. The impact of supervision on bank performance. J. Finance 75 (5), 2765–2808.
Imbens, G.W., Wooldridge, J.M., 2009. Recent developments in the econometrics of program evaluation. J. Econ. Lit. 47, 5–86.
Ivanov, I., Wang, J., 2019. The Impact of Bank Supervision on Corporate Credit. Working Paper.
Kandrac, J., Schlusche, B., 2021. The effect of bank supervision and examination on risk taking: Evidence from a natural experiment. Rev. Financ. Stud. 34 (6), 3181–3212.
Klein, P., Mössinger, C., Pfingsten, A., 2021. Transparency as a remedy for agency problems in securitization? The case of ECB's loan-level reporting initiative. J. Financial Intermediation 46, 100853.
Konietschke, P., Ongena, S., Ponte Marques, A., 2022. Stress Tests and Capital Requirement Disclosures: Do They Impact Banks' Lending and Risk-Taking Decisions?. ECB Working Paper Series, (No. 2679).
Laeven, L., Ratnovski, L., Tong, H., 2016. Bank size, capital, and systemic risk: Some international evidence. J. Bank. Financ. 69, 25–34.
Lazzari, V., Vena, L., Venegoni, A., 2017. Stress tests and asset quality reviews of banks: A policy announcement tool. J. Financial Stab. 32, 86–98.
Leitner, Y., Yilmaz, B., 2019. Regulating a model. J. Financ. Econ. 131 (2), 251–268.
Mariathasan, M., Merrouche, O., 2014. The manipulation of basel risk-weights. J. Financial Intermediation 23 (3), 300–321.
Mirza, H., Zochowski, D., 2017. Macroprudential Policy Analysis and Tools–Stress Test Quality Assurance from a Top-Down Perspective. Macroprudential Bulletin Issue 3, European Central Bank.
Morgan, D.P., Peristiani, S., Savino, V., 2014. The information value of the stress test. J. Money Credit Bank. 46 (7), 1479–1500.
Niepmann, F., Stebunovs, V., 2018. Modeling Your Stress Away. CEPR Discussion Paper DP12624, Center for Economic Policy Research.
Passalacqua, A., Angelini, P., Lotti, F., Soggia, G., 2020. The Real Effects of Bank Supervision: Evidence from On-site Bank Inspections. Working Paper.
Petrella, G., Resti, A., 2013. Supervisors as information producers: Do stress tests reduce bank opaqueness? J. Bank. Financ. 37 (12), 5406–5420.
Pierret, D., Steri, R., 2020. Stressed Banks. Working paper.
Plosser, M.C., Santos, J.A., 2018. Banks' incentives and inconsistent risk models. Rev. Financ. Stud. 31 (6), 2080–2112.
Ponte Marques, A., Pancaro, C., Durrani, A., Giraldo, G., Panos, J., Zaharia, A., 2022. Does the Disclosure of Stress Test Results Affect Market Behaviour? Macroprudential Bulletin Issue 17, European Central Bank.
Quinn, R.E., Rohrbaugh, J., 1983. A spatial model of effectiveness criteria: Towards a competing values approach to organizational analysis. Manage. Sci. 29 (3), 273–393.
Rezende, M., Wu, J., 2014. The Effects of Supervision on Bank Performance: Evidence from Discontinuous Examination Frequencies. Working Paper.
Rosenbaum, P.R., Rubin, D.B., 1983. The central role of the propensity score in observational studies for causal effects. Biometrika 70 (1), 41–55.
Song, F., Thakor, A.V., 2019. Bank culture. J. Financial Intermediation 39, 59–79.
Tarullo, D.K., 2016. Next steps in the evolution of stress testing, speech by Governor Daniel K. Tarullo, at the yale university school of management leaders forum, New Haven, Connecticut 2016-09-26. https://www.federalreserve.gov/newsevents/speech/tarullo20160926a.htm. (Accessed 20 March 2020).