DISCUSSION PAPER SERIES

DP16157

The disciplining effect of supervisory scrutiny in the EU-wide stress test

Christoffer Kok, Carola Müller, Steven Ongena and Cosimo Pancaro

FINANCIAL ECONOMICS

ISSN 0265-8003

The disciplining effect of supervisory scrutiny in the EU-wide stress test

Christoffer Kok, Carola Müller, Steven Ongena and Cosimo Pancaro

Discussion Paper DP16157
Published 15 May 2021
Submitted 14 May 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- Financial Economics

The disciplining effect of supervisory scrutiny in the EU-wide stress test

Abstract

Using a difference-in-differences approach and relying on confidential supervisory data and an unique proprietary data set available at the European Central Bank 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.

JEL Classification: G11

Keywords: Stress Testing, credit risk, Internal Models, Banking Supervision, banking regulation

Christoffer Kok - christoffer.kok@ecb.europa.eu
European Central Bank

Carola Müller - cmuller@cemla.org
CEMLA - Center for Latin American Monetary Studies, Halle Institute for Economic Research (IWH)

Steven Ongena - steven.ongena@bf.uzh.ch
University of Zurich and CEPR

Cosimo Pancaro - cosimo.pancaro@ecb.europa.eu
European Central Bank

Acknowledgements
We thank Rym Ayadi (discussant), Thorsten Beck, 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. Steven Ongena acknowledges financial support from ERC ADG 2016 - GA 740272 lending. All errors are our own. This paper should not be reported as representing the views of the European Central Bank (ECB) or any affiliated institution.

The disciplining effect of supervisory scrutiny in the EU-wide stress test*

Christoffer Kok[†1], Carola Müller[‡2], Steven Ongena[§3], and Cosimo Pancaro[¶1]

[1]*European Central Bank*
[2]*Center for Latin American Monetary Studies , Halle Institute for Economic Research*
[3]*University of Zurich , Swiss Finance Institute , KU Leuven , CEPR*

Abstract

Using a difference-in-differences approach and relying on confidential supervisory data and an unique proprietary data set available at the European Central Bank 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.

JEL Classification: G11, G21, G28.

Keywords: Stress Testing, Credit Risk, Internal Models, Banking Supervision, Banking Regulation.

*We thank Rym Ayadi (discussant), Thorsten Beck, 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.
Steven Ongena acknowledges financial support from ERC ADG 2016 - GA 740272 lending. All errors are our own. This paper should not be reported as representing the views of the European Central Bank (ECB) or any affiliated institution.

†Email: christoffer.kok@ecb.europa.eu

‡Email: cmuller@cemla.org

§Email: steven.ongena@bf.uzh.ch

¶Email: cosimo.pancaro@ecb.europa.eu

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, an important question is whether stress tests contribute to financial stability by promoting risk reduction in the banking sector.

European stress tests often involve interactions between banks and supervisors on banks' risk management practices 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?

In this paper, we provide evidence that supervisory scrutiny that comes with stress testing can have a disciplining effect on bank risk. We look at the EU-wide Stress Test conducted in 2016 by the European Banking Authority and the European Central Bank. Our findings corroborate existing evidence from U.S. stress tests showing reduced bank risk after testing. Exploiting the institutional design of the European stress test, we are able to measure stress-test related scrutiny by creating metrics based on documented interactions between supervisors and tested institutions. The European design offers a good testing ground to highlight the effect of supervisory scrutiny in contrast to other channels through which stress tests can affect bank risk, most notably from effects stemming from stress-test induced capital measures. The use of confidential data allows us to shed light on supervisory interactions which constitute a significant part of the European stress tests. We therefore contribute to the rare evidence on the effectiveness of supervisory scrutiny on bank risk. Our results suggest that stress tests are not merely a *check-the-box* regulatory constraint but rather able to affect bank risk.

Our analysis has two main steps. First, we examine whether banks' participation in the EU-wide stress test can discipline banks' risk. We use a difference-in-difference approach which compares the change in bank risk around the 2016 EU-wide stress test exercise of banks that were tested and banks that were not tested. We use bank-level supervisory data exploiting detailed information on bank-specific balance sheet and profit and loss items and capital requirements for the period between 2015 and 2017. We measure bank credit risk as the aggregate risk-weight of banks' entire credit risk exposures, called risk weight density (RWD).

By controlling for changes in capital requirements that are related to the stress test results[1], we are able to disentangle their effects from those caused by the tighter scrutiny due to the stress test. Overall, our results, which corroborate existing similar evidence on the effect of U.S. stress tests on bank risk (Acharya, Berger, and Roman, 2018; Pierret and Steri, 2018), show that banks that participated in the exercise reduced their average RWD by about 4.2 percentage points relative to banks that were not tested. This effect is economically significant as it amounts to a change in RWD of about 20 percent of the standard deviation of RWD of tested banks.[2] Hence, stress testing can have a significant effect on banks' credit risk and can help "to make banks safer and sounder" (Enria, 2019).

In the analysis, we use the stress test as a treatment that was administered only to a subset of European banks. The main challenge of our analysis is that this subset of banks was not selected randomly from a homogeneous population as it would occur in a perfect experimental setting. In fact, whether or not a bank was tested depended on its status of systemic importance and, more specifically, on its membership to the group of SSM Significant Institutions (SIs).[3] Hence, our treatment group consists of nearly all SIs[4] while our control group comprises a subsample of the SSM Less Significant Institutions (LSIs)[5]. This implies that in our sample of banks there is

[1]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 structure and the scrutiny channel.

[2]Similar results are obtained when using different measures of bank risk such as banks' Expected Default Frequency by Moody's Analytics and the z-score.

[3]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).

[4]Greek SIs were excluded from the 2016 EU-wide stress test as they were materially impacted by the European sovereign debt crisis. Some further SIs were not included in the sample of banks, which participated in the 2016 EU-wide stress test, as they were under restructuring. For the same reasons, these banks are neither included in our treatment group nor in our control group.

[5]Less significant institutions are SSM banks that do not fulfil any of the significance criteria to be qualified

a sizeable difference in the average total assets between the banks in the treatment and control groups. However, as the selection was based on fully observable characteristics, we can control for the selection criteria and are still able to estimate the effect of being stress tested. Yet our results hinge on the notion that the banks in the control group are comparable to the tested banks. Therefore, in order to alleviate the possible concerns that our results are driven by the difference in size in terms of total assets between the banks in the control group and in the treatment group, we rely on two different approaches. 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, Mosk, Ongena, and Wix (2019) which allow to balance the possible differences between the treatment and the control groups. More specifically, the first matching strategy consists in implementing a common support on size by excluding all control banks that are smaller than the smallest tested bank and by excluding all tested banks that are larger than the biggest control group bank while the second strategy consists in reducing the sample by using only the two largest non-tested and two smallest tested banks within each country. Overall, the results of both robustness analyses confirm our baseline findings, i.e. being part of the 2016 EU-wide stress test led to a reduction in stress tested banks' risk with respect to the risk of non-stress tested banks.

In a second step, we examine whether the identified reduction in credit risk is associated with the additional scrutiny, which supervisors exert on banks' stress testing projections and models during the exercise. In order to test this possible channel, we exploit the variation in supervisory scrutiny across banks and 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. The metrics of the intensity of supervisory scrutiny are built relying on ECB proprietary data collected during the 2016 EU-wide stress test.

We focus on the role of the interactions between supervisors and banks about their stress testing models and projections and their potential to reduce banks' risk. 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 for SSM

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.

banks. 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 during the QA process 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 particular, 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. We find that banks that had more subjects to interact on or interacted over a longer period of time with the supervisors reduced their credit risk more than banks with fewer matters or shorter periods of discussions. We also prove that the banks that were involved in more and longer interactions with the supervisors during the stress-test Quality Assurance about their results and models were not the ex-ante riskier indicating that the measure of intensity of the treatment is not endogenously determined by the risk measure.

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

Finally, as a further analysis, we examine if other channels in addition to the supervisory scrutiny channel can potentially explain our baseline result. In particular we consider the capital structure channel and the market discipline channel. The former refers to the possible effects on bank risk related to additional capital requirements or capital distribution limits associated with the stress test results as bank capital is an important determinant of banks' risk choices (Berger and Bouwman, 2013). The latter instead refers to the discipline potentially imposed by markets on banks as a consequence of the disclosure of the stress test results.

Regarding the capital structure 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 structure channel seems highly relevant (Acharya et al., 2018; Pierret and Steri, 2018). This result could, however, reflect the fact that the 2016 EU-wide stress test was not a pass or fail exercise and the capital structure channel is by design less central in the European exercise. Indeed, the results of the 2016 EU-wide stress test were only used to inform supervisory decisions about capital guidance. This 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. Similarly, we cannot find evidence that banks whose results were published and disclosed at a granular level reduced credit risk more than those banks whose results were not published. Hence, we cannot find evidence that the market disciplining channel can explain our baseline result.

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. We complement the existing findings by providing evidence on the effects of stress testing on bank risk in Europe and by shedding more light on the channels through which stress tests can affect bank risk.

The studies, which are closest to ours, are Pierret and Steri (2018) and Acharya et al. (2018) which focus on the effects of stress testing on risk taking in the U.S.. Pierret and Steri (2018) examine the effect of U.S. stress tests on risk-taking and are the first to point out the importance of separating the capital channel from other channels. They find that the supervisory scrutiny, which is placed on banks exposed to the stress test, decreases their risk-taking relative to banks that were not subject to a stress test. They show that this effect is additional to the risk-taking incentives stemming from the increase in capital linked to the stress test results. More specifically, they find that the increase in risk-taking due to the higher capital requirements is mitigated by the enhanced scrutiny associated to the stress test exercise. Our findings are able to refine theirs since in this paper we are able to separately measure both the supervisory scrutiny channel and the capital channel albeit in contrast to the U.S. stress test the European stress tests do not automatically trigger higher capital requirements. Acharya et al. (2018) also support the view that the U.S. stress tests decreased banks' risk-taking. They find that stress tested banks increase spreads on loans and decrease credit supply especially in riskier market segments. Acharya et al. (2018) argue that banks' more prudent behaviour is driven by the channel of higher capital[6].

Other studies confirm that stress tests affect bank behaviour. These studies also rely on U.S. data. For example, Cortés, Demyanyk, Li, Loutskina, and Strahan (2020) find that the banks most affected by stress tests reduce credit supply to riskier borrowers and raise interest rates on

[6]Indeed, the arguments they provide for the risk management hypothesis rely mostly on an increase in capital initiated by the stress test and less by any effect of the stress test itself.

small loans. Calem, Correa, and Lee (2020) provide evidence that stress tests led to a reduction in credit supply in the mortgage market. On the contrary though, Flannery, Hirtle, and Kovner (2017) find no evidence for changes in bank portfolios (risk shifting) in the U.S. after stress test publications.

Furthermore, we 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 (2019) 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. The additional risk took the form of more risky loans, faster asset growth, and a greater reliance on low quality capital. Hirtle, Kovner, and Plosser (2019) 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' losses on risky assets. Bonfim, Cerqueiro, Degryse, and Ongena (2020) find that an inspected bank becomes less likely to refinance zombie firms, immediately spurring their default. 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.

Finally, we contribute to the literature linking internal risk management practices and bank supervision. So far, most attention has been shed 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, Haselmann, and Vig, 2016; Begley, Purnanandam, and Zheng, 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, 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 for SSM banks.

The paper is structured as follows. Section 2 illustrates the possible mechanisms through which the scrutiny exerted by supervisors during the stress test 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 contains an overview of the data sources and describes 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 exerted during the stress test. In Section 6 we show our baseline result on the impact of being stress tested on bank risk and the results of a set of robustness checks. 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 structure 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. Hirtle et al. (2019) 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, Hennessy, Menand, Stiroh, and Tracy, 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, Kalda, and Manela, 2017; Kandrac and Schlusche, 2019) or their hours worked at a supervised banks (Eisenbach, Lucca, and Townsend, 2016; Hirtle et al., 2019).

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., 2020; 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.[7] Their mandatory use for regulatory purposes requires banks to invest resources in the

[7] Bottom-up stress tests in Europe are an important tool to strengthen banks' risk management (Enria, 2019;

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.[8] This additional 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[9].

It is conducted following a constrained bottom-up approach[10]. Under this approach, banks generate stress test projections using their own models, relying on a common predefined macro-financial scenario[11] 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 Au-

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).

[8] 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).

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

[10] Instead, in the U.S., the supervisory stress test is conducted in a top-down fashion.

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

thority 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[12], 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[13].

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

[12]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.

[13]see Mirza and Zochowski (2017) and Kok, Müller, and Pancaro (2019) for further details on the functioning of the ECB QA process.

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 Figure 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[14].

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 capital requirements and Pillar 2 capital guidance[15]. 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,

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

[15]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.

especially in the element of risk governance.[16]

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.

Our treatment group is composed of banks which took part in the 2016 stress test. The sample of banks participating in the EU-wide stress test is selected by the European Banking Authority. Overall, 51 EU banks participated in the 2016 EU-wide stress-test, 37 of which were under the jurisdiction of the SSM. An additional 56 banks participated in the exercise as part of the Supervisory Review and Evaluation Process (SREP). Results were only published for the 51 EBA banks. As the ECB has no supervisory authority on banks outside the SSM, our dataset does not include non-SSM banks that participated in the 2016 EU-wide stress test. Hence, 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 mostly 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

[16] In the U.S., differently than in the euro area, the stress test results contribute to determine capital requirements.

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

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 some 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 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 normalised 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 sub-section 6.2. 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.

[Table 1 around here.]

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. 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:

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

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 otherwise.

$Controls_{i,t-1}$ is a vector of bank-specific controls. $Bank_i$ and $Time_t$ are respectively bank and time-fixed effects. $Country_i \times Time_t$ 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,

[18] Table A12 in the 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 A12 in the Appendix.

it is assigned based on observables.

However, 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-Imbens (2007) 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:

$$
\begin{aligned}
Risk_{i,t} = \quad & Post_t \times Tested_i + Post_t \times Tested_i \times (High)\ QA\ dim_i \\
& + Bank_i + Time_t + Country_i \times Time_t + Controls_{i,t-1} + \epsilon_{i,t} \qquad (2) \\
& \text{where } dim = \{Quantity, Potential\ Impact, Duration\}.
\end{aligned}
$$

where $QA\ dim_i$ is a bank level measure of the intensity of the QA process. More specifically, we use three different measures. 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.[20] 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

[20] All of these three measures refer to flags which were triggered under the stress-test adverse scenario.

were raised and communicated to the banks during the QA process.[21] 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.[22] 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. (2019). Since two of these three measures are continuous and one is ordinal, the regression is estimated using two different approaches. First, all three measures are treated as continuous *QA dim*$_i$. Second, for each of the three measures a binary variable is created, namely, *High QA dim*$_i$, which is equal to 1 for values above the median QA treatment of the respective category (and equal to 0 if below the median). This latter approach eases the interpretation of the triple interaction term and makes the various results comparable.

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) QA dim*$_i$.

6 The effect of being stress-tested on bank risk

6.1 Baseline results

In this section we report and 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

[21] 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.

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

between the control and treatment group. If the outcome variable, credit risk, 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. Figure 3 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 Figure 3 corroborates the findings in table 2.

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 further evidence that the parallel trend assumption is valid.

[Table 2 around here.] [Figure 3 around here.]

Furthermore, columns (1) and (2) of table 2 show the average *RWD* of the treatment and control group in the pre- and post-test period. 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 (at 5% significant) as well as in the post-period (at 1% significant). 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-difference 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.

[Table 3 around here.]

In table 3, we report the estimated results of eq. 1. More specifically, we report the estimated results for four different specifications which include various combinations of explanatory variables. 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, Ratnovski, and Tong, 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. In column (4) 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.

6.2 Robustness: the selection into being stress-tested

The largest concern related to the estimations presented in sub-section 6.1 is on the sample selection and more specifically on whether the banks in the control group are comparable to the banks in the treatment group. The comparability between these two groups of banks can be arguable when considering their size as it is illustrated in Fig. 4 which shows the distribution of the *Log(Assets)* for both the treatment and the control groups.

We address these possible concerns relying on two different approaches. First, we check whether our estimated effect of a 4.2 percentage point decrease in RWD for stress tested banks relative to non-stress tested banks is solely driven by the specific selection of banks included in our sample. Accordingly, we re-estimate equation 1 by gradually reducing the sample on both ends of the distributions. Second, we employ a different estimator and we apply two matching strategies in line with those used in Gropp et al. (2019) which allow to balance the possible differences in terms of asset size between the treatment and the control groups.

[Table A1 around here.]

Table A1 in the Appendix summarizes the results of the first approach used to minimise the possible concerns that our results are purely driven by the difference in asset size between the banks in the two groups. The rationale supporting the use of this approach is that it makes the two distributions shown in Figure 4 converge by gradually excluding the smallest banks in the control group and/or the largest banks in the treatment group. Table A1 reports the coefficient of interest (*Post ST16* $\times$ *Tested*) and number of observations of each regression of eq. 1 in a matrix. The matrix starts in the upper left corner (row 1, column 1) of table A1 with the baseline estimate of the full sample. In column (2) (and respectively in columns (3) and (4)) we repeat the estimation by excluding the 18 (35 and 52) smallest banks in the control group which correspond to the bottom 25th (50th and 75th) percentile of the distribution of the asset size of the control group. Similarly, we exclude the largest banks of the treatment group in the top 25th (50th and 75th) percentile in row (2) (respectively in rows (3) and (4)). The matrix shows that the coefficient and its significance level stay relatively constant despite reducing the number of banks in the sample. The estimate coefficient varies between -3.0 and -5.5 percentage points. Furthermore, the coefficient becomes more negative when we compare stress-tested banks to larger banks in the control group pointing to the fact that including smaller banks works against our finding. The coefficient loses significance once we consider only the 15 smallest stress tested banks (see Row (4)). However, as we approach the lower right corner, the smaller sizes of the samples imply that we are estimating our equation of interest with a reduced amount of degrees of freedom and thereby statistical power. Overall, table A1 shows that our results are not driven by the different asset size of the banks in the treatment and in the control groups.

[Figure 5 around here.]

Nevertheless, we also employ a matching estimator as a second robustness check to show that our results do not depend on bank size. More specifically, we rely on two different identification strategies proposed in Gropp et al. (2019) where a similar problem arises in the search for a comparable control group. First, we reduce the sample similarly to what is done above by excluding very small non-tested and very large tested banks. To be precise we implement a common support on size by excluding all control banks that are smaller than the smallest tested bank and by excluding all tested banks that are larger than the biggest control group bank. We then implement the bias-corrected matching estimator of Abadie and Imbens (2011) using the Mahalanobis distance between the covariates included in eq. 1. Second, we reduce the sample by using only the two largest non-tested and two smallest tested banks within each country and use the aforementioned matching estimator based on an exact match of these banks within each country. Figure 5 illustrates how asset size is distributed after these restrictions are applied. As shown in table A2, both strategies result in an estimate of our coefficient of interest which is negative and significant further collaborating our hypothesis that being part of the stress test leads to a reduction in banks' risk.

6.3 Further robustness analyses

A general possible concern for any analysis on bank risk is related to the use of an appropriate yardstick for risk. In this work we focus on bank credit risk and measure it as the risk-weighted density for credit risk exposures measured as the weighted average risk weight of all credit risk exposures that banks have to report according to Basel guidelines (and their implementation in the SSM) to their supervisory authorities. As such, this yardstick 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., 2016; 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.

In order to address some of these shortcomings, we first assess whether the participation in the 2016 EU-wide stress test differentially affected alternative measures of bank risk which are not solely focused on credit risk and not only based on supervisory reporting. The results are reported in table A3 in the Appendix. We find corroborating evidence indicating that banks that participated in the 2016 EU-wide stress test reduced risk with respect to banks that did not participate when we employ measures related to banks' default probability. In column (1) 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 (2) 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. Instead, we cannot find a significant differential effect on bank leverage or the share of non-performing loans as documented in table A3 in columns (3) and (4) respectively.

Furthermore, we show in table A4 that our results are not driven by the manipulation of risk weights under the IRB approach. Indeed, our results remain almost unchanged if we restrict our analysis only to exposures under the Standardized Approach, as depicted in column (1). We further decompose the dependent variable *RWD* into the numerator (total risk-weighted exposures) and the denominator (total exposures). In column (2) we therefore estimate the differential effect of being stress-tested on *Log(RW Exposure)* and in column (3) on *Log(Total Exposure)*. The effect is not statistically significant but in magnitude indicates that the reduction in *RWD* is driven by a relative increase in total exposures which must be mainly classified in risk buckets below the average RWD in order to decrease the measure.

Finally, we provide evidence in table A5 in the Appendix that our results are neither driven by the choice of the time window, which we defined as relevant to estimate the effect due to the participation in the stress test, nor are biased due to serial correlation present in the panel data, which might lead to the overestimation of significance in difference-in-differences settings according to Bertrand, Duflo, and Mullainathan (2004).

To address the possible concerns related to the selection of the time window for our analysis, we report in columns (1) and (2) of table A5 the estimated results of equation 1 when we change the definition of the *Post ST16* dummy. Unfortunately, we cannot extend our pre-period due to the limited amount of data available before 2015. In column (1) we narrow the window around the execution of the stress test and include the first quarter of 2016 in the pre-period and the last quarter of 2016 in the post-period. In column (2) we further include the first three quarters of 2018 in the post-period, although it must be reckoned that in 2018 the introduction of IFRS9 caused a major change in accounting rules which particularly affected the credit risk exposure accounting. Furthermore, the next stress test exercise was already launched in January 2018. We are therefore cautious about including 2018 in our baseline definition. As shown in table A5, we find a negative and significant coefficient on the interaction of *Post* and *Tested* for different definitions of the pre- and post-period which is of comparable size to our baseline finding and conclude that our result does not depend on the definition of our estimation window.

Finally, Bertrand et al. (2004) illustrate that positive serial correlation present in panel data can lead to a substantial overestimation of the significance in difference-in-differences settings. They therefore propose to eliminate serial correlation by eliminating the time dimension from the data and estimating a simple panel with only two periods: one for the pre-event time and one for the post-event time. We follow this advice in column (3) and find that our result is robust to eliminating the time dimension from the data.

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. Especially, we are interested to know whether the interactions that the ECB had with banks about their stress test models and projections had an effect on their subsequent credit risk. To this end we exploit the variation in the intensity to which stress tested banks were exposed to the 2016 stress test QA process and we examine the stress test supervisory scrutiny channel as described in more detail in section 2 and in section 5.2.

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 this section we assess the hypothesis that the supervisory scrutiny associated with the EU-wide stress tests has an effect on bank risk. As illustrated in section 5.2, we exploit the variation in the intensity of supervisory scrutiny which we measure making use of three metrics built relying on ECB proprietary data collected during the stress test QA process. To this end we estimate eq. 2 interacting these three different metrics once at the time with $Post_t \times Tested_i$. The main results are reported in table A6 and in table 4.

[Table 4 around here.]

The results showed 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 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 A6 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 Figure 6 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 A6 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. *High QA Potential Impact* is measured as above median potential impact of the credit risk QA flags on the final capital depletion while *QA Potential Impact* is measured as the sum of the potential impact of the QA credit risk flags communicated to banks.[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 look at the data, however, reveals that this insignificance is driven by an outlier. The middle panel (b) in Figure 6 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%, Figure 6 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 some evidence that a longer duration had a mildly 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 A6.

Overall, these findings indicate that there is a value 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. 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 structure and the transparency 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

[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 also did not find any significant results.

discipline channel and the capital structure 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.[24] Finally, we assess if only banks that entered the exercise with rather low capital ratios subsequently reduced their risk. The results are reported in table 5.

In order to test the capital structure channel, we associate high stress test intensity with a high impact of the stress test on banks' capitalization. We measure this impact 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) issued in 2017q1. We use supervisory data to construct the dummy variable *High P2G* that indicates above median P2G in 2017q1. 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 devide banks in two groups at the median. Since we expect a stronger effect for banks with low capital buffers, we subsitute *High Intensity* in eq. 2 here with a dummy *Low Voluntary Capital* indicating below median capital buffers before the stress test.

Finally, 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 a disciplining effect on banks' risk (Petrella and Resti, 2013; Morgan, Peristiani, and Savino, 2014; Georgescu, Gross, Kapp, and Kok, 2017; Flannery et al., 2017; Lazzari, Vena, and Venegoni, 2017; Ahnert, Vogt, Vonhoff, and Weigert, 2018; Fernandes, Igan, and Pinheiro, 2020).

[24] 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 5 around here.]

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. To investigate this possible effect we extend our baseline equation 1 by introducing a triple interaction term between $Post_t \times Tested_i$ and a 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. The results reported in column (1) of table 5 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. 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. However, this does not necessarily imply that the publication of the stress test results enhancing transparency does not increase market discipline. We acknowledge that our analysis is possibly not optimally designed to comprehensively answer this question. The 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. Therefore, we go only as far as stating that the baseline effect that we find estimating equation 1 is not significantly associated with the publication of the 2016 EU-wide stress test results.

Furthermore, we investigate the existence of the capital structure channel. 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 (Acharya et al., 2018; Pierret and Steri, 2018). Contrary to the U.S., the results of the 2016 EU-wide stress test were not automatically linked to changes in capital requirements but as clarified in EBA (2016d) they were only used to inform supervisors for setting their capital guidance. 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 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. The P2G, which was informed by the 2016 stress test results, came into effect in the first quarter of 2017. To investigate the existence of the capital structure channel we extend our baseline equation 1 by introducing a triple interaction term between $Post_t \times Tested_i$ and a dummy *High P2G* which is equal to 1 if the increase in the P2G of a bank was above the median in 2017Q1 and otherwise is equal to 0. This dummy variable is constructed relying on confidential supervisory information. In column (2) of table 5 we report the related results. As the estimated coefficient of the triple interaction term is not significant, 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 might have stronger incentives to reduce risk in order to reduce the probability of breaching the regulatory requirements. To assess this possible further channel we extend our baseline equation 1 by introducing a triple interaction term between $Post_t \times Tested_i$ and a dummy *Low Voluntary Capital* which is equal to 1 for banks with a voluntary capital buffer below the median in 2015Q4, i.e. in the period before the stress test begins, and it is otherwise 0. In column (3) of table 5 we show the related results. 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 structure 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 structure 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.

7.3 Robustness for the supervisory scrutiny channel

One concern related to our analysis might be that our measures of supervisory scrutiny are endogenous. Indeed it might be that institutions with higher credit risk and hence have more scope to decrease it receive more scrutiny. To rule this kind of correlation out we reverse our regression and examine whether average RWD levels before the stress test are correlated to the scrutiny measures. The results are shown in table A7 in the appendix. Since we have no variation over time in the scrutiny measures, we average RWD and the control variables over the pre-test period. We check for correlation (without control variables) in columns (1), (3), and (5). We also test for this relationship after controlling for other bank-level variables, such as size, which might also influence the scrutiny measures in columns (2), (4), and (6). We cannot find any evidence that the supervisory scrutiny measures are significantly related to the level of risk nor other variables that we included in the baseline estimation.

To further examine whether our supervisory scrutiny measures capture an effect that stems from another underlying variable, we test the supervisory scrutiny channel jointly with other variables which could be correlated with high scrutiny. In order to test this, we include a second triple interaction term between *Post*, *Tested*, and the variable of interest. We report the results in the upper panel of table A8 where we focus on *QA Quantity*.[25] In column (1) we interact *Post* $\times$ *Tested* with *Log Assets*. Indeed, it could be that larger banks receive more attention during the QA or simply have more QA interactions due to their more material and possible more complex portfolios. For the similar reason of potentially drawing more attention, we interact in column (2) *Post* $\times$ *Tested* with *Retail Ratio* and in column (3) with *Loan Growth*. Supervisors act in the interest of depositors and borrowers and might be incentivized to devote more attention to banks with large retail business or banks that recently experienced faster growth of their lending book. Finally, in column (4) we interact *Post* $\times$ *Tested* with *IRB share* which is the share of the credit risk portfolio that banks report using the internal ratings based approach. This variable serves as a proxy for the sophistication of banks' modelling skills. It could be argued that more sophisticated banks have more scope to twist their projections which then could generate more issues during the QA or that they might engage in more and longer discussions instead of accepting supervisors' recommendations at an earlier stage. However, in

[25] In the lower panel of table A8 we report the same results for *QA Duration*.

all cases the supervisory scrutiny channel is robust to the inclusion of alternative explanatories.[26]

Furthermore, we test the scrutiny channel when jointly controlling for the capital structure and market discipline channels. We report these results in table A9. We first check whether the three scrutiny measures capture the same effect. In column (1) we include therefore the three triple interaction terms capturing the scrutiny effect all at once. Indeed, we can only find a significant effect for *QA Quantity* but not for *QA Duration*.[27] This might point out that *QA Duration* does not have an effect on bank risk separate from the effect of *QA Quantity*. This could mean that both metrics indeed measure the same which is the scrutiny intensity in stress testing. In column (2) we add the three triple interactions with the capital structure channel and the market discipline channel interactions all at once. Still only the yardstick capturing the quantity of supervisory interactions with banks has a significant effect on credit risk. In columns (3) to (5) we do the same for each QA measure separately and find that the *QA Quantity* as well as *QA duration* scrutiny measures affect bank credit risk even when controlling for the capital structure channel and the market discipline channel.

Lastly, to further address the concerns voiced in section 6.2 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 A10 show in column (1) that the effect 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. As is shown in columns (4) and (5) this measurement holds when including alternative scrutiny measures as well as alternative channels between bank credit risk and stress tests. Column (2) and (3) show that the effect for scrutiny is not significant in the sample with stress-tested banks when measured by *High QA Potential Impact* or *High QA Duration*. In line with the robustness of table A8, we also show in table A11 that the effect of *High QA Quantity* in the tested banks sample is robust to including alternative explanations for high scrutiny in the stress test.

[26]We also tested a range of other variables: Liquidity Ratio, Voluntary Capital Buffer, LLP Ratio, Deposit Ratio, and Loan Ratio.

[27]We estimate with the original *QA Potential Impact* measure, i.e., we do not exclude the outlier which is shown in fig. 6.

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.

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 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.

References

Abadie, A. and Imbens, G. W. 2011. Bias-corrected matching estimators for average treatment effects. Journal of Business & Economic Statistics, 29(1):1–11.

Acharya, V. V., Berger, A. N., and Roman, R. A. 2018. Lending implications of US bank stress tests: Costs or benefits? Journal of Financial Intermediation, 34(C):58–90.

Ahnert, L., Vogt, P., Vonhoff, V., and Weigert, F. 2018. The impact of regulatory stress testing on bank's equity and CDS performance. Working Paper.

Begley, T. A., Purnanandam, A., and Zheng, K. 2017. The strategic underreporting of bank risk. The Review of Financial Studies, 30(10):3376–3415.

Behn, M., Haselmann, R., and Vig, V. 2016. The limits of model-based regulation. Working Paper Series 1928, European Central Bank.

Berger, A. N. and Bouwman, C. H. 2013. How does capital affect bank performance during financial crises? Journal of Financial Economics, 109(1):146–176.

Bertrand, M., Duflo, E., and Mullainathan, S. 2004. How much should we trust differences-in-differences estimates? The Quarterly Journal of Economics, 119(1):249–275.

Bonfim, D., Cerqueiro, G., Degryse, H., and Ongena, S. 2020. On-site inspecting zombie lending. CEPR Discussion Paper No. DP14754, Center for Economic Policy Research.

Buch, C. M. and DeLong, G. 2008. Do weak supervisory systems encourage bank risk-taking? Journal of Financial Stability, 4(1):23 – 39.

Calem, P., Correa, R., and Lee, S. J. 2020. Prudential policies and their impact on credit in the United States. Journal of Financial Intermediation, 42:100826.

Chaly, S., Hennessy, J., Menand, L., Stiroh, K., and Tracy, J. 2017. On-site inspecting zombie lending. Working paper, Federal Reserve Bank of New York.

Cortés, K., Demyanyk, Y., Li, L., Loutskina, E., and Strahan, P. E. 2020. Stress tests and small business lending. Journal of Financial Economics, 136(1):260 – 279.

Delis, M. D. and Staikouras, P. K. 2011. Supervisory effectiveness and bank risk. Review of Finance, 15(3):511–543.

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: 2019-07-01.

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: 2019-07-01.

EBA. 2016c. 2016 EU-wide stress test: Frequently Asked Questions. Technical Report 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. Technical Report 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/stress test FAQ.en.html. Accessed: 2019-07-01.

ECB. 2016b. The Supervisory Review and Evaluation Process: what's new?, European Central Bank Banking Supervision. https://www.bankingsupervision.europa.eu/press/publications/newsletter/2016/html/nl161116.en.html. Accessed: 2019-07-01.

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: 2019-07-01.

Eisenbach, T. M., Lucca, D. O., and Townsend, R. M. 2016. The economics of bank supervision. Working Paper 22201, National Bureau of Economic Research.

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 2f 9bdabff9.en.html. Accessed: 2019-12-04.

Fernandes, M., Igan, D., and Pinheiro, M. 2020. March madness in wall street:(what) does the market learn from stress tests? Journal of Banking & Finance, 112:105250.

Flannery, M., Hirtle, B., and Kovner, A. 2017. Evaluating the information in the Federal Reserve stress tests. Journal of Financial Intermediation, 29:1–18.

Georgescu, O. M., Gross, M., Kapp, D., and 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., and Manela, A. 2017. Hub-and-spoke regulation and bank leverage. Working Paper.

Gropp, R., Mosk, T., Ongena, S., and Wix, C. 2019. Banks response to higher capital requirements: Evidence from a quasi-natural experiment. The Review of Financial Studies, 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.sp190904 2 4c8 236275b.en.html. Accessed: 2020-03-20.

Hirtle, B., Kovner, A., and Plosser, M. 2019. The impact of supervision on bank performance. Staff Report No 768, Federal Reserve Bank of New York.

Imbens, G. W. and Wooldridge, J. M. 2009. Recent developments in the econometrics of program evaluation. Journal of Economic Literature, 47:5–86.

Ivanov, I. and Wang, J. 2019. The impact of bank supervision on corporate credit. Working Paper.

Kandrac, J. and Schlusche, B. 2019. The effect of bank supervision on risk taking: Evidence from a natural experiment.

Kok, C., Müller, C., and Pancaro, C. 2019. The disciplining effect of supervisory scrutiny on banks' risk-taking: evidence from the eu-wide stress test. Macroprudential Bulletin Issue 9, European Central Bank.

Laeven, L., Ratnovski, L., and Tong, H. 2016. Bank size, capital, and systemic risk: Some international evidence. Journal of Banking & Finance, 69:S25–S34.

Lazzari, V., Vena, L., and Venegoni, A. 2017. Stress tests and asset quality reviews of banks: A policy announcement tool. Journal of Financial Stability, 32:86–98.

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

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

Mirza, H. and 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., and Savino, V. 2014. The information value of the stress test. Journal of Money, Credit and Banking, 46(7):1479–1500.

Niepmann, F. and Stebunovs, V. 2018. Modeling your stress away. CEPR Discussion Paper DP12624, Center for Economic Policy Research.

Petrella, G. and Resti, A. 2013. Supervisors as information producers: Do stress tests reduce bank opaqueness? Journal of Banking & Finance, 37(12):5406–5420.

Pierret, D. and Steri, R. 2018. Stressed banks. Swiss Finance Institute Research Paper 17-58, Swiss Finance Insitute.

Plosser, M. C. and Santos, J. A. 2018. Banks' incentives and inconsistent risk models. The Review of Financial Studies, 31(6):2080–2112.

Rezende, M. and Wu, J. 2014. The effects of supervision on bank performance: Evidence from discontinuous examination frequencies. Working Paper.

Rosenbaum, P. R. and Rubin, D. B. 1983. The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1):41–55.

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: 2020-03-20.

Tables

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†	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†	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 dagger (†) indicates that the normalized difference is outside of the range ± 0.25 (which serves as a rule of thumb).

Table 2: Summary statistics of the dependent variable RWD.

	(1)	(2)	(3)	(4)	(5)	(6)
	Levels		Diff.	*First Differences*		Diff.
	Control	Tested	(T-C)	Control	Tested	(T-C)
Pre ST16	0.484	0.437	-0.047**	-0.009	-0.002	0.006
	(0.161)	(0.225)	[0.018]	(0.033)	(0.066)	[0.247]
Post ST16	0.472	0.398	-0.074***	-0.001	-0.003	-0.002
	(0.167)	(0.187)	[0.000]	(0.033)	(0.026)	[0.361]
Diff. (Post-Pre)	-0.012	-0.039*	-0.027*	0.008**	-0.001	-0.009
	[0.427]	[0.055]	[0.073]	[0.010]	[0.897]	[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: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 3: Baseline result of stress test participation.

Dependent: RWD	(1) Without Controls	(2) Control for size	(3) Full Controls	(4) With Demand FE
Post ST16 x Tested	-0.027*	-0.035**	-0.040**	-0.042**
	(0.015)	(0.015)	(0.017)	(0.019)
L.Log(Assets)		-0.119***	-0.133***	-0.145***
		(0.036)	(0.029)	(0.039)
L.Regulatory Capital Ratio			-0.130	-0.150
			(0.214)	(0.191)
L.Voluntary Capital Ratio			-0.241*	-0.254*
			(0.125)	(0.144)
L.Retail Ratio			-0.016	0.013
			(0.050)	(0.059)
L.Liquidity Ratio			-0.208**	-0.175**
			(0.085)	(0.078)
L.Loan Loss Provisions Ratio			0.066	0.039
			(0.073)	(0.105)
L.Cost-Income-Ratio			0.001	0.001
			(0.003)	(0.003)
L.Return on Equity			0.218	0.166
			(0.195)	(0.207)
L.Interest Income Ratio			-0.002	-0.001
			(0.004)	(0.004)
Bank Fixed Effects	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes
Country x Time Fixed Effects	No	No	No	Yes
Observations	924	924	924	924
within-R2	0.016	0.069	0.122	0.120

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.

Table 4: The supervisory scrutiny channel.

	(1)	(2)	(3)
	High QA Quantity	High QA Potential Impact	High QA Duration
Post ST16 x Tested	-0.014	-0.031*	-0.008
	(0.016)	(0.016)	(0.024)
Post ST16 x Tested x 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
Time FE	Yes	Yes	Yes
Country x Time FE	Yes	Yes	Yes
Observations	924	924	924
within-R2	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.

Table 5: Alternative channels.

	(1) Market Discipline	(2) Capital Structure	(3) Capitalization
Post ST16 x Tested	-0.031*	-0.049**	-0.048**
	(0.018)	(0.021)	(0.024)
Post ST16 x Tested x High Transparency	-0.026		
	(0.029)		
Post ST16 x High P2G		0.026	
		(0.033)	
Post ST16 x Tested x High P2G		0.003	
		(0.035)	
Post ST16 x Low Voluntary Capital			-0.037*
			(0.022)
Post ST16 x Tested x Low Voluntary Capital			0.016
			(0.030)
L.Log(Assets)	-0.142***	-0.146***	-0.118***
	(0.038)	(0.040)	(0.037)
L.Regulatory Capital Ratio	-0.135	-0.149	0.111
	(0.195)	(0.188)	(0.171)
L.Voluntary Capital Ratio	-0.269*	-0.246*	
	(0.139)	(0.140)	
Bank Controls	Yes	Yes	(Yes)
Bank FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Country Time FE	Yes	Yes	Yes
Observations	924	924	924
within-R2	0.132	0.126	0.127

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 (3). Time fixed effects are dummies for each quarter. Country time fixed effects indicate the country of each bank's headquarters. 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. In Column (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 (3) *Low Voluntary Capital* is a dummy indicating on average below median voluntary capital buffers in the quarters before the stress test.

Figures

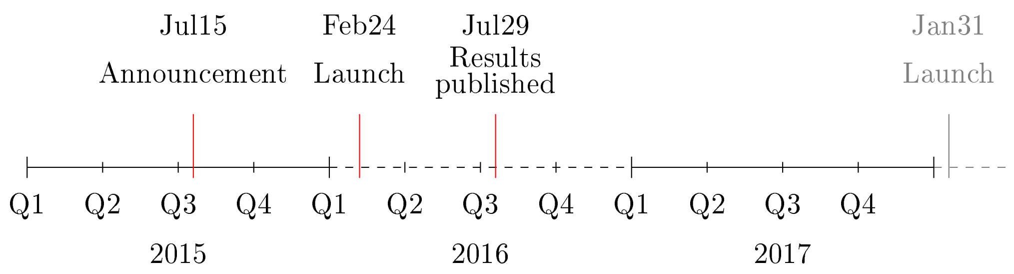


Figure 1: Timeline of the EU-wide 2016 Stress Test.

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).

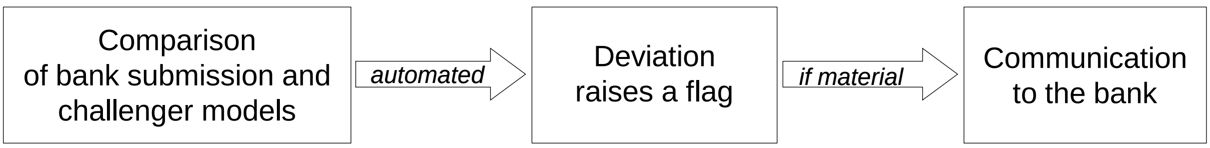


Figure 2: Simplified Quality Assurance process.

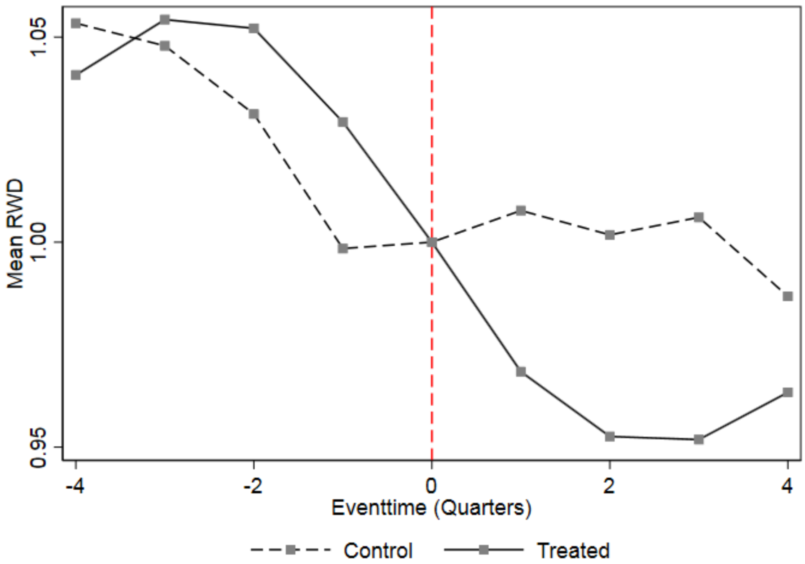


Figure 3: Time trends of RWD around the 2016 stress test by treatment.

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.

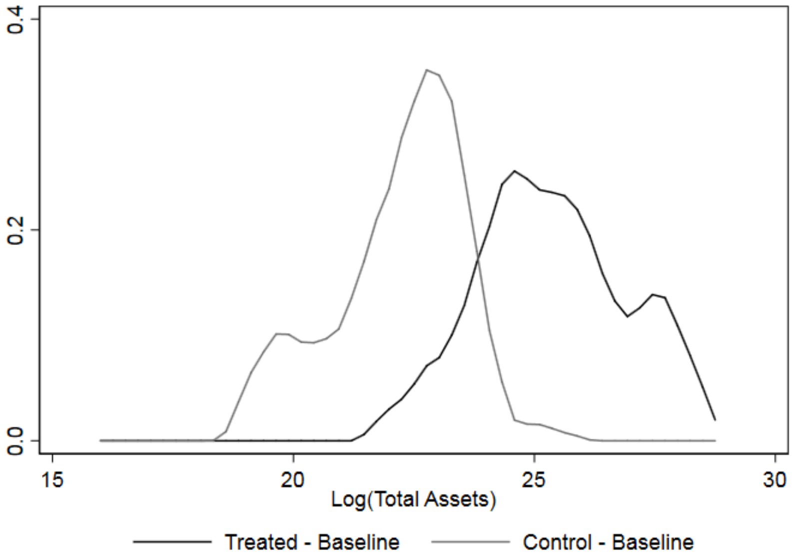
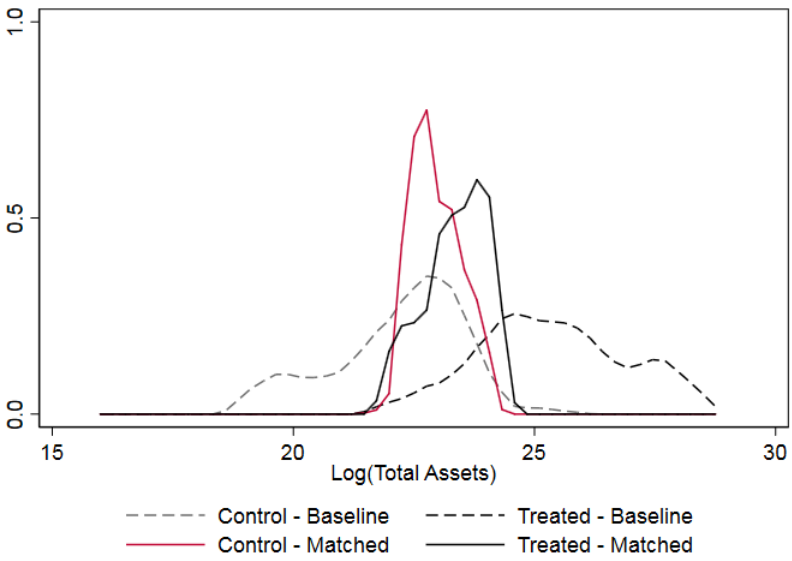
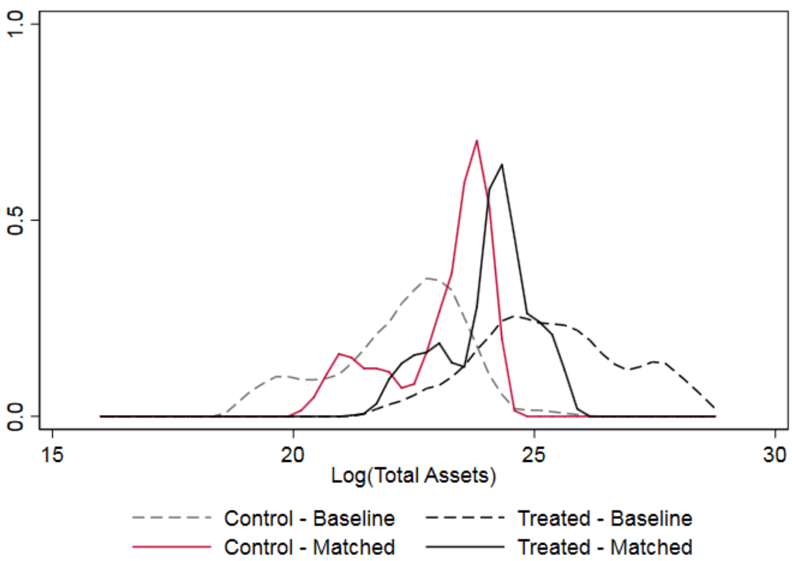


Figure 4: Distribution of Log(Assets) by treatment.

Notes: The figure shows estimated density functions of Log(Assets) for the treatment and control group which enter into the baseline estimation shown in Table 3 using Epanechnikov kernel function. Banks in the treatment group participated in the 2016 stress test. Banks in the control group did not.



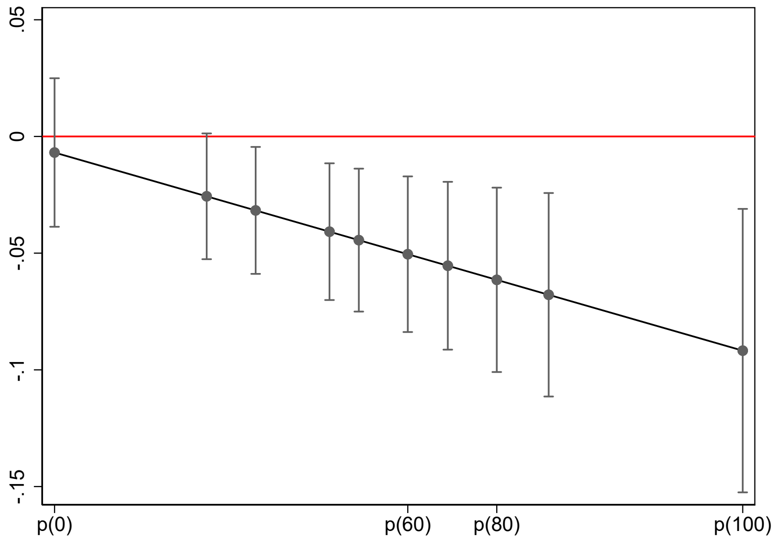
(a) Common Support



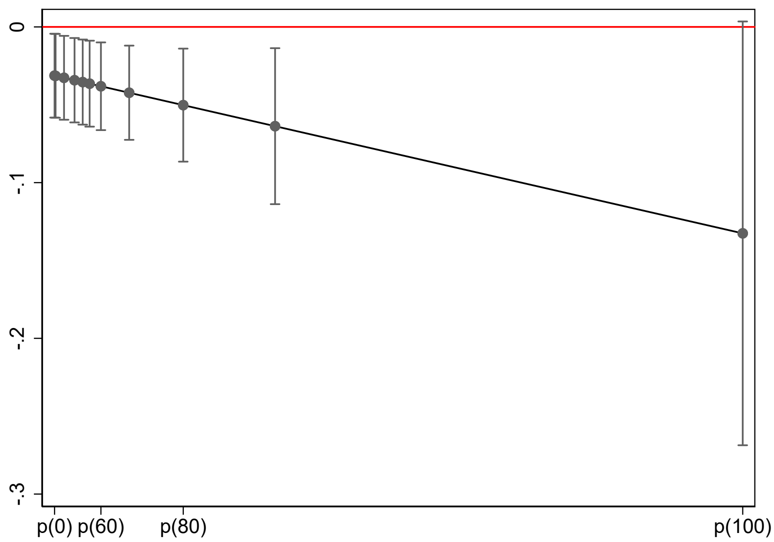
(b) Within Country

Figure 5: Distribution of Log(Total Assets) by treatment in the samples for matching.

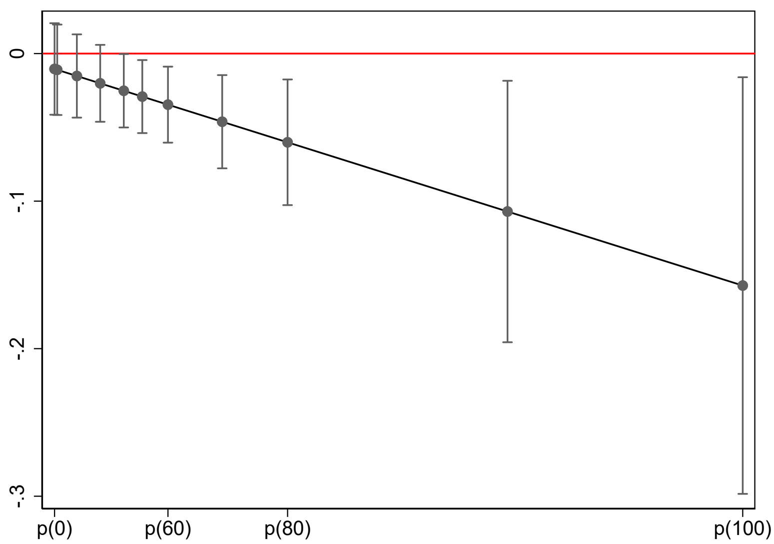
Notes: The figure shows estimated density functions of Log(Assets) by treatment using Epanechnikov kernel function. Dashed graphs show density functions of the groups in the baseline estimation. Banks in the treatment group participated in the 2016 stress test. Banks in the control group did not. Solid graphs show density functions of the treatment group (black) and the control group (red) in the matched samples after (a) discarding all banks without common support and (b) selecting only the two smallest treated and two largest control banks within each country.



(a) QA Quantity



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



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

Figure 6: Marginal effects along percentiles of QA

A Appendix

Table A1: Robustness with gradually decreasing sample sizes.

			(1)	(2)	(3)	(4)
			All Control	excluding Bottom p(25)	excluding Bottom p(50)	excluding Bottom p(75)
		N	69	51	34	17
(1)	All Tested	63	-0.042** (0.019) 924	-0.047** (0.020) 791	-0.051** (0.023) 665	-0.051** (0.021) 539
(2)	excluding Top p(25)	47	-0.041** (0.020) 812	-0.046** (0.022) 679	-0.050** (0.024) 553	-0.055** (0.023) 420
(3)	excluding Top p(50)	31	-0.030* (0.017) 700	-0.032* (0.017) 567	-0.030* (0.017) 441	-0.047** (0.020) 315
(4)	excluding Top p(75)	15	-0.020 (0.024) 581	-0.031 (0.028) 448	-0.028 (0.028) 322	-0.022 (0.025) 203

Notes: Clustered standard errors at the bank-level in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Each matrix entry shows the coefficient and standard error of *Post ST16* $\times$ *Tested*, as well as the number of observations from an estimation of eq. 1 as in Column (4) of Table 3, i.e. a difference-in-differences estimation with a 4-quarters before and after the 2016 stress test where *Post ST16* is a dummy for 2017Q1–2017Q4, and *Tested* is a dummy for stress-tested banks. Each regression includes lagged bank-level control variables, bank fixed effects, time fixed effects, and country$\times$time fixed effects. All regressions in Column (1) include all banks of the control group. Regressions in Column (2) exclude banks in the lower 25th percentile of the distribution of average size of the control group. Regressions in Column (3) exclude the 50th percentile, and in Column (4) the lower 75th percentile of the size distribution of control group banks. Regressions in Row (1) include all tested banks. Regressions in Row (2) exclude the upper 25th percentile of the distribution of average size of the treated banks. Regressions in Row (3) exclude the 50th percentile, and in Row (4) the upper 75th percentile of the size distribution of treated banks.

Table A2: Robustness with matching estimation strategies.

	(1) Common Support Sample	(2) Within Country Sample
Average Treatment Effect	-0.079***	-0.012*
on the Treated	(0.007)	(0.007)
Observations	55	47
Method	Nearest Neighbour	Nearest Neighbour
Metric	Mahalanobis	Exact
Number of matches	1:1	1:1
Variables for Matching	Bank-level covariates	Country

Notes: Bias-adjusted standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. In Column (1) matching was performed only on banks with common support on the *Log(Asset)* distribution, i.e. all control banks smaller than the smallest treated bank and all treated banks larger than the largest control bank are excluded. Banks were matched on bank-level covariates (*Log(Assets), Voluntary Capital, Liquidity, Retail, LLP, CIR, RoE*, and *Interest Income*) using a Mahalanobis distance metric with 1:1 nearest neighbour matching. In Column (2) matching was performed only on a sample of banks in which the two largest control and two smallest treated banks per country were included. Some countries are dropped because there was either no control or no treated bank. Banks were matched exactly on the country with 1:1 nearest neighbour matching.

Table A3: Alternative measures of credit risk.

	(1)	(2)	(3)	(4)
Dependent	Moody's EDF	SNF z-score	Debt Ratio	NPL Ratio
(Specification)	(baseline)	(collapsed)	(baseline)	(baseline)
Post ST16 x Tested	-1.275*	0.674**	0.001	-0.012
	(0.640)	(0.270)	(0.002)	(0.010)
L.Log(Assets)	-1.558	0.047	0.054***	-0.036
	(1.234)	(1.253)	(0.009)	(0.024)
L.Regulatory Capital Ratio	-4.958	4.156	-0.207***	0.120
	(5.966)	(7.034)	(0.055)	(0.155)
L.Voluntary Capital Ratio	-9.696*	3.739	-0.152***	0.140
	(5.387)	(3.909)	(0.043)	(0.169)
L.Retail Ratio	1.014	1.421	0.033***	-0.024
	(2.461)	(1.645)	(0.012)	(0.059)
L.Liquidity Ratio	2.312	0.251	0.020	0.006
	(3.315)	(1.980)	(0.015)	(0.064)
L.Loan Loss Provisions Ratio	6.893	-0.125	0.007	0.594**
	(6.609)	(0.621)	(0.025)	(0.248)
L.Cost-Income Ratio	2.162***	0.092	-0.000	-0.008***
	(0.549)	(0.100)	(0.001)	(0.002)
L.Return on Equity	-16.247*		-0.051	0.230*
	(8.122)		(0.054)	(0.135)
L.Interest Income Ratio	-2.618***	-0.114	0.000	0.009***
	(0.665)	(0.115)	(0.001)	(0.003)
Bank FE	yes	yes	yes	yes
Time FE	yes	yes	yes	yes
Country x Time FE	yes	yes	yes	yes
Observations	299	212	924	918
R2	0.92	0.96	0.98	0.92
Tested Banks	32	51	63	63
Mean *Dependent*	1.347	1.857	0.925	0.12
(SD *Dependent*)	(2.549)	(2.136)	(0.041)	(0.164)
Non-tested Banks	14	55	69	69
Mean *Dependent*	1.551	2.382	0.893	0.139
(SD *Dependent*)	(3.511)	(3.013)	(0.07)	(0.177)

Notes: Clustered standard errors at the bank-level in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The table shows regressions on alternative measures of risk. In Columns (1) to (3) the same estimation as in Column (4) Table 3 is performed according to eq. 1 where *RWD* is replaced as the dependent variable by a different measure indicated in the column heads. In Column (1) *Debt Ratio* is the ratio of total debt to total assets. In Column (2) *NPL Ratio* is defined as non-performing loans over total loans where NPLs are all loans reported as past due over 30 days. In Column (3) Expected Default Frequencies (EDFs) are provided by Moody's Analytics which measure the probability of default within the next year. In Column (4) we estimate relying 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 as a control variable due to collinearity.

Table A4: Decomposition of RWD.

	(1) RWD (SA)	(2) Log(RW Exposure)	(3) Log(Total Exposure)
Post ST16 x Tested	-0.042**	0,042	0.101
	(0.018)	(0.078)	(0.088)
L.Log(Assets)	-0.142***	0.535***	1.003***
	(0.039)	(0.140)	(0.162)
L.Regulatory Capital Ratio	-0.134	-3.023**	-2.128
	(0.186)	(1.486)	(1.758)
L.Voluntary Capital Ratio	-0.247*	-1.057*	-0.070
	(0.142)	(0.595)	(0.586)
L.Retail Ratio	0.009	0.439	0.312
	(0.058)	(0.278)	(0.325)
L.Liquidity Ratio	-0.178**	0.525	1.317**
	(0.078)	(0.458)	(0.544)
L.Loan Loss Provisions Ratio	0.035	-0.201	-0.188
	(0.109)	(0.283)	(0.189)
L.Cost-Income Ratio	0.001	0.006	0.003
	(0.003)	(0.008)	(0.005)
L.Return on Equity	0.166	0.736	-0.245
	(0.205)	(0.712)	(0.573)
L.Interest Income Ratio	-0.001	-0.008	-0.004
	(0.004)	(0.010)	(0.006)
Bank FE	yes	yes	yes
Time FE	yes	yes	yes
Country x Time FE	yes	yes	yes
Observations	924	924	924
R2	0.994	0.993	0.992
Tested Banks	63	63	63
Non-tested Banks	69	69	69

Notes: Clustered standard errors at the bank-level in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The table shows a repetition of the baseline estimation as in Column (4) in Table 3 in which the dependent variable *RWD* is replaced by one of its components. In Column (2) the dependent variable is *Log(RW Exposure)* defined as the logarithm of risk-weighted credit risk exposures, i.e. the nominator of *RWD*. In Column (3) the dependent variable is *Log(Total Exposure)* defined as the logarithm of total credit risk exposures, i.e. the denominator of *RWD*. In Column (3) RWD is measured according to the regulatory approach used to report credit risk exposures. In Column (3) only exposures reported under the Standardized Approach (SA) are included.

Table A5: Robustness to different time spans and averaging over time.

	(1)	(2)	(3)
Post ST16 = 0	15q1-16q1	15q1-16q1	15q1-15q4 averaged
Post ST16 = 1	16q4-17q4	16q4-18q3	17q1-17q4 averaged
Post ST16 x Tested	-0.035**	-0.033**	-0.047**
	(0.015)	(0.014)	(0.021)
L.Log(Assets)	-0.133***	-0.134***	-0.143***
	(0.036)	(0.032)	(0.038)
L.Regulatory Capital Ratio	-0.078	-0.070	0.277
	(0.208)	(0.209)	(0.555)
L.Voluntary Capital Ratio	-0.233*	-0.251**	-0.165
	(0.127)	(0.121)	(0.348)
L.Retail Ratio	0.002	-0.003	0.031
	(0.043)	(0.039)	(0.080)
L.Liquidity Ratio	-0.159**	-0.177***	-0.227**
	(0.065)	(0.059)	(0.107)
L.Loan Loss Provisions Ratio	0.002	-0.030**	-0.257
	(0.023)	(0.011)	(0.181)
L.Cost-Income Ratio	0.002	0.002**	0.003
	(0.003)	(0.001)	(0.009)
L.Return on Equity	0.225	0.224	0.179
	(0.151)	(0.143)	(0.363)
L.Interest Income Ratio	-0.003	-0.002*	-0.003
	(0.003)	(0.001)	(0.010)
Bank Fixed Effects	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes
Country x Time Fixed Effects	Yes	Yes	Yes
Observations	1,188	1,318	264
within-R2	0.097	0.105	0.196

Notes: Clustered standard errors at the bank-level in parentheses in Columns (1) and (2), robust standard errors in parentheses in Column (3): *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Columns (1) to (2) show difference-in-differences estimations with differing definitions of *Post ST16* dummy. In Column (2) we include 2016Q1 in the pre-period and 2016Q4 in the post-period. In Column (2) we include additionally the first three quarters of 2018. In Column (3) we collapse the time dimension in the data to a panel with two periods (pre and post) by averaging all variables according to the baseline definition where the pre-period spans all four quarters of 2015 and the post-period spans all four quarters of 2017. Covariates enter as averages and not lagged into the regression of Column (3). *Tested* is a dummy for stress-tested banks. Bank-level control variables are lagged by one quarter. Time fixed effects are dummies for each time period, i.e. quarters in (1) and (2), pre-dummy and post-dummy in (3). Country-time fixed effects indicate the country of each bank's headquarters.

Table A6: Robustness with continuous QA measures.

	(1) QA Quantity	(2) QA Potential Impact	(3) QA Duration
Post ST16 x Tested	0.012	-0.031*	0.011
	(0.026)	(0.016)	(0.031)
Post ST16 x Tested x QA	-0.027*	-0.333	-0.025*
	(0.014)	(0.268)	(0.014)
Bank Controls	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Country x Time FE	Yes	Yes	Yes
Observations	924	924	924
within R2	0.141	0.133	0.133

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*. Time fixed effects are dummies for each quarter. Country time fixed effects indicate the country of each bank's headquarters. In Column (1) *QA* is a continuous measure of QA Quantity defined as the logarithm of the number of flags communicated to the banks with respect to credit risk. In Column (2) *QA* is a continuous measure of 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) *QA* is a ordinal measure of QA Duration defined as the number of cycles during which a bank received flags with respect to credit risk.

Table A7: Reverse causality.

	(1)	(2)	(3)	(4)	(5)	(6)
	QA Quantity		QA Potential Impact		QA Duration	
Average RWD pre-ST16	0.233	0.132	0.018	0.013	-0.427	-0.574
	(0.435)	(0.474)	(0.024)	(0.032)	(1.233)	(1.710)
Bank Controls	no	yes	no	yes	no	yes
Observations	63	63	63	63	63	63
R2	0.130	0.004	0.070	0.006		

Notes: Robust standard errors at the bank-level in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Columns (1) to (6) show regressions of RWD of stress-tested banks averaged over the four quarters before the stress test on different measures of QA with or without further controls. In Columns (2), (4), and (6) bank-level controls are included as averages over the pre-period. Controls comprise *Log(Assets), Voluntary Capital Ratio, Regulatory Capital Ratio, Liquidity Ratio, Retail Ratio, LLP Ratio, CIR, RoE*, and *Interest Income Ratio*. Columns (1) to (4) show OLS estimations. Columns (5), and (6) estimate ordered logit models.

Table A8: Alternative explanations for high stress-testing intensity.

	(1)	(2)	(3)	(4)
Alternative ST Intensity measures:	Log(Assets)	Retail Ratio	Loan Growth	IRB Share
Post ST16 x Tested	0.023	-0.013	-0.014	0.001
	(0.021)	(0.015)	(0.016)	(0.015)
Post ST16 x Tested x High QA Quantity	-0.059***	-0.059***	-0.056***	-0.045**
	(0.021)	(0.022)	(0.021)	(0.020)
Post ST16 x Tested x *Alternative ST Intensity*	-0.013*	0.025	0.026	-0.063*
	(0.007)	(0.033)	(0.036)	(0.035)
Bank Controls, Bank FE, Time FE, Country x Time FE	Yes	Yes	Yes	Yes
Observations	924	924	923	924
R2	0.945	0.945	0.944	0.946
Post ST16 x Tested	0.016	-0.007	-0.008	0.002
	(0.023)	(0.024)	(0.024)	(0.020)
Post ST16 x Tested x High QA Duration	-0.038*	-0.042*	-0.041*	-0.025
	(0.022)	(0.022)	(0.022)	(0.020)
Post ST16 x Tested x *Alternative ST Intensity*	-0.010	0.010	0.027	-0.075**
	(0.007)	(0.031)	(0.037)	(0.037)
Bank Controls, Bank FE, Time FE, Country x Time FE	Yes	Yes	Yes	Yes
Observations	924	924	923	924
R2	0.943	0.943	0.943	0.944

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*. Time fixed effects are dummies for each quarter. Country time fixed effects indicate the country of each bank's headquarters. In Column (1) *Alternative ST Intensity* is a *Log(Assets)*. In Column (2) *Alternative ST Intensity* is *Retail Ratio*. Both estimations in Column (1) and (2) are robust to excluding the respective variable from the set of lagged control variables. In Column (3) *Alternative ST Intensity* is *Loan Growth*. Loan Growth is defined as the quarter-on-quarter growth rate of credit risk exposures. In Column (4) *Alternative ST Intensity* is *IRB Share*. IRB Share is defined as the share of credit risk exposures that a bank reports under either of the IRB approaches relative to total credit risk exposures.

Table A9: Supervisory scrutiny, capital structure, and market discipline channel.

	(1)	(2)	(3)	(4)	(5)
Post ST16 x Tested	0.004	0.004	-0.008	-0.030	-0.014
	(0.020)	(0.021)	(0.021)	(0.023)	(0.023)
Supervisory scrutiny					
Post ST16 x Tested	-0.052**	-0.050**	-0.055***		
x High QA Quantity	(0.022)	(0.022)	(0.020)		
Post ST16 x Tested	-0.000	-0.001		-0.024	
x High QA Pot. Impact	(0.026)	(0.026)		(0.025)	
Post ST16 x Tested	-0.023	-0.021			-0.043*
x High QA Duration	(0.018)	(0.020)			(0.022)
Market discipline					
Post ST16 x Tested		-0.020	-0.024	-0.020	-0.010
x High Transparency		(0.030)	(0.028)	(0.029)	(0.029)
Capital structure					
Post ST16 x Tested		0.014	0.011	0.021	0.028
x High P2G		(0.021)	(0.020)	(0.021)	(0.021)
Bank Controls	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Country x Time FE	Yes	Yes	Yes	Yes	Yes
Observations	924	924	924	924	924
R2	0.945	0.945	0.945	0.943	0.943

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*. Time fixed effects are dummies for each quarter. Country time fixed effects indicate the country of each bank's headquarters. In Column (1) we include all three QA intensity measures in triple interaction terms. In Column (2) we add interactions with *High Transparency* and *High P2G*. In Columns (3) to (5) we test one of the QA intensity measures jointly with the capital structure and market discipline measures.

Table A10: The supervisory scrutiny channel in the subsample of stress-tested banks.

	(1)	(2)	(3)	(4)	(5)
Supervisory scrutiny					
Post ST16 x High QA Quantity	-0.040**			-0.040*	-0.037*
	(0.019)			(0.023)	(0.019)
Post ST16 x High QA Pot. Impact		-0.014		0.004	
		(0.019)		(0.023)	
Post ST16 x High QA Duration			0.029	-0.020	
			(0.025)	(0.022)	
Capital structure					
Post ST16 x High P2G					0.015
					(0.020)
Market discipline					
Post ST16 x High Transparency					-0.012
					(0.024)
Bank Controls	Yes	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
Country x Time FE	Yes	Yes	Yes	Yes	Yes
Observations	413	413	413	413	413
R-squared	0.947	0.945	0.945	0.947	0.947

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. *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. In Column (5) we add interactions with *High Transparency* and *High P2G*.

Table A11: Alternative explanations for high stress-testing intensity in the subsample of stress-tested banks.

	(1)	(2)	(3)	(4)
Alternative ST Intensity measures:	Log(Assets)	Retail Ratio	Loan Growth	IRB Share
Post ST16 x High QA Quantity	-0.044**	-0.057**	-0.039**	-0.036*
	(0.020)	(0.022)	(0.019)	(0.019)
Post ST16 x *Alternative ST Intensity*	-0.018*	0.132*	0.134	-0.025
	(0.009)	(0.071)	(0.108)	(0.033)
Bank Controls	Yes	Yes	Yes	Yes
Bank FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Country x Time FE	Yes	Yes	Yes	Yes
Observations	413	413	412	413
R-squared	0.948	0.949	0.947	0.947

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. The sample includes only stress-tested 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 Column (1) *Alternative ST Intensity* is a *Log(Assets)*. In Column (2) *Alternative ST Intensity* is *Retail Ratio*. Both estimations in Column (1) and (2) are robust to excluding the respective variable from the set of lagged control variables. In Column (3) *Alternative ST Intensity* is *Loan Growth*. Loan Growth is defined as the quarter-on-quarter growth rate of credit risk exposures. In Column (4) *Alternative ST Intensity* is *IRB Share*. IRB Share is defined as the share of credit risk exposures that a bank reports under either of the IRB approaches relative to total credit risk exposures.

Table A12: Definitions of variables.

Variable	Definition
Variables at the bank-quarter level that enter the baseline estimation	
RWD	Risk-weighted density. Average risk weight of credit risk exposures in the banking book according to standardized and internal-ratings based approach.
Tested	Dummy equal to 1 for financial institutions that took part in the EU-wide stress test 2016.
Post ST16	Dummy equal to 1 in the four quarters after the EU-wide stress test 2016 starting 2017q1 and equal to 0 in the four quarters before the stress test starting 2015q1.
Log(Assets)	Bank size measured as total balance sheet book value of assets in logarithm of EUR mil.
Regulatory Capital Ratio	Tier 1 capital over risk-weighted assets according to CRD IV requirements. This adds up capital bound to comply with Pillar 1 ratios, Pillar 2 requirements and guidance, Capital Conservations Buffer, Countercyclical Buffer, SRB buffer, O-SII buffer, and G-SIIB buffer where applicable.
Voluntary Capital Ratio	Book value of total equity capital minus capital bound to comply with regulation (see above) over total assets.
Retail Ratio	Retail deposits plus retail loans over total assets.
Liquidity Ratio	Liquid assets over total assets. Liquid assets are defined as cash and central bank reserves.
Loan Loss Provisions Ratio	Loan loss provisions over total loans.
Cost-Income Ratio	Total administrative costs over total income.
Return-on-Equity	Net earnings before taxes over total equity calculated using average earnings from a rolling window of 4 quarters.
Interest Income Ratio	Net interest income over total net income.

Notes: This table provides a description of the main variables used for the empirical analysis reported in the paper.

Table A13: Definitions of stress test intensity variables.

Variable	Definition
Supervisory scrutiny channel	
QA Quantity	Sum of the number of flags communicated to banks during QA related to credit risk projections. A flag is raised if the projection of a data point in a credit risk related template in one of the supervisory challenger models deviates in a non-trivial way from the submitted projection of the bank. A flag is communicated to a bank if the deviation is considered to be material. We exclude flags solely related to data quality issues.
High QA Quantity	Dummy equal to 1 for banks with *QA Quantity* above the median.
QA Potential Impact	Sum of the accumulated potential impact of all flags (as in *QA Quantity*) received. Impact of a flag is calculated as the difference between the final CET1 depletion using the supervisory projection causing the flag and final CET1 depletion according to the bank's submission. Accumulated impact sums up these differences by bank.
High QA Potential Impact	Dummy equal to 1 for banks with *QA Potential Impact* above the median.
High QA Duration	Dummy equal to 1 for banks that received flags related to credit risk projections in more than one cycle during QA.
Capital structure channel	
High P2G	Dummy equal to 1 for banks that were subject to Pillar 2 Guidance (P2G) higher than the median P2G in 2017q1. In 2017q1, Pillar 2 Guidance was based on the stress test results as well as the full SREP evaluation.
Low Voluntary Capital	Dummy equal to 1 for banks that had below median *Voluntary Capital Ratios* before the stress test. *Voluntary Capital Ratio* is averaged over the four quarters of 2015 to calculate the median.
Market discipline channel	
High Transparency	Dummy equal to 1 for banks whose stress test results were published on a institutional level on the EBA website. These were Significant Institutions whose accumulated assets account for 60% of the European market.

Notes: This table provides a description of the stress test intensity variables used for the empirical analysis reported in Section 7 in the paper.