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

Christoffer Kok[†], Carola Müller[‡], Cosimo Pancaro[§]

Preliminary — Please do not circulate

Abstract

Using a difference-in-differences approach and relying on confidential supervisory data available at the European Central Bank related to the 2016 EU-wide stress test, this paper presents novel evidence that supervisory scrutiny that comes with stress testing can have a disciplining effect on bank risk. Exploiting the institutional design of the European test, we are able to disentangle the effect of supervisory scrutiny from other channels through which stress tests can affect bank risk, most notably stress-test-induced capital measures. We find that banks that participated in this exercise subsequently reduced their credit risk relative to banks that were not part of the stress test. This disciplining effect is stronger for banks that received tighter scrutiny during the test. We construct several metrics of stress test intensity so that we can shed light into the black box of supervisory action and contribute evidence on the effectiveness of supervisory scrutiny

JEL Classification: G11, G21, G28.

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

*We thank Rym Ayadi (discussant), Diane Pierret (discussant), Thorsten Beck, Enzo Cerletti, Reint Gropp, Anna Kovner, David Marqués-Ibáñez, Juan Francisco Martínez, Steven Ongena, Matías Ossandon Busch, Kasper Roszbach, and participants at the ECB Macroprudential Stress-testing Conference for their helpful comments. All errors are our own. This paper should not be reported as representing the views of the European Central Bank (ECB), Norges Bank, or any affiliated institution.

[†]European Central Bank, Sonnemannstrasse 20, 60314 Frankfurt am Main, Germany.

[‡]Norges Bank, P.O. Box 1179 Sentrum, 0107 Oslo, Norway; Halle Institute for Economic Research (IWH), Kleine Märkerstrasse 8, 06108 Halle, Germany, E-mail: `carola.mueller@iwh-halle.de`.

[§]European Central Bank, Sonnemannstrasse 20, 60314 Frankfurt am Main, Germany.

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. The subtle differences in the design, purposes, and implementation of stress tests around the globe makes it difficult to categorize and evaluate this policy tool. One important question is whether stress tests contribute to financial stability by reducing risk in the banking sector.

Stress tests build a bridge between macroprudential perspectives and microprudential supervision. With the aim of fostering banks' risk management practices, stress tests also gained momentum in microprudential frameworks. Mandatory supervisory stress tests require elaborate methods and know-how to assess institution-wide and system-wide risks. Resources have to be employed to develop the needed techniques on the side of supervisors as well as supervised firms.[1] Stress testing often involves supervisory inspection of banks' risk management as well as confidential communications about techniques and best practice. Mostly due to the confidentiality of supervisory actions, we know little of the effectiveness of these efforts. Do risk management capabilities that are 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 test, we are able to disentangle the effect of supervisory scrutiny from other channels through which stress tests can affect bank risk. Most notably, the European design offers a good testing ground to differentiate effects coming from stress-test-induced capital measures and responses to the procedure of the test itself. The use of confidential data allows us to shed light into the black box of supervisory action which constitutes a great part of stress testing. We therefore contribute to the rare evidence on the effectiveness of supervisory scrutiny. 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 steps. First, we test if stress tests can affect bank outcomes at all. The main contribution of our analysis is displayed in the second part. We derive several hypotheses and test implications about how stress tests might affect bank risk while paying special attention to the supervisory scrutiny channel. We use a difference-in-difference approach which compares the change in risk-taking around the 2016 EU-wide stress test exercise of banks that were

[1] Andrea Enria, Chair of the Supervisory Board of the ECB, pointed out "cost-efficiency (for both supervisors and the industry)" among others as a guiding principle for the future of stress tests (Enria, 2019).

tested and banks that were not tested. We use bank-level supervisory data exploiting detailed information on credit risk exposures, accounting figures, and bank-specific 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).

Our first-stage results 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 also economically significant as it amounts to a change in RWD of about 20 percent of the standard deviation of RWD of tested banks. Hence, stress testing can have a significant effect on bank outcomes, such as 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 Significant Institutions (SIs).[2] Hence, our treatment group consists of nearly all SIs[3] while our control group of the Less Significant Institutions (LSIs). This implies that in our sample of banks there is a sizeable difference in the average total assets between 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 control group of LSIs is comparable to the tested banks. In order to alleviate the concern that our results are driven by the differences between the control and treatment group, we additionally employ matching estimation techniques and we show the robustness of our results when estimating with a more homogeneous subsample of banks.

In the second step, we derive three channels that can potentially explain the result. Stress tests include supervisory inspections and demand a high amount of additional reporting from banks. We hypothesize that stress tests might reduce bank risk because they enhance supervisory scrutiny. We label this the supervisory scrutiny channel. Further, capital measures, such as additional capital requirements or capital distribution limits, might ensue due to stress test results. Bank capital is an important determinant of banks' risk choice (Calem and Rob, 1999; Berger and Bouwman, 2013). Hence stress tests might influence bank risk in as much as they are accompanied with capital measures. We name this the capital structure channel. And lastly, stress test results are published and may contain relevant information. Market participants could use this information in order to demand changes in bank risk or punish risky behaviour. This can be viewed as a market discipline channel of stress tests.

We examine whether the identified reduction in credit risk is associated with the additional scrutiny, which supervisors exert during the exercise, or with one of the alternative channels. In order to test the channels, we make use of variation in scrutiny, stress-test informed capital

[2]Significant institutions are defined as those SSM (Single Supervisory Mechanism) 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).

[3]EBA decided to exclude Greek SIs from the exercise because they were impacted by the sovereign debt crisis in Greek as well as some SIs that were under restructuring. For the same reasons, we also exclude them from our control group.

requirements, and the granularity of published information. We extend our baseline difference-in-differences setting including a triple interaction term which represents the intensity of treatment with respect to any one of the three channels, i.e. we look at the intensive margin of being stress tested.

We find that our baseline result can be explained through the supervisory scrutiny channel but we cannot find evidence for the capital structure or market discipline channel. Our results show that banks that received more supervisory scrutiny reduce risk weighted density more than banks that were under less intense scrutiny. We construct three metrics for supervisory scrutiny making use of data produced during the so-called Quality Assurance during the 2016 stress test. Quality Assurance is a process between bank and supervisor preceding the publication of stress test results. The European macroprudential stress test follows a constrained bottom-up approach. In bottom-up stress tests banks run their own stress test models based on a common methodology and predefined macro-financial scenarios in order produce stress test results.[4] In the Quality Assurance the ECB challenges banks' projections with its own thereby ensuring credibility of the results. If a bank's internal model-based projections materially deviate from the supervisor's, a process to discuss and possibly revise them is launched. We simplify this process by calling it a communication between bank and supervisor. We measure scrutiny as the unweighted number of communications, the importance-weighted number of communications, and the duration of communication cycles. In detail, our results show that stress-tested banks reduced credit risk more, the more communications they had, the more important these communications potentially were for the final result, and the more cycles of communications a bank had to go through.

In light of 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, Berger, and Roman, 2018; Pierret and Steri, 2018). It reflects, however, the fact that the capital structure channel is by design less central in the European exercise. Stress test results of the 2016 EU-wide stress test were used to inform supervisory decisions about capital guidance but the capitalizing the banking sector was less key to the stress test as there were other policies in place that focused on bank capital. This less deterministic relationship between capital requirements and stress test results makes the European stress test an ideal testing ground to study the otherwise hardly distinguishable scrutiny channel. We therefore tailored our analysis to examine the supervisory scrutiny channel. Similarly, we cannot support the notion that banks whose results were published on a more granular level reduced credit risk more than those banks whose results were published only in aggregate. Hence, we cannot find evidence that the market disciplining channel can explain our baseline result. All in all, these findings provide novel evidence that the tighter and more intrusive scrutiny associated to the EU-wide stress-test has the potential to enhance banks' risk management practices and to induce lower risk-taking.

Our study contributes to the literature in several ways. First, we contribute to the growing

[4] This is in contrast to top-down stress tests. In a top-down approach, supervisory authorities run their models in order to produce stress test results.

literature evaluating the effect of stress tests on bank outcomes. Closest to ours are the studies focussing on the effect of stress tests on risk, Pierret and Steri (2018) and Acharya et al. (2018). Pierret and Steri (2018) examine the effect of U.S. stress tests on risk-taking and are the first to point out how decisive it is to separate the capital structure 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 did not participate. They reckon that this effect is additional to risk-taking incentives due to an increase in capital linked to the stress test results. Our findings are able to refine theirs since we are able to measure each of the proposed channels separately due to the distinctive design of the European stress tests. Acharya et al. (2018) also support the view that the U.S. stress test decreased banks' risk-taking. They find that stress tested banks increase spreads on loans and decrease credit supply especially in riskier market segments. While Acharya et al. (2018) argue that banks' more prudent behaviour is driven by the channel of higher capital[5], Pierret and Steri (2018) find an increase of risk-taking through the capital channel which is mitigated through the scrutiny of the stress test exercise. Building on their findings, we control for the capital channel albeit in contrast to the U.S. stress test outcomes of the European tests do not automatically trigger higher capital requirements.

Several other studies confirm that banks reduced risk after being stress tested. These studies are based U.S. data. Cortés, Demyanyk, Li, Loutskina, and Strahan (2020) and Calem, Correa, and Lee (2020) provide evidence of increased spreads and reduced credit supply in the mortgage market. Flannery, Hirtle, and Kovner (2017) find no evidence for changes in bank portfolios (risk shifting) after stress test publications. We complement these findings by providing evidence of the experience in Europe and by shedding more light on the channels through which stress tests can affect risk-taking. Especially, we can comment on the duality of banking supervision and banking regulation, a narrow distinction, in this particular policy tool.[6]

Therewith, we contribute to the literature on the effectiveness of banking supervision - to be understood in opposition to banking regulation in the form of capital requirements - and the interplay between the Basel Pillars (capital adequacy, supervisory review, and market discipline). Based on a cross-country comparison, Levine and Barth (2001) point out that regulatory measures which enhance information disclosure and monitoring can be more effective than capital requirements in promoting financial stability. Buch and DeLong (2008) show that regulatory scrutiny in a more general definition can deter risk-taking. They find that banks shift risks away from countries with strong supervision. Delis and Staikouras (2011) demonstrate that supervisory review and market discipline complement each another in reducing bank risk while capital requirements matter most for those banks close to the boundaries.

Our findings complement several papers that provide micro-evidence for a link of bank risk and supervision by exploiting variation in the intensity of supervisory scrutiny. Gopalan, Kalda, and Manela (2017), Lim, Hagendorff, and Armitage (2016), and Kandrac and Schlusche (2019) use the presence of supervisors' offices; Eisenbach, Lucca, and Townsend (2016), and Hirtle,

[5]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 in itself.

[6]We refer to banking regulation as quantitative rules, mostly capital requirements, and other written guidelines, checking the compliance to which falls in the realm of banking supervision by the means of on- and off-site audits, reporting, inspections, and other qualitative tasks of bank supervisors.

Kovner, and Plosser (2019) look at their hours worked at supervised banks; Rezende and Wu (2014) employ the frequency of inspections; Ivanov and Wang (2019) study the assignment of supervisors for the revision of syndicated loans; and Bonfim, Cerqueiro, Degryse, and Ongena (2020) relate on-site inspections and zombie lending practices. They unanimously advocate for a disciplining effect of supervisory scrutiny. Delis and Staikouras (2011) on the other hand find a more complex relationship between supervision and bank risk. They study data from 17 countries over a decade and point out that supervisory sanctions are associated with a decline in bank risk but supervisory audits at first increase bank risk and only with a certain intensity seem to reduce bank risk. None of the above, however, studies confidential supervisory acts in connections with stress tests.

While the notion that supervisors are able to discipline banks is fairly abstract, one explanation can be found in a change of governance structures through an increased presence of supervisory interests. Chaly, Hennessy, Menand, Stiroh, and Tracy (2017) provide indicative evidence that increased supervision can reduce misconduct risk and contribute to a different risk culture. Another explanation could be that not the conversation itself but the topic of conversation might have a lasting impact. Goldsmith-Pinkham, Hirtle, and Lucca (2017) use linguistic analysis to classify the topic of supervisory actions. They document that especially in large banks over half of supervisory actions concern risk modelling and internal controls. Our measures of supervisory scrutiny would map into this category since the communications we study are about stress testing models and methodologies. In this respect, we comment on 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 (Behn, Haselmann, and Vig, 2016; Plosser and Santos, 2018; Mariathasan and Merrouche, 2014; Begley, Purnanandam, and Zheng, 2017). With respect to 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 scarce evidence for the potential benefits for the use of internal risk management tools in bank supervision. Since the valuable interaction between supervisors and banks that we document here arises within a bottom-up approach and out of the necessity to check the produce of banks' internal models, the results are useful to inform the on-going debate about the future design of stress tests.

The paper is structured as follows. Section 2 derives hypotheses about potential channels through which stress testing might affect bank outcomes, such as credit risk. Section 3 gives an overview of the institutional setting of the EU-wide stress test in 2016. In Section 4 we describe the estimation methodology, the variables we employ, and the metrics we construct to measure stress testing intensity. The following Section 5 contains an overview of the data sources we use and describes the final sample we use in the analysis. In the next Section 6 we show our baseline result of the impact of being stress tested on bank risk and its robustness. Section 7 contains the analysis about the channels which can explain how stress testing can affect bank risk. Finally, we conclude our arguments in Section 8.

2 Hypotheses about stress testing and bank risk

We briefly describe different channels through which stress tests might affect bank risk. We explain how these channels are connected to specific features of stress tests and how we derive testable hypotheses based on those features. The following arguments highlight that stress tests cannot be easily classified. Eisenbach et al. (2016) describe them as an example of the "gray area [...] between supervision and regulation". Their features address each Pillar of the Basel principles: capital adequacy, supervisory review, and market discipline.

2.1 Supervisory scrutiny channel

Stress tests can be viewed as an instrument with intense scrutiny. If successful, more supervisory scrutiny should reduce bank risk. Recent literature provides at least two plausible explanations for the disciplining effect of bank supervision. First, supervision might improve risk management and bank governance practises. Second, it might produce relevant information about risks and malpractices that lead to corrective actions.

Supervision is based on interactions between supervisors and supervised in the form of information exchange (reporting), communications and meetings, as well as on-site and off-site inspections. 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. Incentives favouring more conservative risk attitudes aligned with supervisory views might gain strength. Increased supervision might also reduce misconduct risk and contribute to a different risk culture (Chaly et al., 2017).

The more intrusive supervision gets, the more weight supervisors are bound to 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 act of supervision - without the researcher's further knowledge about the specific issues that are addressed by this supervision - appears to be effective. Several studies document a disciplining effect of more intense supervison. They measure the intensity of supervision by the mere presence of supervisors' offices (Gopalan et al., 2017; Kandrac and Schlusche, 2019) or their hours worked at a supervised banks (Eisenbach et al., 2016; Hirtle et al., 2019).

Furthermore, additional supervisory efforts can produce new information by detecting unrecognised 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) demonstrate that the riskiness of syndicated loans can be affected by the assignment of examiners for their review. Bonfim et al. (2020) show that more on-site inspections reduce the probability to engage in zombie lending at the supervised firm. Further, Delis and Staikouras (2011) document that on-site audits and sanctions affect bank risk. And lastly, DeYoung, Flannery, Lang, and Sorescu (2001) provide evidence that supervisory inspections produce information that affects market prices and can therefore be considered new and valuable.

Several aspects of stress tests imply higher scrutiny of supervisors. One aim of supervisory stress tests is to improve risk management practices. Prior to the great financial crisis, stress tests were primarily employed in the evaluation of trading portfolios, not to assess the situation of a whole banking group, and they were in early stages of development (Baudino, Goetschmann, Henry, Taniguchi, and Zhu, 2018). Their mandatory use for regulatory purposes requires banks to invest resources in the development of stress testing techniques, especially in case of bottom-up stress tests where banks themselves have to run models in order to make loss 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, EBA stress tests evoke discussions of stress testing techniques between supervisors and supervised banks. In the annual Comprehensive Capital Analysis and Review (CCAR), the Federal Reserve paired a quantitative assessment in form of a stress test with a qualitative assessment in which risk management is being scrutinized.[7]

Furthermore, stress tests collect and generate a high amount of new quantitative information. For example, banks had to fill-in 35 templates for the 2016 EBA stress test. Of the results, EBA acknowledged to publish about 16.000 data points per bank which is only a subset of the data actually created per bank (EBA, 2016c). The main aim of this information production is an assessment of risk drivers, in particular with respect to tail risks of exposures to adverse macro-financial developments. Taking stock price reactions as a judge for the information content of stress test results, Petrella and Resti (2013) and Morgan, Peristiani, and Savino (2014) show that stress tests indeed produce insights that investors consider valuable. If vulnerabilities are detected, corrective actions might ensue. Bank management or risk managers themselves might use insights gained from the supervisory exercise to implement new or different strategies.

These arguments underline that stress tests might reduce bank risk by intensifying supervisory scrutiny. But a less engaging story can be told. Banks may view supervision and stress tests in particular as a mere 'check-the-box constraint' (Moody's Analytics, 2013) where the only reason to adjust are sanctions or the threat thereof. Stress tests can have sanctions. Supervisors can obligate banks to withhold capital distribution, raise equity, or adapt risk exposures. We handle these possible sanctions in the following as a distinct channel of how stress tests can affect bank risk. 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.

[7] For example, former Fed Governor Tarullo stressed in a speech the importance of combining the quantitative and qualitative assessment (Tarullo, 2016).

2.2 Capital structure channel

We now turn our attention away from the process and implementation of stress tests and direct it on the potential consequences of stress test results. The overarching question behind stress tests is: Do banks have enough capital to survive severe adverse macroeconomic and financial conditions? If the answer is negative, supervisors respond by imposing capital measures, such as capital requirements or limits to dividend distribution plans. The relationship between capital requirements and bank risk has been intensely studied but remains controversial. Nevertheless, we find it within our realm to state that the intention of the supervisory measures is to make banks safer. Hence, we assume that stress tests might reduce bank risk if they are accompanied by regulatory capital measures.

There are several possible explanations how the level of bank equity capital might influence banks' risk taking decisions. We refer to those channels illustrated in Acharya et al. (2018). To start with, there is (a) a mechanical connection through risk-weighted capital requirements. In order to meet higher requirements, banks can either reduce exposures or reduce the average risk weight of their exposures. Further, there is (b) a moral hazard channel, and (c) a charter value channel. Both reflect that banks have a higher stake in their survival which decreases incentives to exploit deposit insurance and too-big-too-fail guarantees and to gamble on future cash flows by taking more risk. In line with this reasoning, higher bank capital should reduce bank risk. However, according to (d) a reach-for-yield channel, the opposite could also be true. Higher equity stakes reduce the profitability of banks' investments which is why banks might be tempted to compensate this by making riskier investments.[8]

Overall, stress tests make a statement about capital adequacy. The immediate consequences for bank capital plans differ though across jurisdictions and chosen stress test design. These differences can be visible by comparing, e.g., the European and US stress tests. Acharya et al. (2018) call the US stress test (SCAP in 2009 and CCARs) "essentially forward-looking capital requirements". It reflects the fact, that stress tests are very tightly connected to the evaluation of capital plans for the largest Bank Holding Companies in the US. Therefore, a bad stress test outcome will most certainly result in higher capital.[9] For euro area banks on the other hand, the connection is less deterministic. EBA clarified in a statement 2016 that the determination of bank-specific capital requirements would be informed by the stress test results but would also incorporate other information (EBA, 2016d). Furthermore, stress test results factor into Pillar 2 capital guidance, breaches of which do not trigger the same automated legal enforcement actions as breaches of Pillar 1 requirements. This renders the European case a good testing ground for distinguishing the channels described here. In summary, we hypothesize that tested banks that are subjected to more stringent capital requirements because of the stress test would show lower risk after the stress test exercise.

[8]Furthermore, an increase in risk-insensitive capital requirements could also set incentives to reshuffle investments from safe to riskier ones. Stress test-related capital requirements however require banks to increase capital relative to risk-weighted assets.

[9]In detail, CCAR results change the minimum capital level. Capital below this level triggers automatically dividend distribution restrictions.

2.3 Market discipline channel

Stress test might involve another form of scrutiny that is exercised by other stakeholders than supervisory authorities. Conditional on the publication of supervisory information produced in stress tests, market participants that trade equity or debt of stress-tested banks might scrutinize banks. If information asymmetries are reduced through transparency measures, market participants can price bank risk more accordingly which can have a disciplining effect.[10]

As mentioned above, stress tests produce a vast amount of information part of which is regularly shared with the public. Various studies have shown that the information released about the stress test is used and valued by actors in financial markets (Petrella and Resti, 2013; Morgan et al., 2014; Georgescu, Gross, Kapp, and Kok, 2017; Ahnert, Vogt, Vonhoff, and Weigert, 2018; Fernandes, Igan, and Pinheiro, 2020; Flannery et al., 2017; Lazzari, Vena, and Venegoni, 2017). However, the release of too negative information might trigger bank runs or a breakdown of interbank risk sharing agreements as the ultimate disciplining devise of market participants. This trade-off is reflected in the debate on optimal disclosure policy for stress tests (Goldstein and Sapra, 2013; Goldstein and Leitner, 2018). Furthermore, the correct interpretation of stress test results needs a high level of methodological knowledge making the information content hard to process for laypersons. Nevertheless, improving transparency is one of the objectives of stress tests.

Therefore, regulatory authorities publish aggregate or bank-specific results in form of data points as well as more accessible summaries. Parts of stress test results and procedures however remain confidential. In as far as a market disciplining channel is at work, we expect a stronger reaction the more information is available to market participants.

3 The 2016 EU-wide stress test

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

It is conducted following a so called constrained bottom-up approach[12]. Under this approach, banks generate stress test projections using their own models, relying on a common predefined macro-financial scenario[13] and subject to a pre-set methodology. Under these guidelines, banks have to fill-in and submit a number of pre-defined templates. These are structured along risk categories and accounting items. For example, they require separate projections on the development of credit risk exposures and market risk exposures as well as projections of the development of certain items of the profit & loss account or balance sheet. Finally, the templates track the impact of the evolution of risk exposures under the two scenarios on bank capitalization. The

[10] A summary of empirical work on market discipline in banking provide Flannery and Nikolova (2004) for the US and Gropp (2004) for Europe.

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

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

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

final result of the stress test is often summarized in the Core Equity Tier 1 (CET1) capital ratio depletion under the adverse scenario, i.e. the CET1 ratio banks would end up with at the end of the adverse scenario if exposures and balance sheets evolve as projected.

Furthermore, in this exercise, the ECB quality assures the stress test results of the banks under its direct supervision. During this quality assurance (QA) process, the ECB as a competent authority, reviews and challenges banks' projections to ensure their credibility. 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. 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[14], 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 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[15].

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

In total, 107 banks participated in this exercise, 93 of which were SSM Significant Institutions (SI). Among these, 51 banks were part of the EBA stress test sample which only included banks with a minimum of EUR 30 bn in assets(EBA, 2016b). Additional 56 banks, which were also SSM significant institutions, 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[16].

Overall, the 2016 EU-wide stress-test results showed that SSM banks improved their resilience

[14]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.

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

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

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 9.1The 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, 2016). The 2016 SREP process consisted for the first time of two parts: Pillar 2 capital requirements and Pillar 2 capital guidance[17]. 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 determination 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. The qualitative results of the stress test could even have an impact on Pillar 2 requirements.

4 Methodology

We first analyse the effect of banks' participation in the 2016 EU-wide stress test on bank risk. From our hypotheses as well as the existing evidence from US stress tests, we generally expect stress testing to affect bank risk. In a second step, we turn to disentangle different channels derived in section 2 how stress testing might affect bank risk.

4.1 Baseline estimation

We investigate whether banks that were stress tested showed a significantly different risk-taking behaviour after the stress test than banks that were not stress tested. To this end, our empirical strategy relies on a difference-in-differences approach where we use the stress test as a treatment by estimating the following equation:

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

where the dependent variable $Risk_{i,t}$ is the measure of risk for bank i in period (t). Our main yardstick for risk-taking 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. $Treated_i$ 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-

[17] 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.

fixed effects, which is included in the regressions to control for loan demand effects. Being aware that 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 risk-taking.[18] We lag these control variables by one quarter to reduce endogeneity concerns. They comprise the *Regulatory Capital Ratio*, which allows disentangling the effects of supervisory scrutiny and higher capital requirements, and the *Voluntary Capital Ratio*, which means the capital held by banks in addition to the amount required by the regulations 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 *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 control and treatment groups. Given these controls, we require that there are no further unobservable time-varying differences between the control and treatment group banks for our analysis to be valid.

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 significant institution under the ECB direct supervision. Indeed, all banks in our treatment group are significant institutions while for the control group we have to rely on a sample of large SSM Less Significant Institutions (LSIs)[19]. This implies that we cannot claim that our treatment is randomly assigned. Instead, 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). Yet, 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 the matching estimator of Abadie and Imbens (2011) as a robustness check.

[18] Table A8 in the Appendix provides detailed definitions of the variables.

[19] Less significant institutions are SSM banks that do not fulfil any of the significance criteria to be qualified as significant institutions. Less significant institutions are not under the direct supervision of the ECB. They are directly supervised by the National Competent Authorities under the oversight of the ECB which ensures the consistency of the regulatory framework and supervisory practices applied to these banks.

4.2 Channels and stress test intensity measures

After establishing this baseline estimation given in Eq. 1 to test whether there is an external margin in being stress tested, we continue to investigate the internal margin of being tested by defining various intensity of treatment measures related to different design features of the stress test. These measures are constructed to test the three channels outlined in Section 2 through which stress tests could affect bank risk. We exploit variation in supervisory scrutiny, stress test related capital requirements, and stress test transparency by adding a triple interaction term to the baseline regression. Hence, the following regression is estimated:

$$\begin{aligned} Risk_{i,t} = \quad & Post_t \times Tested_i + Post_t \times Tested_i \times High\ Intensity\ _i \\ & + Bank_i + Time_t + Country_i \times Time_t + Controls_{i,t-1} + \epsilon_{i,t} \end{aligned} \tag{2}$$

We are particularly interested in the significance and sign of the estimated coefficient of the interaction term between *Post, Tested* and *High Intensity.* For the ease of interpretation and comparability, we construct *High Intensity* dummy variables that separate tested banks at the median level of the respective intensity measures used. Where applicable, we also present results on continuous increases in stress test intensity.

We focus on exploring the supervisory scrutiny channel. The QA process and the related supervisory scrutiny and interactions between the ECB and the banks provide information about the variation in the intensity of the QA process across banks in the sample that can be exploited, as a measure of the intensity of the treatment. More specifically, three different measures are used to capture the intensity of the QA process. These measures are built relying on ECB proprietary information. The first yardstick, *QA Intensity*, measures the number flags triggered during the QA. We use only flags raised on credit risk related issues that ended up being communicated to the banks participating in the stress test. *High QA Intensity* therefore indicates the group of tested banks that received more than the median number of flags. This measure is a proxy of the amount of interactions, which took place between the supervisors and the banks during the QA. Some flags might be more important than others. A low number of flags might therefore be misleading. The second measure, *QA Effectiveness*, takes this into account. For each flag, we observe the impact of the flagged deviation on final CET1 ratio depletion, i.e. how much lower is CET1 ratio depletion of the bank's projection versus the supervisor's projection. *QA Effectiveness* takes the sum of the potential impact on CET1 ratio depletion of all communicated credit risk flags. Again, to construct *High QA Effectiveness*, we split at the median. This measure, therefore, provides a weighted sum of the intensity of the QA owing to the fact that flags with higher potential impact on the final stress test results might entail more debate or receive more attention. Finally, the third measure, *QA Duration*, is an indicator ranging between 1 and 3 depending on the number of cycles during which a bank was communicated credit risk flags. This indicator reflects the length of the interactions between the supervisors and the banks and corresponds roughly with measures like hours worked per bank as in (Hirtle et al., 2019).

In order to test the capital structure channel, we associate a 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 divide banks in two groups at the median. Since we expect a stronger effect for banks with low capital buffers, we substitute *High Intensity* in Eq. 2 here with a dummy *Low Voluntary Capital* indicating below median capital buffers before the stress test.

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. We exploit the dichotomous distinction in the way stress test results were published. Results of banks in the EBA sample were published on a granular institutional level while only aggregate results of the SREP sample were published that did not allow to extract bank-specific information. Hence, we construct a dummy *High Transparency* indicating banks in the EBA sample.

5 Data

5.1 Data sources and sample restrictions

We use quarterly bank-level data from supervisory databases[20] provided by the European Central Bank for the period 2015q1 to 2017q4. The data comprises credit risk exposures, regulatory capital ratios, and balance sheet as well as profit and loss accounting data. Furthermore, we use proprietary data of the ECB on banks' submissions and the quality assurance process of the 2016 stress test.

The sample of banks participating in the stress test is selected by the European Banking Authority. It comprises systemically important banks in the European Union whereby those that are located in the Euro Area are supervised by the European Central Bank under the Single Supervisory Mechanism. Overall, 51 banks were tested under EBA mandate and an additional 56 banks as part of the Supervisory Review and Evaluation Process (SREP). Results were only published for the 51 EBA banks which were defined by a threshold of EUR 30 bn in consolidated assets. These banks account for a share of over 70% of bank assets in Europe (EBA, 2016b). Because the ECB has no authority to supervise banks outside the Euro Area, our data does not cover EBA banks that participated in the stress test. Hence, we have data on 93 stress tested banks, 39 of which are EBA banks. In terms of geography, we constrain our analysis to exposures and risk-taking in the European Union.

We use banks that did not participate in the stress test as a control group. These are mostly less significant institutions. The sample of LSIs for which we have accounting data from the supervisory dataset is limited due to restricted reporting requirements. Therefore, we can construct covariates from balance sheet information only for 175 of 369 banks for which we have credit risk exposure and regulatory capital ratios, and covariates from profit & loss accounts can only be defined for 81 of 369 banks that potentially serve as control group.

[20] We use Corep and Finrep data collected under SSM mandate.

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 also 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 that were stress tested within or outside of the Euro Area (subsidiaries or branches). Hence, we only consider banks at the highest level of consolidation.

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

5.2 Sample description

Table 1 presents descriptive statistics for covariates of the treatment and control group banks in the pre-treatment period. We present the statistics separately for banks in the treatment group, i.e. banks that participated in the 2016 EU-wide stress test, and banks in the control group. The stress test treatment was not selected randomly. Instead selection is based on observables, namely the status of being systemically important or not. 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 to total assets, and lower loan loss provisions relative to total loans than control group banks. Further, judging from difference in means tests according to Column (3) of Table 1, stress-tested banks seem quite similar in terms of performance measures like the Return-on-Equity and Cost-to-Income ratio as well as economic capital which they hold on top of regulatory required capital provisions. Unsurprisingly, minimum required capital does not differ significantly between both groups, since all are subject to the Single Rule Book and the impact of bank-specific macroprudential buffers is still small compared to core equity ratios in our period of observation.

Imbens and Wooldridge (2009) point out that p-values of difference in means tests can be misleading in large sample sizes 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. We report normalized differences in Column (4) of Table 1. According to this rule, our estimation can handle the aforementioned differences with the exception of the normalized difference in *Liquidity Ratio* which is outside the range with a value of -0.404 and, as expected, the normalized difference in *Log(Assets)* which is 1.584.

Being aware that these stark differences in size could bias our estimates and curtail comparability of control and treatment group, we devote some energy providing robustness regarding the estimators as well as samples. For robustness, we make use of cross-country variation in the distribution of bank size. Because the status of systemic importance is defined by country, cross-country standard variation of total assets within the treatment group is quite high. The smallest stress tested bank has a balance sheet size of 4.2 bn EUR and the median tested bank has a balance sheet size of 94 bn EUR which is below the biggest control bank with 111 bn EUR in total assets. Meanwhile we use logarithm of total assets as control variable in all our regressions as it well captures the selection into treatment.

[Table 1 around here.]

6 The impact of being stress-tested on bank risk

We start with a difference-in-differences analysis examining whether stress-tested banks change their credit risk level relative to banks that were not part of the stress test. In particular, we implement a test of Eq. 1.

6.1 Baseline results

An identifying assumption for this setting is that the change in outcomes, i.e. the trend in credit risk development, is comparable between the control and treatment group in the period before the stress test. If the outcome variable, credit risk, was on a comparable trend before the stress test but diverged between the two groups after the stress test, we 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 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 relative to an average 2016 level. The Figure collaborates 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 tested banks. Column (3) further documents that the mean RWD of stress-tested banks is significantly lower than average RWD of control banks in the pre-period (at 5% significant) as well as 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-taking behaviour. 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.]

We derive our baseline result of estimating Eq. 1 in Table 3. Column (1) shows that the magnitude of the coefficient remains literally unchanged once we add time fixed effects relative to the univariate analysis in Table 2. Nevertheless, since treatment is not assigned randomly to banks, we have reason to believe that this estimation is biased. Hence, we assess the conditional difference-in-difference estimator by including bank size in the form of *Log(Assets)* in Column (2) as a control variable in order to proxy for the status of systemically important banks which is the main variable that drives selection into treatment. We find that size is a relevant determinant for RWD levels which can be seen in the fact that the coefficient is significant at the 1% level as well as an increase of the explanatory power of our estimation with respect to within-bank variation. Further, the coefficient is negative which collaborates existing evidence that larger banks might pose more systemic risk, but are inclined to take less individual risk (Laeven, Ratnovski, and Tong, 2016). Thereby the effect has the same direction as the effect of being stress tested. We see that by conditioning on size, the probability that we have to reject our hypothesis that tested banks reduce credit risk after the stress test decreases. In Column (3) we add further control variables that might influence bank risk. While the signs on all coefficients take the expected direction, only two of them are significant: Banks with higher voluntary capital buffers as well as banks with a higher share of liquid to total assets show lower RWDs. In Column (4) we include fixed effects for the country of banks' headquarters interacted with each time period to capture demand conditions which vary on a country-level in addition to the pan-European trend captured in the time fixed effects. This takes into account that even the international big 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 that influence bank risk and might vary on a more local level of granularity nor for cross-country exposures of these international banks. Notwithstanding, we consider this our preferred specification to explain changes in RWD given the data we have available. We continue to see a negative significant coefficient on the term of interest stating that the reduction of 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 also economically significant as it amounts to a change in RWD of about 20 percent of the standard deviation of RWD of tested banks.

6.2 Robustness against the selection into being stress-tested

[Figure 4 around here.]

The biggest concern about these estimations continues to be the question whether the banks in the control group are comparable to the banks in the treated group. The comparability lacks most in terms of the size as is illustrated in Fig. 4 which shows the distribution of *Log(Assets)* for the treatment and control group. We address these concerns with two approaches. First, we check whether our estimated effect of a 4.2 percentage point decrease in RWD of stress-tested banks relative to non-tested banks is solely driven by this specific sample of banks. Therefore, we re-estimate our results by gradually reducing the sample on both ends of the distributions. Second, we employ a different estimator in applying two matching strategies recently used in

Gropp, Mosk, Ongena, and Wix (2019) that are more capable to balance the differences between the groups.

Table A1 in the Appendix summarizes the results of our first approach to reduce concerns that our results are purely driven by the differences in asset size. The idea is generally to converge the two distribution functions shown in Figure 4 by gradually excluding the smallest banks in the control group and/or the biggest banks in the treatment group. Table A1 reports the coefficient of interest 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) with the baseline estimate of the full sample. In Column (2)[respectively (3)/(4)] we repeat the estimation by excluding the 18 [35/52] smallest banks in the control group which correspond to the bottom 25th [50th/75th] percentile of the distribution of average asset size of the control group. Similarly, we exclude the biggest banks of the treatment group in the top 25th [50th/75th] percentile in Row (2) [(3)/(4)]. The matrix illustrates that the coefficient and its significance level stays relatively constant by reducing the sample in this way. The estimates vary between -3.0 and -5.5 percentage points. Furthermore, the coefficient gets stronger when we compare stress-tested banks to larger banks in the control group pointing to the fact that including small banks works against our finding. The coefficient loses significance once we consider only the 15 smallest stress tested banks (see Row (4)). The small samples also implies that we are estimating with a reduced amount of degrees of freedom thereby statistical power as we approach the lower right corner. Overall, Table A1 shows that our results are neither driven by the very big nor the very small banks.

[Figure 5 around here.]

Nevertheless, given the selection on observables, we employ a matching estimator as the second robustness to show that our results are not coincidental with bank size. For that, we rely on two 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 our approach in Table A1 by excluding very small non-tested and very large tested banks. To be precise we implement 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 Mahalanobis distance between the covariates described in *Controls* 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 introduced. According to the results in Table A2 both estimates are negative and significant further collaborating our hypothesis that being part of the stress test impacts banks' risk-taking behaviour.

6.3 Further robustness

A general concern for any analysis on bank risk is to find an appropriate proxy for risk. We focus on bank credit risk and measure it as the risk-weighted density for credit risk exposures. It reports the average risk weight of all credit risk exposures that banks have to report according to

Basel guidelines (and their implementation in the Euro area) to their supervisory authorities. As such, it is based on reporting for regulatory purposes and might therefore misrepresent bank risk. First, credit risk is only part of overall bank risk. Second, banks have incentives to under-report 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 the first shortcoming, we test whether the participation in the stress test differentially affected alternative measures of bank risk which are not solely focused on credit risk. The results are reported in Table A4 in the Appendix. We find corroborating evidence indicating that banks that participated in the stress test reduce risk when we employ measures related to the default probability of banks. In Column (1) we find a negative significant effect on Expected Default Frequencies (EDFs) provided by Moody's Analytics which measures the probability of default within the next year according the firm (Moody's Analytics, 2011). In Column (2) we show that z-score, i.e. the distance to default, of tested banks increases relative to non-tested banks. To measure bank risk through this z-score we rely on balance sheet data provided by SNL Financials. This data is available on a yearly basis so we estimate Eq. 1 by averaging all covariates in the pre-test and post-test period. However, we cannot find a significant differential effect on bank leverage or the share of non-performing loans as documented in Columns (3) and (4) respectively.

Further, we show in Table A3 that our results are not driven by the manipulation of risk weight under the IRB approach. The results remain almost unchanged if we restrict our analysis only to exposures under the Standardized Approach, as depicted in Column (5). Lastly, we cannot find a medicine for banks unwilling to report risks. The only available data source that might incorporate unreported information is market data which unfortunately is not available for a reasonable number of non-tested banks. Therefore, our analysis cannot account for the possibility that tested banks shifted credit risk exposures into unregulated markets after the stress test exercise. Lastly, we provide evidence in Table A5 in the Appendix that our result is not driven by the choice of the time window we define as relevant to estimate the effect and that our result is not biased due to serial correlation present in the panel data which might lead to overestimation of significance in difference-in-differences settings according to Bertrand, Duflo, and Mullainathan (2004).

In Columns (1) and (2) of Table A5 we change the definition of the *Post ST16* dummy and re-estimate our results. 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. We find a negative and significant coefficient on the interaction of *Post* and *Treated* for more generous 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 in the positive serial correlation in panel data can lead to substantial overestimation of 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 employ 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

Based on the finding from the previous section that stress testing causes a significant difference in bank risk between banks that are tested and those that are not tested, we now continue to investigate how stress testing might bring this difference about. We accomplish this by exploiting variation in the extend to which banks were exposed to different features of the stress test. Taking the stress test as a treatment received by some banks, we turn to define the intensity of this treatment in different dimensions related to different aspects of the stress test. The features that we study are the intensity of the Quality Assurance process, the variation in stress-test induced capital requirements, and the degree of disclosure of stress test results. Thus, we examine what we call the supervisory scrutiny channel, the capital structure channel, and the market discipline channel, respectively, as described in more detail in Section 2.

7.1 The supervisory scrutiny channel

Stress tests are an intense supervisory exercise that lasts over several months. During the quality assurance of the stress test, regulators cross check banks' loss projections against their own top-down and peer benchmark models. In case of material deviations between predictions of the regulator and the bank, a dialogue with the bank is initiated which potentially leads to adjustments of banks projections. The process involves an exchange of opinion, a discourse on best practice, and ultimately supervisors' intrusion into risk management. It also generates a vast amount of information.

[Table 4 around here.]

We test the hypothesis that this fierce supervisory scrutiny associated with stress tests has an effect on bank risk. We make use of variation in the intensity of supervisory scrutiny which we measure with three metrics that we constructed from data produced in the Quality Assurance process. Table 4 shows our main results by estimating Eq. 2. Column (1) presents our main evidence for the existence of a disciplining supervisory scrutiny channel. We find that stress-tested banks that were exposed to high QA intensity significantly reduce their risk-weight density after the stress test relative to stress-tested banks that went through a less intense QA process. *High QA Intensity* indicates here that banks received more than the medium number of communications during the QA process. We estimate that banks decrease RWD by 5.6 percentage points more if they belong to the intense QA group relative to the other stress tested banks which amounts to a differential effect of about 13 percent of their pre-stress-test RWD. It illustrates that the baseline effect can be explained by the subgroup of tested banks that received an above median number of flags. We implement this test as well with a continuous measure of intensity where we use the logarithm of number of flags communicated with banks during the QA process.[21] The result which is shown in Column (1) of Table A7 confirms that banks reduce risk by 2.7 percentage points if they receive 1 percentage point more communications. This 1 percentage point increase corresponds roughly to a quintile in the distribution of the QA intensity measure. The upper panel (a) of Figure 6 depicts the marginal effects along different percentiles of the distribution of *QA Intensity*. As expected, as intensity gets stronger the effect is stronger and stays significantly different from zero.

The result in Column (2) indicates that the effectiveness of these communications do not seem to matter for the risk reduction effect since the coefficient on the triple interaction is insignificant. *High QA Effectiveness* is measured as above median potential impact of the flags on the final capital depletion. [22] This measure should reflect the fact that a bank might have a very intense QA process while receiving only very few communications if these communications can potentially have a very severe impact on the stress test final outcome. Using the underlying continuous measure of *QA Effectiveness* at first glance reinforces the insignificance of the effectiveness of flags. A closer look, 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 Effectiveness* except for the maximum. We cannot disclose the nature of this outlier, but when winsorizing at 5%, Figure 6 shows in the lower panel (c) that banks with a stronger effectiveness of communicated QA flags reduce their risk significantly more than banks with less effective 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.

All in all, these finding seem to indicate that there is a value in the discourse between regulators and banks that is held as part of the stress test exercise. Further the fact that banks are asked to explain or rethink their modelling choices seems to be more relevant than the potential importance of each issue. This might also reflect the fact that each of the issues that enter into the discussion with banks are already considered to have some importance. It also corroborates the idea that the supervisory process itself and the intrusion of supervisors into banks' spheres is the channel through which banks are disciplined.

7.2 The capital structure and the market discipline channel

We now turn to alternative explanations about how the stress test exercise can impact banks' risk-taking choices. We test two hypotheses. First, we test if banks are disciplined through market

[21] Log-levels were chosen due to high non-normality displayed in the absolute value of number of flags according to Shapiro-Wilk test.

[22] We also used the realized and not only the potential impact of QA on CET1 depletion to capture the effectiveness of the QA procedure and did not find any significant results neither.

participants because of the additional information that is made publicly available through the exercise. Second, we test whether the connection of stress tests to the capital structure of banks is the cause of banks' efforts to reduce risk. For that we examine the increase in regulatory capital charges that resulted from the exercise and we check whether only banks that entered with rather low capitalization into the test subsequently reduced their risk. The results are summarized in Table 5.

[Table 5 around here.]

In contrast to the aforementioned interaction that takes place between supervisors and banks under the realm of strict confidentiality during the exercise, stress tests enhance transparency in financial markets with the publication of results at the end of the exercise. They provide more detailed information about risk exposures as e.g. banks' annual reports. Various studies have shown that the information released about the stress test is used and valued by actors in financial markets (Petrella and Resti, 2013; Morgan et al., 2014; Georgescu et al., 2017; Ahnert et al., 2018; Fernandes et al., 2020; Flannery et al., 2017; Lazzari et al., 2017). If information asymmetries are reduced, market participants can price bank risk more accordingly which can have a disciplining effect. As mentioned in section 3, bank-specific information is only published for those banks that were selected by EBA. Hence, we should expect that a disciplining effect through the market is stronger only for these banks. In Column (1) of Table 5 we introduce a triple interaction with a dummy *High Transparency* indicating whether a bank is part of the EBA sample or whether it was stress tested as part of the SREP evaluation. We cannot determine a significant difference between EBA and SREP banks. Therefore, we cannot find evidence that the disciplining effect that we find in our baseline estimation is driven by increased transparency due to the publication of stress test results. However, this does not imply that stress test transparency does not increase market discipline. We acknowledge that our test is not optimally designed to fully answer that. The main limitation of our analysis is our focus on credit risk. While our scrutiny measure directly refers to credit risk issues as does our dependent variable, the stress test results published on the EBA website give a full picture of all risk classes. How market participants value individual parts of this information, we cannot tell. In so much, we go only as far as stating that the baseline effect that we find is not significantly connected to the degree of transparency of the 2016 stress test.

In Column (2) we test explicitly whether the increase in capital guidance after the stress test is driving our main finding. Several studies showed that banks decreased risk-taking after stress tests using U.S. data and connect this to associated increases in capital requirements (Acharya et al., 2018; Pierret and Steri, 2018). Contrary to the U.S., outcomes of the EU-wide stress test are not automatically followed by increases in regulatory capital requirements but as clarified in EBA (2016d) used to inform supervisors for setting their capital guidance. Pillar 2 capital guidances 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" that supervision "expects banks to comply with" (ECB, 2016). Hence, similar to capital requirements they set incentives to lower risk weighted assets in order to comply with the supervisory guidance. Pillar 2 guidance (P2G) that was informed by the 2016 stress test took effect in the first quarter of 2017. In Column (2) we introduce a dummy variable *High P2G* which we construct from supervisory data. It indicates an above median increase in P2G in 2017q1. It relates not to the level but the change in guidance which was informed by stress test outcomes. Since the triple interaction between treated banks, the post period, and banks that received above median P2G is insignificant, we cannot find evidence supporting the hypothesis that the reduction in risk-taking was driven by higher regulatory capital guidance.

Nevertheless, it could be that not the guidance itself but rather banks' ability to comply with additional requirements is relevant for the risk reduction effect. Then banks whose capitalization is closer to the regulatory requirements (including the Pillar 2 guidance), i.e. they have smaller voluntary capital buffers, might have more incentives to reduce risk in order to reduce the chances of breaching the regulatory rules. In Column (3) we therefore create and interact a dummy variable *Low Voluntary Capital* that determines banks with below median voluntary capital buffers in the period before the stress test. While we find that banks with lower capitalization reduced RWD after the stress test, we cannot find a significant difference between tested and non-tested banks. The former 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 others regarding the predominance of the capital structure channel in the US stress testing framework. This could simply reflect the fact that the European stress test design does not center around the evaluation of banks' capital plans. Our results might also be limited due to the fact that our dependent variable reflects only one way for banks to adjust to higher capital requirements, by reducing the average credit risk weight. Banks might therefore have reacted strongly to the stress test but in different dimensions, e.g. by delevering across all risk classes.

8 Conclusions

We examine the effect of the EU-wide Stress Test conducted in 2016 by the European Banking Authority and the European Central Bank on bank risk. We test the hypothesis that the increased supervisory scrutiny associated with the Constrained Bottom-Up approach of the exercise can discipline bank behaviour. In order to identify the effect, we compare stress tested banks with not tested banks in a difference-in-differences design where we examine the change in risk weighted densities (RWD) between the four quarters in 2015 leading up to the test and the four quarters in 2017 after the exercise. We find that stress tested banks reduced their RWD by 4.2 percentage points more than not tested banks. This is a rather sizeable change caused through the participation in the exercise which amounts to a move of about 20 percent of the standard deviation of RWD. We further showed 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 different channels that can possibly explain why banks reduced risk-taking in the aftermath of the stress test exercise. We exploit the institutional set-up and confidential data of the EU-wide stress test in order to construct measures of the internal margin of being stress tested. As part of the Constrained Bottom-Up setting, supervisors interact with banks whose stress scenario projections they check. We measure the intensity, effectiveness, and duration of these interactions. We find that banks that had more issues or issues over a longer time period to discuss with the supervisory authorities reduced RWD more than banks with fewer issues to discuss. These issues relate to the projections that are produced by banks' internal models. In the process of Quality Assurance, banks are asked to explain or revise these projections, respectively the underlying model. We hypothesize that the increased scrutiny through the presence and intrusion of supervisors in banks' risk management functions affects bank risk. We therefore conjecture that banks seem to incorporate some lessons learned during the process of Quality Assurance. We cannot find evidence that supports the hypothesis that banks reduce credit risk due to more market discipline connected to the publication of stress test results nor evidence that the risk reduction is driven by an increase of capital guidance as a result of the bank-specific assessment.

With respect to the discussion of stress test design, our results point out some merit in the use of a constrained Bottom-Up approach. It has to be noted though that this merit is not costless. One of the stress tests primarily objectives is to correctly assess banks' risk profiles. Our findings do not give information about how well this objective is met. The strategic under-reporting of loss projections possible under a Bottom-Up approach works against the credibility of the stress test outcomes. Supervisors have to employ additional effort and resources to detect and restrict this behaviour. Pursuing a more Top-Down approach would 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.

Baudino, P., Goetschmann, R., Henry, J., Taniguchi, K., and Zhu, W. 2018. Stress-testing banks - a comparative analysis. Financial Stability Institute Insights on policy implementation No 12, Bank for International Settlements.

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. and Rob, R. 1999. The impact of capital-based regulation on bank risk-taking. Journal of Financial Intermediation, 8(4):317–352.

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.

DeYoung, R., Flannery, M. J., Lang, W. W., and Sorescu, S. M. 2001. The information content of bank exam ratings and subordinated debt prices. Journal of Money, Credit and Banking, 33(4):900–925.

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. 2016 EU-wide stress test: Frequently Asked Questions. Technical Report 24 February 2016, European Banking Authority.

EBA. 2016c. EBA publishes 2016 EU-wide stress test results, European Banking Authority. https://eba.europa.eu/eba-publishes-2016-eu-wide-stress-test-results. Accessed: 2019-12-04.

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. 2016. 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.

Flannery, M. J. and Nikolova, S. 2004. Market discipline of u.s. financial firms: Recent evidence and research issues. In Borio, C., Hunter, W. C., Kaufman, G. G., and Tsatsaronis, K., editors, *Market Discipline Across Countries and Industries*, chapter 9, pages 87–100. MIT Press.

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.

Goldsmith-Pinkham, P., Hirtle, B., and Lucca, D. O. 2017. Parsing the content of bank supervision. FRB of NY Staff Report No. 770, Federal Reserve Bank of New York.

Goldstein, I. and Leitner, Y. 2018. Stress tests and information disclosure. Journal of Economic Theory.

Goldstein, I. and Sapra, H. 2013. Should banks' stress test results be disclosed? an analysis of the costs and benefits. Foundations and Trends in Finance.

Gopalan, Y., Kalda, A., and Manela, A. 2017. Hub-and-spoke regulation and bank leverage. Working Paper.

Gropp, R. 2004. Bank market discipline and indicators of banking system risk: The european evidence. In Borio, C., Hunter, W. C., Kaufman, G. G., and Tsatsaronis, K., editors, *Market Discipline Across Countries and Industries*, chapter 10, pages 101–118. MIT Press.

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.

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. Working paper.

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.

Levine, R. and Barth, J. 2001. Bank regulation and supervision: what works best? Policy Research Working Papers, The World Bank.

Lim, I., Hagendorff, J., and Armitage, S. 2016. Does distance impede regulatory monitoring? evidence from the banking industry. Working Paper.

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

Moody's Analytics. 2011. Edf overview. https://www.moodysanalytics.com/-/media/products/edf-expected-default-frequency-overview.pdf. Accessed: 2020-03-04.

Moody's Analytics. 2013. The evolution of stress testing in europe by Wilfrid Xoual. Moody's Analytics Risk Perspectives, 1.

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 and differences in means between treatment groups.

		(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	Treated	(T-C)	Control	Treated	(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 here shows 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 Quality Assurance channel.

	(1)	(2)	(3)
	High QA Intensity	High QA Effectiveness	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 Intensity 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 Effectiveness 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)	(2)	(3)
	Market Discipline	Capital Guidance	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

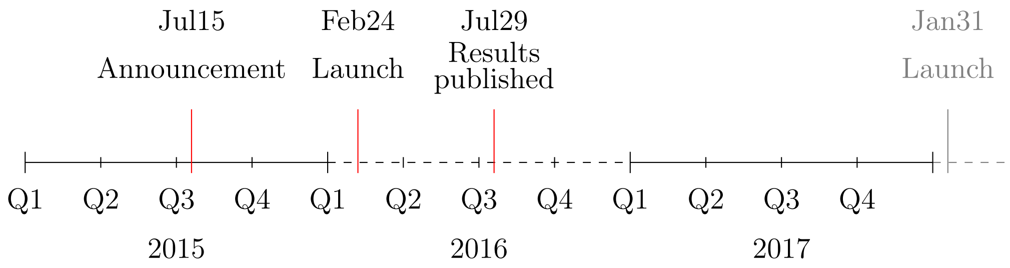


Figure 1: Timeline of the EU-wide 2016 Stress Test.

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).

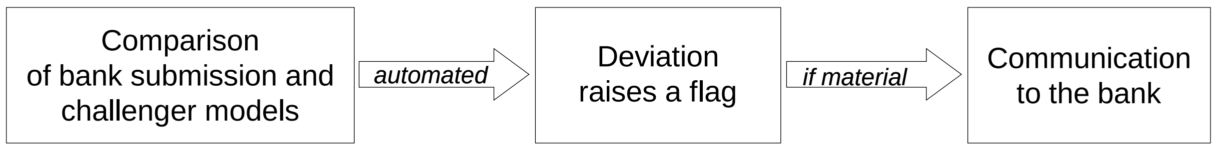


Figure 2: Simplified Quality Assurance process.

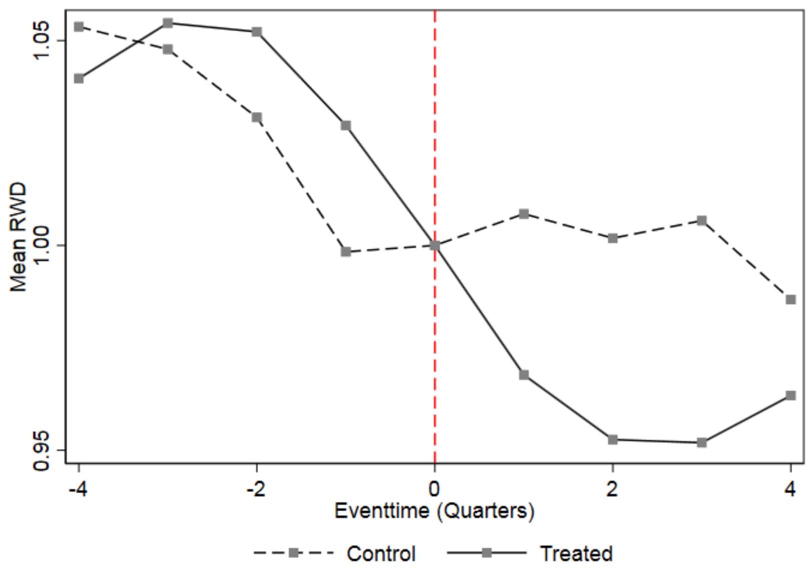


Figure 3: Time trends of RWD around the 2016 stress test by treatment.

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.

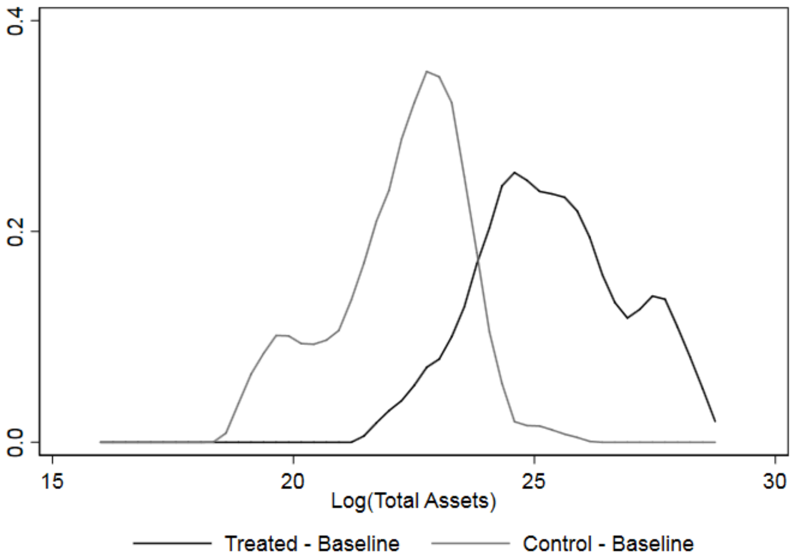
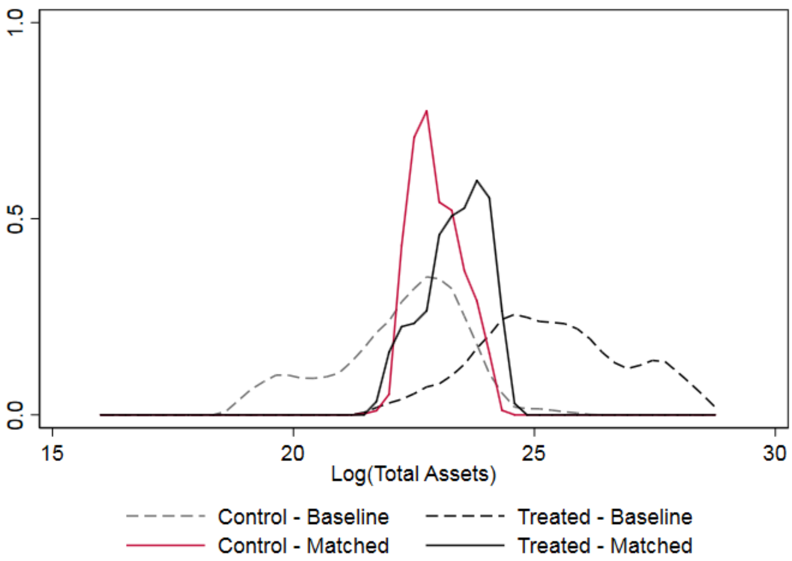
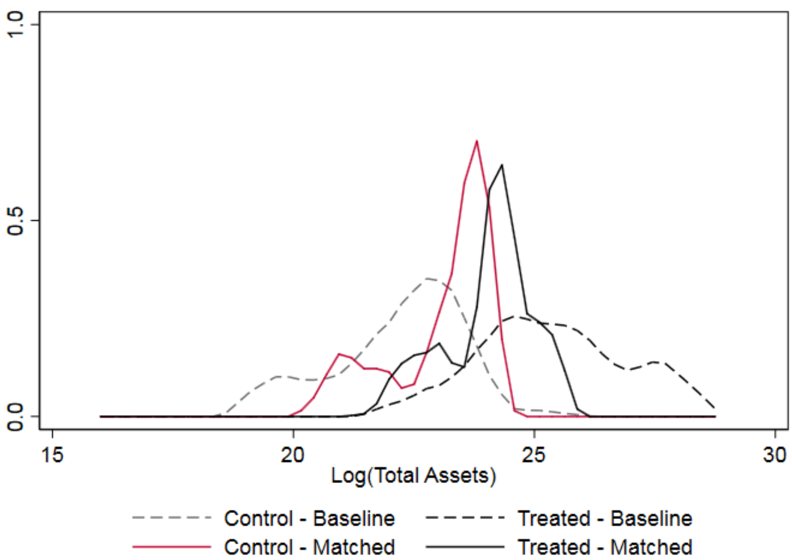


Figure 4: Distribution of Log(Assets) by treatment.

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.



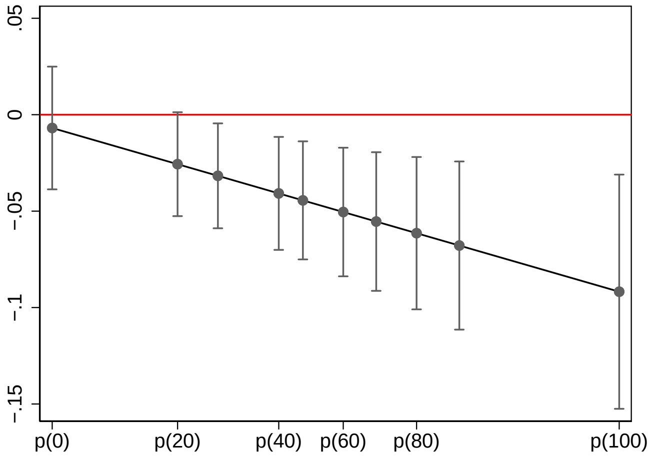
(a) Common Support



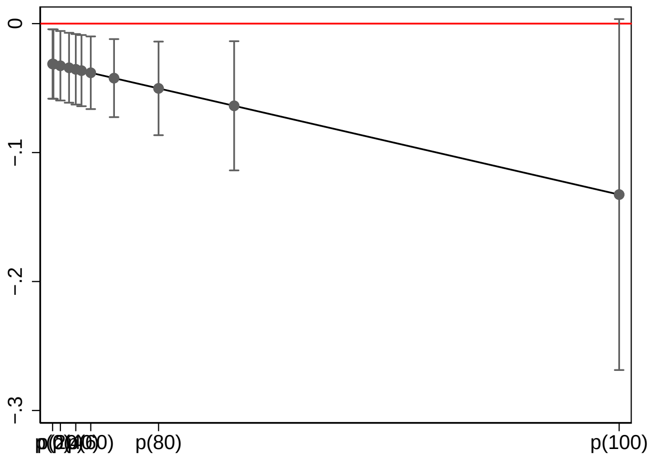
(b) Within Country

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

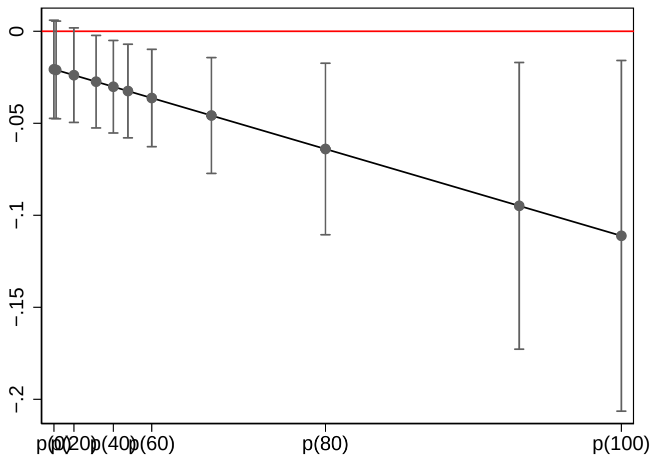
Notes: The figure shows estimated density functions of Log(Assets) by treatment using Epanechnikov kernel function. Dashed graphs show density functions of the groups in the baseline estimation. Banks in the treatment group participated in the 2016 stress test. Banks in the control group did not. Solid 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 intensity



(b) QA Effectiveness - Unwinsorized (baseline).



(c) QA Effectiveness - Winsorized at 5%.

Figure 6: Marginal effects along percentiles of QA measures.

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

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 Treated	63	-0.042**	-0.047**	-0.051**	-0.051**
			(0.019)	(0.020)	(0.023)	(0.021)
			924	791	665	539
(2)	excluding Top p(25)	47	-0.041**	-0.046**	-0.050**	-0.055**
			(0.020)	(0.022)	(0.024)	(0.023)
			812	679	553	420
(3)	excluding Top p(50)	31	-0.030*	-0.032*	-0.030*	-0.047**
			(0.017)	(0.017)	(0.017)	(0.020)
			700	567	441	315
(4)	excluding Top p(75)	15	-0.020	-0.031	-0.028	-0.022
			(0.024)	(0.028)	(0.028)	(0.025)
			581	448	322	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* × *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×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 on the Treated	-0.079***	-0.012*
	(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: Decomposition of RWD.

	(1) Log(RW Exposure)	(2) Log(Total Exposure)	(3) RWD (A-IRB)	(4) RWD (F-IRB)	(5) RWD (SA)
Post ST16 x Tested	0,042	0.101	0.017***	0.136	-0.042**
	(0.078)	(0.088)	(0.005)	(0.114)	(0.018)
L.Log(Assets)	0.535***	1.003***	0.040	-0.320	-0.142***
	(0.140)	(0.162)	(0.029)	(0.295)	(0.039)
L.Regulatory Capital Ratio	-3.023**	-2.128	-0.014	1.029**	-0.134
	(1.486)	(1.758)	(0.085)	(0.403)	(0.186)
L.Voluntary Capital Ratio	-1.057*	-0.070	-0.026	0.790	-0.247*
	(0.595)	(0.586)	(0.061)	(0.483)	(0.142)
L.Retail Ratio	0.439	0.312	-0.027	0.441	0.009
	(0.278)	(0.325)	(0.054)	(0.544)	(0.058)
L.Liquidity Ratio	0.525	1.317**	0.035	-0.541	-0.178**
	(0.458)	(0.544)	(0.086)	(0.517)	(0.078)
L.Loan Loss Provisions Ratio	-0.201	-0.188	-0.060	-0.468	0.035
	(0.283)	(0.189)	(0.085)	(0.886)	(0.109)
L.Cost-Income Ratio	0.006	0.003	0.004	-0.023	0.001
	(0.008)	(0.005)	(0.005)	(0.065)	(0.003)
L.Return on Equity	0.736	-0.245	0.198	0.898	0.166
	(0.712)	(0.573)	(0.187)	(0.684)	(0.205)
L.Interest Income Ratio	-0.008	-0.004	-0.003	0.064	-0.001
	(0.010)	(0.006)	(0.008)	(0.068)	(0.004)
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	262	162	924
R2	0.993	0.992	0.986	0.964	0.944
Tested Banks	63	63	35	22	63
Non-tested Banks	69	69	4	2	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 (1) 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 (2) the dependent variable is *Log(Total Exposure)* defined as the logarithm of total credit risk exposures, i.e. the denominator of *RWD*. In Columns (3) to (5) RWD is decomposed according to the regulatory approach used to report credit risk exposures. In Column (3) only exposures reported under the Advanced Internal Ratings-Based Approach (A-IRB) are included. In Column (4) only exposures reported under the Foundation Internal Ratings-Based Approach (F-IRB) are included. In Column (5) only exposures reported under the Standardized Approach (SA) are included. Not all banks use IRB approaches which reduces observations, especially in the control group.

Table A4: Alternative measures of risk.

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

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 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: Reverse causality.

	(1)	(2)	(3)	(4)	(5)	(6)
	QA intensity		QA effectiveness		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 A7: Robustness with continuous QA measures.

	(1) QA Intensity	(2) QA Effectiveness	(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 Intensity 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 Effectiveness 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 A8: 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 A9: Definitions of stress test intensity variables.

Variable	Definition
Scrutiny channel	
QA Intensity	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 Intensity	Dummy equal to 1 for banks with *QA Intensity* above the median.
QA Effectiveness	Sum of the accumulated potential impact of all flags (as in *QA Intensity*) 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 Effectiveness	Dummy equal to 1 for banks with *QA Effectiveness* 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.