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

Christoffer Kok[a] Carola Müller[b c] Cosimo Pancaro[a]

[a]European Central Bank [b]Norges Bank [c]Halle Institute for Economic Research (IWH)

2020 ECB conference on
Macroprudential Stress-Testing

05 Feb 2020

Stress test assessments and risk-taking behavior

- Regulatory interventions that enhance information disclosure and monitoring can be more effective than capital requirements [Barth et al.,2003]
- Stress-test-informed capital requirements can deter risk taking
- Market discipline and increased transparency can decrease risk-taking incentives [Petrella and Resti, 2013]
- Risk taking can be reduced through more interaction between supervisors and supervised bank or increased regulatory scrutiny [Pierret and Steri, 2018; Hirtle et al., 2018]

(1) Does stress testing affect banks' risk taking?

(2) If so, how effective is supervisory scrutiny associated with stress tests?

Contribution

- Evidence from the U.S. points to lower risk-taking of stress-tested banks [Acharya et al., 2018; Cortés et al., 2017; Calem et al., 2017]

$\Rightarrow$ We evaluate the effect on risk taking in Europe

- Supervisory scrutiny can effectively reduce risk taking [Hirtle et al., 2019]
- Evidence points to a role of supervisory scrutiny in context of U.S. stress tests [Pierret and Steri, 2018]

$\Rightarrow$ We apply a novel measure of supervisory scrutiny and provide evidence that the intensity of supervisory scrutiny in the stress test affects risk taking

Stress test Quality Assurance in the Constrained Bottom-Up approach

The EU-wide stress test follows a Constrained Bottom-Up approach:

- Banks use their own models to generate stress test projections
- Based on a common methodology and a predefined macro-financial scenario
- Supervisory authorities mainly use two complementary challenger approaches (Peer Benchmark and Top-Down) to ensure the credibility of banks' projections

The latter task is carried out by the ECB during the Quality Assurance process:

- Involving cooperation of different stakeholders: ECB, NCAs and banks
- Engaging a material amount of resources
- Lasting about 4 months

(1) Does stress testing affect banks' risk-taking?

(1) Does stress test participation change bank behavior?

$$Risk_{i,t} = \beta_1(Post\ ST16_t \times Tested_i) + \alpha_i + \beta_2 X_{i,t-1} + \gamma_t + \delta_{t,j} + \epsilon_{i,t}$$

i = bank; t = quarter; j = Country of bank HQs

Differences-in-differences setting:

- Difference in treatment: Participation in EU-wide stress test 2016
  - *Tested*= 1 stress-tested banks
  - *Tested*= 0 non-tested banks
- Difference between before and after the stress test 2016
  - *Post ST16*= 1 four quarters after the test (2017)
  - *Post ST16*= 0 four quarters before the test (2015)
- β_1 captures the DID effect of stress-testing on risk taking
- Risk is measured as risk weight density (RWD) for credit risk exposures

$$RWD_{i,t} = \frac{Risk\text{-}Weighted\ Credit\ Risk\ Exposures_{i,t}}{Total\ Credit\ Risk\ Exposure_{i,t}}$$

Timeline

(1) Does stress test participation change bank behavior? (cont.)

$$Risk_{i,t} = \beta_1(Post\ ST16_t \times Tested_i) + \alpha_i + \beta_2 X_{i,t-1} + \gamma_t + \delta_{t,j} + \epsilon_{i,t}$$

Challenges:

- Treatment is not assigned randomly but selected on observables (systemical importantance), banks in control and treatment group might be too different
  - ⇒ We control for SI-status with *Log(Assets)* and observable time-varying differences ($X_{i,t}$)
  - ⇒ We control for all time-invariant differences (α_i)
  - ⇒ We provide indicative evidence of common trends in pre-treatment outcome (*Risk*)
  - ⇒ We provide statistics on the balancing
  - ⇒ We provide a line of robustness: enforce common support, use matching estimators
- Differences in credit risk might be demand-driven
  - ⇒ We estimate within time (γ_t), within home-country-time ($\delta_{t,j}$)
- Endogeneity concerns: Risk might determine control variables
  - ⇒ We lag control variables by one quarter ($X_{i,t-1}$)

Data and Variables

Panel data: quarterly bank-level information (2014Q4 - 2018Q3)

Dependent variable: risk-weight density (RWD)

- Supervisory data (COREP) on credit risk exposures

Control variables: Capital requirements, CAMEL risk measures, business model

- Supervisory data (COREP) on capital ratios and minimum capital requirements
  - *Regulatory Capital*
  - *Voluntary Capital*
- Supervisory data (FINREP) on balance sheet and P& L
  - *Loan Loss Provisions Ratio, CIR, RoE, Liquidity Ratio*
  - *Retail Ratio, Interest Income Ratio*

Treatment and control group - descriptives

Table 1: Summary statistics of covariates by treatment.

	(1)	(2)	(3)	(4)	(5)	(6)
	Non-tested		Stress-tested		Control - Treatment	
	Mean	SD	Mean	SD	Diff. in Means	Normalized Difference†
Log(Total Assets)	22.177	1.334	25.370	1.510	-3.193***	-1.585
Regulatory Capital	0.084	0.020	0.081	0.037	0.003	0.071
CAMEL						
Voluntary Capital	0.090	0.066	0.088	0.046	0.002	0.020
Loan Loss Provisions Ratio	0.019	0.106	0.001	0.015	0.019**	0.173
Cost-Income-Ratio	0.787	2.126	0.653	0.708	0.134	0.060
Return on Equity	0.020	0.036	0.020	0.020	0.000	0.011
Liquidity Ratio	0.118	0.147	0.054	0.060	0.064***	0.404
Business Model						
Retail Ratio	1.239	0.264	1.178	0.233	0.062**	0.176
Interest Income Ratio	0.688	1.589	0.725	1.359	-0.037	-0.018

Significance of two-sided T-test: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. † According to Wooldridge and Imbens (2009) normalized differences are calculated as $\frac{\mu_{ntr} - \mu_{tr}}{\sqrt{\sigma^2_{ntr} + \sigma^2_{tr}}}$.

- Banks in the control group are significantly smaller
- They have higher liquidity ratios, LLPs, and depend more on retail business.

Treatment and control group - parallel trends assumption

Table 2: Summary statistics of the dependent variable RWD by treatment and test for significant differences in pre-period trends.

	(1)	(2)	(3)	(4)	(5)	(6)
		Levels			*First Differences*	
	Non-tested	Tested	Diff.	Non-tested	Tested	Diff.
Pre	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	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.	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 deviation in parentheses. Columns (3) and (6) show the difference in means between groups and p-value of t-test in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

- Negative significant unconditional diff-in-diff effect
- No significant differences in the slopes of risk measure in the pre-period

Table 3: Effect of participating in the stress test on bank risk

Dependent: RWD *Treatment:* Participation	(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			-0.130	-0.150
			(0.214)	(0.191)
L.Voluntary Capital			-0.241*	-0.254*
			(0.125)	(0.144)
L.Retail			-0.016	0.013
			(0.050)	(0.059)
L.Liquidity			-0.208**	-0.175**
			(0.085)	(0.078)
L.LLP			0.066	0.039
			(0.073)	(0.105)
L.CIR			0.001	0.001
			(0.003)	(0.003)
L.RoE			0.218	0.166
			(0.195)	(0.207)
L.Interest Income			-0.002	-0.001
			(0.004)	(0.004)
Bank, Time FE	Yes	Yes	Yes	Yes
Country x Time FE	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: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Stress testing affects risk-taking

Main result

- Tested banks decrease RWD by 4.2 percentage points relative to non-tested banks
- Economic significance: Change of about 20 percent of the standard deviation of RWD of tested banks

Results are robust to

- ✓ using only exposures from Standardized Approach, not IRB (Behn et al., 2016)
- ✓ using only shifts into high-risk buckets (RW>100)
- ✓ using distance-to-default measures for risk (EDFs,z-scores)
- ✓ using different time windows around the stress test
- ✓ using contemporaneous bank-level covariates
- ✓ using different cluster-levels or robust errors
- ✓ eliminating serial correlation bias by collapsing time dimension (Bertrand et al., 2004)

RWD definitions | RWD alternatives | Time windows | Collapse

Different size distributions of control and treatment group

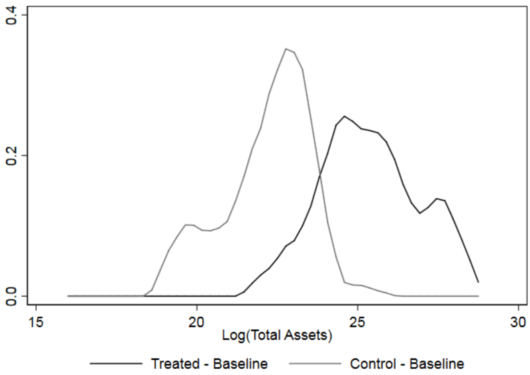


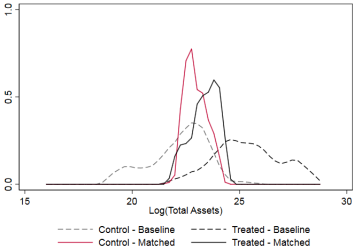
Figure 1: Distribution of Log(Total Assets) by treatment.

Robustness to differences in size

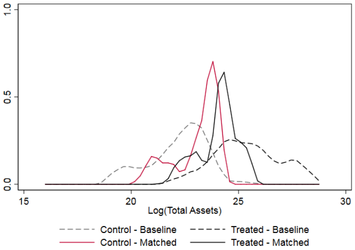
Results are robust to

- ✓ shrinking the sample gradually by excluding the smallest non-tested and the largest tested banks
- ✓ using Abadie and Imbens (2011) bias-adjusted matching estimator
  - on a reduced sample with common support
  - on a reduced sample with up to 2 matches within country

Common support



Within country (2 smallest, 2 largest)



Matching

(2) What role has supervisory scrutiny in stress testing?

(2) Testing the QA scrutiny channel

Differences-in-differences setting testing the scrutiny channel:

$$RWD_{i,t} = \alpha_i + \beta_1(Post\ ST16_t \times Tested_i) + \beta_2(Post\ ST16_t \times Tested_i \times QA_i^{dim}) + \beta_3 X_{i,t-1}\gamma_t + \delta_{t,j} + \epsilon_{i,t}$$

- Intensity of treatment: amount of scrutiny during QA
- Implementation of stress test intensity measure:
  - as a continuous measure (QA_i^{dim})
  - as a discrete measure ($High\ QA_i^{dim} = 1$ for above median intensity of treatment)

Challenges:

- Endogeneity concerns: Risk might determine intensity of scrutiny

⇒ We test if intensity of scrutiny depends on pre-stress test levels of risk

Selection in QA

Intensity of scrutiny measures

3 measures of scrutiny intensity based on data about flags raised during the QA

- *QA Intensity*: Log-Number of communicated flags on credit risk (continuous)
- *QA Effectiveness*: Sum of potential impact on CET1 of flags (continuous)
- *QA Duration*: Number of cycles in which flags were communicated (ordinal)

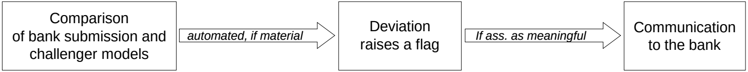


Figure 2: Simplified cycle of the QA process.

Table 6: Effect of supervisory scrutiny on risk-taking

	(1)	(2)	(3)	(4)	(5)	(6)
	QA Intensity		QA Effectiveness		QA Duration	
Post ST16 x Tested	0.012	-0.014	-0.031*	-0.031*	0.011	-0.008
	(0.026)	(0.016)	(0.016)	(0.016)	(0.031)	(0.024)
Post ST16 x Tested x QA	-0.027*		-0.333		-0.025*	
	(0.014)		(0.268)		(0.014)	
Post ST16 x Tested x High QA		-0.056***		-0.023		-0.041*
		(0.020)		(0.024)		(0.022)
Lagged Bank Controls	Yes	Yes	Yes	Yes	Yes	Yes
Bank, Time, Country×Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	924	924	924	924	924	924
within R2	0.141	0.155	0.133	0.126	0.132	0.129

Notes: Clustered standard errors at the bank-level in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Bank-level control variables are lagged by one quarter and comprise *Log(Assets), Voluntary Capital, Regulatory Capital, Liquidity, Retail, LLP, CIR, RoE*, and *Interest Income*.

- More interactions with the supervisor and interactions over a longer period seem to induce a stronger disciplining effect
- Potential impact of each discussed item on average insignificant

QA scrutiny intensity effect

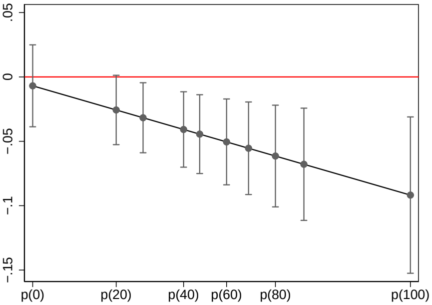
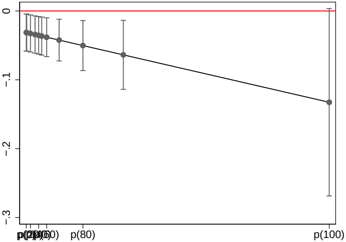


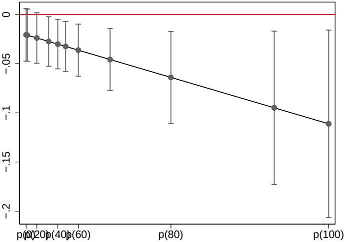
Figure 3: Marginal Effects along percentiles of QA intensity.

Notes: The figure shows marginal effects and 95% confidence interval of being stress-tested while receiving a defined amount of Quality Assurance along the distribution of QA intensity. Marginal effects are calculated for the minimum, the 10th, 20th, 30th, etc. percentile to the maximum.

QA scrutiny effectiveness effect



(a) Unwinsorized (baseline).



(b) Winsorized at 5%.

Figure 4: Marginal Effects along percentiles of QA effectiveness measures.

Notes: The figure shows marginal effects and 95% confidence interval of being stress-tested while receiving a defined amount of Quality Assurance along the distribution of QA Effectiveness. Marginal effects are calculated for the minimum, the 10th, 20th, 30th, etc. percentile to the maximum.

- QA scrutiny effectiveness effect is obfuscated by outlier

Supervisory scrutiny affects risk-taking

Main result

- Stress-tested banks reduce RWD more ...
  - the more flags were communicated to them regarding credit risk projections (intensity)
  - the longer the process of QA lasted for them (duration)

Alternative channels

× Market discipline: Insignificant interaction effect for tested banks whose results were published (EBA banks)

× Capital guidance: Insignificant interaction effect for tested banks that received a higher Pillar 2 capital requirement

× Poor capitalization: Insignificant interaction effect for tested banks that started with poor capitalization into the exercise

⏩ Alternative channels

Conclusions

Main findings

- We find that the 2016 European Stress Test exercise impacted banks behavior
- We find that banks that participated in stress testing reduce their RWD relative to banks that did not participate by 4.2 percentage points, i.e. about 20 percent of one SD of RWD
- Reduction in risk-taking is related to the intensity of supervisory scrutiny applied during QA.
- Banks that had more interactions with the supervisor and over a longer period reduced their RWD more than banks that received less treatment.

Policy implications

- Quality Assurance process may induce banks to adjust their risk-taking behavior
- Scrutiny applied in the constrained bottom-up approach has a disciplining effect on bank credit risk

Appendix

Timeline and sample

Descriptives

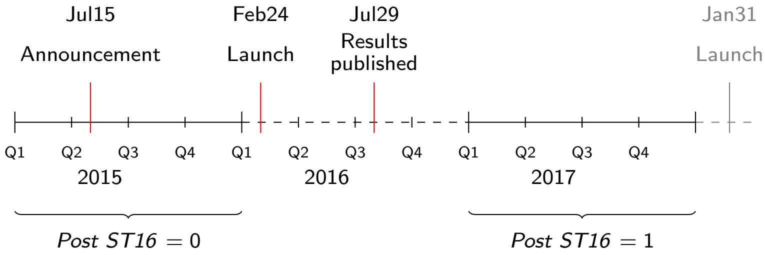


Figure 5: Timeline EU-wide stress test 2016.

Strongly balanced sample:

- *Tested* = 1: 63 SIs (31 EBA, 32 SREP)
- *Tested* = 0: 69 LSIs

Back

Alternative definitions of the dependent variable

Robustness

Table 4: Alternative definitions of RWD.

	(1)	(2)	(3)	(4)
	RWD	RWD only SA	Share (RW>100)	Ratio ($\frac{RW>100}{RW<50}$)
Post ST16 x Tested	-0.042** (0.019)	-0.042** (0.018)	-0.029** (0.014)	-0.091** (0.039)
Lagged Bank Controls	Yes	Yes	Yes	Yes
Bank, Time, Country×Time FE	Yes	Yes	Yes	Yes
Observations	924	924	924	924
within R2	0.120	0.120	0.106	0.109

Notes: Clustered standard errors at the bank-level in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

- Results stem from credit risk exposures under the Standardized Approach (SA), not Internal Ratings Based Approach (IRB).

Back

Alternative measures for risk-taking

Robustness

Table 5: Alternative measures of risk-taking.

Dependent: *(Specification:)*	(1) Debt/Assets (baseline)	(2) NPL/Loans (baseline)	(3) Moody's EDF (baseline)	(4) SNF z-score (time collapsed)
Post ST16 x Treated	0.001	-0.012	-1.275*	0.674**
	(0.002)	(0.010)	(0.640)	(0.270)
lagged bank controls	yes	yes	yes	yes
Bank, Qtr, Country×Qtr FE	yes	yes	yes	yes
Observations	924	918	299	212
R2	0.98	0.92	0.92	0.96
No. Stress-tested	63	63	32	51
Mean	0.925	0.12	1.347	1.857
(SD)	(0.041)	(0.164)	(2.549)	(2.136)
No. Non-tested	69	69	14	55
Mean	0.893	0.139	1.551	2.382
(SD)	(0.07)	(0.177)	(3.511)	(3.013)

Notes: Clustered standard errors at the bank-level in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. In (1) *Debt/Assets* is the ratio of total debt to total assets. In (2) *NPL/Loans* is defined as Non-performing loans over total loans where NPLs are all loans reported as past due over 30 days. In (3) Expected Default Frequencies (EDFs) are provided by Moody's Analytics which measure the probability of default within the next year. In (4) we estimate with a collapsed time dimension relying on yearly data from SNL Financials. *z-score* is 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.

- Results are robust to using distance-to-default measures for a subsample.

Back

Alternative time windows, collapse time dimension

Robustness

Table 6: Robustness of the baseline effect to different time spans around the stress test and a elimination of the time dimension in the data.

	(1)	(2)	(3)	(4)
	Changing the pre- and post-period			Collapse
Pre	15Q1-15Q4	15Q1-16Q1	15Q1-16Q1	No Time
Post	17Q1-17Q4	16Q4-17Q4	16Q4-18Q3	Dimension
Post ST16 x Treated	-0.042**	-0.035**	-0.033**	-0.047**
	(0.019)	(0.015)	(0.014)	(0.021)
Lagged Bank Controls	Yes	Yes	Yes	Yes
Bank, Time, Country x Time FE	Yes	Yes	Yes	Yes
Observations	924	1,188	1,318	264
within R2	0.12	0.097	0.105	0.196

Notes: Clustered standard errors at the bank-level in parentheses in Columns (1) to (3), robust standard errors in parentheses in Column (4): *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

- Results hold when using different time windows around the 2016 stress test
- Results hold when the time dimension and serial correlation is eliminated (Bertrand et al., 2004)

Back

Excluding the smallest not-tested and the biggest tested banks

Robustness

Table 7: Robustness of the interaction effect with gradually decreasing sample size.

				(1) All Control	(2) excl. Bottom p(25)	(3) excl. p(50)	(4) excl. Bottom p(75)
			N	69	51	34	17
Post ST16 × Tested	(1)	All Tested	63	-0.042**	-0.047**	-0.051**	-0.051**
				(0.019)	(0.020)	(0.023)	(0.021)
Obs.				924	791	665	539
Post ST16 × Tested	(2)	excl. Top p(25)	47	-0.041**	-0.046**	-0.050**	-0.055**
				(0.020)	(0.022)	(0.024)	(0.023)
Obs.				812	679	553	420
Post ST16 × Tested	(3)	excl. p(50)	31	-0.030*	-0.032*	-0.030*	-0.047**
				(0.017)	(0.017)	(0.017)	(0.020)
Obs.				700	567	441	315
Post ST16 × Tested	(4)	excl. Top p(75)	15	-0.020	-0.031	-0.028	-0.022
				(0.024)	(0.028)	(0.028)	(0.025)
Obs.				581	448	322	203

Notes: Clustered standard errors at the bank-level in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Back

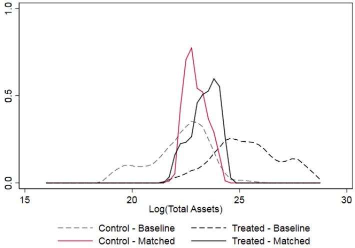
Matching (1)

Robustness

- If selection into treatment is based on observables, matching estimators can be used for causal inference (Rosenbaum and Rubin, 1983)
- We use two matching strategies as in Gropp et al. (2018):

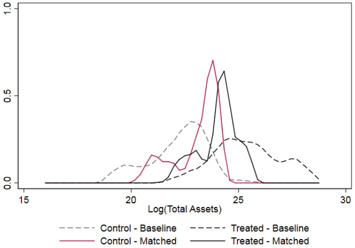
Common support sample:

- Excluding control $<$ Min[tested]
- Excluding tested $>$ Max[control]



Within country sample: For each country

- Includes only 2 biggest control
- Includes only 2 smallest tested



Matching (2)

Robustness

Table 8: Robustness of the baseline effect to matching estimation strategies.

	(1) Common Support Sample	(2) Within Country Sample
Average Treatment Effect on the Treated	-0.079*** (0.007)	-0.012* (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 according to Abadie and Imbens (2011) in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

- Results are robust to using Bias-adjusted matching estimator (Abadie and Imbens, 2011)

Back

Selection into QA

Robustness

Do supervisors pay more attention to riskier banks?

- Processes are organized to ensure impartiality
- Challenger models and methodology are determined before QA
- Implementation of model review (raising of flags) is automated
- Conditional on observables, before the ST we cannot find that risk affects QA

Table 9: The effect of risk-taking on the intensity of treatment.

	(1)	(2)	(3)	(4)	(5)	(6)
	QA intensity		QA effectiveness		QA duration	
RWD	0.132	0.233	0.013	0.018	-0.574	-0.427
	(0.474)	(0.435)	(0.032)	(0.024)	(1.710)	(1.233)
Bank Controls	yes	no	yes	no	yes	no
Constant	1.540	1.966***	-0.017	0.027**		
	(2.732)	(0.231)	(0.134)	(0.013)		
Observations	63	63	63	63	63	63
R-squared	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 (4) are OLS estimations. Columns (5) and (6) estimate an ordered logit model.

Back

Further channels (1)

Market discipline

- Stress tests generate new information for investors and a reduction of information asymmetries gives investors more power to discipline banks [Petrella and Resti, 2013; Georgescu et al., 2017]

⇒ *EBA banks* indicates banks whose results were published by EBA

Capital requirements

- Higher capital requirements resulting from the stress test should deter risk-taking [Acharya et al., 2018; Pierret and Steri, 2018]
- Contrary to the U.S., EU-wide stress tests "*only*" inform Pillar 2 capital guidance

⇒ *High P2G* indicates banks that received above median Pillar 2 guidance in 2017Q1 when P2G was informed by stress test results under the adverse scenario

Capitalization

- Not regulatory requirements but banks' ability to comply with them is relevant

⇒ *Low Voluntary Capital* indicates banks with below median voluntary capital buffers when entering into the stress test

Further channels (2)

Results

Table 7: Effect of being stress tested on risk-taking through alternative channels.

	(1) Market Discipline	(2) Capital Guidance	(3) Capitalization
Post ST16 x Treated	-0.031*	-0.049**	-0.048**
	(0.018)	(0.021)	(0.024)
Post ST16 x Treated x EBA Banks	-0.026		
	(0.029)		
Post ST16 x High P2G		0.026	
		(0.033)	
Post ST16 x Treated x High P2G		0.003	
		(0.035)	
Post ST16 x Low Voluntary Capital			-0.037*
			(0.022)
Post ST16 x Treated x Low Voluntary Capital			0.016
			(0.030)
Lagged Bank Controls	Yes	Yes	(Yes)
Bank, Time, 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: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

- We cannot find evidence that these channels can explain our baseline result

Back