Gaming the Test? Window-dressing and portfolio similarity around the EU-wide stress tests

C. Barbieri, **A. Cuzzola**[1], G. Halaj

SSM - DG-HOL - Stress Test Expert Division

June 26, 2025

DISCLAIMER: the current presentation and article do not reflect in any way the opinion or an official statement of the institutions the authors are affiliated to.

[1]Institute of Economics, Sant'Anna School of Advanced Studies

Background: Stress Tests and their impact

Post-2008 Paradigm Shift: Stress testing emerged as a critical tool in banking supervision following the Global Financial Crisis (➔Basel III and prudential supervision)

Stress Tests Multi-layered Impact on Financial System:

- Direct impact through capital buffer requirements
- Semi-direct market signalling effect
- Indirect effects: e.g. portfolio rebalancing

➢ **Strategy or Risk Management?** Banks optimizing their balance sheets specifically for test scenarios rather than for genuine risk management

- **Window Dressing:** Temporary portfolio adjustments around stress tests to improve reported metrics (Cornett et al., 2020)
- **Beauty Contest Problem:** Focus on outperforming peers rather than identifying actual risks (Quagliariello, 2019)
- **Portfolio Homogenization:** Common scenarios may lead to similar de-risking strategies across institutions (Rhee and Dogra, 2024; Bräuning and Fillat, 2025)

Research Questions

Against this backdrop, our research revolves arount two key questions

- How do EU-wide stress tests influence banks' de-risking behavior around the stress test (not just after)?
  - More precisely: do banks engage in "window dressing" ahead of stress tests? Does this behaviour differ from *regular* year-end dressing?
- What impact do ST-induced banks' behaviours have on systemic risk through portfolio similarity?
  - Do stress tests lead to increased or decreased portfolio similarity?

Relevant literature

- Macropudential policy implementation creates incentives to window-dressing (Garcia et al., 2023; Bassi et al., 2024).
- Impact of stress tests and through which channels:
  - Capital requirements calibrated via STs. Relevant for the US (Acharya et al., 2018), but less for the European exercise (Kok et al., 2023; Konietschke et al., 2022).
  - ST supervisory scrutiny has a disciplining effect (Kok et al., 2023).
  - Market discipline through the disclosure of stress test results provides (Georgescu et al., 2017; Fernandes et al., 2020; Konietschke et al., 2022; Durrani et al., 2024).
- Stress testing implementation led to **increased portfolio similarity** (Bräuning and Fillat, 2025). However, these effects can vary depending on economic conditions and the specific supervisory framework in place (Curry et al., 2008; Kupiec et al., 2017).

Our contribution: key findings

Key Findings:

1. CET1 Ratio: We find **evidence of "dressing" before stress tests**, especially for banks that ultimately score low in the exercises.
2. **Decreased Portfolio Similarity around the stress test**: a first decrease in the "dressing" stage and a further decrease in the supervision phase (2% overall) in portfolio similarity due to the stress tests

Results are confirmed for a set of similarity indicators (including liquidity-weighted similarity) and robustness checks

Data and methodology

Data sources:

- FINREP: main asset categories and balance sheet information
- AnaCredit: loan-level granular information
- Stress Test data: main variable Final Capital Depletion as ST score

Method:

- Difference-in-differences with conditional parallel trends: **De Chaisemartin and d'Haultfoeuille (2024)** for the average treatment effects on the banks subject to the stress test (ATT), aggregated to obtain event study plots around the stress test quarters:
  - Allows relaxing the "parallel trends assumption" after conditioning on observed covariates
  - controls for jurisdiction specific trends in the pretreatment period
- Treatment: 79 banks in 2021 stress test, 95 banks in 2023 stress test
- Control group: selected from LSI using one-to-one matching without replacement using scaled Euclidean distance
- Matching observations by size, business model, and risk profile criteria
- DiD with conditional parallel trends accounts for remaining differences

Leveraging data from two exercises

Parallel ST deployment and stacked estimation

The parallel structure of ST deployment allows us to exploit both exercises using a stacked dataset

Methodology & Templates	Exercise Launch & Scenarios		Results Publication	
2020 Q4 (13.11.20)	2021 Q1 (29.01.21)	2021 Q2	2021 Q3 (30.07.21)	2021 Q4

Methodology & Templates	Exercise Launch & Scenarios		Results Publication	
2022 Q4 (04.11.22)	2023 Q1 (31.01.23)	2023 Q2	2023 Q3 (28.07.23)	2023 Q4

Banks Portfolio Similarity: Two-Step Methodology

Step 1: Computing Pairwise Similarity Between Banks

- Cosine similarity of portfolio shares

$$cs^{D}_{i,j,t} = \sum_{k=1}^{K} \frac{a_{k,i,t} \cdot a_{k,j,t}}{\sqrt{\sum_{l} a^2_{l,i,t}} \sqrt{\sum_{l'} a^2_{l',j,t}}}$$

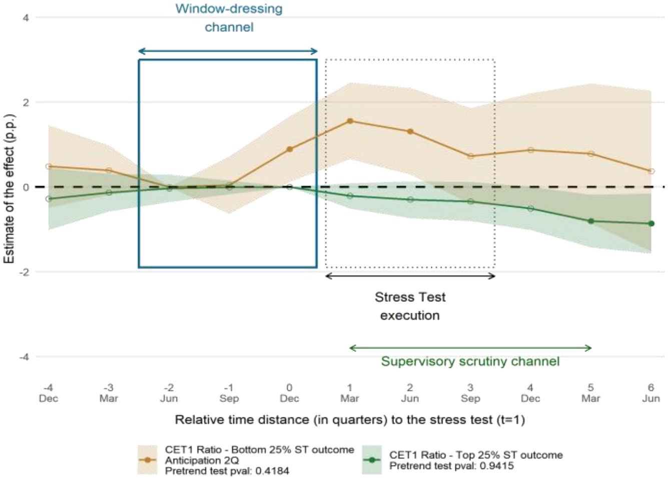
- Liquidity-weighted similarity: cosine similarity weighted by ω_k liquidity weights
- Multiple dimensions of similarity: by instrument type, country of exposure; for loans: by sectors (NACE2), credit quality, maturity buckets

Step 2: Aggregating to Bank-System Similarity Measures

- Simple average similarity: equally weighted nodes
- Size-weighted similarity: accounts for counterparty importance

Key finding 1: Evidence of Window-Dressing

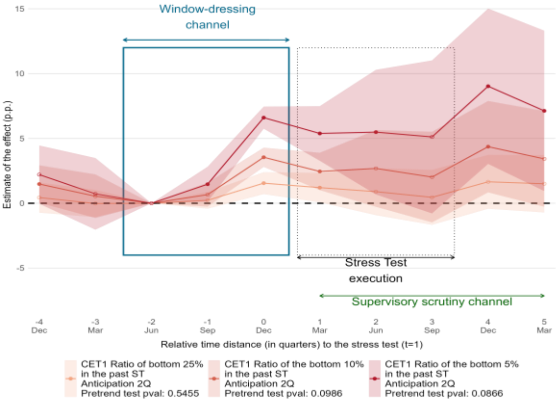
Evidence of Capital Ratio Management up to 2 quarters before the EBA ST



Key finding 1: Window-Dressing of "low-scorer"

Bottom quartile banks increase CET1 by 2 percentage points pre-test, while top quartile banks see no significant anticipatory effects:

- Banks aware of vulnerability to specific stress scenarios adopt strategic behavior to maintain adequate buffers under stress
- Persistence in performance rankings across stress test cycles: 40-50% of bottom quartile banks remain in bottom quartile 50-64% of top quartile banks remain in top quartile

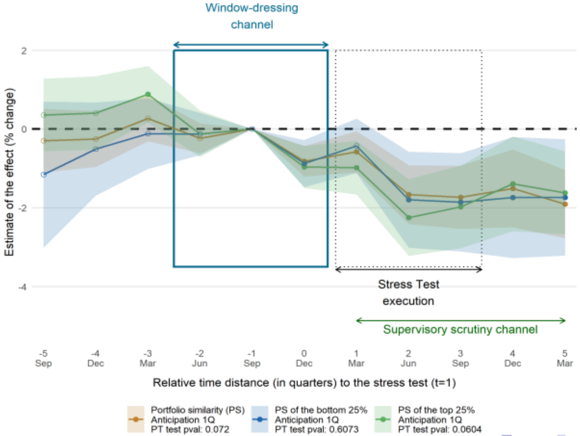


Key finding 2: Decreasing Portfolio Similarity

We measure the ATE of the stress test on the bank similarity with respect to the other banks in the system.

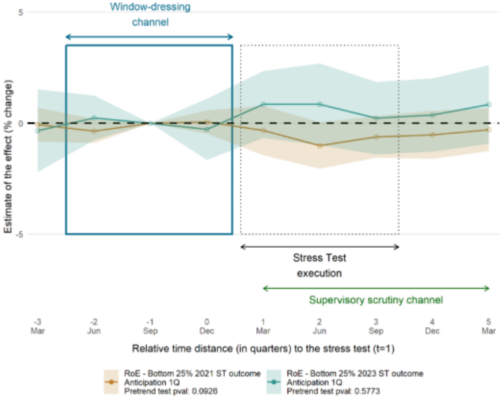
- Window-dressing period: ∼1% decrease in similarity
- Execution/post-test: More substantial ∼2% persistent reduction

Heterogeneous effects by performance



Robustness checks

- Consistent decrease across asset categorizations: decrease in similarity for multiple dimensions of asset similarity (robustness to step 1 similarity computation)
- Strongest and most consistent evidence of decreased similarity appears within business model groups, not by country (advantage of testing different aggregations of similarity as step 2)
- Result confirmed by Liquidity-Weighted Portfolio Similarity and Loan Portfolio Similarity (robustness to alternative measures of similarity)
- Window-dressing around EBA ST is significant beyond routine year-end reporting practices



Conclusion

Key Finding 1: Ex-Ante window-dressing drives de-risking

- Significant de-risking occurs before stress tests, not after
- Primary mechanism: 5% reduction in risk-weighted assets
- Effect strongest for poor-performing banks: CET1 increases up to 2pp
- Strategic adjustments to optimize ST starting point data
- Evidence robust to routine end-of-year accounting adjustments

Key Finding 2: Decreased portfolio similarity reduces systemic risk

- No convergence toward common portfolio structures
- $\sim$1% decrease during window-dressing, $\sim$2% during/after tests Banks pursue idiosyncratic de-risking strategies
- Effect holds across different asset categorizations and across country and business model clusters
- Bottom performers drive the strongest decrease in similarity

Policy Implications

Positive Financial Stability Effects:

- Current stress test design effectively minimizes negative externalities
- Tests contribute to financial stability by reducing systemic risk
- No evidence that supervisory follow-up creates conditions for convergence
- Limited risk of synchronized reactions to common shocks
- Lower potential for amplification through coordinated adjustments or fire sales

Monitoring and refinement needs:

- Banks with weaker fundamentals require closer monitoring, also in light of the fact that these institutions anticipate more strongly stress tests
- Strategic window-dressing may undermine the informativeness of test results

Gaming the Test? Window-dressing and portfolio similarity around the EU-wide stress tests

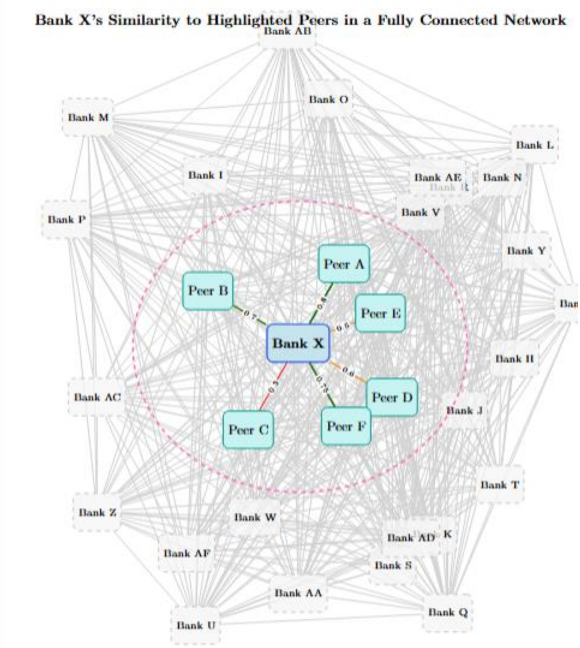
ANNEX

Sample Characteristics: Before vs. After Matching

Key observations:

- **Improved homogeneity after matching:**
  - Capital ratios: Nearly identical (CET1, Tier 1)
  - Return metrics: Reduced differences in RoE, RoA
  - Leverage ratio: Gap eliminated (503-484 vs. 506-516)
  - Loans-to-assets ratio: Well balanced (74-77% in both groups)
- **Persistent differences after matching:**
  - Bank size: Treated banks remain significantly larger (TA log: 25.4 vs. 22.5)
  - Risk density: Lower in ST banks (35% vs. 44-46%)
  - Deposit funding: Lower reliance in ST banks (74% vs. 84-87%)

- Standardized mean differences (SMD) show improved balance after matching → Irreducible difference because of selection into treatment (by policy design)
- DiD with conditional parallel trends accounts for remaining differences



Bank X's Similarity to Highlighted Peers in a Fully Connected Network
Bank AB
Bank O
Bank M
Bank L
Bank I
Bank AE
Bank N
Bank P
Bank V
Bank Y
Peer A
Peer B
Peer E
Bank X
Bank H
Peer D
Bank AC
Bank J
Peer C
Peer F
Bank T
Bank Z
Bank W
Bank AD
Bank AF
Bank S
Bank AA
Bank Q
Bank U

Key Findings: Impact of EU-wide Stress Tests

Stress test-induced derisking via window-dressing

- Major risk adjustments occur *before* tests (window-dressing), not after
- Banks with poorer projected performance engage more actively in ex-ante derisking

Decline in portfolio similarity reduces systemic risk

- Risk management actions are idiosyncratic with no evidence of herding
- Effect strongest among banks with higher capital depletion under stress scenarios
- Asynchronous adjustments by vulnerable banks reduce potential for amplifying shocks

No evidence of local similarity increases and robustness

- Similarity does not increase within country clusters or business model groups
- Results robust to geographical, legal, historical, and business model connections

Evidence of Window-Dressing: Capital Ratio Management

Banks strategically boost capital ratios before tests

- CET1 ratio increases by ~20 basis points in quarter before stress test
- Standard model (no anticipation) shows violation of parallel trends
- Anticipation model reveals significant pre-test adjustments
- Primary adjustment through reduction in risk-weighted assets (~5%)
- Bottom quartile banks: Increase CET1 by 2 percentage points pre-test
- Top quartile banks: No significant anticipatory effects

Explaining the pattern:

- Banks aware of vulnerability to specific stress scenarios
- Strategic behavior to maintain adequate buffers under stress
- Persistence in performance rankings across stress test cycles:
    - 40-50% of bottom quartile banks remain in bottom quartile
    - 50-64% of top quartile banks remain in top quartile

Additional findings

Consistency across stress test cycles:

- Similar anticipatory patterns in both 2021 and 2023 exercises
- 2021: Larger and more persistent effects
- 2023: Higher anticipatory behavior in quarter immediately before test
- Both show inverse U-shaped pattern (effects dissipate during test)

Balance sheet trade-offs:

- Liquidity Coverage Ratio decreases by ~30% during window-dressing
- Strategic reallocation toward capital optimization
- No significant impact on profitability (ROE) despite adjustments

Post-publication effects:

- Poor performers: 0.55% decrease in CET1 per 1pp additional depletion
- Top performers: 1.03% increase in ROE per 1pp better performance
- Confirms market discipline as secondary channel of influence

Portfolio Similarity Around Stress Tests

Competing Hypotheses:

- **"Beauty Contest" Hypothesis:** Banks converge toward similar portfolios
  - Common evaluation framework leads to synchronized de-risking
  - Increased portfolio similarity as banks minimize vulnerability to scenarios
- **"Idiosyncratic Adjustment" Hypothesis:** Banks pursue unique strategies
  - De-risking reflects specific constraints, business models, and market conditions → Decreased similarity coupled with individual risk reduction

Multi-dimensional Analysis Approach:

- **Overall portfolio similarity** - Complete picture of balance sheet overlap
- **Securities portfolio** - Liquidity-weighted similarity of tradable assets
- **Loan portfolio** - Granular analysis of sectoral exposures via AnaCredit

Portfolio Similarity Around Stress Tests: Key Findings

Main Finding: Stress tests lead to decreased portfolio similarity

- Two distinct phases of adjustment:
  - Window-dressing period: ∼1% decrease in similarity
  - Execution/post-test: More substantial ∼2% persistent reduction
- Heterogeneous effects by performance:
  - Bottom performers: Strongest decrease in similarity
  - Top performers: More modest decrease
- Consistent result across different asset categorizations
  - Challenges the "beauty contest" hypothesis
  - Banks pursue idiosyncratic de-risking strategies
  - Potentially enhances system stability by reducing synchronized reactions

Robustness 1: Consistent Decrease Across Asset Categorizations

Asset Dimension	Estimate	SE	95% CI	Pre-trend p
Type of Counterparty (Cpt.)	**-0.013****	0.005	[-0.024, -0.003]	0.348
Type of Instrument (Inst.)	**-0.008****	0.003	[-0.015, -0.001]	0.706
Type of Cpt. $\times$ Type of Instr.	**-0.022****	0.006	[-0.034, -0.009]	0.148
Asset Type	**-0.021****	0.006	[-0.033, -0.009]	0.177

Testing different levels of granularity in portfolio categorization reveals consistent decrease in similarity regardless of how assets are classified. The effect is strongest when using more granular categorizations (combined dimensions), indicating that banks differentiate their portfolios across multiple characteristics simultaneously.

Robustness 2: Within-Cluster Similarity Results

Asset Dimension	Estimate	95% CI	Pre-trend p
Within Business Model Similarity			
Type of Counterparty	**-0.013****	[-0.023, -0.004]	0.276
Type of Cpt × Type of Instr	**-0.022****	[-0.033, -0.011]	0.396
Type of Instrument	-0.006	[-0.014, 0.001]	0.978
Asset Class	**-0.021****	[-0.032, -0.009]	0.416
Within Country Similarity			
Type of Counterparty	-0.008	[-0.025, 0.009]	0.552
Type of Cpt × Type of Instr	-0.014	[-0.033, 0.004]	0.664
Type of Instrument	-0.005	[-0.012, 0.003]	0.000
Asset Class	-0.013	[-0.031, 0.005]	0.601

Note: Table shows stress test effects on within-cluster similarity across different asset categorizations. Bold estimates with ** indicate statistical significance at 95% confidence level. Strongest and most consistent evidence of decreased similarity appears within business model groups.

Robustness 3: Additional Similarity Analyses Consistent Pattern

Liquidity-Weighted Portfolio Similarity:

- Focuses on tradable assets subject to fire-sale dynamics
- Larger effects (up to 5% decrease) for 2021 stress test
- Strongest decrease in similarity of instrument types
- Reduces potential for synchronized fire sales during market stress

Loan Portfolio Similarity:

- Based on granular AnaCredit data (sector-country exposures)
- 2021: Significant decrease in similarity (∼4% for bottom performers)
- 2023: Mixed results with temporary increases

Key Insights from Local and Weighted Similarity Analysis

Business Model Clusters:

- Strongest evidence of decreased similarity within similar business models
- Most significant for counterparty and combined categorizations
- Challenges the notion that banks might converge toward peers with similar models

Geographic Clusters:

- No statistically significant country-specific patterns
- Idiosyncratic adjustments appear independent of geographical considerations
- Suggests harmonized impact of stress tests across the EU financial system

Size-Weighted Similarity:

- No evidence of convergence toward larger institutions
- Pattern consistent with overall findings of decreased similarity
- Reinforces conclusion that stress tests promote portfolio differentiation

Bibliography I

Acharya, V. V., A. N. Berger, and R. A. Roman (2018): "Lending implications of U.S. bank stress tests: Costs or benefits?" *Journal of Financial Economics*, 129, 136–157.

Bassi, C., M. Behn, M. Grill, and M. Waibel (2024): "Window dressing of regulatory metrics: evidence from repo markets," *Journal of Financial Intermediation*, 58, 101086.

Bräuning, F. and J. L. Fillat (2025): "The Impact of Regulatory Stress Tests on Banks' Portfolio Similarity and Implications for Systemic Risk," *Journal of Money, Credit and Banking*, https://doi.org/10.1111/jmcb.13239.

Cornett, M. M., K. Minnick, P. J. Schorno, and H. Tehranian (2020): "An examination of bank behavior around Federal Reserve stress tests," *Journal of Financial Intermediation*, 41, 100789.

Curry, T. J., G. S. Fissel, and C. D. Ramirez (2008): "The impact of bank supervision on loan growth," *North American Journal of Economics and Finance*, 19, 192–211.

Bibliography II

De Chaisemartin, C. and X. d'Haultfoeuille (2024): "Difference-in-differences estimators of intertemporal treatment effects," *Review of Economics and Statistics*, 1–45.

Durrani, A., S. Ongena, and A. P. Marques (2024): "Decoding market reactions: The certification role of EU-wide stress tests," *Economic Modelling*, 139, 106828.

Fernandes, M., D. Igan, and M. Pinheiro (2020): "March madness in Wall Street:(What) does the market learn from stress tests?" *Journal of Banking & Finance*, 112, 105250.

Garcia, L., U. Lewrick, and T. Sečnik (2023): "Window Dressing and the Designation of Global Systemically Important Banks," *Journal of Financial Services Research*, 64, 231–264.

Georgescu, O. M., M. Gross, D. Kapp, and C. Kok (2017): "Do stress tests matter? Evidence from the 2014 and 2016 stress tests," .

Kok, C., C. Müller, S. Ongena, and C. Pancaro (2023): "The disciplining effect of supervisory scrutiny in the EU-wide stress test," *Journal of Financial Intermediation*, 53, 100971.

Bibliography III

Konietschke, P., P. Löw, S. Ongena, and O. Ricci (2022): "The disciplining effect of supervisory scrutiny on banks' risk-taking: Evidence from the EU-wide stress test," *ECB Working Paper*.

Kupiec, P., Y. Lee, and C. Rosenfeld (2017): "Does bank supervision impact bank loan growth?" *Journal of Financial Stability*, 28, 29–48.

Quagliariello, M. (2019): "Are stress tests beauty contests?" *EBA Staff Papers*, 4.

Rhee, K. and K. Dogra (2024): "Stress tests and model monoculture," *Journal of Financial Economics*, 152, 103760.

Gaming the Test? Window-dressing and portfolio similarity around the EU-wide stress tests

Angelo Cuzzola[1,2], Claudio Barbieri[2], and Grzegorz Halaj[2]

[1] *Scuola Superiore Sant'Anna*

[2] *European Central Bank*

March 10, 2025

Authorship:

Angelo Cuzzola: European Central Bank & Scuola Superiore Sant'Anna: angelo.cuzzola@santannapisa.it.

Claudio Barbieri: European Central Bank: cluadio.barbieri@ecb.europa.eu.

Grzegorz Halaj: European Central Bank: grzegorz.halaj@ecb.europa.eu.

Disclaimer: This paper should not be reported as representing the views of the European Central Bank (ECB). The opinions expressed are those of the authors and do not necessarily reflect those of the ECB.

Contents

Abstract

This study investigates the impact of supervisory stress testing on banks' behaviors and their systemic risk implications. Utilizing confidential supervisory data from the European Banking Authority's EU-wide stress tests in 2021 and 2023, we employ a difference-in-differences framework to analyze how these exercises influence portfolio management decisions among European banks. This methodology allows us to compare stress-tested banks with similar non-tested institutions before and after the stress test events, isolating the effects specifically associated with the EU-wide assessments.

Our findings reveal significant patterns of anticipatory behavior, with banks strategically window-dressing their capital ratios before stress tests begin. This behavior is particularly pronounced among institutions that subsequently receive the lowest scores in terms of capital depletion. We document that these anticipatory adjustments lead to decreased portfolio similarity across banks, an effect that persists after the stress tests and remains consistent across different similarity measures. Importantly, such a decrease in similarity does not spin off into more granular business model or country clusters, thus limiting potential systemic risk through portfolio synchronization.

Our results, while considering how financial institutions incorporate stress test considerations into their strategic decision-making, highlight the dual role of stress tests in enhancing individual bank resilience and reducing systemic vulnerabilities. These findings contribute to the ongoing debate on effective banking supervision and the design of regulatory stress testing frameworks.

Non-technical summary

In the aftermath of the Global Financial Crisis, stress testing has emerged as a crucial tool for banking supervisors to assess financial institutions' resilience to severe macro-financial shocks. As these stress test exercises have gained prominence in the supervisory toolkit, they have begun to influence bank behavior in ways that extend beyond their primary purpose to assess and improve the resilience of single banks and the banking sector as a whole. Our paper examines how banks' strategic responses to stress tests affect both individual risk profiles and broader financial stability.

To this end, we assess the impact of the EU-wide stress test — conducted jointly by the European Banking Authority and the ECB Banking Supervision (SSM) — on banks' risk-taking behavior and systemic risk combining three main data sources: detailed balance sheet information from FINREP, including granular data on regional exposures and asset classes; loan-level information from AnaCredit; and confidential supervisory datasets containing detailed stress test submissions, intermediate and final results, and documentation of bank-supervisor interactions during the quality assurance process. We use this rich dataset to construct precise measures of portfolio similarities and analyze systematically how stress test influence portfolio management decisions around stress tests, including ex ante positioning to mitigate the impact of forthcoming stress scenarios or ex post adjustments following the interaction with supervisors or publication of the stress test results.

Our analysis highlights three key findings.

First, we document significant increase in the CET1 capital ratios primarily occurring through ex-ante window-dressing, with banks managing risk-weighted assets rather than capital levels. This anticipatory behavior is particularly pronounced among institutions that subsequently receive low scores in the exercises, suggesting that poorly performing banks engage in more aggressive balance sheet adjustments in preparation of the stress test.

Second, we find that these risk-mitigating actions lead to a decline in portfolio similarity, contributing positively to systemic risk reduction. The management actions by vulnerable banks prove beneficial for financial stability also through the lens of liquidity risk, reducing the risk of shock amplification through coordinated fire sales. Importantly, we find no evidence that supervisory follow-up creates conditions for banks to converge in the composition of their balance sheets. This result holds consistently when measuring similarity using portfolio shares across several dimensions, encompassing sector, region, counterparty, and type of instrument.

Third, we find a decrease in similarity among banks with the same business model but there is no significant effect on the similarity of banks within the same country. This pattern supports a certain homogeneity of EU financial systems, with a more limited role for jurisdictional specificities in influencing exposure profiles and risk management strategies.

From a policy standpoint, our results suggest that the general set-up of the supervisory stress tests should not incentivise banks to develop common strategies allowing the banks to game the exercise and leading to an increased overlap of banks' exposures. It is per se reassuring that the design of the exercises diminishes systemic risk that might arise through portfolio similarity. However, the behavior of banks with weaker fundamentals should be monitored, since these institutions might tend to react more strongly—either to the mere fact of having a looming supervisory stress test on the horizon or to the publication of negative results—potentially increasing idiosyncratic risk even as they contribute to systemic risk reduction.

Our findings contribute to ongoing discussions about the design and implementation of effective supervisory frameworks that balance disclosure, market discipline, and financial stability considerations.

1 Introduction

In the aftermath of the Global Financial Crisis (GFC), stress testing has emerged as a crucial analytical tool in banking supervision (Schuermann, 2014; Guindos, 2019; McCaul, 2021). The severe financial turmoil of 2007-2009 revealed that banks were inadequately prepared for major systemic shocks, lacking sufficient buffers against both credit and market risks. This prompted a fundamental transformation in the regulatory approach to risk assessment, with stress testing becoming mandatory and more rigorous (Hirtle and Lehnert, 2015). The Basel III framework formalized and harmonized stress testing methodologies, making them integral to both prudential supervision and banks' internal risk management frameworks.

The implementation of stress tests has had a dual impact: directly influencing banks' solvency levels and providing market assurance about the resilience of the financial system (Goldstein and Sapra, 2014). Central banks and supervisors now use these exercises to calibrate capital buffer requirements, enhancing banks' resilience to adverse but plausible economic scenarios. Prominent examples include the Federal Reserve's Comprehensive Capital Analysis and Review (CCAR) and Dodd-Frank Act Stress Tests (DFAST) in the United States, and the European Banking Authority's (EBA) biannual stress tests in Europe.

Beyond regular testing cycles, supervisors also deploy stress tests for specific institutional purposes. A notable example is the 2014 Comprehensive Assessment conducted in preparation for the Single Supervisory Mechanism (SSM) in Europe (ECB, 2014; Jabbour and Sridharan, 2020; Carboni et al., 2017a). This exercise served as a foundational health check of bank capitalization levels for the new supervisory framework in the eurozone, identifying capital shortfalls that banks were expected to address promptly following the SSM's inception. Regular stress tests have since become instrumental in evaluating banks' resilience to evolving macroeconomic risks (Kapinos et al., 2018; Baudino et al., 2018).

Although supervisory stress tests do not have explicit pass-or-fail thresholds, banks subject to these exercises may adjust their portfolios to improve their performance metrics. Banks might engage in window dressing ahead of planned stress tests to present favorable results to investors analyzing the outcomes. While such portfolio adjustments can be costly and may be reversed shortly after the stress tests, they can temporarily distort the true picture of banks' financial standing by artificially improving their solvency ratios. This phenomenon has sparked an important debate in the literature, with recent work highlighting how such behavior creates a "beauty contest" problem, where banks focus more on outperforming peers than identifying actual risks (Quagliariello, 2019). This raises broader questions about how to better align stakeholder incentives and enhance the overall effectiveness of stress testing exercises.

More importantly, the application of common scenarios across participating institutions might create unintended systemic risks through banks' coordinated adjustments. Indeed, when facing similar stress scenarios, banks may adopt similar derisking strategies, such as reshuffling their bond holdings to less volatile instruments or those with lower risk weights (e.g., sovereign bonds). This behavioral response leads to increased similarity of bank balance sheets, a concerning development highlighted in the literature on systemic risk related to overlapping portfolios (Allen et al., 2012; Cai et al., 2018). This commonality of assets (and liabilities) creates structural vulnerabilities in the financial system, as banks with similar portfolios tend to react uniformly when faced with shocks. The synchronized deleveraging of similar assets creates a "crowd at the exit door" scenario, where excess asset supply leads to adverse price impacts (Greenwood et al., 2015). In extreme cases, these dynamics can trigger fire sales, amplifying systemic risk rather than reducing it – an outcome that contradicts the very purpose of stress testing.

This study examines how EU-wide stress tests, conducted jointly by the European Banking Authority and the ECB Banking Supervision (SSM), influence banks' de-risking behavior and, subsequently, the systemic risk. Our analysis goes beyond simply examining the results of the stress test or the intensity of supervisory scrutiny during the quality assurance process. Specifically, we also consider banks' management actions taken before these exercises, referring to them as anticipatory management actions. The empirical analysis leverages a unique and comprehensive supervisory dataset covering the 2021 and 2023 stress test cycles, encompassing respectively 79 and 95 significant institutions that participated in the exercises.

We combine three main data sources: detailed balance sheet information from FINREP, including granular data on regional exposures and asset classes; loan-level information from AnaCredit; and confidential supervisory datasets containing detailed stress test submissions and documentation of bank-supervisor interactions during the quality assurance process. This rich dataset enables us to construct precise measures of portfolio overlap and systematically examine how supervisory stress testing activities influence banks' portfolio management decisions.

Our identification strategy employs a difference-in-differences framework around the EU-wide stress test exercises, following the literature analyzing regulatory effects on bank behavior (Gropp et al., 2018; Jiménez et al., 2017; Bräuning and Fillat, 2024), where significant institutions serve as the treatment group and matched LSIs form the control group (Kok et al., 2023a). We adopt the estimator recently proposed by De Chaisemartin and d'Haultfoeuille (2024) to estimate the dynamic effects of the stress test at a quartely frequency, while accounting for potential anticipation. Notably, this approach relaxes the classical parallel trend assumption by allowing it to hold conditionally on observed covariates while controlling for jurisdiction-specific trends in the pretreatment period, thus rducing to the minimum possible concerns about non-random selection of the treated units.

Our identification strategy employs a difference-in-differences framework around the EU-wide stress test exercises, following the literature analyzing regulatory effects on bank behavior (Gropp et al., 2018; Jiménez et al., 2017; Bräuning and Fillat, 2024), where significant institutions serve as the treatment group and matched LSIs form the control group Kok et al. (2023a). We use the estimator proposed recently by (De Chaisemartin and d'Haultfoeuille, 2024) to estimate dynamic treatment effects accounting for potential anticipation and which allows to relax the parallel trend assumption by allowing it to hold conditionally on observed covariates and to control for jurisdiction-specific trends in the pretreatment period.

Our findings reveal several important patterns in bank behavior around stress tests. First, we document signs of anticipatory capital ratio management (in line with previous evidence on Federal Reserve stress test by Cornett et al., 2020), particularly among banks that subsequently receive low scores in stress tests. This strategic window-dressing behavior indicates that institutions anticipate regulatory scrutiny and preemptively adjust their capital positions to present more favorable profiles when entering the assessment period.

Second, we find evidence of decreased portfolio similarity in anticipation of and after stress tests. This result holds consistently when measuring similarity using portfolio shares across several dimensions (encompassing sector and country of exposure, and type of instrument), or when restricting the analysis to marketable assets weigthed by their relative liquidity and when cosidering similarity in loan portfolios. The persistence of this effect beyond the stress test period suggests a structural rather than temporary adjustment in portfolio allocation strategies. These findings demonstrate that stress tests not only serve as point-in-time assessments but also drive longer-term changes in banks' risk management approaches and asset allocation decisions.

Third, we further analyze similarity measures within local clusters, finding a decrease in similarity among banks with the same business model but no significant effect on the similarity of banks within the same country. This nuanced result suggests that while institutions maintain certain country-specific exposures shaped by domestic market conditions and regulations, they increasingly differentiate their portfolios from peers with similar business models. Our results suggest that regulatory pressure related to the recent stress test exercises has contributed to limiting systemic risk associated with portfolio synchronization, indicating that these supervisory exercises can enhance financial stability not only through improved individual bank positions but also imrpoving the resilience of the financial ecosystem as a whole.

These results contribute to our understanding of how regulatory stress testing affects bank behavior and systemic risk (Gandhi and Purnanandam, 2023), while also informing the ongoing policy debate about the design of effective banking supervision frameworks that balance microprudential and macroprudential objectives (Orlov et al., 2023).

The paper is organized as follows. Section 2 summarizes the relevant literature. Section 3.1 outlines a description of the EBA stress test exercises, the definitions of the similarity measures, and the specification of our identification strategy. The results are presented in Section 4, where we first provide evidence of window-dressing behavior, followed by the reduction in portfolio similarity, and finally, the absence of local clustering as a possible unwarranted side effect. Lastly, Section 5 summarizes our main findings and discusses them in the context of the ongoing policy debate.

2 Literature Review

First, this paper contributes to the extensive literature on how bank supervision interacts with capital requirements. Early work by Hancock et al. (1995) and Peek and Rosengren (2000) establishes that changes in bank equity capital affect credit supply, with the latter providing causal evidence from Japanese banks' U.S. operations during a banking crisis in Japan. This fundamental relationship has been extensively documented in different contexts: Jiménez et al. (2017) show the credit effects of countercyclical capital regulation in Spain, while Gropp et al. (2018) find that banks meet higher capital requirements primarily by reducing risk-weighted assets rather than raising new capital. The supervisory dimension is explored by Hirtle et al. (2020), who show that increased oversight leads banks to maintain less risky portfolios without compromising profitability, and Kupiec et al. (2017), who document that poor examination ratings restrict bank lending. Curry et al. (2008) provide complementary evidence that CAMEL ratings affect loan growth differently across credit categories and economic conditions. In the

European context, Couaillier and Henricot (2023) find that higher capital requirements improve bank solvency without significantly affecting shareholder value, suggesting effective regulatory transmission. While this literature establishes the importance of capital requirements and supervision, our paper moves beyond these traditional channels to examine how banks respond to stress testing parameters, particularly when capital levels approach regulatory constraints. This aspect of how management actions impact solvency ratios facing extreme stress parameters prescribed in stress test scenarios is crucial to fully appreciate the dynamics of banks' solvency.

Second, this study connects to research examining how post-crisis banking regulation, particularly stress testing, affects bank portfolio allocation. Bräuning and Fillat (2024) document increased portfolio similarity among U.S. banks subject to stress tests, providing evidence that regulatory pressure can lead to synchronized portfolio choices. Cortés et al. (2020) show effects on credit supply to small businesses, while Konietschke et al. (2022) find that stress-tested banks reallocate credit toward safer borrowers, particularly in retail portfolios. The supervisory channel is examined by Kok et al. (2023a), who provide evidence that scrutiny during stress testing has a disciplining effect on bank risk, distinct from capital requirements. This supervisory effect appears particularly strong when complemented by robust risk management practices. In the European context, Durrani et al. (2022b, 2024) document that stress test performance affects bank behavior through market discipline, while Cappelletti et al. (2024) show that ECB stress tests lead to reduced credit supply to riskier borrowers. These findings built on theoretical work by Farhi and Tirole (2012), who show that imperfect policy instruments generate incentives for portfolio convergence, and empirical evidence from Gandhi and Purnanandam (2023), who document increased comovement in stress-tested banks' returns. Our contribution is to show that banks adjust their portfolios not just in response to capital constraints, but specifically to the stress sensitivities embedded in regulatory stress testing models, with important implications for financial stability. Specifically, we document anticipatory capital ratio management, particularly among banks that subsequently receive low stress test scores. More importantly, while U.S. studies find increased portfolio similarity following stress tests, we reveal decreased portfolio similarity across European banks in anticipation of and following stress tests, suggesting that regulatory pressure in Europe has actually limited systemic risk associated with portfolio synchronization.

Third, our work relates to studies of systemic risk stemming from asset portfolio similarity. This literature highlights fundamental tensions between individual portfolio diversification and system-wide risk (Allen et al., 2012; Caccioli et al., 2014; Goldstein and Leitner, 2020). Cai et al. (2018) measure interconnectedness through loan portfolio overlap based on industry and region, demonstrating that institution-level risk reduction through diversification can generate negative systemic externalities. Abbassi et al. (2017) examine how market risk measures correlate with portfolio similarity, finding that commonality in asset holdings predicts joint crash risk. Recent work by Cai et al. (2022) develops sophisticated measures of bank herding and its impact on systemic risk, while Acharya et al. (2018) show how regulatory pressure can lead to coordinated portfolio adjustments. The European evidence from Loipersberger (2017) suggests that harmonized supervision might inadvertently contribute to increased synchronization. Brunnermeier et al. (2020) provide a theoretical framework explaining how individual banks' optimal responses to regulation can generate systemic vulnerabilities. Our analysis extends this literature by demonstrating specific mechanisms through which stress testing contributes to portfolio convergence, providing a clear link to the increased return comovement documented in Gandhi and Purnanandam (2023) and the systemic risk implications highlighted by Adrian and Brunnermeier (2016).

3 Data and methodology

3.1 Institutional setting and empirical design

In crafting our empirical design, we focus on specific elements of the EU stress test exercises that directly inform our analysis of dynamic portfolio adjustments and strategic behavior.

First, the EU stress test timeline creates distinct windows for bank responses or strategic actions. The exercises follow a predictable cycle, with the European Banking Authority (EBA) typically announcing methodology several months before launch, followed by data collection over Q1-Q2, and results published mid-year [1]. This sequential process enables us to identify anticipatory effects in the quarters preceding the exercise, contemporaneous responses during implementation, and persistent adjustments following publication.

Second, several features of the design facilitate anticipatory behavior. Banks have advance knowledge of the methodology and general scenario design before the official launch[2], allowing

[1]See: Stress test shows euro area banking sector could withstand severe economic downturn, Press release, 28 July 2023, source: `https://www.bankingsupervision.europa.eu/press/pr/date/2023/html/ssm.pr230728~a10851714c.en.html`

[2]Altough the scenario in details is published only at the launch idustry consultations take between supervisory

them to simulate potential impacts and adjust portfolios preemptively. Additionally, since the reference date for the exercise is December 31st of the preceding year, banks have strong incentives to optimize balance sheets specifically for this snapshot date. This explains our focus on detecting anticipation effects in Q3-Q4 of the year prior to the stress test.

Third, regarding portfolio synchronization, the standardized methodological constraints and common scenarios could theoretically drive convergence in bank portfolios. However, the bottom-up approach—where banks use their own models to project losses—combined with bank-specific quality assurance by supervisors may counteract homogenization pressures. This tension between standardization and bank-specific implementation informs our examination of portfolio similarity dynamics.

Importantly, the direct connection between stress test results and supervisory capital requirements—notably through Pillar 2 Guidance[3] calibration—creates powerful incentives for strategic portfolio management both before and after the exercise. This regulatory consequence mechanism amplifies the significance of any detected portfolio adjustments in our empirical analysis.

The key elements of the EU-wide stress tests are further described in the subsequent subsection, providing a foundation for understanding the empirical strategy of our analysis.

3.1.1 The EU-wide Stress Test Framework

The EU-wide stress test is a biennial supervisory exercise coordinated by the European Banking Authority (EBA) to evaluate the resilience of financial institutions under adverse market conditions. This comprehensive assessment involves collaboration with key stakeholders, including the European Systemic Risk Board (ESRB), the European Central Bank (ECB), and national competent authorities.

The stress test employs a constrained bottom-up approach, whereby banks utilize their internal models to project the impact of a predefined scenario on their balance sheets. The exercise projects into a three-year horizon and encompasses two scenarios: a baseline scenario reflecting standard economic conditions and an adverse scenario designed to evaluate banks'

authorities and the banking industry allow banks to infer general trends and focus of the upcoming exercise. Moreover banks have full access to the documents and figures detailing past stress test.

[3]"The level of the Pillar 2 guidance for each bank is based on how it performs in the regular EU-wide stress tests" (source: https://www.bankingsupervision.europa.eu/activities/srep/ as of 03.01.25). For Pillar 2 Requirement see also https://www.bankingsupervision.europa.eu/activities/srep/pillar-2-requirement

resilience under severe yet plausible economic stress.

The implementation of the stress test involves several critical steps. Initially, banks receive standardized templates that cover major risk categories: credit risk, market risk, net interest income, operational risk, and other profit and loss components. Using these templates, banks must forecast how both baseline and adverse scenarios would influence their financial position, adhering to methodological constraints such as the static balance sheet assumption.[4] These constraints are essential to ensure comparability across institutions while accommodating each bank's unique risk profile.

A cornerstone of the process is the rigorous quality assurance (QA) conducted by supervisors. For banks under its direct supervision, the ECB[5] undertakes a thorough review of submissions through multiple validation rounds. This involves benchmarking banks' projections against top-down supervisory models and peer institutions. If material discrepancies arise, banks must either adjust their projections or provide robust justifications for their estimates. This iterative process is vital to ensure the credibility and consistency of the final results.

The outcomes of the stress test serve multiple purposes: they provide critical input for the Supervisory Review and Evaluation Process (SREP) and guide banks' capital planning. In addition, as demonstrated by Kok et al. (2023a), the intensity of supervisory scrutiny during the QA process, gauged by the number and type of challenges posed by supervisors, exerts a disciplining effect on banks' risk-taking behavior.

The timing and granularity of result disclosures are also instrumental in reinforcing market discipline. Durrani et al. (2022a) and Durrani et al. (2023) document significant market reactions to stress test announcements, particularly for banks with weaker performance, underscoring how the disclosure framework amplifies market discipline. This conclusion is further supported by Carboni et al. (2017b), who illustrates that market responses are contingent on the exercise's credibility.

[4]Under the static balance sheet assumption, banks are required to replace assets and liabilities that mature or amortize within the exercise's time horizon with similar financial instruments. Furthermore, no workout or cure of Stage-3 assets is presumed, nor are capital measures taken post-reference date. For further details, see: EBA methodology 2025.

[5]While the EBA stress test is conducted at the European Union level, the ECB's jurisdiction is confined to banks within the Euro Area, whereas non-Euro Area banks are overseen by their respective national authorities.

3.1.2 The 2021 and 2023 Exercises

In comparison to previous exercises, the 2021 and 2023 stress tests marked significant advancements in the European Banking Authority's (EBA) approach, both in scope and methodology. The 2021 exercise, originally scheduled for 2020 but postponed due to the COVID-19 pandemic, assessed 50 EU banks representing approximately 70% of total EU banking sector assets.[6] Within this sample, 38 banks fell under ECB supervision. Concurrently, the ECB conducted the same stress test for an additional 51 SSM banks, with results published only in aggregate form. The adverse scenario was designed to address prevailing concerns about the pandemic's trajectory in a prolonged low-interest-rate environment. It projected negative confidence shocks that could exacerbate the economic downturn, with EU GDP contracting by a cumulative 3.6% by 2023 and unemployment rising by 4.7 percentage points. A distinct feature of the 2021 exercise was the explicit consideration of COVID-19-related factors, such as the treatment of loan moratoria and public guarantees in credit risk assessments.

The 2023 exercise represented a substantial expansion in both scope and severity. The sample size increased to 70 banks, covering 75% of EU banking sector assets, with 20 new institutions added compared to prior exercises. The ECB supervised 57 of these banks and also included 41 SSM banks in the exercise. The adverse scenario was considerably more severe, incorporating hypothetical geopolitical tensions that could lead to persistent inflation and elevated interest rates. This scenario forecasted a 6% cumulative decline in EU GDP and a 6.1 percentage point rise in unemployment over the three-year period, marking it as the most severe scenario in terms of GDP decline used in EU-wide stress tests to date.[7]

Several methodological innovations were introduced in the 2023 exercise. Most notably, the stress test included, for the first time, a sectoral decomposition with data on Gross Value Added (GVA) growth across 16 economic sectors. This granular approach allowed for a more precise assessment of banks' vulnerabilities to sector-specific shocks. Additionally, the 2023 exercise featured enhanced consideration of inflation risks, with inflation remaining above baseline throughout the scenario horizon (3 percentage points higher in 2023 and 1.5 percentage points higher in 2025).

Another key distinction between the two exercises lies in their respective foci. The 2021

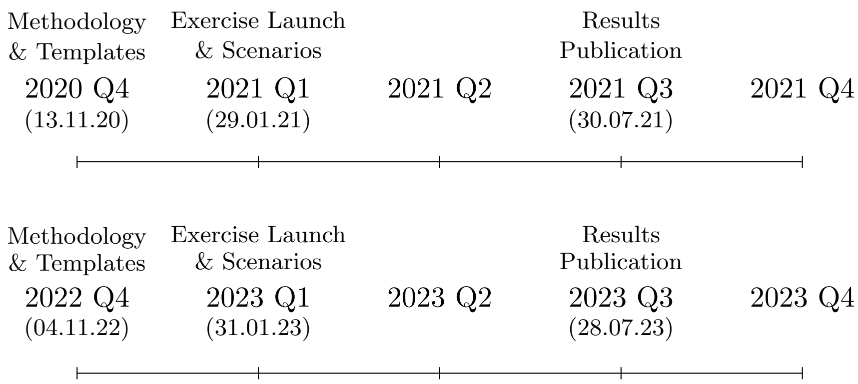
[6] See: hrefhttps://www.eba.europa.eu/risk-and-data-analysis/risk-analysis/eu-wide-stress-testing/stress-tests-2021EBA - Stress Test 2021.

[7] See: hrefhttps://www.eba.europa.eu/risk-and-data-analysis/risk-analysis/eu-wide-stress-testing/stress-test-2023EBA - Stress Test 2023.

stress test concentrated on the ongoing impact of the pandemic and recovery scenarios, while the 2023 exercise focused on emerging risks such as geopolitical tensions, energy price shocks, and persistent inflation. The 2023 scenario also incorporated more severe market risk shocks, with assumed equity price declines of 50% in advanced economies and 65% in emerging economies during the first year.

Both exercises maintained rigorous transparency standards, with detailed bank-by-bank results published to enhance market discipline. However, the disclosure framework evolved between the two exercises, with the 2023 exercise offering more detailed information about sectoral exposures and their performance under stress.

Figure 1: Timeline of the 2021 and 2023 EU-wide Stress Tests



Note: For both exercises, the timeline shows three key phases: methodology publication, exercise launch with scenario release, and results disclosure. The 2021 exercise was initially planned for 2020 but postponed due to COVID-19, while the 2023 exercise represents the most severe scenario to date in terms of GDP decline.

Notably, as illustrated in Figure 1, the timelines of the 2021 and 2023 stress test exercises had parallel structures, with the publication of methodology, launch, and results disclosure occurring at regular intervals.[8] This parallelism provides significant benefits for our empirical analysis: it allows for a consistent examination of the evolution of key variables before and after the stress tests using a stacked dataset where the evolution of the bank-specific outcome variables is analyzed using time relative to stress test execution, thereby enabling us to average the effects across both stress tests increasing thus the number of observations and the robustness of the estimates. However, it will be always possible and form the basis for our heterogeneity analysis to isolate and look at the specific impacts of each stress test vintage, helping to identify whether

[8]Historically, the execution of stress tests extended up to three quarters from launch to completion. In 2021 and 2023, however, the timeline was condensed to two quarters, enhancing the efficacy and timeliness of the stress tests. This streamlined timeline also afforded the opportunity for more targeted thematic stress tests to be conducted during the interim years of the EBA's EU-wide exercises.

the effects, irrespectively of the channels, vary with the severity of the scenarios or the scope of the exercises.

3.2 Data sources and variables

Our study utilizes data from the two most recent European Banking Authority (EBA) stress test exercises conducted in 2021 and 2023[9]. Each exercise's execution phase is organized into three cycles of data quality assurance. These cycles begin with supervisors identifying potential irregularities and inconsistencies in banks' data submissions. Such irregularities trigger a series of flags: data quality checks and "challenging view", where supervisors assess the conservativeness of banks' submissions through peer benchmarking and top-down models. Throughout these cycles, supervisors engage with banks, initially gathering additional materials to clarify detected anomalies. Ultimately, supervisors decide whether the banks' submissions are acceptable or require adjustments towards more conservative projections. Each cycle spans approximately one month.

Stress test results and supervisor-bank interactions. For our study, we use data from the first and third cycles of each stress test, as these two phases are separated by roughly one quarter, aligning with the frequency of regular supervisory data. The primary variable of interest is the Common Equity Tier 1 (CET1) depletion, defined as the difference between a bank's CET1 ratio before and after the stress test simulation, including corrections requested by supervisors. We also have access to confidential data on the number of flags raised and evaluated for each bank, serving as a proxy for supervisory intensity.

Banks balance sheet and financial indicators. The supervisory data, collected quarterly through the Finrep and Corep templates, provide comprehensive insights into banks' balance sheets and regulatory requirements. These datasets are compiled by national supervisory authorities and aggregated by the European Central Bank (ECB) to monitor banks' financial stability and regulatory compliance. Key variables from this data include CET1 and Tier 1 capital ratios, the volume of non-performing loans, profitability measures, and liquidity ratios. Our analysis incorporates detailed information on banks' asset composition by country and instrument type.

[9] We also have access to data from the 2018 stress test exercise, which we use to analyze the persistence of stress test results and retrieve information we deem relevant to understand banks' portfolio dynamics and strategic behavior across consecutive stress tests.

Loans information from AnaCredit. The granularity of supervisory data on loans to non-financial corporations is further enhanced through AnaCredit. AnaCredit, or Analytical Credit Datasets, offers detailed information on individual bank loans within the euro area, harmonized across all Member States. Like supervisory data, AnaCredit is collected as credit registry data from national central banks and subsequently harmonized by the ECB. Before release, the data undergo vetting and verification by the ECB's Statistics Department. AnaCredit includes data on all loans and credit exposures of individual firms and companies exceeding a consolidated position of €25,000, encompassing details on loan amounts, maturities, interest rates, and borrower characteristics. Data collection commenced at the end of 2018, with monthly releases since then. The harmonization and verification process takes approximately three months, ensuring that the data is vetted for each month after this lag.

Our analysis examines banks at a consolidated level, specifically focusing on the European perimeter for banking groups where the balance sheet must be consolidated for the EBA stress tests. After excluding banks with incomplete data, we analyze 79 banks for the 2021 stress test and 95 banks for the 2023 stress test. Given that the sample of Less Significant Institutions (LSI) banks, which forms the majority of the control group, includes up to 3,000 entities, we employ a sample selection procedure to reduce its size. This procedure retains only those non-participating banks which are the most similar to the participating banks with respect to a series of relevant indicators (further details are provided below).

3.2.1 Measuring banks' portfolio similarity

Following Bräuning and Fillat (2024), we construct measures of portfolio similarity to assess the degree of synchronization among banks' asset allocations. We develop these measures along multiple dimensions and complement them with i) a liquidity-weighted metric that move the focus to common exposures on marketable assets accounting for their relative liquidity; ii) a measure of similarity restricted to loan exposures only, using AnaCredit data on loans to non-financial corporations. We also exploit the network structure naturally induced by pairwise similarities to propose a series of centrality-like measures that highlight size effects and identify local clusters of synchronization.

Our analysis, following Bräuning and Fillat (2024), relies exclusively on banks' portfolio exposures rather than market-based measures of comovement as in Sahin et al. (2020)[10] The

[10]Our portfolio-based measures correlate with the average bank pairwise correlation between changes in credit-

former measures and their generalizations aim to identify the underlying portfolio-reallocation mechanisms that drive comovement in bank stock returns (which drive the latter), revealing structural characteristics of banks' investment strategies that become crucial during market distress.

Benchmark pairwise similarity measure. Our baseline approach constructs pairwise bank similarity measures based on the normalized distance between vectors of portfolio shares. For a given dimension D, we decompose portfolio exposures into K categories and compute the portfolio value $a_{k,i,t}$ for each bank i and category k at time t. We then define the D-cosine similarity between the portfolio exposures of banks i and j as:

$$\mathrm{cs}^{\mathrm{D}}_{i,j,t} = \sum_{k=1}^{K} \frac{a_{k,i,t} \cdot a_{k,j,t}}{\sqrt{\sum_{l} a_{l,i,t} \cdot a_{l,i,t}}\sqrt{\sum_{l'} a_{l',j,t} \cdot a_{l',j,t}}} \tag{1}$$

The granularity of our data allows us to consider several years of asset categorization, which define the dimension and level of aggregation of Equation 1. From Finrep, assets are distinguished by type of instrument, country of exposure, and tradable vs. non-tradable assets. Through AnaCredit, we add industrial sectors (NACE2), credit quality, and maturity buckets to the asset dimensions for loans to non-financial corporations. The cosine similarity index defined in equation 1 ranges from 0 to 1, with higher values indicating greater portfolio similarity[11]. We choose this measure for several reasons. First, it readily generalizes to accommodate additional dimensions, such as the liquidity-weighted overlaps discussed below. Second, the measure is computed using normalized vectors (portfolio shares) and thus abstracts from absolute exposure sizes, allowing us to focus purely on portfolio composition similarities.[12]. However, the absolute size of exposures remains relevant for systemic risk considerations, as similarities with larger banks may have greater systemic implications. We decided to address this, not when defining the pairwise-similarity measures, but when aggregating the similarity at the bank level, introducing at that stage the possibility to weight pairwise-similarity by total assets (see Section 4 below).

default swaps (CDS) spreads (see Bräuning and Fillat, 2024), which respond to stress test execution (Sahin and de Haan, 2020; Gandhi and Purnanandam, 2023)

[11]For comparability with existing literature, we note that our cosine similarity measure is mathematically equivalent to the Euclidean distance on normalized vectors, as used in Bräuning and Fillat (2024).

[12]This property is particularly relevant when comparing institutions of different sizes, as it prevents larger institutions from mechanically displaying higher similarity measures.

Liquidity-weighted pairwise similarity. A key advantage of cosine similarity is its flexibility in incorporating parameters that reflect relevant properties of the assets under consideration, as for example the differential liquidity. Following the work by Cont and Schaanning (2019), we can modify equation 1 to account for asset liquidity, providing economically meaningful interpretations of portfolio overlaps in terms of their potential market impact. For each bank pair (i, j):

$$\text{lwcs}^{\text{D}}_{i,j,t} = \sum_{k=1}^{K} \frac{a_{k,i,t} \cdot \omega_k \cdot a_{k,j,t}}{\sqrt{\sum_l a_{l,i,t} \cdot \omega_l \cdot a_{l,i,t}}\sqrt{\sum_{l'} a_{l',j,t} \cdot \omega_{l'} \cdot a_{l',j,t}}} \tag{2}$$

where $a_{k,i,t}$ represents the portfolio value (not the share) in category k and ω_k denotes a liquidity weight reflecting the marketability of assets in that category.[13]

We construct the liquidity weights using the inverse of market depths, following the methodology of Cont and Schaanning (2019), and adapted to FINREP exposures classes by Cuzzola et al. (2023). Market depth for each asset class determines the potential price impact during fire sales, with lower market depth indicating greater price sensitivity to liquidation pressure.[14]

This liquidity-weighted index of pairwise similarity, despite being model-dependend and restricted to a portion of the exposures in the portfolio, improve the measurement of systemic risk by assigning greater importance to common exposures in less liquid assets and providing a direct link to potential fire sale losses. For this reason, it becomes particularly effective at identifying vulnerable portfolio commonalities during market stress periods and quantify the increase in projected expected loss given an increase in similarity.[15]

Bank similarity, clusters, and network structures The importance of measuring portfolio similarity extends beyond individual institutions. As observed by Girardi et al. (2021) for the insurance sector, portfolio overlaps, measured via cosine similarity, can amplify market impacts through coordinated sales, potentially depressing asset prices and affecting the value of financial holdings. These amplification effects persist regardless of whether the initial shock originates within or outside the financial sector.[16]

[13] Notice that imposing $\omega_k = 1$, $\forall k$, Equation 2 becomes 1. Moreover, the denominator ensures normalization of the measure, maintaining its range between 0 and 1, similar to the standard cosine similarity.

[14] The full specification of market depths depends on the choice of liquidation model and market conditions. For a comprehensive treatment, see Cont and Schaanning (2019).

[15] This approach aligns with recent literature emphasizing the role of asset liquidity in amplifying financial system vulnerabilities (see, e.g., Brunnermeier and Pedersen, 2009).

[16] This mechanism relates to the broader literature on fire sales and strategic complementarities in financial markets (see, e.g., Shleifer and Vishny, 2011; Greenwood et al., 2015).

It is then possible to rely on pairwise similarities to construct bank-specific measures of similarity with respect to the banking system or a relevant (economically meaningful) subset. The pairwise similarities defined by equations (1) and (2) naturally generate a fully connected weighted network of the European banking institutions.[17] Within this framework, we construct a bank-level centrality measure by computing the average similarity with all other institutions:

$$\text{Similarity}_{i,t}^{\text{D}} = \frac{1}{N-1} \sum_{j \in \mathcal{B} \setminus i} \text{cs}_{i,j,t}^{\text{D}} \tag{3}$$

where $\mathcal{B} \setminus i$ denotes the set of all banks excluding bank i. This measure captures each institution's degree of portfolio alignment with the broader banking system.[18]

The network structure offers a flexible framework to construct economically meaningful measures of similarity between banks and the financial system. One natural extension incorporates the relative importance of counterparties by weighting the average similarity:

$$\text{Wgt-Similarity}_{i,t}^{\text{D}} = \sum_{j \in \mathcal{B} \setminus i} w_{j,t} \cdot \text{cs}_{i,j,t}^{\text{D}} \tag{4}$$

where $w_{j,t}$ represents the weight of bank j at time t, typically defined as its total assets relative to the banking system. Additionally, we can restrict the similarity measure to specific subnetworks of interest:

$$\text{SubNet-Similarity}_{i,t}^{\text{D}} = \frac{1}{|\mathcal{S}|-1} \sum_{j \in \mathcal{S} \setminus i} \text{cs}_{i,j,t}^{\text{D}} \tag{5}$$

where $\mathcal{S} \subset \mathcal{B}$ denotes a subset of banks sharing common characteristics (e.g., business model, country, or size category) and $|\mathcal{S}|$ is the number of banks in the subset. These variations provide targeted insights into portfolio commonalities within economically relevant peer groups.

3.3 Empirical strategy

Following Kok et al. (2023a), our identification strategy exploits the institutional framework of European banking supervision, wherein some Significant Institutions (SIs) are subject to stress

[17]In the following, we formalize the construction of bank-specific similarity measure only using pairwise similairity from Equation 1. Derived formulas can be easily generalized to the liquidity-weighted version.

[18]This network-based approach aligns with recent literature on systemic risk measurement through interconnectedness (see Elliott et al., 2021).

testing while Less Significant Institutions (LSIs) are not[19]. In this setting, we implement a difference-in-differences (DiD) approach using stress test participation, both in 2021 and 2023, as the treatment and computing the dynamic average treatments effect. Using relative time to stress test inception, the parallel structure of both stress test exercise allows to represent the outome and the treatment variables for a given bank around the stress tests as those of two different banks undergoing the same treatment. It is thus possible to use the obervations in both stress tests for the computation of the dynamic average effects of being into treatment in the quarters around the exercise and report those dynamic effects in the form of event studies for a set of bank fundamentals and a set of similairty measures as defined in Eqs. 3 - 5 above.

The validity of our DiD estimator would require the treatment to be exogenous and randomly distributed among a homogeneous group of banks. However, participation in EU-wide stress tests is determined by a bank's SI status under ECB direct supervision, which is contingent upon meeting at least one of several criteria (European Central Bank, 2024): size (assets exceeding €30 billion), economic importance, cross-border activities (assets exceeding €5 billion with over 20% cross-border ratio), direct public financial assistance, or being one of the three most significant banks in a participating country.

While we acknowledge that our treatment assignment is neither random nor unexpected for the treated units, our institutional setting provides a unique advantage to handle selection bias (Rubin, 1973; Abadie, 2005). We know the observable characteristics determining treatment selection, allowing us to account for structural differences between stress-tested banks and non-significant institutions by controlling for covariates that correlate with both treatment status and our outcomes of interest, whether bank fundamentals or systemic-risk indices.

Technically, since classical DiD estimators rely on two key assumptions - (i) the parallel trends assumption and (ii) no anticipation by treated units - we implement a set of methodological precautions to enhance our identification strategy. Concerning the first assumption, these precautions consist of two main steps. First, we restrict our sample by selecting LSIs that most closely match the characteristics of treated institutions using matching techniques (see

[19]The supervision of Less Significant Institutions (LSIs) falls under national authorities, with the European Central Bank (ECB) primarily serving a coordinating role and facilitating the sharing of results and best practices. Due to the continued heterogeneity in stress testing practices among different national authorities, LSIs cannot be assumed to undergo stress tests of the same rigor as the EBA EU-wide stress test. Nonetheless, some national authorities do conduct stress test analyses for LSIs, applying a proportional approach. Consequently, when estimating the impact of the EBA stress test on Significant Institutions (SIs) compared to LSIs, our findings can be interpreted as reflecting the effect of the EBA stress test in addition to the routine level of supervision, which may include some form of stress testing, at the national level.

Heckman et al., 1997). Second, we leverage recent advances in the DiD literature by employing the estimator proposed by De Chaisemartin and d'Haultfoeuille (2024), which allows the parallel trends assumption to hold conditionally on a set of covariates and incorporates additional controls to minimize structural differences between treated and control groups, including the possibility that parallel treds hold only within group of units. This approach helps to reduce to the minimum the possibility that the observed effects are due to the fact the treated and non-treated units are in different paths for the outcome variable under analysis. Details on both steps are provided below. Concerning the anticipation, we explicitly model its potential effects by running the esimtation assuming different treatment windows allowing them to begin up to the three quarters before the official stress test date.

Finally, we conclude this section discussing how the DiD framework and event studies contribute to identifying the impact of stress tests. This approach inherently controls for shocks affecting both treated and control units. While macroeconomic shocks that exclusively impact treated units could potentially confound our results, this concern is mitigated when the effects of such shocks vary with observable characteristics that we control for, such as jurisdiction and size proxies.[20] Furthermore, our estimation of dynamic treatment effects would highlight anomalous deviations if non-contemporaneous shocks were driving our results. Nonetheless, we acknowledge that our identification strategy cannot disentangle the effects of stress tests from macroeconomic shocks that affect treated units exclusively and in perfect synchronization with the treatment. In such a case, these two phenomena would be perfectly confounded, leaving no room for separate identification. However, we can rule out this concern for both the time windows and the stress test exercises under consideration.

3.3.1 Sample construction

Our proposed approach aligns with recent advancements in evaluating regulatory policies, particularly where treated and control units may differ substantially in key characteristics (Blackman et al., 2018; Arkhangelsky et al., 2021; Kok et al., 2023a).

We construct our sample using the universe of significant institutions subject to stress testing and less significant institutions within the ECB's jurisdiction. By integrating multiple data sources, we retrieve data for 79 treated banks in 2021 and 95 treated banks in 2023. We

[20]Notice, that widespread difference-in-difference estimators. The 'generalized' difference-in-differences estimator does not adjust for everything. The unit fixed effects adjust for all factors that do not change over time within a panel unit, whereas the time fixed effects adjust for the common shocks affecting all units within a time period.

compose the control gourp selecting one LSI for each significant institution assigned through a matching procedure based on a nearest-neighbor approach using the scaled Euclidean distance, without replacement, and conditioned on pre-treatment characteristics. These characteristics include bank size measures (total assets, risk-weighted assets), business model indicators (loan-to-deposit ratio, trading assets ratio), risk profile measures (NPL ratio, CET1 ratio), as well as geographic location and jurisdictional indicators. The use of scaled Euclidean distance ensures that matched pairs are similar across multiple dimensions simultaneously, mitigating the curse of dimensionality typically associated with exact matching on multiple covariates.[21] Following Abadie and Imbens (2006), we perform matching without replacement to maintain independence between matched pairs.

Our descriptive statistics are presented in Table 1. We compare the covariates of banks in our sample during the pre-treatment period, distinguishing between banks in the treatment and control groups. Due to the selection criteria of the stress test sample, banks in the treatment group are substantially larger than those in the control group. As significant institutions, the treated banks have, on average, significantly higher total assets compared to the control banks, which are predominantly less significant institutions.

The analysis reveals that banks subjected to stress tests exhibit a lower risk density, measured by the ratio of risk-weighted assets to total assets. Additionally, these banks tend to have a lower average return on assets and a lower leverage ratio. Conversely, stress-tested banks possess a higher proportion of tradable assets relative to total assets and a greater reliance on short-term funding. Furthermore, they depend less on deposits as a funding source, particularly household deposits.

Table 1 highlights the differences in descriptive statistics before and after implementing our sample matching procedure. Following Imbens and Wooldridge (2009), we also examine the standardized mean differences, as p-values from difference-in-means tests can be misleading in large samples. Standardized mean differences (SMD) close to zero, ideally below 0.2-0.25, indicate a good balance between the treatment and control groups. The test confirms most of our previous findings, with significant differences persisting between the two samples in terms of total assets, risk density, and reliance on deposit funding. Post-matching, the mean-distances

[21] Direct distance-based matching methods often outperform propensity score matching as they ensure paired units have similar values across all covariates. In contrast, propensity score matching may result in pairs with similar treatment probabilities but substantial differences in individual covariates, a key limitation highlighted by King and Nielsen (2019). However, as documented by Ripollone et al. (2018), the relative performance of distance-based methods is context-dependent and may vary across empirical settings.

between the treatment and control groups become statistically insignificant, or at least these differences are notably reduced as highlighted by the variation in the SMD. This is evident in variables such as return on assets, leverage ratio, and the relative size of tradable assets and short-term funding.

Table 1: Summary statistics and univariate analysis of the stress test sample compared to the control group.

	Pre-matching				Post-matching			
	Control group Mean (Std.Dev)	ST sample Mean (Std.Dev)	p-val	SMD	Control group Mean (Std.Dev)	ST sample Mean (Std.Dev)	p-val	SMD
Stress Test 2021								
n	1502	79			79	79		
Total Asset (log)	21.08 (1.26)	24.37 (0.50)	0.001**	3.444	22.46 (1.30)	25.41 (1.33)	0.001**	2.234
CET1 ratio	0.18 (0.05)	0.18 (0.06)	0.973	0.004	0.18 (0.05)	0.18 (0.05)	0.356	0.151
Tier1 ratio	0.18 (0.05)	0.19 (0.05)	0.188	0.154	0.19 (0.05)	0.19 (0.05)	0.903	0.020
Risk Density	53.45 (12.98)	35.74 (13.33)	0.001**	1.346	45.68 (15.87)	35.28 (13.71)	0.001**	0.701
RoE	1.37 (1.02)	1.34 (1.03)	0.774	0.033	1.42 (1.04)	1.36 (0.97)	0.719	0.059
RoA	0.13 (0.11)	0.10 (0.09)	0.007**	0.345	0.12 (0.10)	0.10 (0.09)	0.114	0.258
LCR	243.73 (162.40)	230.67 (121.97)	0.487	0.091	235.03 (95.83)	221.76 (89.63)	0.382	0.143
Leverage ratio	552.89 (145.77)	503.52 (130.84)	0.004**	0.356	506.42 (149.23)	503.68 (127.85)	0.904	0.020
Loans to asset ratio	72.12 (18.59)	76.61 (9.62)	0.054	0.304	76.16 (13.97)	76.55 (9.49)	0.862	0.032
Tradable asset to asset ratio	1.16 (2.85)	4.04 (4.84)	0.001**	0.725	2.87 (7.26)	5.07 (7.45)	0.129	0.299
Short-term funding ratio	0.03 (0.16)	0.22 (0.35)	0.001**	0.687	0.23 (0.69)	0.41 (0.80)	0.223	0.243
Deposit ratio	86.15 (20.41)	73.76 (21.53)	0.001**	0.591	84.42 (20.55)	73.78 (21.47)	0.011*	0.506
Stress Test 2023								
n	1533	95			95	95		
Total Asset (log)	21.11 (1.34)	24.44 (0.54)	0.001**	3.261	22.31 (1.18)	25.39 (1.26)	0.001**	2.527
CET1 ratio	0.18 (0.05)	0.18 (0.05)	0.410	0.088	0.18 (0.06)	0.18 (0.06)	0.716	0.054
Tier1 ratio	0.18 (0.05)	0.19 (0.05)	0.331	0.104	0.19 (0.06)	0.19 (0.06)	0.673	0.062
Risk Density	53.02 (13.68)	36.12 (14.06)	0.001**	1.218	43.35 (16.06)	35.64 (14.21)	0.001**	0.509
RoE	2.78 (1.69)	2.37 (1.51)	0.022*	0.258	2.73 (1.87)	2.38 (1.54)	0.171	0.203
RoA	0.28 (0.18)	0.18 (0.14)	0.001**	0.604	0.23 (0.16)	0.18 (0.14)	0.034*	0.315
LCR	228.46 (151.05)	217.48 (108.77)	0.491	0.083	235.34 (86.53)	209.52 (78.69)	0.035*	0.312
Leverage ratio	530.00 (138.45)	482.41 (121.31)	0.001**	0.366	515.65 (147.21)	483.52 (117.94)	0.104	0.241
Loans to asset ratio	72.66 (19.59)	74.32 (14.81)	0.475	0.096	74.49 (16.59)	74.31 (14.66)	0.947	0.011
Tradable asset to asset ratio	0.78 (2.22)	4.16 (5.01)	0.001**	0.871	1.98 (6.09)	6.24 (9.68)	0.004**	0.526
Short-term funding ratio	0.02 (0.11)	0.26 (0.37)	0.001**	0.895	0.06 (0.30)	0.56 (1.00)	0.001**	0.668
Deposit ratio	86.71 (21.02)	73.84 (20.79)	0.001**	0.616	86.99 (14.49)	73.96 (20.40)	0.001**	0.737

Source: Stress test and supervisory data. Note: ***, **, * indicate $p < 0.001$, $p < 0.01$, $p < 0.05$ respectively. Risk density is calculated as Risk-Weighted Assets (RWA) divided by total assets. The short-term funding ratio is determined by dividing short-term funding by total liabilities. The deposit ratio is calculated as total deposits divided by total liabilities. Additionally, the deposit ratios for financial institutions, non-financial corporations, and households are computed relative to total liabilities.

Recognizing that the substantial difference in bank characteristics between the treatment and control groups could potentially bias our estimates and undermine their comparability, we take additional precaustions in the design of the difference-in-difference analysis. In particular we assume parallel trends to hold conditionally on the set of covariates which inherently mark the differences between our treated and control groups.

3.3.2 Difference-in-differences estimator, conditional parallel trends and anticipation

We employ the estimator proposed by De Chaisemartin and d'Haultfoeuille (2024), which belongs to a new wave of DiD estimators designed to handle dynamic and heterogeneous treatment effects allowing for the classic parallel trend assumption to hold conditionally on a set of covariates and proposing systematic testing for this assumption in the pre-treatment period.[22]

We define a time index realtive to the treatment l (with $l = 0$ at treatment inception) and for a given outcome Y, we compare the difference in Y of a given bank i in the treatet group ($\mathcal{T}$) at l having the period -1 as baseline[23]:

$$\mathrm{did}_{i,t}(Y) = (Y_{i,t-1} - Y_{i,-1}) - \frac{1}{N_{\mathcal{C}}} \sum_{j \in \mathcal{C}} (Y_{j,t-1} - Y_{j,-1}) \tag{6}$$

where $\mathcal{C}$ is the group of control banks and $N_{\mathcal{C}}$ the number of units therein.

We then aggregate the dynamic bank-specific effects to obtain the estimator[24]:

$$\mathrm{DID}_t(Y) = \frac{1}{N_{\mathcal{T}}} \sum_{i \in N_{\mathcal{T}}} \mathrm{did}_{i,t}(Y) \tag{7}$$

where $\mathcal{T}$ is the group of treated banks and $N_{\mathcal{T}}$ the number of units therein.

Notice here that this estimator, given the overlapping timing structure of stress test exercises, can accomodate the estimation of the effects of both stress tests or only one. As anticipated, it is just necessary to create a stacked panel structure where units are defined by the combination of bank identifier and exercise label (2021 or 2023) and timing is aligned with 2021-Q1 corresponding to 2023-Q3 (and so on).

Moreover, the estimator allows to estimate aggregated post-treatment coefficients and to accomodate treatments which are not necessarily binary, i.e. to represent the tretment using

[22] Altough the validity of the assumption is not testable by definition, as it concerns a non-observed counterfactual (i.e. that the path of the treated and untreated units would have been parallel in the absence of the treatment), it is accepted to test for pretreatment differences in trends ("pre-trends") as a way of assessing the plausibility of the parallel trends assumption(see Roth, 2022, for a critical analysis).

[23] We keep this baseline for comparability with standard two-way fixed effects (TWFE) estimators. Concerning other options, one could average the units' outcomes from the starting observation to -1, giving rise to another unbiased estimator for the dynamic effects. See De Chaisemartin and d'Haultfoeuille (2024) and also Borusyak et al. (2024) for a thorough discussion.

[24] The proposed equation substantially differs from the classical TWFE regressions with the treatment indicator interacted with post-treatment period fixed effects (see Cornett et al., 2020; Kok et al., 2023a, for applications to a context similar to ours). The conditions under which these estimators are equivalent to ours can be found in Section 4.1 of De Chaisemartin and d'Haultfoeuille (2024).

discrete and continuos variables[25][26].

In our study, heterogeneity in treatment timing is not a primary concern. However, we choose this estimator for its flexibility in the identification assumptions which we summarize here.

Conditional parallel trends. The estimator facilitates formal testing of pre-trends to evaluate the plausibility of parallel trend assumption. More importantly, as anticipated, it allows the parallel trends assumption to hold conditionally on a set of covariates. This feature is particularly valuable in our context, as stress-tested banks differ fundamentally from Less Significant Institutions (LSIs) across several dimensions, even after matching. In particular, the asset size threshold that determines stress test participation creates inherent differences between treated and control banks. The conditional parallel trends framework allows treated and control units to experience differential trends, provided those differential trends are fully explained by changes in some observed covariates (see De Chaisemartin and d'Haultfoeuille, 2024, Appendix 1.2 for details). Formally this means defining:

$$\begin{aligned} \mathrm{did}_{i,t}^{\mathbf{X}}(Y) &= (Y_{i,t-1} - Y_{i,-1}) - (\mathbf{X}_{i,t-1} - \mathbf{X}_{i,-1})' \, \widehat{\boldsymbol{\vartheta}} \\ &- \frac{1}{N_{\mathcal{C}}} \sum_{j \in \mathcal{C}} \Big((Y_{j,t-1} - Y_{j,-1}) - (\mathbf{X}_{j,t-1} - \mathbf{X}_{j,-1})' \, \widehat{\boldsymbol{\vartheta}} \Big) \end{aligned} \tag{8}$$

where $\mathbf{X}_t$ is a vector of selected bank observables and $\widehat{\boldsymbol{\vartheta}}$ denote the coefficient of $\mathbf{X}_t - \mathbf{X}_{t-1}$ in the OLS regression of $Y_t - Y_{t-1}$ on $\mathbf{X}_t - \mathbf{X}_{t-1}$ and time fixed effects, in the sample of all untreated units and treated units before the treatment. $\mathrm{did}_{i,t}^{\mathbf{X}}(Y)$ is different from $\mathrm{did}_{i,t}(Y)$ for the fact that instead of comparing banks' outcome evolution, it compares the part of that evolution that is not explained by a change in the covariates.

Group-specific trends. In some cases, controlling for covariates may be insufficient to account for differences in trends between units. A common solution in static or dynamic two-way

[25]This is possible generalizing equation 6 to the case of continuos or discrete treatment. Covering the details here is beyond the scope of the paper, thus we refer to De Chaisemartin and d'Haultfoeuille (2024) for further details.

[26]These features facilitate the comparison with existing literature by allowing the computation of aggregated post-treatment effects over a specified horizon and to estimate the marginal effects associated to the differential exposure to the treatment, when fully enconed in some observable variable. In some parts of our analysis, this will enable the direct comparison with previous findings in the literature on supervisory impact. Tipically this has been exploited in past applications to uncover the effects of specific channels as the supervisory scrutiny, market discipline and capital channels (see Durrani et al., 2023; Kok et al., 2023b).

fixed effect regressions consists in including interactions between time FE and FE for sets of units. For instance, referring to our application, one can allow for European banking jurisdiction-specific (country-specific) trends. A similar idea can be pursued in our context. Let $s \in \{1, ..., S\}$ denote the sets partitioning treated and control units, we can define the estimator

$$\mathrm{did}^s_{i,t}(Y) = (Y_{i,t-1} - Y_{i,-1}) \; - \; \frac{1}{N^s_{\mathcal{C}}} \sum_{j \in \mathcal{C}^s} (Y_{j,t-1} - Y_{j,-1}) \tag{9}$$

for any s and any i belonging to the treted units and to the partitioning set s. $\mathrm{did}^s_{i,t}$ is similar to $\mathrm{did}^s_{i,t}$, except that it only compares the outcome evolution of units in the same set s. In our application, $\mathrm{did}^s_{i,t}$ compares the outcome evolution of banks in the same countries. Then dynamic effects can be easily recovered in the form of Equation 7 by aggregation. Notice also that Equations 8 and 9 can be combined to accomodate the parallel trend assumption assuming both conditionality on a set of covariates or validity only for units within the same set of the partition. This allows us to construct the main estimators aggregating jurisdiction-specific estimates, thereby increasing the plausibility of the parallel trends assumption by eliminating potential violations arising from comparisons between institutions operating in different countries. This is especially relevant in the EU banking system, which, despite being integrated, maintains significant jurisdictional specificities in terms of business cycles, regulatory frameworks, and market structures (Anna-Lena Högenauer and Quaglia, 2023).

Anticipation and Dynamic Effects. A key challenge in our context is the possibility of anticipation behaviours, which might violate the no-anticipation assumption of DiD estimators. Anticipation occurs when banks adjust their behavior in advance of the stress test, knowing that they will be treated. This is particularly relevant in our context, as the list of stress-tested institutions is disclosed well in advance. As noted by Malani and Reif (2015), anticipation effects have substantial implications for the interpretation of pre-trends, as they may reflect forward-looking behavior rather than endogeneity. For example, anticipation effects that mirror the post-treatment direction of effects can lead to underestimation of the treatment effect if ignored.

The problem of strategic anticipation is well-documented in the literature on stress tests. Notably, Quagliariello (2019) discuss how stress-tested institutions engage in "beauty contests" adjusting their portfolios to appear more resilient in the eyes of regulators and markets. This dynamic clearly emerges in our findings, as banks with potentially lower performances in the stress test engage in significant de-risking and rebalancing at the year-end preceding the stress test. These pre-treatment adjustments raise concerns about the validity of estimated treatment effects in studies that fail to account for this anticipation behavior.

To address this challenge, we explicitly model anticipation effects by defining treatment windows that begin before the official stress test date. Specifically, we estimate models with anticipation periods of one, two, and three quarters before the stress test. This approach allows us to: i) Identify the optimal pre-treatment window where parallel trend assumptions are most likely to hold, testing for the presence of anticipation effects using our conditional parallel trends framework; ii) quantify the magnitude of anticipatory responses. This way we account for anticipation while ensuring robust estimation of treatment effects during the execution of the stress test and at later stages. Notably, this is the first study to address the problem of stress test strategic anticipation analytically. Our evidence of anticipatory behavior poses a challenge to existing studies of stress test effects, as their choice of pre-treatment window may project a bias onto estimates if pre-treatment adjustments are not considered.

4 Results

The paper investigates whether banks' participation in the 2021 and 2023 EU-wide stress tests has an impact on individual bank risk and on the systemic risk implied by banks' portfolio syncronization.

Our analysis focuses on the dynamic effects of the stress tests in an attempt to isolate the timing and channels through which each exercise can influence banks strategy and portfolio choices, and in particular to highlight and quantify the effects of anticipating behaviours.

Throughout, our benchmark exercises are complemented by a set of heterogeneity analyses and robustness checks to uncover whether results are driven by specific subsamples of banks or by distinctive features of different stress test vintages.

Our analysis proves three main results:

Stress test-induced derisking via window-dressing. We find evidence of banks' de-risking, with major changes occuring ex-ante (window-dressing channel), rather than ex-post (triggered by supervisory scrutiny channel or market discipline channel). Moreover, banks that poorly perform in the stress tests tend to engage more in ex-ante de-risking.

Aggregate decline in similarity and positive effect on systemic risk. The strategic management of risk is idiosyncratic, i.e., there is no evidence of herding behaviour. The effect is more pronounced among banks with higher capital depletion under the stress test scenarios. Asynchronous management actions of more vulnerable banks are even more important for the financial stability as they reduce the risk that they may amplify shocks through similar rebalancing of their portfolios or fire sales.

Decrease in similarity robust to local clusters. Locally, i.e., at a country level or within the same business model, the similarity does not increase. Our conclusion on systemic risk, therefore, holds for local environment (i.e., geographical, legal or historical, etc.) and business model connections possibly affecting banks. Concerning this aspect of the EU financial system, the impact of the stress tests on similarity of significant institutions appears harmonised and does not depend on specificities of banks' domicile.

4.1 Bank behaviours around stress tests: the window-dressing channel

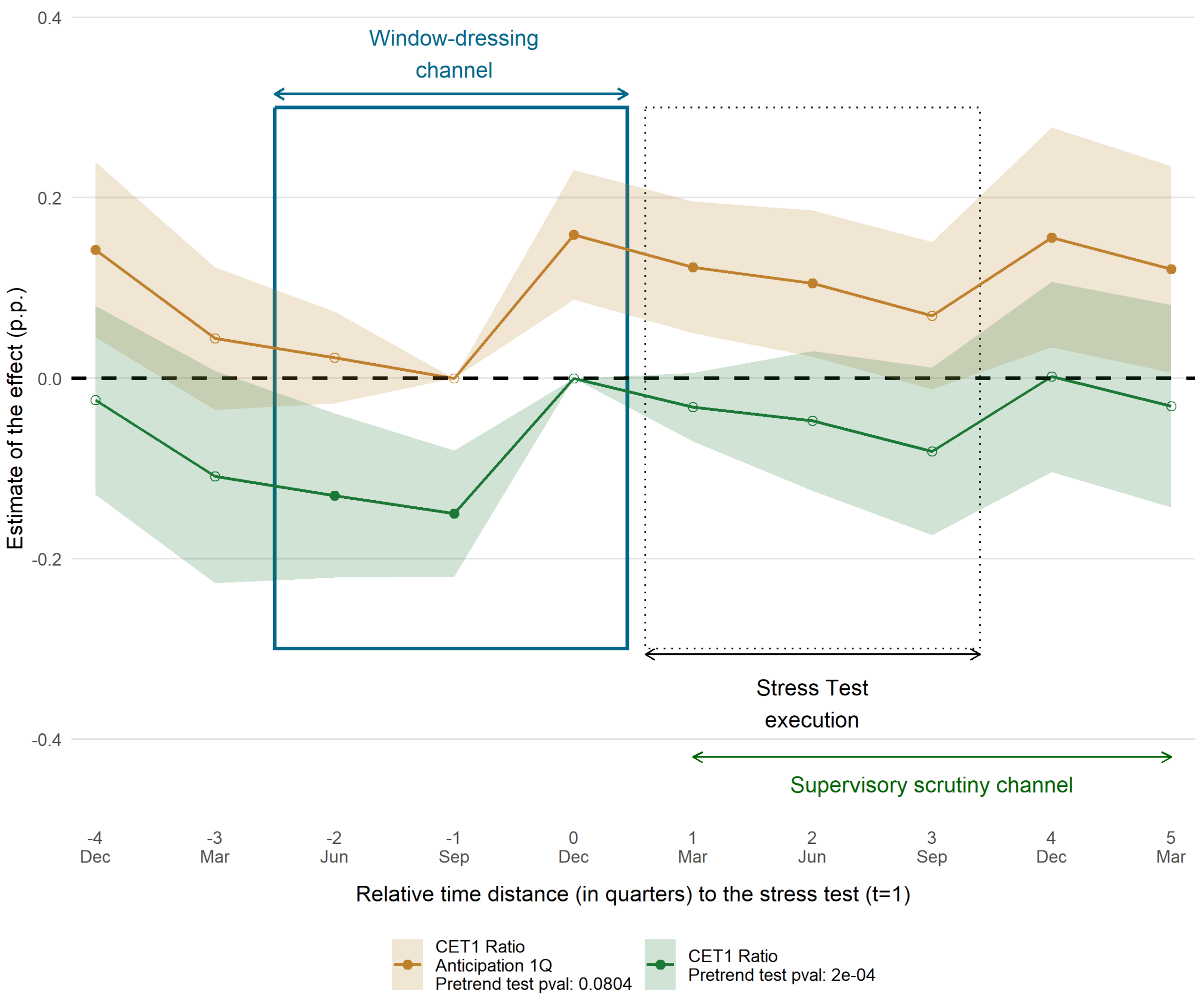
To investigate whether stress-tested European banks exhibit divergent trajectories in their fundamentals compared to non-stress-tested banks, we employ a difference-in-difference empirical design, using stress test participation as a treatment variable. For each output variable, such as capital ratios and profitability, we compute dynamic effects at quarterly intervals.

The identification of causal effects in our methodological setup is based on parallel trends and no-anticipation assumptions (see Section 3.3.2 for details). Given concerns about banks' strategic anticipatory behavior(Quagliariello, 2019), we consider that the treatment effect may begin up to three quarters before the stress test's official inception date (i.e., Q1 of the stress test year). Specifically, we consider treatment set at three, two, and one quarters lag prior to the start of the stress test. We then select the lag that most strongly supports the assumption of parallel trends as verified by the absence of divergent trends between the tested and untested banks prior to treatment (also relying on standard parallel pre trend testing as explained by Roth, 2022; De Chaisemartin and d'Haultfoeuille, 2024).

Analysing banks' output dynamics around the stress test exercises, we can identify the specific channels of influence on bank behaviors and balance sheets. Specifically, we distinguish between the *window-dressing channel*, encompassing year-end adjustments made by banks to manipulate the starting point data of the stress test, and other channels affecting bank fundamentals during the execution of the stress test and the in the immediate aftermath such as the *supervisory scrutiny* channel (Kok et al., 2023b)[27].

Our estimates across various bank-level outputs reveal systematic anticipatory behavior by banks preparing for the stress test.

Figure 2: Stress Test Impact on the CET1 Ratio



Note: The plot shows dynamic treatment effects for the CET1 Ratio (in percentage points) estimated using (De Chaisemartin and d'Haultfoeuille, 2024), with confidence intervals at 95% level. Full dots highlight significant effects and pre-trend test p-values are reported in the legend. All specifications use bank-level clustering for standard errors and assume parallel trend to hold within country and conditional on total asset, return on equity, and levels and growth rate of the CET1 ratio before the treatment. The timeline is expressed in quarters relative to stress test initiation (t=1). The estimation is provided in two specification: one (in green) assumes the treatment starts at Q1 in the year of the stress test; one (in yellow) accounts for anticipation effects by using quarter -1 as the reference point for comparing pre-treatment trends and estimating the effects of the stress test already in december (t=0).

[27]Extant literature focused also on the *market discipline* and *capital* channels (see e.g. Durrani et al., 2024).

The evolution of the capital ratio around the stress test provides a representative example of the importance of accounting for anticipatory effects. Figure 2 presents estimates from two specifications of our dynamic difference-in-differences model (see subsection 3.3.2). The first and standard specification treats the launch of the stress test (Q1 in the test year) as the intervention date, after which the two groups of banks operate under different regimes, conditional on a set of characteristics we control for. The second specification accounts for anticipation by setting the treatment one quarter prior to the stress test execution.

The standard specification, which assumes that treatment begins at the launch, shows a violation of parallel trends (pre-trend p-value = 0.0002)[28], indicating significant pre-treatment divergence between treated and control banks. The estimated coefficients before treatment reveal that this divergence is primarily attributed to a significant jump between September and December of the year preceding the stress test. When adjusting for anticipation by setting the treatment start as of December, we achieve parallel trends (p-value = 0.0804) and identify a statistically significant increase in CET1 ratios of 20 basis points during the last quarter before the stress test year. This pattern is consistent with strategic balance sheet management in preparation for the stress test.

Notice that our identification strategy, relying on the stress time window prior to the stress tests, isolates any anticipatory effect from other concurrent regulatory changes and market conditions. The timing and magnitude of the observed capital adjustments, therefore, suggest that banks attempt to optimize their positions before the execution phase of the stress tests begins.

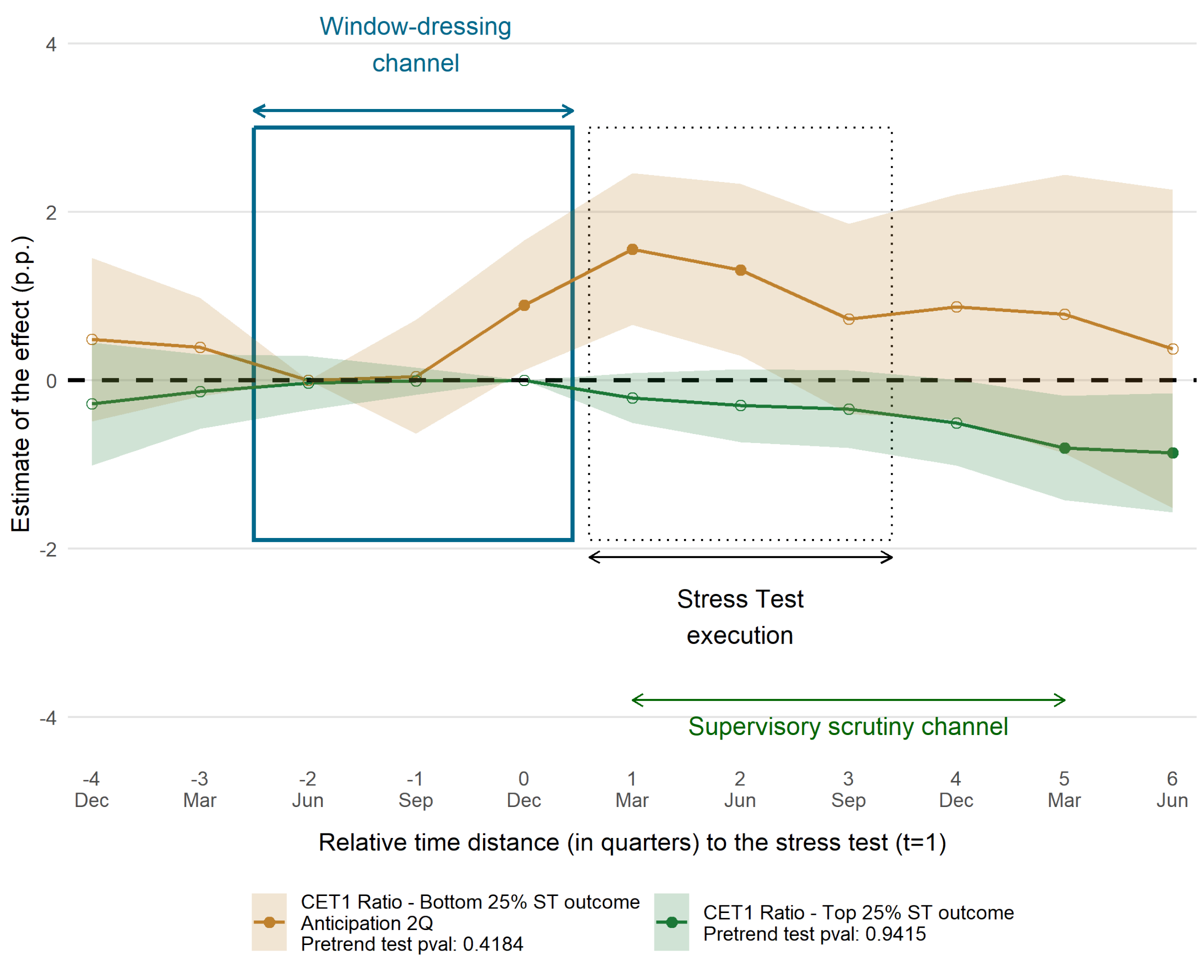
The baseline results of Figure 3 capture the average anticipatory response across all stress-tested banks. Aggregate effects, however, may cover underlying heterogeneity in both timing and magnitude of balance sheet adjustments across different institutions. To test such heterogeneity, we show that performace in the stress test is a key dimension to explain such a heterogeneity, as banks' anticipatory behavior differs systematically depending on banks' expected performance under stress scenarios.

Anticipation effects are stronger for banks that perform worse in the exercises. Figure 3 reveals substantial variation in anticipatory behavior between banks in the top and bottom

[28] Concerning the interpretation of the p-values for the pretrend tests here and in the following, please notice that the null hypothesis assumes the joint nullity of the placebos, namely that there are no differences - conditionally on the chosen covariates - in the path of the observed variables for the treatment and control units. In this is setting an higher p-value points in the direction of the non rejection of the null.

quartiles of stress test outcomes. Banks that ultimately rank in the bottom quartile exhibit significant capital adjustments ahead of the execution phase of the stress tests, i.e., 2 quarters in advance, and increase their CET1 ratios by approximately 2 percentage points during the window-dressing period. In contrast, top performers do not show a significant anticipatory effect.

Figure 3: Heterogeneous Effects on CET1 Ratio for Top and Botton Stress Test Scorers



Note: The plot shows dynamic treatment effects for the CET1 Ratio (in percentage points) estimated using De Chaisemartin and d'Haultfoeuille (2024) for two different samples where treated units are the banks in the bottom 25% in terms of projected capital depletion in the stress test or bank in the top 25%. Control group is maintained constant for the two estimates as per Table 1. Confidence intervals are at 95%, full dots hihglight significant effects and pre-trend test p-values are reported in the legend. All specifications use bank-level clustering for standard errors, and assume parallel trend to hold within country and conditional on total asset, return on equity, and levels and growth rate of the CET1 ratio before the treatment. The timeline is expressed in quarters relative to stress test initiation (t=1). The estimation for the bottom 25% sample accounts for anticipation effects by using quarter -2 as the reference point for comparing pre-treatment trends. The results indicate that, during the window-dressing period, banks in the bottom quartile experience an average increase of 2 percentage points in their CET1 Ratio relative to the pre-treatment period.

At first glance, an observation that banks with the strongest anticipatory adjustments end up performing poorly in the stress tests appears counterintuitive, as we would expect window dressing to pay off when conducted.

Further considerations explain this result. First, the pattern suggests that the stress tests are effective in identifying the risk of poor performers despite their efforts to adjust the starting point data which feed into banks' models projecting the stress test impact. Second, banks are aware of their vulnerability to the specific stress test scenarios based on their knowledge of the stress testing methodology and take pre-emptive actions to ensure that they maintain adequate capital buffers even under stressed conditions. This latter interpretation would suggest that banks engage in strategic behavior to minimize the impact of stress test losses, especially when they expect to approach critical capital buffer levels in the hypothetical depletion implied by the scenarios. Figure 4 confirms this pattern by showing that the poorer the performance of banks in previous stress tests the stronger are the anticipatory responses.[29]

Third, and perhaps more compelling, the anticipatory behavior reflects the fact that banks which have the highest depletion in the stress test were likely in the same tail of the performance distribution in the previous stress exercise. This is highlighted by the transition probabilities across stress test cycles reported in Table 2 revealing substantial stability in performance rankings. For the 2018-2021 cycle, almost 40% of banks in the bottom quartile remained there in the subsequent test, while this persistence increased to 50% in the 2021-2023 cycle. Similar stability characterizes the top performers, with 64% and 50% remaining in the top quartile for the respective cycles.

As an additional robustness check, detailed in Appendix A, we control for potential heterogeneity of these effects across stress test exercises, comparing separately the dynamics around the EBA stress test in 2021 and 2023. We find that the anticipatory patterns are similar in 2021 and 2023 exercises (Figure 9). The positive effects from the 2021 stress test are slightly larger and more persistent in time. The 2023 stress test elicited a higher anticipatory behavior in the quarter preceding the 2023 stress test. The increased CET1 ratios in both exercises, however, rapidly set-off already within the stress test execution, displaying an inverse U-shaped pattern.

[29]Importantly, even though the performance in the past stress tests appears to be an important marker for strategic anticipatory behaviours, we decided to keep the performance in the current stress test as the primary dimension for heterogeneity analysis in the following analyses. This way, we aim to capture both the persistence in banks projected financial vulnerability and the intention to anticipate due to scenario-specific (or, broadly speaking, exercise-specific) considerations.

Table 2: Transition Matrices for the deplition quantiles

Transitions 2018-2021

		2021			
		Bottom 25%	Middle 50%	Top 25%	Not ST
	Bottom 25%	9 (39.13%)	12 (52.17%)	2 (8.7%)	0 (0%)
2018	Middle 50%	3 (10%)	21 (70%)	6 (20%)	0 (0%)
	Top 25%	0 (0%)	9 (36%)	16 (64%)	0 (0%)

Transitions 2021-2023

		2023			
		Bottom 25%	Middle 50%	Top 25%	Not ST
	Bottom 25%	6 (50%)	3 (25%)	1 (8.33%)	2 (16.67%)
2021	Middle 50%	3 (7.14%)	21 (50%)	9 (21.43%)	9 (21.43%)
	Top 25%	2 (8.33%)	10 (41.67%)	12 (50%)	0 (0%)

Note: The tables report the transitions from different positions in the distribution of the final stress test capital depletion.

4.1.1 Banks fundamentals around the stress test

We complement the observed exercises on CET1 Ratio with a set of estimates including the most relevant bank-level outcomes, detailed in Appendix B.

First, Our analysis shows that banks primarily achieve the observed increases in capital ratios through strategic reductions in risk-weighted assets rather than through increases in capital levels. In Appendix B.1, we decompose the CET1 ratio effects and find that during the window-dressing period, banks in the bottom quartile of stress test performance reduce their risk-weighted assets by approximately 5%, while maintaining relatively stable capital levels (Figures 12 and 13). This asymmetric response suggests that banks find it more efficient to manage regulatory ratios through the denominator rather than the numerator.

Second, further analysis in Appendix B.2 reveals interesting trade-offs in banks' strategic adjustments. We observe a significant decrease in the Liquidity Coverage Ratio (LCR) during the window-dressing period (Figure 14), with an average reduction of 30% across all stress-tested banks. This reduction in liquidity buffers, while seemingly counterintuitive before a supervisory exercise, aligns with a strategic reallocation of resources toward capital ratio optimization, especially given that stress tests primarily focus on solvency metrics rather than liquidity requirements. Despite these substantial adjustments to balance sheet composition, we find no significant impact on banks' profitability as measured by Return on Equity (Figure 15). This suggests that the anticipatory risk-weighted asset reductions are implemented in ways that

Figure 4: Stress test effects on CET1 Ratio: Bottom-performers in the past Stress Test

Note: The plot shows dynamic treatment effects estimated using De Chaisemartin and d'Haultfoeuille (2024) for the CET1 Ratio (in percentage points) of banks which where at the bottom of the distribution in the previous stress test exercises. Effects are displayed as points (full if significant) with confidence intervals (95% level) shown as bands. All specifications use bank-level clustering for standard errors, and assume parallel trend to hold within country and conditional on total asset, return on equity, and levels and growth rate of the CET1 ratio before the treatment. The timeline is expressed in quarters relative to stress test launch (t=1). The estimation accounts for anticipation effects by using quarter -2 as the reference point for comparing pre-treatment trends. The results indicate that during the window-dressing period banks that were most penalized in the preceding stress test anticipate that the incoming stress test increase their CET1 Ratio up to 2 percentage points.

preserve earnings capacity, at least in the short term.

Finally, for methodological comparability with other contributions in the literature, particularly Durrani et al. (2024), we also try to detect the effects of the publication of the stress test on bank indicators, specifically CET1 Ratio and Return on Equity, in Appendix C. Exploiting the differential disclosure policy in European stress tests (granular results for EBA banks versus aggregated results for SSM banks), we find that poor-performing banks with publicly disclosed results experience a 0.55% decrease in CET1 ratio for each percentage point of additional capital depletion, while top performers see a 1.03% increase in ROE for each percentage point of better performance. Notably, we do not observe significant symmetric effects: the CET1 ratio of top performers and the ROE of poor performers show no statistically significant changes following publication. This confirms that beyond window dressing, stress tests also influence bank behavior through market discipline, though we caution that our quarterly data and the indicator used may not fully capture short-term market reactions that could be detected with higher-frequency observations.

4.2 General decline in portfolio similarity

In turn, we examine whether banks' balance sheet management around the stress test exercises increases synchronization in portfolio allocations. Two competing hypotheses follow from our findings on de-risking described in section 4.1. On one hand, since banks face a common evaluation framework through the stress test scenario, they might converge toward similar portfolio structures – a "beauty contest" effect where institutions optimize with the same set of constraints. This hypothesis suggests increased portfolio similarity as banks attempt to minimize their vulnerability to the anticipated scenario.

On the other hand, our earlier evidence that poor performers make persistent adjustments across the stress test cycles suggests an alternative hypothesis. Rather than converging to a common portfolio structure, banks might pursue unique derisking strategies that reflect their specific constraints, business models, and local market conditions. Under this hypothesis, while all banks might reduce risk, they would do so in different ways aligning with their distinct starting positions and characteristics.

Our analysis of portfolio synchronization develops along three main dimensions, where each dimension insights into different sections of banks' balance sheets. First, we examine overall portfolio similarity to provide a complete picture of portfolio overlap. This metric leverages data on the most granular balance sheet structure available in the Finrep supervisory reporting, which combines exposures across several dimensions: type of instrument, type of counterparty, country. Second, we zoom into the securities portfolios of the balance sheet, focusing on tradable assets only and using a liquidity-weighted similarity measure. Through this approach we aim at capturing the systemic risk arising from common exposures to assets vulnerable to fire sale dynamics during financial stress. Third, we further increase the granularity of the analysis by looking at the composition of the loan portfolios. Such an analysis is made possible by merging data from AnaCredit to the superviory reporting dataset. This allows us to investigate overlaps in loan to non-financial sectors using exposure values in NACE2-country cells and analyse their changes following stress tests.

4.2.1 Decreasing portfolio similarity

Our first analysis examines whether stress tests lead to increased portfolio similarity across banks. Following Bräuning and Fillat (2024), in the benchmark exercise, when computing similaritie we always compute similarity of stress tested banks with stress tested banks and, for the control group, banks of non stress tested banks with their peers[30]. This approach allows us to identify whether the stress test exercise induces convergence or divergence in portfolio allocation strategies among participating institutions.

The results reveal two key patterns (Figure 5). First, during the window-dressing period, portfolio similarity decreases by approximately 1 percentage point relative to control banks. This decline is statistically significant and economically meaningful, representing a substantial deviation from the synchronized derisking hypothesis. Poorly performing banks, while actively manage their risk-weighted assets, do so through idiosyncratic strategies rather than converging toward common portfolio structures.

Second, the decline in similarity persists and even intensifies during the execution of the stress test and the following quarters, indicating that the divergence in portfolio structures is not merely a temporary adjustment. This persistence suggests that stress tests may encourage banks to develop more distinctive portfolio strategies aligned with their individual strengths and constraints, rather than promoting homogenization of risk profiles[31].

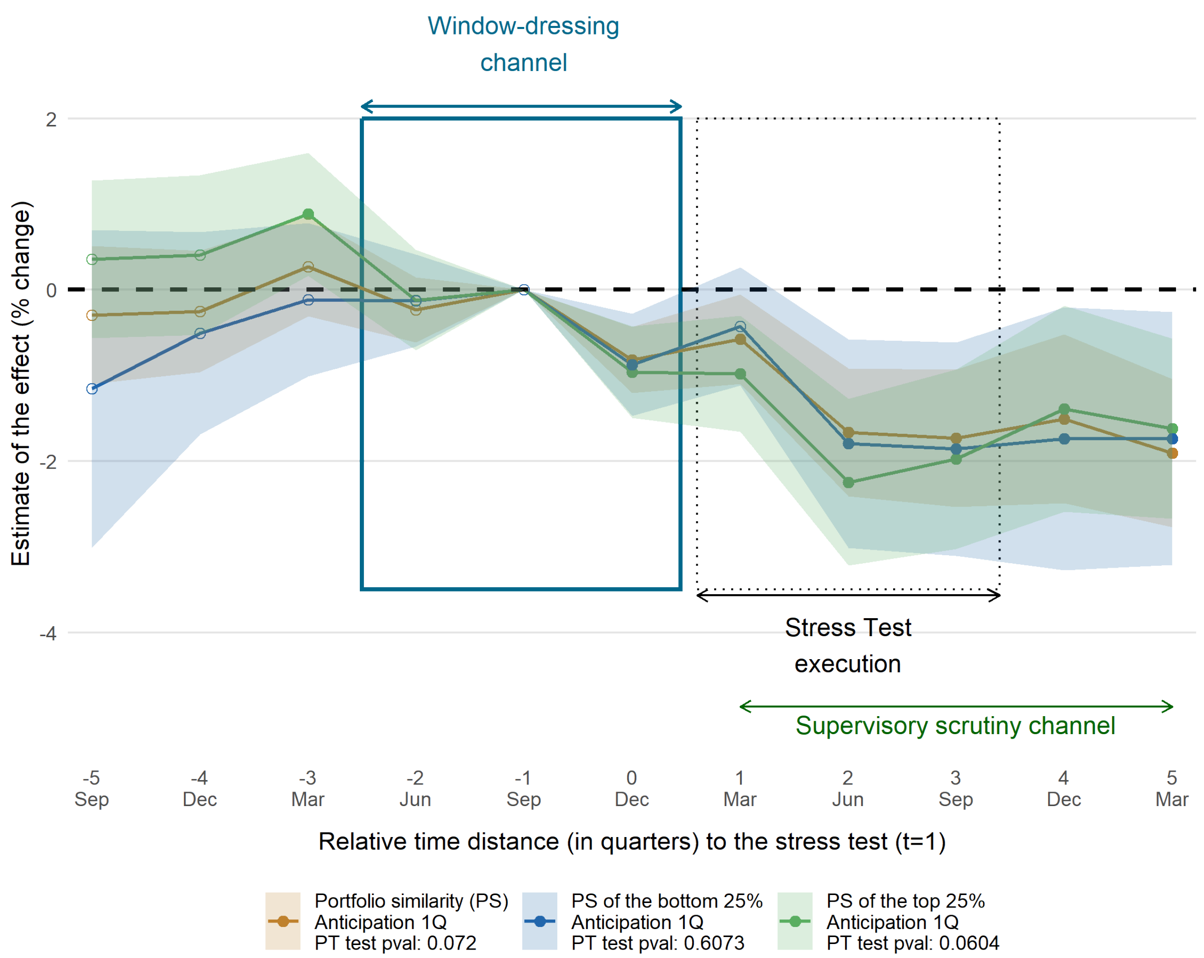
Moreover, in Table 3, we present the results obtained gauging the granularity in the definition of the banks' portofolio items. We separately test similarity relying on coarser partitions of assets, we take exposures value by types of counterparties, types of instruments, and combinations of the two categories, and at last we combine them with the categorisation of tradable vis-à-vis non-tradable assets that is additionally available in Finrep.[32] As shown in Table 3, for

[30]Formally this is obtained restricting the average in Equation 3 to banks each of the two groups.

[31]As we did for the estimates of the effects on the capital ratio, in order to strenghten our finding, we test whether the decrease in similarity is consistent across stress test exercises finding overall consistence (see Figure 10 in Appendix A).

[32]The distinction comes from the accounting rules that Finrep inherits from the IFRS 9 accounting practices. Tradable assets in IFRS accounting and Finrep are called "Held for trade", roughly equivalent to the trading

Figure 5: Stress test impact on Portfolio Similarity



Note: The plot shows dynamic treatment effects of the stress test on banks' Portfolio Similarity computed using baseline pairwise-similarity as per 1. Effects are displayed as points (filled when significant) with confidence intervals (95% level) shown as bands, expressed as percentage changes from the pre-treatment period assumed to end at time step -1. The analysis presents three series: baseline effects for the complete sample (orange), and heterogeneous effects for banks in the bottom (blue) and top (green) quartiles of the distribution of the performances in the stress test. All specifications use bank-level clustering for standard errors and assume parallel trend to hold within country and conditional on total asset, return on equity, and Tier1 capital. The timeline spans the window-dressing period and stress test execution. The results reveal two distinct phases of portfolio adjustment: an initial decline of approximately 0.9 percent of bank-specific porfolio similarity during the window-dressing period, followed by a more substantial and persistent reduction of 2 percent throughout the stress test execution phase. These patterns suggest a sequential adjustment in portfolio similarity across banks in response to the stress test exercise.

every specification of asset dimension, we observe a decrease in the portfolio similarity after the stress test exercise.

These findings challenge the conventional wisdom that stress tests might lead to increased

book of banks. All other assets are non-Held for trade and can accounted as amortised cost, fair value

Table 3: ST Effects on Similarity for all banks - Similarity based on different asset categorization

Asset Dimension	Estimate	SE	95% CI	Pre-trend p
Type of Counterparty (Cpt.)	**-0.013****	0.005	[-0.024, -0.003]	0.348
Type of Instrument (Inst.)	**-0.008****	0.003	[-0.015, -0.001]	0.706
Type of Cpt. $\times$ Type of Instr.	**-0.022****	0.006	[-0.034, -0.009]	0.148
Asset Type	**-0.021****	0.006	[-0.033, -0.009]	0.177

Notes: ** indicates significance at 95% confidence level, that is the standard confidence interval used for bootstrapped calculated standard errors. Bold estimates indicate statistical significance. SE: Standard Error. Asset Type is the combination of type of counterparty and type of instrument distinguished by tradable and non-tradable asset. The estimates show the average treatment effect on portfolio similarity over a one-year horizon following stress test inception. The estimation follows De Chaisemartin and d'Haultfoeuille (2024) and accounts for anticipation effects by using quarter -2 as the reference point for comparing pre-treatment trends. All specifications use bank-level clustering for standard errors and assume parallel trend to hold within country and conditional on total asset, return on equity, and Tier1 capital. The treatment effect captures the overall portfolio adjustment starting from September the year before the stress test, through various channels, including window-dressing, regulatory pressure, market discipline, and internal risk management adjustments. Pre-trend p-values test the parallel trends assumption in the pre-treatment period. Confidence intervals are computed using bootstrapping methods.

portfolio synchronization through the adoption of common risk management practices or "model monoculture" (Rhee and Dogra, 2024). Instead, our results suggest that stress tests may actually promote diversity in bank portfolio strategies, potentially enhancing system stability by reducing the likelihood of synchronized portfolio adjustments during stress periods.

4.2.2 Analysis of Liquidity-Weighted Portfolio Similarity

We now turn our attention to portfolio similarities weighted by asset liquidity, focusing specifically on tradable assets that could be subject to fire sales during periods of market stress. This analysis is particularly relevant as it captures potential systemic vulnerabilities arising from common exposures to assets whose liquidity typically deteriorates during market turbulence.

Figure 6 shows the effect of the stress test on liquidity-weighted portfolio similarity. We observe here a significant heterogeneity in the effects for different stress test exercises. While the 2021 exercise shows a pronounced divergence in portfolio structures when accounting for asset liquidity, with effects approximately twice as large as those observed for the overall portfolio similarity (up to 5 percent decline during the stress test execution period), the 2023 exercise exhibits a statistically non-significant evolution. The interpretation of the 2023 results, accordingly, requires caution due to the non-rejection of the existence of a trend in the pretreatment period (the pre-trend test null hypothesis), casting doubts on the plausibility of the parallel trends assumption.

The stronger effect on liquidity-weighted similarity in the 2021 exercise suggests that, at that time, banks' portfolio adjustments were particularly focused on their more liquid and tradable assets. The timing of these changes — intensifying during and after the stress test execution rather than during the window-dressing period — also indicates that these adjustments represent strategic portfolio restructuring rather than temporary window-dressing behavior. The existing literature has provided ample quantitative evidence on the risks arising from liquidity-weighted overlaps, which we exploit to define approximate measures of the gains or losses that can be imputed to this decreased synchronization around stress test exercises (see Cont and Schaanning (2019) for details). The substantial reduction in liquidity-weighted similarity observed in 2021 suggests that stress tests may contribute to reducing systemic risk by decreasing the potential for synchronized fire sales during market stress episodes, although this effect is not granted and appears to vary across different exercises.

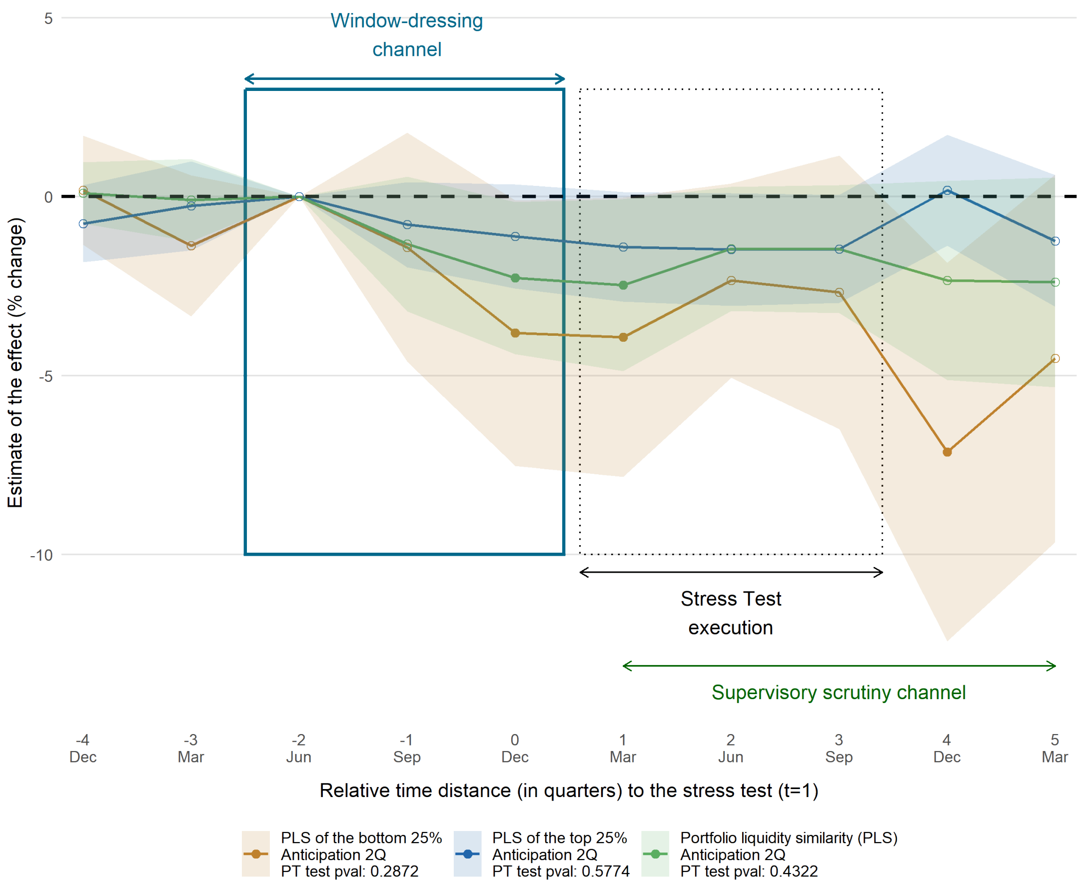
In addition, in the case of liquidity-weighted similarity, we challenge our results applying different aggregations of asset categories (see Table 4). Overall, we confirm that the liquidity similarity does not increase irrespectively of the aggregation level, i.e., across type of counterparty, instrument, or combinations of the two. Interestingly, and perhaps not surprisingly, we find that the decrease in liquidity similarity is mostly driven by the decrease in similarity across types of instruments, also when combined with the type of counterparty. The regulatory liquidity categorisation into L1-2A-2B, indeed, is assigned at instrument level, with little scope for the type of counterparty (typically, the issuer in the FINREP reporting). Therefore, the results shown in Table 4 appear plausible from liquidity similarity perspective, and support our conclusion that portfolio diversification followed the 2021 stress test exercise and was achieved through idiosynchratic reallocation of assets to different types of financial instruments.

4.2.3 Loan Portfolios Similarity

The third dimension of similarity resulting from the common behaviours of banks and potentially contributing to systemic risk concerning similarities of their loan portfolios.[33]. Specifically, we analyze common structures in banks' loan portfolios aggregated along the sectoral (NACE2) and country dimension. The analysis is relevant as it captures potential systemic vulnerabilities arising from common exposures to economic activities. Figure 7 presents our findings on

[33]Notice that loan similarity is computed as in Eq. 1, but differently from Section 4.2.1 we use here a granular categorisation of loans from AnaCredit.

Figure 6: Stress test impact on Liquidity-Weighted Similarity



Note: The plot compares dynamic treatment effects on banks' liquidity-weighted similaity cmoputed using the paiwise-similarity as defined by Eq. 2 on the most granular asset categorization available in Finrep. Effects for the bottom 25% and the top 25% of the performances in the ST are reported together with the estimates for the full sample. Confidence intervals (95% level) are shown as bands and coefficient can be interpreted as percentage changes from the pre-treatment period. Pre-trend test p-values are reported in the legend. Anticipation is accounted for assuming treatment starts at September the year of the stress test. All specifications use bank-level clustering for standard errors and assume parallel trend to hold within country and conditional on total asset, return on equity, and Tier1 capita.l

changes in sector-country overlap patterns following stress tests, using granular AnaCredit data to provide the most detailed view of similarities in banks' credit exposures towards non-financial corporations.

The results suggest heterogeneous effects across different stress test exercises, with notable differences in both the magnitude and direction of portfolio adjustments. Like in case of the liquidity similarity presented in the previous section, we observe that two quarters in anticipation to the 2021 stress test exercise, banks decreased their similarity in loans. Differently, loan

Table 4: ST Effects on Liquidity Similarity for all banks - Similarity based on different asset categorization

Asset Dimension	Estimate	SE	95% CI	Pre-trend p
Type of Counterparty (Cpt.)	0.010	0.008	[-0.006, 0.026]	0.058
Type of Instrument (Inst.)	**-0.024****	0.007	[-0.037, -0.010]	0.069
Type of Cpt. × Type of Inst.	**-0.036****	0.011	[-0.058, -0.014]	0.076

Notes: ** indicates significance at 95% confidence level, that is the standard confidence interval used for bootstrapped calculated standard errors. Bold estimates indicate statistical significance. SE: Standard Error. Asset Class is the combination of type of counterparty and type of instrument distinguished by tradable and non-tradable asset. The estimates show the average treatment effect on portfolio similarity over a one-year horizon following stress test inception. The estimation follows De Chaisemartin and d'Haultfoeuille (2024) and accounts for anticipation effects by using quarter -2 as the reference point for comparing pre-treatment trends. All specifications use bank-level clustering for standard errors and assume parallel trend to hold within country and conditional on total asset, return on equity, and Tier1 capital. The treatment effect captures the overall portfolio adjustment through various channels, including regulatory pressure, market discipline, and internal risk management adjustments. Pre-trend p-values test the parallel trends assumption in the pre-treatment period. Confidence intervals are computed using bootstrapping methods.

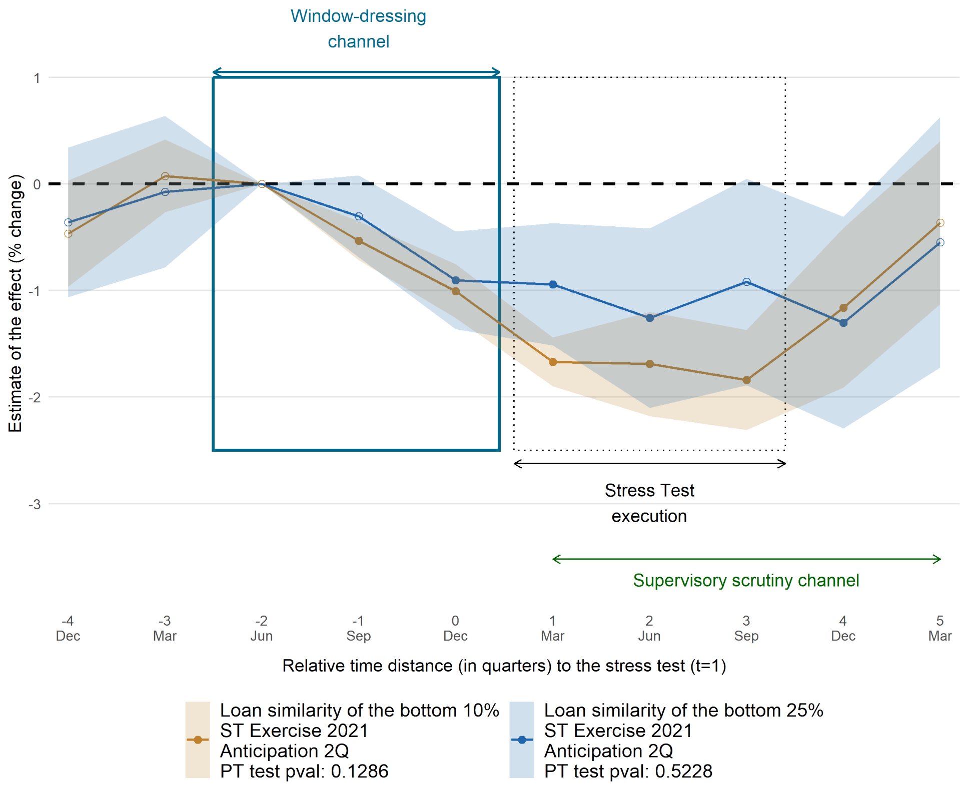
similarity seems not to vary around the 2023 exercise, altough when examining banks' behaviour three months before the 2023 stress test, we find, on the contrary, that similarity increases for a window of only one quarter before the stress test, and rises again after the stress test.

4.3 Local clusters and other structures of similarity

In Section 4.2, we have examined similarity varying the measurment of the baseline pairwise-similarity as described in Section 3.2.1, also exploring how similarity varies depending on the level of granularity and the considered asset categories. We turn now to variations in the way pairwise-similarities are aggregated to compute bank-specific similarities, examining clustering patterns and an alternative metric defined using size-dependent weights.

Our baseline similarity measure (defined in Equation 3) represents network centrality in a fully connected undirected weighted network, where banks are edges and pairwise similarities are weights. Building on this network representation, we enhance the analysis in two ways. First, we do so by focusing on specific country and business model subnetworks that might harbor localized financial risks (see Equation 5). These local risk concentrations, while potentially masked when applying aggregate centrality measures, could serve as triggers for broader systemic events propagating through the network. Specifically, we examine within-country similarity to capture potential national risk clusters and within-business-model similarity to identify model-specific vulnerabilities. This latter analysis, building on the approach of Bräuning and Fillat

Figure 7: The impact of 2021 Stress Test on Loan Portfolio Similarity



Note: The plot shows dynamic treatment effects on Loan Portfolio Similarity obtained computing benchmark pairwise-similairties on loans exposures using (1). Estimates are provided for banks at the bottom of the performance distribution for the 2021 stress test exercise. Effects are displayed as points (filled when significant) with confidence intervals (95% level) shown as bands, expressed as percentage changes from the pre-treatment period. All specifications use bank-level clustering for standard errors and assume parallel trend to hold within country and conditional on total asset, return on equity, and Tier1 capital. Potential anticipation effects are assumed to take place up to two-quarters before the start of the stress test. Pre-trend test p-values are reported in the legend.

(2024), investigates whether banks adjust their portfolios to mimic successful peers, potentially creating new forms of systemic risk through coordinated portfolio adjustments.

Next, we set edge weights to bank sizes, measured by total assets, to capture the systemic importance of specific linkages (see 4). In addition to local clusters, the weighted-aggregation method controls whether portfolio adjustments occur towards or from the largest financial institutions, allowing us to identify if banks are mimicking systemic institutions or if convergence patterns are driven by broader industry trends.

4.3.1 Country and business model clusters

While our previous analyses revealed an overall decline in system-wide similarity, examination of within-group patterns could potentially uncover local clusters where similarity increases or highlight specific drivers of the diversification effect. Table 5 presents a comprehensive analysis of portfolio convergence patterns along both geographical and business model dimensions, using various asset categorization schemes.

The results strongly confirm our main finding of a decreasing portfolio similarity, with several notable patterns emerging. Regarding the overall similarity, the strongest and most consistent evidence of decreased similarity appears within business model groups, with statistically significant reductions across multiple asset categorization schemes. Similarity especially decreases across type of counterparty and type of instrument. This suggests that a substantial portion of the diversification effect occurs among banks with similar business models, challenging the notion that banks might converge toward peers during stress tests.

The overall similarity appears, on the contrary, to be independent from country specificities. When examining different asset categorisations across country, indeed we find no statistically significant features. We conclude that idiosynchratic balance sheet adjustments before the stress test do not trigger a country-specific fallback of banks in the euro area.

For liquidity-weighted similarity, we find negative significant effects on similarity particularly for the similarity in exposures to the same instruments within business models and for combined counterparty-instrument measures within countries. This aligns with our previous findings while suggesting that the intensity of portfolio adjustments may vary across different dimensions of similarity.

In the loan portfolio analysis, while the coefficients are predominantly negative, they lack statistical significance across most specifications. This consistency with our benchmark exercise suggests that credit portfolio adjustments, while directionally aligned with the overall trend toward differentiation, may be more constrained or gradual in nature.

These additional analyses provide large-scale evidence that the decrease in similarity is robust to varying definitions of asset categorization, with particularly strong effects observed within business model groupings. The results reinforce our main conclusion that stress tests appear to encourage portfolio differentiation rather than convergence, even when examining more localized network structures.

Table 5: ST Effects on Within Cluster Similarity for all banks

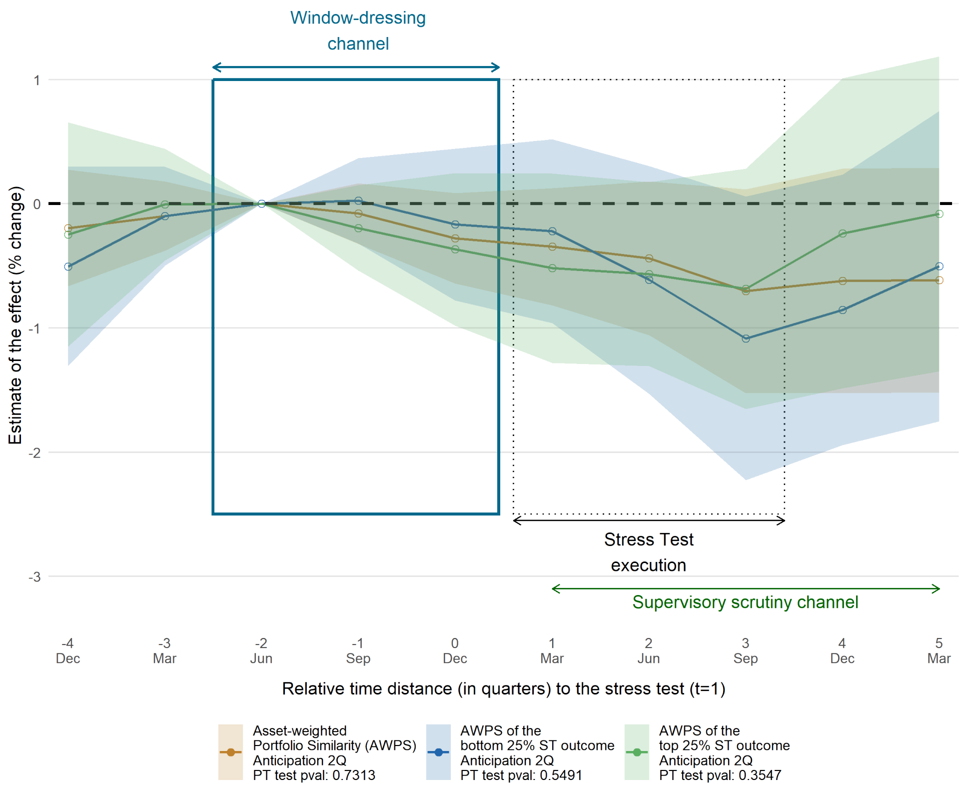
Asset Dimension	Estimate	95% CI	Pre-trend p
1. Overall Similarity			
Within Business Model Similarity			
Type of Counterparty	**-0.013****	[-0.023, -0.004]	0.276
Type of Counterparty × Type of Instrument	**-0.022****	[-0.033, -0.011]	0.396
Type of Instrument	-0.006	[-0.014, 0.001]	0.978
Country	-0.012	[-0.027, 0.003]	0.380
Asset Class	**-0.021****	[-0.032, -0.009]	0.416
Within Country Similarity			
Type of Counterparty	-0.008	[-0.025, 0.009]	0.552
Type of Counterparty× Type of Instrument	-0.014	[-0.033, 0.004]	0.664
Type of Instrument	-0.005	[-0.012, 0.003]	0.000
Country	-0.012	[-0.026, 0.002]	0.448
Asset Class	-0.013	[-0.031, 0.005]	0.601
2. Liquidity-Weighted Similarity			
Within Business Model Similarity			
Type of Counterparty	0.016	[-0.003, 0.035]	0.025
Type of Counterparty× Type of Instrument	-0.029	[-0.062, 0.005]	0.715
Type of Instrument	**-0.024****	[-0.041, -0.007]	0.815
Country	0.004	[-0.016, 0.025]	0.131
Within Country Similarity			
Type of Counterparty	0.009	[-0.015, 0.032]	0.227
Type of Counterparty× Type of Instrument	**-0.044****	[-0.076, -0.011]	0.001
Type of Instrument	0.003	[-0.014, 0.020]	0.587
Country	-0.007	[-0.033, 0.019]	0.172
3. Loan Similarity			
Within Business Model Similarity			
Credit Quality Step	-0.025	[-0.059, 0.008]	0.000
Credit Quality Step × Maturity	-0.006	[-0.044, 0.032]	0.227
Maturity	0.011	[-0.013, 0.035]	0.178
Country	-0.008	[-0.021, 0.005]	0.000
Sector (NACE2)	-0.003	[-0.019, 0.013]	0.047
Within Country Similarity			
Credit Quality Step	-0.016	[-0.041, 0.008]	0.000
Credit Quality Step × Maturity	-0.001	[-0.035, 0.033]	0.159
Maturity	0.004	[-0.018, 0.025]	0.064
Country	-0.013	[-0.038, 0.011]	0.578
Sector (NACE2)	-0.008	[-0.043, 0.026]	0.001

Notes: ** indicates significance at 95% confidence level, that is the standard confidence interval used for bootstrapped calculated standard errors. Bold estimates indicate statistical significance. SE: Standard Error. Asset Class is the combination of type of counterparty and type of instrument distinguished by tradable and non-tradable asset. The estimates show the average treatment effect on portfolio similarity over a one-year horizon following stress test inception. The estimation follows De Chaisemartin and d'Haultfoeuille (2024) and accounts for anticipation effects by using quarter -2 as the reference point for comparing pre-treatment trends. All specifications use bank-level clustering for standard errors and assume parallel trend to hold within country and conditional on total asset, return on equity, and Tier1 capital. The treatment effect captures the overall portfolio adjustment through various channels, including window-dressing, regulatory pressure, market discipline, and internal risk management adjustments. Pre-trend p-values test the parallel trends assumption in the pre-treatment period. Confidence intervals are computed using bootstrapping methods.

4.3.2 Size-Weighted Similarity Analysis

To deepen our understanding of portfolio synchronization patterns, we examine one last dimension of similarity to testconvergence toward larger banks via size-weighted similarity (4), using weights derived from the value of total assets. This analysis helps determine whether the observed patterns in overall similarity remain consistent when accounting for bank size. In particular, we want to test whether smaller-sized and possibly weaker performing banks tend to converge to bigger banks, which potentially also have more knowledge and resources for performing well in a stress test exercise.

Figure 8: Stress test impact on Size-Weighted Similarity



Note: The plot shows dynamic treatment effects of the stress test on size-weighted portfolio similarity computed using benchmark pairwise-similarity as per Eq. 1 and then aggregated at the bank-level using weights defined using total assets via Eq. 4. Samples include the bottom and top 25%, as well as all the ST banks), effects are displayed as points (filled when significant) with confidence intervals (95% level) shown as bands, expressed as percentage changes from the pre-treatment period. All specifications use bank-level clustering for standard errors and assume parallel trend to hold within country and conditional on total asset, return on equity, and Tier1 capital, Anticipation is accounted for assuming treatment start its effectstwo-quarter in advance of the stress test launch. Pre-trend test p-values are reported in the legend.

The asset-weighted similarity analysis in Figure 8 reveals no significant convergence in similarity toward bigger banks, with a pattern broadly consistent with our previous findings. During the window-dressing period, we observe a relatively stable relationships, with no significant shifts in similarity patterns. The results suggest that during the stress test execution phase, asset-weighted similarity should be reduced, although the effects are not statistically significant. In other words, when accounting for bank size, stress tests do not induce convergence toward larger institutions' portfolio structures.

Importantly, these findings reinforce our main results by showing that even when considering bank size, we do not observe any evidence of increased portfolio synchronization. Instead, the directional effect continues to point toward differentiation, albeit with more muted magnitudes compared to our baseline specifications.

5 Conclusions

We assess the impact of the 2021 and 2023 EU-wide stress tests on banks' risk-taking behavior and systemic risk using confidential supervisory data from ECB Banking Supervision. Given the growing importance of stress tests in the financial system and their integration into internal risk management systems and prudential supervision, banks can be expected to incorporate stress test considerations into their strategic decisions about balance sheet composition. This could manifest either through ex ante positioning to mitigate the impact of forthcoming stress scenarios or through ex post adjustments following the interaction with supervisors or publication of the stress test results. These adjustments, even though aiming at derisking indivual banks' positions, could potentially create negative externalities, particularly through the synchronization of portfolio exposures, fuelling systemic risk.

Through a difference-in-difference econometric approach, we examine both individual bank risk and portfolio synchronization effects. Our analysis reveals three key findings.

First, we document significant derisking behavior primarily occurring through ex-ante window-dressing rather than ex post adjustments, with banks managing risk-weighted assets rather than capital levels. This effect is particularly pronounced among poor-performing institutions in the stress tests.

Second, we find that those risk-mitigating action by poor-performing banks lead to an aggregate decline in portfolio similarity, contributing positively to systemic risk reduction. These management actions by vulnerable banks prove beneficial for financial stability also through the lens of liquidity risk, reducing the risk of shock amplification through coordinated balance sheet adjustments or fire sales. Importantly, we find no evidence that supervisory follow-up creates conditions for banks to converge in the composition of their balance sheets.

Third, we document how the decrease in similaity is robust within countries and similar business models. This pattern supports a certain homogeneity of EU financial systems, with a more limited role for jurisdictional specificities in influencing exposure profiles and risk management strategies.

From a policy standpoint, our results of weak negative externalities suggest that the general set-up of the supervisory stress tests does not result into prescribed strategies that would allow banks to game the exercises. It is *per se* a reassuring finding that the design of the exercises diminishes systemic risk that might arise through portoflio similarity. However, the behaviour of banks with weaker fundamentals should be monitored, since these banks might tend to react more strongly – either to the mere fact of having looming supervisory stress test on the horizon or to a publication of some negative results – increasing idiosynchratic risk.

References

Abadie, A. (2005). Semiparametric difference-in-differences estimators. *Review of Economic Studies*, 72(1):1–19.

Abadie, A. and Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, 74(1):235–267.

Abbassi, P., Brownlees, C., Hans, C., and Podlich, N. (2017). Market risk measures and portfolio similarity. *Journal of Financial Economics*, 125(2):384–406.

Acharya, V. V., Berger, A. N., and Roman, R. A. (2018). Stress tests and small business lending. *Journal of Financial Economics*, 129(1):136–157.

Adrian, T. and Brunnermeier, M. K. (2016). Covar. *American Economic Review*, 106(7):1705–41.

Allen, F., Babus, A., and Carletti, E. (2012). Financial connections and systemic risk. *Management Science*, 58(10):1831–1883.

Anna-Lena Högenauer, D. H. and Quaglia, L. (2023). Introduction to the special issue: the persistent challenges to european banking union. *Journal of European Integration*, 45(1):1–14.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., and Wager, S. (2021). Synthetic difference-in-differences. *American Economic Review*, 111(12):4088–4118.

Baudino, P., Goetschmann, R., Henry, J., Taniguchi, K., and Zhu, W. (2018). *Stress-testing banks: a comparative analysis*. Bank for International Settlements, Financial Stability Institute.

Blackman, A., Pfaff, A., and Robalino, J. (2018). Strict versus mixed-use protected areas: Guatemala's maya biosphere reserve. *Ecological Economics*, 144:88–99.

Borusyak, K., Jaravel, X., and Spiess, J. (2024). Revisiting event-study designs: robust and efficient estimation. *Review of Economic Studies*, page rdae007.

Bräuning, F. and Fillat, J. L. (2024). The impact of regulatory stress tests on banks' portfolio similarity and implications for systemic risk. *Journal of Money, Credit and Banking*. Forthcoming.

Brunnermeier, M. K., Feng, G., and Palia, D. (2020). Banks' risk taking and creditors' bargaining power. *Review of Financial Studies*, 33(1):88–121.

Brunnermeier, M. K. and Pedersen, L. H. (2009). Market liquidity and funding liquidity. *Review of Financial Studies*, 22(6):2201–2238.

Caccioli, F., Shrestha, M., Moore, C., and Farmer, J. D. (2014). Stability analysis of financial contagion due to overlapping portfolios. *Journal of Banking & Finance*, 46:233–245.

Cai, F., Han, S., Li, D., and Li, Y. (2018). Institutional herding and its price impact: Evidence from the corporate bond market. *Journal of Financial Economics*, 127(1):91–108.

Cai, M., Correa, R., Gao, P., and Rosa, C. (2022). The effect of regulatory stress tests on bank lending and its macroeconomic consequences. *Journal of Financial Economics*, 146(2):467–487.

Cappelletti, G., Marques, A. P., and Varraso, P. (2024). Impact of higher capital buffers on banks' lending and risk-taking in the short-and medium-term: Evidence from the euro area experiments. *Journal of Financial Stability*, 72:101250.

Carboni, M., Fiordelisi, F., Ricci, O., and Lopes, F. S. S. (2017a). Surprised or not surprised? the investors' reaction to the comprehensive assessment preceding the launch of the banking union. *Journal of Banking & Finance*, 74:122–132.

Carboni, M., Fiordelisi, F., Ricci, O., and Stentella Lopes, F. S. (2017b). The effect of bank stress tests on bank lending. *Journal of Financial Stability*, 33:61–80.

Cont, R. and Schaanning, E. (2019). Monitoring indirect contagion. *Journal of Banking & Finance*, 104:85–102.

Cornett, M. M., Minnick, K., Schorno, P. J., and Tehranian, H. (2020). An examination of bank behavior around federal reserve stress tests. *Journal of Financial Intermediation*, 41:100789.

Cortés, K. R., 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.

Couaillier, C. and Henricot, D. (2023). How do markets react to tighter bank capital requirements? *Journal of Banking and Finance*, 148:106723.

Curry, T. J., Fissel, G. S., and Ramirez, C. D. (2008). The impact of bank supervision on loan growth. *North American Journal of Economics and Finance*, 19(2):192–211.

Cuzzola, A., Barbieri, C., and Bindseil, U. (2023). Stress testing with multi-faceted liquidity: the central bank collateral framework as a financial stability tool.

De Chaisemartin, C. and d'Haultfoeuille, X. (2024). Difference-in-differences estimators of intertemporal treatment effects. *Review of Economics and Statistics*, pages 1–45.

Durrani, A., Ongena, S., and Marques, A. P. (2024). Decoding market reactions: The certification role of eu-wide stress tests. *Economic Modelling*, 139:106828.

Durrani, A., Ongena, S., and Ponte Marques, A. (2022a). The certification role of the eu-wide stress testing exercises. *ECB Working Paper*, (2711).

Durrani, A., Ongena, S., and Ponte Marques, A. (2022b). The certification role of the EU-wide stress testing exercises in the stock market. what can we learn from the stress tests (2014-2021)? Working Paper Series 2711, European Central Bank.

Durrani, A., Ponte Marques, A., Giraldo, G., Pancaro, C., Panos, J., and Zaharia, A. (2023). Does the disclosure of stress test results affect market behaviour? *ECB Macroprudential Bulletin*, 21.

ECB (2014). Aggregate report on the comprehensive assessment. *Technical Report, European Central Bank*.

Elliott, M., Georg, C.-P., and Hazell, J. (2021). Systemic risk shifting in financial networks. *Journal of Economic Theory*, 191:105157.

European Banking Authority (2021). Eba confirms eu banks' solid overall liquidity position but warns about low foreign currency liquidity buffers. Technical report, European Banking Authority.

European Banking Authority (2023). EU banks' liquidity coverage ratio declined but remains well above the minimum requirement. Technical report, European Banking Authority.

European Central Bank (2024). Supervisory manual. Banking supervision guide, European Central Bank, Frankfurt am Main. Accessed: January 2024.

Farhi, E. and Tirole, J. (2012). Collective moral hazard, maturity mismatch, and systemic bailouts. *American Economic Review*, 102(1):60–93.

Gandhi, P. and Purnanandam, A. (2023). Do bank stress tests reduce risk-taking? *Journal of Financial Economics*, 147(2):284–302.

Girardi, G., Hanley, K. W., Nikolova, S., Pelizzon, L., and Sherman, M. G. (2021). Portfolio similarity and asset liquidation in the insurance industry. *Journal of Financial Economics*, 142(1):69–96.

Goldstein, I. and Leitner, Y. (2020). Banking stability and regulation: What role for market discipline? *Journal of Financial Intermediation*, 44:100847.

Goldstein, I. and Sapra, H. (2014). Should banks' stress test results be disclosed? an analysis of the costs and benefits. *Foundations and Trends in Finance*, 8(1):1–54.

Greenwood, R., Landier, A., and Thesmar, D. (2015). Vulnerable banks. *Journal of Financial Economics*, 115(3):471–485.

Gropp, R., Mosk, T., Ongena, S., and Wix, C. (2018). Banks response to higher capital requirements: Evidence from a quasi-natural experiment. *The Review of Financial Studies*, 31(1):266–299.

Guindos, L. d. (2019). The evolution of stress-testing in Europe, Keynote speech by Luis de Guindos, Vice-President of the ECB, at the annual US-EU Symposium organised by the Program on International Financial Systems, Frankfurt.

Hancock, D., Laing, A. J., and Wilcox, J. A. (1995). Bank capital shocks: dynamic effects on securities, loans, and capital. *Journal of Banking & Finance*, 19(3-4):661–677.

Heckman, J. J., Ichimura, H., and Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies*, 64(4):605–654.

Hirtle, B., Kovner, A., and Plosser, M. (2020). The impact of supervision on bank performance. *The Journal of Finance*, 75(5):2765–2808.

Hirtle, B. and Lehnert, A. (2015). Supervisory stress tests. *Annual Review of Financial Economics*, 7:339–355.

Imbens, G. W. and Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1):5–86.

Jabbour, R. S. and Sridharan, N. (2020). The ECB's in-comprehensive SSM-ent: the higher they go, the harder they fall. *The European Journal of Finance*, 26(15):1506–1528.

Jiménez, G., Ongena, S., Peydró, J.-L., and Saurina, J. (2017). Macroprudential policy, countercyclical bank capital buffers, and credit supply: Evidence from the spanish dynamic provisioning experiments. *Journal of Political Economy*, 125(6):2126–2177.

Kapinos, P., Martin, C., and Mitnik, O. (2018). Stress testing banks: Whence and whither? *Journal of Financial Perspectives*, 6(3):68–87.

King, G. and Nielsen, R. (2019). Why propensity scores should not be used for matching. *Political Analysis*, 27(4):435–454.

Kok, C., Müller, C., Ongena, S., and Pancaro, C. (2023a). The disciplining effect of supervisory scrutiny in the eu-wide stress test. *Journal of Financial Intermediation*, 53:100971.

Kok, C., Müller, C., Ongena, S., and Pancaro, C. (2023b). The disciplining effect of supervisory scrutiny in the eu-wide stress test. *Journal of Financial Intermediation*, 53:100971.

Konietschke, P., Ongena, S., and Ponte Marques, A. (2022). Stress tests and capital requirement disclosures: do they impact banks' lending and risk-taking decisions? *Swiss Finance Institute Research Paper*, (60).

Kupiec, P., Lee, Y., and Rosenfeld, C. (2017). Does bank supervision impact bank loan growth? *Journal of Financial Stability*, 28:29–48.

Loipersberger, F. (2017). The effect of supranational banking supervision on the financial sector: Event study evidence from europe. *Munich Discussion Paper*, (2017-5).

Malani, A. and Reif, J. (2015). Interpreting pre-trends as anticipation: Impact on estimated treatment effects from tort reform. *Journal of Public Economics*, 124:1–17.

McCaul, E. (2021). The evolution of stress testing in banking supervision. Joint EIB-ESM-SRB Conference "Bank Resolution and the Common Backstop for the Single Resolution Fund: A new decisive step to completing the Banking Union".

Orlov, D., Zryumov, P., and Skrzypacz, A. (2023). The design of macroprudential stress tests. *The Review of Financial Studies*, 36(11):4460–4501.

Peek, J. and Rosengren, E. S. (2000). Collateral damage: Effects of the japanese bank crisis on real activity in the united states. *American Economic Review*, 90(1):30–45.

Quagliariello, M. (2019). Are stress tests beauty contests? *EBA Staff Papers*, 4.

Rhee, K. and Dogra, K. (2024). Stress tests and model monoculture. *Journal of Financial Economics*, 152:103760.

Ripollone, J. E., Huybrechts, K. F., Rothman, K. J., Ferguson, R. E., and Franklin, J. M. (2018). A comparison of distance metrics for balancing covariates: A simulation study and application using a new england medical claims database. *American Journal of Epidemiology*, 187(7):1451–1458.

Roth, J. (2022). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights*, 4(3):305–322.

Rubin, D. B. (1973). Matching to remove bias in observational studies. *Biometrics*, pages 159–183.

Sahin, C. and de Haan, J. (2020). Bank stress testing: A stochastic simulation framework to assess banks' financial fragility. *Journal of Banking & Finance*, 120:105818.

Sahin, C., de Haan, J., and Neretina, E. (2020). Banking stress test effects on returns and risks. *Journal of Banking & Finance*, 117:105843.

Schuermann, T. (2014). Stress testing banks. *International Journal of Forecasting*, 30(3):717–728.

Shleifer, A. and Vishny, R. (2011). Fire sales in finance and macroeconomics. *Journal of economic perspectives*, 25(1):29–48.

A Heterogeneity analysis: effects of different stress test exercises

As a key robustness check for our main findings, we examine whether the anticipatory behaviors identified in the pooled sample are consistent across different stress test exercises. While our main analysis uses a stacked dataset to maximize statistical power, this approach could potentially mask heterogeneity between the 2021 and 2023 stress tests.

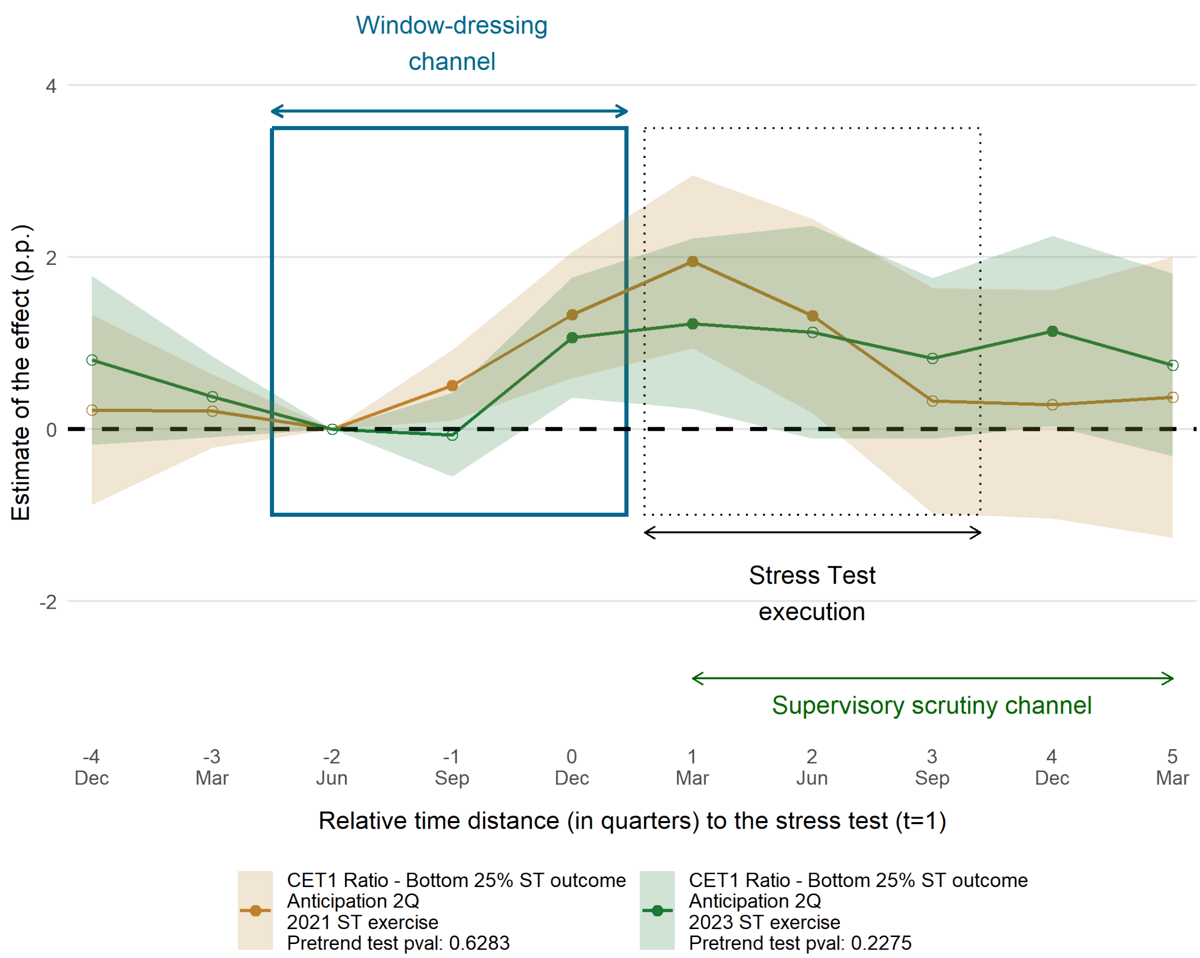
To address this concern, we separate our sample and conduct distinct estimations for the 79 banks participating in the 2021 stress test and the 95 banks participating in the 2023 stress test (noting that some institutions participated in both exercises). For each subsample, we implement our difference-in-differences framework using exercise-specific control groups assigned through our matching procedure. This approach allows us to determine whether our findings are driven primarily by one stress test or whether they represent a consistent pattern across exercises.

CET1 Ratio Figure 9 presents the comparison of CET1 ratio effects across the two stress tests. The results confirm that the anticipatory patterns identified in our main analysis are present in both the 2021 and 2023 exercises, though with some notable differences in magnitude and timing. The 2021 stress test shows slightly larger effects that persist longer throughout the stress test execution phase. In contrast, the 2023 stress test exhibits a more pronounced anticipatory response concentrated in the quarter immediately preceding the exercise, followed by a quicker reversal.

These differences may reflect evolving bank strategies, variations in stress scenario severity, or changes in the regulatory environment between the two exercises. Despite these differences, the fundamental pattern of pre-emptive capital adjustments remains consistent across both stress tests. This consistency strengthens our conclusion that the window-dressing behavior documented in the main analysis represents a systematic response to stress testing rather than an artifact of a specific exercise.

The comparable magnitude of effects in both stress tests also suggests that banks' anticipatory behaviors have become an established feature of their strategic response to regulatory stress testing, with institutions consistently adjusting their capital positions in advance of these exercises. This finding has important implications for the design and implementation of future stress tests, as it indicates that the starting point data used in these exercises may be systematically

Figure 9: CET1 Ratio Effects: 2021 versus 2023 Stress Tests



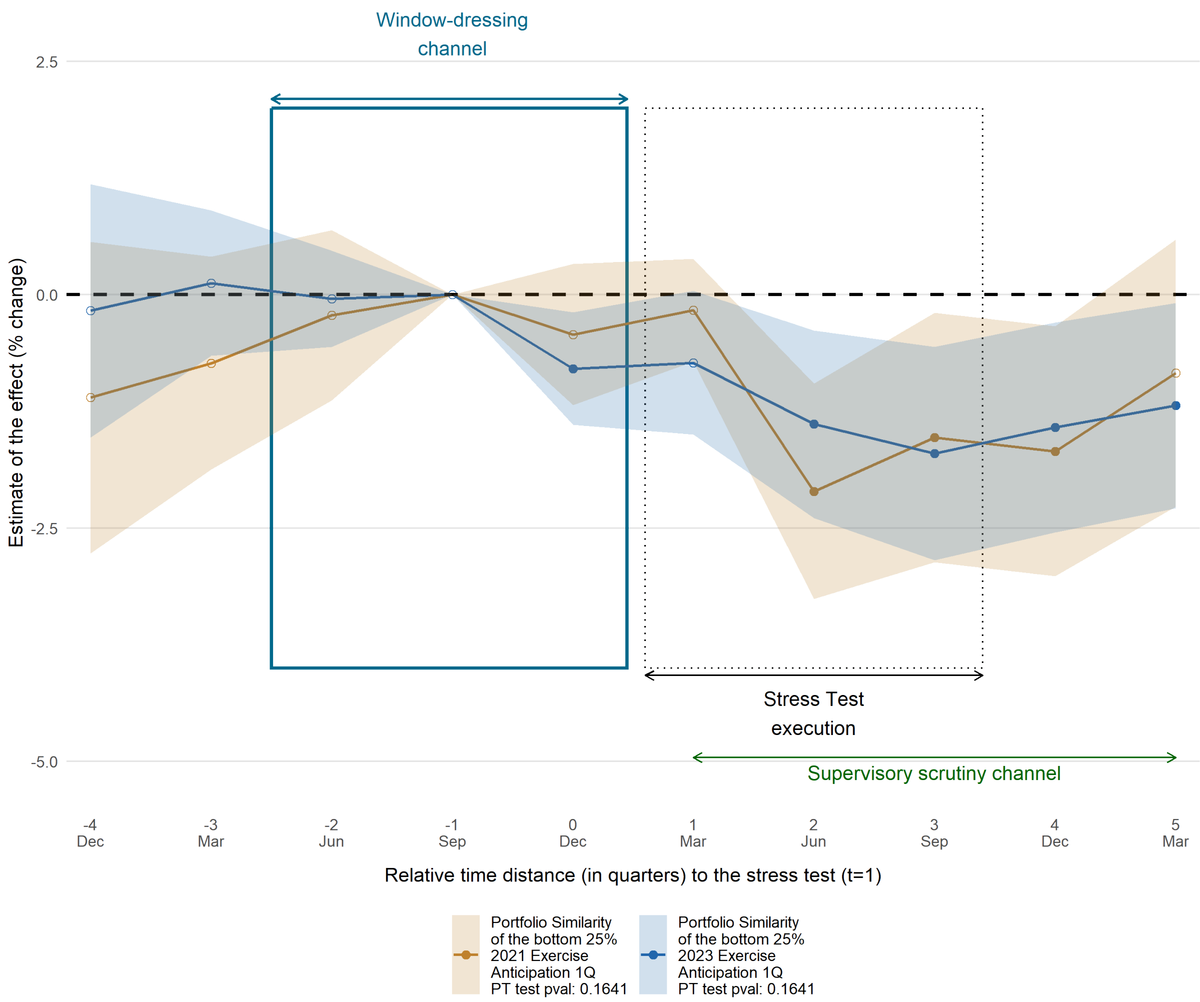
Note: The plot shows dynamic treatment effects estimated using De Chaisemartin and d'Haultfoeuille (2024) for banks in the bottom 25% of stress test outcomes, calculated separately for the 2021 and 2023 exercises. Effects are displayed as points (filled when significant) with confidence intervals (95% level) shown as bands. All specifications use entity-level clustering for standard errors, and include country-specific non-parametric trends controlling for total assets and ROE. The timeline is expressed in quarters relative to stress test initiation (t=1). The estimation accounts for anticipation effects, with the anticipation window optimized to maximize pre-trend p-values (2 quarters both for 2021 and 2023). Pre-trend test p-values are reported in the legend. Results indicate that anticipatory behavior is consistent across exercises.

influenced by banks' strategic behaviors.

Portfolio similarity Figure 10 additionally shows a split sample analysis to verify whether the effects may vary for different stress test exercises. The results provide evidence of a remarkable consistency across stress test vintages, with both the 2021 and 2023 exercises showing portfolio similarity declining by approximately 2%. While the timing of effects varies slightly — with the

2023 exercise showing a stronger initial decline during window-dressing and the 2021 exercise exhibiting more pronounced effects during execution — the overall magnitude and direction of the effects remain stable. This consistency across different stress test exercises strengthens our confidence in the robustness of the main findings.

Figure 10: Comparison of Portfolio Similarity Effects: 2021 vs 2023 Stress Tests



Note: The plot compares dynamic treatment effects on portfolio similarity between the 2021 and 2023 stress test exercises, estimated via difference-in-differences. Effects are displayed as points (filled when significant) with confidence intervals (95% level) shown as bands, expressed as percentage changes from the pre-treatment period assumed to end at time step -1. The analysis presents three series: baseline effects for the complete sample (orange), and heterogeneous effects for banks in the bottom (blue) and top (green) quartiles. All specifications use entity-level clustering for standard errors, include country-specific non-parametric trends, and control for total assets, RoE and Tier1 capital. The timeline spans the window-dressing period and stress test execution for both exercises. Both stress tests show similar overall effects, with portfolio similarity declining by approximately 2% (slightly more pronounced for 2021 by the end of the period). The timing of the effects differs somewhat: the 2023 exercise shows a more significant initial decline during the window-dressing phase, while the 2021 exercise exhibits a stronger reduction during the stress test execution period.

Loan similarity. In contrast to the consistent patterns observed for CET1 ratio and portfolio similarity, our analysis of loan similarity reveals more heterogeneous effects across stress test vintages. Figure 7 in the main text shows significant negative effects for the 2021, while the estimation with the stacked dataset including both the exercise and the esitmates for 2023 are inconclusive (see Figure 11). Limited significance appears to be due to the lack of homogeneity in trends before and after the stress tests across vintages. The pre-trend test for the 2023 exercise indicates a potential violation of the parallel trends assumption, which likely reflects in the stacked analysis despite the clear effects observed in the 2021 exercise.

This mixed pattern is further corroborated by comparing the results across different dimensions of similarity (Table 6). Despite analyzing assets in AnaCredit across various dimensions, we find a statistically significant decrease in similarity only across countries, yet with pvalue of the pretrend test below 0.0001. For other asset dimensions, such as different credit quality steps, maturity, and NACE2 sector, we do not obtain statistically significant results.

Table 6: ST Effects on Loan Similarity for all banks - Similarity based on different asset categorization

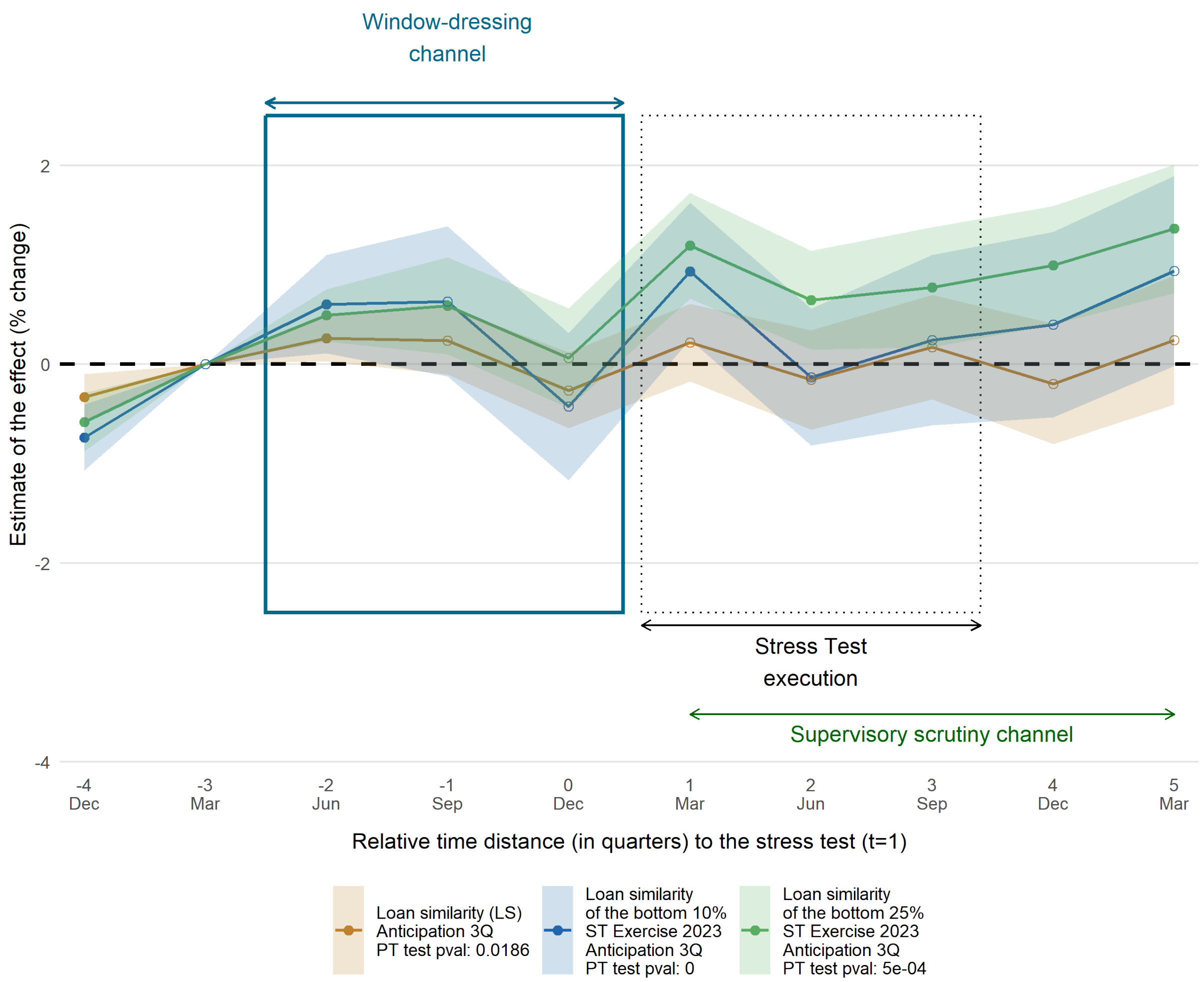
Asset Dimension	Estimate	SE	95% CI	Pre-trend p
Credit Quality Step	-0.028	0.016	[-0.060, 0.005]	0.000
Credit Quality Step × Maturity	0.004	0.017	[-0.030, 0.037]	0.872
Maturity	0.017	0.009	[-0.000, 0.035]	0.042
Country	**-0.009****	0.003	[-0.015, -0.003]	0.000
Sector (NACE2)	0.011	0.011	[-0.011, 0.033]	0.037

Notes: ** indicates significance at 95% confidence level, that is the standard confidence interval used for bootstrapped calculated standard errors. Bold estimates indicate statistical significance. SE: Standard Error. Asset Class is the combination of type of counterparty and type of instrument distinguished by tradable and non-tradable asset. The estimates show the average treatment effect on portfolio similarity over a one-year horizon following stress test inception. The estimation follows De Chaisemartin and d'Haultfoeuille (2024) and accounts for anticipation effects by using quarter -2 as the reference point for comparing pre-treatment trends. All specifications control for total assets and their growth rates in the pre-treatment period. The treatment effect captures the overall portfolio adjustment through various channels, including regulatory pressure, market discipline, and internal risk management adjustments. Pre-trend p-values test the parallel trends assumption in the pre-treatment period. Confidence intervals are computed using bootstrapping methods.

B Results on bank fundamentals around stress tests

This appendix complements our main analysis by examining the dynamics of various bank fundamentals around stress tests. We investigate how the increase in capital ratios documented in the main text is achieved, specifically focusing on the decomposition of CET1 ratio effects

Figure 11: The Impact of Stress Test on Loan Portfolio Similarity



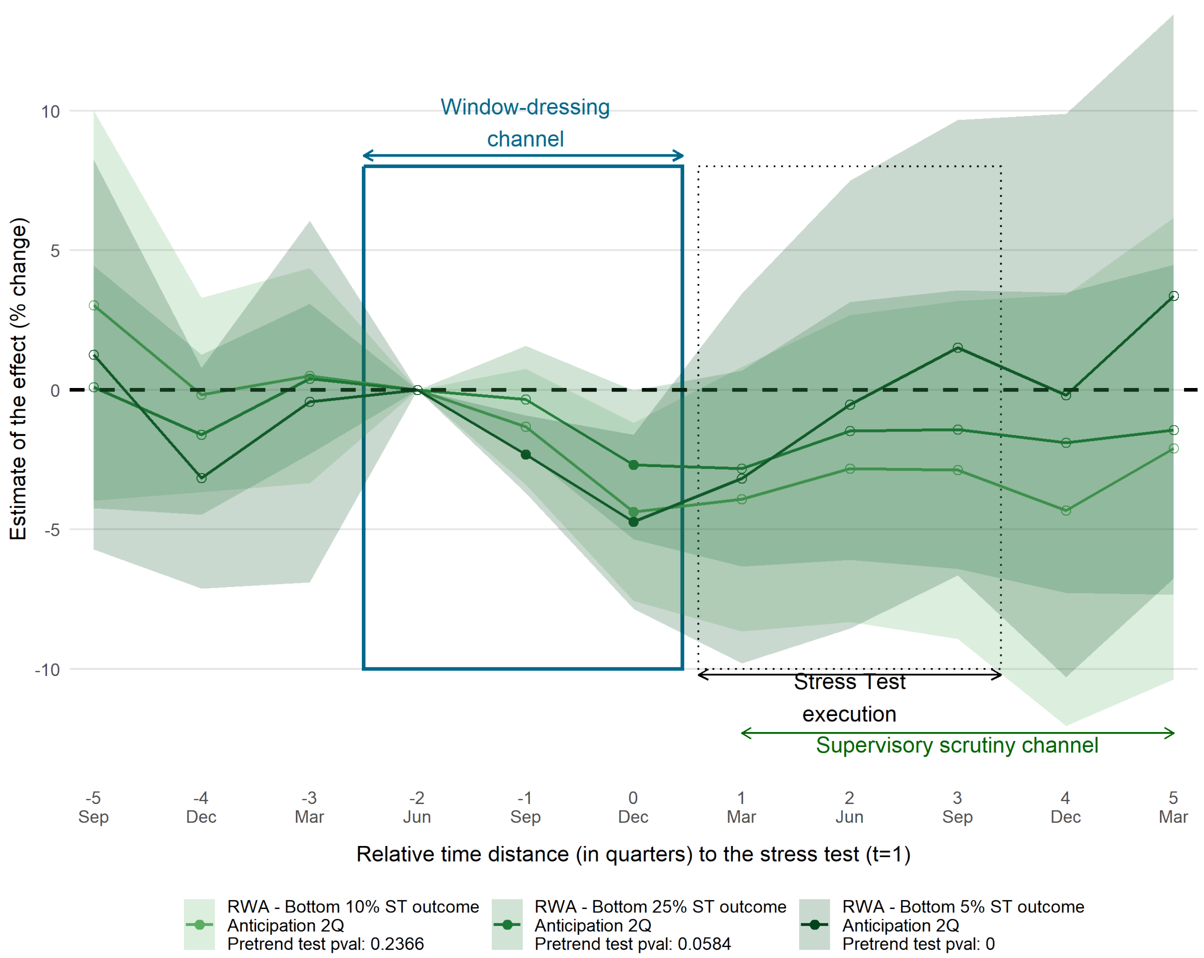
Note: The plot compares dynamic treatment effects on Loan Portfolio Similarity, estimated via difference-in-differences. Effects are displayed as points (filled when significant) with confidence intervals (95% level) shown as bands, expressed as percentage changes from the pre-treatment period. The analysis presents baseline effects for both exercises and heterogeneous effects for banks in the top and bottom quartiles. All specifications use entity-level clustering for standard errors, include country-specific non-parametric trends, and account for one-quarter anticipation effects. Pre-trend test p-values are reported in the legend.

into changes in risk-weighted assets and capital levels. We also explore related effects on bank liquidity and profitability to provide a comprehensive understanding of how banks adjust their balance sheets in anticipation of stress tests.

B.1 Decomposition of the effects on CET1 ratio

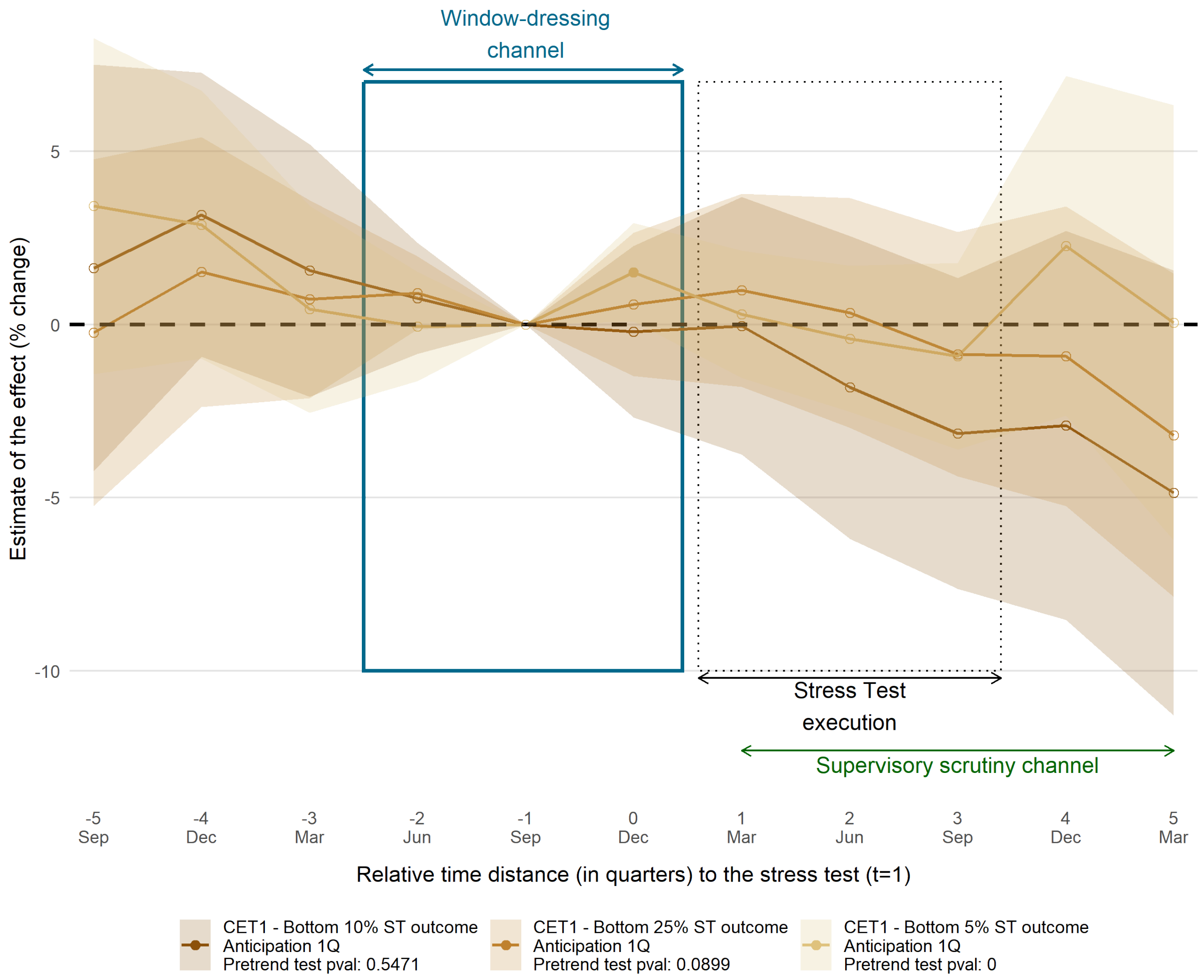
Having established that banks in the bottom quartile of stress test performance engage in anticipatory capital management, we examine the components of capital ratios separately to uncover the drivers of the underlying adjustment mechanisms.

Figure 12: Stress test effects on Risk-Weighted Assets



Note: The plot shows dynamic treatment effects estimated using De Chaisemartin and d'Haultfoeuille (2024) for the logarithm of Risk-Weighted Assets. Effects are displayed as points (filled when significant) with confidence intervals (95% level) shown as bands. All specifications use bank-level clustering for standard errors, and include country-specific non-parametric trends and CET1 capital as conditional parallel trend control. The timeline is expressed in quarters relative to stress test initiation (t=1). The estimation accounts for anticipation effects by using quarter -2 as the reference point for comparing pre-treatment trends. The coefficient can be interpreted as percentage with respect to benchmark period (-2). Results indicate that during the window-dressing period, a significant reduction in Risk-Weighted Assets ranging between the 5 and 10% relative to a counterfactual without stress test, suggesting substantial portfolio derisking in anticipation of the stress test.

Figure 13: Stress test effects on CET 1 Capital



Note: The plot shows dynamic treatment effects estimated using De Chaisemartin and d'Haultfoeuille (2024) for CET1 Capital, comparing banks in the bottom and top performance groups. Effects are displayed as points (filled when significant) with confidence intervals (95% level) shown as bands, expressed as percentage changes from the pre-treatment period. All specifications use entity-level clustering for standard errors, include country-specific non-parametric trends, and control for total assets. The timeline is expressed in quarters relative to stress test initiation (t=1). The estimation accounts for anticipation effects by using quarter -1 as a reference point for comparing pre-treatment trends. The results suggest a decline of approximately 5% of Tier 1 capital during the window-dressing period, though these effects are not statistically significant.

We find that adjustments to the CET1 ratio occur through a contraction of Risk-Weighted Assets (RWA) (Figure 12) rather than capital increases. During the window-dressing period, treated banks reduce their RWAs by 5%. This substantial decrease in risk-weighted assets is evident primarily in the two quarters preceding the stress test. Also, consistently with our previous findings, the lower the stress test outcome of the bank, the more extensive are reductions of the RWA.

Capital, differently, does not significantly contribute to the window-dressing behaviour (Figure 13). Changes in the absolute level of CET1 capital are modest and statistically insignificant during the window-dressing period. The point estimates suggest a slight decline in capital levels, though confidence intervals are wide. This asymmetric response — i.e., large RWA reductions coupled with stable capital levels — indicates that banks manage their regulatory ratios through the denominator rather than the numerator. The preference for RWA management likely reflects the relative flexibility and lower cost of adjusting asset composition compared to raising new capital. More specifically, banks can reduce RWAs through various channels, including portfolio reallocation to assets with lower risk-weights, increased collateralization, and the use of credit risk mitigation techniques.

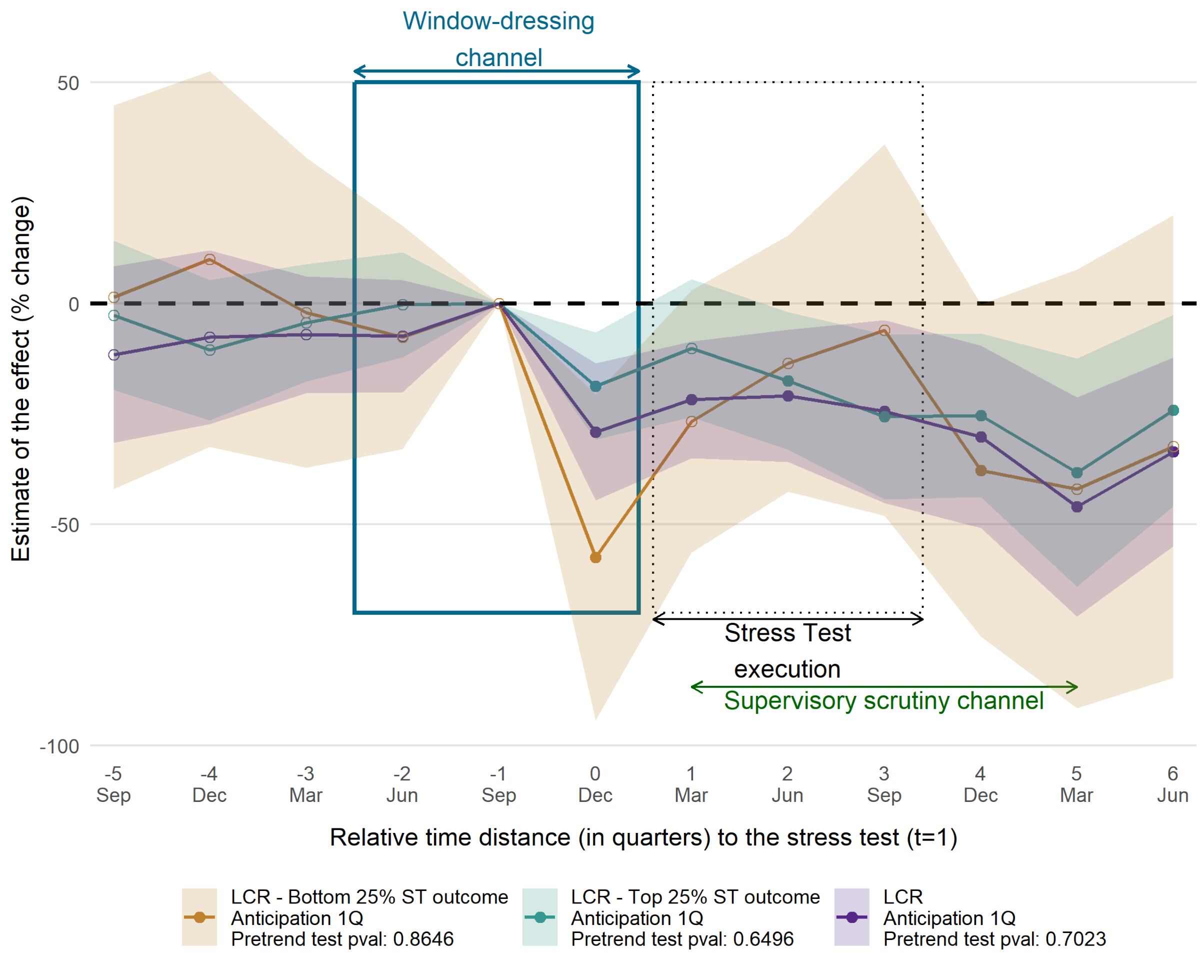
B.2 Liquidity and profitability

We complement our results by reflecting on possible implications of window dressing on the liquidity and profitability of banks. Figure 14 describes a decrease in the Liquidity Coverage Ratio (LCR) during the window-dressing period for all banks subject to the stress test and irrespective of their final outcome. While a reduction in liquidity buffers might appear counterintuitive at first, as it incurs the risk of reducing regulatory requirements right before a supervisory exercise, it aligns with a strategic reallocation of resources toward capital optimization.

Two key considerations rationalize the tendency to reduce LCR before an EBA stress test. First, banks typically maintain LCR levels well above regulatory requirements, providing them with substantial flexibility to adjust their liquidity positions (see European Banking Authority, 2021, 2023). Next, stress tests focus primarily on solvency metrics. Therefore, banks are more likely to strategically deploy excess liquidity buffers to optimize their position with respect to capital indicators, which are the key aspects of the stress test evaluation prominently featured in the official publications of the stress test results.

Turning to profitability, figure 15 reports the evolution of Return on Equity (RoE) around stress test announcements and reveals no significant changes in RoE during the window-dressing period. Anticipatory adjustments to risk-weighted assets, therefore, do not substantially impact banks' earnings capacity in the short term.

Figure 14: Stress test effects on Liquidity Coverage Ratio

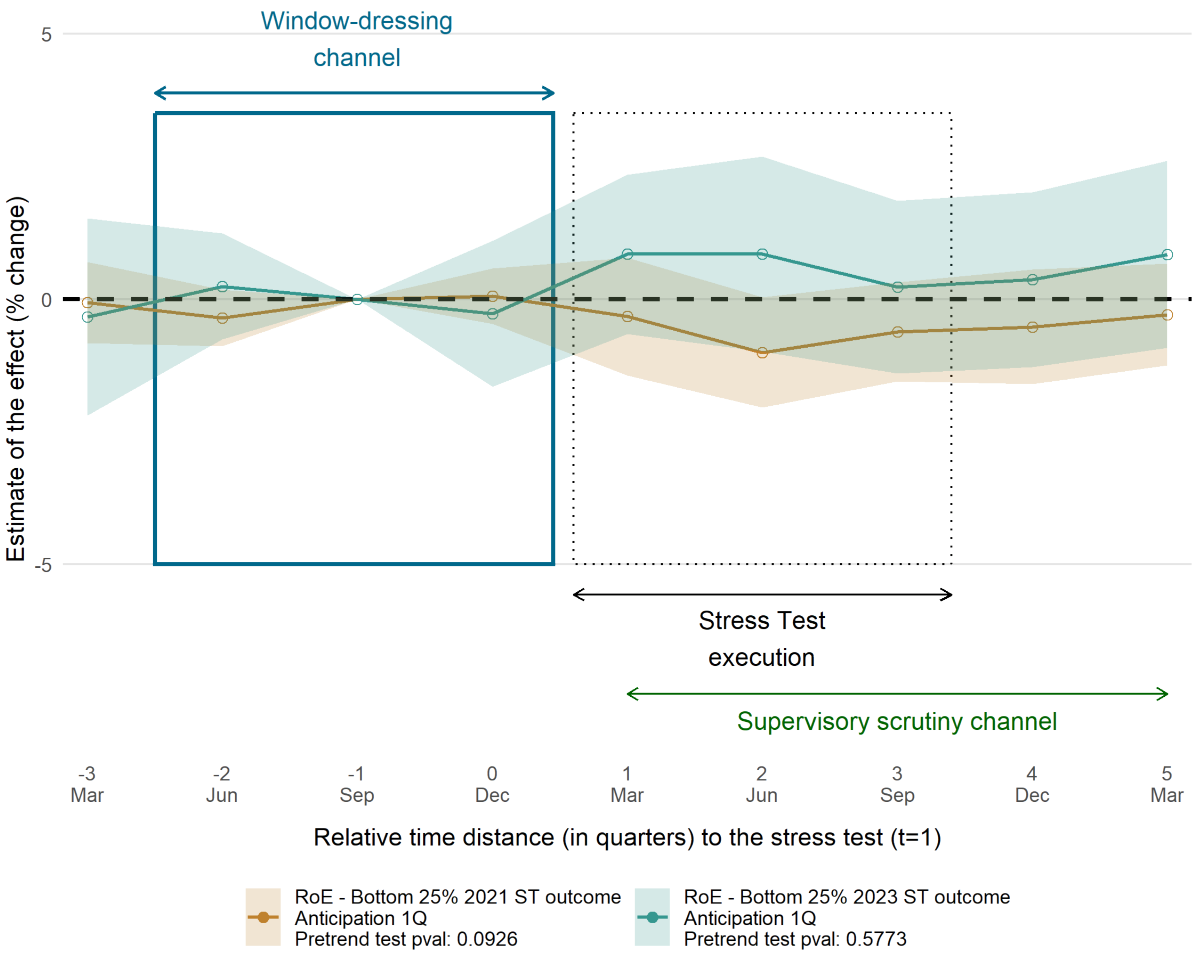


Note: The plot shows dynamic treatment effects estimated using De Chaisemartin and d'Haultfoeuille (2024) for the Liquidity Coverage Ratio (LCR, expressed in log levels). Effects are displayed as points (filled when significant) with confidence intervals (95% level) shown as bands, interpreted as percentage changes from the pre-treatment period. All specifications use entity-level clustering for standard errors, include country-specific non-parametric trends, and control for total assets and ROE. The timeline is expressed in quarters relative to stress test initiation (t=1). The estimation accounts for anticipation effects by using quarter -1 as the reference point for comparing pre-treatment trends. The results indicate a substantial decline in LCR during the window-dressing period, with an average reduction of 30% across all banks, with the effect being particularly pronounced among banks in the bottom quartile with respect to performance in the stress tests.

C Effects of the publication of the results

This appendix explores the impact of stress tests on banks' market-based performance, complementing our main analysis on window dressing behavior. While the primary focus of our paper is on anticipatory balance sheet adjustments, examining market reactions provides a more comprehensive understanding of how stress tests influence bank behavior through different channels.

Figure 15: Stress test effects on Return on Equity



Note: The plot shows dynamic treatment effects estimated using De Chaisemartin and d'Haultfoeuille (2024) for Return on Equity (RoE) comparing banks in the bottom 25% of performers across the 2021 and 2023 stress test exercises. Effects are displayed as points (filled when significant) with confidence intervals (95% level) shown as bands. All specifications use entity-level clustering for standard errors, include country-specific non-parametric trends, and control for total assets. The timeline is expressed in quarters relative to stress test initiation (t=1). The estimation accounts for anticipation effects by using quarter -1 as the reference point for comparing pre-treatment trends. The results reveal no significant impact neither within the window-dressing period nor through other channels.

We analyze whether the publication of stress test results affects bank fundamentals through market discipline, though we note that the quarterly frequency of our data may limit our ability to capture short-term market reactions.

C.1 Market reaction to stress test publications

The impact on market performance may operate through different channels than window dressing, especially through the market discipline effect. Leveraging our estimation methodology, we explore additional transmission channels identified in the literature to corroborate the robustness of the identified window dressing behaviour. Specifically, we examine market reactions, though we acknowledge that our quarterly data may not capture more granular effects documented, e.g., in Durrani et al. (2024) using higher-frequency observations.

To investigate the market reaction channel, we shift the focus to those quarters close to the stress test publication rather than those around its inception (Figure 16). We compare banks with detailed stress test results published by the EBA (EBA sample) against those with aggregated results only (SSM sample).

Table 7 presents these effects, separately analyzing institutions in the top 25% and below the median of the published capital depletion. Such an asymmetric split is not arbitrary: it allows us to isolate the effects for the best performers while maintaining sufficient sample size for the poor performers. The analysis reveals an asymmetric response to the publication of results. For institutions below the median, each additional percentage point of capital depletion relative to the median is associated with a 0.5% decrease in CET1 ratio. Conversely, institutions in the top quartile experience a 1.0% increase in ROE for each percentage point of lower depletion relative to the median. For robustness, we provide a complementary analysis using a symmetric split around the median in Section C.2.

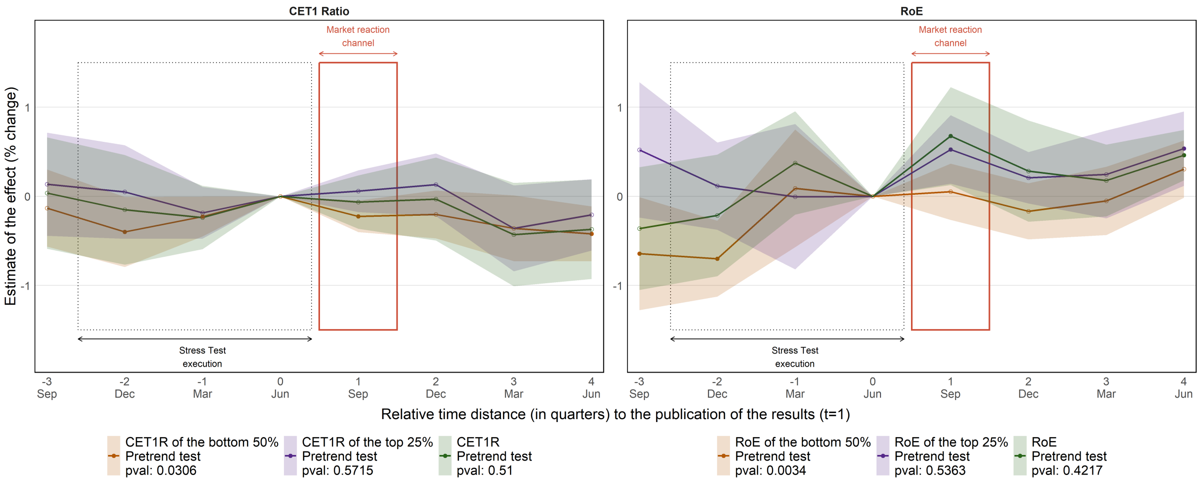
Table 7: Marginal Effects of Stress Test Performance on Bank Fundamentals

	CET1 Ratio			ROE		
Sample	Full Sample	Top Quartile	Below Median	Full Sample	Top Quartile	Below Median
Estimate	-0.009	-0.259	-0.554*	0.016*	1.032*	0.062
95% CI	[-0.024, 0.007]	[-0.969, 0.451]	[-0.923, -0.186]	[0.003, 0.029]	[0.190, 1.875]	[-0.401, 0.525]

The table presents marginal effects estimated over four quarters following stress test result publication, using a continuous treatment variable defined as the absolute distance from the median CET1 depletion. For banks below the median, the treatment measures additional percentage points of CET1 depletion relative to the median; for banks in the top 25%, it measures the extent to which projected depletion was lower than the median. Estimation methodology and controls follow the specification detailed in Figure 16. The results reveal significant effects depending on banks' performance in the stress tests: for banks below the median capital depletion in the stress tests, each additional percentage point of CET1 depletion relative to the median is associated with a 0.55% reduction in actual CET1 Ratio, i.e., disclosed in the supervisory reporting. For banks in the top quartile, each percentage point of better stress test performance (measured as lower projected depletion in the stress test scenarios) is associated with a 1.03% increase in RoE.

Interestingly, in addition to the substantial adjustments in capital and risk-weighted assets

Figure 16: Publication effects on CET1 Ratio and RoE



Note: The figure shows differential effects estimated using De Chaisemartin and d'Haultfoeuille (2024) comparing granular (EBA) versus aggregated (SSM) stress test result publication on CET1 Ratio (left panel) and RoE (right panel). For internal comparability outcome variable are measured in log, so that the coefficient can be interpreted as elasticities to the treatment status. Effects are displayed as points (filled when significant) with confidence intervals (95% level) shown as bands, expressed as percentage changes from the baseline period. The analysis compares banks with published results (EBA sample), split by above and below median performance, against banks with non-published results (SSM sample) to isolate market reaction effects. All specifications use entity-level clustering for standard errors, include country-specific non-parametric trends, and control for total assets. The timeline spans four quarters after result publication, with the reference period set at the conclusion of the stress test exercise. The results reveal heterogeneous effects: banks with below-median stress test results experience a small but significant decline in CET1 Ratio of 0.3-0.4%, while top-performing banks with published results show a significant increase in RoE of approximately 1%. Other differential effects are not statistically significant.

documented earlier, we also find limited evidence of significant profitability effects.[34] The RoE within the quarter of stress test results publication increases for banks with publication of granular stress test results, i.e. EBA sample banks, that achieve high scores in the outcomes. This positive effect for top performers likely reflects market recognition of their resilience. We do not observe, however, a symmetric negative effect for banks with the highest depletion.

Concluding, the absence of broader profitability effects suggests that banks manage to implement their precautionary capital buffers and risk-weighted asset reductions without significantly compromising their earnings capacity. This finding is particularly noteworthy given the substantial de-risking observed among poor performers, indicating that banks can adapt their balance sheets to regulatory requirements while maintaining profitability.

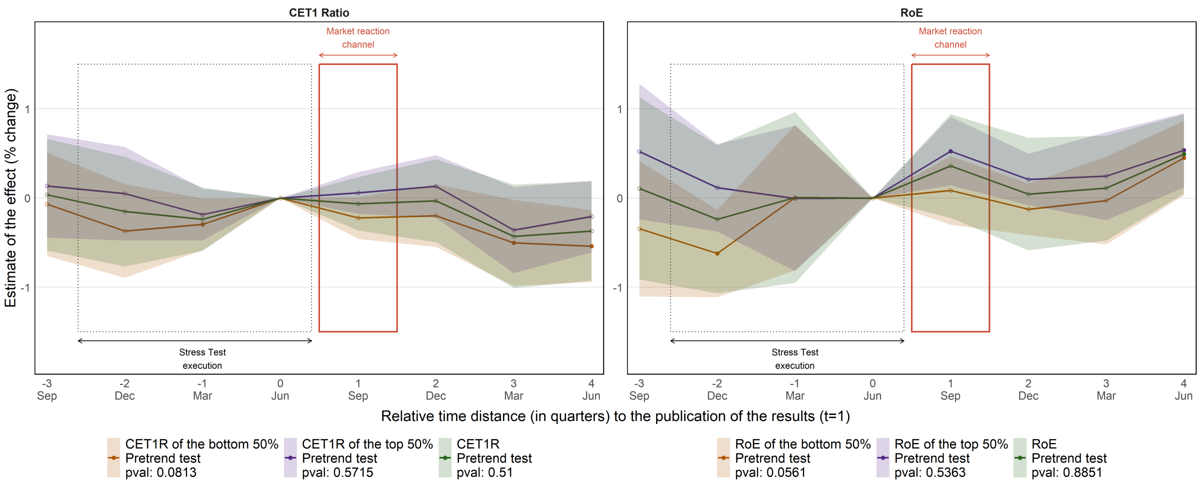
[34]Admittedly, it is possible that more granular short-term market reactions remain uncovered by the available quarterly frequency of the data. Unfortunately, for the purpose of this study, more granular data are not available. However, detailed investigations on market reactions are already available in the literature (e.g. (Durrani et al., 2024)), while our main focus in the paper is the study of the window-dressing channel.

C.2 Publication results: robustness with symmetric sample split

In our main analysis of publication effects, we used an asymmetric split of the sample, comparing the top quartile against banks below the median in the distribution of projected capital depletion. Here, we verify the robustness of our findings using a symmetric split that compares the top and bottom quartiles of the published depletion. This complementary analysis helps ensure our results are not driven by the specific sample partition chosen in the main analysis.

Figure 17 presents the differential effects using this alternative specification. The results largely confirm our main findings. The magnitudes of the effects are actually larger when focusing on the extreme quartiles: banks in the bottom quartile show a more pronounced negative CET1 Ratio effect (-0.67% vs -0.5% in the main analysis), while the positive RoE effect for top quartile performers trivially remains unchanged at 1.03%.

Figure 17: Publication effects on CET1 Ratio and RoE (Symmetric Quartile Split)



Note: The figure shows differential effects estimated using De Chaisemartin and d'Haultfoeuille (2024) comparing granular (EBA) versus aggregated (SSM) stress test result publication on CET1 Ratio (left panel) and RoE (right panel). For internal comparability outcome variable are measured in log, so that the coefficient can be interpreted as elasticities to the treatment status. Effects are displayed as points (filled when significant) with confidence intervals (95% level) shown as bands, expressed as percentage changes from the baseline period. The analysis compares banks with published results (EBA sample), split by top and bottom quartile performance, against banks with non-published results (SSM sample) to isolate market reaction effects. All specifications use entity-level clustering for standard errors, include country-specific non-parametric trends, and control for total assets. The timeline spans four quarters after result publication, with the reference period set at the conclusion of the stress test exercise. The results reveal heterogeneous effects: banks with bottom quartile stress test results experience a small but significant decline in CET1 Ratio of approximately 0.7%, while top-performing banks with published results show a significant increase in RoE of approximately 1%. Other differential effects are not statistically significant.

This alternative analysis continues to reveal asymmetric responses to result publication.

Table 8 shows that for institutions in the bottom quartile, each additional percentage point of capital depletion relative to the median is associated with a 0.67% decrease in CET1 ratio. Conversely, institutions in the top quartile experience a 1.03% increase in ROE for each percentage point lower depletion relative to the median. The ROE effect for poor performers (0.173%) is not statistically significant, suggesting that market discipline primarily operates through negative capital effects for the worst performers and positive profitability effects for the best performers.

Table 8: Marginal Effects of Stress Test Performance on Bank Fundamentals (Symmetric Quartile Split)

	CET1 Ratio			ROE		
Sample	Full Sample	Top Quartile	Bottom Quartile	Full Sample	Top Quartile	Bottom Quartile
Estimate	-0.009	-0.259	-0.670*	0.010	1.032*	0.173
95% CI	[-0.024, 0.007]	[-0.969, 0.451]	[-1.147, -0.193]	[-0.006, 0.027]	[0.190, 1.875]	[-0.392, 0.739]

The table presents marginal effects estimated over four quarters following stress test result publication, using a continuous treatment variable defined as the absolute distance from the median CET1 depletion. In order to have a positive measure and interpret the results consistently, for banks in the bottom 25%, the treatment measures additional percentage points of CET1 depletion relative to the median; for banks in the top 25%, it measures the extent to which projected depletion was lower than the median. Estimation methodology and controls follow the specification detailed in Figure 17. The results reveal significant effects depending on banks' performance in the stress tests: for banks in the bottom quartile, each additional percentage point of CET1 depletion relative to the median is associated with a 0.67% reduction in actual CET1 Ratio. For banks in the top quartile, each percentage point of better stress test performance is associated with a 1.03% increase in RoE.

This robustness check strengthens our main conclusions about the asymmetric nature of market reactions to stress test publications, showing that these effects are most pronounced when comparing banks at the extremes of the performance distribution.