Informational Value and Abnormal Stock Returns:

An Event Study of EU-Wide Stress Tests from 2010 to 2018

Kolja Gauer

A thesis submitted in partial fulfilment of the requirements of Edinburgh Napier University,
for the award of Doctor of Philosophy

February 2023

This thesis is dedicated to my children Paul and Hanna.
My wish for you is that you will always keep your innate curiosity and continue to cultivate your critical thinking as you grow up.
You will always be on my mind. I love you!

Declaration

I hereby declare that this thesis is the result of my own independent work and that it has not been submitted for any other degree or professional qualification.

[redacted]

Kolja Gauer

24 February 2023

Acknowledgements

Motivated by personal experiences in asset management and corporate finance during the recent global financial crisis (2007-2009), I embarked on my PhD journey on capital market reactions to supervisory stress tests – a journey that would soon become my very own personal stress test. I confidently accepted the challenges and uncertainties that lay ahead, but then, halfway through my PhD, life happened: I married a wonderful woman, became a father of twins, bought and renovated a family home, changed jobs, and moved to Brussels. Given these circumstances and the challenges inherent in scientific research, my PhD journey would not have been possible without the company and support of various groups of people. All of them contributed in one way or another to the success of this thesis. Now it is time to formally acknowledge their contribution and express my gratitude to them.

First and foremost, I would like to thank the members of my supervisory team, Dr Marizah Minhat, Dr Nazam Dzolkarnaini, Dr Piotr Jaworski, and Prof. Simon Gao for their guidance and support throughout my journey. I am grateful for their invaluable comments and suggestions, as well as for their academic rigor and determination to keep pushing the boundaries. My thanks also go to Dr Morrison Handley-Schachler and Prof. Richard Whitecross, who acted as my independent panel chairs and oversaw the entire process. It goes without saying that any remaining errors and oversights are my sole responsibility.

I am also indebted to the faculty and staff at Edinburgh Napier University for providing an excellent research environment. This specifically includes the librarians and the organisers of the Business School's annual research conference. My gratitude also goes to my fellow PhD students and my good friend Dr Stefan Ehlers for many intense and controversial discussions. It is also important for me to thank the people in Edinburgh, who have always welcomed me with open arms from the very first moment. This is particularly true of the management and staff at the Braid Hills Hotel in Morningside, who have made Edinburgh my second home.

Last – but certainly not least – I would like to express my personal gratitude to my family who had to endure a father and husband who was at times mentally and physically absent for several years. I am especially grateful to my wife Alice, without whom none of this would have been possible. Given the special challenges of a global pandemic, this cannot be appreciated enough. Thank you for your love, patience, and continued support in pursuing my dreams. I also want to thank the rest of my family for their moral support and their unwavering belief in me. Thank you!

Abstract

Since the global financial crisis (2007-2009), supervisory stress testing has become increasingly important. Previous studies of banks' abnormal stock returns in response to EU-wide stress test results have produced inconsistent and contradictory results, while leaving important aspects unexplored. The aim of this study was therefore to address these shortcomings. Specifically, the study examined the five EU-wide stress tests conducted between 2010 and 2018 with the objective of: (1) testing the impact of stress test results on bank stock prices, (2) analysing the relationship between stress test results and abnormal stock returns, and (3) determining how the informational value of stress test results has changed over time. The study's original contribution to knowledge are more comprehensive insights into the informational value of EU-wide stress test results.

The study was conducted on five cross-sectional samples (*n*=33 to *n*=59) and one longitudinal sample (*n*=28) of banks. Data were collected using structured direct observation and analysed following a quasi-natural experimental strategy. Based on an event-study approach, (absolute) cumulative abnormal returns were determined for three event-window types to conduct research question-specific analyses.

The findings showed that EU-wide stress test results generally had a significant impact on banks' stock prices, suggesting that they provided investors with valuable new information, in most cases. Further analysis revealed a counterintuitive U-shaped relationship between stress test results and abnormal stock returns, implying that banks with particularly positive and negative stress test results experienced disproportionately positive abnormal stock returns. The longitudinal analysis found no discernible trend in the informational value of stress test results across the five EU-wide stress tests examined.

These insights contribute to filling empirical gaps in understanding abnormal stock returns in response to EU-wide stress test results. They could therefore be used by policymakers and investors to refine their disclosure policies or to develop profitable investment strategies.

"The plan aimed to impose transparency on opaque financial institutions and their opaque assets in order to reduce the uncertainty that was driving the panic. It would help markets distinguish between viable banks that were temporarily illiquid and weak banks that were essentially insolvent."

Timothy F. Geithner (2014, p. 286),
looking back on the Supervisory Capital Assessment Program (SCAP)*

* Timothy F. Geithner served as president of the Federal Reserve Bank of New York (2003-2009) and US Secretary of the Treasury (2009-2013). In both functions, he played a key role in the development and implementation of the SCAP, which was the first supervisory stress test to be pioneered during the global financial crisis (2007-2009). The SCAP not only marked a critical turning point in the crisis, but also defined the beginning of a new era in prudential policy.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AIC .... Akaike Information Criterion
AMA .... Advanced Measurement Approach
ANOVA .... Analysis of Variance
|AR| .... Absolute Abnormal Return
BCBS .... Basel Committee on Banking Supervision
BHAR .... Buy-and-Hold Abnormal Return
BHC .... Bank Holding Company
CAP .... Capital Assistance Program
CAPM .... Capital Asset Pricing Model
CAR .... Cumulative Abnormal Return
$\overline{\text{CAR}}$ .... Average Cumulative Abnormal Return
$|\text{CAR}|$ .... Absolute Cumulative Abnormal Return
$|\overline{\text{CAR}}|$ .... Average Absolute Cumulative Abnormal Return
CBOE .... Chicago Board Options Exchange
CCAR .... Comprehensive Capital Analysis and Review
CDS .... Credit Default Swap
CEBS .... Committee of European Banking Supervisors
CET 1 .... Common Equity Tier 1
CGFS .... Committee on the Global Financial System
COVID-19 .... Corona Virus Disease 2019
CRD .... Capital Requirements Directive
CRR .... Capital Requirements Regulation
DFAST .... Dodd-Frank Act Stress Test
EBA .... European Banking Authority
EC .... European Commission
ECB .... European Central Bank
ECOFIN .... Economic and Financial Affairs Council
EFSF .... European Financial Stability Facility

EIOPA .... European Insurance and Occupational Pensions Authority
EMH .... Efficient Market Hypothesis
EONIA .... Euro Overnight Index Average
ESFS .... European System of Financial Supervision
ESM .... European Stability Mechanism
ESMA .... European Securities and Markets Authority
ESRB .... European Systemic Risk Board
EU .... European Union
EURIBOR .... Euro Interbank Offered Rate
FDIC .... Federal Deposit Insurance Corporation
Fed .... Federal Reserve
FF3F .... Fama-French Three-Factor Model
FF5F .... Fama-French Five-Factor Model
FSAP .... Financial Stability Assessment Program
FWER .... Family-Wise Error Rate
GARCH .... Generalised Autoregressive Conditional Heteroskedastic
GDP .... Gross Domestic Product
G-SIB .... Global Systemically Important Bank
HQIC .... Hannan-Quinn Information Criterion
I/B/E/S .... Institutional Brokers' Estimate System
IMA .... Internal Models Approach
IMF .... International Monetary Fund
IPO .... Initial Public Offering
IRBA .... Internal Ratings-Based Approach
JST .... Joint Supervisory Team
JTB .... Justified True Belief
LCR .... Liquidity Coverage Ratio
LIBOR .... London Interbank Offered Rate
LIL .... Law of the Iterated Logarithm
MAM .... Market-Adjusted Model
MAR .... Mean-Adjusted Return Model
MCP .... Multiple Comparisons Problem
MM .... Market Model
MPI .... Macroprudential Indicator
MSE .... Mean Squared Error
NASDAQ .... National Association of Securities Dealers Automated Quotations

NCA .... National Competent Authority
NSFR...Net Stable Funding Ratio
NYSE ....New York Stock Exchange
OCC .... Office of the Comptroller of the Currency
OCR .... Overall Capital Requirement
OIS .... Overnight Index Swap
OLS ....Ordinary Least Squares
O-SSI.. Other Systemically Important Institution
OTC...Over the Counter
P2G.. Pillar 2 Guidance
P2R... Pillar 2 Requirement
QAP.. Quality Assurance Process
RegTech .... Regulatory Technology
RePEc.. Research Papers in Economics
RMSE...Root Mean Squared Error
RWA ....Risk-Weighted Assets
S&P ....Standard & Poor’s
SA...Standardised Approach
SCAP...Supervisory Capital Assessment Program
SIC .... Schwarz Information Criterion
SML ....Security Market Line
SREP .... Supervisory Review and Evaluation Process
SSM.. Single Supervisory Mechanism
SSE...Sum of Squared Errors
US...United States
VaR ....Value-at-Risk
VIX...CBOE Implied Volatility Index
VSTOXX ....Euro Stoxx 50 Volatility Index

Chapter 1
Introduction

1.1 Background

Since the global financial crisis (2007-2009), bank supervisors have honed their financial stability monitoring tools and significantly expanded the use of stress testing. Supervisory stress tests assess banks' capital adequacy under adverse macro-financial conditions based on hypothetical scenarios. They examine the impact of deteriorating macro-financial indicators (such as gross domestic product, inflation, interest rates, or asset prices) on banks' trading and banking books. In other words, stress tests translate adverse macro-financial scenarios into hypothetical asset losses on banks' balance sheets (Acharya *et al.* 2014). The key result of a stress test is a stressed capital ratio, which can be compared with the bank's actual capital ratio from the last financial statement for assessment purposes. Sometimes bank supervisors set thresholds that use the stressed capital ratio to determine whether a bank has passed or failed a stress test. Banks that fall below the threshold are typically expected to fill the capital shortfall or face appropriate supervisory actions such as capital distribution restrictions.[1]

Supervisory stress tests as used today are a direct response to lessons learned from the global financial crisis (2007-2009): low investor confidence and uncertainty about the quantity and quality of bank capital paralyzed international capital markets; in the absence of a clear understanding of banks' solvency, investors have been reluctant to allocate capital (Morgan *et al.* 2014, Schuermann 2014, Wall 2014a). To counteract these developments through a credible assessment of banks' solvency, bank supervisors, particularly in the US and EU, began conducting system-wide stress tests. The purpose of supervisory stress testing is thus to test banks' resilience to severe but

[1] For an overview of possible supervisory actions in the EU, see Article 16(2) of the SSM Regulation (EU) 1024/2013.

plausible shocks (*i.e.* their ability to absorb losses over periods of macro-financial distress), to improve transparency about bank risks, and to promote market discipline through public disclosure of stress test results (BCBS 2009, CEBS 2009a, Fed 2009a). By publicly disclosing bank-level stress test results, stress testing deliberately departs from the established supervisory practice of keeping bank-specific supervisory information confidential (Fed 2009a). This allows investors unprecedented insights into bank risks and their sensitivity to changing macro-financial indicators. It should also allow investors to better price-discriminate between sound and unsound banks.

The US Supervisory Capital Assessment Program (SCAP), conducted in early 2009, was the first ever system-wide supervisory stress test (Fed 2009a). It marked a critical turning point in the crisis (Bernanke 2013, Bookstaber *et al.* 2014, Langley 2013) and opened a new evolutionary stage in the design and application of supervisory stress tests (Hirtle and Lehnert 2015, Kapinos *et al.* 2018, Schuermann 2014). The SCAP was followed by the Comprehensive Capital Analysis and Review (CCAR) in 2011 and the Dodd-Frank-Act Stress Test (DFAST) in 2013, which have been carried out regularly ever since.

Similarly, in May 2009, the Economic and Financial Affairs Council (ECOFIN) of the EU mandated the Committee of European Banking Supervisors (CEBS) to conduct the first in a series of EU-wide stress tests (CEBS 2009b). Since 2011, EU-wide stress tests have been carried out regularly by the European Banking Authority (EBA) under Article 32 of the EBA Regulation (EU) 1093/2010. The results feed into the Supervisory Review and Evaluation Process (SREP) where they are used to provide Pillar 2 Guidance (P2G)[2] under the Basel III framework. The five EU-wide stress tests carried out between 2010 and 2018 are the subject of this thesis. Table 1 provides an overview of the main features of the EU-wide stress tests examined.

[2] Pillar 2 Guidance (P2G) is a supervisory capital expectation above the Overall Capital Requirement (OCR) consisting of the minimum own funds requirement (Pillar 1), the additional own funds requirement (Pillar 2 Requirement – P2R), and, if relevant, the combined buffer requirement or the leverage ratio buffer requirement. The European Central Bank (ECB) and national competent authorities (NCAs) may provide Pillar 2 Guidance for banks to meet capital requirements under stress scenarios (Capital Requirements Directive (EU) 2019/878 (CRD V)). For more details on the institutional framework of EU-wide stress tests, see Section 2.4.3, and in particular Figure 4. For more information on the link between EU-wide stress tests and financial stability policy, see Ebner (2018).

Table 1
Overview of EU-Wide Stress Tests Conducted from 2010 to 2018

Stress Test	Banks (total assets)[a]	Risks Covered	Threshold (adverse)	Capital Shortfall[b]	Data Points per Bank	Disclosure (t_0)[c]
CEBS 2010	91 (65%)	Credit risk, market risk, sovereign risk	6% Tier 1	€ 3.5 bn. (7 banks)	149	Fri., 23.07.2010 5:30 p m.
EBA 2011	90 (65%)	Credit risk, market risk, sovereign risk	5% CET 1	€ 26.8 bn. (20 banks)	3,200	Fri., 15.07.2011 5:00 p m.
EBA 2014	123 (70%)	Credit risk, market risk, sovereign risk, cost of funding	5.5% CET 1	€ 24.2 bn. (24 banks)	12,000	Sun., 26.10.2014 11:00 a.m.
EBA 2016	51 (70%)	Credit risk[d], market risk, counterparty credit risk (CCR), credit valuation adjustment (CVA), operational risk[e]	NA	€ 269 bn. (NA)	16,000	Fri., 29.07.2016 09:00 p.m.
EBA 2018	48 (70%)	Credit risk[d], market risk, counterparty credit risk (CCR), credit valuation adjustment (CVA), Operational risk[e]	NA	€ 246 bn. (NA)	17,200	Fri., 02.11.2018 05:00 p.m.

Note. This table provides an overview of the five EU-wide stress tests conducted from 2010 to 2018 and their main features. Data is compiled from EBA (2020a). CEBS = Committee of European Banking Supervisors. EBA = European Banking Authority. CET 1 = Common Equity Tier 1 ratio. NA = not applicable. a.m. = ante meridiem. p.m. = post meridiem.

[a] The term "total assets" denotes the percentage of the EU banking system's total assets covered by the group of banks subject to any given stress test. To be selected into a stress test, banks must have at least € 30 bn. in assets as of their most recent annual statement prior to the exercise. Furthermore, jurisdiction-specific selection thresholds apply (for further details on the selection rule, see Section 5.3.1.2). [b] For the stress tests CEBS 2010, EBA 2011, and EBA 2014, the term "Capital Shortfall" refers to the aggregate capital gap of the banks failing to meet the defined thresholds; in contrast, for EBA 2016 and EBA 2018 "Capital Shortfall" refers to the total transitional CET 1 capital depletion across all banks subject to the stress tests (as the EBA did not define pass/fail thresholds for these exercises). [c] London time. [d] Including securitisation. [e] Including conduct risk.

The regularity with which supervisory stress tests are conducted today (both in the EU and the US) indicates that they have evolved from a crisis resolution tool into an important and well-established financial stability monitoring tool. However, there are still a number of problems and open research questions related to supervisory stress testing; this is discussed in more detail below.

1.2 Problem Statement

The novel role of stress tests in the macroprudential toolkit of banking supervision has sparked considerable research interest. Since its inception in 2009, numerous researchers have studied various aspects of supervisory stress testing. Over time, five main lines of research have emerged: (1) Stress-Test Methodology (*e.g.* Acharya *et al.* 2014, Borio *et al.* 2014, Schuermann 2020), (2) Scenario Selection (*e.g.* Breuer and Csiszár 2013, Flood and Korenko 2015, Glasserman *et al.* 2015), (3) Stress-Test Governance (*e.g.* Hirtle and Lehnert 2015, Ong and Pazarbasioglu 2014, Wall 2014a), (4) Results Disclosure Policy (*e.g.* Faria-e-Castro *et al.* 2017, Goldstein and Leitner 2018, Pacicco *et al.* 2020), and (5) Capital Market Reactions (*e.g.* Ahnert *et al.* 2020, Flannery *et al.* 2017, Petrella and Resti 2013).

This study is located in the last-mentioned research line (with some links to results disclosure policy). This particular line of research examines the extent to which stress test-related events (*e.g.* stress test announcements, methodological clarifications, or results disclosures) were of informational value for investors. In other words, market reaction studies are concerned with whether such events have caused statistically significant abnormal returns in the securities of affected banks. The problem is that, despite considerable research, there are still significant research gaps regarding abnormal bank stock returns in response to the disclosure of EU-wide stress test results. These gaps are described in more detail below.

First, there is as yet no coherent understanding of the informational value contained in the results of EU-wide stress tests. This is because the existing body of research is a patchwork of different EU-wide stress tests and research method specifications (see the review of previous studies in Section 3.3.2, in particular the overview in Table 3). It is therefore not surprising that previous studies have produced mixed and sometimes contradictory results (*e.g.* Candelon and Sy 2015 and Petrella and Resti 2013).[3] The collective evidence from existing research is therefore inconclusive and difficult to compare across different EU-wide stress tests. What is still missing is a study that determines the informational value of all EU-wide stress tests available for research based on a systematic and consistent methodology. This also includes a systematic model selection procedure, which is crucial for the internal validity of event studies (Fama 1970) but has been neglected in all previous studies.

Second, the current knowledge about the extent to which EU-wide stress test results have helped investors to price-discriminate between financially sound and unsound banks is insufficient. In other words, it is not yet clear whether the disclosure of EU-wide stress test results has supported market discipline as defined in Pillar 3 of the Basel Capital Accords – a key goal of EU-wide stress testing (EBA 2020c, Enria 2018,

[3] The studies by Candelon and Sy (2015) and Petrella and Resti (2013) are striking examples of contradictory results. Both studies have examined the 2011 EU-wide stress test using the standard event study approach with the same five-day event window (-2, +2) and the same 200-day estimation period. In fact, the only difference was the asset pricing model used to estimate normal (expected) returns: Candelon and Sy (2015) used the Capital Asset Pricing Model (CAPM), while Petrella and Resti (2013) used the Market Model. Despite their very similar research design, Candelon and Sy (2015) found an average cumulative abnormal return ($\overline{CAR}$) of -0.019%, which was significant at the 5% level, while Petrella and Resti (2013) found a $\overline{CAR}$ of 0.8% which was not statistically significant. This contradiction is consistent with Fama's (1970, 1991) joint-hypothesis problem, according to which abnormal returns may reflect market inefficiencies or inappropriate asset pricing models (or both).

Quagliariello 2020). Previous studies have been largely limited to examining differences in market response between two dichotomous groups of banks, typically taking a "pass vs. fail" perspective (Ahnert *et al.* 2020, Georgescu *et al.* 2017, Petrella and Resti 2013).[4] As a result, there is still no differentiated understanding of how investors have revised their previous risk assessments of banks in light of EU-wide stress test results. In order to improve the current state of knowledge, the functional relationship between stress test results and corresponding abnormal stock returns needs to be examined at bank level to determine whether it is linear or non-linear in nature.

Third, after more than a decade of EU-wide stress tests, there is still no longitudinal study on whether and how the informational value of EU-wide stress test results has changed across the various exercises. This is particularly surprising given that a number of studies have found that the informational value of US stress test results has declined over time (Candelon and Sy 2015, Fernandes *et al.* 2020, Sahin *et al.* 2020). It has been argued that this is due to perverse incentives in US stress tests (Cornett *et al.* 2020, Glasserman and Tangirala 2016, Goldstein and Sapra 2014) giving rise to Goodhart's (1975) law. Similar concerns have also recently been raised in the European context, but have not been further explored (Kok *et al.* 2019, Quagliariello 2020). This calls for a longitudinal analysis to empirically examine whether the informational value of EU-wide stress test results has been intertemporally stable.

The above gaps in the existing research motivated the formulation of the following research aim, which guided and directed this study.

[4] Ahnert *et al.* (2020), for example, examined whether the market reactions of banks that "passed" or "failed" an EU-wide stress test differed significantly from one another. Alves *et al.* (2015) carried out a similar analysis, but added a third group of banks that "barely passed" an EU-wide stress test. Georgescu *et al.* (2017) and Petrella and Resti (2013), on the other hand, based their analyses on two groups of banks, which were composed according to the level of the stressed capital ratio (large vs. small capital ratio impact and top 20% vs. bottom 20%, respectively).

1.3 Research Aim

The aim of this quantitative event study was to examine banks' abnormal stock returns in response to the results of the five EU-wide stress tests carried out by the CEBS and the EBA between 2010 and 2018.

The analytical lens through which this aim was pursued consisted in an extended event-study approach, which implied a distinctly quantitative perspective. The methodological and philosophical perspective from which this study was conducted is described in more detail in Chapter 5. The above aim unfolded into the following research questions and objectives.

1.4 Research Questions and Objectives

To achieve the aim of this study and to fill the gaps in existing research (Section 1.2), the following three research questions were developed. Each of them targeted a specific aspect of the impact of EU-wide stress test results on the abnormal stock returns of affected banks, *i.e.* the average intervention effect, the functional relationship between cause and effect, and the intertemporal dynamics of the average intervention effect.

Research Question 1: *What is the average value of the information contained in the results of EU-wide stress tests measured in terms of abnormal stock returns?*

The main objective of this research question was to quantify the average informational value of EU-wide stress test results. At its most basic level, the question asked whether the results disclosures actually conveyed valuable new information to investors or whether they were simply non-events (*i.e.* events with no informational value). At a higher level, the question aimed to determine the value of the information conveyed (average intervention effect), which is usually measured in average abnormal returns in an event-study setting (Campbell *et al.* 1997, Kothari and Warner 2007, MacKinlay 1997). The scope of this objective extended to each of the five EU-wide stress tests available for research (Table 1).

Another objective was therefore to develop a systematic methodology that could be universally applied to all EU-wide stress tests in order to ensure an unbiased and consistent analysis. This involved extending the existing standard approach to event studies (Campbell *et al.* 1997, MacKinlay 1997) with a systematic model selection procedure for the normal return-generating model. The purpose of this objective was twofold: first, to reduce uncertainties arising from methodological weaknesses in previous studies related to Fama's (1970, 1991) joint-hypothesis problem (see also Sections 3.23.3.1 and 3.5.3);[5] and second, to facilitate comparability between EU-wide stress tests through the use of a consistent methodology.

The final objective of this research question was to develop a dedicated theoretical framework for studying market reactions to supervisory transparency measures in the banking sector. This was an urgent need for this study and beyond, since previous studies have failed to construct a clearly specified theoretical framework for such investigations (Section 3.2). In general, previous studies have made little reference to the theory behind their research. The idea was therefore to synthesize an extensible theoretical framework that enables the investigation of this research question and opens up avenues for further research inside (Research Questions 2 and 3) and outside of this study.

Research Question 2: *What is the functional relationship between new information from EU-wide stress test results and corresponding abnormal stock returns?*

The main objective of this research question was to determine, for the first time ever, the functional relationship between stress test results and abnormal stock returns at bank level. That is, the function or curve that best fitted the empirically observed data points in the plane, *i.e.* the idealised assignment of units of return (abnormal returns) to units of risk (stress test results). This made it possible to examine whether the stock prices of banks have adjusted in a linear or non-linear way, or, in other words, proportionally or disproportionally to their stress test results. Thus, while Research Question 1 was concerned with the average intervention effect of EU-wide stress tests, this research question aimed to decompose this aggregated figure by revealing its underlying structure or shape.

---

[5] It should be noted that there is considerable uncertainty in all previous studies about the origin of the abnormal returns found. This is due to the joint-hypothesis problem (Fama 1970, 1991) and the fact that all previous studies have failed to address this problem with a systematic model selection procedure. Any abnormal return found can therefore be arbitrarily attributed to the new information from the EU-wide stress test results or the selection of an inappropriate normal return-generating model. For more information, see the review of previous studies in Section 3.2 and the discussion of problems in testing for market efficiency in Section 3.5.3.

In more practical terms, the objective was to examine whether the goal of enhancing market discipline was actually achieved by disclosing the results of EU-wide stress tests. Market discipline (Pillar 3 of the Basel Capital Accords) is a regulatory mechanism that delegates monitoring and disciplining tasks to the market participants concerned (De Ceuster and Masschelein 2003). Thus, if EU-wide stress tests were effective in enhancing market discipline, the functional relationship between stress test results and abnormal stock returns should reveal the following transmission mechanism: the new information from the stress test results prompted investors to revise their prior risk assessments and thus increased their ability to price-discriminate between banks with different risk profiles.

Research Question 3: *How has the informational value of EU-wide stress test results, measured in abnormal stock returns, changed over time?*

The main objective of this research question was to create a better understanding of the dynamics of the average intervention effect (Research Question 1). In other words, this research question aimed to determine whether the economic and statistical significance of the new information in EU-wide stress test results was intertemporally stable or subject to a certain trend. In particular, whether the informational value of EU-wide stress test results has decreased over time, as has been repeatedly found in US supervisory stress tests (Cornett *et al.* 2020, Glasserman and Tangirala 2016, Goldstein and Sapra 2014). This has never been investigated before in a European context and required, for the first time ever, a longitudinal analysis of banks' abnormal stock returns in response to the disclosure of EU-wide stress test results.

Building on this, another objective was to contribute to the ongoing debate on the results disclosure policy of supervisory stress tests. This debate concerns the extent and circumstances under which sensitive supervisory information should be disclosed to the public, taking into account the potential impact on financial stability and welfare. So far the debate has been largely theoretical (*e.g.* Faria-e-Castro *et al.* 2017, Goldstein and Leitner 2018, Goncharenko *et al.* 2018).[6] In addition, the few empirical analyses were limited to the situation in the US (Goldstein and Yang 2019, Pacicco *et al.* 2020).

[6] For further theoretical studies, see, for example, Berlin (2015), Gick and Pausch (2012), Goldstein and Sapra (2014), and Schuermann (2014, 2016).

The objective was therefore to expand the debate with the empirical findings of a longitudinal analysis that had a distinct European focus and covered both economically stable and unstable times (*e.g.* the European sovereign debt crisis (2010-2013)).

To operationalise the research questions, testable hypotheses are developed in Section 4.3 on the basis of the literature review (Chapter 3) and the newly developed theoretical framework (Section 4.2). The significance of this study is highlighted in the next section.

1.5 Significance of the Study

The research problems in Section 1.2 indicate that this study has significant implications for both supervisors and investors (see Sections 7.3 and 8.4.3). A key lesson for supervisors from the global financial crisis was the need for more thorough testing of banks' resilience to unexpected changes in macro-financial conditions, leading to the rise of supervisory stress testing. Investors, on the other hand, have an obvious interest in knowing the risks of the banks in which they invest, but their ability to make such assessments themselves is limited by the information that is publicly available. The decision of supervisors to depart from established supervisory practice to keep bank-specific information confidential and to disclose supervisory stress tests results at bank level therefore offers an interesting area of research.

This study makes an original contribution to knowledge by providing a comprehensive and unprecedented insight into the informational value of EU-wide stress test results for bank stock pricing. It significantly extends and advances the findings of previous studies by complementing the usual analysis of the average intervention effect with additional investigations of mediating and moderating effects. More specifically, the study integrates rational choice theory and the risk-return tradeoff of investments, as well as Goodhart's law on financial policy indicators, into a classic test of semi-strong form efficiency.

To implement this, a new dedicated theoretical framework for studying market reactions to supervisory transparency measures was developed. The framework was deliberately designed to be extensible and applicable beyond this study, opening avenues for future research. In addition, the existing standard event study approach of Campbell *et al.* (1997) and MacKinlay (1997) was extended to include a systematic model selection procedure and other methodological advances. Furthermore, this is one of the first studies to additionally use the average absolute cumulative abnormal return ($|\overline{CAR}|$) measure proposed by Flannery *et al.* (2017) in a European context. This study thus also makes several contributions to theory and the development of methodology and methods (Sections 8.4.1 and 8.4.2).

Finally, to the best of the author's knowledge, this is the first study ever to conduct a longitudinal analysis of EU-wide stress tests and examine the relationship between stress test results and abnormal stock returns at the bank level. The empirical results of this study therefore contribute to a better understanding of the overall intervention effect of disclosures of EU-wide stress test results, its dynamics over time, and the functional relationship between EU-wide stress test results and abnormal stock returns. The study thus facilitates the further development and refinement of EBA's stress test results disclosure policy and enables investors to develop opportunistic or event-driven investment strategies targeting disclosures of EU-wide stress test results.

1.6 Scope and Delimitations

This study concerns the abnormal stock returns of affected banks in response to the disclosure of the results of the five EU-wide stress tests listed in Table 1. However, the scope of the study can be further defined in terms of the stress test exercises, events, security types, and banks covered. Therefore, the relevant scope of each of these parameters is defined and delimited in more detail below.

The study covers the five EU-wide stress tests conducted by CEBS and EBA between 2010 and 2018. That is, all exercises available for this research. In addition, there was another EU-wide stress test before and after this study period that could not be taken into account: the first-ever EU-wide stress test in 2009 and the latest EU-wide stress test originally scheduled for 2020 but postponed to 2021 due to the COVID-19 pandemic. Given the time constraints of this study, the latter exercise could no longer be considered due to the late publication of the results (30 July 2021). In contrast, the 2009 EU-wide stress test could not be included for logistical reasons. In light of the ongoing global financial crisis, the exercise was deliberately designed to provide only aggregated information on the European banking system; accordingly, the CEBS did not disclose any results at bank level (CEBS 2009a). In fact, not even the names of the banks subjected to the 2009 EU-wide stress test were disclosed. However, because the research design (Section 5.3) of this study relied on bank-level results and the corresponding abnormal stock returns, the 2009 EU-wide stress test had to be excluded from the study. Furthermore, for the avoidance of doubt, this study does not cover the EU Capital Exercises carried out by the EBA between 2011 and 2014, nor the Comprehensive Assessments carried out by the ECB together with national competent authorities. Especially not the 2014 Comprehensive Assessment, which was performed in parallel with the 2014 EU-wide stress test before the ECB assumed its supervisory role in the Single Supervisory Mechanism (SSM).[7]

The scope of this study can also be delimited with regard to the type of event analysed. Some previous studies have examined banks' abnormal returns in relation to a variety of stress test-related events, such as the stress test announcement, various methodological clarifications, or the specification of dates and the overall timeline (*e.g.* Candelon and Sy 2015, Cardinali and Nordmark 2011, and Gerhardt and Vander Vennet 2017). In contrast, this study focuses solely on the main event of any EU-wide stress test, *i.e.* the disclosure of the results. This approach is consistent with other previous studies such as Alves *et al.* (2015), Georgoutsos and Moratis (2021), and Petrella and Resti (2013). The goal was to produce focused and meaningful research on the main event rather than on secondary side issues.

Regarding the type of securities covered, this study concentrates on the analysis of stocks. This is consistent with most previous studies on supervisory stress tests. However, studies on US stress tests in particular have also examined bonds, credit default swaps, and market microstructures such as implied volatilities and stock trading volumes (*e.g.* Fernandes *et al.* 2020, Flannery *et al.* 2017, and Morgan *et al.* 2014). The focus on stocks is based on the following conceptual reasoning. A basic requirement for studying price reactions is that the securities used are sensitive to new information. Stocks rank last in the capital structure in the order of repayment, making stock investors particularly sensitive to new information and prompting them to adjust prices quickly. This is in line with Fama's (1970) efficient market hypothesis and has been supported by empirical studies which have found that stocks are more responsive

[7] For more information on the EU Capital Exercises and the ECB Comprehensive Assessments, see ECB (2022) and EBA (2022), respectively. For specific analyses of abnormal stock returns in response to the 2014 Comprehensive Assessment, see Carboni *et al.* (2017), Georgescu *et al.* (2017), and Sahin and de Haan (2016).

to new information than credit default swaps (CDS) and bonds (Dang *et al.* 2015, Kajurová and Hvozdenská 2016, Norden and Weber 2009).

The decision to focus on the analysis of stocks means that the scope of the study had to be limited to those banks that have issued such securities. Therefore, of all banks that have been subjected to the relevant EU-wide stress tests, this study only covers banks that are stock corporations with publicly traded stocks. Accordingly, banks with other corporate forms (*e.g.* savings banks or cooperative banks) and banks whose stocks are not publicly traded (*e.g.* state banks, nationalised banks, or captive banks) were excluded from the analysis. This approach is formalised in the sampling procedure (Section 5.3.2.1.2), of which it forms an integral part.

1.7 Outline of the Study

This study is divided into eight chapters. After this introductory chapter, further information on the regulation and supervision of bank capital is given in Chapter 2. This includes the evolution of the Basel Capital Accords, the measurement of bank capital, and the rise of supervisory stress testing as a policy instrument, especially since the global financial crisis (2007-2009). The purpose of this background chapter is to present the regulatory environment for this study in its historical context.

The relevant literature is critically reviewed in Chapter 3. This concerns the body of empirical research on capital market reactions to supervisory stress test results and the theories, constructs, and debates underlying this study. The four theoretical areas covered in the literature review are: bank opacity and information uncertainty, informational efficiency of capital markets, rational choice theory and the risk-return tradeoff of investments, and Goodhart's law on financial policy indicators.

Building on this, two key steps are taken in Chapter 4. First, a new and coherent theoretical framework is created to link the different theoretical areas that underpin this study. Second, testable null and alternative hypotheses are developed on the basis of this framework in order to operationalise the research questions (Section 1.4).

The methodology used to test the hypotheses and answer the research questions is presented in Chapter 5. This chapter consists of two main parts: the research philosophy and the research design of this study. The first part describes, among other things, the objectivist ontological position and the empirical-positivist epistemological stance of this study. The second part explains and justifies the strategy (quasi-natural experimentation) and the methods used to collect and analyse the research data. This comprises the sampling procedure, the data-collection method, the extended event-study approach (which forms the common methodological basis of this study), and other research-question specific analysis and test methods. The second part also covers the confounding controls and robustness checks used to improve the validity and reliability of the study.

In Chapter 6, the empirical results are presented separately for each of the research questions. In addition to the economic and statistical significance of the results, this also includes a final statement as to whether the respective null hypothesis can be rejected in favour of the alternative hypothesis. The chapter also reports the results of the robustness checks and provides basic descriptive statistics of the dependent variables for all samples used in the study.

The empirical results are then discussed in Chapter 7. This involves interpreting the meaning of the results and discussing their significance and implications for research, supervisory policy, and investment practice. To ensure a thorough and complete discussion, the results are contextualised with the limitations of this study and synthesised with theory and with the findings from previous research. The discussion chapter thus paves the way for the final conclusions.

Finally, the study is concluded in Chapter 8. This includes a clear and definitive answer to each of the research questions posed in Section 1.4. In addition, the overall research process is summarised and reflected upon with emphasis on the contribution of this study to theory, methodology and methods, and supervisory policy and investment practice. The study concludes with recommendations for future research. Figure *1* illustrates the overall research process of this study.

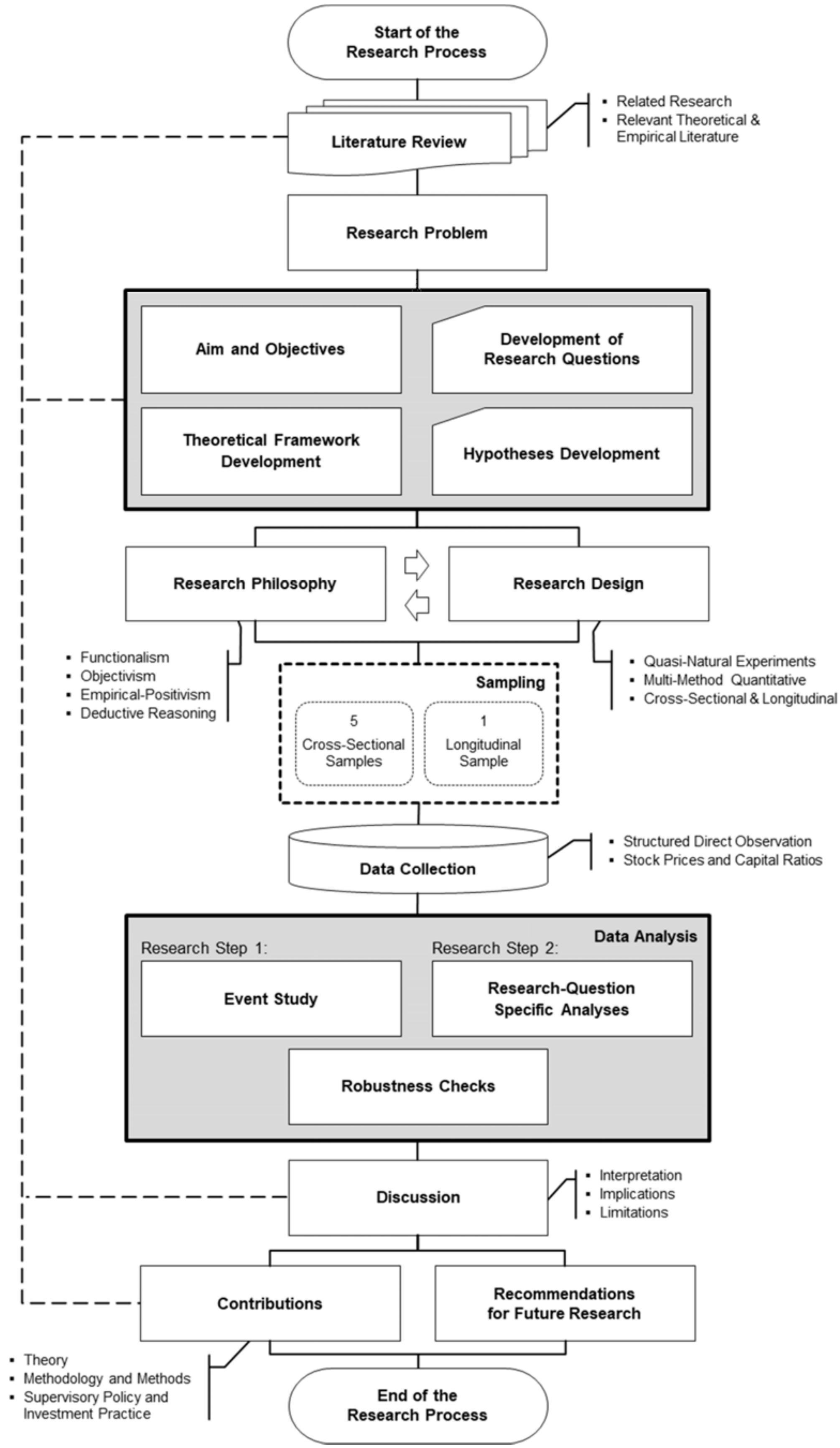


Figure 1. Overall research process of this study

Chapter 2
Regulation and Supervision of Bank Capital

2.1 Introduction

This chapter provides background information on the regulation and supervision of bank capital. This includes the evolution of the Basel Capital Accords, the measurement of bank capital, and the rise of supervisory stress testing as a policy instrument, especially since the global financial crisis (2007-2009). It also provides a broad understanding of the Basel concepts and principles relating to regulatory capital. The purpose of this chapter is to present the regulatory environment for this study in its historical context.

2.2 The Basel Capital Accords

This section outlines the evolution of the Basel Capital Accords as the international prudential standards for the measurement of bank capital. The Basel Capital Accords are a series of international prudential standards related to bank capital (commonly referred to as Basel I, II, and III) developed by the Basel Committee on Banking Supervision (BCBS). The Basel Capital Accords are not themselves binding or enforceable, but are implemented by national competent authorities (or through EU directives and regulations) and form the basis for national capital requirements. In order to introduce the basic concepts and principles of Basel capital regulation, the evolution of the Basel Capital Accords is outlined below.

Basel I

The original Basel Capital Accord was introduced in 1988 to set minimum capital requirements for banks. It had three main areas of regulation, which (1) introduced commonly accepted definitions for the constituents of regulatory capital, (2) linked capital requirements to risk through the introduction of risk weights (percentage factors used to weight the risk of bank assets based on five broad asset categories when calculating capital requirements), and (3) established a minimum capital requirement of eight percent of risk-weighted assets for internationally active banks. The BCBS (1988) decided that bank capital for supervisory purposes should be defined in two tiers: Tier 1 capital, also referred to as a bank's "core capital", was defined as equity and reported reserves from retained earnings after tax, *i.e.* the most permanent and loss-absorbing instruments. Tier 2 capital, also known as "supplementary capital", was defined as undisclosed reserves, revaluation reserves, general provisions or general loan-loss reserves, hybrid debt capital instruments, and subordinated term debt. The constituents of Tier 1 and Tier 2 capital, as well as their limits and restrictions, are further defined in Annex 1 to BCBS (1988). The Basel I minimum capital requirement mentioned above was to be met by Tier 1 capital and – up to a maximum of 50 percent of the total capital – by Tier 2 capital (BCBS 1988). This was the first attempt to establish a risk-based capital regime. Due to its widespread adoption by national competent authorities, Basel I quickly became the *de facto* standard for almost all banks.

One of the early criticisms of Basel I was that it only focused on credit risk. Therefore, the 1996 Market Risk Amendment required banks to calculate and apply capital charges for their market risk in addition to their credit risk (BCBS 1996). The BCBS (1996) defined market risk as the risk of losses in both on- and off-balance sheet items resulting from movements in market prices. To measure their market risk, banks could choose between the standardised approach (SA) and the internal models approach (IMA), which allowed banks to use their own internal risk management models, subject to supervisory approval.

To ensure consistency with the calculation of minimum capital requirements for credit risk, risk-weighted asset equivalents[8] were used to calculate market risk. That is, banks using the standardised approach had to apply percentage factors to

[8] The BCBS (1996) also refers to risk-weighted asset equivalents as "trading book notional risk weighted assets".

weight and calculate the general market risk and specific risk of each security across five categories: (1) interest rate related instruments, (2) equities, (3) foreign exchange, (4) commodities, and (5) options. Details of the risk weights and the overall calculation can be found in Part A of the 1996 Market Risk Amendment. To finally calculate their capital charge for market risk, banks had to arithmetically sum the risk-weighted market risks across the five categories above and multiply the result by 12.5 (*i.e.* the reciprocal of the eight percent minimum capital requirement), creating an explicit numerical link between the capital requirement calculations for credit and market risks.

Similarly, banks using the internal models approach had to use risk measures derived from their own internal risk management models to calculate their market risk. The internal models used had to meet certain qualitative and quantitative standards, cover an appropriate set of market risk factors, and be subjected to regular internal stress tests. Despite some discretion in specifiying risk factors, the BCBS (1996) required banks to cover at least the following risk factors: (1) interest rates, (2) equity prices, (3) exchange rates, (4) commodity prices, and (5) the risks associated with options. However, the total specific risk charge applied to debt securities or equities should in no case be less than half of the specific risk charge calculated under the standardised approach (BCBS 1996). Details on the required qualitative and quantitative standards, as well as the requirements for the specification of risk factors and internal stress tests can be found in Part B of the 1996 Market Risk Amendment. As with the standardised approach, banks had to arithmetically sum the risk measures obtained and multiply the result by 12.5 to ensure consistency with the calculation of minimum capital requirements for credit risk.

Regardless of whether banks opted for the standardized approach or the internal models approach, they could – at the discretion of national competent authorities – use an additional tier of short-term subordinated debt capital (Tier 3 capital) to meet part of the capital requirements for market risk (BCBS 1996).

Basel II

With the introduction of Basel II in 2004, the existing regulations were completely revised (BCBS 2004). While retaining the original capital definitions and the established minimum capital requirement of eight percent, Basel II introduced a three-pillar concept consisting of: minimum capital requirements (Pillar 1), supervisory review (Pillar 2), and market discipline (Pillar 3). Basel II also responded more thoroughly to criticism by including operational risk (alongside credit and market risk) and introducing more granular risk weights.

In addition, Basel II also reflected feedback from large and complex banks, who indicated that the risk weights and capital requirements under Basel I bore little resemblance to their internal risk assessment and capital allocation in risk management. Basel II responded to this feedback by introducing the internal ratings-based approach (IRBA) and advanced measurement approaches (AMA), which allowed banks to use internal models to determine capital requirements for credit and operational risk, subject to supervisory approval. Consistent with market risk, banks using the IRBA to calculate their capital requirements were required to stress test credit risk.

Basel III

The regulations introduced by Basel III in 2010 were clearly shaped by the recent global financial crisis. Accordingly, the new regulations provided for more extensive and stringent requirements and responded to regulatory deficiencies revealed by the crisis (BCBS 2010). More specifically, Basel III introduced two global liquidity standards (liquidity coverage ratio (LCR) and net stable funding ratio (NSFR)), a leverage ratio, better risk coverage, and significantly higher and better quality minimum capital requirements. Figure 2 illustrates the evolution of minimum capital requirements under Basel I, II, and III.

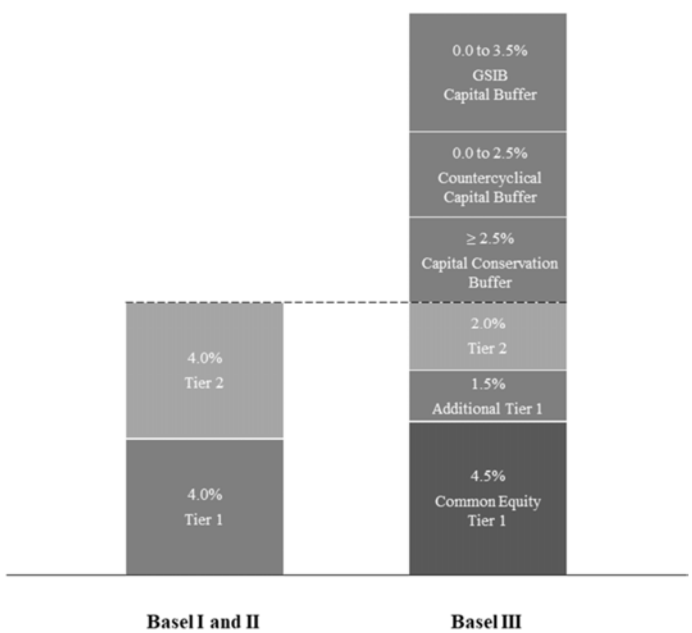


Figure 2. Evolution of minimum capital requirements under Basel I, II, and III

In 2017, additional standards were introduced to finalise the post-crisis reforms and complement the initial Basel III regulations (BCBS 2017a). These additional standards focused on reducing variability in banks' risk-weighted asset (RWA) calculations. To date, the Basel III regulations are the last published international capital requirements; due to repeated extensions, their implementation by the national competent authorities is currently still ongoing.

2.3 Basel Capital Ratios and Stress Testing

The Basel Capital Accords provide methodological guidance for calculating regulatory capital and regulatory capital ratios. In the EU, the relevant standards are implemented by the Capital Requirements Directives (CRD) and the Capital Requirements Regulation (CRR); their currently valid versions are codified in Directive 2013/36 (CRD IV) and Regulation 575/2013 (CRR). The Basel capital ratios serve, among other things, as a measure for bank-internal and supervisory stress tests. They thus represent the link between a bank's capital requirements and its performance in an EU-wide stress test, with the stressed (projected) capital ratio representing the bank's stress test result.

Since the beginning of EU-wide stress testing, Basel capital ratios have been used as a starting point for determining the banks' stressed (projected) capital ratios, which represent the result of the stress test. While the Tier 1 ratio was used in the 2010 EU-wide stress test, the Common Equity Tier 1 (CET 1) ratio has been used continuously since then (for the different capital ratio thresholds applied in the various stress tests, see Table 1). For this reason, the calculation of the CET 1 ratio is outlined below as an example (but the calculation of Basel capital ratios is very generic and can easily be adapted to the Tier 1 ratio or any other capital ratio). As a risk-based capital ratio, the CET 1 ratio is an expression of the relationship between capital and risk. It is formally calculated as

$$CET\ 1\ ratio = \frac{CET\ 1}{RWA}, \tag{1}$$

where $CET\ 1$ is a bank's Common Equity Tier 1 capital and RWA is its Risk-Weighted Assets.

Common Equity Tier 1 capital in the numerator of the CET 1 ratio is defined in Article 26 CRR as capital instruments, share premium accounts, retained earnings, accumulated other comprehensive income, other reserves, and funds for general banking risk. In order for most of these items to qualify as CET 1 capital, certain conditions need to be met, as set out in the remainder of Article 26 and Article 28 CRR. In general, CET 1 capital has a high loss-absorbing capacity as it does not have to be repaid, does not require dividend or interest payments, and ranks last in the order of repayment in the event of bankruptcy or insolvency proceedings.

The Risk-Weighted Assets in the denominator of the CET 1 ratio are defined in Article 92 CRR and include many different positions, mainly related to credit risk, but also to market, operational, and other risks. The applicable risk weights typically range between zero and 100 percent, but can also be several times higher for some particularly risky assets. As the name suggests, a bank's Risk-Weighted Assets result from the weighted average of the bank's assets. Banks that have been permitted to use internal models to calculate capital requirements (Section 2.2) use their own internal risk models to estimate the applicable risk measures.

This is also where EU-wide stress tests come into play. The methodology of EU-wide stress tests provides for the translation of macro-financial variables from their baseline and adverse scenarios into bank balance sheet losses, using a constrained bottom-up approach. Under this approach, banks use, within certain limits, their own internal risk models to map the macro-financial impact of the stress scenarios on their Basel capital ratio (typically the CET 1 ratio) assuming a static balance sheet.[9] The resulting stressed capital ratios (*i.e.* the stress test results) are used by the CEBS and the EBA to determine which banks have failed the stress test and what supervisory or recapitalisation actions should be taken based on the identified capital gaps. Once disclosed, a bank's stressed capital ratio can also be used by investors for assessment by comparing it to the bank's actual capital ratio from the most recent financial statement. The recent 2018 EU-wide stress test provides a useful example that can be used for illustrative purposes: Lloyds Banking Group, for example, reported a CET 1 ratio of 14.06% as of 31 December 2017 in its last financial statement before the stress test. This actual capital ratio compares to a stressed CET 1 ratio of 8.55% under the adverse

[9] Assuming a static balance sheet means that the size, composition, and risk profile of a bank's balance sheet are invariant throughout the time horizon of the exercise. In particular, no capital measures taken after the reference starting date are to be assumed.

scenario of the stress test (EBA 2020a).[10] That is, the macro-financial stress assumed in the adverse scenario of the 2018 EU-wide stress test translated into hypothetical balance sheet losses and a 5.51% reduction in the bank's CET 1 ratio. Since EU-wide stress tests examine a large number of banks at the same time, investors are not limited to isolated before-and-after comparisons, but can also make horizontal comparisons between a large number of banks. Figure 3 visualises the stress transmission model of EU-wide stress tests.

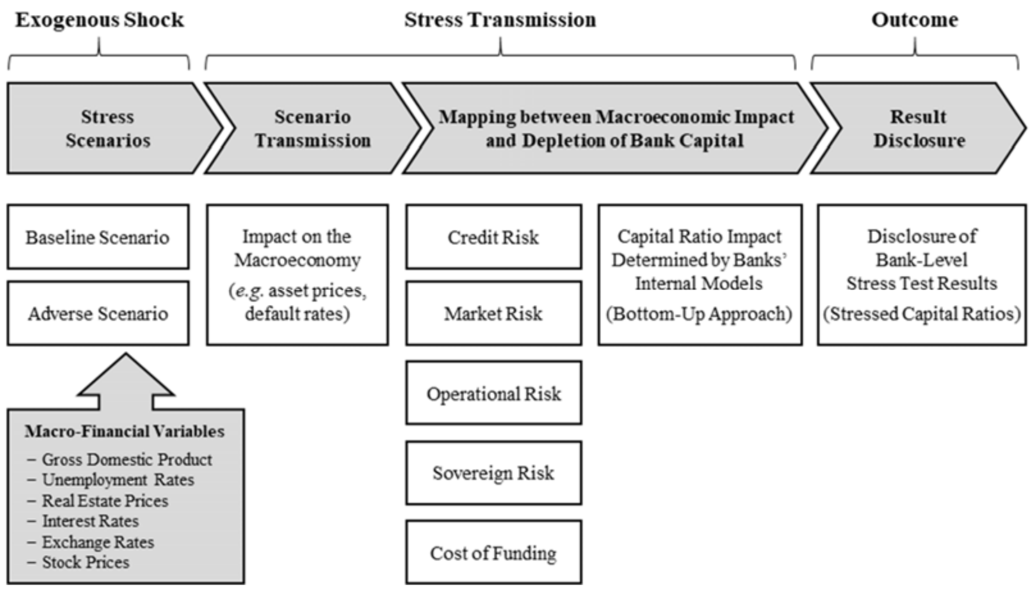


Figure 3. Stress transmission model

2.4 Supervisory Stress Testing

Stress tests *per se* are not a recent phenomenon. However, a distinction must be made between the risk management stress tests that were gradually introduced with the Basel Capital Accords, micro stress tests of individual banks, and modern supervisory (macro) stress tests that were established during the recent global financial crisis. Therefore, Section 2.4.1 first outlines the origins and development of stress testing. This is followed by an overview of modern supervisory stress testing in Section 2.4.2 and an outline of EU-wide stress testing in Section 2.4.3.

[10] In order to keep this illustrative example concise and informative, it is limited to presenting the stressed capital ratio under the adverse scenario and refrains from presenting the result under the milder baseline scenario.

2.4.1 The Origins and Development of Stress Testing

The first stress testing methods emerged in the risk management departments of large US banks in the early 1980s when risk managers began stressing interest rate risk in the banking book (Carhill 2009, Houpt and Embersit 1991, Sierra and Yeager 2004). Due to the gradual introduction of stress testing requirements under the Basel Capital Accords (Section 2.2), bank-internal stress tests became more established, but often did not go beyond the business or divisional level. This was confirmed by the Committee on the Global Financial System's (CGFS) occasional surveys of stress testing practices, which found that most banks did not introduce bank-level stress testing until after the 1996 Market Risk Amendment to Basel I (CGFS 2000, 2001).[11] A later survey found that the development and application of stress tests for credit risk significantly lagged those for market risk (CGFS 2005). In addition, sparse supervisory guidance led to inconsistencies in the design of stress scenarios across banks, thus limiting the ability of supervisors to make horizontal comparisons (Bookstaber *et al.* 2014, Schuermann 2014).

The shortcomings and lack of comparability between stress tests used by different banks contributed to the need and gradual development of macro stress tests (Lester *et al.* 2012, Schuermann 2014). The use of stress testing for macroprudential purposes started in 1999, when the International Monetary Fund (IMF) and the World Bank launched their joint Financial Sector Assessment Program (FSAP) in response to the Asian financial crisis (IMF 2020, World Bank 2020).[12] The crisis had shown that financial stress can easily spill over from a limited number of banks to other domestic and foreign institutions and spread rapidly through the global banking system (Anderson *et al.* 2018). The difference between micro- and macroprudential stress tests is that the former assess the idiosyncratic risks of individual banks, while the latter take a systemic risk perspective and simultaneously subject a number of banks to common stress factors (Anderson *et al.* 2018, Borio *et al.* (2014), Hirtle *et al.* 2009).

---

[11] For a summary of the main findings of the 2001 CGFS survey, see Fender *et al.* (2001). A more recent review of bank and supervisory stress testing practices can be found in BCBS (2017b).

[12] The mandate of the FSAP extends to IMF and World Bank member states and has a twofold objective: first, to assess the resilience of member states' financial sectors against adverse macro-financial conditions and, second, to assess the potential contribution of the financial sector to economic growth and development (IMF 2019).

Assessments under the FSAP are comprehensive exercises that rely on several key methods.[13] Stress testing was selected as the leading quantitative method because of its unique forward-looking nature, which distinguished it from balance sheet-based indicators such as CAMELS[14] that were available at the time (Adrian *et al.* 2020). While early FSAP stress tests were mere single-factor sensitivity analyses (Blaschke *et al.* 2001, Čihák 2007, Moretti *et al.* 2008), more recent exercises have used a multitude of sophisticated methods (IMF 2012, IMF 2014a, Jobst *et al.* 2013).[15] Over time, FSAP stress tests and the related literature have repeatedly played a leading role in the development of innovative stress testing features. One of the distinguishing features compared to similar exercises is the fairly broad scope of FSAP stress tests, which can include parts of the non-banking sector such as insurance, pension funds, corporates, and households (Adrian *et al.* 2020, Ong and Čihák 2014).

As an integral part of the FSAP, stress tests have been carried out continuously since the start of the programme (IMF 2000, Jones *et al.* 2004, Baudino *et al.* 2018). In light of the global financial crisis, the IMF integrated the FSAP into its ongoing surveillance of the international monetary system in 2010 and required member states with systemically important financial sectors to undergo FSAP assessments every five years (IMF 2010). In 2013, the method for determining systemic importance was revised, placing greater emphasis on banks' interconnectedness rather than on their size. As a result, the number of member states with systemically important financial sectors increased from 25 to 29 (IMF 2014b).[16] By mid-2018, IMF and World Bank had carried out a total of 346 FSAP assessments in 173 member states (Baudino *et al.* 2018).[17]

---

[13] The key methods of the FSAP include stress testing, macroprudential indicators (MPI), and advanced methods for the assessment of standards and financial sector codes (IMF 2000).

[14] The CAMELS rating system is used by various banking authorities to assess the overall condition of banks. It it based on a ratio analysis of banks' financial statements and assesses the following six areas that form the acronym: capital adequacy, asset quality, management, earnings, liquidity, and sensitivity (Lopez 1999).

[15] Čihák (2007) offers a useful entry point into the literature on early FSAP stress testing. For a comprehensive overview of more recent FSAP stress tests, see Adrian *et al.* 2020 and Ong (2014). Caprio (2018) provides an independent assessment of the FSAP including stress testing.

[16] The initial 2010 list comprised the following IMF member states: Australia, Austria, Belgium, Brazil, Canada, China, France, Germany, Hong Kong, India, Ireland, Italy, Japan, Korea, Luxembourg, Mexico, Netherlands, Russia, Singapore, Spain, Sweden, Switzerland, Turkey, United Kingdom, and United States. After the 2013 methodological revision the initial list was complemented by Denmark, Finland, Norway, and Poland. For further information on the effective list of IMF member states with systemically important financial sectors, see IMF (2014c).

[17] Although the FSAP is a joint program of IMF and World Bank, FSAP missions in advanced economies are the sole responsibility of the IMF, whereas missions in developing and emerging economies are the joint responsibility of IMF and World Bank (IMF 2019).

Encouraged by the stress-testing experience gained from their participation in FSAP assessments, national banking authorities started to develop their own, independent stress-testing frameworks in the early 2000s (Baudino 2009, Dent *et al.* 2016, Sorge 2004). At that time, the main objective was to create microprudential stress tests that were workable at the supervisory level and the would allow for the analysis of a bank's sensitivity to certain stress factors. The literature on early supervisory stress testing approaches is extensive, for example Bunn *et al.* (2005), Čihák (2004), Kalirai and Scheicher (2002), Mawdsley *et al.* (2004), and Sorge and Virolainen (2006). Useful entry points to the literature are the Risk Assessment Model for Systemic Institutions (RAMSI) of the Bank of England (Alessandri *et al.* 2009), the Systemic Risk Monitor (SRM) of the Austrian Central Bank (Boss *et al.* 2006), and the macro stress testing framework of the ECB (Dees *et al.* 2017, Henry and Kok 2013).

2.4.2 Modern Supervisory Stress Testing

Modern supervisory stress testing, as it is used today, dates back to the Supervisory Capital Assessment Program (SCAP). The SCAP was conducted by US federal banking supervisors[18] from February to May 2009 (Fed 2009a) at the height of the global financial crisis shortly after the bankruptcy of Lehman Brothers in September 2008. The time was characterised by a high level of uncertainty and a lack of confidence among investors about the capital adequacy of individual banks and about the overall stability of the banking system (Hirtle *et al.* 2009, Hirtle and Lehnert 2015, Morgan *et al.* 2014, Schuermann 2014). Existing regulatory approaches that used to be informative (Basel capital ratios) revealed methodological problems[19] and were therefore no longer credible and heavily discounted by the market (Hirtle and Lehnert 2015, Morgan *et al.* 2014, Schuermann 2014).[20] In a speech on the launch of the SCAP then-Fed Chairman Ben Bernanke said:

[18] The design, implementation, and execution of the SCAP was a joint effort of the Board of Governors of the Federal Reserve System, the Federal Reserve Banks, the Federal Deposit Insurance Corporation (FDIC), and the Office of the Comptroller of the Currency (OCC) (Fed 2009a).

[19] For a detailed discussion of methodological weaknesses in the calculation of Basel capital ratios and the role of supervisory stress testing in mitigating these weaknesses, see Wall (2014a, 2014b).

[20] Furlong (2011) and Haldane (2011) have provided insights into the extent to which investors discounted banks' reported equity capital by contrasting book-value capital ratios (as promoted through the Basel Process) with market-value capital ratios; both studies have shown that capital ratios based on market values were generally much lower than those based on book values.

> The loss of confidence we have seen in some banking institutions has arisen not only because market participants expect the future loss rates on many banking assets to be high, *but because they also perceive the range of uncertainty surrounding estimated loss rates as being unusually wide.* (Bernanke 2009, para. 5, emphasis added)

As can be seen from Bernanke's quote, the inherent opacity of banks (Section 3.4.2) was particularly pronounced during the crisis. This is consistent with the stated aim of the SCAP to reduce investor uncertainty about the amount and quality of banks' capital (Fed 2009a). More precisely, the SCAP was designed to assess whether the 19 largest domestic banks[21] were sufficiently capitalised to survive a period of macro-financial stress that would be longer and more adverse than expected at that time (Flannery *et al.* 2017, Hirtle *et al.* 2009, Hirtle and Lehnert 2015, Morgan *et al.* 2014). The ultimate goal of the SCAP was to restore investor confidence in individual banks and the banking system as a whole, in order to facilitate the recapitalisation of banks found to be undercapitalised (Hirtle and Lehnert 2015, Morgan *et al.* 2014, Petrella and Resti 2016).

Achieving this goal required an unprecedented level of transparency (Flannery *et al.* 2017, Hirtle *et al.* 2009, Morgan *et al.* 2014), a novel combination of micro- and macroprudential stress-testing approaches (Hirtle *et al.* 2009, Hirtle and Lehnert 2015, Ong and Pazarbasioglu 2014),[22] and the introduction of a number of innovative stress testing features (Hirtle *et al.* 2009, Hirtle and Lehnert 2015, Schuermann 2014). Table 2 provides a list of the SCAP's most important innovative features.

[21] All US-owned bank holding companies (BHC) with assets exceeding USD 100 billion as of 31 December 2008. Together, the 19 SCAP banks accounted for two-thirds of the assets and for more than half of the loans in the US banking system (Fed 2009a).

[22] The macroprudential objective of the SCAP to mitigate systemic tail risk relied on the microprudential assessment of each participating bank's idiosyncratic risk, which allowed for differentiated analyses and bank-specific supervisory actions (Hirtle *et al.* 2009, Hirtle and Lehnert 2015, Ong and Pazarbasioglu 2014).

Table 2
Innovative Stress-Testing Features introduced by the SCAP

Design Element	**Pre-SCAP**	**Post-SCAP**
Prudential policy	Microprudential	Micro- and macroprudential
Stress factors	Single-factor sensitivity analysis	Macro-financial scenarios with multiple stress factors
Relevant cash flows	Losses	Losses and revenues
Time horizon	Static snapshot of the *status quo*	Dynamic and forward-looking analysis
Result measure	Losses	Stressed capital ratios
Recapitalization	Private capital	Private and public capital
Disclosure policy	Results were kept confidential	Results are publicly disclosed

Note. This table highlights the most innovative features of the Supervisory Capital Assessment Program (SCAP) by contrasting common pre- and post-SCAP stress test design elements. The information in this table is based on Hirtle *et al.* (2009), Hirtle and Lehnert (2015), and Schuermann (2014).

The most notable innovation of the SCAP was the public disclosure of bank-level stress-test results,[23] which marked a significant departure from the established supervisory practice of treating bank-specific information confidential (Bernanke 2013, Hirtle *et al.* 2009, Hirtle and Lehnert 2015, Schuermann 2014). This is also reflected in the white paper on the SCAP results:

> The decision to depart from the standard practice of keeping examination information confidential stemmed from the belief that greater clarity around the SCAP process and findings will make the exercise more effective at reducing uncertaintiy and restoring confidence in our financial institutions. (Fed 2009b, p. 1)

In this context, Morgan *et al.* (2014) highlighted the horizontal nature of the SCAP – that is the simultaneous assessment of all participating banks based on a common set of scenarios and analysis methods – which made the bank-level results particularly useful for determining the *relative* value of banks.[24] Similarly, Hirtle and Lehnert (2015) argued that the assessment of banks' current capitalisation against forward-looking stress scenarios helped investors to distinguish resilient banks from banks that are more vulnerable to deteriorating macro-financial conditions.

In general, there is a broad consensus among researchers that the SCAP was successful in achieving its aim to reduce uncertainty and restore investor confidence (Bernanke 2013; Bookstaber *et al.* 2014; Candelon and Sy 2015; Fernandes *et al.* 2020;

[23] For the results of the SCAP, see Fed (2009b).

[24] In contrast, ordinary supervisory assessments take place over time and offer little opportunity for comparison across banks (Morgan *et al.* 2014).

Flannery *et al.* 2013, 2017; Hirtle *et al.* 2009; Sahin *et al.* 2020; Morgan *et al.* 2014). This success in reducing the opacity of banks made the SCAP a role model for many subsequent stress-testing programmes, particularly in the US and the EU.[25]

2.4.3 EU-Wide Stress Testing

The US SCAP not only marked a critical turning point in the global financial crisis (Bernanke 2013, Bookstaber *et al.* 2014, Langley 2013), but also led to a permanent change in the utilisation and design of supervisory stress tests (Hirtle and Lehnert 2015, Kapinos *et al.* 2018, Schuermann 2014). Shortly after the completion of the SCAP, the EU's Economic and Financial Affairs Council (ECOFIN) mandated the Committee of European Banking Supervisors (CEBS) to develop and coordinate an EU-wide stress test (CEBS 2009a). In this first attempt at EU-wide stress testing, the CEBS had already adopted most of the innovative features introduced by the SCAP (Table 2), but did not disclose any bank-level results (CEBS 2009a, 2009b; Hirtle and Lehnert 2015).[26] As noted by the CEBS (2009a, para. 5) the purpose of the exercise was "to increase the level of aggregate information among policy makers", not "to identify individual banks that may need recapitalization" (CEBS 2009a, para. 3).

This approach was already modified in December 2009, when ECOFIN – in view of the worsening European sovereign debt crisis (Ong and Pazarbasioglu 2014, Schuermann 2014)[27] – mandated the CEBS to carry out another "extended" EU-wide stress test in 2010, which also took into account the recapitalisation of individual banks

[25] The SCAP was never repeated, but developed into the Comprehensive Capital Analysis and Review (CCAR) in early 2011 and was supplemented by the Dodd-Frank Act Stress Tests (DFAST) in 2013; since then, both programmes have been carried out regularly (Flannery *et al.* 2017, Hirtle and Lehnert 2015, Petrella and Resti 2016). For a comprehensive overview of the prevailing approach to supervisory stress testing in the US, see Fed (2020a, 2020b) and Lesambo (2020).

[26] There was also no common public capital backstop that would be equivalent to the Capital Assistance Program (CAP) of the US Department of the Treasury (Enria 2018, Ong and Pazarbasioglu 2014, Petrella and Resti 2016). For aggregate results of the 2009 EU-wide stress test, see CEBS (2009b).

[27] In light of the European sovereign debt crisis, several studies have focused on the relationship between banks' exposure to sovereign debt and EU-wide stress tests: For a critical view on how sovereign debt exposure was accounted for in the 2010 EU-wide stress test, see Blundell-Wignall and Slovik (2010). Based on the 2010 and 2011 EU-wide stress test results, Bischof and Daske (2013) examine the effect of bank-specific information disclosure on the level of subsequent voluntary disclosure. Using exposure data from the 2011 EU-wide stress test, Greenwood *et al.* (2015) show how a sovereign debt shock can cause distress for individual banks and how financial contagion can propagate through the banking system.

(CEBS 2010a, 2010b).[28] As a result, all EU-wide stress tests since 2010 have continuously disclosed both aggregate *and* bank-level results.

Since the dissolution of the CEBS in 2011 and its replacement by the newly founded European Banking Authority (EBA), EU-wide stress tests have been carried out under Article 32 of the EBA Regulation (EU) 1093/2010. Under this new regime, the EBA has established an integrated stress-testing programme that foresees running EU-wide stress tests roughly every two years. Beginning with the 2011 EU-wide stress test, the EBA has steadily increased the number of bank-specific data points that have supplemented the disclosure of bank-level results (Table 1), thus enabling “the market to ‘check the math’.” (Schuermann 2014, p. 728). These supplementary data included banks’ exposures by asset class, geography, and duration band (Petrella and Resti 2013, Schuermann 2014, Wall 2014a). In this context, Quagliariello (2020) noted that detailed disclosure of exposure data improves the understanding of analysts and investors of bank-specific risk drivers. According to former EBA chairman Andrea Enria, “the disclosure of bank data has been consistently commended by market participants” (Enria 2018, p. 10).[29] In addition, the EBA prepared convenience presentations for analysts in which the results of EU-wide stress tests were explained and which were published at the time the results were disclosed (EBA 2011a, 2016a, 2018a).[30] In general, the EBA aims for transparency throughout the stress testing process by providing timely information and publishing numerous technical documents.[31]

All of this is evidence that the EBA does indeed place more emphasis on market discipline when conducting EU-wide stress tests, as has been repeatedly signaled to the market (EBA 2011b, 2020c; Enria 2018). In a recent discussion paper the EBA stated that “transparency allows markets…to gain information from the supervisory stress test. It also fosters market discipline by enabling market participants to review the stress test results of banks” (EBA 2020c, p. 13).

---

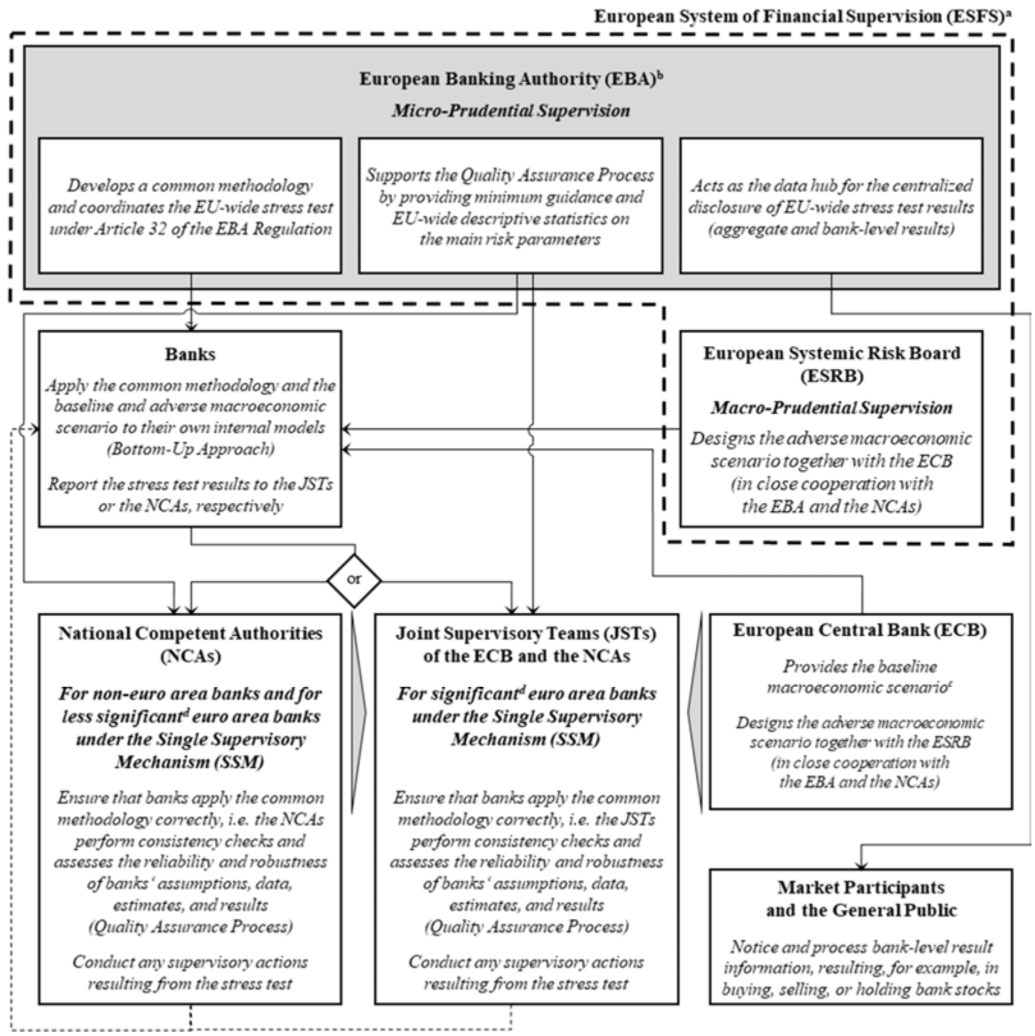
28 The ECOFIN mandate explicitly aimed at “the dependence of EU banks on public support and on the amount of capital available for further lending” (ECOFIN 2009, p. 16).

29 For a similar view, see EBA (2020c).

30 The results of the 2014 EU-wide stress test (which was conducted in parallel with the 2014 comprehensive assessment of the ECB before assuming its new SSM supervisory role) were conveniently summarised in a database and explained by interactive tools on the EBA website (EBA 2020b).

31 For example, by announcing new EU-wide stress tests or the date and time of result disclosure and by publishing methodological notes, templates, and summary reports.

Although the EBA, and previously the CEBS, performs an important coordination role, it draws on a large institutional framework to conduct EU-wide stress tests. This framework includes, among others, the European Systemic Risk Board (ESRB), the European Central Bank (ECB), and national competent authorities (NCAs). Figure *4* shows the components and functions of the institutional framework of EU-wide stress tests.



[a] The European System of Financial Supervision (ESFS) as bordered above is complemented by the European Securities and Markets Authority (ESMA) and the European Insurance and Occupational Pensions Authority (EIOPA) [b] Before the EBA was established on 1 January 2011, EU-wide stress test exercises were coordinated by the Committee of European Banking Supervisors (CEBS) [c] Until the 2018 EU-wide stress test, the macroeconomic baseline scenario was based on the European Commission's (EC) biannual economic forecasts [d] A bank that meets any of the following four conditions is deemed significant under the SSM: (1) the bank's assets exceed € 30 bn , (2) the bank's assets exceed 20% of its home country's gross domestic product (GDP), unless the assets are below € 5 bn , (3) the bank is considered significant by the ECB with regard to its home country's economy, and (4) the bank has requested or received direct public financial assistance from the European Financial Stability Facility (EFSF) or the European Stability Mechanism (ESM)

Figure 4. Institutional framework of EU-wide stress tests

2.5 Summary

In this chapter, background information on the regulation and supervision of bank capital was provided to place the regulatory environment for this study in its historical context. This included an overview of the evolution of the Basel Capital Accords and an outline of the calculation of Basel capital ratios. In addition, the origins and development of stress testing were presented. This was followed by an introduction to modern supervisory stress tests, starting with the SCAP, which not only marked a critical turning point in the global financial crisis, but also led to a lasting change in the utilisation and design of supervisory stress tests. Finally, an overwiew of EU-wide stress testing was given, with a focus on the institutional framework and EBA's stress test results disclosure policy.

Chapter 3

Literature Review

3.1 Introduction

This chapter critically reviews the literature relevant to this study. The purpose of the study was to examine banks' abnormal stock returns in response to the disclosure of EU-wide stress test results (Section 1.3). More specifically, the objectives of the study were threefold. First, to quantify the abnormal stock returns caused by the results disclosures in the cross-section of the sample banks. Second, to determine the shape of the functional relationship curve between stress test results and abnormal stock returns at the bank level. Third, to determine whether the informational value of EU-wide stress test results has been intertemporally stable or has been subject to a downward trend across the various exercises.

Against this background, five bodies of literature were reviewed: (1) related research on capital market reactions to supervisory stress tests, (2) bank opacity and information uncertainty, (3) informational efficiency of capital markets, (4) rational choice theory and the risk-return tradeoff of investments, and (5) Goodhart's law and the intertemporal instability of financial policy indicators. To organise the relevant literature in the context of this study, an overview is first given in Section 3.2. This is followed by detailed thematic reviews in Sections 3.3 to 3.7. Section 3.8 summarises the literature review.

3.2 Overview of the Relevant Literature

This section provides a basic overview of the five bodies of literature relevant to this study. In addition, the literature search process is outlined. This includes the academic databases and search engines used, as well as the literature search methods and the time period of the literature review.

The review begins with related research on capital market reactions to supervisory stress tests conducted in the US and the EU (Section 3.3). Emphasis is placed on presenting the current state of research and identifying contradictions and research gaps that deserve more attention. This also helps position this study relative to previous studies in the field. Since previous studies have made little reference to the theoretical foundations of their research, a fundamental review of the relevant theories, constructs, and debates was carried out next (see also the development of the dedicated theoretical framework in Section 4.2).

The entry point for the theory-oriented review was the long-standing debate on the opacity of banks and the uncertainty of information about bank risks (Section 3.4). At the heart of this debate is whether banks' balance sheets are more opaque to outside investors than those of firms in other sectors. There is now a broad consensus among researchers that banks are indeed particularly opaque (see Table 5). This arguably justifies the regulation of bank transparency through supervisory stress testing as a means of generating and disclosing information on bank risk to the public.

Another area of review was therefore how new public information is reflected in stock prices. In other words, the extent to which stock prices are efficient according to Fama's (1970) efficient market hypothesis (Section 3.5). The review of the market efficiency literature provides the theoretical basis for stock pricing and discusses how market efficiency can be tested and the problems involved. This forms the basis for testing stock price adjustments (abnormal returns) to new public information (results of EU-wide stress tests) about otherwise opaque banks.

Building on this, rational choice theory and the risk-return tradeoff of investments are reviewed to create a theoretical construct that can be used to fit the functional relationship curve between EU-wide stress test results and corresponding abnormal stock returns (Section 3.6). That is, a construct that can be used to theoretically model how rational investors might use and respond to the disclosure of EU-wide stress test results (*i.e.* enhanced insights into the risks and prospects of their stock investments) to maximise their utility.

Finally, the literature on Goodhart's (1975) law on financial policy indicators was reviewed to provide a theoretical basis for investigating the intertemporal stability of the informational value of EU-wide stress test results (Section 3.7). Goodhart's law states that statistical relationships break down when used for regulatory purposes. It therefore provides an appealing theoretical basis for investigating whether the value of the information contained in EU-wide stress test results began to decline once the repetitive nature of the exercises became apparent.

The literature reviewed consists of a variety of primary and secondary sources, including textbooks, regulatory documents, dissertations, speech transcripts, and most importantly, leading academic journals. To identify the relevant literature, two systematic search methods were used: building blocks search and cited reference search. The literature was mainly obtained from libraries and the following academic databases and search engines: Research Papers in Economics (RePEc), ResearchGate, EconBiz, SSRN, JSTOR, Science Direct, and Google Scholar. No specific delimitation period was used for the literature review in order not to prevent the inclusion of relevant sources due to an arbitrary time criterion.

In the following sections, the above bodies of literature are reviewed in detail. That is, they identify the relevant research gaps and discuss the theories, constructs, and debates underlying this study. In this respect, they also pave the way for the development of the theoretical framework and the formulation of empirically testable hypotheses in Chapter 4.

3.3 Related Research

The literature review begins with a critical discussion of existing research related to this study. That is, studies within the relevant “Capital Market Reactions” line of research (see Section 1.2). In order to create a basis for the discussion, the characteristic features of this research line are first outlined in Section 3.3.1. This is followed by a discussion of existing studies on US and EU-wide stress tests in Section 3.3.2. The review also covers studies on US stress tests in order to provide a complete picture of the current state of research and to help identify open research problems in EU-wide stress tests via direct comparison. The purpose of this section is therefore to discuss the current state of the relevant research line and to substantiate the research problems already announced in Section 1.2.

3.3.1 Characteristic Features of the Research Line

The research line "Capital Market Reactions" (see Section 1.2) can be roughly divided into three groups: studies on US stress tests (Fernandes *et al.* 2020, Flannery *et al.* 2017, Morgan *et al.* 2014, Sahin *et al.* 2020), on both US and EU-wide stress tests (Ahnert *et al.* 2020, Candelon and Sy 2015), and on EU-wide stress tests only (Alves *et al.* 2015, Cardinali and Nordmark 2011, Georgescu *et al.* 2017, Georgoutsos and Moratis 2021, Gerhard and Vander Vennet 2017, Petrella and Resti 2013). The latter two groups most closely resemble this study and are therefore collectively referred to as "previous studies".

The public nature of supervisory stress tests in the US and EU has prompted researchers to examine the informational value of stress test-related events (*e.g.* stress test announcements, methodological and policy clarifications, and results disclosures) by analysing the corresponding price reactions of various capital market instruments. The overarching hypothesis was that new, formerly confidential, supervisory information would be reflected in the securities prices of the affected banks once it became public. Although this implies testing the semi-strong form of Fama's (1970) efficient market hypothesis, existing studies on US and EU-wide stress tests have made hardly any reference to the underlying theory. Accordingly, there is still no theoretical framework upon which empirically testable hypotheses could be based – a gap that is addressed in Chapter 4 of this study.

Methodologically, there is consensus among researchers that event studies are the method of choice. With few exceptions, most studies have based their analysis on CDS and stock prices (Fernandes *et al.* 2020, for example, also used stock and bond bid-ask spreads, stock option implied volatilities, and stock trading volumes). The event study design requires the selection of an appropriate normal return-generating model (asset pricing model) so that actual (observed) returns can be compared against normal (expected) returns. However, despite the problems associated with an inappropriate choice of normal return-generating model, none of the existing studies on US and EU-wide stress tests have performed a proper model selection procedure.[32] Previous studies have thus typically chosen their asset pricing models arbitrarily or resorted

[32] Specifically, validity problems can arise from the *joint-hypothesis problem* (Fama 1970, 1991), also known as the *bad-model problem* (Butler and Wan 2010, Jarrow and Larsson 2012, Min and Kim 2012). For more information, see Section 3.5.3.

to the convenient Market Model, leading to a "model monoculture" (see Table 3). Only Georgoutsos and Moratis (2021) used both a single-factor model (Makret Model) and the Five-Factor Model by Fama and French (2015), albeit only for robustness check purposes and with no apparent model selection procedure. In order to avoid validity problems due to an inappropriate normal return-generating model, a systematic model selection procedure was carried out in this study for the first time ever in the context of stress test-related investigations (Section 5.3.2.1.3).

This is the common ground on which studies in the "Capital Market Reactions" line of research have been built. All relevant studies are presented and discussed in the following section.

3.3.2 Studies on US and EU-Wide Stress Tests

After the SCAP was conducted in 2009, Morgan *et al.* (2014) was the first study to examine whether the SCAP produced valuable new information for investors. Using a standard event study methodology across a variety of stress test-related events, they found that CDS and stock prices responded only to policy clarification events and the disclosure of the results. They also found that while investors could largely predict which banks would have capital shortfalls, investors were surprised by the size of the shortfalls and used this information to re-evaluate the securities of affected banks. The observed security price reactions thus corresponded to the unexpected component of the information generated by the SCAP.

Subsequent studies of US stress tests have typically included the SCAP, but have gradually added the CCAR and DFAST stress tests that emerged from it. Under this setup, Candelon and Sy (2015), Fernandes *et al.* (2020), and Sahin *et al.* (2020) have found that market reactions to the SCAP were generally much stronger than those to subsequent US stress tests. The latter argued that this result could be explained by investors' increased need for credible information in times of financial distress. This argument resonates with the ongoing debate about stress test results disclosure policy and whether it should differ in normal times and times of crisis (Alvarez and Barlevy 2021, Goldstein and Leitner 2018, Goldstein and Yang 2019, Ong and Pazarbasioglu 2014, Schuermann 2014). An alternative explanation is that perverse incentives in US stress tests led banks to exploit learning effects, window dressing, and other suboptimal myopic behaviour to "game the system", making US stress test results less informative over time (Cornett *et al.* 2020, Glasserman and Tangirala 2016, Goldstein and Sapra 2014). Kok *et al.* (2019) and Quagliariello (2020) have recently raised similar concerns in the European context but have not explored them further. To date, there is not a single longitudinal study that has empirically examined the change or dynamics of the informational value of EU-wide stress test results across the various exercises.

In contrast to Candelon and Sy (2015), Fernandes *et al.* (2020), and Sahin *et al.* (2020), the results of Flannery *et al.* (2017) showed that disclosures of US stress test results have consistently provided investors with significant amounts of valuable new information. They argued that these divergent findings may be due, at least in part, to a conceptual flaw in the standard approach to event studies used by nearly all other studies in the field. This flaw consists in the assumption that the securities of all affected banks react in the same direction, so that average cumulative abnormal returns ($\overline{CARs}$) of zero imply no impact. Therefore, standard event studies fail to distinguish between non-events and significant events with opposite-signed security reactions that cancel each other out. To address this flaw, Flannery *et al.* (2017) used, among other things, average *absolute* cumulative abnormal returns ($|\overline{CARs}|$) as a non-directional measure that accounts for positive and negative information effects without offsetting them. This approach was later adopted by Georgoutsos and Moratis (2021) and applied in the European context, albeit only to a small sample of two stress tests (*i.e.* the 2016 and 2018 EU-wide stress tests). Despite the above differences in the magnitudes of $\overline{CARs}$ and $|\overline{CARs}|$ across the various US stress tests, the studies overall agreed that public disclosure of stress test information can help reduce information asymmetries and increase transparency in the banking sector (Candelon and Sy 2015, Fernandes *et al.* 2020, Flannery *et al.* 2017, Morgan *et al.* 2014, Sahin *et al.* 2020).

While US studies have typically gradually incorporated new stress tests in order to produce coherent analyses, the body of research on EU-wide stress tests is a patchwork of different stress tests and event-study specifications. As a result, the collective evidence from previous studies on EU-wide stress tests is inconclusive and sometimes contradictory (see summary of key findings in Table 4). For example, while Cardinali and Nordmark (2011) and Candelon and Sy (2015) have found little impact from disclosing the results of the 2010 EU-wide stress test, Alves *et al.* (2015) reported large and statistically significant $\overline{CAR}s$ for the same event.[33] Another striking example is the contradiction in the results of Candelon and Sy (2015) and Petrella and Resti (2013) for the results disclosure of the 2011 EU-wide stress test. Despite their very similar research design, Candelon and Sy (2015) found $\overline{CAR}s$ of -0.019% that were significant at the .05 level, while Petrella and Resti (2013) reported $\overline{CAR}s$ of 0.8% that were not statistically significant.[34] Thus, there is as yet no coherent understanding of the informational value contained in the results of EU-wide stress tests. This suggests the need for a study that determines the informational value of all EU-wide stress tests available for research, based on a systematic and consistent methodology. Table 3 provides an overview of the different EU-wide stress tests and event-study specifications used in previous studies.

[33] More specifically, Alves *et al.* (2015) found $\overline{CAR}s$ of 4.37% for a six-day event window (0, +5) that were significant at the .01 level, and $\overline{CAR}s$ of 3.90% for an 11-day event window (0, +10) that were significant at the .05 level.

[34] In fact, the only difference between the two studies was the asset pricing model used to estimate normal (expected) returns: Candelon and Sy (2015) used the Capital Asset Pricing Model (CAPM), while Petrella and Resti (2013) used the Market Model. This is further evidence of the importance of model selection and choosing an appropriate normal return-generating model (see Section 3.3.1).

Table 3
Overview of Previous Studies on EU-Wide Stress Tests

Study	EU-Wide Stress Test	Asset Pricing Model	Estimation Period[a]	Event Window[b]
Ahnert *et al.* (2020)	2010, 2011, 2014, 2016, 2018[c]	MM	120	3 days (-1, +1)
Alves *et al.* (2015)	2010, 2011	MM	120	5 days (-1, -5) 6 days (0, +5) 9 days (-9, -1) 11 days (0, +10)
Candelon and Sy (2015)	2010, 2011[d]	CAPM	200	5 days (-2, +2)
Cardinali and Nordmark (2011)	2010[e]	MM	262[f]	3 days (-1, +1) 11 days (-5, +5) 21 days (-10, +10)
Georgescu *et al.* (2017)	2016[g]	MM	30[h]	2 days (+1, +2)
Georgoutsos and Moratis (2021)	2016, 2018	MM, FF-5F	60, 120	2 days (0, +1) 3 days (-1, +1) 8 days (0, +7) 15 days (-7, +7)
Gerhardt and Vander Vennet (2017)	2011	MM	120	1 day (0) 2 days (0, +1) 3 days (0, +2)
Petrella and Resti (2013)	2011	MM	200	3 days (0, +2) 4 days (-2, +1) 5 days (-2, +2)

Note. This table provides an overview of previous studies on EU-wide stress tests, in particular the event-study specifications used. MM = Market Model. CAPM = Capital Asset Pricing Model. FF-5F = Fama and French (2015) Five-Factor Model.

[a] In trading days. [b] In trading days, where the parentheses specify the number of trading days before and after the event date t_0. [c] Ahnert *et al.* (2020) also examined the seven annual CCAR stress tests conducted from 2012 to 2018. [d] Candelon and Sy (2015) also examined the 2012 EU Capital Exercise, the SCAP, and the two CCAR stress tests conducted in 2012 and 2013. [e] Cardinali and Nordmark (2011) also examined events from the 2011 EU-wide stress test, but not the results disclosure event, which was too late for the study to consider. [f] This corresponds to a full trading year. [g] Georgescu *et al.* (2017) also examined the 2014 ECB Comprehensive Assessment. [h] This is a rolling 30 trading-day estimation window.

Since Ahnert *et al.* (2020) stands out from the overview in Table 3 due to the large number of EU-wide stress tests examined, it should be noted that they used an integrated research approach to analyse a combined cross-jurisdictional sample of seven US and five EU-wide stress tests. Therefore, their results cannot be compared directly with those of other studies. In addition, they focused on how market reactions differed for banks that passed the stress tests versus banks that failed the stress tests (as measured by the official capital ratio thresholds, which is problematic because the EBA has dispensed with pass/fail exercises since the 2016 EU-wide stress test and has not used capital ratio thresholds since then). Their analysis showed that, on average, passing banks saw significantly positive abnormal stock returns of 0.50%, while failing banks saw significantly negative abnormal stock returns of -1.74% on the results disclosure date. Similarly, on average, passing banks experienced tightening abnormal CDS spreads of -0.52%, while failing banks experienced widening abnormal CDS spreads of 0.83% on that particular date. This suggests a linear or proportional relationship between stress test results and abnormal returns.

The results of Ahnert *et al.* (2020) are largely consistent with those of Morgan *et al.* (2014) and Fernandes *et al.* (2020) for US stress tests and those of Alves *et al.* (2015) for EU-wide stress tests. Morgan *et al.* (2014) showed that banks that were found to have larger capital shortfalls experienced more negative abnormal returns. Similarly, Fernandes *et al.* (2020) found that the direction of capital market reactions tended to depend on the nature of the stress test information disclosed, *e.g.* whether banks have passed or failed a stress test or whether announced stress scenarios were more or less severe than expected by the market. In the Eurpean context, Alves *et al.* (2015) showed that banks that clearly passed a stress test experienced stronger positive cumulative abnormal stock returns, while banks that narrowly passed a stress test experienced weaker positive cumulative abnormal stock returns.

However, the above studies only weakly support an assumed linear relationship between stress test results and abnormal returns (see Ahnert *et al.* 2020). First, studies that have based their analysis on comparisons between passing and failing banks suffer from low statistical representativeness due to the small number of banks that failed a stress test (in total, Ahnert *et al.* (2020) listed 31 failing banks versus 361 passing banks). Second, using two categorical dichotomous variables (such as pass/fail or pass/near pass) suggests a linear relationship by definition. Third, several studies of US and EU-wide stress tests have provided evidence that contradicts a proportional relationship (Sahin *et al.* 2020, Georgescu *et al.* 2017, Georgoutsos and Moratis 2021).

Specifically, Sahin *et al.* (2020) found conflicting abnormal stock returns in response to disclosures of US stress test results. They showed that some bank stock prices increased in response to the disclosure of the SCAP results, regardless of the individual banks' stress test result. On the other hand, the stock prices of some banks that passed the 2014 CCAR stress test decreased in response to the disclosure of the results. However, with respect to the 2011 CCAR stress test – for which no bank-level results were disclosed – no abnormal returns were observed, indicating that there is some stock price formation process associated with the disclosure of supervisory stress test results. In this regard, Georgescu *et al.* (2017) found that abnormal returns in CDS spreads and stock prices in response to the 2016 EU-wide stress test results were stronger for banks with weaker stress test results, from which a non-linear or disproportional relationship can be assumed. Similarly, using quantile regressions for the 2016 and 2018 EU-wide stress tests, Georgoutsos and Moratis (2021) showed that Common Equity Tier 1, leverage, and profitability ratios were important determinants of abnormal stock returns and that there was a non-linear relationship between them and the observed abnormal returns (but only for a specific subset of banks). What is still missing is a study that determines the shape of the functional relationship curve between EU-wide stress test results and abnormal stock returns at bank level. A better understanding of this relationship is important in order to be able to assess whether the disclosure of EU-wide stress test results has indeed improved market discipline, as intended by the EBA.

The work closest to this study in terms of research approach is that of Georgoutsos and Moratis (2021). Their evidence is based on event studies and quantile regressions. Besides this study, they are so far the only ones in the European context that have used $|\overline{CARs}|$ as an additional non-directional measure for the informational value of stress test results (beyond the directional $\overline{CARs}$). They have also taken first, but insufficient, steps to mitigate the joint-hypothesis problem, which poses the greatest threat to the internal validity of event studies. A fundamental difference is that this study covers all EU-wide stress tests available for research, while the study by Georgoutsos and Moratis (2021) is limited to the 2016 and 2018 EU-wide stress tests. Furthermore, this is the first study ever on supervisory stress testing that has performed a systematic model selection procedure to address the joint-hypothesis problem (including confounding control). Regarding the research questions, Georgoutsos and Moratis (2021) limited themselves to examining the informational value of EU-wide stress test results and the determinants of the corresponding abnormal stock returns. This study provides a more comprehensive analysis, also covering the functional relationship between stress test results and corresponding abnormal returns, as well as the change or dynamics of the informational value of EU-wide stress test results over time. The results of the two studies are consistent and mutually reinforcing. Table 4 provides a comprehensive overview of key findings from previous studies on EU-wide stress tests (including studies covering both US and EU-wide stress tests).

Table 4
Key Findings from Previous Studies

Study	EU-Wide Stress Test	Key Findings
Ahnert *et al.* (2020)	2010, 2011, 2014, 2016, 2018[a]	Banks that passed a stress test saw positive abnormal stock returns and tightening CDS spreads, while banks that failed a stress test saw sharp declines in stock prices and widening CDS spreads. In addition, strong market reactions were observed at the stress test announcement events. Despite different institutional designs, similar capital market reactions were observed for US and EU-wide stress tests.
Alves *et al.* (2015)	2010, 2011	Both stress tests have had an impact on bank stock prices. The 2010 EU-wide stress test reduced volatility in bank stock prices, while the 2011 EU-wide stress test increased volatility. The stress test results were not anticipated by the stock market, but were partly anticipated by the CDS market. Banks that clearly passed the stress tests experienced stronger positive cumulative abnormal stock returns, while banks that narrowly passed the stress tests experienced weaker positive cumulative abnormal stock returns. In addition, abnormal stock returns were stronger in banks with a higher risk profile.
Candelon and Sy (2015)	2010, 2011[b]	Both US and EU-wide stress tests have affected bank stock prices. The SCAP had a strong positive impact, while the impact of subsequent US stress tests decreased over time. Among the European exercises, the 2011 EU-wide stress test stood out as the only exercise with significantly negative market reactions. Stock market reactions were largely driven by qualitative factors in stress test governance.
Cardinali and Nordmark (2011)	2010[c]	The results disclosure event of the 2010 EU-wide stress test and the clarification event of the 2011 EU-wide stress test produced only small and statistically insignificant cumulative abnormal stock returns, suggesting that these events were rather uninformative for investors. In contrast, the methodology event of the 2011 EU-wide stress test was found to be highly informative. A breakdown of the stress tested banks into regional portfolios did not yield any new insights.
Georgescu *et al.* (2017)	2016[d]	The announcement of the stress test and the disclosure of the results have caused significant reactions in the stock prices and CDS spreads of the banks concerned. Market reactions to the results disclosure were stronger for weaker-performing banks, suggesting that the stress test results improved price discrimination. Additional evidence suggests that the stress test has also affected sovereign CDS spreads.
Georgoutsos and Moratis (2021)	2016, 2018	The 2018 EU-wide stress test was relatively more informative for investors than the 2016 EU-wide stress test, but only for a subset of banks based on sovereign debt exposure and non-eurozone countries. Quantile regressions on the determinants of abnormal stock returns of banks suggest a non-linear relationship between abnormal returns and Common Equity Tier 1, leverage, and profitability ratios, but only for the same subset of banks as above.
Gerhardt and Vander Vennet (2017)	2011	Overall, the 2011 EU-wide stress test was informative for investors. However, its informational value changed over the course of the six official information events of the stress test, suggesting that the stress test announcement and some of the clarification events were more informative than the disclosure of the results. In addition, sovereign debt exposures had a significant impact on banks' abnormal stock returns in light of the looming European sovereign debt crisis.
Petrella and Resti (2013)	2011	The first and detailed announcement as well as the capital definition event of the 2011 EU-wide stress test were informative for investors. The results disclosure event, on the other hand, produced economically significant but statistically insignificant cumulative abnormal stock returns. Investors did not anticipate the results of the stress test, indicating the lack of transparency in the banking sector. Liquidity and model risk were important determinants of abnormal stock returns. Sovereign debt exposure was only significant in a univariate setting, but not in a multivariate setting.

Note. This table provides an overview of key findings from previous studies of EU-wide stress tests, including studies that examined both US and EU-wide stress tests. [a] Ahnert *et al.* (2020) also examined the seven annual CCAR stress tests conducted from 2012 to 2018. [b] Candelon and Sy (2015) also examined the 2012 EU Capital Exercise, the SCAP, and the two CCAR stress tests conducted in 2012 and 2013. [c] Cardinali and Nordmark (2011) also examined events from the 2011 EU-wide stress test, but not the results disclosure event, which was too late for the study to consider. [d] Georgescu *et al.* (2017) also examined the 2014 ECB Comprehensive Assessment.

3.4 Bank Opacity and Information Uncertainty

Research on bank opacity examines the extent to which banks and the risks associated with their business are transparent to outside investors.[35] To some extent, all firms are opaque and suffer from information asymmetry between insiders and outsiders. However, there is a long-standing debate as to whether banks are more opaque than firms in other sectors (see, for example, Berlin and Loeys 1988; Blau *et al.* 2017, 2020; Calomiris and Mason 1997; Campbell and Kracaw 1980; Diamond 1989, 1991; Flannery *et al.* 2013; Haggard and Howe 2012; Jones *et al.* 2012, 2013; Morgan 2002, Morgan and Stiroh 2001). Today, there is a broad consensus among researchers that banks are indeed particularly opaque (see Table 5). This is problematic because opacity-induced information uncertainty impairs the ability of investors to accurately determine the fundamental value of banks, thus impeding price efficiency and price discrimination between sound and unsound banks (Blau *et al.* 2017, 2020; Dewally and Shao 2013; Jones *et al.* 2012, 2013). In the following, Section 3.4.1 discusses the theoretical sources of bank opacity, while Section 3.4.2 reviews the corresponding empirical evidence.

3.4.1 Theoretical Sources of Bank Opacity

The theoretical literature suggests that bank opacity arises from three main sources specific to banking. First, banks may be prohibited by law from disclosing confidential information about their customers, such as credit and loan relationships, which are typically covered by banking secrecy (Bartlett 2012, BCBS 2006, Jones *et al.* 2012). Second, banks may deliberately choose to withhold sensitive information such as the composition of their assets or their proprietary trading strategies (Bartlett 2012, Chamley *et al.* 2012, Wagner 2007).[36] Third, the complexity of banking and the nature of the

[35] Jones *et al.* (2012, p. 383), for example, defined opacity as "the uncertainty that even the most sophisticated investors face in accurately assessing the fundamental value of a firm." In a related study, Jiang *et al.* (2005) referred to information uncertainty as "value ambiguity" and defined it as "the degree to which a firm's value can be reasonably estimated by even the most knowledgable investors at reasonable costs."

[36] Wagner (2007) argued that bankers value opacity because it offers protection against unwanted disciplinary action. For theoretical studies that have modelled different opacity and transparency strategies using standard signaling games pitting bank insiders against outside investors or regulators, see Besancenot and Vranceanu (2011), Jungherr (2018), and Spargoli (2013).

underlying assets can also create opacity (*e.g.* Blau *et al.* 2017; Flannery *et al.* 2013; Jones *et al.* 2012, 2013).

While the first two sources of bank opacity are intuitive and compelling, there is some debate about the opacity of bank assets. Since the main function of banks is financial intermediation (*i.e.* accepting deposits and granting loans), much of the debate focuses on loans as the main asset of most banks. Several researchers have suggested that bank loans are opaque by definition because banks are better informed about the loans they make than outside investors are (Allenspach 2009, Flannery *et al.* 2004, Haggard and Howe 2012). This reasoning is supported by former Fed Chairman Alan Greenspan (1996, pp. 1-2), who said:

> [B]ank loans are customized, privately negotiated agreements that, despite increases in availability of price information and in trading activity, still quite often lack transparency and liquidity. This unquestionably makes the risks of many bank loans rather difficult to quantify and to manage.

Greenspan's quote above is consistent with Campbell and Kracaw's (1980) theory of financial intermediation, which is based on banks having privileged information about the characteristics of loan agreements and the creditworthiness of borrowers.[37] Their model suggests that while oustside investors are generally able to estimate the fundamental value of a bank loan, the inherent opacity of the financial intermediation process makes generating such information inefficient and costly. A number of subsequent theoretical studies have confirmed the view that bank lending and the process of financial intermediation are inherently opaque and therefore create information uncertainty for outside investors (*e.g.* Berlin and Loeys 1988; Diamond 1989, 1991; Heider *et al.* 2015; Kwan and Carleton 2010).

However, the empirical evidence for the opacity of bank loans and other bank assets is rather mixed. It is widely recognised that asset composition is a key determinant of bank opacity (see, for example, Blau *et al.* 2017, Flannery *et al.* 2013, Morgan 2002, Morgan and Stiroh 2001). In particular, Morgan (2002) provided evidence that bank loans and trading assets are major contributors to bank opacity because their risks are difficult to observe and easy to change. However, several studies have shown that

[37] The theory builds on Campbell's (1979) earlier work on the purpose of the financial intermediary function. According to this work, the function of the financial intermediary is to protect the confidentiality of information related to investment projects of borrowers, the knowledge of which could be advantageous for competing firms. Leland and Pyle (1977) offered a broader rationale by arguing that financial intermediation resolves the information asymmetry between borrowers and lenders.

stock investors were able to correctly identify the banks that were lending to government borrowers at the time of the Mexican (1982) and Brazilian (1987) debt moratoria (Bruner and Simms 1987, Musumeci and Sinkey 1990, Smirlock and Kaufhold 1987). Notably, Musumeci and Sinkey (1990) and Smirlock and Kaufhold (1987) have found that the stock prices of affected banks declined *in proportion* to their loan exposure to the respective countries. This is consistent with rather low opacity levels.

The study by Haggard and Howe (2012) represented the first attempt to examine the extent to which different loan types contribute to overall bank opacity. Their results show that agricultural and consumer loans are relatively less opaque than commercial and industrial loans, real estate loans, and loans to other deposit-taking institutions.[38] Given the overall high opacity of bank loans, it should be noted that an increasing number of studies have used the size of bank loan books as a proxy for bank opacity (*e.g.* Blau *et al.* 2017, 2020; Flannery *et al.* 2013; Jones *et al.* 2012).

Haggard and Howe (2012, p. 52) also induced that if bank loans are opaque, then "bank assets, which are composed primarily of bank loans, are also opaque." This reasoning may be obvious in hindsight, but the opacity of securitised loan products (such as mortgage-backed securities and collateralised debt obligations) was a major cause of the global financial crisis and unprecedented uncertainty about the stability of individual banks and the banking system as a whole (Bernanke 2009, Blanchard 2009, Schuermann 2014). This view is also supported by Flannery *et al.* (2013), who highlighted the significant disagreement between bank insiders and outside investors about the economic value of such assets. They found that this information asymmetry has caused "outside investors to undervalue the average banking firm's equity in a pooling equilibrium" (Flannery *et al.* 2013, p. 56).

The combined evidence from the review of the above studies suggests that bank lending and the process of financial intermediation are likely to contribute to overall

[38] Haggard and Howe (2012) attribute the comparatively low opacity of agricultural and consumer loans to structural reasons that allow outside investors to include more public information in their valuations than with other types of loans. Specifically, they argue that outside investors can infer how the proceeds of an agricultural loan might have been used from the season of lending and the geographic location of a bank. They further argue that price and growth information about the relevant agricultural product is publicly available from spot and futures markets, local weather reports, and agricultural reports. In addition, government-backed crop insurance mitigates some of the risks associated with agricultural loans. Regarding the relative transparency of consumer loans, they argue that this type of loan is dominated by auto loans and that the active secondary car market provides readily available market prices for the underlying assets.

bank opacity. The empirical literature on whether and to what extent banks are actually more opaque than firms from other sectors is reviewed in the next section.

3.4.2 Empirical Evidence for the Opacity of Banks

The empirical literature on bank opacity is extensive and diverse. However, because opacity is not directly observable, various measurable bank characteristics have been used as proxies for bank opacity. As noted by Dahiya *et al.* (2017), these proxies can be broadly classified into three categories: analyst-based measures, market microstructure-based measures, and stock return-based measures. Each of these lines of research is discussed separately below.

Analyst-Based Measures

A number of studies have used analyst-based measures, such as analyst forecasts or rating agency ratings, to proxy the opacity of banks (see, for example, Anolli *et al.* 2014, Bannier *et al.* 2010, Dahiya *et al.* 2017, Flannery *et al.* 2004, Fosu *et al.* 2017, Iannotta 2006, Morgan 2002, Van Roy 2013). The rationale behind these studies is that *ceteris paribus*, analysts' forecast errors and disagreements between different rating agencies (split ratings) should be positively related to the level of opacity. Similarly, a lower number of analysts following a particular bank should be indicative of higher opacity. Most of these studies have found that banks are indeed more opaque than firms in other sectors, but there has also been some conflicting evidence.

The results of Flannery *et al.* (2004) showed that analysts had no more difficulty forecasting bank earnings than non-bank earnings, suggesting that banks are not particularly opaque.[39] In contrast, Anolli *et al.* (2014) and Fosu *et al.* (2017) have found that errors in analysts' earnings forecasts for banks are associated with bank-specific risks and opacity.[40] Taking a slightly different approach, Dahiya *et al.* (2017) showed that differences in analyst coverage also provide evidence that banks are more opaque

[39] More specifically, Flannery *et al.* (2004) found that the errors in analysts' earnings forecasts for large banks (traded on the NYSE) were statistically indistinguishable from those for matched non-banking firms. Notably, analysts' earnings forecasts for small banks (traded on the NASDAQ) were even *more accurate* than the forecasts for the control group.

[40] In a follow-up study, Fosu *et al.* (2018) provided evidence that banks with higher market power and banks that operate in a less competitive environment have lower analyst forecast errors and therefore appear to be less opaque.

than non-banks. Morgan (2002) and Iannotta (2006) have shown for the US and Europe, respectively, that split ratings between Moody's and Standard & Poor's are more common for bank bonds than for non-bank bonds.[41] These disagreements imply that rating agencies struggle to make reasonable assessments of banks' risks, which is consistent with bank opacity. Similarly, Bannier *et al.* (2010) and Van Roy (2013) have found evidence of a downward bias in unsolicited bank ratings compared to solicited bank ratings. Both studies have attributed this excessive conservatism to the opacity of banks, *i.e.* the fact that unsolicited ratings must rely on opaque public information, while solicited ratings may involve both public and private information.

Market Microstructure-Based Measures

Another line of research has proxied bank opacity using various market microstructure variables such as stock trading volumes, bid-ask spreads, insider trades, and Amihud's (2002) illiquidity measure (*e.g.* Dahiya *et al.* 2017; Flannery *et al.* 2004, 2013; Spargoli and Upper 2018). Despite some conflicting results, the combined evidence from these studies suggests that banks are not particularly opaque.

Building on the theories of informed trading by Grossman and Stiglitz (1980) and Kyle (1985), Spargoli and Upper (2018) tested whether insider trades benefited from an informational advantage over outsider trades. Their results provided no evidence of an informational advantage of bank insiders and therefore suggest that banks are no more opaque than non-banks. Using a range of different market microstructure variables Flannery *et al.* (2004) showed that neither small nor large banks are particularly opaque. This was basically confirmed by Flannery *et al.* (2013). However, their results also showed that bank opacity varies over time regardless of bank size, and that banks become more opaque than non-banks in times of crisis. This finding is in conflict with Spargoli and Upper (2018) who also examined bank opacity over time but found no such evidence. Using Amihud's (2002) illiquidity measure, Dahiya *et al.* (2017) provided evidence that banks are actually more opaque than non-banks, contradicting the findings of most other market microstructure-based studies.

[41] The results of Iannotta (2006) also indicate that bank opacity is associated with the size, capital structure, and asset composition of a bank. In addition, bank bond seniority and bank opacity are found to be negatively related. Livingston *et al.* (2007) reconfirm that split ratings indicate firm opacity using a sample consisting only of non-banking firms.

Stock Return-Based Measures

Finally, a third line of research has used stock return characteristics as a proxy for bank opacity (*e.g.* Blau *et al.* 2017, 2020; Dewally and Shao 2013; Haggard and Howe 2012; Jones *et al.* 2012, 2013). There is a broad consensus among these studies that banks are indeed particularly opaque. Most of them have relied on the theoretical models of Jin and Myers (2006) and Veldkamp (2006), where outside investors have incomplete firm-specific information and fill in the gaps with expected values (conditional on the information available) or with common market and industry signals. An important consequence of a lack of firm-specific information is greater co-movement of stock prices, also known as "price synchronicity" (Morck *et al.* 2000).

Haggard and Howe (2012) showed that market returns explain a higher proportion of the variance in bank stock returns than in matched non-bank stock returns, confirming Veldkamp's (2006) theoretical predictions. In other words, they showed that bank stocks contain a higher proportion of market information (systematic risk) versus firm-specific information (idiosyncratic risk) compared to non-bank stocks. This suggests that the different risk profiles of banks are not adequately reflected in their stock prices. The results also point to stronger co-movement in bank stocks.[42] This is consistent with the results of more specific studies, which have shown that bank opacity is associated with price synchronicity, or greater co-movement of bank stocks relative to non-bank stocks (Blau *et al.* 2020; Dewally and Shao 2013; Jones *et al.* 2012, 2013). Notably, Jones *et al.* (2012) also showed that a bank's level of opaque assets is positively related to the level of stock price revaluation following an exogenous information shock. They also provided evidence that bank stocks incorporate industry-specific information signals. Using Hou and Moskowitz's (2005) measure of price delay, Blau *et al.* (2017) showed that bank stocks take longer than non-bank stocks to incorporate new information into their prices. They argued that bank opacity creates information uncertainty, making it difficult for investors to interpret new information and reflect it in the stock prices of banks.

All of the above suggests that bank stock pricing is informationally inefficient, as opacity prevents outside investors from accurately assessing the specific risks of

[42] For a general discussion of the co-movement of stock prices and its links to firm-specific information disclosure, see Haggard *et al.* (2008).

individual banks. It has therefore been argued that opacity leaves bank stocks vulnerable to crashes and sharp revaluations triggered by changes in outside investors' perceptions of risk (Dewally and Shao 2013; Haggard and Howe 2012; Jones *et al.* 2012, 2013), confirming Jin and Myers' (2006) theoretical concerns.

Summary

Based on the combined evidence from the three lines of research above, it seems reasonable to accept banks as particularly opaque entities. There is a broad consensus among analyst and stock return-based studies that banks are more opaque than non-banks and that opacity is associated with bank-specific risks. Only market microstructure-based studies have provided some evidence to the contrary. Table 5 summarises the empirical literature on bank opacity and lists the key findings.

Table 5
Summary of the Empirical Literature on Bank Opacity

Study	Opacity Proxy	Sample	Study Period	Key Findings
Anolli *et al.* (2014)	AB	411 banks	2003-2009	Opacity is associated with bank-specific risks
Bannier *et al.* (2010)	AB	26,413 ratings[a]	1996-2006	Banks are more opaque than non-banks.
Fosu *et al.* (2017)	AB	402 banks	1995-2013	Opacity is associated with bank-specific risks
Iannotta (2006)	AB	2,473 bonds[b]	1993-2003	Banks are more opaque than non-banks.
Morgan (2002)	AB	7,862 bonds[c]	1983-1993	Banks are more opaque than non-banks.
Van Roy (2013)	AB	169 bank ratings	2004	Banks are opaque.
Dahiya *et al.* (2017)	AB, MMB	72,833 firm-years[d]	1981-2011	Banks are more opaque than non-banks.
Flannery *et al.* (2004)	AB, MMB	320 banks	1990-1997	Banks are *not* more opaque than non-banks
Flannery *et al.* (2013)	MMB	48,000 stock-months	1993-2009	Banks are more opaque than non-banks in "crisis" times but not in "normal" times.
Spargoli and Upper (2018)	MMB	743 bank stocks	1990-2015	Banks are *not* more opaque than non-banks
Blau *et al.* (2017)	SRB	361 bank stocks	1996-2008	Banks are more opaque than non-banks and bank stocks have a significantly higher price delay than non-bank stocks.
Blau *et al.* (2020)	SRB	25,000 stocks[e]	1980-2012	Banks are more opaque than non-banks and the prices of bank stocks co-move more than the prices of non-bank stocks.
Dewally and Shao (2013)	SRB	98 bank stocks	1995-2010	Banks are opaque and the prices of bank stocks move synchronously.
Haggard and Howe (2012)	SRB	243 bank stocks	1993-2002	Banks are more opaque than non-banks.
Jones *et al.* (2012)	SRB	357 non-merger banks and 80 merger-banks	2000-2006	Banks are opaque and the prices of bank stocks incorporate industry-specific information.
Jones *et al.* (2013)	SRB	8,152 bank-quarters	2000-2007	Banks are opaque and the prices of bank stocks move synchronously.

Note. This table summarises the empirical literature on bank opacity and lists the key findings. AB = analyst-based measure. MMB = market microstructure-based measure. SRB = stock return-based measure.

[a] Of which 5,990 were bank ratings. [b] Of which 2,051 were bank bonds. [c] Of which 848 were bank bonds. [d] Of which 5,183 were bank firm-years. [e] Of which 2,039 were bank stocks.

The findings from the empirical literature presented in this section suggest that banks are particularly opaque, meaning that the information asymmetry between bank insiders and outside investors is unusually high. The resulting information uncertainty makes it difficult for investors to determine the fundamental value of banks, leading to informationally inefficient stock prices that tend to move in sync. This implies that investors' ability to differentiate and price-discriminate between sound and unsound banks is impaired. Several studies have therefore argued that the problems associated with bank opacity justify regulating bank transparency through information-generating measures such as supervisory stress tests (Jordan *et al.* 2000, Morgan 2002, Petrella and Resti 2013).

These are important interim review results that form the basis for three further threads of discussion. First, about the impact of new public information (*e.g.* EU-wide stress test results) on stock prices (Section 3.5). Second, about the risk-return tradeoff of stocks and its dynamics under exogenous information shocks (Section 3.6). Third, about the intertemporal stability of the informational value of supervisor-generated information (Section 3.7).

3.5 Informational Efficiency of Capital Markets

The relative opacity of banks to outside investors (Section 3.4) and the arguable need to regulate bank transparency through supervisory stress testing and public disclosure of results at bank level point to the informational efficiency of capital markets. This is the subject of the thematic review in the following sections. More specifically, the extent to which stock prices are efficient in incorporating new public information according to Fama's (1970) efficient market hypothesis. In other words, the impact of newly disclosed information (*e.g.* bank-level results of EU-wide stress tests) on the stock prices of affected banks. The review begins by outlining the efficient market hypothesis (Section 3.5.1) and describing tests for market efficiency using event-study analysis (Section 3.5.2). It also discusses the problems of testing for market efficiency (Section 3.5.3), in particular the joint-hypothesis problem (Fama 1970, 1991). The thematic review concludes with a discussion of existing empirical evidence for market efficiency in its weak, strong and semi-strong forms (Section 3.5.4).

3.5.1 The Efficient Market Hypothesis (EMH)

In a seminal study, Fama (1970) synthesised the existing literature on the informational efficiency of capital markets and formalised it in his efficient market hypothesis. According to the EMH, a market is efficient when "security prices at any point in time 'fully reflect' *all* available information" (Fama 1970, p. 388). Therefore, the price of a security should be an unbiased reflection of all information available at the time, including the risk associated with holding the security (Reilly and Brown 2012).

The EMH is based on three main assumptions. First, investors are rational and value securities independently based on maximum expected utility. Second, new information about securities come to the market randomly; that is, the timing of a new piece of information is generally independent of that of others. Third, investors' buying and selling decisions cause security prices to adjust rapidly to reflect the new information (Reilly and Brown 2012).[43]

Fama (1970, 1991) divided empirical testing of the EMH into three forms, depending on the subset of information considered: *weak-form*, *semi-strong form*, and *strong-form efficiency*. Each of these forms is briefly discussed below; the empirical evidence is reviewed in Section 3.5.4, with a focus on semi-strong form efficiency, the form of efficiency relevant to this study.

Weak-form efficiency assumes that prices fully reflect all past *market* information about a security, *i.e.* historical prices and other market-generated information (Fama 1970, 1991; Reilly and Brown 2012). This implies that there is no connection between past and future security prices (random walk) since the current security price already reflects all historical information. Therefore, future security prices should not be predictable, and technical analysis should not yield returns that systematically exceed those of the market (Fama 1970, Reilly and Brown 2012).

Semi-strong form efficiency, on the other hand, assumes that security prices adjust rapidly to all new *public* information concerning a security (Fama 1970, 1991), where "public information" includes all non-market information, such as financial announcements, accounting ratios, as well as political, economic, and firm-specific news (Reilly and Brown 2012). Semi-strong form efficiency therefore implies that fundamental analysis should not be able to systematically generate excess returns, since current security prices should immediately reflect all relevant information (Reilly and Brown 2012). It is important to note that the disclosure of bank-level supervisory stress test results is part of the information subset relevant to semi-strong form efficiency (Ahnert *et al.* 2020). In addition, semi-strong form efficiency includes weak-form efficiency since all past market information is public by definition.

---

[43] For a detailed discussion of the EMH, see Malkiel (1989).

Finally, *strong-form efficiency* assumes that security prices fully reflect all market, public, and *private* information pertaining to a security (Fama 1970, 1991). This implies that strong-form efficiency includes both weak-form and semi-strong form efficiency. Strong-form efficiency also extends the assumptions of the EMH to include the assumption of a perfect market where all information is available to everyone at the same time and free of charge (Reilly and Brown 2012). That is, no group of investors (including corporate insiders) should have monopolistic access to information relevant to security pricing (Fama 1970, 1991). As a result, no one should be able to systematically outperform the market (Fama 1970, 1991; Reilly and Brown 2012).

The theoretical basis provided by the efficient market hypothesis was important for all three research questions of this study (Section 1.4) as it formed the key theory of the study's theoretical framework (Section 4.2). What all three research questions have in common is that they examined the effects of new public information (EU-wide stress test results) on the stock prices of the banks concerned. In this respect, the efficient market hypothesis offered a suitable theoretical foundation. In addition, the interrelationship between market efficiency and asset pricing should be mentioned in this context, which is explained in more detail in Section 3.5.3. Since market efficiency and asset pricing are two sides of the same coin (Fama 2014), it was essential to also address the various asset pricing models that can be used in event studies to estimate normal returns (Section 3.5.2). This was particularly important for this study as it introduced systematic model selection into the standard event study approach developed by Campbell *et al.* (1997) and MacKinlay (1997). The aim was to improve the content validity when measuring abnormal stock returns and thus achieve more robust answers to the research questions than previous studies. The thematic review in this section (and in the other sections of Chapter 3) contributed to the literature by providing a stronger theoretical and methodological basis for studies of capital market responses to supervisory transparency measures. This finally allowed this study to answer research questions on important aspects left unexplored by previous studies of EU-wide stress tests (see Section 1.2). For more details on the theoretical, methodical, and practical contributions of this study, see Section 8.4.

With regard to the overall state of knowledge, the empirical evidence for the efficient market hypothesis is mixed, depending on the form of efficiency examined (see Section 3.5.4). It therefore seems reasonable to briefly discuss alternative theories on the efficiency of capital markets. As mentioned above, both technical and fundamental analysis are diametrically opposed to the EMH. Technical analysis is a collective term for a set of techniques that go back to the Dow Theory (Edwards *et al.* 2018, Levy 1966).[44] In general, technical analysis assumes that security prices adjust gradually (rather than rapIdly) to new information, thereby creating persistent trends that can be exploited (Edwards *et al.* 2018, Levy 1966). Another assumption of technical analysis is that historical price patterns repeat themselves in the future and are therefore suitable for predicting prices (Edwards *et al.* 2018; Fama 1965a, 1965b; Levy 1966). Both assumptions are in direct contradiction to weak-form efficiency. However, almost all weak-form efficiency tests have shown that security prices do *not* move in trends and that technical trading rules are unable to systematically generate excess returns, thus confirming the EMH.

Fundamental analysis, in turn, assumes that there is an intrinsic value for every security that depends on economic and firm-specific information and may differ from the security's current market price (Fama 1965b, Malkiel 2003, Reilly and Brown 2012). Investors should therefore be able to systematically generate excess returns by buying undervalued securities and selling overvalued securities.[45] This assumption is in direct contradiction to semi-strong form efficiency, which states that all economic and firm-specific information is already reflected in the price of a security; therefore, fundamental analysis should not be able to generate returns that systematically exceed those of the market.

In addition, the EMH has been challenged by a variety of studies that emphasise the psychological and behavioural elements in security pricing, for example Barberis (2018), Daniel and Titman (1999), Daniel *et al.* (1998, 2002), De Bondt and Thaler (1985, 1987), Hirshleifer (2015), Shefrin and Statman (1985), Shiller (1984, 1990, 1999, 2003), and Thaler (1999, 2005).[46] More specifically, these behavioural finance studies have criticised the EMH for assuming that investors behave rationally and have argued that investor behaviour is subject to various cognitive and social biases. As

---

[44] The Dow Theory asserts that security prices move in trends and describes three types of price movement: primary, secondary, and minor trends (Edwards *et al.* 2018). For a comprehensive discussion of the Dow Theory, see Edwards *et al.* (2018). For a conceptual justification of technical analysis in general, see Levy (1966).

[45] Provided that the difference between a security's intrinsic value and its market price is large enough to cover transaction costs (Reilly and Brown 2012).

[46] For surveys of the behavioural finance literature and its relations to the EMH, see Barberis and Thaler (2003), Hirshleiffer (2001), Malkiel (2003), and Shiller (1999).

noted by Olsen (1998), behavioural finance recognises that established finance theory, which assumes rational and expected utility-maximizing investors, can be true within certain limits, but claims that it is incomplete because it does not take into account the behaviour of individual investors. However, the existing contributions of behavioural finance are largely limited to anecdotal evidence on specific market anomalies[47] and do not provide a unified theory of behavioural finance (Joo and Durri 2018, Olsen 1998, Reilly and Brown 2012).[48] The most promising approaches are attempts to reconcile behavioural finance with neoclassical economics, such as Lo's (2004, 2005, 2012, 2019) adaptive markets hypothesis.

The next section reviews how market efficiency can be tested empirically, with a focus on semi-strong form tests.

3.5.2 Testing for Market Efficiency: Event-Study Analysis

As mentioned above, the three forms of efficiency of the EMH are based on alternative subsets of information. Since only new public information (disclosure of EU-wide stress test results) is relevant in the context of this study, the following review is limited to empirical tests of semi-strong form efficiency.

The purpose of semi-strong form tests is to examine whether security prices adjust efficiently to new public information (Fama 1970, 1991). The most widely used method for testing semi-strong form efficiency is event study analysis.[49] As noted by Fama (1991, p. 1602) "[e]vent studies are the cleanest evidence we have on efficiency" because they allow to isolate the abnormal return associated with a specific information

---

[47] For example the weekend effect (Smirlock and Starks 1986), the January effect (Thaler 1987), and the correlation between stock returns and morning sunshine (Hirshleifer and Shumway 2003).

[48] Hirshleifer (2015, p. 151) acknowledeges that "[m]ore theoretical…study is needed of how feelings affect financial decisions, and the implications of such effects for prices and real outcomes." For an early attempt to develop a capital asset pricing theory grounded in behavioural finance, see Shefrin and Statman (1994).

[49] A census of event-study articles found that 565 event studies were published in five leading academic journals from 1974 to 2000 alone (*Journal of Finance*, *Journal of Financial Economics*, *Review of Financial Studies*, *Journal of Financial and Quantitative Analysis*, and *Journal of Business*); the number of published event-study articles per year increased in the 1980s and has since remained stable (Kothari and Warner 2007). Given the overwhelming increase in event studies, Fama even changed the category title he had originally used for semi-strong form tests and started to „use the now common title, *event studies*" (Fama 1991, p. 1577); in his seminal 1970 article, Fama had originally used the the collective term „semi-strong form tests of efficient markets models" (Fama 1970, p. 404). See also the survey study by Atanasov and Black (2016), which identified event studies as a leading research method for shock-based causal inference.

event. The abnormal return is the actual (observed) return of a security minus the normal (expected) return of the security that would be expected if the information event had not occurred, measured over a specified event window (Campbell *et al.* 1997, Kothari and Warner 2007, MacKinlay 1997). Equations (2) and (3) illustrate this relationship. For each security i and each event window τ, the actual (observed) return $R_{i\tau}$ is given by:

$$R_{i\tau} = E(R_{i\tau}|X_\tau) + AR_{i\tau}, \tag{2}$$

where $E(R_{i\tau})$ is the normal (expected) return, X_τ is the conditioning information for the normal return-generating model, and $AR_{i\tau}$ is the abnormal (unexpected) return. Given this decomposition of returns, Equation (2) can be rearranged as follows:

$$AR_{i\tau} = R_{i\tau} - E(R_{i\tau}|X_\tau). \tag{3}$$

As a result, the abnormal (unexpected) return $AR_{i\tau}$ is the difference between the actual (observed) return $R_{i\tau}$, conditional on the information event, and the normal (expected) return $E(R_{i\tau})$, not conditional on the information event, but on the normal return-generating model X_τ. The abnormal return $AR_{i\tau}$ is therefore a direct measure of the impact of a particular information event on the price of a security, *i.e.* the impact that is not explained by the normal return-generating model (Kothari and Warner 2007). The null hypothesis to be tested in event studies typically assumes that abnormal returns (potentially aggregated over time or across a sample) are statistically indifferent from zero.

Two important inferences can be drawn from the above. First, every event study relies on some kind of return-generating model to estimate a security's normal return. Second, the normal return estimate is the key parameter since the only other parameter (the actual return) is a given observation. In other words, selecting an appropriate asset pricing model to estimate a security's normal return is critical to the internal validity of any event study. Consequently, estimating normal returns is one of the most debated issues in the event-study methodology literature.

Event studies have a long history, dating back to Dolley (1933), which is probably the first event study ever published. Early event studies have relied on the simple mean-adjusted returns (MAR) model (Ashley 1962; Barker 1956, 1957, 1958; Myers and Bakay 1948), where the normal (expected) return $E(R_i)$ of a given security i is estimated as the arithmetic mean of the security's actual (observed) return $\bar{R}_i$ over a given estimation period:[50]

$$E(R_i) = \bar{R}_i. \tag{4}$$

However, the MAR model has been criticised for its implicit assumption that the normal (expected) return of a given security is constant through time (Brown and Warner 1980, Fama 1991, MacKinlay 1997). Additionally, the model does not explicitly consider the risk of the security or its relationship to market return (Binder 1998).

Over time, researchers have recognised the need to adjust the normal (expected) return $E(R_i)$ of a given security i by the market return R_m (Binder 1998, Brown and Warner 1980, 1985; Levis 1993; Ritter 1991):

$$E(R_i) = R_m. \tag{5}$$

This Market-Adjusted Model addressed the concern that normal (expected) return estimates should vary over time. Using market return instead of historical mean returns removes the portion of normal (expected) return related to variations in market return and thus reduces the variance of abnormal return (see Equation (3)).

In the late 1960s, the seminal studies by Ball and Brown (1968) and Fama *et al.* (1969) have introduced the basic event-study framework that is essentially still in use today (Campbell *et al.* 1997, Campbell 2014, MacKinlay 1997). In light of the theoretical studies of the time (Sharpe 1964; Lintner 1965a, 1965b; Mossin 1966) that have led to the capital asset pricing model (CAPM), Ball and Brown (1968) and Fama *et al.* (1969) have recognised that the normal (expected) return $E(R_i)$ of a given security i should not only vary over time, but should also reflect the security's sensitivity to market risk, *i.e.* the security's beta (β_i):[51]

---

50 That is, $\bar{R}_i = \frac{1}{n}\sum_{i=1}^{n} R_i = \frac{R_1+R_2+ \quad +R_n}{n}$, where R_i is the actual (observed) return of security i on a given trading day during a specified estimation period, and n is the number of trading days in the estimation period.

51 Expressed more formally, the beta β_i of a given security i is the sensitivity of the expected excess return of the security (*risk premium*) $E(R_i) - R_f$ to the expected excess return of the market (*market premium*) $E(R_m) - R_f$, that is $\beta_i = \frac{Cov(R_i,R_m)}{Var(R_m)} = \rho_{i,m}\frac{\sigma_i}{\sigma_m}$, where $\rho_{i,m}$ is the correlation coefficient between security i and the market m, σ_i is the standard deviation of security i, and σ_m is the standard deviation of the market m.

$$E(R_i) = R_f + \beta_i\big(E(R_m) - R_f\big), \tag{6}$$

where R_f is the risk-free rate of return, and $E(R_m)$ is the expected return of the market.

Following Ball and Brown (1968) and Fama *et al.* (1969), a large number of event studies have used a variety of asset pricing models to estimate normal (expected) returns. The most prominent are the CAPM, Ross' (1976) arbitrage pricing theory, the Fama and French (1993) three-factor model, and the simple market model.[52]

Previous studies examining the impact of EU-wide stress test events on banks' abnormal stock returns have typically resorted to the convenient market model (see Table 3). This created a "model monoculture" and may have contributed to the inconclusiveness of previous studies (Section 3.3.2). In this study, a systematic model selection procedure (Section 5.3.2.1.3) was carried out as the basis for the event-study analysis and the research-question specific analyses, which required to review the connection between market efficiency and asset pricing in a historical context. Since the purpose of semi-strong form tests is to determine whether security prices adjust efficiently to new public information (*i.e.* whether abnormal stock returns can be found in response to EU-wide stress test results), it is important to understand the existing models for estimating normal returns. This is relevant for all three research questions of this study, but especially for Research Question 1 (*The Informational Value Hypothesis*), which comes closest to the design of the previous studies and can therefore be used for cross-study comparisons. The specific problems associated with event studies as tests for semi-strong form efficiency are discussed in the next section.

[52] The market model is the predominant model in event studies for estimating normal (expected) returns. In a meta-analysis of 400 event studies, Holler (2012) found that 79.1% of the studies have used the market model (amongst others). The market model is a simple one-factor model that relates the return of a security to the return of the market without any economic content, *i.e.* it is a purely statistical model (MacKinlay 1997, Stapleton and Subrahmanyam 1983). More formally, for any given security i the market model is: $E(R_i) = \alpha_i + \beta_i R_m$, where R_i is the return of security i, and R_m is the return of the market m. For a more detailed discussion of the market model, see Rudd and Rosenberg (1980).

3.5.3 Problems in Testing for Market Efficiency

This section reviews the problems associated with testing market efficiency, particularly semi-strong form efficiency. Many of these problems arise from the fact that market efficiency *per se* is not testable and must always be tested jointly with some kind of asset pricing model (Fama 1970, 1991, 1998; Jarrow and Larsson 2012). This is because measuring market efficiency – *i.e.* a security's abnormal (unexpected) return – requires an estimate of the normal (expected) return in order to compare it to the actual (observed) return. The need for a combined test of market efficiency and asset pricing theory is commonly referred to as the *joint-hypothesis problem* (Fama 1970, 1991). Any abnormal (unexpected) return may therefore be due to market inefficiency or modelling errors (Fama 1970, 1991, 1998; Kothari and Warner 2007), or a combination of both.[53] Empirical studies on market efficiency and asset pricing models are therefore two sides of the same coin.

The joint-hypothesis problem is also known as the *bad-model problem* (Butler and Wan 2010, Jarrow and Larsson 2012, Min and Kim 2012) because every test of market efficiency is contaminated by a certain amount of estimation error inherent in any asset pricing model (Fama 1998). It is generally accepted that all models are simplifications or approximations of reality (Burnham and Anderson 2002, Fama 1998, Nester 1996, Reiss 2012). However, different models produce different estimation errors. Choosing an inappropriate asset pricing model is therefore likely to result in large estimation error and consequently spurious abnormal returns (Fama 1998). In other words, "event study tests are well-specified only to the extent that the assumptions underlying their estimation are correct" (Kothari and Warner 2007, pp. 12-13). This underlines the importance of model selection to minimise error in estimating normal (expected) returns. In this context, Box and Draper (1987, p. 424) concisely noted that "[e]ssentially, all models are wrong, but some are useful."[54] Therefore, for a model selection approach to be effective, the set of candidates must include models that approximate the risk factors supported by empirical evidence (Burnham and Anderson 2002).

[53] In other words, event studies jointly test whether the abnormal (unexpected) returns of a given security are zero and whether the asset pricing model used for estimating the security's normal (expected) returns is correct (Campbell 2014, Kothari and Warner 2007).

[54] It should be noted that Box's famous aphorism "all models are wrong" was first mentioned in his 1976 article on science and statistics (Box 1976, p. 792) and was only later supplemented with the expansion "but some are useful"(Box 1979, p. 202).

In the late 1970s, empirical studies began to examine the existence of risk factors that influence the formation of security prices and complement the explanation of the traditional market risk factor β. Early studies in this area include Basu (1977), Reinganum (1981), and Stambaugh (1982). Over the years, empirical research has identified a number of additional risk factors, including firm size (Banz 1981; Barber and Lyon 1997a; Brown *et al.* 1983; Fama and French 1992, 2012; Keim 1983; Reinganum 1981, 1983), market value ratios (Barber and Lyon 1997a; De Bondt and Thaler 1985; Fama and French 1992, 2012; Lakonishok *et al.* 1994), momentum (Fama and French 2012; Jegadeesh and Titman 1993, 2001, 2011), and profitability and investment (Fama and French 2006, 2008). Given the empirical evidence, researchers have developed multi-factor models such as the Fama and French (1993) three-factor model, the Carhart (1997) four-factor model, and the Fama and French (2015) five-factor model that extend the CAPM and intend to better explain the sample-specific return variation (Barber and Lyon 1997b, Fama 1998).[55] However, given the approximate nature of all asset pricing models, bad-model problems can only be mitigated to a certain extent, but never completely avoided (Fama 1998).

The next section reviews the empirical evidence on market efficiency with a focus on event studies and semi-strong form efficiency.

3.5.4 Empirical Evidence on Market Efficiency

To recall, for the EMH to hold, the prices of securities must fully reflect all available information at any point in time (Fama 1970, 1991). Any test of market efficiency therefore involves an extreme null hypothesis[56] which, as Fama (1970) recognised, is not expected to be literally true. Instead, the three forms of market efficiency (see Section 3.5.1) "serve the useful purpose of allowing us to pinpoint the level of information at which the hypothesis breaks down" (Fama 1970, p. 388). The overall empirical evidence on market efficiency is best described as mixed. However, depending on the form of efficiency tested and the time of the study, considerable differences can be observed.

---

[55] For detailed descriptions of the asset pricing models used in this study, see Section 5.3.2.1.3.

[56] Event study tests of semi-strong form efficiency, for instance, assume a null hypothesis of zero abnormal (unexpected) returns (see Equation (3)).

3.5.4.1 Tests of Weak-Form Efficiency

All weak-form efficiency tests are carried out in the context of a "fair game" model or the even more specific random walk hypothesis (see Section 3.5.1) and essentially date back to the work of Mandelbrot (1963, 1966) and Samuelson (1965).[57] Empirical studies on weak-form efficiency can therefore be categorised in two groups of tests: first, statistical tests of independence between rates of return (mainly autocorrelation tests and runs tests) and, second, risk-return comparisons between certain trading rules and a simple buy-and-hold strategy.[58]

Early statistical studies have tended to support weak-form efficiency (*e.g.* Fama 1965a, Godfrey *et al.* 1964, Solnik 1973),[59] while later studies have rejected it for portfolios of small-cap stocks (Lo and MacKinlay 1988, Conrad and Kaul 1988) and for long (multi-year) observation periods over which stock prices may *temporarily* deviate from their fundamental values (Shiller 1984, Summers 1986). However, this contradictory evidence must be treated with caution. While the small-cap portfolio anomaly may be confounded by non-synchronous trading effects (Fisher 1966), the long-run fundamental value deviations found are of low statistical power and, due to their temporary nature, are likely to reverse over time (Fama and French 1988, Poterba and Summers 1988). In a broad study of eighteen national stock markets with a 32-year study period from 1961 to 1992, Chan *et al.* (1997) found that stock markets are generally weak-form efficient.

Similarly, early trading-rule studies have generally confirmed weak-form efficiency, especially after accounting for trading costs; more recent studies, however, tend to produce different results, partly due to the general decline in trading costs (Reilly and Brown 2012). Several studies have suggested that certain past-return-based trading rules can produce returns in excess of a simple buy-and-hold strategy (*i.e.* the market) and therefore contadict weak-form efficiency. The best-known examples are probably the studies by De Bondt and Thaler (1985, 1987), which are based on reversals of past "winner" and "loser" stocks, and by Jegadeesh and Titman (1993, 2001),

---

[57] However, the first formulation and testing of the random walk hypothesis was carried out by Bachelier (1900), who stated that speculation should be a fair game and that the speculator should expect zero profits.

[58] For a comprehensive discussion of the random nature of capital markets, including many issues relevant to weak-form efficiency (*e.g.* long-horizon effects, non-synchronous trading, and price reversals), see Lo and MacKinlay (2011).

[59] For early statistical studies that contradict weak-form efficiency, see, for example, Fisher (1966) or Niederhoffer and Osborne (1966).

which, somewhat contradictorily, use relative strength strategies to generate momentum profits. The alleged winner-loser effect has been criticised for its failure to risk-adjust the returns (Ball and Kothari 1989, Chan 1988) and for not reducing the confounding impact of Banz' (1981) size effect (Zarowin 1989). The momentum effect, however, seems to be robust across different time periods, markets, and asset classes (Bhojraj and Swaminathan 2006, Griffin *et al.* 2003, Rouwenhorst 1998) even after risk-adjustment (Fama and French 1996). This caused Fama and French (2008, p. 1653) to accept momentum as "[t]he premier anomaly".

3.5.4.2 Tests of Strong-Form Efficiency

Strong-form efficiency tests, on the other hand, examine whether certain groups of investors (*e.g.* corporate insiders, security analysts, and professional fund managers) who are believed to have private information can systematically generate excess returns above the market return (Fama 1970, 1991; Reilly and Brown 2012). Evidence of consistent above-market returns would contradict strong-form efficiency.

Over the years, a number of studies have found that transactions of corporate insiders are associated with significant excess returns (Fidrmuc *et al.* 2006, Seyhun 1986, Pettit and Venkatesh 1995).[60] Chowdhury *et al.* (1993) offer a contrary view by suggesting that insider transactions react to market movements, and not vice versa. Notably, Spargoli and Upper (2018), who examined the transactions of bank insiders, found that sales did not produce excess returns, while purchases did, but less than purchases by insiders from non-banking firms.

Similarly, most analyst recommendation studies have found that security analysts possess private information and thus contradict strong-form efficiency (Davies and Canes 1978, Goff *et al.* 2008, Jegadeesh *et al.* 2004, Womack 1996). Ivković and Jegadeesh (2004, p. 462) conclude that "the value of analysts'…recommendations stems more from their independent collection of information than from their interpretation of public information." Analyst recommendation studies that support strong-form efficiency are generally limited to analyses of smaller and less developed markets (Lidén 2006, Yazici and Muradoğlu 2002). This is consistent with Jegadeesh and

[60] This is consistent with firm opacity, *i.e.* information asymmetry between corporate insiders and outside investors, see Section 3.4.

Kim (2006), who found that larger markets tend to react stronger to revisions of analyst recommendations than smaller markets.

In contrast, studies on the performance of professional fund managers have found almost unanimously that mutual funds have been unable to consistently generate excess returns, especially after risk adjustment and deduction of costs (Friend and Vickers 1965; Jensen 1968, 1969; Kon 1983; Sharpe 1966; Williamson 1972). In fact, after taking all costs into account, the majority of mutual funds performed worse than the market. In a second wave, several studies have examined alternative measures for the excess performance of mutual funds (Ashton 1996, Ferson and Schadt 1996, Kothari and Warner 2001, Lehmann and Modest 1987). The overall evidence from these studies suggests that alternative measures tend to improve mutual fund performance relative to the market, but generally do not reverse the findings of previous studies.

3.5.4.3 Tests of Semi-Strong Form Efficiency

The body of literature on semi-strong form tests of market efficiency is vast and diverse. In financial economics and related disciplines, event studies have been used extensively to test whether certain information events are associated with abnormal (unexpected) security returns.[61] The information events examined are very diverse; popular events include initial public offerings (IPOs), exchange listings, stock splits, accounting and economic information, and regulatory actions. The following review of the event-study literature is organised by information event or item of public information. Notably, the results of these studies have almost unanimously supported semi-strong form efficiency.

A large number of studies have examined the short-term behaviour of stock prices after initial public offerings (IPOs), *e.g.* Aggarwal *et al.* (2002), Beatty and Ritter (1986), Derrien and Womack (2003), Ibbotson (1975), Ibbotson *et al.* (1994), Loughran *et al.* (1994), Miller and Reilly (1987), Reilly and Hatfield (1969), Ritter

[61] Event study tests of semi-strong form efficiency have also had a significant impact on legal practice, as litigators routinely use event studies to infer fundamental information from changes in the prices of securities (Campbell 2014). For more information on market efficiency and the application of event studies in a legal context, see, for example, Baker (2016), Gilson and Kraakman (1984, 2003), and Torchio (2009).

(1984). A recurring observation in these studies is that newly issued stocks yielded significant positive abnormal returns in the short run – in particular on the day of listing.[62] Early evidence on IPO returns suggests an average abnormal return of 18.3% (Reilly and Hatfield 1969). This is consistent with Miller and Reilly (1987) and Ibbotson *et al.* (1994), who have shown that most of the stock price adjustment happens within one day after the IPO. These results were re-confirmed more recently by Derrien and Womack (2003), who reported an average first-day IPO return of 13.23% compared to a market return of 1.55%. This substantial and rapid stock price adjustment around IPOs lends considerable support to semi-strong form efficiency. Notably, an equally large number of studies examining long-term post-IPO returns have shown that investments in newly issued stocks *after* the initial price adjustment do *not* generate positive abnormal returns, *e.g.* Carter *et al.* (1998), Espenlaub *et al.* (2000), Levis (1993), Loughran and Ritter (1995), Ritter (1991), Schultz (2003), Stehle *et al.* (2000), Teoh *et al.* (1998).[63]

Researchers have also examined stock price adjustments due to exchange listing events, usually changes in listing from over-the-counter (OTC) markets to organised exchanges (*e.g.* Baker and Edelman 1990, McConnell and Sanger 1984, Sanger and McConnell 1986, Ying *et al.* 1977) or from one exchange to another (*e.g.* Baker and Edelman 1992; Dharan and Ikenberry 1995; Edelman and Baker 1990, 1993, 1994; Elyasiani *et al.* 2000; Lau *et al.* 1994). Some of these studies have observed two common patterns around exchange listing events: the stocks concerned had, on average, positive (negative) abnormal returns in the period immediately before (after) the listing date (Lau *et al.* 1994, McConnell and Sanger 1984, Sanger and McConnell 1986, Ying

[62] The most frequently cited reason for the occurrence of short-term abnormal returns in newly issued stocks is information asymmetry. The literature has suggested two basic theories that work on different levels: Baron (1982) models the IPO market as a principal-agent problem, where the compensation agreement between the issuing firm (principal) and the underwriting investment bank (agent) is a function of the IPO proceeds and the stock price and where the investment bank has superior information in terms of market demand. This theory was convincingly refuted by Muscarella and Vetsuypens (1989) and has been cited less frequently since then. The other, more widely accepted, model was developed by Rock (1986), where a group of investors has information superior to that of any other investor as well as that of the issuer. Due to the issuer's uncertainty about the appropriateness of the offer price, informed investors crowd out uninformed investors for newly issued stocks with favourable prospects in order to maximise their profits. This causes the price to adjust from the offer price to the "true" value of the stock.

[63] Brav (2000) and Gompers and Lerner (2003) argue that the underperformance often observed in long-term post-IPO studies is, at least partially, due to the methodology used. While Brav (2000) suggests a Bayesian approach for statistical inference, Gompers and Lerner (2003) note that the underperformance observed using buy-and-hold abnormal returns (*BHAR*) disappears when cumulative abnormal returns (*CARs*) are used.

et al. 1977). While Edelman and Baker (1990) and Baker and Edelman (1990) have found no such pattern, other exchange listing studies have shown mixed results (Baker and Edelman 1992; Dharan and Ikenberry 1995; Edelman and Baker 1993, 1994; Elyasiani *et al.* 2000). In a modification of the original research design, several studies have examined stock price adjustments due to the start of dual listings on domestic or foreign stock exchanges (Baker *et al.* 1994, Bris *et al.* 2012, Roosenboom and Van Dijk 2009). The results of Roosenboom and Van Dijk (2009) and Baker *et al.* (1994) have confirmed the original pre- and post-listing pattern, respectively, while Bris *et al.* (2012) found evidence contradicting the post-listing pattern. The overall evidence from exchange listing studies on semi-strong form efficiency is therefore rather mixed.[64]

The seminal stock split study by Fama *et al.* (1969) impressively demonstrated that stock splits are associated with positive abnormal returns in the stocks concerned. Given the monthly data used in their study, Fama *et al.* (1969, p. 20) argued that the stock price adjustment due to the information inherent in the stock splits is "fully reflected in the price of a share at least by the end of the split month but most probably almost immediately after the announcement date."[65] This finding has sparked considerable interest in research on the behaviour of stock prices around stock split events. Subsequent studies have generally confirmed the initial results, *e.g.* Bar-Yosef and Brown (1977), Charest (1978), Hausman *et al.* (1971), Nichols and Brown (1981). Given, however, that these studies did not control for potential confounding information releases, Copeland (1979) presumed that the abnormal returns observed might

---

64 Exchange listing studies have attributed the occurrence of abnormal returns to various effects, including improved disclosures (Roosenboom and Van Dijk 2009), increased analyst and media coverage (Baker *et al.* 2002), and, in particular, changes in stock liquidity (Baker *et al.* 1994, Edelman and Baker 1993, Elyasiani *et al.* 2000).

65 The reason why stock splits cause the price of the stocks concerned to adjust has been a mystery to researchers for some time. This is because stock splits merely increase the number of shares outstanding by dividing the existing shares into smaller units. Therefore, a stock split *per se* has no economic effect on a firm's fundamentals. Fama *et al.* (1969) argue that stock splits have historically been associated with significant dividend increases, and that the positive stock price adjustments observed therefore reflect the implicit effects on the level of dividends. Nayak and Prabhala (2001) found support for this relationship, but also found that a significant portion of stock split-induced abnormal returns cannot be attributed to implied dividend effects. Over time, a number of alternative explanations have been put forward by other researchers. Leland and Pyle (1977), for instance, argue that stock splits signal favourable information about a firm's prospects from corporate insiders to outside investors. Lakonishok and Lev (1987) suggest that positive abnormal returns represent a market reward for stock splits because they keep the prices of the affected stocks in a "normal range". Similarly, Muscarella and Vetsuypens (1996) argue that positive stock price adjustments are due to improved liquidity effects. Lamoureux and Poon (1987) argue differently and suggest that the increased idiosyncratic volatility of a stock raises the tax-option value and therefore the price of the stock. In contrast, Grinblatt *et al.* (1984) attribute the abnormal returns observed to the increased attention drawn to splitting stocks, which causes positive revaluation effects.

have been distorted by other concurrent information events. However, this presumption was refuted by subsequent studies that have eliminated confounding information releases and have still found evidence of significant abnormal returns around stock splits (Grinblatt *et al.* 1984, Desai and Jain 1997, Lamoureux and Poon 1987, Muscarella and Vetsuypens 1996, Nichols and McDonald 1983).[66] More recent studies have focused more on the long-term effects of stock split events on stock prices, but have not provided substantial evidence against semi-strong form efficiency (*e.g.* Byun and Rozeff 2003, Hwang *et al.* 2008, Ikenberry and Ramnath 2002).

Similar to stock-split studies, research into stock price adjustments to accounting information was motivated by early seminal work; that is, the studies by Ball and Brown (1968) and Beaver (1968) on abnormal stock returns in response to firms' annual earnings anouncements. Beaver (1968) found that the magnitude of changes in the stock price (and trading volume) was much greater in the week of the earnings announcement than in other weeks.[67] This finding suggests that earnings announcements have information value for investors and is consistent with semi-strong form efficiency. At a more detailed level, Ball and Brown (1968) showed that the information contained in annual earnings announcements is valuable when the actual earnings differ from the expected earnings; then stock prices typically responded in the same direction (*post-earnings announcement drift*). However, Ball and Brown (1968) also showed that most (85 to 90%) of the information contained in annual earnings anouncements is already covered by more timely media, including interim reports, and is therefore known to investors before the annual earnings report is published. The post-earnings announcement drift was later confirmed by a large number of studies on quarterly earnings announcements, *e.g.* Bhushan (1994), Chen *et al.* (2017), Jones and Litzenberger (1970), Joy *et al.* (1977), Ke and Ramalingegowda (2005), Latané and Jones (1977, 1979), Rendleman *et al.* (1987), Truong (2011), and Watts (1978). Depending on the duration of the drift, abnormal stock returns in response to earnings announcements can be arbitrarily classified as efficient adjustments or inefficient underreactions.

---

66 Nichols and McDonald (1983) controlled for unexpected changes in corporate earnings. While they confirmed semi-strong form efficiency for stock splits of firms with moderate unexpected changes in corporate earnings, they identified anomalies for stock splits of firms that experienced large unexpected increases in corporate earnings.

67 It is worth mentioning that Beaver (1968) used a non-directional measure by squaring the abnormal returns and dividing them by their variance.

3.6 Rational Choice Theory and the Risk-Return Tradeoff

The opacity of banks (Section 3.4) creates information asymmetry between bank insiders and outside investors, which can lead to a mismatch between banks' risk profiles and stock prices. This is reflected in stronger co-movement or "price synchronicity" of bank stocks (Blau *et al.* 2020; Dewally and Shao 2013; Jones *et al.* 2012, 2013). The result is an informationally inefficient pooling equilibrium (Flannery *et al.* 2013, Myers and Majluf 1984) that represents quality uncertainty in the Akerlofian (1970) sense. Fama's (1970) semi-strong form efficiency (Section 3.5) asserts that new public information is immediately reflected in security prices. Therefore, disclosure of supervisory stress test results at bank level may enhance investors' ability to price-discriminate between sound and unsound banks (Jordan *et al.* 2000, Morgan 2002, Petrella and Resti 2013). Rational choice theory and the risk-return tradeoff of investments suggest that rational investors use such information in nuanced ways to maximise their utility; that is, to rebalance their prior risk-return expectations on individual bank stocks and act accordingly in the market to increase return or reduce risk. As a result, the quality uncertainty in bank stocks and the resulting price synchronicity that prevailed prior to the disclosure of supervisory stress test results could be temporarily resolved. To substantiate these considerations, the relevant literature on the risk-return tradeoff of investments (Section 3.6.1) and rational choices after the disclosure of stress test results (Section 3.6.2) is discussed below.

3.6.1 The Risk-Return Tradeoff of Investments

The notion that risk is rewarded with return is fundamental to many economic theories and models, such as modern portfolio theory (Markowitz 1952), the CAPM (Lintner 1965a, 1965b, Mossin 1966, Sharpe 1964), or the Sharpe (1966) ratio. Any deviation in a security's price from its level of risk should be quickly reversed by arbitrageurs and other market participants until the price is deemed commensurate with the level of risk. This is consistent with Fama's (1970) efficient market hypothesis, which implies that investors can be confident that a security's market price reflects a reasonable tradeoff between its risk and return. The principles of this tradeoff generally apply across and within asset classes (Campbell and Viceira 2005). Given the focus of this study on stocks, the evidence from the empirical literature on the risk-return tradeoff in this particular asset class is discussed below.

Overall, empirical evidence on the risk-return tradeoff in stocks is mixed, while recent studies clearly support the positive relationship proposed by economic theory. French *et al.* (1987), for example, found a positive and statistically significant relationship between stock returns and stock market volatility. Similarly, Chou (1988) and Chan *et al.* (1992) have found a small but significant positive risk-return relationship in the US stock market. In a test of the CAPM, Harvey (1989) confirmed that high returns were indeed associated with high conditional covariances. In contrast, Campbell (1987) and Glosten *et al.* (1993) have found a negative relationship between stock returns and their conditional variance. Campbell and Hentschel (1992) found that the relationship between stock returns and stock market volatility was positive for one part of the sample period and negative for another, but neither was statistically significant. This is consistent with Baillie and DeGennaro (1990) who concluded that any relationship between mean return and variance in a stock portfolio is weak, suggesting that investors consider other measures of risk to be more important than variance.

More recent studies have attempted to resolve the inconsistencies of previous research. Ghysels *et al.* (2005) introduced a new variance estimator that yielded a positive and significant risk-return tradeoff for the stock market. They also showed that the new estimator was more powerful in forecasting stock market variance than previously used rolling windows or GARCH estimators. Guo and Whitelaw (2006) decomposed the expected return into two separate components (a risk component and a hedge component) to estimate a relative risk aversion coefficient. They found a significantly positive risk-return relationship for the aggregate stock market and argued that omitting the hedge component was partly responsible for the conflicting results of previous studies. Based on a long sample period (1836-2003), Lundblad (2007) found a positive and statistically significant risk-return relationship in US and UK stock markets. He argued that a long data set was required to reliably identify the true risk-return tradeoff (which was typically not the case in previous studies). He also raised the possibility of a time-varying risk-return tradeoff based on the cyclicality of risk aversion and the increasing importance of idiosyncratic risk in stock pricing. Bali (2008) examined the relationship between expected return and risk for a wide variety of stock portfolios constructed based on industry, size, book-to-market ratio, and beta. His results showed a positive and highly significant risk-return relationship. Using the value premium of Fama and French (1996) as a proxy for time-varying investment opportunities, Guo *et al.* (2009) found a positive and significant risk-return tradeoff in the stock market.

They argued that previous specifications may have suffered from an omitted variable problem, leading to a downward bias in the estimate of the risk-return tradeoff. Kanas (2014) provided evidence for a strong and positive risk-return relationship in the S&P 100 index when the CBOE implied volatility index (VIX) is included as an exogenous variable in the conditional variance equation.

The above combined empirical evidence, particularly the more recent studies, suggests a positive relationship between risk and return in stocks, thereby confirming theoretical predictions. Building on this interim conclusion of the review, the next section discusses the rational choices available to stock investors in light of the results of supervisory stress tests, taking into account the risk-return tradeoff.

3.6.2 Information Shocks and Rational Investor Choices

In this section, the disclosure of supervisory stress test results is considered as an exogenous information shock that divides stock pricing into two periods: before and after the disclosure. First, the literature on the pre-disclosure equilibrium state is reviewed. Then, the existing literature on the post-disclosure period is discussed and related to rational choice theory. The basic assumption is that the new information from stress test results is immediately reflected in the affected banks' stock prices, consistent with Fama's (1970) semi-strong form efficiency.

The pre-disclosure or "normal" state of the bank stock market is characterised by a pooling equilibrium (Flannery *et al.* 2013) in which bank stocks exhibit a high degree of co-movement or price synchronicity (Blau *et al.* 2020; Dewally and Shao 2013; Jones *et al.* 2012, 2013). As Haggard and Howe (2012) showed, this is because investors compensate for the opacity-related lack of bank-specific information through common market and industry signals, thereby confirming Veldkamp's (2006) theoretical predictions.[68] Bank stocks therefore suffer from quality uncertainty in the Akerlofian (1970) sense, where asymmetric information causes adverse selection problems (Flannery *et al.* 2013, Myers and Majluf 1984). In other words, in the normal equilibrium state, investors' ability to price-discriminate between sound and unsound banks is impaired, leading to a more homogeneous movement in stock prices. This implies

[68] In Veldkamp's (2006) theoretical model, investors face high information costs and therefore only buy information for a subset of assets and then use that information to value other related assets. As a result, many investors rely on the same information, resulting in excessive asset co-movement.

that the risk-return tradeoff in the bank stock market is (to some extent) decoupled and informational efficiency is compromised (Blau *et al.* 2020; Jones *et al.* 2012, 2013). As a result, several studies have argued that bank stocks are prone to crashes and sharp revaluations triggered by changing investor risk perceptions (Dewally and Shao 2013; Haggard and Howe 2012; Jones *et al.* 2012, 2013). These studies have empirically supported Jin and Myers' (2006) theoretical model, which predicts that high opacity is associated with greater risk of extreme stock price movements.

Conversely, Jin and Myers' (2006) model suggests that better availability of firm-specific information (*e.g.* through information-generating measures such as supervisory stress tests) can improve the informational efficiency of stock prices and reduce their tendency to move in sync. This is consistent with Fama's (1970) semi-strong form efficiency and with Blackwell's (1951, 1953) theorem that more information is always better. Numerous studies have empirically shown that bank stock prices react to exogenous information shocks from supervisory authorities (*e.g.* Berger and Davies 1998, Blau *et al.* 2020, Bushee and Leuz 2005, Flannery and Houston 1999, Jordan *et al.* 2000). Given the objective of EU-wide stress tests to enhance market discipline (Section 2.4.3), differentiated price reactions in the stocks of affected banks to disclosed stress test results are politically desirable. In the post-disclosure state, in light of the bank-level stress test results, investors are presented with a set of three 9 for each bank stock: $A = \{\text{buy, hold, sell}\}$. Rational choice theory suggests that investors use the new information from the stress test results to revise their prior risk-return expectations and act accordingly in the market to reduce risk or increase return. This would lead to differentiated stock price reactions until the market believes that the levels of risk and return of each affected bank are commensurate. As a result, the quality uncertainty and the homogeneous movement of stock prices, which prevail in the banking market in the pre-disclosure state, could be temporarily resolved in the post-disclosure state, until the new stress test result information is fully reflected in the banks' stock prices and a new equilibrium is established. Such stock price adjustments are also consistent with Fama's (1970) efficient market hypothesis (Section 3.5).

Similar to the risk-return tradeoff of investments (Section 3.6.1), empirical evidence is mixed in the context of supervisory stress test results. The results of Ahnert *et al.* (2020) and Alves *et al.* (2015) on EU-wide stress tests, for example, have suggested a linear proportional relationship between banks' stress test results and abnormal stock returns. Fernandes *et al.* (2020) and Morgan *et al.* (2014) found similar results for US stress tests. However, several studies have provided evidence against a linear proportional relationship. For example, Sahin *et al.* (2020) showed that some banks' stock prices increased in response to the results of the US SCAP, regardless of their stress test result. In the European context, Georgescu *et al.* (2017) counterintuitively found that banks with weaker results in the 2016 EU-wide stress test experienced stronger positive abnormal stock returns. Georgoutsos and Moratis (2021) showed for the 2016 and 2018 EU-wide stress tests that CET 1, leverage, and profitability ratios were important determinants of abnormal stock returns and that there was a non-linear relationship for a specific subset of banks. However, previous studies have typically used simple dichotomous "pass vs. fail" comparisons to describe how the stocks of specific groups of banks responded to their stress test results. What is still missing is a study determining the general shape of the relationship curve between EU-wide stress test results and abnormal stock returns at bank level. Gaining a better understanding of whether investors have made nuanced use of EU-wide stress test results and whether the EBA's market discipline objective has been met is important for both investors and supervisors. This issue was addressed in Research Question 2 (*The Functional Relationship Hypothesis*) of this study.

3.7 Goodhart's Law on Financial Policy Indicators

Goodhart's law, as originally formulated, states that "any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes" (Goodhart 1975, p. 5). One way this can occur is when economic agents pursue their own goals and act independently of regulators in ways that undermine the goals and purposes of a particular regulation (Manheim and Garrabrant 2019, Sheng and Looi 2003). In this regard, Sheng and Looi (2003, p. 237) stated that "the setting of any particular regulatory rule will invite regulatory arbitrage or encourage innovation to circumvent the rules." Sheng and Looi (2003) therefore concluded that financial regulation involves a complex game between regulators and regulated institutions with multiple feedback mechanisms. These considerations might also be applicable to supervisory stress tests, where the supervisory intent to assess banks' resilience to adverse macro-financial conditions might be compromised by short-sighted bank actions to manipulate and improve their stress test performance. To review the relevant literature, the principles of Goodhart's law and other closely related concepts (Goodhart-like phenomena) are discussed in Section 3.7.1. Building on this, Section 3.7.2 reviews the existing evidence that these principles are effective in supervisory stress testing.

3.7.1 Goodhart-Like Phenomena

The origin of Goodhart's law lies in monetary policy. Goodhart originally formulated his law to explain why the Bank of England had failed to control inflation by adjusting the money supply in the 1970s (Goodhart 1975). Since then, Goodhart's law has greatly influenced monetary policy-making and the actions of central banks.[69] It has been repeatedly tested in monetary policy research (see, for example, Evans 1985, Fontana *et al.* 2020, Issing 1997), but has spread greatly to other areas, including finance and risk (Acharya and Thakor 2016, Daníelsson 2002, Thornton 2008), public policy (Hood and Piotrowska 2021, Tanzi 2013, Wellink 1996), and information processing (Beaulac 2022, Freeman and Soete 2009, Teney *et al.* 2020).

The broader interpretation of Goodhart's law is probably due to its generalisation by Strathern (1997, p. 308), who paraphrased it as "when a measure becomes a target, it ceases to be a good measure". Another explanation could be other concepts closely related to Goodhart's law that appeared around the same time. These concepts include Campbell's (1979) law and the Lucas (1976) critique. The former concerns unintended adverse effects of public policies and other government interventions and was formulated by Campbell (1979, p. 85) as follows:

> The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor.

Similarly, Lucas (1976) criticised econometric policy evaluation for estimating statistical relationships from historical data to predict the effects of a new policy, since correlations between aggregated variables tend to change when policy is changed. Lucas (1976, p. 41) summarised his critique as follows:

[69] For an overview on the origin, meaning and implications of Goodhart's law for monetary policy, see Chrystal and Mizen (2003).

> Given that the structure of an econometric model consists of optimal decision rules of economic agents, and that optimal decision rules vary systematically with changes in the structure of series relevant to the decision maker, it follows that any change in policy will systematically alter the structure of econometric models.

All three concepts mentioned above are largely congruent and hardly distinguishable. Therefore, Manheim and Garrabrant (2019) made an attempt to specify and structure the different underlying dynamics. They distinguished between four different effects: (1) regressional effects, where the selection of an imperfect proxy for a measured signal inevitably also selects noise, (2) extremal effects, where the selection of a metric shifts the state distribution into a region in which previous relationships no longer exist, (3) causal effects, where an action by the policy maker unintentionally causes a metric to break down, and (4) adversarial effects, where an economic agent with goals different from those of the policy maker causes a metric to break down. The adversarial effect is the only effect that takes multiple actors into account and thus best describes the dynamic space of supervisory stress tests.

In a generalised form, the adversarial effect can be described as a situation where a policy maker acts on an economic agent by using a metric to align the agent's goals with its own regulatory goals; the agent then responds by altering the previous causal structure due to imperfectly aligned goals in a way that creates a Goodhart effect (Manheim and Garrabrant 2019). In the case of supervisory stress tests, the supervisory goal of assessing banks' resilience to adverse macro-financial conditions could be undermined by banks' attempts to manipulate their stress test results to their advantage in order to avoid appropriate supervisory action. The above description by Manheim and Garrabrant (2019) implies that adversarial effect are associated with perverse incentives, *i.e.* unintended and undesirable effects that run counter to the intention of the policy maker. The existing empirical evidence for the presence of perverse incentives and their impact on the informational value of supervisory stress test results is reviewed in the next section.

3.7.2 Evidence Related to Supervisory Stress Testing

There is a broad consensus among researchers that the banking sector is subject to a variety of problems, including principal-agent problems, moral hazard, and asymmetric information (see, for example, Alexander 2006, Dell'Ariccia 2001, Nier and Baumann 2006). Daníelsson (2002) showed that such problems also have perverse effects on the properties of risk measures used for internal and external (regulatory) purposes. Using a sample of 2,500 data points (on average) for each major asset class (stocks, bonds, foreign exchange, and commodities), he found that risk measures were excessively volatile and lacking in robustness. In particular, he found that the regulatory value-at-risk (VaR) measure can provide misleading information and lead to perverse increases in idiosyncratic and systemic risk. In this context, Daníelsson (2002, p. 1276) also noted the following corollary to Goodhart's law: "A risk model breaks down when used for regulatory purposes." Consequently, he was also rather pessimistic about the feasibility of risk-based regulation (including Basel II and supervisory stress testing), where bank capital becomes a direct function of bank risk (Daníelsson 2003).

Goodhart (2016) raised similar concerns about the reliability of supervisory stress test results. While acknowledging the potential of supervisory stress tests, he emphasised their reliance on stressed capital ratios as a key metric for presenting the results, and reiterated the well-known problems of such ratio controls (Section 3.7.1). He succinctly argued that "banks will try to game the exercise by setting their resources at levels that will just satisfy the authorities's presumed requirements" (Goodhart 2016, p. 145). This is consistent with former Fed-Chairman Ben Bernanke's (2013) concern about attempts by banks to learn about and reverse-engineer supervisory stress testing models (see also Hirtle and Lehnert 2015, Leitner and Williams 2017, Schuermann 2020). Indeed, a number of studies have argued that perverse incentives in US stress tests have led banks to exploit learning effects, window dressing, and other suboptimal myopic behaviour, making US stress test results less informative over time (Cornett *et al.* 2020, Glasserman and Tangirala 2016, Goldstein and Sapra 2014). Similar concerns have also been raised recently in the European context, but have not been further explored (Kok *et al.* 2019, Quagliariello 2020). The findings of the above US studies are discussed in more detail below.

Goldstein and Sapra (2014) examined the relationship between *ex post* disclosure of stress test results and banks' *ex ante* behaviour. They argued that the banking sector's second-best environment – in which risks are opaque, difficult to verify, and prone to asset substitution – prompts banks to window dress their stress test performance through suboptimal myopic behaviour. In other words, they were concerned about *ex ante* behaviour that makes a bank appear sound in a stress test, but which is unsustainable and reduces the bank's long-term value. They also argued that this behaviour worsens over time as banks become more familiar with a stress test and the associated procedures.

This is consistent with Glasserman and Tangirala (2016) who argued that routine stress testing leads banks to optimise their decisions towards a certain supervisory capital ratio threshold, thereby implicitly creating new, harder to detect, risks. They also showed that the projected losses in the 2013 and 2014 DFAST stress tests were almost perfectly correlated for banks that participated in both exercises. Therefore, they concluded that the results of supervisory stress tests have become more predictable and thus less informative over time. This finding is consistent with the results of several capital market reaction studies that have discovered a decreasing trend in the informational value of US stress test results (Candelon and Sy 2015, Fernandes *et al.* 2020, Sahin *et al.* 2020).[70]

Examining a number of key figures, Cornett *et al.* (2020) found convincing evidence of manipulative bank behaviour in the context of US stress tests. Specifically, they found that stressed banks increased their capital ratios significantly more than non-stressed banks before the start of a stress test; this trend was completely reversed after the stress test. In addition, the different behavioural patterns of stressed and non-stressed banks could not be observed in 2010, when no stress test was conducted in the US. Their results also showed that stressed banks reduced their dividend payments significantly more when entering a stress test than non-stressed banks. Finally, they provided evidence that stressed banks invested significantly more in lobbying than non-stressed banks. Therefore, Cornett *et al.* (2020) concluded that stress-tested banks managed their financial performance and invested in political spending to improve their stress test performance.

[70] In contrast, using a non-directional approach, Flannery *et al.* (2017) showed that US stress test results have consistently provided investors with significant amounts of valuable new information (as discussed in detail in Section 3.3.2).

In a similar study, Bouwman *et al.* (2018) examined bank behaviour around the Dodd-Frank Act size thresholds, which entail different regulatory requirements for different bank sizes and imply higher net regulatory costs for above-threshold banks. They found that banks just below the size threshold reduced asset, risk-weighted asset, and loan growth to try to avoid or delay the higher regulatory costs associated with exceeding the threshold. In a very recent study, García and Steele (2022) found somewhat contradictory results. They showed that large US banks did not manipulate their bank size to avoid inclusion in CCAR stress tests, which are associated with increased regulatory oversight and stricter capital and transparency requirements.

In a related context, Garcia *et al.* (2021) examined the year-end balance sheet behaviour of global systemically important banks (G-SIBs) from the EU. They found that some G-SIBs contracted their year-end balance sheets to formally reduce their systemic importance, thereby mitigating the impact of G-SIB capital surcharges or avoiding G-SIB designation altogether. Notably, other systemically important institutions (O-SSIs) saw significantly smaller year-end balance sheet contractions, indicating window dressing activities among G-SIBs in the EU.

Overall, the above literature suggests that there is potential for perverse incentives in EU-wide stress tests, which could degrade the informational value of the stress test results over time. Developing an understanding of this is important for investors and supervisors alike in order to be able to respond appropriately. Therefore, it is worth examining the intertemporal stability of the informational value of EU-wide stress test results across the various excercises. This issue was addressed in Research Question 3 (*The Intertemporal Stability Hypothesis*) of this study.

3.8 Summary

In this chapter, the literature relevant to this study was reviewed. The literature review revealed significant gaps in previous studies, which constituted the research problems of this study (Section 1.2) and led to the formulation of its research questions and objectives (Section 1.4). This was based on the characteristics of the relevant line of research and the detailed review of previous studies on capital market reactions to supervisory stress test results in the EU and the US.

Thematically, the literature review showed that banks are particularly opaque entities where information asymmetry between insiders and outsiders is unusually high. This makes it difficult for outside investors to assess their fundamental value and price-discriminate between sound and unsound banks. It has also been shown that opacity-induced information uncertainty causes bank stocks to co-move more than stocks from other sectors ("price synchronicity").

Information-generating measures such as EU-wide stress tests can provide investors with new relevant information which, according to the semi-strong form of Fama's (1970) efficient market hypothesis, should be immediately reflected in the stock prices of the affected banks, thus improving their informational efficiency. As noted by Fama (1991, p. 1602) "[e]vent studies are the cleanest evidence we have on efficiency" because they allow for isolating the abnormal return associated with a specific information event. Indeed, the review of previous studies has confirmed that there is a broad consensus among researchers that event studies are the method of choice to study abnormal stock returns in response to supervisory stress test results. This justifies the use of event studies as the key method of this study (Section 5.3.2.1).

Rational choice theory and the risk-return tradeoff of investments suggest that investors revise their prior risk-return expectations about individual banks in the light of new information (*e.g.* from EU-wide stress test results). Economic theory suggests a linear shift along the Security Market Line (SML). Indeed, some empirical studies (*e.g.* Ahnert *et al.* 2020, Alves *et al.* 2015, Fernandes *et al.* 2020) have pointed to a linear proportional relationship between banks' stock returns and the risk represented by their stressed capital ratio (*i.e.* a bank's result on a supervisory stress test). However, other empirical studies have found contradictory results, indicating a non-linear risk-return relationship (*e.g.* Georgescu *et al.* 2017, Georgoutsos and Moratis 2021, Sahin *et al.* 2020).

Finally, several studies have found a decreasing trend in informational value of US stress test results across the various exercises. This has been attributed to perverse incentives in US stress tests that created Goodhart effects and caused stress test results to become less informative over time. Similar concerns have also been raised recently in the European context but have not been further explored (Kok *et al.* 2019, Quagliariello 2020).

Given that previous studies on US and EU-wide stress tests have hardly made any reference to the underlying theory, the theories, constructs, and debates presented in this chapter have been used to develop a dedicated theoretical framework for studying abnormal bank stock returns in response to supervisory transparency measures. This framework is presented in the following chapter.

Chapter 4
Theoretical Framework and Hypotheses Development

4.1 Introduction

This chapter develops a dedicated theoretical framework for studying abnormal bank stock returns in response to EU-wide stress test results and formulates empirically testable hypotheses for each of the research questions. The framework is developed in Section 4.2 based on the definitions and principles of Grant and Osanloo (2014) and Imenda (2014). It builds on the literature review and was deliberately designed to be extensible and applicable beyond this study, to open avenues for future research. This is because previous studies have failed to construct a clearly specified theoretical framework that could be used to study banks' abnormal stock returns in response to supervisory transparency measures (Section 3.3.1). The framework provided in this chapter also facilitates the development of empirically testable hypotheses. Building on this, Section 4.3 formulates a set of null and alternative hypotheses for each research question. Section 4.4 summarises the chapter.

4.2 Theoretical Framework

The theoretical framework builds on and connects the theories, constructs, and debates discussed in the literature review (Chapter 3) to guide the investigation of the research problems (Section 1.2). It contains the following elements: bank opacity, information uncertainty, the risk-return tradeoff of investments, rational choice theory, Goodhart's law, and the efficient market hypothesis as the key formal theory.

The graphical representation of the framework extends over a horizontal and a vertical dimension. The horizontal "spine" of the framework is based on a classic test of semi-strong form efficiency (*i.e.* the impact of the new information from EU-wide stress test results on bank stock prices) with bank opacity and information uncertainty as antecedents, representing the *status quo* prior to the disclosure of new supervisory information. This provided the theoretical foundation for Research Question 1. Additional vertical connections formed the basis for Research Questions 2 and 3. These connections introduced rational choice theory and the risk-return tradeoff of investments as mediators that could explain the process by which new supervisory information (*e.g.* EU-wide stress test results) affects bank stock prices. At the same time, Goodhart's law was introduced as a moderator that could affect the relationship between recurring supervisory information disclosures (*e.g.* regular EU-wide stress tests) and banks' corresponding abnormal stock returns over time. Since the internal decision-making process of investors and the stress-test related behaviour of banks cannot be observed directly, they are treated as latent variables and are inferred from the observed stock price behaviour. Figure 5 illustrates the structure of the framework.

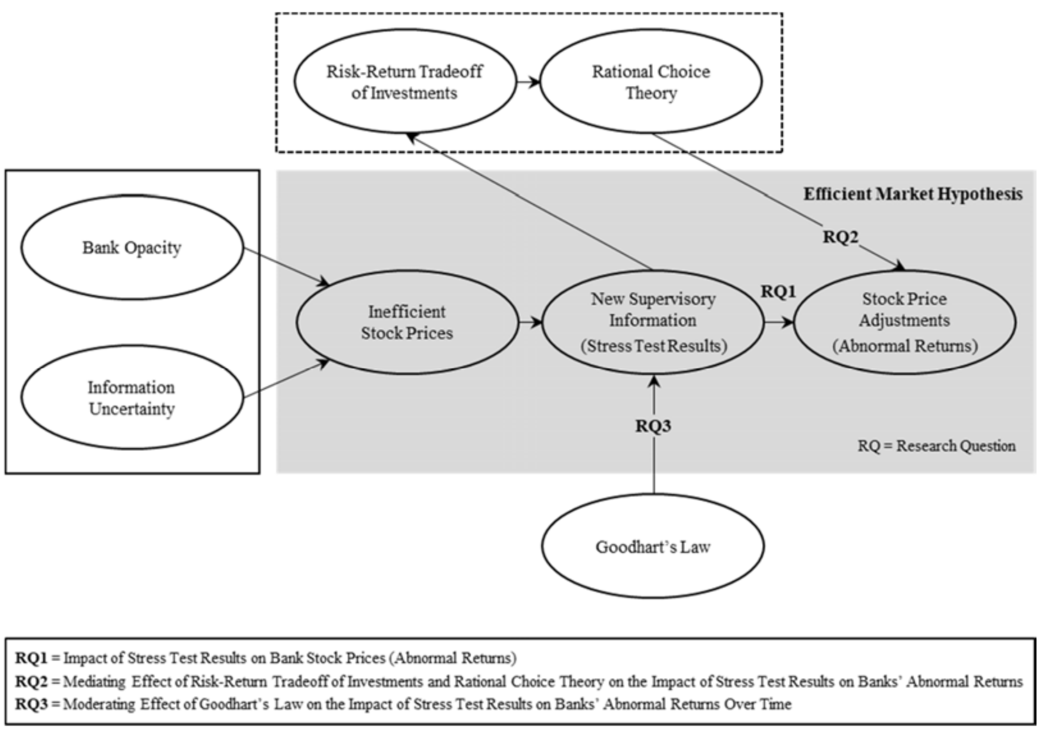


Figure 5. Theoretical framework for studying abnormal stock returns of banks in response to supervisory transparency measures

Although the framework was developed to meet the needs of this study, it was deliberately designed to be applicable and extensible for future research. For example, by adding behavioural elements such as investor sentiment and risk perception. In the following, the current state of the framework is described in more detail by explaining the connections between the underlying theories, constructs, and debates. The explanations start with the horizontal spine of the framework, *i.e.* with the theoretical basis for Research Question 1. Building on this, the vertical connections extending upwards and downwards are explained as the foundation for Research Questions 2 and 3.

There is a broad consensus among researchers that banks are opaque and that this opacity is significantly greater than that of non-banking firms (see the summary of the bank opacity literature in Table 5). This means that the information asymmetry between insiders and outside investors is particularly high in the banking sector. As a result, outsiders are faced with a high degree of uncertainty about the prospects and the actual risk exposure of banks (Anolli *et al.* 2014, Fosu *et al.* 2017, Morgan 2002). This affects the ability of investors to accurately determine the fundamental value of a bank, and makes them less likely to discriminate between sound and unsound banks (Blau *et al.* 2020, Dewally and Shao 2013, Jones *et al.* 2013). The resulting co-movement, or "price synchronicity", means that bank stocks tend to reflect average quality in the Akerlofian sense rather than the specific characteristics of a bank. All of this suggests that bank stock pricing is not informationally efficient. It has therefore been argued that there is justification for regulating bank transparency through information-generating measures such as supervisory stress tests (Jordan *et al.* 2000, Morgan 2002, Petrella and Resti 2013). Indeed, supervisory stress tests in the US and EU have disclosed unprecedented amounts of formerly confidential supervisory information to the public (Hirtle and Lehnert 2015, Petrella and Resti 2013, Schuermann 2014). If these disclosures did in fact provide investors with new relevant information, then according to the semi-strong form of Fama's (1970) efficient market hypothesis, this should have caused significant price adjustments in the stocks of the banks concerned.

In extension to that, rational choice theory and the risk-return tradeoff of investments suggest that rational investors would use the new information from EU-wide stress tests to maximise their utility. That is, to rebalance their prior risk-return expectations on individual bank stocks and act accordingly in the market to increase return or reduce risk. Based on their improved ability to discriminate between sound and unsound banks, investors can choose from a set of three possible alternatives for each bank stock: $A = \{\text{buy, hold, sell}\}$. The rationality assumption of rational choice theory predicts that the direction and magnitude of the corresponding abnormal stock return should be linearly related to each bank's actual stress test result. Assigning units of return (cumulative abnormal returns ($CARs$)) to units of risk (capital ratio differences (ΔCRs) between stressed and actual capital ratios) yields a relationship curve whose general shape can be used to functionally describe how EU-wide stress test results and abnormal stock returns are related (*e.g.* whether the relationship is linear or non-linear).[71] Figure 6 schematically illustrates two selected possible outcomes.

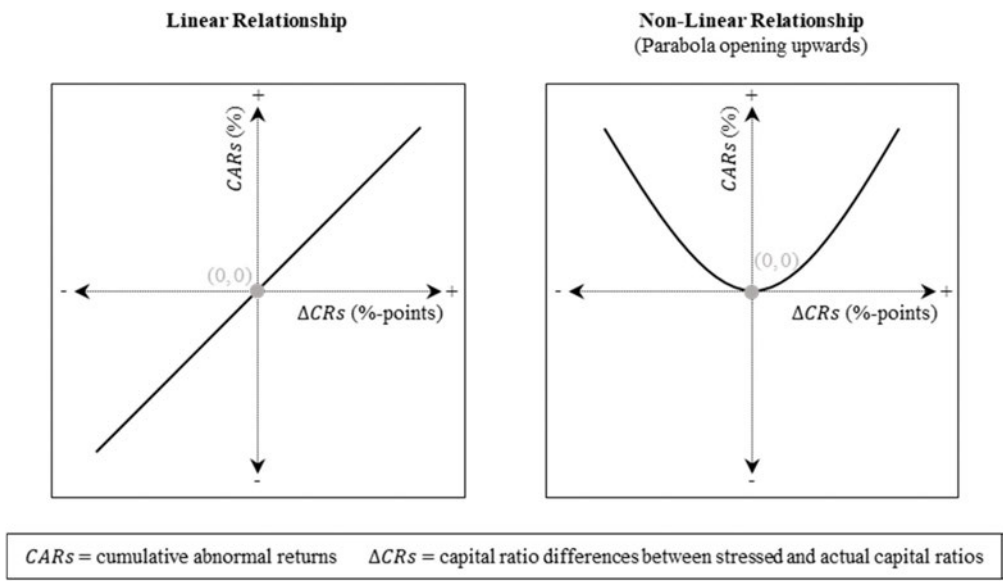


Figure 6. Schematic representation of two selected possible functional relationships between capital ratio differences (ΔCRs) on the x-axes and cumulative abnormal returns ($CARs$) on the y-axes

Another extension is the introduction of the time dimension by considering Goodhart's law on financial policy indicators. In essence, Goodhart's law suggests that any attempt to use statistical data or models for regulatory purposes creates perverse incentives that render those data and models uninformative over time (Goodhart 1975). This effect has been observed for internal and external (regulatory) risk models, indicating a general lack of time-robustness in such approaches (Daníelsson 2002). More specific studies have shown that the informational value of US stress tests has decreased over time (Candelon and Sy 2015, Fernandes *et al.* 2020, Sahin *et al.* 2020). This has been attributed to perverse incentives in US stress tests and corresponding

[71] For more information on the specific variables that were related, see Section 5.3.2.3.1.

attempts by banks to improve their stress test results by exploiting learning effects, window dressing, and other suboptimal myopic behaviour (Cornett *et al.* 2020, Glasserman and Tangirala 2016, Goldstein and Sapra 2014). In the presence of such effects, it is almost impossible for investors to distinguish whether a stress test result reflects a bank's actual capital adequacy or the result of manipulative actions. Therefore, such Goodhart effects cause the informational value of supervisory stress test results to deteriorate over time. In the EU, the risk of perverse incentives is arguably lower than in the US due to the Constrained Bottom-Up Appraoch and the Quality Assurance Process inherent in EU-wide stress tests (Quagliariello 2020). However, the mere possibility of perverse incentives raises the question of whether and how the informational value of EU-wide stress test results has changed over time. More specifically, whether the informational value of EU-wide stress test results has been intertemporally stable or has been subject to a decreasing trend.

4.3 Hypotheses Development

Based on the above theoretical framework and the literature review more generally, a set of null and alternative hypotheses was developed for each research question. This provided the starting point for a series of *a priori* hypothesis tests. To facilitate readability, the hypotheses sets associated with the research questions were given unique names. The hypotheses associated with Research Questions 1, 2, and 3 are referred to as the *Informational Value Hypothesis*, the *Functional Relationship Hypothesis*, and the *Intertemporal Stability Hypothesis*, respectively. They are all described in more detail below.

4.3.1 The Informational Value Hypothesis

Research Question 1 asked whether the results of EU-wide stress tests provided bank stock investors with new relevant information and what the *average* value of this information actually was. This was consistent with testing semi-strong form efficiency in an opaque and uncertain environment (Section 4.2). In short, if sample banks were opaque and EU-wide stress tests generated valuable new information, then, according to semi-strong form efficiency, stock prices should have adjusted upon disclosure. Accordingly, the average informational value of EU-wide stress test results was examined independently of the individual stress test results of the sample banks.

In order to turn the above into testable hypotheses, a quantitative definition of informational value was required. Following common practice in event studies, the average cumulative abnormal return ($\overline{CAR}$) of the sample banks was defined as a proxy for the informational value and was tested against zero. Furthermore, as suggested by Flannery *et al.* (2017), the average *absolute* cumulative abnormal return ($|\overline{CAR}|$) was used as an additional non-directional measure.[72] Testing the $|\overline{CAR}|$ against zero would have been inappropriate because of its absolute value. Following Flannery *et al.* (2017) and Georgoutsos and Moratis (2021), they were therefore tested against the average absolute estimation error of the normal return-generating model used (for a more detailed explanation, see Section 5.3.2.2.3). On this basis, the following set of null hypotheses $\{H_{0_1}, H_{0_2}\}$

$$H_{0_1}: \overline{CAR} = 0 \tag{7}$$

$$H_{0_2}: |\overline{CAR}| = |\gamma| \tag{8}$$

was tested against the corresponding set of alternative hypotheses $\{H_{A_1}, H_{A_2}\}$

$$H_{A_1}: \overline{CAR} \neq 0 \tag{9}$$

$$H_{A_2}: |\overline{CAR}| \neq |\gamma|, \tag{10}$$

where $\overline{CAR}$ is the average cumulative abnormal return and $|\overline{CAR}|$ is the average absolute cumulative abnormal return of the sample banks' stocks, and $|\gamma|$ is the average absolute estimation error of the normal return-generating model used to estimate normal (expected) returns.

4.3.2 The Functional Relationship Hypothesis

Research Question 2 asked about the functional relationship between the new information contained in EU-wide stress test results and the corresponding abnormal stock returns of the sample banks. It was therefore a test of rational choice theory in the context of the risk-return tradeoff of investments (Section 4.2). In short, if disclosure of EU-wide stress test results did improve price-discrimination and market discipline,

[72] For more information, see Section 5.3.2.1.5.

then rational investors would have revised their prior risk-return expectations and stock prices would have adjusted linearly in proportion to the stress test result of each bank. Accordingly, the informational value of EU-wide stress test results was examined in relation to the individual stress test results of the sample banks. More specifically, the sample banks' cumulative abnormal returns ($CARs$) were related to the differences between their stressed and actual capital ratios (ΔCRs), and then subjected to polynomial curve fitting (see Section 5.3.2.3). The stressed capital ratios corresponded to the bank-level results of a given EU-wide stress test under the adverse scenario, while the actual capital ratios were taken from the financial statements of the banks at the end of the fiscal year before the respective stress test. The reason for taking the capital ratio difference was to capture the impact of the stress tests in units that could be compared across the sample banks. Further details on the related variables and data collection are provided in Section 5.3.2.3.1 and Section 5.3.2.1.2, respectively.

Turning the above into testable hypotheses required defining the bounds of the permissible results. In order to avoid unstable oscillation (Runge's phenomenon) and to keep the relationship economically interpretable, the permissible results were constrained to first- and second-degree polynomials (*i.e.* linear or quadratic relationships with parabolas opening upwards or downwards). Based on the predictions of rational choice theory and the risk-return tradeoff of investments, it seemed reasonable to assume a linear relationship as the most likely outcome and define it as the null hypothesis. The alternative hypothesis represented the logical complement to this and assumed a non-linear relationship, *i.e.* that extreme (positive or negative) stress test results (measured as differences between banks' stressed and actual capital ratios (ΔCR)) were associated with disproportionate (positive or negative) cumulative abnormal returns. Accordingly, the following null hypothesis (H_0)

$$H_0\colon s = ax + b \tag{11}$$

was tested against the corresponding alternative hypothesis (H_A)

$$H_A\colon s \neq ax + b, \tag{12}$$

where $s = \left\{\widehat{CAR}_1, \dots, \widehat{CAR}_n\right\}$ is a set of fitted cumulative abnormal returns $\left(\widehat{CAR}\right)$ of each sample bank i, (*i.e.* the fitted stock price reactions), $x = \{\Delta CR_1, \dots, \Delta CR_n\}$ is a set of corresponding capital ratio differences (ΔCR) between the stressed and actual capital ratios of the sample banks, and a and b are the slope and intercept coefficients.

Because of the above constraint on first- and second-degree polynomials, any rejection of the null hypothesis (H_0) in favour of the alternative hypothesis (H_A) implied that the best permissible fit was a quadratic relationship (second-degree polynomial) with a parabola opening upwards or downwards, *i.e.*:

$$s = ax^2 + bx + c \quad \begin{cases} -a, & a < 0 \\ a, & a > 0' \end{cases} \tag{13}$$

where ax^2, bx, and c are the quadratic, linear, and constant terms of the polynomial, respectively.

4.3.3 The Intertemporal Stability Hypothesis

Research Question 3 asked about the change in the informational value of EU-wide stress test results over time. This implied testing Goodhart's law on the informational value of EU-wide stress test results, *i.e.* whether the value of that information has decreased over the course of the five EU-wide stress tests (Section 4.2). In short, if the results of the EU-wide stress tests were subject to Goodhart's law, then their informational value should have decreased from earlier to later stress tests, showing an overall decreasing trend. Accordingly, it was examined whether the informational value of EU-wide stress test results was intertemporally stable.

To turn the above into testable hypotheses, the elements of the two finite sequences of $\overline{CAR}_j$ and $|\overline{CAR}|_j$ with the time-ordered set of five EU-wide stress tests $j =$ {CEBS 2010, EBA 2011, EBA 2014, EBA 2016, EBA 2018} were defined as proxies for the informational value of the EU-wide stress tests (see Section 4.3.1). In order to reflect the longitudinal nature of Research Question 3, the analysis was based on longitudinal panel data and a corresponding sample (Section 5.3.2.1.2). Building on the above, the following set of null hypotheses $\{H_{0_1}, H_{0_2}\}$

$$H_{0_1}: \overline{CAR}_{j,t} > \overline{CAR}_{j,T} \tag{14}$$

$$H_{0_2}: |\overline{CAR}|_{j,t} > |\overline{CAR}|_{j,T} \tag{15}$$

was tested against the corresponding set of alternative hypotheses $\{H_{A_1}, H_{A_2}\}$

$$H_{A_1}: \overline{CAR}_{j,t} \leq \overline{CAR}_{j,T} \tag{16}$$

$$H_{A_2}: |\overline{CAR}|_{j,t} \leq |\overline{CAR}|_{j,T}, \tag{17}$$

where $\overline{CAR}$ is the average cumulative abnormal return and $|\overline{CAR}|$ is the average absolute cumulative abnormal return of the sample banks' stocks for a given EU-wide stress test from the time-ordered set $j = \{$CEBS 2010, EBA 2011, EBA 2014, EBA 2016, EBA 2018$\}$, where t indicates precedence and T indicates succession in time.

4.4 Summary

In the first part of this chapter, a dedicated theoretical framework was developed to provide a theoretical basis for studying abnormal bank stock returns in response to supervisory transparency measures. Although the framework was developed for the specific needs of this study, it was deliberately designed to be applicable and extensible beyond that. The framework is based on the literature review and on the definitions and principles of Grant and Osanloo (2014) and Imenda (2014). It also provides an inventory of relevant theories, constructs, and debates, *i.e.* the building blocks of this study. More specifically, the framework builds on and connects the following elements: bank opacity, information uncertainty, the risk-return tradeoff of investments, rational choice theory, Goodhart's law, and the efficient market hypothesis. Each element of the framework was derived directly from the research questions (Section 1.4). This facilitated the formulation of testable hypotheses, guided the design and conduct of the study, and provided an organisational structure for reporting the results.

In the second part of this chapter, the framework was used to transform the research questions into empirically testable hypotheses. Research Questions 1 and 2 were *state* problems examining the state of the informational value of EU-wide stress test results at a specific point in time, *i.e.* closely around the result disclosure dates of the five EU-wide stress tests. Specifically, the *Informational Value Hypothesis* (Research Question 1) assumed that the new information from EU-wide stress test results was valuable for investors and thus, according to semi-strong form efficiency, caused statistically significant abnormal stock returns in the cross-section of the banks concerned. The *Functional Relationship Hypothesis* (Research Question 2) took a more granular view and assumed that the stress test results and abnormal stock returns of the individual banks were related in a non-linear disproportional way, *i.e.* that banks with more extreme (positive or negative) results experienced disproportionate (positive or negative) abnormal returns. Finally, Research Question 3 was a *process* problem examining how the informational value of EU-wide stress test results has changed over time. Specifically, the *Intertemporal Stability Hypothesis* (Research Question 3) assumed that the informational value of EU-wide stress test results was not subject to Goodhart's law and thus did not decrease over the course of the five exercises.

Chapter 5
Methodology

5.1 Introduction

The purpose of this chapter is to present the methodology of this study. The discussion is divided into two main parts: the research philosophy (Section 5.2) and the research design (Section 5.3). These two parts are related because the underlying philosophical assumptions inform and guide the research design of a study (Saunders *et al*. 2009a, 2009b). Figure 7 signposts a visual summary of the research philosophy and the research design of this study.

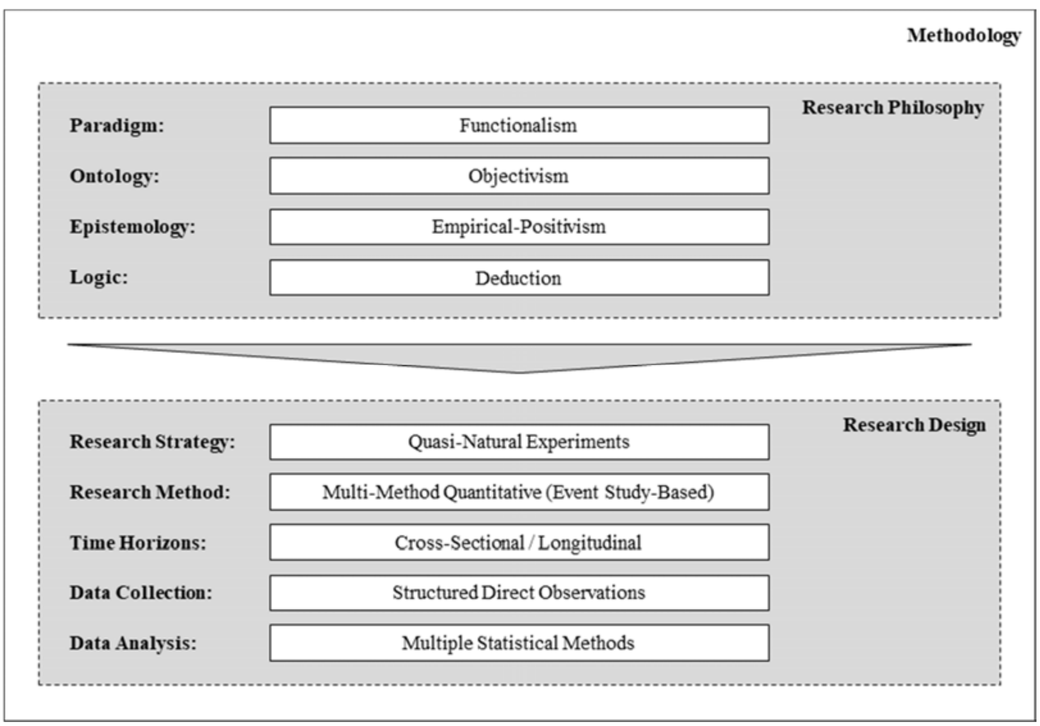


Figure 7. Research philosophy and research design

The organisation of this chapter follows the arrangement of the elements in Figure 7 from top to bottom, with the last three elements presented together with the research method in Section 5.3.2. In other words, the research philosophy and the research design of this study are introduced gradually in successive sections. The structure of the chapter therefore resembles the structure of the well-known "research onion" by Saunders *et al.* (2009a, 2009b) when viewed from the outside in. Finally, the methodological particularities of each of the three research questions are described. The chapter concludes with a brief discussion of alternative research paradigms that could have been used in this study.

5.2 Research Philosophy

This section specifies the philosophical positions of this study and shows how they underpin the research design (Section 5.3). To achieve this objective, the philosophical choices made during the research process are discussed. The discussion begins with the research paradigm chosen and continues with the decisions made for three interrelated philosophical concepts: ontology, epistemology, and logic (O'Gorman and MacIntosh 2015, Saunders *et al.* 2009a). This section is structured accordingly, presenting each philosophical choice separately below.

5.2.1 Functionalist Paradigm

In the social sciences, observations of phenomena, meanings, and interpretations are influenced by researchers' beliefs and assumptions about the nature of the world, the place of the individual in it, and the range of possible relationships between them (Lincoln *et al.* 2017). In this respect, regardless of whether or not researchers explicitly acknowledge paradigmatic assumptions, they make them and use them to develop and apply theory (Laughlin 1995, Schultz and Hatch 1996).

However, the term "paradigm" has been used ambiguously in scientific research and therefore requires clarification. The way it is used in this study is consistent with the well-established definitions of Kuhn (1962) and Burrell and Morgan (1979). In his seminal book, Kuhn (1962, p. 45) defined a paradigm as "the set of common beliefs and agreement shared between scientists about how problems should be understood and addressed." Similarly, Burrell and Morgan (1979, p. 23) described a paradigm as "the commonality of perspective which binds the work of a group of theorists together in such a way that they can be usefully regarded as approaching social theory within the bounds of the same problematic."[73] Based on this description, Burrell and Morgan (1979) developed a coherent framework for classifying paradigms in the social sciences along researchers' meta-theoretical assumptions about two dimensions: the nature of social science (subjective-objective dimension) and the nature of society (regulation-radical change dimension).[74] Figure 8 shows how the two dimensions are arranged to form four distinct paradigms and lists the names of philosophers who have made major contributions to each of the paradigms.

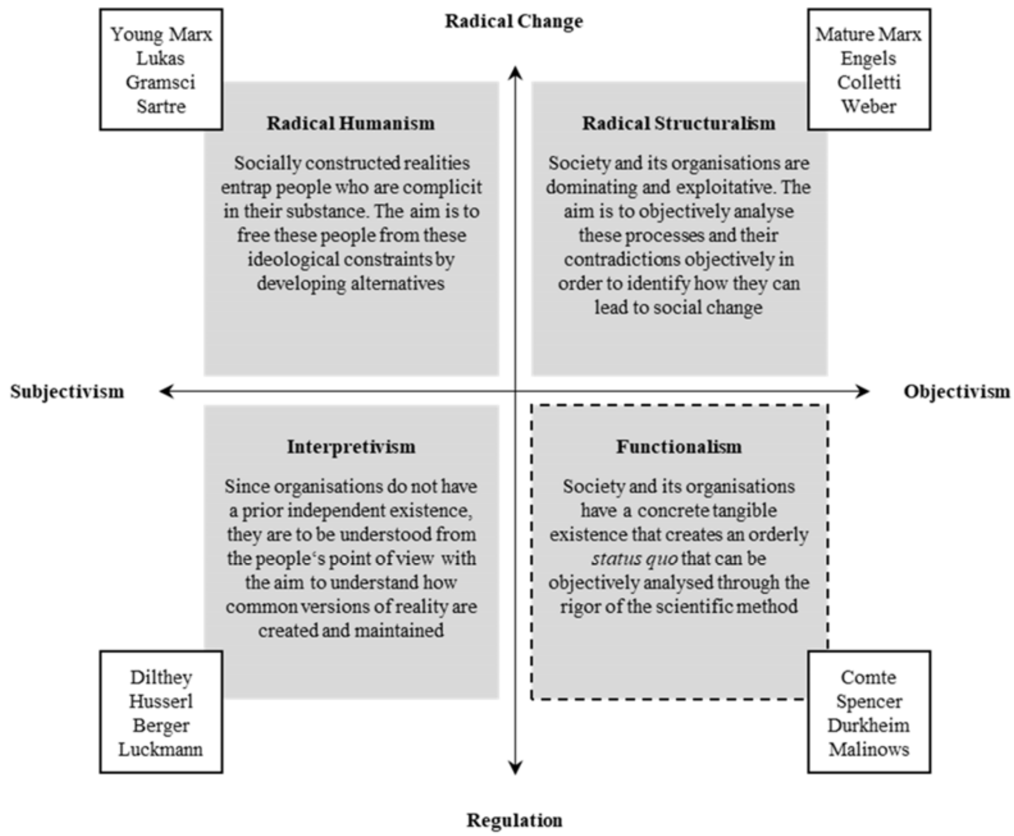


Figure 8. Four paradigms for the analysis of social theory

73 More recent work has used similar definitions, Saunders *et al.* (2009a, p. 118), for example, described a paradigm as "a way of examining social phenomena from which particular understandings of these phenomena can be gained and explanations attempted."

74 For a suggestion of two complementary paradigms (*i.e.* radical emergence and radical verificationism), see Callaghan (2017).

Burrell and Morgan (1979, p. 23) presented the four paradigms as "continuous but separate"; that is, one dimension of the framework is shared while the other is different. The subjective-objective dimension indicates the ontological position of a researcher, *i.e.* whether it is believed that reality depends on the consciousness and cognition of the individual or is hard and objective. The regulation-radical change dimension, in turn, indicates the researcher's assumption about society, *i.e.* whether it is believed that society is seeking to maintain the *status quo* or aiming for radical change.

This study is based on functionalism, which is the dominant paradigm in finance research (Ardalan 2003, Gendron and Smith-Lacroix 2015, Rao 2019) and in business and management research in general (Burrell and Morgan 1979, Saunders *et al.* 2009a). The dominance of the functionalist paradigm in finance suggests that it is particularly well suited to address financial research questions. This is due to its objectivist and regulatory assumptions (see Figure 8). More specifically, the functionalist paradigm attempts to explain the socio-economic *status quo* with an objectivist ontology, which views reality as external to the individual. In addition, functionalists typically take a positivist epistemological stance and embrace the scientific method by focusing on empirical evidence, hypothesis testing, and the use of quantitative methods to collect and analyse data (Burrell and Morgan 1979, Goles and Hirschheim 2000, Rao 2019). This study examines the formation of empirically observable stock prices in response to EU-wide stress test results from 2010 to 2018 under the currently prevailing conditions of reality (see Sections 1.3 and 1.4). It takes an objectivist ontological stance (Section 5.2.2) and an empirical-positivist epistemological perspective (Section 5.2.3) and uses a range of quantitative methods and techniques to collect and analyse data (Section 5.3). The functionalist paradigm is therefore consistent with the purpose of this study, its philosophical positions, and its research design. The adoption of the functionalist paradigm forms the philosophical basis of this study. Building on this fundamental decision, the philosophical and methodological choices of this study are explained in more detail in the following sections.

5.2.2 Objectivist Ontology

A researcher's ontological position is an expression of assumptions about the nature of reality (O'Gorman and MacIntosh 2015, Saunders *et al.* 2009a). In accordance with its functionalist paradigm, this study took an objectivist ontological stance. This means that the study was carried out from the position of a detached observer who stands outside the research situation and takes a "view from nowhere" (Nagel 1989, p. 70).[75] In other words, the study was conducted without any interaction with the research subjects. This was consistent with the study's quasi-natural experimental research strategy (Section 5.3.1), which is based on exogenous events that do not involve interaction or manipulation by the researcher. In concrete terms, this meant that the study assumed that EU-wide stress tests (exogenous events) exist and are carried out independently of this study, *i.e.* they occur quasi-naturally. The implications of the study's ontological position on the research design and process are described below.

This study assumed that "social entities exist in reality external to social actors" (Saunders *et al.* 2009a, p 110). That is, reality was viewed as existing of "objects that can be measured and tested, and which exist even when we are not directly perceiving or experiencing them" (O'Gorman and MacIntosh 2015, p. 56). For example, the banks that were subjected to EU-wide stress tests were viewed as social entities that exist independently. Similarly, the stocks representing ownership of these banks and the prices assigned to them exist in reality before and outside of this study and are therefore objective. While stock prices are viewed as concrete entities falling within the ontological category of *being*, changes in them (*i.e.* stock returns) are viewed as abstract entities associated with the ontological category of *becoming* (Rosen 2020). This is consistent with Heraclitus' aphorism *panta rhei* ("everything flows"), which is a metaphor for the processuality of the world.

The investors whose actions in the market cause stock prices to change are assumed to act rationally and according to the principle of profit maximisation or, to put it philosophically, according to *instrumental rationality* (Kolodny and Brunero 2020). This implies that investors who were buying or selling bank stocks at the time of the EU-wide stress tests were not affected in any way by this study.

[75] Williams (2011, p. 153) offers a similar construct, which he refers to as "absolute conception". For a critical perspective on Nagel's (1989, p. 70) "view from nowhere", see Metzinger (2011).

The research process of this study follows the ideal of the scientific method and assumes that it is possible to establish scientific facts through robust and reproducible methods. According to Popper (2002, p. 22), "the objectivity of scientific statements lies in the fact that they can be inter-subjectively tested." This can be understood as a call to procedural objectivity as a necessary condition for reproducibility, testability, and criticism. Throughout the entire research process, this study endeavours to eliminate subjectivity and bias as far as possible through the use of quantitative rules and control measures (Porter 1995). Examples include the study's systematic approach to model selection (Section 5.3.2.1.3), the statistically determined length of post-event windows (Section 5.3.2.1.1), and the extensive use of robustness checks. Consequently, this research follows the Fisherian (frequentist) approach to inference, as the alternative Bayesian approach is viewed as non-objective and prone to bias due to its dependence on a subjective prior.[76] In contrast, this study is intended to be conducted in a value-free and unbiased manner, with the researcher being independent of the data and maintaining an objective ontological stance. This attitude facilitates reproducibility and testability in the Popperian sense and is supported by the use of quantitative methods for collecting and analysing the data.

5.2.3 Empirical-Positivist Epistemology

Epistemology is concerned with a researcher's view on the nature, origin, and scope of knowledge. This study takes an empirical-positivist epistemological position. A useful way to explain this position is through the traditional analysis of knowledge, commonly known as Justified True Belief (JTB), which was introduced in Plato's *Theaetetus* and is still widely used today (*e.g.* Powell 2020, Thicke 2018, White 2017). According to the JTB account of knowledge, justification, truth, and belief are conditions for knowledge that are individually necessary and jointly sufficient (Audi 2011, Shope 2017).[77]

---

[76] For a detailed discussion of the objections to Bayesian statistics, see Gelman (2008).

[77] For an entirely different approach, see Williamson (2002). In this context, it is also worth mentioning the Gettier Problem, according to which there may be cases where the JTB account of knowledge is insufficient to make a claim to knowledge, since the reason for the belief, although justified, might turn out to be false (Gettier 1963).

From the tripartite analysis of the JTB conditions follows that: (1) if a knowledge claim is true, and (2) if the researcher believes that the claim is true, and (3) if the researcher is justified to believe that the claim is true, then and only then is the claim accepted as true knowledge (Audi 2011, Shope 2017). The latter two conditions, truth and belief, are hardly controversial. This is because "truth" is a metaphysical rather than an epistemological concept, *i.e.* it is about how things *are*, not about how they can be *seen* (Ichikawa and Steup 2018). The belief condition, on the other hand, requires outright belief; failure to fully believe a claim precludes knowledge (Ichikawa and Steup 2018). A true belief can be justified by several concepts.[78] In this study the sufficient likelihood approach to justification is adopted, according to which a researcher is justified to believe a knowledge claim if, and only if, the researcher believes the claim in a way that makes the belief sufficiently likely to be true (Steup and Neta 2020). There are, again, several views on how sufficient likelihood can be assessed. This study follows the reliabilist view, which goes beyond the evidentialist view by not only requiring evidence, but also that such evidence results from a reliable process or source (Goldman 1986, Ichikawa and Steup 2018). The reliabilist approach to knowledge and justification resembles the scientific method best and therefore corresponds to the empirical-positivist epistemological stance taken in this study.

The above epistemological positions of this study become evident throughout the entire research process. First, the research strategy (quasi-natural experiments, see Section 5.3.1) is consistent with the positivist assumption that "the researcher is independent of and neither affects nor is affected by the subject of the research" (Remenyi *et al.* 1998, p. 33). This assumption establishes an important connection with the objectivist stance of the study and its intention to conduct the research in a value-free and unbiased manner (Section 5.2.2). This is also reflected in the data collection process. The data used in this study are passively collected through empirical observations of stock prices and bank-level results of EU-wide stress tests, that is, through structured direct observations of social reality (Section 5.3.2.1.2). This is consistent with the empirical-positivist view that only observable phenomena can provide factual data (Saunders *et al.* 2009a). The analysis of the data follows a highly structured process and includes the use of event studies and multiple statistical methods to test the formulated

[78] For an overview of concepts of epistemic justification, see Alston (1989) and Steup and Neta (2020).

hypotheses (Sections 4.3 and 5.3). In order to establish a sufficient likelihood that justifies to believe an empirical result to be true, the structured methodological approach of this study is supplemented by a set of significance tests and robustness checks (see Section 5.3.2). All of these are key components of the scientific method (Kosso 2011) and characteristic of a positivist methodological approach to research (Gill and Johnson 2010, Saunders *et al.* 2009a). As a result, based on its empirical-positivist epistemological stance, this study aims to contribute to the advancement of theory by producing positive *a posteriori* knowledge that is justifiably believed to be true.

5.2.4 Deductive Logic

Logic is the branch of philosophy that studies valid rules of inference or, in other words, the way in which formal reasoning can be used to produce logically valid arguments. In this study, deductive logic, or deductive reasoning, is used to answer the research questions. Deductive reasoning is the process of inferring logically certain conclusions from a given set of true premises (Schechter 2013). In accordance with the *modus ponens* rule of inference, the research questions can be expressed deductively as follows:

$$\frac{P \rightarrow Q, P}{\therefore Q}, \tag{18}$$

where $P \rightarrow Q$ is the conditional premise that P implies Q, P is the antecedent, and Q is the consequent of the conditional premise.

In less formal terms, if $P \rightarrow Q$ is true and if P is true, then it can be inferred that Q must also be true. When Equation (18) is applied to the research questions of this study, the following can be inferred using *modus ponens*, provided that the conditional premise $P \rightarrow Q$ and the antecedent P are true.

> **Research Question 1:** *What is the average value of the information contained in the results of EU-wide stress tests measured in terms of abnormal stock returns?*
>
> If new relevant public information causes stock prices to adjust $(P \rightarrow Q)$, and if EU-wide stress test results represent new relevant public information (P), then the stock prices of the banks concerned will adjust to this information (Q), the value of which can be measured in abnormal returns.

Research Question 2: *What is the functional relationship between new information from EU-wide stress test results and corresponding abnormal stock returns?*

If risk is an important determinant of stock prices $(P \rightarrow Q)$, and if EU-wide stress test results provide new relevant public information about banks' risks (P), then the stocks of affected banks will adjust to their new risk-return equilibrium (Q).

Research Question 3. *How has the informational value of EU-wide stress test results, measured in abnormal stock returns, changed over time?*

If EU-wide stress test results have an informational value, the stocks of the banks concerned show abnormal returns in every exercise $(P \rightarrow Q)$, and if the magnitude of the abnormal returns does not remain constant on average (P), then the informational value will change over time (Q).

The above *modus ponens* rule of inference was applied to answer the research questions based on the empirical results obtained in this study (Section 8.3).

5.2.5 Alternative Research Paradigms

Although this study is based on functionalism, the dominant paradigm in finance research (Ardalan 2003, Gendron and Smith-Lacroix 2015, Rao 2019), and an objectivist ontology and empirical-positivist epistemology, it would have been possible to resort to alternative research paradigms. The purpose of this section is to briefly discuss these alternatives.

The framework of paradigms in the social sciences (Figure 8) developed by Burrell and Morgan (1979) suggests three distinct paradigms besides functionalism: interpretivism, radical humanism, and radical structuralism. These paradigms have recently been complemented by radical emergence and radical verificationism (Callaghan 2017). However, all of these paradigms – with the exception of interpretivism – are on the "radical change" side of the regulation-radical change dimension of the framework. This dimension indicates the researcher's assumptions about the nature of society, *i.e.* whether one believes that society seeks to maintain the *status quo* or seeks radical change (Burrell and Morgan 1979). It is not part of the belief system of this study to believe that society seeks radical change, but rather to maintain the *status quo*. Therefore, most of the aforementioned alternative research paradigms do not appear to be well suited for this study. However, it must be acknowledged that interpretivism (which, like functionalism, is on the "regulation" side of the regulation-radical change dimension of the paradigm framework) is to be considered as a possible alternative research paradigm.

Taking an interpretivist approach would have meant shifting the ontological perspective of the study from an objectivist to a subjectivist view along the subjective-objective dimension of the paradigm framework (see Figure 8). As a result, other, qualitative, research methods (such as surveys or interviews) should have been used to account for this change in perspective. A potential benefit of an interpretivist approach could have been higher content validity of the findings compared to a positivist approach, as it would have been able to uncover the meaning and motivation of investor behaviour. However, since reality and the subjective life experiences of its observer are inextricably linked in interpretivist research, the results are difficult to reproduce and may be less reliable than those of positivist research. This reflects the well-known tradeoff between validity and reliability: the stronger the basis for validity, the weaker the basis for reliability (and *vice versa*). As a result, interpretivist research is sometimes considered unscientific because its findings are hardly falsifiable due to its subjective influences and potential biases (Collis and Hussey 2003). Given the inhomogeneous groups of investors in the stock market, an interpretivist approach would also have raised the practical question of which investor group(s) the study should have targeted (*e.g.* private or institutional investors).

After thoroughly considering the advantages and disadvantages of the different research paradigms as well as the conventions in finance research, it was decided to adopt a functionalist paradigm with objectivist ontological and empirical-positivist epistemological perspective to conduct this study.

5.2.6 Summary

In this section, the research philosophy of the study was described. It is based on the functionalist research paradigm, which is the dominant paradigm in finance research (Ardalan 2003, Gendron and Smith-Lacroix 2015, Rao 2019) and is particularly well suited to answering financial research questions. This is because functionalism attempts to explain the socio-economic *status quo* with an objectivist ontological perspective and an epistemological stance rooted in positivism (Figure 8). The choice of a research paradigm guides and informs subsequent philosophical decisions.

Accordingly, this study takes an objectivist ontological view. More precisely, the study was carried out from the position of a detached observer who stands outside the research situation and does not interact with the research subjects. This was consistent with the study's quasi-natural experimental strategy (Section 5.3.1), which is based on exogenous events that do not involve interaction or manipulation by the researcher and follows the ideal of the scientific method. The objectivist ontological view of this study is expressed through the use of quantitative methods, rules, and control measures (Porter 1995) as well as the intention to conduct the study in a value-free and unbiased manner to facilitate reproducibility and testability.

Similarly, the empirical-positivist epistemology of this study follows from its functionalist paradigm. In accordance with the Justified True Belief account of knowledge, this study accepts empirical, *a posteriori* evidence as true knowledge only if it is justifiably believed to be true (Audi 2011, Shope 2017). This corresponds to the ideal of the scientific method and is demonstrated in this study by its highly structured research process and extensive use of controls and robustness checks. Furthermore, the data used in this study are collected passively through empirical observations of stock prices and EU-wide stress test results, *i.e.* through structured direct observations of social reality. All of this is characteristic of an empirical-positivist epistemology (Gill and Johnson 2010, Saunders *et al.* 2009a) and corresponds to the objectivist ontology of the study.

Consequently, deductive reasoning is used to answer the research questions of this study. This becomes evident in this study through the development of a dedicated theoretical framework, the formulation of empirically testable hypotheses, and the application of the *modus ponens* rule of inference.

To conclude, the research philosophy of this study is internally consistent as all of its components are compatible and mutually supportive. The combination of objectivist ontology and empirical-positivist epistemology follows from the adopted functionalist paradigm and is often used together with deductive logic (O'Gorman and MacIntosh 2015, Saunders *et al.* 2009a). The above philosophical choices are also consistent with the research questions and purpose of the study. They thus formed a coherent basis for the research design of this study, which is presented in the following section. The adopted paradigm and approach to research has been carefully weighed against other alternative research paradigms, taking into account advantages and disadvantages as well as conventions in finance research.

5.3 Research Design

This section describes the research design of this study. The research design includes the research strategy and the methods used to collect and analyse the data; it also covers the time horizons used in the study (see Figure 7). This study was based on a quasi-natural experimental strategy (Section 5.3.1), which was implemented through an extended event study approach and subsequent research-question specific analyses (Section 5.3.2). Research Questions 1 and 2 were based on cross-sectional samples, while Research Question 3 used a longitudinal sample. The methods used to collect and analyse the data are outlined in Section 5.3.2 and explained in detail in the subsequent sections on the event study and the research-question specific analyses.

5.3.1 Research Strategy

A research strategy is a general plan of action that enables a study to be carried out systematically in order to answer its research questions (Saunders *et al.* 2009b). The research strategy used in this study is quasi-natural experimentation and is based on an extended one-group pretest-posttest design.[79] This strategy is specified and detailed in Section 5.3.1.2. However, in order to provide the necessary background, quasi-natural experiments are first introduced in Section 5.3.1.1 and contrasted with "true" (randomised controlled) experiments. Finally, the controls used to minimise the potential

[79] For a discussion of quasi-experimental designs, see Shadish *et al.* (2002).

impact of confounding and extraneous factors are presented in Section 5.3.1.3 (confounding control). This also includes the extensions to the traditional one-group pretest-posttest design of this study.

5.3.1.1 Quasi-Natural Experiments

Quasi-natural experiments have a long history (see, for example, Lind 1753) and are a subtype of "true" (randomised controlled) experiments, often referred to as the gold standard against which alternative strategies must be assessed (O'Gorman and MacIntosh 2015, Saunders *et al.* 2009b, Shadish *et al.* 2002).

In "true" experiments, which are often conducted under laboratory conditions by the natural sciences, research subjects are randomly assigned to an intervention controlled by the researcher; the aim is to manipulate one or more independent variables in order to observe a causal effect on the dependent variable (Collis and Hussey 2003, Saunders *et al.* 2009b, Shadish *et al.* 2002). In contrast, quasi-natural experiments deviate from this ideal in two respects: first, the intervention is not randomly assigned, and second, the intervention is controlled by a force other than the researcher (Meyer 1995, Shadish *et al.* 2002).

When used separately, the two deviations form independent research strategies known as *quasi*-experiments (first deviation) and *natural* experiments (second deviation); they can be used in research situations in which random assignment or intervention control by the researcher is not possible, respectively.[80] Accordingly, combined *quasi-natural* experiments can be used to address specific research situations that neither allow for random assignment nor intervention control by the researcher, *e.g.* the analysis of abnormal stock returns in response to EU-wide stress test results.

In the next section, the quasi-natural experimental strategy is specified to the context of this study. This also includes the justification for the choice of this particular research strategy.

---

[80] It should be noted, however, that the literature has been using the terminology for these research strategies in inconsistent and sometimes contradicting ways. Therefore, when identifying a research strategy, the main focus must be on its design and application, and not on its name. For an overview of how the terms *natural experiment* and *quasi-natural experiment* have been used in the literature, see DiNardo (2008).

5.3.1.2 Strategy Specification

The context of EU-wide stress tests intuitively lends itself to quasi-natural experiments: CEBS and EBA have imposed EU-wide stress tests (interventions) on a number of banks using a size-based selection rule (non-random assignment) and disclosed bank-level results (independent variable) to the public, including bank stock investors, whose aggregate response to the new information can be hypothesised to cause abnormal stock returns at the affected banks (dependent variable). Quasi-natural experiments have also recently been used in similar financial regulation contexts, see, for example, Gropp *et al.* (2019), Hu *et al.* (2019), and Wang and Chou (2018).

The *natural* and *quasi*-experimental design elements of this research strategy and their implementation in this study are elaborated below. The focus is on describing how the performed quasi-natural experiments differ from "true" experiments.

Natural-Experimental Elements

The research strategy used in this study is *natural* in that the studied interventions (*i.e.* the EU-wide stress tests) are controlled by CEBS or EBA and are therefore beyond the control of the researcher. Meyer (1995, p. 151) explained that "[g]ood natural experiments are studies in which there is a transparent exogenous source of variation in the explanatory variables". In the social sciences, changes in laws, regulations, and policies are among the most frequently cited sources of such variations (Dunning 2012, Meyer 1995, Shadish *et al.* 2002). Supervisory stress tests arguably fall into the same category of sovereign public intervention (Ellahie 2012, Atanasov and Black 2016)

Due to the strict separation between research and control of the intervention, Shadish *et al.* (2002) argued that the intervention in natural experiments is often not even potentially manipulable by the researcher. The use of natural experiments is therefore consistent with the objectivist ontology and empirical-positivist epistemology of this study, which emphasises the importance of value-free and unbiased study conduct. This consistency between research philosophy and research strategy is also supported by the fact that all experimental research strategies are positivist in nature (Collis and Hussey 2003).

Dunning (2012) pointed out that the use of natural experiments is best when a well-defined population is exposed to a particular intervention. This is the case with EU-wide stress tests, as it is precisely defined and publicly disclosed which banks are subject to an exercise. The selection of banks into EU-wide stress tests by CEBS and EBA follows a size-based selection rule, which is described below.

Quasi-Experimental Elements

The research strategy of this study is also *quasi*-experimental, since the interventions were not randomly assigned to the research subjects.[81] This means that the banks that were subjected to EU-wide stress tests were not selected at random. Instead, CEBS and EBA applied a size-based selection rule, according to which banks are selected into EU-wide stress tests based on their total consolidated assets.[82] Such authority-controlled selection is known as *administrator selection* (Shadish *et al.* 2002).

A negative consequence of non-random assignment is that research subjects may be exposed to factors other than the intended intervention in many systematic (non-random) ways (Dunning 2012, Shadish *et al.* 2002). Any such extraneous or confounding factors could be a possible alternative explanation for the observed intervention effect. In "true" (randomised controlled) experiments, the potential impact of such factors is reduced by the offsetting effect of randomisation; therefore, confounding control is inherently built into the design of "true" experiments (Dunning 2012, Shadish *et al.* 2002). However, quasi-experiments, by definition, lack random assignment and must therefore rely on other methods to control for extraneous and confounding factors. Shadish *et al.* (2002) suggested the following means: design, measurement, and logic. In this study, all of these means were used to implement effective controls for extraneous and confounding factors (confounding control). These controls are described in the next section.

---

[81] The term *quasi-experiment* was coined by Campbell and Stanley (1963) in their influencial textbook on experimental and quasi-experimental research designs. For a more contemporary discussion of the matter, see Shadish *et al.* (2002).

[82] More precisely, banks are selected into EU-wide stress tests according to the following rule: for every relevant jurisdiction, banks are ranked in descending order based on their total consolidated assets at the end of the financial year preceding an EU-wide stress test. Banks are then selected top-down until a certain threshold level of total banking sector assets has been reached for each relevant jurisdiction. The threshold level depends on the respective exercise; for the EU-wide stress tests carried out in 2010, 2011, and 2014, the threshold level was 50% (CEBS 2010a, EBA 2011c, EBA 2014a), while for the EU-wide stress tests conducted in 2016 and 2018, the threshold level was 70% (EBA 2016b, EBA 2018b). Since the threshold level for the relevant jurisdictions is usually not exactly met, but exceeded, this approach means that the coverage of the total EU banking sector assets can sometimes be significantly above the threshold set at the jurisdiction level (see Table 1).

5.3.1.3 Confounding Control

The result of a quasi-natural experiment can be causally linked to the intervention to the extent that alternative explanations are implausible (Shadish *et al.* 2002). Therefore, this study used multiple methods to control for extraneous and confounding factors. They are based on Shadish *et al.* (2002) and aim to make the observed intervention effects robust to alternative explanations. Table 6 provides an overview of the controls used in this study.

Table 6
Controls for Extraneous and Confounding Factors

Design Control	Measurement Control	Logic Control
Multiple Pretests (intead of a single pretest) and a systematic model selection procedure as part of an extended one-group pretest-posttest design	Inclusion of asset pricing models in the set of candidate models that explicitly take into account the factors supported by the empirical evidence (firm size)	Exclusion of banks from the research population that were exposed to known extraneous factors (*ad hoc* disclosures, director dealings, and ex-dividend days) during the event window

Note. This table provides an overview of the extraneous and confounding factor controls used in this study. The potential influence of extraneous and confounding factors that may have remained after the above controls were mitigated through the use of average values (average cumulative abnormal returns $\overline{CARs}$ and average absolute cumulative abnormal returns $|\overline{CARs}|$) in the analyses of Research Questions 1 and 3. In the analyses for Research Question 2, no averaging was carried out.

While the above measurement and logic controls aim to minimise known extraneous and confounding factors, the design control aims to capture unknown factors through the use of multiple pre-tests in conjunction with a systematic model selection procedure. Each of these controls is described in more detail below.

Design Control

The design of the quasi-natural experiments used in this study extends the traditional one-group pretest-posttest design to include multiple pretests and a systematic model selection procedure. This means that, in contrast to the traditional design, multiple candidate pretests are performed and subjected to systematic model selection. The purpose of this extension is to identify and select the pretest model that minimises the estimation error (information loss) inherent in any estimation procedure, *i.e.* the difference between the estimated parameter value and the true parameter value. The extended design thereby helps to eliminate the systematic component of the estimation error introduced by unknown extraneous and confounding factors. In addition, using multiple pretests together with a systematic approach to model selection also helps to counter experimenter bias. The use of multiple prestests and other design extensions has been advocated by Meyer (1995) and Shadish *et al.* (2002), among others.

The research strategy of this study is implemented through event studies (Section 5.3.2.1). This is consistent as both pretest-posttest designs and event studies use counterfactual inference to study the effect of an intervention (event). That is, they compare the actual observed state of the research subjects with their normal state, which would have been expected if the intervention (event) had not occurred (Campbell *et al.* 1997, Shadish *et al.* 2002, see also Section 3.5.2). In event studies, pretesting is represented by the estimation of the normal return, while posttesting is represented by the observation of the actual return. As shown in Section 3.5.2, the normal return estimate is the key parameter for the abnormal return calculation (Equation (3)). Therefore, careful pretesting is crucial for the internal validity of any event study. Posttesting, on the other hand, is generally less problematic in event studies, since direct observation of the actual return is less prone to error. Figure 9 shows a schematic representation of the extended pretest-posttest design used in this study.[83]

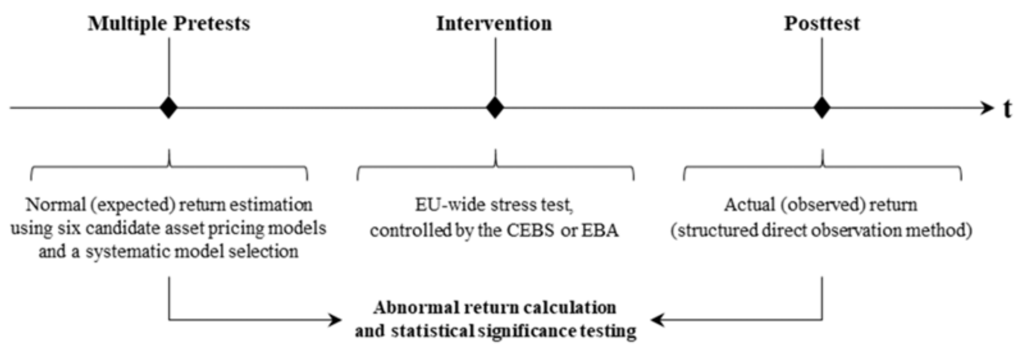


Figure 9. Pretest-posttest design with multiple pretests in conjunction with a systematic model selection procedure

The multiple-pretest design is implemented through the use of six candidate asset pricing models and a systematic model selection procedure. After running the candidate models, the model selection is performed based on their goodness-of-fit (Section 5.3.2.1.3). The aim is to select the most accurate asset pricing model (in terms

[83] A more detailed visualisation of the methodological implementation of the research strategy is shown in Figure 12, including the estimation period (which is used to estimate the parameters for the candidate asset pricing models (normal return estimates)) and the event windows (over which the actual returns are observed.

of estimation error) in order to use its normal return estimates for the subsequent abnormal return calculation. This approach reduces the risk that the use of an inappropriate asset pricing model introduces extraneous or confounding factors into the normal return estimates and consequently into the abnormal returns. The multiple-pretest design of this study therefore explicitly addresses the joint-hypothesis problem (also known as bad-model problem), which is a critical and well-known problem in testing market efficiency (see detailed discussion in Section 3.5.3).

Measurement Control

Another control for extraneous and confounding factors is measurement. In this study, measurement primarily means the specification of the asset pricing models considered for model selection. Measurement control is therefore closely related to design control and aims to ensure that the set of candidate models includes asset pricing models that fit the specific research context of this study. In other words, the goal of the measurement control is to specifically include asset pricing models that optimise the signal-to-noise ratio of the pretests, *i.e.* the ratio between meaningful (signal) and meaningless (noise) output. Without proper specification, the asset pricing models are likely to produce large estimation errors or, more specifically, spurious normal return estimates contaminated by extraneous or confounding factors.

In the context of this study, firm size is an obvious confounding factor due to the size-based selection rule applied by the CEBS and EBA. On the other hand, firm size is also a well-known factor in asset pricing (Banz 1981; Barber and Lyon 1997a; Brown *et al.* 1983; Fama and French 1992, 2012; Keim 1983; Reinganum 1981, 1983).[84] In their seminal textbook on model selection and multi-model inference, Burnham and Anderson (2002) emphasised that any set of candidate models must include models that take into account the factors supported by empirical evidence. This general consideration is consistent with the more specific statement by Kothari and Warner (2007, pp. 12-13), that "event study tests are well-specified only to the extent that the assumptions underlying their estimation are correct." Therefore, the set of can-

[84] For reviews of the theoretical and empirical literature on the size effect on stock returns, see Schwert (1983) and Van Dijk (2011).

didate models used in this study included two asset pricing models that explicitly consider firm size as one of the asset pricing factors, *i.e.* the Fama and French (1993) Three-Factor Model and the Fama and French (2015) Five-Factor Model.[85]

The size-based selection of banks into EU-wide stress tests by CEBS and EBA is also the main reason for the one-group element in the one-group pretest-posttest design of this study. Using a one-group design, or in-sample comparison, means analysing the intervention effect of EU-wide stress tests on the participating banks (intervention group) without a control group. This is appropriate in this context since using any control group would inevitably confound the intervention effect with pre-existing differences between the intervention group and the control group. This is for two reasons. First, any control group selected from European banks would necessarily consist of banks smaller than the size-selected banks in the intervention group. This is because the size-based selection rule chooses the largest European banks into EU-wide stress tests (intervention group) so that the remaining European banks that could be used to form a control group are therefore necessarily smaller in size. Consequently, any causal inference drawn from comparisons between the intervention group and such a control group would be confounded by differences in firm size. Second, any alternative control group selected from banks outside Europe would almost certainly introduce a variety of other confounding factors, such as differences in banking regulations (Bruno *et al.* 2018, Francis *et al.* (2015), Hoque *et al.* 2015), monetary policy regimes (Chen and Chan 1989, Dinenis and Staikouras 1998, Flannery and James 1984), and business cycles (Choudhry *et al.* 2016, Corradi *et al.* 2013, Hamilton and Lin 1996).

Logic Control

As a final control, those banks found to be exposed to a known extraneous factor during an event window were excluded from the analysis (Section 5.3.2.1.2). The logic control was based on the following known extraneous factors that could not be addressed by the design or measurement control: *ad hoc* disclosures, director dealings, and ex-dividend days. These factors are explained in more detail below.

85 For a complete list of the asset pricing models used in this study, see Section 5.3.2.1.3.

Ad hoc disclosure requirements refer to the legal or regulatory obligation of securities issuers to immediately report and publish information that could affect the market price of their securities.[86] Therefore, any *ad hoc* disclosure is by definition an extraneous factor. This view is generally supported by empirical evidence showing that timely disclosure of *ad hoc* information is associated with significant abnormal returns (Baule and Tallau 2016, Lerman and Livnat 2010, McMullin *et al.* 2019). A more detailed intraday analysis by Muntermann and Guettler (2007) suggests that the stock price adjustment process after an *ad hoc* disclosure only takes ten price fixings or 30 minutes. Similarly, Muntermann (2005) showed that the stock price adjustment is completed within the first five price fixings or an average of 23.2 minutes. Bank and Baumann (2016, p. 640) even report that "the average time until the market accounts for most of the information is around 3-5 min."

Director dealings, or insider trades, are widely regarded as signals to the market (Ajlouni and Toms 2008, Del Brio and De Miguel 2010, Hillier and Marshall 2002). This is because corporate insiders are assumed to be better informed about their firm's prospects and affairs than outside investors (Lakonishok and Lee 2001). Empirical evidence indeed suggests that director dealings tend to yield significant positive abnormal returns (Fidrmuc *et al.* 2006, Seyhun 1986, Pettit and Venkatesh 1995). In order to prevent market abuse, director dealings are typically subject to prompt reporting and disclosure requirements. In the EU, the Market Abuse Regulation (EU) 596/2014 requires immediate notification within three business days of the transaction.[87] As a result, the public disclosure of director dealings is perceived as conveying superior information to the market that is quickly incorporated into security prices by outside investors imitating the actions of corporate insiders (Hillier and Marshall 2002, Lakonishok and Lee 2001, Lorie and Niederhoffer 1968).[88] Director dealings therefore

---

[86] The purpose of *ad hoc* disclosures is to mitigate the abuse of insider information and to increase market transparency.

[87] Lenkey (2014) even modeled the potential effects of advance disclosure of insider trading. He found that disclosing insider trading prior to the deal increases the wellfare of both corporate insiders and outside investors. Somewhat surprisingly, however, he also found that advance disclosure of insider trading would lead to lower market efficiency. This can be explained by the relative strength of the two observed signals: in his model, outside investors react to the signal contained in the advance disclosure of insider trading and in the stock price. However, since the advance disclosure signal turned out to be noisier than the stock price signal, an advance disclosure regime would have a negative impact on market efficiency.

[88] For a study examining the imitation of director dealings as a stand-alone investment strategy, see Moodley *et al.* (2016).

have a significant impact on stock prices and must be considered as an extraneous factor in this study.

Dividend payments, on the other hand, can intervene more technically in the stock price formation process by lowering the price of a stock on its ex-dividend day by approximately the amount of the dividend per stock (*ceteris paribus*). This relationship is well established in the literature (*e.g.* Barclay 1987, Barker 1959, Boyd and Jagannathan 1994, Campbell and Beranek 1955, Kalay 1982) and suggests that stock investors, on average, view dividend payments and capital gains as perfect substitutes (excluding tax effects such as differential taxation of dividends and capital gains).[89] This indifference is theoretically based on Miller and Modigliani's (1961) dividend irrelevance theorem; whereas the characteristic price drop on the ex-dividend day is based on the fact that investors who buy a stock on or after that day are not entitled to the upcoming dividend payment. Consequently, ex-dividend days were considered as an extraneous factor in this study and controlled accordingly.

5.3.2 Research Methods

This study used a two-step research process to address the research questions. First, an extended event study was carried out, on the basis of which research-question specific analyses were performed. The necessary capital ratio and stock price data for the sample banks were collected by means of structured direct observation. Various statistical methods were used to analyse the data, including *t*-tests and Wilcoxon signed-rank tests (Research Question 1), curve fitting, OLS regressions, and error metrics (Research Question 2), and repeated measures ANOVAs, Friedman tests, and the corresponding *post hoc* tests (Research Question 3). In the following sections, the methods used are presented in detail.

[89] For studies that have decomposed the total trading activity around ex-dividend days in order to identify the (opposing) trading strategies of different investor groups, see Felixson and Liljeblom (2008) and Koski and Scruggs (1998). For tax effects on the ex-dividend day behaviour of stock prices, see Eades *et al.* (1984), Elton and Gruber (1970), and Kalay (1982).

5.3.2.1 Event Study

The basic idea of an event study is to determine the impact of a common information event on the prices of securities, as measured by abnormal returns. The abnormal (unexpected) return of a given security i is defined as the difference between its actual (observed) return, conditional on a specific event, and its normal (expected) return that would have been expected in the absence of the event (Campbell *et al.* 1997, Kothari and Warner 2007, MacKinlay 1997, see also Section 3.5.2). This relationship was previously defined in Equation (3) and is repeated here for convenience and better understanding of the following equations:

$$AR_{i\tau} = R_{i\tau} - E(R_{i\tau}|X_\tau), \tag{19}$$

where $AR_{i\tau}$ is the abnormal (unexpected) return, $R_{i\tau}$ is the actual (observed) return, $E(R_{i\tau})$ is the normal (expected) return, and X_τ is the conditioning information for the normal return-generating model used to estimate the normal (expected) return.

When calculating abnormal returns, both the actual return and the normal return are cumulated over a specific event window τ (*time-series aggregation*), which in this study can span one to five trading days (Section 5.3.2.1.1). Accordingly, the cumulative abnormal return (CAR) of security i over event window τ (from time t_1 to time t_2) is defined as

$$CAR_{i\tau} = \sum_{t=t_1}^{t_2} AR_{it}. \tag{20}$$

In addition, the $CARs$ of individual securities can also be aggregated across the sample to form average cumulative abnormal returns ($\overline{CARs}$). The goal is to examine whether the event under investigation is, on average, associated with a change in the price of securities (*cross-sectional aggregation*). For a sample of N securities, the $\overline{CAR}$ over any event window τ is

$$\overline{CAR}_\tau = \frac{1}{N}\sum_{i=1}^{N} CAR_{i\tau}. \tag{21}$$

While cross-sectional aggregation is only used in the analyses for Research Questions 1 and 3 (Section 1.4), time-series aggregation is used throughout this study.

The common methodological basis for all three research questions largely follows the structure developed by Campbell *et al.* (1997) and MacKinlay (1997), which has become the standard approach to event studies. This approach has been used, for example, by Ahnert *et al.* (2020), Morgan *et al.* (2014), and Petrella and Resti (2013) to examine market reactions to supervisory stress test events in the US and the EU.[90] The event study structure of Campbell *et al.* (1997) and MacKinlay (1997) is an integrated research methodology that covers all relevant stages and includes, among other things, sampling, data collection, and data analysis.

Building on that, this study extends the existing standard approach in four important ways. First, by a new method for statistically determining the length of event windows (Section 5.3.2.1.1). Second, by implementing extensive confounding controls in the sampling and model selection stages (Sections 5.3.2.1.2 and 5.3.2.1.3). Third, by introducing a systematic model selection procedure (Section 5.3.2.1.3). Fourth, by additionally examining *absolute* abnormal returns as a non-directional measure (Section 5.3.2.1.5). Figure 10 illustrates and summarises the methodological structure of this study.

[90] Other studies that have used the standard approach to event studies include Candelon and Sy (2015), Cardinali and Nordmark (2011), and Georgescu *et al.* (2017).

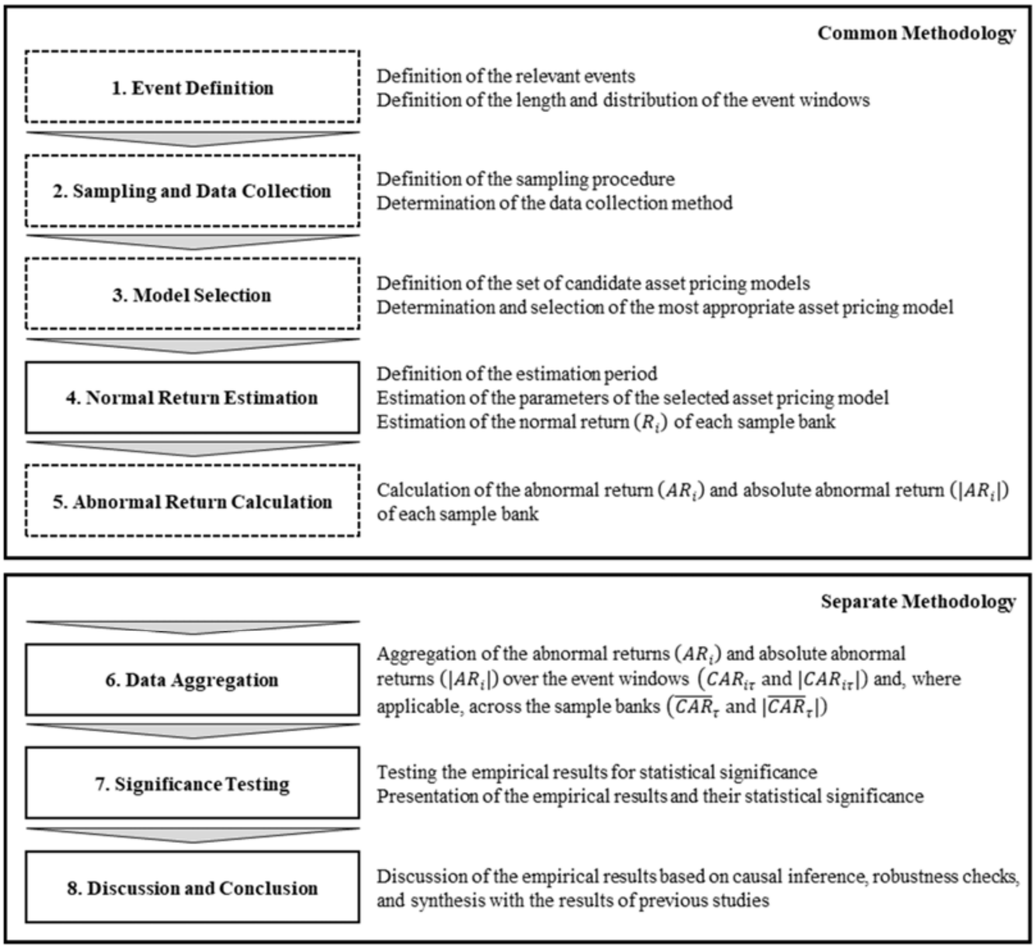


Figure 10. Methodological structure. Dotted lines indicate stages with extensions of the standard event study approach developed by Campbell *et al.* (1997) and MacKinlay (1997)

The first five stages shown in Figure 10 are the same for all three research questions and are presented in the following Sections 5.3.2.1.1 to 5.3.2.1.5. These stages therefore form the common methodological basis of this study. Any methodological specifics (stages six to eight) are described separately for each research question in Sections 5.3.2.2 to 5.3.2.4, along with further specific analysis methods.

5.3.2.1.1 *Event Definition*

This event study began by defining the relevant events and the different event windows over which the abnormal returns were calculated.

Definition of the Relevant Events

The events relevant for this study are the result disclosures of the five EU-wide stress tests that were carried out from 2010 to 2018 (Section 1.5). The results were disclosed centrally on the CEBS or EBA website. Table 7 provides an overview of the relevant results disclosure events.

Table 7
Relevant Results Disclosure Events

	Results Disclosure Events		
EU-Wide Stress Test	**Event Date**	**Day of the Week**	**Time[a]**
CEBS 2010[b]	23 July 2010	Friday	18.30 CEST
EBA 2011[c]	15 July 2011	Friday	18.00 CEST
EBA 2014[d]	26 October 2014	Sunday	12.00 CET
EBA 2016[e]	29 July 2016	Friday	22.00 CEST
EBA 2018[f]	2 November 2018	Friday	18.00 CET

Note. This table provides an overview of the results disclosure events of the five EU-wide stress tests examined in this study. Data are from CEBS (2010b)[b], EBA (2011d)[c], EBA (2014b)[d], EBA (2016c)[e], and EBA (2018c)[f]. CEBS = Committee of European Banking Supervisors. EBA = European Banking Authority. CEST = Central European Summer Time. CET = Central European Time.

[a] The times of the result disclosures were standardised to CEST and CET because the CEBS and EBA have used inconsistent time zones to indicate the times at which the results were disclosed.

As shown in Table 7, the results of EU-wide stress test are regularly disclosed at the end of a trading week and after the markets have closed in Europe.[91] This observation has a significant impact on the definition of event windows, since abnormal stock returns can only be causally determined from the first trading day *after* the EU-wide stress test results have been disclosed. This fact has been neglected in previous studies, except by Petrella and Resti (2013), who deemed the trading day after the results disclosure to be the event date.

[91] On European stock exchanges, the markets usually close at 17.30 CET/CEST, with a few uncritical exceptions, *e.g.* Amsterdam (17.40 CET/CEST), Copenhagen (17.00 CET/CEST), and Warsaw (16.50 CET/CEST).

Using a similar approach, this study keeps the actual event date for consistency, but calculates daily stock returns from closing to closing prices rather than from opening to closing prices (Equation (23)). In this way, the study also captures the overnight return (in addition to the intraday return) that occurs over night between any two trading days, including the night between the event date and the next trading day.[92] This is crucial in the context of this study, as the efficient market hypothesis suggests that stock prices adjust immediately to new information (see discussions in Sections 3.5.1 and 3.5.4.3). The stock returns used in this study therefore represent *total* daily returns, which include all return components and thus reflect the entire stock price adjustment process.

Definition of the Event Windows

In addition to the event dates, the different event windows must be defined. An event window is the period of time over which the sample banks' stock prices are examined and abnormal returns are calculated. Carefully defining the length of an event window is critical to ensure that the abnormal return reflects the event-related signal as closely as possible, rather than unrelated noise. Similarly, an informed decision must be made about the distribution of the event window around the event date.

In this study, abnormal returns are cumulated over three different types of event windows: standard, pre-event, and post-event windows. While it is common in event studies to use multiple event windows, the rationale for using these three particular event windows lies is in their specific distributions around the event date and their resulting functions. These are explained in more detail below. The length and distribution of the event windows are specified in terms of trading days, using the event date (t_0) as the reference time. Figure 11 visualises the design of the three event windows used in this study.

[92] For a study on the characteristics of overnight returns, see Riedel and Wagner (2015). They found that overnight returns have significant tail risks, which can be attributed to a lack of market functionality and liquidity during non-trading hours, and which can manifest themselves in large price movements between the closing and opening prices.

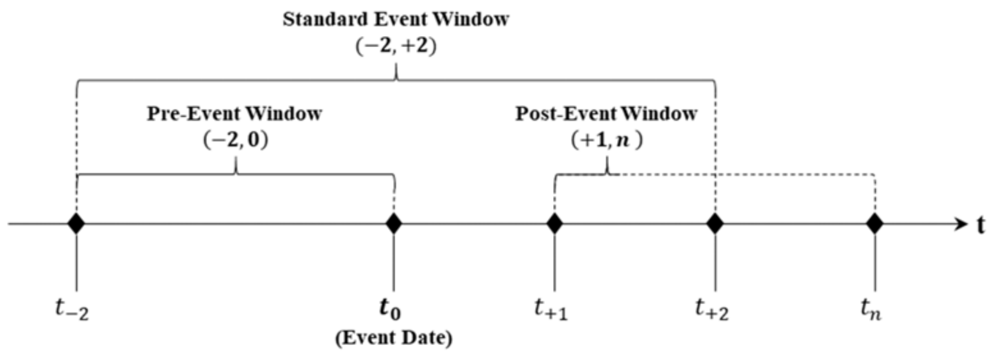


Figure 11. Event window design

The standard event window is a five-day event window that is evenly distributed around the event date (-2, +2).[93] It is the best possible proxy for a typical or conventional event window and is used to facilitate comparison and synthesis with previous studies. The standard event window was estimated from the central tendency (in terms of median) of the event windows used in previous studies of supervisory stress tests in the EU and the US, both separately and jointly (for detailed results, see Appendix A). Previous studies have used a wide range of different event windows.[94] This heterogeneity in the length and distribution of event windows can be explained by the specific research questions at hand. However, it is probably also due to the fact that the methodological literature (*e.g.* Campbell *et al.* 1997, MacKinlay 1997, Kothari and Warner 2007) does not provide clear guidance on the definition of event windows. Instead, it is left to the discretion of the researcher to define the length and distribution of an event window. Researchers therefore often resort to heuristics using fixed-length event windows with arbitrary distributions around the event date (Krivin *et al.* 1997, Lev 1989). The fact that the standard event window is distributed around the event date means that it captures abnormal returns for two different reasons: first, the actual disclosure of EU-wide stress test results on the event date (post-event part) and second,

93 This is consistent with the review by Thompson (1995), which suggests the use of event windows with one to five days in length. However, the review does not provide guidance on the distribution of event windows around the event date.

94 This is true for previous studies on market reactions to supervisory stress tests events in the EU and the US (Appendix A), but also for event studies in general. Lev (1989) surveyed the event studies published in three major accounting journals (*Accounting Review*, *Journal of Accounting and Economics*, and *Journal of Accounting Research*) from 1980 to 1988 and reported that event window lengths ranged from two days to one year. According to Kothari and Warner (2007), about half of these event studies can be classified as short-term (< 1 year) or long-term (≥ 1 year). For summaries of long-term event studies, see Barber and Lyon (1997b; Table 3) and Kothari and Warner (1997; Table 10).

the speculative pre-disclosure positioning of investors and potential information leaks (post-event part).

The pre-event window is a three-day event window that includes the two trading days prior to the event date and the event date itself (-2, 0). Its function is to isolate abnormal returns that may occur prior to disclosure of EU-wide stress test results in order to quantify the impact of speculative pre-disclosure positioning by investors and potential information leakage. The definition of the pre-event window followed the same procedure as the standard event window but was limited to the trading days prior to the disclosure of the stress test results including the event date (Appendix A). Therefore, like the standard event window, the pre-event window is a simple fixed-length event window.

The post-event window is an event window of variable length with a fixed start date and a variable end date, which was individually determined for all bank-year observations using an innovative statistical approach.[95] The post-event window is therefore formally defined as $(+1, n)$, with the start date (t_{+1}) being the first trading day after the event date and the end date being either the same trading day or a later trading day (t_n).[96] In the former case, the length of the event window is one trading day; in the latter case, more than one trading day. The individual end dates were determined by recontextualising the Ljung-Box (1978) test, which is normally used to evaluate the independence of model residuals. However, the Ljung-Box (1978) test can also be used to determine the lag k up to which a time series *does* exhibit serial correlation. The logic behind this approach is simple: based on the initial stock price reaction to the disclosure of EU-wide stress test results on t_{+1}, the stock price adjustment process continues until the first trading day on which the serial correlation is no longer significant. In other words, a post-event window extends until the trading day when the stock

---

95 This new approach differs significantly from the few other existing methods that have been used in the literature to determine the appropriate length of variable event windows, see, for example, De Franco *et al.* (2007) and Lins *et al.* (2013), who have used trigger events based on quarterly and monthly intervals, respectively. The technique that comes closest to the new approach presented here is that of Krivin *et al.* (1997), who determined the length of variable event windows based on the number of trading days that showed significant abnormal returns. However, this technique is prone to volatility as it does not distinguish between positive and negative abnormal returns. This means that both the original stock price reaction to the event and any potential price reversal are included in the determination of the event window length (provided they represent abnormal returns). As a result, the event window lengths determined by this technique tend to be longer than appropriate. This is not the case with the new approach proposed in this study.

96 It is worth reiterating that abnormal returns that occur during the night between the event date (t_0) and the first trading day after the event date (t_{+1}) are captured by the post-event window, as stock returns are calculated from closing to closing prices (see discussion above).

price has fully incorporated the new information and has thus reached efficiency. The function of the post-event window is therefore to optimise the extent to which the intervention effect is captured in the abnormal returns by maximising the event-related signal and minimising the unrelated noise. Formally, the Ljung-Box (1978) test is defined as

$$Q = T(T+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{T-k}, \tag{22}$$

where T is the length of the time series, $\hat{\rho}_k$ is the serial correlation coefficient at lag k, and h is the number of lags being tested.

The results show that the lengths of the individual post-event windows range from one day (83% of all bank-year observations) to a maximum of four days. These results are robust to variations in the length of the time series and the number of lags being tested (for detailed results and descriptive statistics, see Appendix B). This finding is consistent with the semi-strong form of the efficient market hypothesis, which suggests that stock prices adjust quickly to new information. It also agrees with the empirical observation of previous studies, which found that shorter event windows are more informative than longer event windows (Georgescu *et al.* 2017, Morgan *et al.* 2014).

According to Krivin *et al.* (1997), variable event windows are more appropriate than fixed event windows in studies with relatively small sample sizes (as in this study, see Section 5.3.2.1.2). This is because of the following consideration: the smaller the sample, the lower the probability that a noise-induced price change in one stock will be offset by a noise-induced price change in another stock in the opposite direction, since the law of large numbers is not (fully) effective. However, small sample sizes make it possible to determine the length of each window of events individually. This is a major advantage that allows using actual price data and statistical methods (instead of heuristics) to determine the appropriate length of event windows and thus optimize the overall signal-to-noise ratio of a study.

5.3.2.1.2 Sampling and Data Collection

The next stage in this event study was to define the research samples. This includes a description of the sampling procedure and the method used for subsequent data collection.

Sampling

This study uses five cross-sectional samples (one for each EU-wide stress test) and one longitudinal sample (over the entire study period from 2010 to 2018) to address the different nature of the research questions (see Sections 1.4 and 4.3). Research Questions 1 (the *Informational Value Hypothesis*) and 2 (the *Functional Relationship Hypothesis*) use pooled (cross-sectional) data from the five EU-wide stress tests. In contrast, Research Question 3 (the *Intertemporal Stability Hypothesis*) uses panel (longitudinal) data from those banks that were continuously subjected to EU-wide stress tests over the entire study period. Otherwise, the sampling procedure is identical for all samples used in this study.

The samples were constructed using census sampling (also known as *total population sampling*). This is a purposive non-probability sampling method that selects as many elements as possible from the population into the sample. There are two reasons for choosing this sampling method: first, the population (*i.e.* the banks that have been subjected to EU-wide stress tests) is finite, and second, the size of the population is relatively small (N = 48 to N = 123).[97] Given this research setting, the advantage of census sampling over alternative sampling methods is that it maximises the coverage of the population of interest and thus the representativeness of the final samples. Census sampling therefore, by definition, minimises the sampling error. This allows deeper insights into the phenomena under investigation and reduces the risk of missing out on potential insights from unsampled banks.

The sampling begins with defining the population as the set of banks that were subjected to any of the five EU-wide stress tests carried out during the study period from 2010 to 2018. This yields a total of 403 bank-year observations. A list of the population is provided in Appendix C. However, since the population includes banks that are non-stock corporations or that are not publicly traded, these bank-year observations were excluded due to a lack of data. This applies to state banks, cooperative

[97] For more information about the population and the final samples, see Table 8.

banks, and captive banks, but also to banks that were merged or nationalised and delisted in the wake of the global financial crisis (2007-2009) or the subsequent European sovereign debt crisis (2010-2013). In addition, as stated in the research strategy (Section 5.3.1.3), bank-year observations of banks that were exposed to known confounding factors during an event window were excluded from the analysis.[98] This results in a total of 227 final bank-year observations of 72 unique bank stocks and cross-sectional sample sizes from n = 33 to n = 59. The cross-sectional samples are detailed in Appendix D.

Since the longitudinal analysis only contains those banks that were continuously subjected to all five EU-wide stress tests over the entire study period, further bank-year observations had to be excluded to construct the longitudinal sample. This results in a total of 140 final bank-year observations and a longitudinal sample size of n = 28. The longitudinal sample is detailed in Appendix E. Table 8 reconciles the population with the cross-sectional and longitudinal samples.

Table 8

Reconciliation of the Population and the Cross-Sectional and Longitudinal Samples

	EU-Wide Stress Tests					
	CEBS 2010	**EBA 2011**	**EBA 2014**	**EBA 2016**	**EBA 2018**	**Bank-Year Observations**
N	91	90	123	51	48	403
Panel A: Cross-Sectional Samples						
Less exclusions	41	39	64	17	15	176
n	50	51	59	34	33	227
Panel B: Longitudinal Sample						
Less exclusions	63	62	95	23	20	263
n	28	28	28	28	28	140

Note. This table summarises the sampling procedure used in this study for the five cross-sectional samples and the longitudinal sample. The starting point was the population of all banks that were subjected to EU-wide stress tests during the study period from 2010 to 2018. This population was adjusted for banks that were non-stock corporations or whose stocks were not publicly traded (unavailability of data). In addition, banks that were exposed to a known extraneous or confounding factor during a relevant event window were excluded from the analysis. The remaining banks formed the five cross-sectional samples (one for each EU-wide stress test) with sample sizes from n = 33 to n = 59. Basically the same sampling procedure was used to construct the longitudinal sample. In addition, however, every bank that was not continuously subjected to all five EU-wide stress tests during the study period was also excluded. This resulted in a longitudinal sample with n = 28 sample banks. CEBS = Committee of European Banking Supervisors. EBA = European Banking Authority. N = population size. n = sample size.

[98] In the context of this study, known confounding factors include *ad hoc* disclosures, director dealings, and ex-dividend days, as described in detail in Section 5.3.1.3. Confounding events were identified using the DGAP database of RegTech provider EQS Group.

Data Collection

Two main types of data were used in this study: the capital ratios (independent variable) and stock prices (dependent variable) of the sample banks. The capital ratios were collected from the official bank-level result reports published by the CEBS and EBA after each EU-wide stress test. More precisely, the data collected included the *stressed* capital ratio of each sample bank and its *actual* capital ratio at the end of the fiscal year before the respective stress test. Daily stock prices for all sample banks were collected from Bloomberg; the complete time series extends from 1 October 2009 to 30 November 2018. Based on these stock prices, simple (discrete) returns were calculated from closing to closing prices as follows:

$$R_t = \frac{P_t}{P_{t-1}} - 1, \tag{23}$$

where R_t is the daily simple return between trading days t and $t-1$, P_t is the closing price at trading day t, and P_{t-1} is the closing price at trading day $t-1$.

All of the above data were collected using structured direct observation. This systematic method is best suited for collecting standardised quantitative data such as capital ratios from bank supervisory reports or stock prices from financial data providers. The transformed stock prices (simple returns) and stressed capital ratios were used throughout the study, while the actual capital ratios (along with the other data) were only used to address the *Functional Relationship Hypothesis* (Research Question 2).

In addition, input data were required to run the asset pricing models for the normal return estimation (Section 5.3.2.1.3). This includes proxies for the market rate (Euro Stoxx Price Index) and the risk-free rate (ECB yield curve spot rate, one-year maturity), as well as the Fama-French factors for Europe. The data sets were collected from Bloomberg, the ECB Statistical Data Warehouse, and the Data Library on Kenneth French's research website, respectively.[99] The time series cover the period from 1 October 2009 to 30 November 2018. In accordance with the stock prices of the sample banks, all data sets were collected on a daily basis using structured direct observation. The data are described in more detail below.

[99] Kenneth French's website at the Tuck School of Business at Dartmouth College is available at: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/index html

The Euro Stoxx Index is a broad and liquid benchmark index that represents a large part of the European stock market. It currently consists of 288 stocks from a wide range of countries and sectors.[100] This makes the Euro Stoxx Index one of the most diversified stock market indices in Europe. The Euro Stoxx Index is therefore a suitable proxy for the market rate in studies on EU-wide stress tests. In this study, the price return variant of the index (Euro Stoxx Price Index) was used, which only reflects the price movements (*i.e.* capital gains or losses) of the index stocks.[101] That is, the Euro Stoxx Price Index represents an isolated measure of "clean" stock price movements, free from other influencing factors such as dividends. This was important in the context of this study as it concerned market efficiency and the adjustment of stock prices to new information from EU-wide stress test results. In other words, the price returns used were a highly content-valid measure that measured what they were supposed to measure. The decision to use the price return variant of the Euro Stoxx Index for this study was therefore based on the phenomena of interest and avoiding negative impacts on content validity. Using a price return index as a proxy for the market rate when estimating normal (expected) returns was also consistent with the exclusion of banks with ex-dividend days during event windows (Section 5.3.1.3). In this way, dividend effects were eliminated from both the estimated normal returns and the observed actual returns, ensuring that all return components of the abnormal return calculation (Equation (3)) were free of dividend effects and represented uncontaminated stock price movements.

The one-year spot rate of the ECB yield curve was used as a proxy for the risk-free rate, where this annual rate was recalculated to match the number of days in the event windows. The ECB yield curve is calculated using the Svensson (1994) method, an extension of the parametric model developed by Nelson and Siegel (1987). The basic idea is to fit an exponential yield curve to the yield of government securities that are considered risk-free. This approach is well established and is widely used by central banks to approximate domestic risk-free rates for different maturities (Nymand-

---

[100] For a list of the components of the Euro Stoxx Index with country, sector, and weight information, see: https://www.stoxx.com/document/Indices/Factsheets_Components/2021/July/SXXGT.pdf
The total weight of the sample banks in the benchmark index varied from stress test to stress test, ranging from 7.30% (EU-wide stress tests 2010 and 2014) to 7.48% (EU-wide stress test 2018).

[101] In contrast, the total return variant of the index also includes dividends, interest, subscription rights, and other distributions.

Andersen 2018).[102] In the literature, on the other hand, it is more common to use interbank offered rates (*e.g.* LIBOR or EURIBOR) or the yield on US Treasury bills as proxies for the risk-free rate. However, there are several problems associated with these alternative approaches. First, interbank offered rates tend to reflect higher risks than government securities and are therefore generally less "risk-free". Second, the time series required for this study is distorted by manipulated interbank offered rates (LIBOR and EURIBOR scandals) and is therefore not reliable.[103] Finally, the alternative use of US Treasury bill yields would inappropriately reflect the European interest rate and currency environment implied in this study. In contrast, ECB yield curve spot rates are not affected by any of these problems, as they are based on AAA-rated government securities from the euro zone and are not known to have been manipulated.

The Fama-French factors denote the risk factors required to run the multi-factor asset pricing models developed by Eugene Fama and Kenneth French. These are the Fama and French (1993) Three-Factor Model (Equation (28)) and the extended Fama and French (2015) Five-Factor Model (Equation (29)). Both models add risk factors to the market risk factor in order to reflect stock return patterns that remain unexplained by single-factor models such as the Market Model or the CAPM. The Fama-French factors are: *firm size* and *book-to-market ratio* (Three-Factor Model) plus *profitability* and *investment* (Five-Factor Model). All Fama-French factors are available on a daily basis and for a range of markets including Europe. The factors are constructed from diversified stock portfolios, which are formed according to factor-specific criteria and breakpoints. On the basis of these portfolios, the equally-weighted average return of the portfolios with high factor exposures is subtracted from the equally-weighted average return of the portfolios with low factor exposures. More specifically, this procedure results in the following factor returns: *SMB* (small minus big firm size), *HML* (high minus low book-to-market ratio), *RMW* (robust minus weak profitability), and *CMA* (conservative minus aggressive investment).[104] These factor returns, along with the corresponding factor coefficients (s_i, h_i, r_i, and c_i), are finally used as input pa-

102 The yield curve modelling methods of Svensson (1994) and Nelson and Siegel (1987) are used, for example, by the Deutsche Bundesbank, the Banco de España, the Banca d'Italia, and the Banque de France.

103 For more information on the LIBOR and EURIBOR scandals, see, for example, Ashton and Christophers (2015), Hou and Skeie (2014), and McConnell (2013).

104 For more detailed information on the construction procedure, see Fama and French (1993, 2015).

rameters in the Fama-French asset pricing models. The factor coefficients are determined by linear OLS regression of the sample banks' stock returns on the factor returns.

5.3.2.1.3 Model Selection

The next stage in this event study was to select the asset pricing model that was used to estimate the normal (expected) return of each sample bank. This involved defining a set of candidate models and performing a systematic model selection procedure based on the recommendations of Burnham and Anderson (2002). Since the standard event study approach of Campbell *et al.* (1997) and MacKinlay (1997) lacks a formal model selection part, the introduction of a systematic model selection procedure represented an extension of the standard approach. In addition, the multiple-pretest design described in Section 5.3.1.3 was operationalised by running multiple asset pricing models to estimate the sample banks' normal returns. The systematic model selection was also crucial for the overall validity of this study, since the normal return estimates played a decisive role in the calculation of the abnormal returns (see Equation (3)).

Definition of the Set of Candidate Models

As a starting point for defining the set of candidate models, the following three premises were established. First, any asset pricing model considered should be well established and widely used in the econometric literature. Second, the set of candidates should include statistical and economic models ranging from simple single-factor models to more complex multi-factor models in order to reflect a wide range of different approaches. Third, given the size-based selection by the CEBS and EBA, the candidate set should include models that explicitly consider firm size as one of the asset pricing factors (Section 5.3.1.3).

Based on these premises, a set of $R = 6$ asset pricing models was defined.[105] The six candidate models considered included: (1) the Mean-Adjusted Return Model, (2) the Market-Adjusted Model, (3) the Market Model, (4) the Capital Asset Pricing Model, (5) the Fama and French (1993) Three-Factor Model, and (6) the Fama and

[105] In their seminal textbook on model selection and multi-model inference, Burnham and Anderson (2002) advocate including at least four models in the set of candidates.

French (2015) Five-Factor Model. The individual models are described in more detail below and are presented in increasing order of complexity.

The *Mean-Adjusted Return Model (MAR)*, also known as Constant-Mean Return Model, assumes that the expected return of a stock is equal to its historical mean return. The resulting estimate of the expected return is therefore a constant that does not take market-wide (systematic) factors into account. The MAR is the simplest of the six candidate models. However, Brown and Warner (1980, 1985) have argued that it can produce results that are qualitatively similar to those of more complex models. Examples of econometric studies that have used the MAR include Hundt *et al.* (2017), Kalay and Loewenstein (1985), and Lahey and Conn (1990). Formally, the MAR estimates the expected return $E(R_{it})$ of stock i on trading day t as

$$E(R_{it}) = \bar{R}_i + \epsilon_{it}, \tag{24}$$

where $\bar{R}_i$ is the mean return of the stock[106] and ϵ_{it} is a normally distributed error term with zero mean and constant variance σ^2.

The *Market-Adjusted Model (MAM)*, also known as the Market Index Model, assumes that the expected return of a stock is equal to the market return as represented by a benchmark index. Since the MAM does not take into account any firm-specific (unsystematic) factors, the expected return is constant across different stocks, but not necessarily across time. The MAM can thus be interpreted as a restricted Market Model (Equation (26)), with α_i constrained to be zero and β_i constrained to be one. It has been used, for example, by Larsen and Resnick (1999), Maynes and Rumsey (1993), and Rajgopal *et al.* (2002). The MAM estimates the expected return $E(R_{it})$ of stock i on trading day t as

$$E(R_{it}) = R_{mt} + \epsilon_{it}, \tag{25}$$

where R_{mt} is the market return on trading day t and ϵ_{it} is a normally distributed error term with zero mean and constant variance σ^2.

106 In more formal terms, $\bar{R}_i$ can be expressed as $\bar{R}_i = \frac{1}{n}\sum_{i=1}^{n} R_i = \frac{R_1+R_2+ \quad +R_n}{n}$, where R_i is the return of stock i on a given trading day and n is the number of trading days in the estimation period.

The *Market Model (MM)*, first proposed by Sharpe (1963), establishes a linear relationship between the expected return of a stock and the market return.[107] It assumes that a stock's expected return can be expressed as the sum of its excess return and its proportional market return. The MM therefore covers both market-wide and firm-specific factors. The Security Market Line (SML) implied in the MM formed the empirical basis for the later development of the CAPM (Stapleton and Subrahmanyam 1983). According to Armitage (1995), the MM is the most commonly used model for estimating normal (expected) returns in event studies. This general observation also applies to previous studies of EU-wide stress tests (see Table 3). The expected return $E(R_{it})$ of stock i on trading day t is estimated by the MM as

$$E(R_{it}) = \alpha_i + \beta_i R_{mt} + \epsilon_{it}, \tag{26}$$

where α_i is the excess return, β_i is the beta coefficient,[108] R_{mt} is the market return, and ϵ_{it} is a normally distributed error term with zero mean and constant variance σ^2.

The *Capital Asset Pricing Model (CAPM)* was developed independently by Sharpe (1964), Lintner (1965a, 1965b), and Mossin (1966). It is one of the most widely used asset pricing models. Similar to the MM, the CAPM assumes that a stock's expected return can be estimated based on its sensitivity to the market. However, the beta coefficient of the CAPM is not applied to the market return, but to the market risk premium (*i.e.* the difference between the expected market rate and the risk-free rate). To compensate for this adjustment, the risk-free rate is then added again, as shown in Equation (27). The expected returns estimated by the CAPM can thus be decomposed into the risk-free rate and the proportional market risk premium. Examples of studies that have used the CAPM are Chan *et al.* (1992), Eisdorfer and Giaccotto (2014), and Frank and Shen (2016). The CAPM was also used by Candelon and Sy (2015) in a study of supervisory stress tests in the EU and the US (Table 3). The CAPM estimates the expected return $E(R_{it})$ of stock i on trading day t as

---

107 Sharpe (1963) originally referred to the Market Model as the „Diagonal Model".

108 More formally, β_i can be expressed as $\beta_i = \frac{Cov(R_i, R_m)}{Var(R_m)} = \rho_{i,m} \frac{\sigma_i}{\sigma_m}$, where $\rho_{i,m}$ is the correlation coefficient between stock i and market m and σ is the standard deviation.

$$E(R_{it}) = R_{ft} + \beta_i\left[E(R_m) - R_{ft}\right] + \epsilon_{it}, \tag{27}$$

where R_{ft} is the risk-free rate, β_i is the beta coefficient, $E(R_m) - R_{ft}$ is the market risk premium, $E(R_m)$ is the expected market return, and ϵ_{it} is a normally distributed error term with zero mean and constant variance σ^2.

The *Fama and French (1993) Three-Factor Model (FF3F)* introduced two additional risk factors to explain stock return patterns that remain unexplained by single-factor models, which only consider the market risk factor. These additional risk factors are the size and book-to-market ratio of a firm (as described in Section 5.3.2.1.2). They are based on the empirical observation that small stocks and stocks with a high book-to-market ratio (also known as "value stocks") tend to outperform other stocks.[109] The FF3F therefore covers both market-wide and firm-specific factors. It has been used, for example, by Adrian and Rosenberg (2008), Feng *et al.* (2021), and Tetlock *et al.* (2008). Formally, the FF3F estimates the expected return $E(R_{it})$ of stock i on trading day t as follows:

$$E(R_{it}) = R_{ft} + a_i + \beta_i\left(R_{mt} - R_{ft}\right) + s_i SMB_t + h_i HML_t + e_{it}. \tag{28}$$

In this equation, R_{ft} is the risk-free rate, α_i is the excess return, β_i is the beta coefficient, R_{mt} is the market return, s_i and h_i are the size and book-to-market factor coefficients, respectively, SMB_t and HML_t are the size and book-to-market factor returns, respectively, and ϵ_{it} is a normally distributed error term with zero mean and constant variance σ^2.

The *Fama and French (2015) Five-Factor Model (FF5F)* extended the FF3F to include two additional risk factors: the profitability and investment patterns of firms. The inclusion of these additional risk factors built on their previous work on asset pricing anomalies (Fama and French 2006, 2008), which suggested that a high (low) exposure of firms to profitability and investment factors can explain the out- (under-) performance of their stocks compared to other stocks. The construction of the profita-

[109] For more information on the size effect, see, for example, Banz (1981), Brown *et al.* (1983), and Fama and French (1992). For details on the book-to-market ratio effect, see, for example, Barber and Lyon (1997a), Fama and French (1992), and Lakonishok *et al.* (1994). See also Section 3.5.3 for a general discussion of risk factors that influence the formation of stock prices.

bility and investment factors is analogous to the other Fama-French factors, as explained in Section 5.3.2.1.2. Examples of studies that have used the FF5F are Ashton and Trinh (2018), Hossain and Kryzanowski (2021), and Roussanov *et al.* (2021). The FF5F was also recently used by Georgoutsos and Moratis (2021) to estimate normal (expected) returns in a study of EU-wide stress tests (Table 3). The FF5F estimates the expected return $E(R_{it})$ of stock i on trading day t as

$$E(R_{it}) = R_{ft} + a_i + \beta_i\big(R_{mt} - R_{ft}\big) + s_i SMB_t + h_i HML_t \\ + r_i RMW_t + c_i CMA_t + e_{it}. \tag{29}$$

In this equation, R_{ft} is the risk-free rate, α_i is the excess return, β_i is the beta coefficient, R_{mt} is the market return, s_i, h_i, r_i, and c_i are the size, book-to-market, profitability, and investment factor coefficients, respectively, SMB_t, HML_t, RMW_t, and CMA_t are the size, book-to-market, profitability, and investment factor returns, respectively, and ϵ_{it} is a normally distributed error term with zero mean and constant variance σ^2.

Systematic Model Selection Procedure

The goal of the model selection procedure was to identify the asset pricing model from the set of candidate models whose normal return estimates best fitted the actual returns of the sample banks. In other words, the goal was to determine the model that best approximated the real data-generating process (*i.e.* the *true model*). The model selection procedure described below was therefore crucial in order to address the joint-hypothesis problem (Fama 1970, 1991). That is, to minimise the risk that any abnormal returns found are due to a bad model and not to the disclosure of EU-wide stress test results.

The model selection procedure was carried out systematically based on the goodness-of-fit of the candidate models. Measures of goodness-of-fit typically summarise the difference between the observed values (actual returns) and the expected values (normal returns) estimated by the model under investigation. Accordingly, the sample banks' normal returns were estimated from each of the candidate models as described in Section 5.3.2.1.4. The model-specific normal return estimates were then tested against a common set of actual return data, *i.e.* the full set of $n = 227$ available bank-year observations (see Table 8). This represented the largest possible data set for model selection.[110] In order to take into account the three different types of event windows, the return data were divided by event-window type (*pre-event window*, *standard event window*, and *post-event window*). The resulting subsets formed the basis for the goodness-of-fit model selection procedure.

To assess the goodness-of-fit of the candidate models, linear OLS regressions and analyses based on statistical information criteria were performed. More precisely, the Akaike Information Criterion (AIC), the Schwarz Information Criterion (SIC), and the Hannan-Quinn Information Criterion (HQIC) were employed. The use of these more sophisticated methods became necessary because the results of the initial regression analysis were not very meaningful. This was probably due to the limitations of correlation-based measures of goodness-of-fit, such as the coefficient of determination (R^2) and adjusted R^2, which are very sensitive to outliers and insensitive to additive and proportional differences between model estimates and actual observations (see Willmott 2009).

In the regression analysis, the normal return estimates $E(R)$ of each candidate model were regressed on the actual returns R in order to determine the extent to which a candidate model explained the actual (observed) returns. The measures used to evaluate the goodness-of-fit of the candidate models were the sum of squared errors (SSE), R^2, and adjusted R^2. However, the results did not favour any particular model as the goodness-of-fit measures were similar for five of the six candidate models across all event-window types. These models were the MAM, the MM, the CAPM, the FF3F, and the FF5F. Conversely, the results suggested removing at least the remaining model (MAR) from the set of candidates. Notably, the historical-return-based MAR was the simplest of the six candidate models, and its poor goodness-of-fit was consistent with weak-form market efficiency, according to which security prices change randomly and are independent of past returns (Fama 1970, 1991). The detailed results of the regression-based goodness-of-fit tests are given in Appendix F. In order to obtain a more reliable basis for the model selection, the regression analysis was followed up with statistical information criteria.

---

[110] In order to ensure consistency throughout this study, no separate model selection was carried out based on the subset of the $n = 140$ bank-year observations of the longitudinal sample (see Table 8).

Information criteria were used to assess the goodness-of-fit of a candidate model relative to the other models in the candidate set. The basic idea is to identify and select the model that minimises the information loss (Kullback and Leibler (1951) (KL) Divergence) between the candidate model and the *true model*. This model is known as the KL best model.[111] The aim of using information criteria in this study was therefore to identify the candidate model that minimises the estimation error, *i.e.* the error terms in Equations (24) to (29)(1). A common feature of information criteria is that they reward not only the goodness-of-fit of a candidate model, but also its parsimony in terms of the number of parameters estimated (k). This is because the goodness-of-fit of a model increases monotonically with the number of parameters it contains, *i.e.* it can always be improved by adding more parameters, since each additional parameter leads to a reduction in the sum of squared errors. Formally, information criteria follow the principle of parsimony (Occam's razor) by including a penalty term in order to discourage overfitting (similar to the adjusted R^2 measure). Information criteria therefore aim to reconcile the goodness-of-fit of a model with its parsimonious use of parameters. In this way, they simultaneously address the risk of under- and overfitting. As shown in the following equations, the maximum likelihood function was the same for all information criteria used in this study, while their penalty terms were different and varied in severity.[112] All information criteria values were calculated based on the common sample size of $n = 227$ bank-year observations and the individual number of parameters (k) estimated by the candidate models.

The Akaike (1973, 1974) Information Criterion is probably the most widely used measure of goodness-of-fit model selection. In the OLS framework, the *AIC* value for each candidate model i is estimated as follows:[113]

---

[111] For more details on information criteria-based model selection, see Burnham and Anderson (2002).

[112] In every analysis with a sample size of $n \geq 16$, the SIC penalises the most and the AIC the least, while the penalisation level of the HQIC lies between those of the other two information criteria. If the sample size is $n < 16$, the order changes between AIC and HQIC, and if the sample size is $n < 8$, the AIC becomes the information criterion that penalises the most.

[113] In a more general form, the maximum likelihood function can also be expressed as $-2\log(\mathcal{L}(\hat{\theta}|y))$, where $\log(\mathcal{L}(\hat{\theta}|y))$ is the numerical value of the log-likelihood at its maximum. This more general form can also be employed in the other information criteria used in this study (SIC and HQIC).

$$AIC_i = n * \ln\left(\frac{SSE}{n}\right) + 2k. \tag{30}$$

In this equation, $n * \ln\,(SSE/n)$ is the maximum likelihood function of the model, n is the sample size, SSE is the sum of squared errors $\sum(Y_i - \widehat{Y_i}\,)^2$, and $2k$ is the penalty term with k being the number of parameters estimated by the model.

The Schwarz (1978) Information Criterion estimates the *SIC* value for each candidate model i as

$$SIC_i = n * \ln\left(\frac{SSE}{n}\right) + \ln(n)\,k, \tag{31}$$

where $n * \ln\,(SSE/n)$ is the maximum likelihood function of the model, n is the sample size, SSE is the sum of squared errors $\sum(Y_i - \widehat{Y_i}\,)^2$, and $\ln(n)\,k$ is the penalty term with k being the number of parameters estimated by the model.

The Hannan and Quinn (1979) Information Criterion uses the law of the iterated logarithm (LIL) and estimates the *HQIC* value for each candidate model i as

$$HQIC_i = n * \ln\left(\frac{SSE}{n}\right) + \ln(\ln(n))2k, \tag{32}$$

where $n * \ln\,(SSE/n)$ is the maximum likelihood function of the model, n is the sample size, SSE is the sum of squared errors $\sum(Y_i - \widehat{Y_i}\,)^2$, and $\ln(\ln(n))2k$ is the penalty term with k being the number of parameters estimated by the model.

The information criterion values resulting from Equations (30) to (32) were used as a direct measure for the goodness-of-fit, but also formed the basis for further analyses. More precisely, for the analysis of information criterion differences (ΔIC_i), relative likelihoods (L_i), and Akaike weights (w_i). These associated metrics are specified in more detail below. Since they were calculated for all three information criteria (*AIC*, *SIC*, and *HQIC*), the metrics are described using a generic information criterion.

Information criterion differences (ΔIC_i) build on the fact that all information criteria favour the candidate model that minimises the information loss. This KL best model is defined as the candidate with the lowest information criterion value (IC_{min}). Building on that, the information criterion difference of candidate model i is calculated as the difference between its own information criterion value (IC_i) and IC_{min}. Formally, the ΔIC of candidate model i is calculated as follows:

$$\Delta IC_i = IC_i - IC_{min}. \tag{33}$$

Accordingly, the ΔIC of the KL best model is zero. The use of information criterion differences supports the analysis, since it is not the absolute level of the information criterion values that is decisive, but the relative values and thus the differences. Burnham and Anderson (2002) provide some guidance on how to interpret ΔIC in terms of the level of empirical support for a given candidate model i. According to them, a ΔIC_i between 0 and 2 represents *substantial* empirical support, while a ΔIC_i between 4 and 7 represents *considerably less*, and a $\Delta IC_i > 10$ represents *essentially none* empricial support. The information criterion difference of a candidate model also forms the basis for calculating its relative likelihood.

The *relative likelihood* (L_i) of a candidate model i can be interpreted as the relative probability that this model is the KL best model. For any candidate model i, the relative likelihood L_i is calculated as

$$L_i = \exp\left(-\frac{1}{2} * \Delta IC_i\right), \tag{34}$$

where $\exp(.)$ is the exponential function and ΔIC is the information criterion difference. The relative likelihoods of the candidate models can also be used to calculate the Akaike weight of each of the models.

The *Akaike weight* (w_i) of a candidate model i is the weight of evidence that the model is actually the KL best model, given that one of the candidate models must be the KL best model of the candidate set. Akaike weights thus improve interpretability by normalising the relative likelihoods of the candidate models to 1. Given the data and a set of R candidate models, the Akaike weight w_i of candidate model i is calculated as follows:

$$w_i = \frac{\exp\left(-\frac{1}{2} * \Delta IC_i\right)}{\sum_{r=1}^{R} \exp\left(-\frac{1}{2} * \Delta IC_r\right)}. \tag{35}$$

In this equation, $\exp(.)$ is the exponential function and ΔIC is the information criterion difference.

To illustrate the relationship between these metrics, it can be stated that for the KL best model (IC_{min}) from the set of R candidate models, $\Delta IC_{min} = 0$. Therefore, for this model, $L = \exp\,(-1/2 * \Delta IC_{min}) = 1$. The probability that candidate model i is actually the KL best model is thus $\exp\,(-1/2 * \Delta IC_i)$ to 1. To facilitate interpretation, it is convenient to transform the probabilities for each of the candidate models into Akaike weights (w_i). The smaller the ΔIC of a candidate model i, the greater its w_i, and the higher the probability that this model is the actual KL best model for the data and set of R candidate models used. The results of the model selection procedure are summarises below.

Model Selection Results

The results of the information criteria values and metrics confirmed the basic results of the regression analysis, *i.e.* that the MAR was not particularly well suited for this study. However, the information criteria analysis also provided clear evidence as to which of the remaining candidate models should be used to estimate normal returns. The results of the AIC, SIC, and HQIC showed almost unanimously for all event-window types that the FF3F was the KL best model given the data and the candidate set considered. The only exceptions were the AIC results for the pre-event window and the standard event window, where the FF5F was identified as the KL best model, closely followed by the FF3F.[114] In view of the third starting premise for the definition of the set of candidate models, it was not entirely surprising that the information criteria favoured the two models that explicitly considered firm size as an asset pricing factor.

[114] These results are laregely consistent with those of Fama and French (2016), who tested the FF5F against the FF3F and found that the FF5F generally performed better than the FF3F, except for portfolios formed on firm size or accruals. In this context, it should be mentioned again that the banks subjected to EU-wide stress tests (and thus also the sample banks) were selected by the EBA and CEBA according to a size-based selection rule (Section 5.3.1).

More specifically, apart from the two above exceptions, the FF3F consistently showed the lowest information loss (IC_{min}) of all candidate models, across all information criteria (AIC, SIC, and HQIC) and event-window types. The FF3F thus also consistently showed the smallest possible information criterion difference of $\Delta IC = 0$, which can be interpreted as substantial empirical support for the model (Burnham and Anderson 2002). As a result, the relative likelihood (L) and Akaike weight (w) of the FF3F were consistently the highest of all candidate models, with L always being 1 and w ranging between .510 and .995. More detailed results of the information criteria-based goodness-of-fit tests are provided in Appendix G.

Given the results of the information criteria and regression analysis, it seemed reasonable to accept the FF3F as the most appropriate model for estimating normal returns in this study. This is especially true when the parsimony of the candidate models is emphasized and the results of the SIC and the HQIC are therefore given more weight than those of the AIC. Based on these considerations, the FF3F was selected as the asset pricing model for estimating the normal returns of the sample banks.

The selection of the FF3F model and the underlying model selection procedure represent an original contribution to knowledge in two ways. First, previous studies of abnormal returns in response to EU-wide stress tests have never performed systematic model selection, but have chosen an arbitrary asset pricing model or resorted to the convenient market model (Table 3). Second, to date, the selected FF3F model has never been used in the context of EU-wide stress tests and was therefore used for the first time in this study to estimate normal returns. The estimation procedure is described in more detail in the next section.

5.3.2.1.4 Normal Return Estimation

The estimation of the sample banks' normal returns was nested in the preceding model selection stage, since the estimates were necessary to assess the goodness-of-fit of the candidate models. However, the normal return estimates of the selected FF3F model were also needed to eventually calculate the sample banks' abnormal returns (see Equation (3)), which is explained in more detail in the following section. The main concern of this section, however, is the definition of the estimation period over which the parameters of the candidate models were estimated.

When defining estimation periods in the context of event studies, there are two aspects to consider: the calendar time in which the estimation period falls and the length of the estimation period. These two aspects are explained in more detail below.

First, it is important to avoid overlap between the estimation period and the event window to prevent event-induced volatility from affecting the estimation of reliable model parameters (Campbell *et al.* 1997, MacKinlay 1997). This is especially important when analysing multiple events that are spread out over time, as in this study. The maximum length of the estimation period was therefore limited by the time available between the event windows of any two consecutive EU-wide stress tests. In addition, a common buffer period of two trading days was used between the end of an estimation period and the start of the following *pre-event window*; the buffer periods between the end of a *post-event window* and the beginning of the subsequent estimation period were even longer.[115] The 120-day estimation period that was finally chosen (see below) also had the advantage that it allowed the estimation periods and event windows to be defined without any overlapping problems. Figure 12 illustrates these relationships.

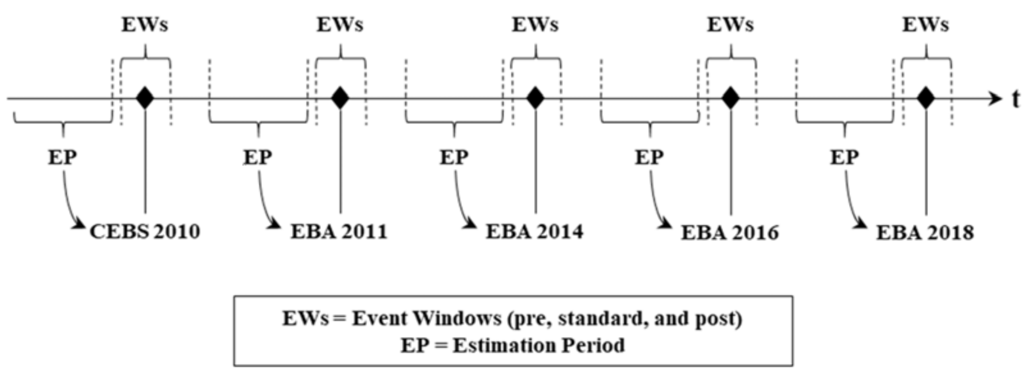


Figure 12. Overall event study timeline with multiple events

[115] All buffer periods between the end of a post-event window and the start of the subsequent estimation period resulted from the common length of the estimation period (120 trading days) and the calendar date on which the EU-wide stress test results were disclosed. The shortest of these buffer periods was between the end of the post-event window of the CEBS 2010 and the start of the estimation period for the EBA 2011 and was 126 trading days (all other buffer periods were considerably longer).

The fact that the estimation periods ended almost immediately before the start of the corresponding event windows helped also to reduce the impact of potential confounding events by including them in the estimation of the model parameters. Examples of confounding events affecting stock prices include the global financial crisis (2007-2009) and the subsequent European sovereign debt crisis (2010-2013).

Second, the statistical reliability of model parameters is also influenced by the length of the estimation period. Both long and short estimation periods have certain advantages and disadvantages that are mutually exclusive. While a long estimation period usually benefits from a larger database, it is also more susceptible to noise and confounding factors than a short estimation period (and *vice versa*). Therefore, there is a tradeoff in the length of the estimation period that must be taken into account.

There appears to be no consensus in the econometric literature about the appropriate length of the estimation period. For example, Fiordelisi *et al.* (2020) used a 252-day estimation period, while Kolaric *et al.* (2021) and Mukhtarov *et al.* (2021) have used estimation periods of 120 and 60 days, respectively. In this context, Thompson (1995, p. 973) noted that "[t]here is discretion over the choice of [*sic*] nonevent estimation period for most empirical investigations." This discretion can also be seen in previous studies of EU-wide stress tests, whose estimation periods have ranged from 30 to 262 days. However, despite this heterogeneity, previous studies of EU-wide stress tests have tended to use a medium-long estimation period of 120 days (Table 3). This estimation period has also been suggested in the methodological literature for event studies based on daily data (Campbell *et al.* 1997, MacKinlay 1997). Following this suggestion and most previous studies of EU-wide stress tests, a common 120-day estimation period was used in this study to estimate the normal returns for each of the candidate models. The normal return estimates of the FF3F model selected in the model selection stage were then used to calculate the abnormal returns of the sample banks. The abnormal return calculation is detailed in the following section.

5.3.2.1.5 Abnormal Return Calculation

The final stage of this event study was to calculate the abnormal return (AR) of every sample bank using the actual (observed) return and the normal (expected) return estimated by the selected FF3F model. As a robustness check, abnormal returns were also calculated using the normal return estimates from the MM, which has been the most popular choice in previous studies of EU-wide stress tests (Table 3). In addition, this study also calculated the *absolute* abnormal return ($|AR|$) of every sample bank as a non-directional measure. The following calculation methods are model-independent and are therefore identical for both asset pricing models.

The abnormal return AR_i of a given security i is defined as the simple difference between its actual (observed) return R_i and its normal (expected) return $E(R_i|X)$. Any abnormal return found can therefore be interpreted as an unexpected return plausibly attributable to the event under investigation. The formal equation for calculating the abnormal return has already been given in Equation (3).

As an extension of the standard event study approach, the abnormal returns AR_i obtained for each sample bank i were converted into absolute abnormal returns $|AR_i|$ using the absolute function $\mathrm{abs}(.)$:

$$|AR_i| = \mathrm{abs}(AR_i). \tag{36}$$

Using absolute abnormal returns as an additional measure can greatly increase the analytical power of an event study. This is because the individual abnormal returns of the sample banks are usually aggregated across the sample to obtain the average cumulative abnormal return ($\overline{CAR}$), *i.e.* the average effect of the event on the value of the sample banks. This also applies to parts of this study (see Sections 5.3.2.2 and 5.3.2.4). The problem with averaging directional abnormal returns is that the analysis does not distinguish between positive and negative information effects. It is therefore not possible to distinguish whether a $\overline{CAR}$ close to zero results from small effects across the sample or from large positive and negative effects that almost cancel each other out on average. In contrast, when aggregating and averaging non-directional absolute abnormal returns, the resulting $|\overline{CAR}|$ should be large if investors respond to the information event, regardless of the distribution of positive and negative effects across the sample. Therefore, the larger the $|\overline{CAR}|$, the higher the value of the information disclosed (*ceteris paribus*).

The use of absolute abnormal returns has only recently emerged and can be attributed to a study by Flannery *et al.* (2017) on the information value of supervisory stress tests in the US. Since then, studies of US supervisory stress tests have repeatedly relied on absolute abnormal returns (*e.g.* Bird *et al.* 2020, Fernandes *et al.* 2020, and Fung and Loveland 2020). In contrast, so far only one study (Georgoutsos and Moratis 2021) has used this non-directional measure in the European context and only for the 2016 and 2018 EU-wide stress tests. In this respect, the use of absolute abnormal returns in a large number of EU-wide stress tests is a first for this study.

This final stage of the event study represents the end of the common methodological basis, which was the same for all three research questions. Methodological specifics and further analysis methods are described separately for each research question in the following sections.

5.3.2.2 The Informational Value Hypothesis

The *Informational Value Hypothesis* (Research Question 1) asked whether the disclosures of EU-wide stress test results provided investors with valuable new information. To answer this question, it was examined for each event-window type whether the result disclosures caused, on average, significant (absolute) abnormal returns in the cross-section of the samples. For this analysis, the individual (absolute) abnormal returns of the sample banks first had to be aggregated twice: over time and over each of the five samples. Subsequently, one-sample *t*-tests and Wilcoxon signed-rank tests were used to test whether the resulting $\overline{CAR}s$ and $|\overline{CAR}s|$ were statistically significant. The methods used to aggregate the data and test for statistical significance are described in more detail below.

5.3.2.2.1 Time-Series Aggregation

In the first aggregation step, the daily (absolute) abnormal returns of each sample bank were aggregated over time to obtain their $CARs$ and $|CARs|$ over the three different event-window types. In other words, the return data determined in Section 5.3.2.1.5 for each sample bank was aggregated over the event windows defined in Section 5.3.2.1.1, *i.e.* the *pre-event window*, the *standard event window*, and the *post-event window* of every EU-wide stress test.

Since the simple (discrete) returns used in this study[116] are not time additive, the time-series aggregation required multiplications rather than simple summations. Thus, the $CAR_{i\tau}$ for each sample bank i and each event window τ was calculated as

$$CAR_{i\tau}(t,T) = \prod_{t=1}^{T}(1 + AR_i) - 1, \tag{37}$$

where AR is the (daily) abnormal return and where t and T denote the beginning and the end of event window τ, respectively.

Similarly, the $|CAR_{i\tau}|$ for each sample bank i and each event window τ was calculated as

$$|CAR_{i\tau}|(t,T) = \prod_{t=1}^{T}(1 + |AR_i|) - 1, \tag{38}$$

where $|AR|$ is the (daily) absolute abnormal return and where t and T denote the beginning and the end of event window τ, respectively.

The operations shown in Equations (37) and (38) resulted in $CARs$ and $|CARs|$ for each sample bank and each combination of EU-wide stress test and event-window type. Building on this, the individual $CARs$ and $|CARs|$ were aggregated across the five cross-sectional samples defined in Section 5.3.2.1.2.

5.3.2.2.2 Cross-Sectional Aggregation

In the second aggregation step, the $CARs$ and $|CARs|$ of all sample banks were aggregated and averaged across the samples to obtain the $\overline{CAR}_{j\tau}$ and $|\overline{CAR}_{j\tau}|$ for each combination of EU-wide stress test and event-window type.

---

[116] Simple (discrete) returns were used throughout this study; this applies to both the actual (observed) returns and any derived returns (*i.e.* the normal (expected) returns and the (absolute) abnormal returns) of the sample banks, see Section 5.3.2.1.2.

Formally, the $\overline{CAR}_{j\tau}$ for each EU-wide stress test j and each event window τ was calculated as

$$\overline{CAR}_{j\tau} = \frac{1}{n}\sum_{i=1}^{n} CAR_{i\tau}, \tag{39}$$

where $CAR_{i\tau}$ is the individual cumulative abnormal return of each sample bank i over event window τ, and n is the number of sample banks in the sample.

Similarly, the $|\overline{CAR}_{j\tau}|$ for each EU-wide stress test j and each event window τ was calculated as

$$|\overline{CAR}_{j\tau}| = \frac{1}{n}\sum_{i=1}^{n} |CAR_{i\tau}|, \tag{40}$$

where $|CAR_{i\tau}|$ is the individual absolute cumulative abnormal return of each sample bank i over event window τ, and n is the number of sample banks in the sample.

The $\overline{CAR}s$ and $|\overline{CAR}s|$ resulting from Equations (39) and (40) provided preliminary evidence whether the disclosure of EU-wide stress test results conveyed valuable new information to investors. For a final evaluation, however, each individual $\overline{CAR}$ and $|\overline{CAR}|$ was tested for statistical significance. This is described in more detail in the following section.

5.3.2.2.3 Significance Testing

The most common method to determine whether the mean of a sample is significantly different from a specified value is the one-sample *t*-test. Its non-parametric equivalent is the one-sample Wilcoxon signed-rank test (*T*-test). Both tests were used to complete the analysis of the $\overline{CAR}s$ and $|\overline{CAR}s|$.[117]

[117] It should be noted that the z-test could not be used in this study because the standard deviation of the population (*i.e.* all banks that were subjected to EU-wide stress tests) was not known due to lack of data and exclusions due to known confounding factors (Section 5.3.2.1.2). In addition, the minimum sample size ($n > 30$) required for the z-test was barely achieved by the EBA 2016 ($n = 34$) and the EBA 2018 ($n = 33$).

The rationale for using both parametric and non-parametric significance tests was based on the results of the classic assumption tests for the one-sample t-test. The results indicated that not all assumptions were reasonably satisfied across all combinations of EU-wide stress tests and event-window types. Specifically, the assumption of normality was not met in most cases. This became evident from the descriptive statistics shown in Section 6.2 and from the results of two formal normality tests (Lilliefors test and Shapiro-Wilk test) that were carried out to supplement the descriptive statistics. The results of the Lilliefors test and the Shapiro-Wilk test are provided in Appendix H.

Although the results of the formal normality tests suggested the use of a non-parametric significance test (such as the Wilcoxon signed-rank test, which does not require normality), the t-test is fairly robust to deviations from its assumptions (Boneau 1960, Cicchitelli 1989, Posten 1979). Therefore, both the one-sample Wilcoxon signed-rank test and the one-sample t-test were applied to all combinations of EU-wide stress tests and event-window types. In this respect, the use of parametric and non-parametric significance tests represented a mutual robustness check.

In one-sample significance tests, the mean of the sample is compared to a standard value, *i.e.* a constant. This constant was zero for every $\overline{CAR}$, since the specific null hypothesis to be tested for each combination of EU-wide stress test and event-window type was whether the $\overline{CAR}$ was equal to zero (Section 4.3.1). In contrast, it would have been inappropriate to compare $|\overline{CAR}|$ to zero as this is an absolute-value measure. Following Flannery *et al.* (2017) and Georgoutsos and Moratis (2021), the significance of each $|\overline{CAR}|$ was instead assessed by comparing it to its average value over the corresponding estimation period (*i.e.* the average absolute estimation error of the asset pricing model used). Consequently, the null hypothesis for the $|\overline{CARs}|$ was adapted accordingly (Section 4.3.1).

The statistical significance tests concluded the analysis of the *Informational Value Hypothesis* (Research Question 1). The results presented in Section 6.3.1 provide new insights into the informational value of EU-wide stress test results by showing how the stock prices of the banks concerned responded, on average, to the newly disclosed information. This analysis was continued at a more detailed level by examining the *Function Relationship Hypothesis* (Research Question 2). The methods used to answer this research question are described in the next section.

5.3.2.3 The Functional Relationship Hypothesis

The *Functional Relationship Hypothesis* (Research Question 2) asked about the functional relationship between changes in the capital ratios and changes in the stock prices of the sample banks. To determine this risk-return relationship, a curve-fitting procedure was used to fit functions of the form $y = f(x)$. This required specifying the related variables at the bank level and fitting the functions to the empirically observed data for each of the five samples (*i.e.* for each EU-wide stress test). The resulting fits were then evaluated for the quality of their approximations using regression analyses and error metrics in order to determine the best-fitting function. It should be noted, however, that the fitted functions were not tested against each other, but rather independently against the actual, empirically observed, data. The related variables, the curve-fitting procedure, and the methods used to evaluate the performance of the fits are described in more detail below.

5.3.2.3.1 Related Variables

The specific variables that were related for each sample were: (1) the $CARs$ over the post-event window[118] and (2) the differences between the stressed and actual capital ratios (ΔCRs) of the sample banks. Formally, the capital ratio difference ΔCR_i for each sample bank i was calculated as

$$\Delta CR_i = \widehat{CR}_i - CR_i, \tag{41}$$

where $\widehat{CR}$ is the stressed capital ratio and CR is the actual capital ratio.

The $CARs$ formed the dependent variable and were taken from the time-series aggregated results of the event study (Section 5.3.2.2.1). The descriptive statistics on the $CARs$ can therefore be found in Section 6.2.1. The stressed and actual capital ratios, on the other hand, were as defined in Section 5.3.2.1.2. However, it should be noted

[118] The determination of a functional relationship requires that all relevant variables are available so that potential causal effects can be observed. Consequently, the *Functional Relationship Hypothesis* was constrained to examine the post-event window (*i.e.* the event window that only included times *after* the EU-wide stress test results, and thus the stressed capital ratios, had been disclosed). In contrast, the pre-event and standard event windows could not be meaningfully examined because the causality requirement that the cause precedes the effect was not met for these event-window types.

In addition, the analysis had to be limited to $CARs$, since the $|CARs|$ could not be meaningfully investigated because the relationship between ΔCRs and $|CARs|$ had been distorted by the absolute value transformation (Equation (36)).

that the stressed capital ratios $(\widehat{CR}s)$ used were those resulting from the adverse scenario of the stress tests. The rationale was that the stock price reactions represented by the $CARs$ were more likely due to the sample banks' results under the adverse scenario than under the milder baseline scenario. However, the stressed capital ratios from the baseline scenario were used as a robustness check along with other alternative risk measures (for further details, see Section 5.3.2.3.3).

In the next step, five sets of paired variables were formed from the $CARs$ and ΔCRs, one for each sample. In addition, scatter plots were generated from these sets to visualize the sample-specific relationship between the $CARs$ and ΔCRs. The sets of paired variables served as the basis for the curve-fitting procedure, which is explained in more detail in the following section.

5.3.2.3.2 Curve-Fitting Procedure

Curve fitting is the process of optimising a curve or a parametrised functional form to fit a set of data points that describe a relationship. More specifically, the curve-fitting procedure used in this study involved fitting polynomial functions to each of the five sets of paired variables (empirically observed $CARs$ and ΔCRs for each sample) defined in the previous section. However, the fitted functions were constrained to first- and second-degree polynomials to avoid unstable oscillation (Runge's phenomenon) and to keep the response function economically interpretable.

The polynomial functions considered were therefore linear ($y = ax + b$) and quadratic ($y = ax^2 + bx + c$), with the dependent variable y being the fitted $\widehat{CAR}s$ and the independent variable x being the ΔCRs. The linear (LF) and quadratic function (QF) above could therefore be rewritten for each sample bank i as

$$\widehat{CAR}_{i,LF} = a * \Delta CR_i + b, \tag{42}$$

and

$$\widehat{CAR}_{i,QF} = a * \Delta CR_i^2 + b * \Delta CR_i + c. \tag{43}$$

In order to fit these functions to the data, the variables a, b, and c were determined in such a way that the fitted $\widehat{CAR}s$ best represented the data points of the observed $CARs$. This procedure was performed separately for each of the five samples and involved least-squares minimisation of the SSE, *i.e.* the sum of squared errors between the observed $CARs$ and the fitted $\widehat{CARs}$. The resulting fitted value curves described the best possible linear and quadratic fits for the data. The next step in the analysis was to evaluate whether the linear or the quadratic fits performed better in describing the actual relationship between the ΔCRs and $CARs$. The methods used to conduct this evaluation are described in the next section.

5.3.2.3.3 Performance Evaluation

Every measure of performance is based on the comparison between the empirically observed and the fitted values of the dependent variable. In this study, the observed values were the $CARs$ (as defined in Section 5.3.2.2.1), while the fitted values were the $\widehat{CARs}$ predicted by the linear and quadratic fits, respectively, described in the previous section. The goal of the performance evaluation was to determine the extent to which the fitted $\widehat{CARs}$ explained the actually observed $CARs$. This implied answering the question of whether the linear or the quadratic fits provided a better explanation overall. It is important to note that the linearly and quadratically fitted $\widehat{CARs}$ were not tested against each other but independently against the empirically observed $CARs$ to evaluate the quality of their approximations. That is, no test for differences between the models was carried out.

The performance of the linear and quadratic fits was evaluated using regression analysis and error metrics. In particular, the Coefficient of Determination (R^2), the Mean Squared Error (MSE), and the Root Mean Squared Error ($RMSE$) were considered. These performance measures are described in more detail below.

In the basic regression analysis, the empirically observed $CARs$ of each sample were regressed on the fitted $\widehat{CARs}$ predicted by the linear and quadratic fits. The resulting R^2 values indicated the proportion of the total variance in the $CARs$ that was explained by the $\widehat{CARs}$. In other words, R^2 indicated how well the linear and quadratic fits approximated the actual data from each sample, with $R^2 = 1$ indicating an exact fit. This logic was reversed for the error metrics, which evaluated the performance of the linear and quadratic fits by quantifying the prediction error between the observed and fitted values. Error metrics such as MSE and $RMSE$ are therefore inverse measures of explanatory power and can be used to compare the performance of competing functions.

The MSE is probably the most commonly used error metric for continuous dependent variables like $\widehat{CAR}$. It measures the mean of the squared errors between the observed values ($CARs$) and the fitted values ($\widehat{CARs}$). The MSE thus aggregates the magnitude of the prediction errors for the various data points into a single performance measure. As a result, the MSE is almost always strictly positive with $MSE = 0$ for a perfect function that fits all measured values exactly. In this study, the MSE for each sample was calculated as

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(CAR_i - \widehat{CAR}_i\right)^2, \tag{44}$$

where CAR_i and $\widehat{CAR}_i$ are the empirically observed and fitted cumulative abnormal returns of sample bank i, respectively, and n is the sample size.

The $RMSE$ builds on the above as it is defined as the square root of the MSE. Both metrics are often used together. This is because the MSE is not readily interpretable, as the scale on which it is constructed is different from the scale of the dependent variable. The $RMSE$ corrects this problem and thus facilitates interpretation. Formally, the $RMSE$ for each sample was calculated as

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(CAR_i - \widehat{CAR}_i\right)^2}, \tag{45}$$

where CAR_i and $\widehat{CAR}_i$ are the empirically observed and fitted cumulative abnormal returns of sample bank i, respectively, and n is the sample size.

As a robustness check, the performance evaluation was repeated using a set of alternative risk measures in the risk-return relationship. More precisely, (1) the ΔCRs based on the baseline scenario of the stress tests, (2) the stressed capital ratios ($\widehat{CRs}$) resulting from the adverse scenario, and (3) the stressed capital ratios ($\widehat{CRs}$) resulting from the baseline scenario. As with the main analysis, each of these alternative risk measures was related to the $CARs$ over the post-event window; then the performances of the linear and quadratic fits were evaluated for each of the five samples, analogous to the procedure described above.

The performance evaluation of the linear and quadratic fits was the last step in the analysis of the *Functional Relationship Hypothesis* (Research Question 2). The results shown in Section 6.3.2 established for the first time a functional relationship between the stress test results and the corresponding abnormal stock returns of the banks subjected to the various EU-wide stress tests. The final *Intertemporal Stability Hypothesis* (Research Question 3), on the other hand, used a longitudinal approach to investigate the time-dependence of abnormal stock returns in response to the disclosure of EU-wide stress test results. The methods used to answer this research question are explained in more detail in the following section.

5.3.2.4 The Intertemporal Stability Hypothesis

The *Intertemporal Stability Hypothesis* (Research Question 3) asked whether the informational value of EU-wide stress test results had decreased over time. To answer this question, it first had to be analysed whether the $\overline{CARs}$ and $|\overline{CARs}|$ of the longitudinal sample were overall significantly different across the various EU-wide stress tests. This analysis was performed separately for each event-window type using conventional one-way repeated measures Analyses of Variance (ANOVAs) and a series of Friedman tests. These omnibus tests were followed up with multiple comparison *post hoc* tests to identify exactly which EU-wide stress tests were significantly different from each other. Based on these findings, it was determined how the informational value of EU-wide stress test results had developed over time. The above methods are explained in more detail below.

5.3.2.4.1 Data Aggregation

Prior to the analysis, the daily (absolute) abnormal returns of the longitudinal sample banks had to be aggregated into $CARs$ and $|CARs|$ in order to obtain the input data for the subsequent tests. This was done separately for each EU-wide stress test and event-window type using the same procedure as described for the cross-sectional samples in Section 5.3.2.2.1. Based on the resulting aggregates, standard descriptive statistics were calculated in order to gain insight into the basic characteristics of the longitudinal sample (Section 6.2.2). The individual $CARs$ and $|CARs|$ of the sample banks also formed the basis for identifying statistically significant differences between the means (*i.e.* the $\overline{CARs}$ and $|\overline{CARs}|$) of the various EU-wide stress tests. This is described in more detail in the next two sections.

5.3.2.4.2 Omnibus Tests

The common way to determine whether the means of the dependent variable are significantly different across three or more repeated measures of the same sample is to run a one-way repeated measures ANOVA. This approach was also followed in this study by testing whether the $\overline{CARs}$ and $|\overline{CARs}|$ of the longitudinal sample differed significantly across the five different EU-wide stress tests.

Prior to the analyses, classic assumption tests were performed. They showed that the assumptions of normality and sphericity were not met in many cases (the descriptive statistics for the longitudinal sample are shown in Section 6.2.2, for the results of the formal normality tests, see Appendix H). These assumption violations have been accounted for as described below.

In general, the ANOVA is quite robust against violations of the assumption of normality (Glass *et al.* 1972, Harwell *et al.* 1992, Schmider *et al.* 2010). Like the *t*-test (Section 5.3.2.2.3), it can therefore still be used if the dependent variable is only approximately normally distributed. However, the normality tests showed mixed results in this regard, as some of the significant distributions were clearly non-normal while others were in fact close to normal. The analyses were therefore performed using both the conventional one-way repeated measures ANOVA and its non-parametric equivalent, the one-way Friedman test, which does not require normality.[119] The use of parametric and non-parametric tests (see also Section 5.3.2.4.3) thus again represented a mutual robustness check.

The violations of the sphericity assumption, however, were addressed using the Greenhouse-Geisser correction. This is probably the most widely used approach to correcting deviations from sphericity and involves reducing the degrees of freedom of the F-distribution. The result is a more accurate p-value while maintaining the original

[119] The assumptions of the Friedman test are a subset of those of the repeated measures ANOVA. That is, the Friedman test makes similar but less or less strict assumptions. In particular, due to the ranking involved, it does not require normality or sphericity or the absence of outliers, and the dependent variable can also be ordinally scaled. For this reason, all assumptions of the Friedman test were implicitly tested with the assumption tests of the repeated measures ANOVA and therefore did not have to be tested separately.

F-statistic. The idea of the Greenhouse-Geisser correction is to compensate for the fact that the repeated measures ANOVA F-test is too liberal when sphericity is lacking. In contrast, the Friedman test does not require any correction as it uses rank data instead of the raw data and therefore does not assume sphericity.

The results of the repeated measures ANOVAs and the Friedman tests are presented in Section 6.3.3. However, since these are omnibus tests, their results were limited to showing whether there was an *overall* statistically significant difference between the different measurements. In order to identify the EU-wide stress tests whose $\overline{CARs}$ or $|\overline{CARs}|$ differed significantly from those of the other stress tests, each significant omnibus test result was followed up with multiple comparison *post hoc* tests. These are described in the next section.

5.3.2.4.3 Multiple Comparison post hoc Tests

Every significant difference found by an omnibus test was further analysed using multiple comparison *post hoc* tests. More precisely, these tests were pairwise comparisons between all possible combinations of $\overline{CARs}$ or $|\overline{CARs}|$ from the given set of EU-wide stress tests. In this way, significant differences could be identified between specific pairs of $\overline{CARs}$ or $|\overline{CARs}|$, *i.e.* between any two specific EU-wide stress tests. The results of all pairwise comparisons together gave an overall view of the economic and statistical significance of the informational value of each individual EU-wide stress test result. Building on this, it was possible to determine whether the informational value of the EU-wide stress test results was subject to a time-dependent trend.

Methodologically, each pairwise comparison consisted of a paired-sample t-test or a paired-sample Wilcoxon signed-rank test, depending on whether a repeated measures ANOVA or Friedman test was followed up. The fulfillment of the assumptions of these *post hoc* tests was already implicitly tested in the assumption tests for the repeated measures ANOVA. This is because the assumptions of the paired-sample t-test and the paired-sample Wilcoxon signed-rank test are subsets of the assumptions underlying the repeated measures ANOVA.[120] It was therefore not necessary to carry out separate assumption tests.

[120] The paired-sample t-test assumptions include a categorical independent variable with two repeated measures and a continuous dependent variable that is normally distributed and has no significant outliers. Similarly, the paired-sample Wilcoxon signed-rank test assumes a categorical independent

However, given the series of simultaneous pairwise comparisons involved in the *post hoc* analysis, the multiple comparisons problem (MCP) had to be addressed. The MCP is the increased risk of erroneously rejecting the null hypothesis due to the accumulation of alpha (type 1) error that is inherent in multiple comparisons with the same data. In other words, the family-wise error rate (FWER) across multiple comparisons always exceeds the error rate per comparison.[121] To address this problem and to reduce the risk of false positives, the Bonferroni correction was used. This simple but conservative and widely used correction method compensates for the higher FWER by adjusting the observed *p*-values. More specifically, this meant that the *p*-values obtained from the paired-sample *t*-tests and the paired-sample Wilcoxon signed-rank tests had to be multiplied by the number of multiple comparisons in order to correct the *p*-values. For completeness, both the observed and the corrected *p*-values are given in the results in Section 6.3.3.

The multiple comparison *post hoc* tests concluded the analysis of the *Intertemporal Stability Hypothesis* (Research Question 3). For the first time, the results provided insights into the time-dependency of the informational value of EU-wide stress test results. The conclusion of this investigation also represented the end of the overall analysis. Therefore, the research strategy and methods used to answer the research questions are summarised again in the next section before the results are presented in Chapter 6.

variable with two repeated measures and a dependent variable that is at least ordinally scaled. For comparison, the assumptions of the repeated measures ANOVA are listed in Section 5.3.2.4.2.

121 This can be illustrated as follows: The five EU-wide stress tests used in this study resulted in ten pairwise comparisons. Based on a significance level per comparison of $\alpha_{PC} = .05$ and $c = 10$ pairwise comparisons, the FWER would have been $\alpha_{FW} = 1 - (1 - \alpha_{PC})^c = 1 - (1 - .05)^{10} = .401$. This means that the probability of erroneously rejecting the null hypothesis at least once across the family of all pairwise comparisons would have been 40.1%, while the accepted error rate per comparison would have been only 5%.

5.3.3 Summary

In this section, the research design of this study was described. The research design comprises the research strategy and the methods used to collect and analyse the data. This study was based on a quasi-natural experimental strategy with a multiple-pretest design. This strategy was implemented through an event study, which introduced several advances and extensions to the standard event study approach developed by Campbell *et al.* (1997) and MacKinlay (1997). In particular, the multiple-pretest design of the research strategy was operationalised by introducing a systematic model selection procedure, which represents a major extension to the standard event study approach. Further important extensions include (1) the introduction of a new method for statistically determining the length of event windows, (2) the implementation of extensive confounding controls, and (3) the additional analysis of absolute abnormal returns, a non-directional measure recently proposed by Flannery *et al.* (2017).

Building on this common methodological basis, further research-question specific analyses were conducted. Various statistical methods were used to perform these analyses, including *t*-tests, Wilcoxon signed-rank tests, curve fitting, OLS regressions, error metrics, repeated measures ANOVAs, Friedman tests, and the corresponding *post hoc* tests. All of these methods were described in detail in the sections dedicated to the respective research questions.

Chapter 6
Empirical Results

6.1 Introduction

This chapter presents the empirical results of the event study and the research-question specific analyses. The starting point are the time-series aggregated results of the event study from Section 5.3.2.2.1; that is, the $CARs$ and $|CARs|$ for the various EU-wide stress tests and event-window types. Since these variables formed the dependent variables in the subsequent analyses, basic descriptive statistics of the $CARs$ and $|CARs|$ are first shown in Section 6.2; separately for the cross-sectional samples (Research Questions 1 and 2) and the longitudinal sample (Research Question 3).

The results of the main analysis are presented in Section 6.3. That is, the results of the research-question specific analyses from Sections 5.3.2.2 to 5.3.2.4 and the corresponding robustness checks. This includes the economic and statistical significance of the results and a final statement as to whether the respective null hypothesis can be rejected in favour of the alternative hypothesis. A summary of the results is given in Section 06.4.

6.2 Descriptive Statistics of the Event-Study Results

Basic descriptive statistics were calculated for the time-series aggregated results of the event study, *i.e.* for the $CARs$ and $|CARs|$ determined in Section 5.3.2.2.1. The results are shown separately for the cross-sectional samples (Research Questions 1 and 2) and the longitudinal sample (Research Question 3) and include the mean, median, standard deviation, range, skewness, and kurtosis.

The main purpose of the descriptive statistics was to describe the distributional properties of the $CARs$ and $|CARs|$ (which served as dependent variables in the subsequent analyses). Therefore, all location, scale, and shape parameters of the sample distributions are reported in order to allow comparisons with the parameters of the standard normal distribution (M = 0, SD = 1, skew = 0, kurtosis = 3). The results of formal normality tests are given in Appendix H.

6.2.1 Cross-Sectional Samples

The descriptive statistics presented below are based on the five cross-sectional samples defined in Section 5.3.2.1.2 (one sample for each of the five relevant EU-wide stress tests). The samples were constructed using census sampling; that is, all banks subjected to a given EU-wide stress test, minus the exclusions due to data unavailability or exposure to known extraneous factors (see Logic Control in Section 5.3.1.3). The resulting size of the cross-sectional samples ranged from n = 33 to n = 59 sample banks.[122] Since the descriptive statistics were calculated from two different sample distributions ($CARs$ and $|CARs|$), their results are reported in two separate tables (Table 9 and Table 10), each covering all event-window types.

Cumulative Abnormal Returns ($\boldsymbol{CARs}$)

First, Table 9 presents the descriptive statistics of the $CARs$ of the cross-sectional samples, *i.e.* their distributional properties.

[122] For a complete overview of the sample banks included in the different cross-sectional samples, see Appendix D.

Table 9

Descriptive Statistics of Cumulative Abnormal Returns ($CARs$) *Based on the Cross-Sectional Samples*

Sample	*n*	*M*	*Mdn*	*SD*	Range	Skew	Kurtosis
Pre-Event Window $(-2, 0)$							
CEBS 2010	50	-0.93	-0 97	2.43	16.61	0 19	3.97
EBA 2011	51	-1.50	-1 33	4.04	26.33	-2.36	11.59
EBA 2014	59	0 92	0.73	2.99	23.22	-2.17	14.68
EBA 2016	34	-0.42	0.00	2.39	13.49	-0.37	2.88
EBA 2018	33	1 96	2.51	2.88	15.33	-0.56	1.90
Standard Event Window $(-2, +2)$							
CEBS 2010	50	2 54	1.81	5.60	27.31	1 14	1.75
EBA 2011	51	-0.80	-1 59	6.29	49.92	2 50	14.97
EBA 2014	59	0.47	0.83	4.52	24.78	-1.04	3.35
EBA 2016	34	-1.06	-0.46	3.60	21.74	-0.34	4.37
EBA 2018	33	1 52	1.61	2.83	14.47	-0.52	1.42
Post-Event Window $(+1, n)$							
CEBS 2010	50	1 56	0.90	2.75	17.50	3.00	12.78
EBA 2011	51	0 29	-0 32	4.06	24.25	2.01	7.55
EBA 2014	59	0.44	0.52	4.58	35.41	-1.03	8.92
EBA 2016	34	-0.85	-0 31	2.37	13.29	-1.68	5.21
EBA 2018	33	-0.51	-0 52	1.56	9.50	-1.26	5.75

Note. This table shows the descriptive statistics of the cumulative abnormal returns ($CARs$) of the five cross-sectional samples for each event-window type. The parentheses behind the event-window types specify the number of trading days before and after the event date t_0. The term n in the definition of the post-event window $(+1, n)$ refers to a statistical determination of variable event window lengths based on the Ljung-Box (1978) test introduced in this study. Based on census sampling, the cross-sectional samples consisted of all banks subjected to a given EU-wide stress test minus the exclusions due to unavailability of data or exposure to known extraneous factors, resulting in variable sample sizes ranging from $n = 33$ to $n = 59$ sample banks. The normal returns used to calculate the $CARs$ were estimated using the Fama and French (1993) Three-Factor Model. n = sample size. *M* = mean. *Mdn* = median. *SD* = standard deviation. CEBS = Committee of European Banking Supervisors. EBA = European Banking Authority.

The descriptive statistics in Table 9 show that the mean and median were non-zero in almost all cases, regardless of the event-window type. That is, the center of the distribution was typically shifted from zero, with left and right shifts occurring approximately equally often. Thus, mean and median typically deviated from the location parameter value (zero) of the standard normal distribution.

Standard deviation (SD > 1) and range indicate that the samples were, in most cases, highly dispersed. This is consistent with the (strongly) leptokurtic shape of most sample distributions (kurtosis > 3), suggesting the presence of outliers. Conversely, only a few of the sample distributions showed a platykurtic (kurtosis < 3) or mesokurtic (kurtosis ≈ 3) shape.

Most of the sample distributions were highly skewed (skew $\geq |1|$), with the majority of them skewed to the left across all event-window types. Exceptions were some slightly (skew $\leq |0.5|$) to moderately ($|0.5| <$ skew $< |1|$) skewed sample distributions over the pre-event and the standard event window, which tended to be associated with mesokurtic or platykurtic shapes.

Overall, the descriptive statistics in Table 9 suggested that the $CARs$ of the cross-sectional samples were generally not normally distributed. This was confirmed by the results of the Lilliefors test and the Shapiro-Wilk test in Appendix H.1.

Absolute Cumulative Abnormal Returns ($|CARs|$)

The description of the $CARs$ is followed by a description of the non-directional (absolute value-based) $|CARs|$ of the cross-sectional samples. Table 10 shows the corresponding descriptive statistics.

Table 10
Descriptive Statistics of Absolute Cumulative Abnormal Returns ($|CARs|$) *Based on the Cross-Sectional Samples*

Sample	*n*	*M*	*Mdn*	*SD*	Range	Skew	Kurtosis
Pre-Event Window $(-2, 0)$							
CEBS 2010	50	1.89	1.61	1.77	8.75	2 10	5.68
EBA 2011	51	2.85	2.62	3.21	21.64	4 16	23.81
EBA 2014	59	1.84	0.94	2.53	15.22	3 27	13.60
EBA 2016	34	1.64	1.11	1.77	7.82	1 91	3.96
EBA 2018	33	2.69	2.65	2.19	8.20	0.76	-0.60
Standard Event Window $(-2, +2)$							
CEBS 2010	50	4 36	2.63	4.32	20.36	2.00	4.46
EBA 2011	51	3.69	2.29	5.12	31.85	3 93	19.03
EBA 2014	59	3 16	1.87	3.24	15.14	2.00	4.27
EBA 2016	34	2 27	1.18	2.97	11.44	2 10	3.72
EBA 2018	33	2 52	1.83	1.96	7.43	0 96	0.24
Post-Event Window $(+1, n)$							
CEBS 2010	50	1.87	1.06	2.54	15.39	3 59	16.42
EBA 2011	51	2.78	2.15	2.95	18.73	3 39	16.90
EBA 2014	59	2 57	1.03	3.80	18.16	3.02	9.54
EBA 2016	34	1.62	1.09	1.91	9.66	2.69	9.44
EBA 2018	33	1.06	0.56	1.25	6.33	2.68	9.37

Note. This table shows the descriptive statistics of the absolute cumulative abnormal returns ($|CARs|$) of the five cross-sectional samples for each event-window type. The parentheses behind the event-window types specify the number of trading days before and after the event date t_0. The term *n* in the definition of the post-event window $(+1, n)$ refers to a statistical determination of variable event window lengths based on the Ljung-Box (1978) test introduced in this study. Based on census sampling, the cross-sectional samples consisted of all banks subjected to a given EU-wide stress test minus the exclusions due to unavailability of data or exposure to known extraneous factors, resulting in variable sample sizes ranging from $n = 33$ to $n = 59$ sample banks. The normal returns used to calculate the $|CARs|$ were estimated using the Fama and French (1993) Three-Factor Model. n = sample size. *M* = mean. *Mdn* = median. *SD* = standard deviation. CEBS = Committee of European Banking Supervisors. EBA = European Banking Authority.

Since $|CARs|$ are an absolute value measure, all means and medians shown in Table 10 were necessarily shifted to the right. However, the magnitude of these shifts did not generally exceed the average absolute estimation error over the corresponding estimation period used as the benchmark for the significance tests (Table 16).

The standard deviation and range show that the samples were moderately to highly dispersed across all event-window types. Similar to the $CARs$, most distributions of $|CARs|$ had a (strongly) leptokurtic shape. The only platykurtic distributions occurred in the EBA 2018 over the pre-event and the standard event window, while the only mesokurtic distribution was arguably in the EBA 2016 over the standard event window (kurtosis = 3.72).

All of the sample distributions were skewed to the right, with almost all of them being highly skewed. The only exceptions occurred again in the EBA 2018, whose distributions over the pre-event and the standard event window were only moderately skewed to the right.

Similar to the $CARs$, the descriptive statistics in Table 10 also indicated for the $|CARs|$ that the cross-sectional samples were generally not normally distributed. This was formally confirmed by the results of the Lilliefors test and the Shapiro-Wilk test in Appendix H.2.

6.2.2 Longitudinal Sample

Since the size and composition of the longitudinal sample differed from those of the cross-sectional samples, the descriptive statistics had to be recalculated for this sample. Like the cross-sectional samples, the longitudinal sample was constructed using census sampling, but consisted only of those banks subject to *all* five EU-wide stress tests examined, minus the exclusions due to unavailability of data or exposure to known extraneous factors (see Logic Control in Section 5.3.1.3). The resulting size of the longitudinal sample was n = 28 sample banks. A more detailed definition of the construction of the longitudinal sample can be found in Section 5.3.2.1.2.[123] The results of the descriptive statistics are again reported separately for the two different sample distributions ($CARs$ and $|CARs|$), for all three event-window types.

[123] For a complete overview of the sample banks included in the longitudinal sample, see Appendix E.

Cumulative Abnormal Returns ($CARs$)

As above, the descriptive statistics of the $CARs$ of the longitudinal sample are presented first. Table 11 shows the corresponding descriptive statistics.

Table 11
Descriptive Statistics of Cumulative Abnormal Returns ($CARs$) Based on the Longitudinal Sample

Sample	*n*	*M*	*Mdn*	*SD*	Range	Skew	Kurtosis
Pre-Event Window $(-2, 0)$							
CEBS 2010	28	-0.44	-0 27	1.71	6.06	-0.16	-0.86
EBA 2011	28	-0.85	-0.74	2.36	10.37	-0.09	0.01
EBA 2014	28	0.63	0.66	1.36	7.18	-0.11	1.82
EBA 2016	28	-0.68	-0 31	2.31	12.47	-0.86	2.93
EBA 2018	28	1 98	2.58	3.06	15.33	-0.59	1.68
Standard Event Window $(-2, +2)$							
CEBS 2010	28	1.80	2.04	4.10	16.15	-0.03	-0.38
EBA 2011	28	-1.48	-1.66	3.06	15.99	-0.31	1.83
EBA 2014	28	0 27	0.45	2.92	16.51	0.62	3.58
EBA 2016	28	-0.87	-0 28	3.90	21.74	-0.47	3.82
EBA 2018	28	1.43	1.59	2.96	14.47	-0.44	1.33
Post-Event Window $(+1, n)$							
CEBS 2010	28	1.08	0.95	1.71	9.23	1.43	4.94
EBA 2011	28	-0.73	-1 31	3.06	12.49	0.86	0.44
EBA 2014	28	0.41	0.27	1.54	8.76	1.02	4.24
EBA 2016	28	-0.94	-0 31	2.41	11.53	-2.05	5.86
EBA 2018	28	-0.57	-0 54	1.60	9.50	-1.25	6.30

Note. This table shows the descriptive statistics of the cumulative abnormal returns ($CARs$) of the longitudinal sample for each event-window type. The parentheses behind the event-window types specify the number of trading days before and after the event date t_0. The term *n* in the definition of the post-event window $(+1, n)$ refers to a statistical determination of variable event window lengths based on the Ljung-Box (1978) test introduced in this study. Based on census sampling, the longitudinal sample consisted of only those banks that were the subject of all five EU-wide stress tests examined, minus the exclusions due to unavailability of data or exposure to known extraneous factors, resulting in a sample size of *n* = 28 sample banks. The normal returns used to calculate the $CARs$ were estimated using the Fama and French (1993) Three-Factor Model. *n* = sample size. *M* = mean. *Mdn* = median. *SD* = standard deviation. CEBS = Committee of European Banking Supervisors. EBA = European Banking Authority.

The means and medians in Table 11 are similar to those reported in Table 9 for the cross-sectional samples. That is, they were generally shifted from zero across all event-window types, with left and right shifts occurring about equally often. In addition, all means and medians of the longitudinal sample had the same sign as those of the cross-sectional samples (except for EBA 2011 over the post-event window and EBA 2016 over the pre-event window). However, the distribution shifts of the longitudinal sample tended to be smaller than those of the cross-sectional samples.

The standard deviation shows that the $CARs$ of the longitudinal sample (like those of the cross-sectional samples) were in all cases more dispersed than the standard normal distribution. However, the longitudinal sample was relatively less dispersed and therefore closer to the standard normal distribution than the cross-sectional samples. The shape of the $CARs$ of the longitudinal sample was mostly leptokurtic over the post-event window, while in most other cases it was (strongly) platykurtic.

Similar to the cross-sectional samples, most longitudinal sample distributions were skewed to the left (notably, however, they were mostly skewed to the right over the post-event window). In addition, the distributions over the post-event window were generally highly skewed in both the longitudinal and cross-sectional samples. However, the longitudinal sample was overall less skewed than the cross-sectional samples, especially over the pre-event and the standard event window, where the distributions were only slightly to moderately skewed.

In summary, the descriptive statistics in Table 11 showed that the $CARs$ of the longitudinal sample approximated the standard normal distribution reasonably well. This conclusion was formally confirmed by the Lilliefors test and the Shapiro-Wilk test in Appendix H.3, which found that the $CARs$ of the longitudinal sample were normally distributed in about half of the cases. This also implies that the $CARs$ of the longitudinal sample approximated the standard normal distribution better than those of the cross-sectional samples.

Absolute Cumulative Abnormal Returns ($|CARs|$)

Finally, Table 12 presents the descriptive statistics of the non-directional (absolute value-based) $|CARs|$ of the longitudinal sample.

Table 12

Descriptive Statistics of Absolute Cumulative Abnormal Returns ($|CARs|$) *Based on the Longitudinal Sample*

Sample	*n*	*M*	*Mdn*	*SD*	Range	Skew	Kurtosis
Pre-Event Window $(-2, 0)$							
CEBS 2010	28	1.43	1.41	1.00	3.63	0.49	-0.03
EBA 2011	28	1 94	1.50	1.57	5.99	0 94	0.24
EBA 2014	28	1 13	0.80	0.96	4.06	1.46	2.24
EBA 2016	28	1.66	1.27	1.72	7.82	2 13	5.48
EBA 2018	28	2.80	2.67	2.30	8.20	0.67	-0 30
Standard Event Window $(-2, +2)$							
CEBS 2010	28	3.62	2.73	2.58	9.20	0.86	-0 28
EBA 2011	28	2.46	1.90	2.31	9.60	1 56	2.39
EBA 2014	28	2.03	1.42	2.08	9.19	2 36	5.91
EBA 2016	28	2 33	1.08	3.22	11.44	2.01	2.96
EBA 2018	28	2 50	1.74	2.10	7.43	0 97	-0.01
Post-Event Window $(+1, n)$							
CEBS 2010	28	1.43	1.10	1.42	7.12	2 51	9.20
EBA 2011	28	2 53	2.37	1.82	6.83	0.62	-0.26
EBA 2014	28	1.06	0.70	1.18	5.46	2.47	6.93
EBA 2016	28	1.62	0.93	2.00	9.64	2.87	10.06
EBA 2018	28	1.09	0.59	1.29	6.31	2.78	9.72

Note. This table shows the descriptive statistics of the absolute cumulative abnormal returns ($|CARs|$) of the longitudinal sample for each event-window type. The parentheses behind the event-window types specify the number of trading days before and after the event date t_0. The term *n* in the definition of the post-event window $(+1, n)$ refers to a statistical determination of variable event window lengths based on the Ljung-Box (1978) test introduced in this study. Based on census sampling, the longitudinal sample consisted of only those banks that were the subject of all five EU-wide stress tests examined, minus the exclusions due to unavailability of data or exposure to known extraneous factors, resulting in a sample size of *n* = 28 sample banks. The normal returns used to calculate the $|CARs|$ were estimated using the Fama and French (1993) Three-Factor Model. *n* = sample size. *M* = mean. *Mdn* = median. *SD* = standard deviation. CEBS = Committee of European Banking Supervisors. EBA = European Banking Authority.

All means and medians in Table 12 were necessarily shifted to the right due to the absolute value of the input data used to calculate the $|CARs|$. In most cases, however, the magnitude of these shifts did not exceed the average absolute estimation error over the corresponding estimation period.[124]

[124] The average absolute estimation error over the estimation periods ranged from M = 0.92 to M = 7.03 and from Mdn = 0.80 to Mdn = 6.09. It should be noted, however, that these errors were not used as a benchmark in the further analysis of the longitudinal sample (*i.e.* in the investigation of Research Question 3). This is in contrast to the average absolute estimation errors in the cross-sectional samples, which served as a benchmark in the analysis of Research Question 1.

The standard deviation and range show that the $|CARs|$ of the longitudinal sample were consistently less dispersed than the $CARs$ of the same sample across all event-window types. In general, the standard deviation of the $|CARs|$ was rather moderate and came relatively close to that of the standard normal distribution (with the exception of the standard event window and a few exceptions in the other event windows). The distributional shape of the $|CARs|$ largely resembled that of the $CARs$. That is, in most cases the distribution was (strongly) platykurtic, while over the post-event window it was mostly (strongly) leptokurtic.

All sample distributions were skewed to the right, with most of them being highly skewed. However, distributions with moderate skewness were also found in each of the event-window types. The only slightly skewed distribution occurred in the CEBS 2010 over the pre-event window (skew = 0.49). Notably, all slightly and moderately skewed distributions were associated with strongly platykurtic shapes.

Overall, the descriptive statistics in Table 12 indicated that the $|CARs|$ of the longitudinal sample were generally not normally distributed. This was confirmed by the formal results of the Lilliefors test and the Shapiro-Wilk test in Appendix H.4.

6.3 Research-Question Specific Results

In this section, the results of the research-question specific analyzes are presented, which were methodically described in Sections 5.3.2.2 to 5.3.2.4. A common feature of these analyses is that they built on the time-series aggregated results of the preceding event study. That is, their dependent variables were the $CARs$ and $|CARs|$ determined in Section 5.3.2.2.1 (see also the descriptive statistics above). Identified outliers were deliberately retained in the analyses as they did not represent errors or invalid data but natural variability in measuring abnormal stock returns and should thus represent true and meaningful values. It should also be mentioned in this context that the non-parametric tests used are robust to all types of outliers.

The results of the research-question specific analyses are presented separately in Sections 6.3.1 to 6.3.3. In addition to the economic and statistical significance of the results, this also includes a final statement as to whether the respective null hypothesis could be rejected in favour of the alternative hypothesis. Furthermore, the results of the robustness checks are provided in the relevant sections. At the beginning of each section, the respective research question is repeated to facilitate coherent reading.

6.3.1 The Informational Value Hypothesis

The specific research question to be answered based on the following results was: *what is the average value of the information contained in the results of EU-wide stress tests measured in terms of abnormal stock returns?* This question was examined separately for each of the relevant five EU-wide stress tests using cross-sectional samples. The samples consisted of all banks subjected to a given EU-wide stress test, minus the exclusions due to unavailability of data or exposure to known extraneous factors (see Logic Control in Section 5.3.1.3), and ranged from n = 33 to n = 59 sample banks. The results are reported separately for the directional $\overline{CARs}$ and the non-directional $|\overline{CARs}|$, including the corresponding robustness checks.

6.3.1.1 Average Cumulative Abnormal Returns ($\overline{CARs}$)

A series of one-sample t-tests and one-sample Wilcoxon signed-rank tests were used to test whether the $\overline{CAR}$ of each sample was significantly different from zero. That is, whether the results of EU-wide stress tests had sufficient informational value to cause statistically significant reactions in the average stock price. Depending on the significance test used, the sample averages were calculated using the mean (t-test) or the median (Wilcoxon signed-rank test). The normal returns used to calculate the underlying $CARs$ of each sample bank were estimated using the Fama and French (1993) Three-Factor Model. Table 13 shows the $\overline{CARs}$ and significance test results across all samples and event-window types.

Table 13

Informational Value of the EU-Wide Stress Test Results Measured in Average Cumulative Abnormal Returns ($\overline{CAR}s$)

			$\overline{CAR}s$		Significance	
Sample	***n***	***df***	***M***	***Mdn***	***t***	***T***
Pre-Event Window $(-2, 0)$						
CEBS 2010	50	49	-0.93	-0.97	-2.693**	-2.978***
EBA 2011	51	50	-1.50	-1.33	-2.648**	-2.700***
EBA 2014	59	58	0.92	0.73	2.356**	3.842***
EBA 2016	34	33	-0.42	0.00	-1.025	-0.983
EBA 2018	33	32	1.96	2 51	3.894***	3.511***
Standard Event Window $(-2, +2)$						
CEBS 2010	50	49	2.54	1.81	3.210***	2.775***
EBA 2011	51	50	-0.80	-1.59	-0.905	-2.278**
EBA 2014	59	58	0.47	0.83	0.798	1.487
EBA 2016	34	33	-1.06	-0.46	-1.714*	-1.958*
EBA 2018	33	32	1.52	1.61	3.088***	3.064***
Post-Event Window $(+1, n)$						
CEBS 2010	50	49	1.56	0 90	4.005***	4.436***
EBA 2011	51	50	0.29	-0.32	0.503	-0.291
EBA 2014	59	58	0.44	0 52	0.730	2.023**
EBA 2016	34	33	-0.85	-0.31	-2.091**	-1.906*
EBA 2018	33	32	-0.51	-0.52	-1.885*	-2.296**

Note. This table shows the average cumulative abnormal returns ($\overline{CAR}s$) and significance test results of the five cross-sectional samples for each event-window type. The parentheses behind the event-window types specify the number of trading days before and after the event date t_0. The term n in the definition of the post-event window $(+1, n)$ refers to a statistical determination of variable event window lengths based on the Ljung-Box (1978) test introduced in this study. The normal returns used to calculate the underlying cumulative abnormal returns ($CARs$) for each individual sample bank were estimated using the Fama and French (1993) Three-Factor Model. *n* = sample size. *df* = degrees of freedom. *M* = mean. *Mdn* = median. *t* = *t*-test statistic. *T* = Wilcoxon test statistic. CEBS = Committee of European Banking Supervisors. EBA = European Banking Authority.

* $p < .10$. ** $p < .05$. *** $p < .01$.

The $\overline{CAR}s$ in Table 13 were generally different from zero across all samples and event-window types (both mean and median-based). In most cases these differences were economically significant. Statistically, however, only about half (t-test) to two thirds (Wilcoxon signed-rank test) of all cases were significant at the .05 level or better. The highest overall significance was found for the pre-event window, while the significance of the standard event window and the post-event window were lower and approximately the same. Notably, only the $\overline{CAR}s$ of the CEBS 2010 and the EBA 2018 were statistically significant across all event-window types (although the $\overline{CAR}s$ of the EBA 2018 over the post-event window were only significant at the .10 level according to the t-test). The results of the two significance tests were fairly consistent, showing that they were robust across parametric and non-parametric approaches (despite some inconsistencies in the post-event window).

Overall, the results provided evidence against the null hypothesis and for the alternative hypothesis, in particular for CEBS 2010 and EBA 2018 as well as for the pre-event window. That is, in most cases the average informational value contained in the results of EU-wide stress tests was both economically and statistically significant. Table 14 summarises the hypothesis testing results by detailing each acceptance and rejection decision.

Table 14
Summary of Hypothesis Testing Results at the .05 Significance Level Based on Average Cumulative Abnormal Returns ($\overline{CARs}$)

Sample	*n*	Null Hypotesis	Alternative Hypothesis
Pre-Event Window $(-2, 0)$			
CEBS 2010	50	Rejected	Accepted
EBA 2011	51	Rejected	Accepted
EBA 2014	59	Rejected	Accepted
EBA 2016	34	Accepted	Rejected
EBA 2018	33	Rejected	Accepted
Standard Event Window $(-2, +2)$			
CEBS 2010	50	Rejected	Accepted
EBA 2011	51	**Rejected**	**Accepted**
EBA 2014	59	Accepted	Rejected
EBA 2016	34	Accepted	Rejected
EBA 2018	33	Rejected	Accepted
Post-Event Window $(+1, n)$			
CEBS 2010	50	Rejected	Accepted
EBA 2011	51	Accepted	Rejected
EBA 2014	59	**Rejected**	**Accepted**
EBA 2016	34	**Rejected**	**Accepted**
EBA 2018	33	**Rejected**	**Accepted**

Note. This table summarises the acceptance and rejection of the relevant hypotheses at the .05 significance level or better for the analyses performed on the average cumulative abnormal returns ($\overline{CARs}$) of the five cross-sectional samples for each event-window type. The parentheses behind the event-window types specify the number of trading days before and after the event date t_0. The term *n* in the definition of the post-event window $(+1, n)$ refers to a statistical determination of variable event window lengths based on the Ljung-Box (1978) test introduced in this study. In case of discrepancy between the *t*-test and the Wilcoxon signed-rank test, the more significant test result is shown. Statements in bold indicate discrepancies between the two tests. *n* = sample size. CEBS = Committee of European Banking Supervisors. EBA = European Banking Authority.

As a robustness check, the analysis was repeated with $\overline{CARs}$ based on normal (expected) returns obtained from the Market Model. The Market Model was by far the most frequently used normal return-generating model in previous studies (Table 3). In addition, the Market Model performed reasonably well in the goodness-of-fit tests used for model selection (Appendix G). Apart from the different normal returns, the $\overline{CARs}$ in the robustness check were calculated in the same way as in the main analysis. In addition, the same methods were used in both cases to test for significance. Table 15 presents the results of the robustness check.

Table 15
Robustness Check of the Specification of the Normal Return-Generating Model Used to Calculate the Average Cumulative Abnormal Returns ($\overline{CARs}$)

			$\overline{CARs}$		Significance	
Sample	***n***	***df***	***M***	***Mdn***	***t***	***T***
Pre-Event Window $(-2, 0)$						
CEBS 2010	50	49	-1.36	-1.43	-4.280***	-4.175***
EBA 2011	51	50	-1.41	-1.15	-2.371**	-2.465**
EBA 2014	59	58	1.14	0.79	2.856***	4.212***
EBA 2016	34	33	0.45	0.14	0.973	0.983
EBA 2018	33	32	1.87	2.30	3.694***	3.386***
Standard Event Window $(-2, +2)$						
CEBS 2010	50	49	5.08	3.33	6.056***	5.343***
EBA 2011	51	50	-0.71	-0.70	-0.743	-1.987**
EBA 2014	59	58	0.09	0.35	0.168	0.785
EBA 2016	34	33	-1.49	-0.82	-2.203**	-2.556**
EBA 2018	33	32	1.79	1.82	3.424***	2.993***
Post-Event Window $(+1, n)$						
CEBS 2010	50	49	2.43	1.87	5.759***	5.565***
EBA 2011	51	50	0.25	-0.65	0.442	-0.300
EBA 2014	59	58	-0.49	-0.16	-0.835	-0.747
EBA 2016	34	33	-1.53	-0.82	-3.404***	-3.120***
EBA 2018	33	32	0.02	-0.04	0.124	0.080

Note. This table shows the results of a robustness check performed to check the robustness of the results from the main analysis against an alternative specification of the normal return-generating model. Specifically, the table shows the average cumulative abnormal returns ($\overline{CARs}$) and significance test results of the five cross-sectional samples for each event-window type, with the normal returns used to calculate the underlying cumulative abnormal returns ($CARs$) being estimated using the Market Model. The parentheses behind the event-window types specify the number of trading days before and after the event date t_0. The term *n* in the definition of the post-event window $(+1, n)$ refers to a statistical determination of variable event window lengths based on the Ljung-Box (1978) test introduced in this study. *n* = sample size. *df* = degrees of freedom. *M* = mean. *Mdn* = median. *t* = *t*-test statistic. *T* = Wilcoxon test statistic. CEBS = Committee of European Banking Supervisors. EBA = European Banking Authority.
* $p < .10$. ** $p < .05$. *** $p < .01$.

Overall, the robustness check produced qualitatively similar results to the main analysis. However, it also revealed significant differences in individual cases. A comparison of Table 13 and Table 15 shows that the $\overline{CARs}$ in the vast majority of cases had the same sign and were generally similar in size. In two cases, however, the robustness check and the main analysis yielded mean-based $\overline{CARs}$ of almost identical size, but with opposite signs (EBA 2014 over the post-event window and EBA 2016 over the pre-event window). Similarly, the size of the $\overline{CARs}$ of the CEBS 2010 differed significantly over the standard event window. Specifically, the result of the robustness check exceeded that of the main analysis by more than 2.5 percentage points in the mean and by more than 1.5 percentage points in the median. The results of the significance tests were largely consistent between the robustness check and the main analysis, with the most important differences occurring in the post-event window. In summary, while the evidence obtained is generally robust, it can also be sensitive to the specification of the normal return-generating model, leading to significantly different results.

6.3.1.2 Average Absolute Cumulative Abnormal Returns ($|\overline{CARs}|$)

As with the $\overline{CARs}$, one-sample *t*-tests and one-sample Wilcoxon signed-rank tests were used to test the statistical significance of the $|\overline{CAR}|$ of each sample. Due to their absolute value, however, the $|\overline{CARs}|$ were not compared with zero, but with their average value over the corresponding estimation period ($|\gamma|$), as explained in Section 5.3.2.2.3. This means that a $|\overline{CAR}|$ was considered significant if it differed significantly from the average absolute estimation error of the normal return-generating model. The model used to estimate normal returns was the Fama and French (1993) Three-Factor Model. Table 16 presents the $|\overline{CARs}|$, the corresponding $|\gamma|$, and the results of the significance tests for all samples and event-window types.

Table 16

Informational Value of the EU-Wide Stress Test Results Measured in Average Absolute Cumulative Abnormal Returns ($|\overline{CAR}s|$)

Sample	*n*	*df*	$\lvert\overline{CAR}s\rvert$ *M*	$\lvert\overline{CAR}s\rvert$ *Mdn*	$\lvert\gamma\rvert$ *M*	$\lvert\gamma\rvert$ *Mdn*	Significance *t*	Significance *T*
Pre-Event Window $(-2, 0)$								
CEBS 2010	50	49	1.89	1.61	4.56	3.80	-10.663***	-5.063***
EBA 2011	51	50	2.85	2.62	4.46	3.03	-3 580***	-1 922*
EBA 2014	59	58	1.84	0 94	2.52	1.99	-2.073**	-2.815***
EBA 2016	34	33	1.64	1 11	4.49	3.72	-9.409***	-4.248***
EBA 2018	33	32	2.69	2.65	2.75	2.56	-0.157	-0.045
Standard Event Window $(-2, +2)$								
CEBS 2010	50	49	4.36	2.63	7.74	6.41	-5 544***	-3.828***
EBA 2011	51	50	3.69	2 29	7.60	5.11	-5.443***	-3.946***
EBA 2014	59	58	3.16	1.87	5.12	4.01	-4.642***	-3.004***
EBA 2016	34	33	2.27	1 18	7.63	6.27	-10.517***	-4.539***
EBA 2018	33	32	2.52	1.83	4.63	4.31	-6 187***	-3.922***
Post-Event Window $(+1, n)$								
CEBS 2010	50	49	1.87	1.06	1.72	1.32	0.423	0.169
EBA 2011	51	50	2.78	2 15	2.04	1.26	1.780*	3.646***
EBA 2014	59	58	2.57	1.03	1.80	1.09	1.563	1.925*
EBA 2016	34	33	1.62	1.09	2.27	1.58	-1.977*	-1 103
EBA 2018	33	32	1.06	0 56	1.13	0.93	-0.327	-0.652

Note. This table shows the average absolute cumulative abnormal returns ($|\overline{CAR}s|$), the average absolute estimation errors ($|\gamma|$) of the normal return-generating model, and the significance test results of the five cross-sectional samples for each event-window type. The parentheses behind the event-window types specify the number of trading days before and after the event date t_0. The term *n* in the definition of the post-event window $(+1, n)$ refers to a statistical determination of variable event window lengths based on the Ljung-Box (1978) test introduced in this study. The normal returns used to calculate the underlying absolute cumulative abnormal returns ($|CARs|$) for each individual sample bank were estimated using the Fama and French (1993) Three-Factor Model. *n* = sample size. *df* = degrees of freedom. *M* = mean. *Mdn* = median. $|\gamma|$ = average absolute estimation error. *t* = *t*-test statistic. *T* = Wilcoxon test statistic. CEBS = Committee of European Banking Supervisors. EBA = European Banking Authority.
* $p < .10$. ** $p < .05$. *** $p < .01$.

The $|\overline{CAR}s|$ shown in Table 16 cannot be interpreted in an economically meaningful way (due to the underlying absolute values). From their magnitude, however, it can be seen that the absolute sum of positive and negative information effects tended to be relatively high, which is generally indicative of a high informational value. This was basically true for all samples and event-window types. However, in almost none of the cases did the $|\overline{CAR}|$ exceed the benchmark of the average absolute estimation error. The only two exceptions were the $|\overline{CAR}s|$ of the CEBS 2010 over the post-event window (mean-based) and the EBA 2018 over the pre-event window (median-based). Notably, almost all $|\overline{CAR}s|$ were found to be statistically significant at the .01 level in the pre-event and standard event windows. The $|\overline{CAR}s|$ in the post-event window, on the other hand, were hardly significant (with the exception of the EBA 2011 according to the Wilcoxon signed-rank test). In general, the results of the two significance tests were quite consistent and therefore robust.

Overall, the results provided mixed evidence for the different event-window types. For the pre-event and standard event windows, the null hypothesis was almost universally rejected in favour of the alternative hypothesis. In contrast, the null hypothesis could generally not be rejected for the post-event window (except in one single case). Table 17 provides a detailed summary of the acceptance and rejection decisions.

Table 17
Summary of Hypothesis Testing Results at the .05 Significance Level Based on Average Absolute Cumulative Abnormal Returns ($|\overline{CARs}|$)

Sample	*n*	Null Hypotesis	Alternative Hypothesis
Pre-Event Window $(-2, 0)$			
CEBS 2010	50	Rejected	Accepted
EBA 2011	51	**Rejected**	**Accepted**
EBA 2014	59	Rejected	Accepted
EBA 2016	34	Rejected	Accepted
EBA 2018	33	Accepted	Rejected
Standard Event Window $(-2, +2)$			
CEBS 2010	50	Rejected	Accepted
EBA 2011	51	Rejected	Accepted
EBA 2014	59	Rejected	Accepted
EBA 2016	34	Rejected	Accepted
EBA 2018	33	Rejected	Accepted
Post-Event Window $(+1, n)$			
CEBS 2010	50	Accepted	Rejected
EBA 2011	51	**Rejected**	**Accepted**
EBA 2014	59	Accepted	Rejected
EBA 2016	34	Accepted	Rejected
EBA 2018	33	Accepted	Rejected

Note. This table summarises the acceptance and rejection of the relevant hypotheses at the .05 significance level or better for the analyses performed on the average absolute cumulative abnormal returns ($|\overline{CARs}|$) of the five cross-sectional samples for each event-window type. The parentheses behind the event-window types specify the number of trading days before and after the event date t_0. The term *n* in the definition of the post-event window $(+1, n)$ refers to a statistical determination of variable event window lengths based on the Ljung-Box (1978) test introduced in this study. In case of discrepancy between the *t*-test and the Wilcoxon signed-rank test, the more significant test result is shown. Statements in bold indicate discrepancies between the two tests. *n* = sample size. CEBS = Committee of European Banking Supervisors. EBA = European Banking Authority.

Analogously to the $\overline{CARs}$, the robustness of the results obtained for the $|\overline{CARs}|$ was checked against an alternative specification of the normal return-generating model using the Market Model. Table 18 presents the results of the robustness check.

Table 18

Robustness Check of the Specification of the Normal Return-Generating Model Used to Calculate the Average Absolute Cumulative Abnormal Returns ($|\overline{CAR}s|$)

			$\|\overline{CAR}s\|$		$\|\gamma\|$		Significance	
Sample	***n***	***df***	***M***	***Mdn***	***M***	***Mdn***	***t***	***T***
Pre-Event Window $(-2, 0)$								
CEBS 2010	50	49	1.93	1 58	4.55	3.89	-10.451***	-5.237***
EBA 2011	51	50	3.02	2.46	4.66	3.57	-3 546***	-2.709***
EBA 2014	59	58	2.02	1.08	2.65	2.13	-1.867*	-2.370**
EBA 2016	34	33	1.91	1 59	4.74	3.92	-8.463***	-3.992***
EBA 2018	33	32	2.69	2 37	3.08	3.00	-1.054	-1 331
Standard Event Window $(-2, +2)$								
CEBS 2010	50	49	5.49	3 37	7.72	6.56	-2.845***	-1.974**
EBA 2011	51	50	3.99	2.63	7.96	6.02	-5.092***	-4.518***
EBA 2014	59	58	3.02	1 98	5.38	4.31	-5.876***	-3.465***
EBA 2016	34	33	2.70	1 55	8.06	6.62	-9.732***	-4.334***
EBA 2018	33	32	2.73	1 99	5.19	5.04	-6 551***	-4.279***
Post-Event Window $(+1, n)$								
CEBS 2010	50	49	2.52	1.87	1.72	1.32	1.932*	2.563**
EBA 2011	51	50	2.76	2 10	2.12	1.37	1.614	3.403***
EBA 2014	59	58	2.55	1.44	1.88	1.11	1.385	2.755***
EBA 2016	34	33	2.05	1.73	2.42	1.63	-0.970	-0 231
EBA 2018	33	32	0.80	0 55	1.26	1.00	-3 240***	-2.278**

Note. This table shows the results of a robustness check performed to check the robustness of the results from the main analysis against an alternative specification of the normal return-generating model. Specifically, the table shows the average absolute cumulative abnormal returns ($|\overline{CAR}s|$), the average absolute estimation errors ($|\gamma|$) of the normal return-generating model, and the significance test results of the five cross-sectional samples for each event-window type, with the normal returns used to calculate the underlying absolute cumulative abnormal returns ($|CARs|$) being estimated using the Market Model. The parentheses behind the event-window types specify the number of trading days before and after the event date t_0. The term *n* in the definition of the post-event window $(+1, n)$ refers to a statistical determination of variable event window lengths based on the Ljung-Box (1978) test introduced in this study. *n* = sample size. *df* = degrees of freedom. *M* = mean. *Mdn* = median. $|\gamma|$ = average absolute estimation error. *t* = *t*-test statistic. *T* = Wilcoxon test statistic. CEBS = Committee of European Banking Supervisors. EBA = European Banking Authority.

* $p < .10$. ** $p < .05$. *** $p < .01$.

A comparison of Table 16 and Table 18 shows that the robustness check generally yielded results qualitatively similar to those of the main analysis. That is, the $|\overline{CAR}s|$ and $|\gamma|$ of the robustness check and the main analysis showed quite similar magnitudes overall. In individual cases, however, there were again significant differences. This is particularly true for the $|\overline{CAR}s|$ of the CEBS 2010 over the standard and post-event windows and for the median-based $|\gamma|$ of the EBA 2011 over the standard event window. The results of the significance tests were largely consistent between the robustness check and the main analysis. However, there were again important differences, particularly in the post-event window. Overall, both significance tests tended to assign more and higher significance to the $|\overline{CAR}s|$ of the robustness check than to those of the main analysis. In summary, the robustness of the evidence was generally confirmed, with individual cases again demonstrating the sensitivity of the results to the specification of the normal return-generating model.

6.3.2 The Functional Relationship Hypothesis

The research question to be answered based on the results below was: *what is the functional relationship between new information from EU-wide stress test results and corresponding abnormal stock returns?* This question was examined separately for each of the five relevant EU-wide stress tests using the same cross-sectional samples as the previous research question (Section 6.3.1). The size of the samples ranged from n = 33 to n = 59 sample banks. Due to the specifics of the research question, the analysis was constrained to the $CARs$ over the post-event window. The results are reported separately for the $CARs$-ΔCRs relationship defined in Section 5.3.2.3.1 and the robustness checks using alternative specifications of the risk measure involved.

6.3.2.1 The $CARs$-ΔCRs Relationship

A series of simple OLS regressions were used to determine the proportion of the total variance in the empirically observed $CARs$ that was explained by the fitted $\widehat{CAR}s$ predicted from linear and quadratic fits to the $CARs$-ΔCRs relationship. That is, to determine whether EU-wide stress test results and the corresponding abnormal stock returns were better characterised by a linear or a quadratic relationship. The regression analysis was complemented by the calculation of error metrics (*MSE* and *RMSE*) to quantify the prediction error between the observed and fitted values. Both analyses together formed the performance evaluation. It is important to note that the $\widehat{CAR}s$ predicted by the linear and quadratic fits were not tested against each other but, independently, against the actual $CARs$ observed for each sample to evaluate the quality of the approximations. Table 19 shows the performance evaluation results for the linear and quadratic fits to the $CARs$-ΔCRs relationship under the adverse stress test scenario.

Table 19

Performance Evaluation Results of the Linear and Quadratic Fits to the Relationship Between $CARs$ and ΔCRs (Adverse Scenario)

		Regression Analysis		Error Metrics	
Sample	***n***	R^2	***F***	***MSE***	***RMSE***
Linear Fit					
CEBS 2010	50	.013	(1, 48) 0.64	0.000732	0.0270
EBA 2011	51	.245	(1, 49) 15.87***	0.001221	0.0349
EBA 2014	59	< .001	(1, 57) 0.02	0.002062	0.0454
EBA 2016	34	.008	(1, 32) 0.26	0.000540	0.0232
EBA 2018	33	.087	(1, 31) 2.94*	0.000217	0.0147
Quadratic Fit					
CEBS 2010	50	**.017**	**(1, 48) 0.83**	**0.000729**	0.0270
EBA 2011	51	**.344**	**(1, 49) 25.74*** **	**0.001060**	**0.0326**
EBA 2014	59	**.077**	**(1, 57) 4.75** **	**0.001904**	**0.0436**
EBA 2016	34	**.220**	**(1, 32) 9.03*** **	**0.000425**	**0.0206**
EBA 2018	33	**.259**	**(1, 31) 10.85*** **	**0.000176**	**0.0133**

Note. This table shows the performance evaluation results of the linear and quadratic fits to the relationship between the cumulative abnormal returns ($CARs$) and the capital ratio differences (ΔCRs) between the sample banks' stressed and actual capital ratios under the adverse stress test scenario for the five cross-sectional samples, cumulated over the post-event window $(+1, n)$. The term n in the definition of the post-event window $(+1, n)$ refers to a statistical determination of variable event window lengths based on the Ljung-Box (1978) test introduced in this study. Specifically, the table shows the results of the regression analyses and the error metrics that were conducted to evaluate the performance of the linear and quadratic fits. It is important to note that the fitted functions were not tested against each other, but independently against the actual, empirically observed data. The normal returns used to calculate the $CARs$ that fed into the $CARs$-ΔCRs relationship were estimated using the Fama and French (1993) Three-Factor Model. Values in bold indicate better performance in the direct comparison between the linear and quadratic fits. *n* = sample size. R^2 = coefficient of determination. *F* = *F*-value (numbers in brackets are the degrees of freedom). *MSE* = mean squared error. *RMSE* = root mean squared error. CEBS = Committee of European Banking Supervisors. EBA = European Banking Authority.

* $p < .10$. ** $p < .05$. *** $p < .01$.

The R^2 values in Table 19 show that the quadratically fitted $\widehat{CARs}$ explained up to 34.4 percent of the total variance in the $CARs$ (EBA 2011), with R^2 being rather high in most cases. The corresponding F-tests indicate that the regression results were statistically significant at the .05 level or better in all cases except the CEBS 2010. In contrast, the explanatory power of the linearly fitted $\widehat{CARs}$ was generally much lower and tended not to be statistically significant. The error metrics confirmed the results of the regression analysis, showing that the prediction errors of the linear fits generally exceeded those of the quadratic fits. Overall, the performance evaluation suggests that the relationship between EU-wide stress test results and the corresponding abnormal stock returns was quadratic rather than linear. Figure 13 visualizes this finding using sample-specific scatter plots including the associated linear and quadratic fit curves.

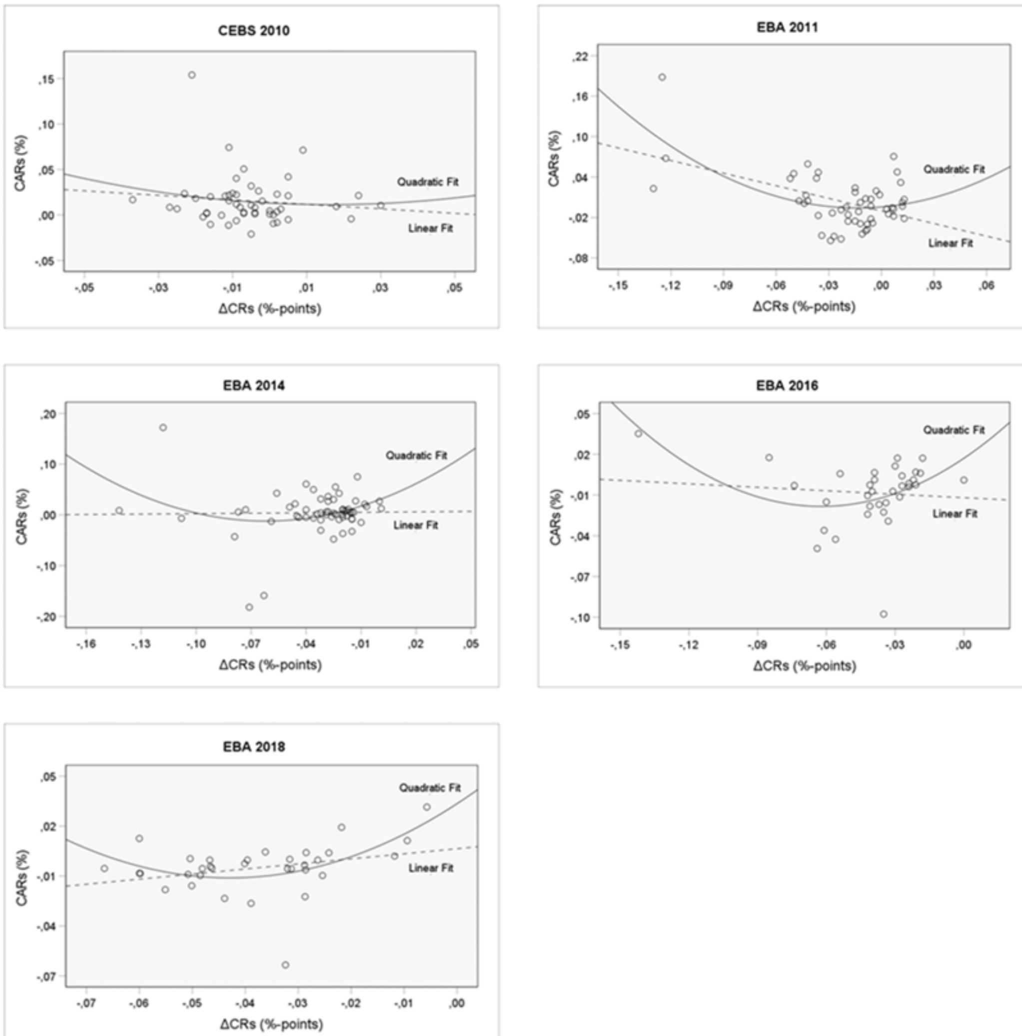


Figure 13. Scatter plots of the observed $CARs$-ΔCRs relationships (adverse scenario) overlaid with the fitted linear and quadratic curves

The results of the performance evaluation provided evidence against the null hypothesis and in favour of the alternative hypothesis in almost all cases. That is, the quadratic fits generally provided a better description of the relationship between EU-wide stress test results and the corresponding abnormal stock returns than the linear fits. Table 20 summarises the hypothesis testing results by detailing each acceptance and rejection decision.

Table 20
Summary of Hypothesis Testing Results at the .05 Significance Level Based on the Relationship Between $CARs$ and ΔCRs (Adverse Stress Test Scenario)

Sample	*n*	Null Hypotesis	Alternative Hypothesis
CEBS 2010	50	Accepted	Rejected
EBA 2011[a]	51	Rejected	Accepted
EBA 2014	59	Rejected	Accepted
EBA 2016	34	Rejected	Accepted
EBA 2018	33	Rejected	Accepted

Note. This table summarises the acceptance and rejection of the relevant hypotheses at the .05 significance level or better for the analyses performed on the relationship between the cumulative abnormal returns ($CARs$) and the capital ratio differences (ΔCRs) between the sample banks' stressed and actual capital ratios under the adverse stress test scenario for the five cross-sectional samples, cumulated over the post-event window $(+1, n)$. The term n in the definition of the post-event window $(+1, n)$ refers to a statistical determination of variable event window lengths based on the Ljung-Box (1978) test introduced in this study. *n* = sample size. CEBS = Committee of European Banking Supervisors. EBA = European Banking Authority.

[a] Since the significance levels were the same, the accept-reject decision regarding the EBA 2011 was made based on the exact *p*-values of the linear (*p* = .000225) and the quadratic (*p* = .000006) fit.

6.3.2.2 Robustness Checks

To check the robustness of the results, the performance evaluation was repeated with several alternative specifications of the risk measure that feeds into the risk-return relationship. Specifically, these alternative risk measures were: (1) the ΔCRs based on the baseline scenario of the stress tests, (2) the stressed capital ratios ($\widehat{CR}s$) resulting from the adverse scenario, and (3) the stressed capital ratios ($\widehat{CR}s$) resulting from the baseline scenario. Each alternative measure of risk was again related to the $CARs$ over the post-event window, and the observed values were compared to the fitted values from the linear and quadratic fits. As with the main analysis, the fitted functions were not tested against each other, but independently against the actual, empirically observed data. Table 21 summarises the results of the individual robustness checks in a single overview.

Table 21

Robustness Checks of the Specification of the Risk Measure Used in the Risk-Return Relationship

		(1) ΔCRs (Baseline Scenario)				(2) $\widehat{CR}s$ (Adverse Scenario)				(3) $\widehat{CR}s$ (Baseline Scenario)			
		Regression Analysis		**Error Metrics**		**Regression Analysis**		**Error Metrics**		**Regression Analysis**		**Error Metrics**	
Sample	***n***	R^2	***F***	***MSE***	***RMSE***	R^2	***F***	***MSE***	***RMSE***	R^2	***F***	***MSE***	***RMSE***
Linear Fit													
CEBS 2010	50	< .001	(1, 48) 0.02	0.000741	0.0272	.148	(1, 48) 8 32***	0.000632	0.0251	.176	(1, 48) 10.22***	0.000611	0.0247
EBA 2011	51	.201	(1, 49) 12.33***	0.001292	0.0359	.150	(1, 49) 8.63***	0.001375	0.0371	.031	(1, 49) 1.54	0.001567	0.0396
EBA 2014	59	< .001	(1, 57) 0.02	0.002062	0.0454	.046	(1, 57) 2.76	0.001968	0.0444	.058	(1, 57) 3.52*	0.001943	0.0441
EBA 2016	34	.019	(1, 32) 0.60	0.000534	0.0231	.041	(1, 32) 1.36	0.000522	0.0229	.119	(1, 32) 4 32**	0.000480	0.0219
EBA 2018	33	.001	(1, 31) 0.04	0.000237	0.0154	.169	(1, 31) 6.28**	0.000197	0.0140	.074	(1, 31) 2.49	0.000220	0.0148
Quadratic Fit													
CEBS 2010	50	**.002**	**(1, 48) 0.09**	**0.000740**	0.0272	**.155**	**(1, 48) 8.83*** **	**0.000626**	**0.0250**	**.210**	**(1, 48) 12.76*** **	**0.000586**	**0.0242**
EBA 2011	51	**.341**	**(1, 49) 25.38*** **	**0.001065**	**0.0326**	**.248**	**(1, 49) 16.18*** **	**0.001215**	**0.0349**	**.081**	**(1, 49) 4.33** **	**0.001485**	**0.0385**
EBA 2014	59	**.036**	**(1, 57) 2 15**	**0.001988**	**0.0446**	**.066**	**(1, 57) 4.02** **	**0.001927**	**0.0439**	**.116**	**(1, 57) 7.45*** **	**0.001824**	**0.0427**
EBA 2016	34	**.026**	**(1, 32) 0 87**	**0.000530**	**0.0230**	**.165**	**(1, 32) 6.34** **	**0.000454**	**0.0213**	**.123**	**(1, 32) 4.47** **	**0.000478**	0.0219
EBA 2018	33	**.042**	**(1, 31) 1.36**	**0.000227**	**0.0151**	**.238**	**(1, 31) 9.68*** **	**0.000181**	**0.0134**	**.131**	**(1, 31) 4.69** **	**0.000206**	**0.0144**

Note. This table summarises the results of three robustness checks performed to check the robustness of the main analysis results against alternative specifications of the risk measure in the risk-return relationship. It should be noted that the fitted functions were not tested against each other, but independently against the actual, empirically observed data. The normal returns used to calculate the cumulative abnormal returns ($CARs$) involved in each risk-return relationship were estimated using the Fama and French (1993) Three-Factor Model. Values in bold indicate better performance in the direct comparison between the linear and quadratic fits. ΔCRs = capital ratio differences between the stressed and actual capital ratios. $\widehat{CR}s$ = stressed capital ratios. *n* = sample size. R^2 = coefficient of determination. *F* = *F*-value (numbers in brackets are the degrees of freedom). *MSE* = mean squared error. *RMSE* = root mean squared error. CEBS = Committee of European Banking Supervisors. EBA = European Banking Authority.

* $p < .10$. ** $p < .05$. *** $p < .01$.

Overall, the robustness check results in Table 21 (R^2, MSE, $RMSE$) show that the quadratic fit consistently performed better than the linear fit, regardless of the risk measure used. This confirmed the results of the main analysis. However, the outperformance was marginal in many cases. In addition, for the ΔCRs calculated from the baseline scenario (Robustness Check 1), the significance levels of the linear and quadratic fits were the same for all samples (as implied in Section 5.3.2.3.1). This was not the case for Robustness Checks 2 and 3, where the significance of the quadratic fits typically exceeded that of the linear fits (with all results being significant at the .01 or .05 level). However, these results may have been affected by another effect, namely by relating the $CARs$ to stressed capital ratios ($\widehat{CR}s$) rather than rates of change (ΔCRs) and thus to a more static measure. In summary, the robustness checks provided moderate but consistent support for the results obtained from the main analysis.

6.3.3 The Intertemporal Stability Hypothesis

The research question to be answered based on the following results was: *how has the informational value of EU-wide stress test results, measured in abnormal stock returns, changed over time?* This question was examined using a single longitudinal sample (*n* = 28) composed only of those banks that were subject to *all* five relevant EU-wide stress tests, minus the exclusions due to unavailability of data or exposure to known extraneous factors (see Logic Control in Section 5.3.1.3). The analysis consisted of two consecutive parts and included (1) parametric and non-parametric omnibus tests and (2) multiple comparison *post hoc* tests. The results are reported separately for the directional $\overline{CAR}s$ and the non-directional $|\overline{CAR}s|$, including the corresponding robustness checks.

6.3.3.1 Average Cumulative Abnormal Returns ($\overline{\boldsymbol{CAR}}\boldsymbol{s}$)

A series of one-way repeated measures ANOVAs and one-way Friedman tests were used to determine whether there were statistically significant changes in the $\overline{CAR}s$ over the course of the five EU-wide stress tests. That is, whether the informational value contained in EU-wide stress test results was intertemporally stable. The analyses were performed for each of the three event-window types. The normal returns used to calculate the underlying $CARs$ of each sample bank were estimated using the Fama and

French (1993) Three-Factor Model. Table 22 presents the results of both tests including effect size.

Table 22

Changes in Informational Value of EU-Wide Stress Test Results (2009 to 2018) Measured in Average Cumulative Abnormal Returns ($\overline{CARs}$)

	One-Way Repeated Measures ANOVA			**One-Way Friedman Test**		
Event Window[a]	F^b	p	η^2	χ^2_F	p	W
Pre-Event Window $(-2, 0)$	(4, 108) 7.61***	< .001	.22	(4) 20.46***	< .001	.18
Standard Event Window $(-2, +2)$	(4, 108) 4.97***	.001	.16	(4) 19.00***	< .001	.17
Post-Event Window $(+1, n)$	(2.88, 77.77) 5.02***	.004	.16	(4) 20.17***	< .001	.18

Note. This table shows the results of two omnibus tests (one-way repeated measures ANOVA and one-way Friedman test) for the longitudinal sample of average cumulative abnormal returns ($\overline{CARs}$) for each event-window type. The parentheses under the event-window types specify the number of trading days before and after the event date t_0. The term n in the definition of the post-event window $(+1, n)$ refers to a statistical determination of variable event window lengths based on the Ljung-Box (1978) test introduced in this study. The normal returns used to calculate the underlying cumulative abnormal returns ($CARs$) for each individual sample bank were estimated using the Fama and French (1993) Three-Factor Model. F = F-value (numbers in brackets are the degrees of freedom). p = p-value. η^2 = effect size. χ^2_F = χ^2-value (numbers in brackets are the degrees of freedom). W = Kendall's W (effect size). n = sample size.

[a] n = 28. [b] Mauchly's test of sphericity showed that the sphericity assumption was met for the pre-event window ($\chi^2(9)$ = 13.29, p = 151) and the standard event window ($\chi^2(9)$ = 5.03, p = .832), but it was violated for the post-event window ($\chi^2(9)$ = 20.09, p = .018). Therefore, the Greenhouse-Geisser correction (ε = 0.720) was applied to the post-event window.

* $p < .10$. ** $p < .05$. *** $p < .01$.

The results of the ANOVAs and Friedman tests in Table 22 consistently show that the $\overline{CARs}$ observed in response to EU-wide stress test results have experienced statistically significant changes over time (regardless of the event window used).[125] This suggests that the informational value contained in the results of EU-wide stress tests was not intertemporally stable. Notably, the changes found by both tests were significant at the .01 level in all cases. The high agreement between the two tests shows that the results were robust across parametric and non-parametric approaches. A major difference, however, was the magnitude of the effect size found. Based on commonly used guidelines for interpreting effect size, the time effect was found to be large by the ANOVAs and small by the Friedman tests.[126]

---

[125] For the difference in $\overline{CARs}$ between any two EU-wide stress tests, see the multiple comparisons in Table 23. For the mean, median, and standard deviation of the $CARs$ from each individual EU-wide stress test, see the descriptive statistics in Table 11.

[126] The thresholds used for effect size classification were as follows: η^2 = .01 (small effect), η^2 = .06 (moderate effect), η^2 = .14 (large effect), and W = .10 (small effect), W = .30 (moderate effect), W = .50 (large effect).

However, since both tests are omnibus tests, they were limited to determining whether there were overall significant changes in the $\overline{CARs}$ over time. In order to identify the *exact* EU-wide stress tests that caused the changes, the omnibus tests were followed up with multiple comparison *post hoc* tests (*i.e.* paired-sample *t*-tests and paired-sample Wilcoxon signed-rank tests). Table 23 summarises the results of all *post hoc* tests for each event-window type.

Table 23

Multiple Comparison post hoc Tests of EU-wide Stress Tests (2009 to 2018) Based on Average Cumulative Abnormal Returns ($\overline{CAR}s$)

Comparison		Pre-Event Window $(-2, 0)$			Standard Event Window $(-2, +2)$			Post-Event Window $(+1, n)$		
			p^*			p^*			p^*	
Sample 1	**Sample 2**	$\Delta\overline{CAR}$	***t*-test**	**Wilcoxon**	$\Delta\overline{CAR}$	***t*-test**	**Wilcoxon**	$\Delta\overline{CAR}$	***t*-test**	**Wilcoxon**
CEBS 2010	EBA 2011	0.41	1.000	1.000	3.28	**.039**	**.013**	1.81	.171	**.007**
	EBA 2014	-1.07	.159	.910	1.53	.954	1.000	0.67	1.000	1.000
	EBA 2016	0.24	1.000	1.000	2.67	.132	.088	2.02	**.003**	**.023**
	EBA 2018	-2.42	**.009**	**.007**	0.37	1.000	1.000	1.65	**.003**	**.002**
EBA 2011	CEBS 2010	-0.41	1.000	1.000	-3.28	**.039**	**.013**	-1.81	.171	**.007**
	EBA 2014	-1.48	.131	.280	-1.75	.442	1.000	-1.14	.734	.346
	EBA 2016	-0.17	1.000	1.000	-0.61	1.000	1.000	0.21	1.000	1.000
	EBA 2018	-2.83	**.002**	**.001**	-2.91	**.010**	**.005**	-0.16	1.000	1.000
EBA 2014	CEBS 2010	1.07	.159	.910	-1.53	.954	1.000	-0.67	1.000	1.000
	EBA 2011	1.48	.131	.280	1.75	.442	1.000	1.14	.734	.346
	EBA 2016	1.31	.113	1.000	1.14	1.000	1.000	1.35	.241	.759
	EBA 2018	-1.35	.483	.910	-1.16	1.000	.630	0.98	.214	.142
EBA 2016	CEBS 2010	-0.24	1.000	1.000	-2.67	.132	.088	-2.02	**.003**	**.023**
	EBA 2011	0.17	1.000	1.000	0.61	1.000	1.000	-0.21	1.000	1.000
	EBA 2014	-1.31	.113	1.000	-1.14	1.000	1.000	-1.35	.241	.759
	EBA 2018	-2.66	**.013**	**.010**	-2.30	.182	**.041**	-0.37	1.000	1.000
EBA 2018	CEBS 2010	2.42	**.009**	**.007**	-0.37	1.000	1.000	-1.65	**.003**	**.002**
	EBA 2011	2.83	**.002**	**.001**	2.91	**.010**	**.005**	0 16	1.000	1.000
	EBA 2014	1.35	.483	.910	1.16	1.000	.630	-0.98	.214	.142
	EBA 2016	2.66	**.013**	**.010**	2.30	182	**.041**	0 37	1.000	1.000

Note. This table summarises the results of all multiple comparison *post hoc* tests based on the average cumulative abnormal returns ($\overline{CAR}s$) of the longitudinal sample for each event-window type. The parentheses behind the event-window types specify the number of trading days before and after the event date t_0. The term *n* in the definition of the post-event window $(+1, n)$ refers to a statistical determination of variable event window lengths based on the Ljung-Box (1978) test introduced in this study. The normal returns used to calculate the underlying cumulative abnormal returns ($CARs$) for each individual sample bank were estimated using the Fama and French (1993) Three-Factor Model. Values in bold indicate statistical significance at the .05 level or better. Bonferroni-corrected *p*-values (p^*) can be > 1 due to backward correction used by SPSS; such values are therefore shown as p^* = 1.000. $\Delta\overline{CAR}$ = difference in $\overline{CAR}s$ (Sample 1 – Sample 2). p^* = Bonferroni-corrected *p*-value (adjustment for multiple comparisons). CEBS = Committee of European Banking Supervisors. EBA = European Banking Authority.

The results in Table 23 show that the *post hoc* tests uncovered several pairs of EU-wide stress tests whose $\overline{CARs}$ were significantly different at the .05 level or better. This confirmed the results of the omnibus tests (Table 22) and identified the exact EU-wide stress tests that caused the overall significant changes. With only two exceptions, the *t*-tests and Wilcoxon signed-rank tests gave very consistent results, showing that their results were robust.[127] The key findings of the *post hoc* tests are summarised below for each event-window type.

For the pre-event window, three of the 10 pairwise comparisons were found to be statistically significant.[128] Notably, all three were comparisons to the EBA 2018 and included the CEBS 2010, EBA 2011, and EBA 2016. None of the preceding significant stress tests had a $\overline{CAR}$ level that exceeded that of the EBA 2018. For the standard event window, two (*t*-test) or three (Wilcoxon signed-rank test) pairwise comparisons were identified as statistically significant. This affected all EU-wide stress tests, with the exception of EBA 2014, and mainly involved EBA 2011 and EBA 2018. The level of $\overline{CARs}$ decreased or increased from preceding to succeeding significant stress tests and showed no discernible trend over time. Finally, for the post-event window, two (*t*-test) or three (Wilcoxon signed-rank test) pairwise comparisons proved statistically significant. All of them concerned the CEBS 2010 and involved EBA 2011, EBA 2016, and EBA 2018. The $\overline{CAR}$ level of all subsequent significant stress tests was lower than that of the CEBS 2010. Figure 14 illustrates the above findings by plotting the data and highlighting statistically significant pairs of EU-wide stress tests.

127 The only two exceptions where the *t*-test and the Wilcoxon signed-rank test disagreed were the pairwise comparisons between EBA 2016 and EBA 2018 (standard event window) and between CEBS 2010 and EBA 2011 (post-event window) and *vice versa*, see Table 23.

128 Due to the two-sided order of the pairwise comparisons in Table 23, all information on the number of comparisons has been corrected for double counting.

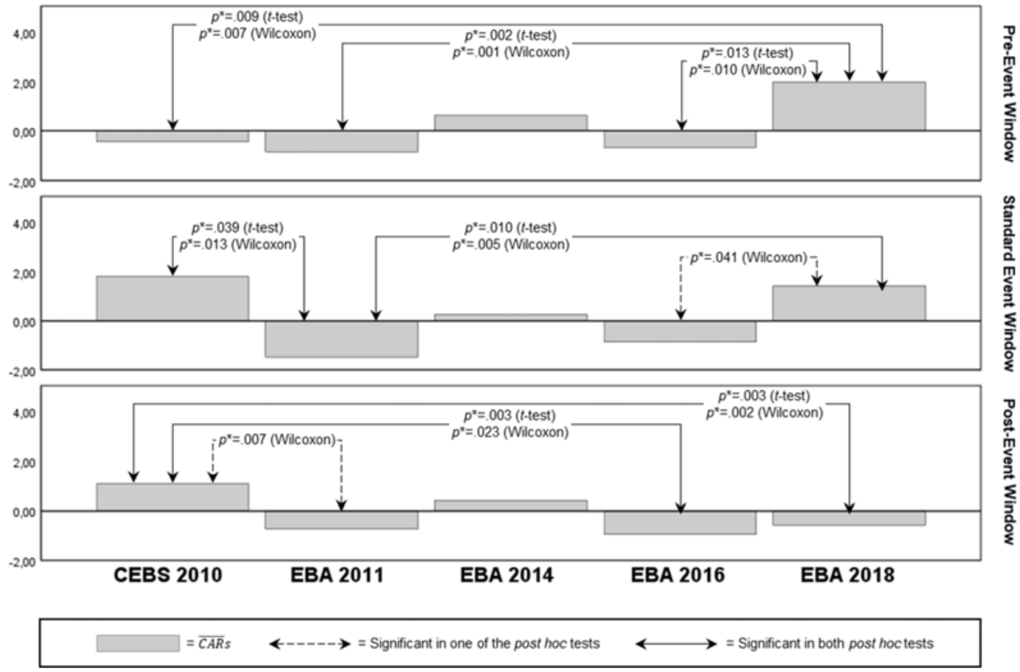


Figure 14. Significant pairs of EU-wide stress tests ($\overline{CAR}s$)

Overall, the chronological sequence of EU-wide stress tests showed no statistically significant downward trend in the informational value ($\overline{CAR}$) except for the post-event window. The results of the *post hoc* tests thus provided evidence against the null hypothesis and in favour of the alternative hypothesis in the case of the pre-event and standard event windows. However, the null hypothesis could not be rejected for the post-event window. Table 24 summarises the results of the hypothesis tests.

Table 24
Summary of Hypothesis Testing Results at the .05 Significance Level Based on Average Cumulative Abnormal Returns ($\overline{CAR}s$)

Event Window[a]	Null Hypotesis	Alternative Hypothesis
Pre-Event Window $(-2, 0)$	Rejected	Accepted
Standard Event Window $(-2, +2)$	Rejected	Accepted
Post-Event Window $(+1, n)$	Accepted	Rejected

Note. This table summarises the acceptance and rejection of the relevant hypotheses at the 05 significance level or better for the analyses performed on the average cumulative abnormal returns ($\overline{CAR}s$) of the longitudinal sample for each event-window type The parentheses behind the event-window types specify the number of trading days before and after the event date t_0 The term n in the definition of the post-event window $(+1, n)$ refers to a statistical determination of variable event window lengths based on the Ljung-Box (1978) test introduced in this study n = sample size
[a] n = 28

As a robustness check, the above analyses were repeated with $\overline{CAR}s$ calculated with a different specification of the normal return-generating model. For this purpose, the Market Model was used again (as in Section 6.3.1.1). Table 25 shows the results of the corresponding omnibus tests.

Table 25

Robustness Check of the Model Specification for the Calculation of the $\overline{CAR}s$ Underlying the Omnibus Tests

	One-Way Repeated Measures ANOVA			**One-Way Friedman Test**		
Event Window[a]	F[b]	p	η^2	χ^2_F	p	W
Pre-Event Window $(-2, 0)$	(4, 108) 6.78***	< .001	.20	(4) 21.60***	< .001	.19
Standard Event Window $(-2, +2)$	(4, 108) 12.40***	< .001	.32	(4) 33.51***	< .001	.30
Post-Event Window $(+1, n)$	(2.84, 76.63) 10.43***	< .001	.28	(4) 31.54***	< .001	.28

Note. This table shows the results of a robustness check performed to check the robustness of the results from the main analysis against an alternative specification of the normal return-generating model. Specifically, the table shows the results of two omnibus tests (one-way repeated measures ANOVA and one-way Friedman test) for the longitudinal sample of average cumulative abnormal returns ($\overline{CAR}s$) for each event-window type, with the normal returns used to calculate the underlying cumulative abnormal returns ($CARs$) being estimated using the Market Model. The parentheses under the event-window types specify the number of trading days before and after the event date t_0. The term n in the definition of the post-event window $(+1, n)$ refers to a statistical determination of variable event window lengths based on the Ljung-Box (1978) test introduced in this study. F = F-value (numbers in brackets are the degrees of freedom). p = p-value. η^2 = effect size. χ^2_F = χ^2-value (numbers in brackets are the degrees of freedom). W = Kendall's W (effect size). n = sample size.

[a] n = 28. [b] Mauchly's test of sphericity showed that the sphericity assumption was met for the pre-event window ($\chi^2(9)$ = 11.91, p = 220) and the standard event window ($\chi^2(9)$ = 4.42, p = .882), but it was violated for the post-event window ($\chi^2(9)$ = 20.89, p = .013). Therefore, the Greenhouse-Geisser correction (ε = 0.710) was applied to the post-event window.

* $p < .10$. ** $p < .05$. *** $p < .01$.

The results of the robustness check in Table 25 were very similar to those of the main analysis (Table 22), indicating that they were robust. To further assess the robustness, the above results were followed up with multiple comparison *post hoc* tests (as in the main analysis). The results are reported in Table 26.

Table 26

Robustness Check of the Model Specification for the Calculation of the $\overline{CAR}s$ Underlying the Multiple Comparison post hoc Tests

		Pre-Event Window $(-2, 0)$			**Standard Event Window $(-2, +2)$**			**Post-Event Window $(+1, n)$**		
Comparison			p^*			p^*			p^*	
Sample 1	**Sample 2**	$\Delta\overline{CAR}$	***t*-test**	**Wilcoxon**	$\Delta\overline{CAR}$	***t*-test**	**Wilcoxon**	$\Delta\overline{CAR}$	***t*-test**	**Wilcoxon**
CEBS 2010	EBA 2011	-0.29	1.000	1.000	5.46	**< .001**	**< .001**	2.51	**.039**	**< .001**
	EBA 2014	-1.81	**.002**	.053	4.22	**< .001**	**.010**	2.27	**< .001**	**.013**
	EBA 2016	-1.00	.678	1.000	5.70	**< .001**	**< .001**	3.60	**< .001**	**< .001**
	EBA 2018	-2.71	**.002**	**< .001**	2.73	.059	1.000	1.98	**.001**	.112
EBA 2011	CEBS 2010	0.29	1.000	1.000	-5.46	**< .001**	**< .001**	-2.51	**.039**	**< .001**
	EBA 2014	-1.51	.118	.425	-1.24	1.000	1.000	-0.24	1.000	1.000
	EBA 2016	-0.71	1.000	1.000	0.24	1.000	1.000	1.10	1.000	1.000
	EBA 2018	-2.42	**.012**	**.007**	-2.73	**.024**	.053	-0.53	1.000	.910
EBA 2014	CEBS 2010	1.81	**.002**	.053	-4.22	**< .001**	**.010**	-2.27	**< .001**	**.013**
	EBA 2011	1.51	.118	.425	1.24	1.000	1.000	0.24	1.000	1.000
	EBA 2016	0.81	1.000	1.000	1.49	1.000	.910	1.33	.209	.425
	EBA 2018	-0.91	1.000	1.000	-1.49	.686	.630	-0.29	1.000	1.000
EBA 2016	CEBS 2010	1.00	.678	1.000	-5.70	**< .001**	**< .001**	-3.60	**< .001**	**< .001**
	EBA 2011	0.71	1.000	1.000	-0.24	1.000	1.000	-1.10	1.000	1.000
	EBA 2014	-0.81	1.000	1.000	-1.49	1.000	.910	-1.33	.209	.425
	EBA 2018	-1.71	.275	.112	-2.98	**.050**	**.004**	-1.62	**.032**	.068
EBA 2018	CEBS 2010	2.71	**.002**	**< .001**	-2.73	.059	1.000	-1.98	**.001**	.112
	EBA 2011	2.42	**.012**	**.007**	2.73	**.024**	.053	0 53	1.000	.910
	EBA 2014	0.91	1.000	1.000	1.49	.686	.630	0 29	1.000	1.000
	EBA 2016	1.71	.275	.112	2.98	**.050**	**.004**	1.62	**.032**	.068

Note. This table shows the results of a robustness check performed to check the robustness of the results from the main analysis against an alternative specification of the normal return-generating model. Specifically, the table shows the results of all multiple comparison *post hoc* tests based on the average cumulative abnormal returns ($\overline{CAR}s$) of the longitudinal sample for each event-window type, with the normal returns used to calculate the underlying cumulative abnormal returns ($CARs$) being estimated using the Market Model. The parentheses behind the event-window types specify the number of trading days before and after the event date t_0. The term n in the definition of the post-event window $(+1, n)$ refers to a statistical determination of variable event window lengths based on the Ljung-Box (1978) test introduced in this study. Values in bold indicate statistical significance at the .05 level or better. Bonferroni-corrected p-values (p^*) can be > 1 due to backward correction used by SPSS; such values are therefore shown as p^* = 1.000. $\Delta\overline{CAR}$ = difference in $\overline{CAR}s$ (Sample 1 – Sample 2). p^* = Bonferroni-corrected p-value (adjustment for multiple comparisons). CEBS = Committee of European Banking Supervisors. EBA = European Banking Authority.

The robustness check produced *post hoc* test results (Table 26) that were qualitatively similar to those of the main analysis (Table 23). No statistically significant downward trend could be identified for the pre-event and standard event windows. The post-event window of the robustness check generally showed a similar pattern to that of the main analysis; however, the increase in the informational value ($\overline{CAR}$) in the statistically significant comparison between EBA 2016 and EBA 2018 meant that a downward trend had to be rejected. The robustness check therefore generally confirmed the results of the main analysis and showed that they were robust to an alternative specification of the normal return-generating model.

6.3.3.2 Average Absolute Cumulative Abnormal Returns ($|\overline{CARs}|$)

As with the $\overline{CARs}$, one-way repeated measures ANOVAs and one-way Friedman tests were used to determine whether there were statistically significant changes in the $|\overline{CARs}|$ over the course of the five EU-wide stress tests. The analyses were performed for each of the three event-window types. The normal return-generating model used to calculate the underlying $|CARs|$ of each sample bank was the Fama and French (1993) Three-Factor Model. Table 27 shows the results of the two tests including effect size.

Table 27
Changes in Informational Value of EU-Wide Stress Test Results (2009 to 2018) Measured in Average Absolute Cumulative Abnormal Returns ($|\overline{CARs}|$)

	One-Way Repeated Measures ANOVA			One-Way Friedman Test		
Event Window[a]	F^b	p	η^2	χ^2_F	p	W
Pre-Event Window $(-2, 0)$	(4, 108) 4.61***	.002	.15	(4) 8.03*	.091	.07
Standard Event Window $(-2, +2)$	(4, 108) 1.60	.180	.06	(4) 15.51***	.004	.14
Post-Event Window $(+1, n)$	(4, 108) 4.18***	.003	.13	(4) 10.06**	.039	.09

Note. This table shows the results of two omnibus tests (one-way repeated measures ANOVA and one-way Friedman test) for the longitudinal sample of average absolute cumulative abnormal returns ($|\overline{CARs}|$) for each event-window type. The parentheses under the event-window types specify the number of trading days before and after the event date t_0. The term n in the definition of the post-event window $(+1, n)$ refers to a statistical determination of variable event window lengths based on the Ljung-Box (1978) test introduced in this study. The normal returns used to calculate the underlying absolute cumulative abnormal returns ($|CARs|$) for each individual sample bank were estimated using the Fama and French (1993) Three-Factor Model. F = F-value (numbers in brackets are the degrees of freedom). p = p-value. η^2 = effect size. $\chi^2_F = \chi^2$-value (numbers in brackets are the degrees of freedom). W = Kendall's W (effect size). n = sample size.

[a] $n = 28$. [b] Mauchly's test of sphericity showed that the sphericity assumption was met for the pre-event window ($\chi^2(9) = 11.78$, $p = .227$), the standard event window ($\chi^2(9) = 4.22$, $p = .897$), and the post-event window ($\chi^2(9) = 7.67$, $p = .568$).

* $p < .10$. ** $p < .05$. *** $p < .01$.

The results in Table 27 show that the ANOVAs and Friedman tests provided mixed evidence, both in direct comparison and across the various event-window types. Accordingly, the determined effect sizes ranged between small (Friedman tests) and moderate to large effects (ANOVAs).[129] The subsequent *post hoc* tests provided a more differentiated perspective based on multiple pairwise comparisons. Table 28 summarises the results of the *post hoc* tests for each event-window type.

[129] The thresholds used for effect size classification were again: η^2 = .01 (small effect), η^2 = .06 (moderate effect), η^2 = .14 (large effect), and W = .10 (small effect), W = .30 (moderate effect), W = .50 (large effect).

Table 28

Multiple Comparison post hoc Tests of EU-wide Stress Tests (2009 to 2018) Based on Average Absolute Cumulative Abnormal Returns ($|\overline{CAR}s|$)

Comparison		Pre-Event Window $(-2, 0)$			Standard Event Window $(-2, +2)$			Post-Event Window $(+1, n)$		
			p^*			p^*			p^*	
Sample 1	**Sample 2**	$\Delta\lvert\overline{CAR}\rvert$	***t*-test**	**Wilcoxon**	$\Delta\lvert\overline{CAR}\rvert$	***t*-test**	**Wilcoxon**	$\Delta\lvert\overline{CAR}\rvert$	***t*-test**	**Wilcoxon**
CEBS 2010	EBA 2011	-0.51	1.000	.759	1.16	.800	.759	-1.10	.083	.346
	EBA 2014	0.30	1.000	1.000	1.59	.241	**.005**	0.37	1.000	1.000
	EBA 2016	-0.23	1.000	1.000	1.29	.888	**.013**	-0.19	1.000	1.000
	EBA 2018	-1.37	.096	.910	1.12	.930	.910	0.34	1.000	1.000
EBA 2011	CEBS 2010	0.51	1.000	.759	-1.16	.800	.759	1.10	.083	.346
	EBA 2014	0.81	.205	.280	0.43	1.000	.910	1.47	**.046**	**.041**
	EBA 2016	0.28	1.000	1.000	0.13	1.000	1.000	0 91	.791	180
	EBA 2018	-0.86	.821	1.000	-0.04	1.000	1.000	1.44	**.023**	.180
EBA 2014	CEBS 2010	-0.30	1.000	1.000	-1.59	.241	**.005**	-0.37	1.000	1.000
	EBA 2011	-0.81	.205	.280	-0.43	1.000	.910	-1.47	**.046**	**.041**
	EBA 2016	-0.53	1.000	1.000	-0.30	1.000	1.000	-0.56	1.000	1.000
	EBA 2018	-1.67	**.026**	.346	-0.47	1.000	.759	-0.03	1.000	1.000
EBA 2016	CEBS 2010	0.23	1.000	1.000	-1.29	.888	**.013**	0.19	1.000	1.000
	EBA 2011	-0.28	1.000	1.000	-0.13	1.000	1.000	-0.91	.791	180
	EBA 2014	0.53	1.000	1.000	0.30	1.000	1.000	0 56	1.000	1.000
	EBA 2018	-1.14	.240	1.000	-0.17	1.000	1.000	0 53	1.000	1.000
EBA 2018	CEBS 2010	1.37	.096	.910	-1.12	.930	.910	-0.34	1.000	1.000
	EBA 2011	0.86	.821	1.000	0.04	1.000	1.000	-1.44	**.023**	.180
	EBA 2014	1.67	**.026**	.346	0.47	1.000	.759	0.03	1.000	1.000
	EBA 2016	1.14	.240	1.000	0.17	1.000	1.000	-0.53	1.000	1.000

Note. This table summarises the results of all multiple comparison *post hoc* tests based on the average absolute cumulative abnormal returns ($|\overline{CAR}s|$) of the longitudinal sample for each event-window type. The parentheses behind the event-window types specify the number of trading days before and after the event date t_0. The term n in the definition of the post-event window $(+1, n)$ refers to a statistical determination of variable event window lengths based on the Ljung-Box (1978) test introduced in this study. The normal returns used to calculate the underlying absolute cumulative abnormal returns ($|CARs|$) for each individual sample bank were estimated using the Fama and French (1993) Three-Factor Model. Values in bold indicate statistical significance at the .05 level or better. Bonferroni-corrected p-values (p^*) can be > 1 due to backward correction used by SPSS; such values are therefore shown as p^* = 1.000. $\Delta|\overline{CAR}|$ = difference in $|\overline{CAR}s|$ (Sample 1 – Sample 2). p^* = Bonferroni-corrected p-value (adjustment for multiple comparisons). CEBS = Committee of European Banking Supervisors. EBA = European Banking Authority.

The *post hoc* test results in Table 28 show that most pairwise comparisons were not statistically significant. Nevertheless, the *t*-tests and Wilcoxon signed-rank tests sporadically identified pairs of EU-wide stress tests whose $|\overline{CARs}|$ were significantly different at the .05 level or better. Notably, however, the two tests agreed on only two pairwise comparisons while disagreeing on four. This indicates that the results were not robust across parametric and non-parametric approaches. The following summarises the key findings of the *post hoc* tests for each event-window type.

For the pre-event window, the *t*-test found only the $|\overline{CARs}|$ of the EBA 2014 and the EBA 2018 statistically different, while the Wilcoxon signed-rank test did not identify any pairwise comparison as significant. In general, the differences between the $|\overline{CARs}|$ of the individual stress tests were rather small and only increased substantially in comparison with the EBA 2018. With the standard event window, the situation was reversed. While the *t*-test did not find a single pair of significant $|\overline{CARs}|$, the Wilcoxon signed-rank test identified two significant pairwise comparisons. Notably, both involved the CEBS 2010. This was consistent with overall rather small differences between the $|\overline{CARs}|$ of the individual stress tests, apart from comparisons with the CEBS 2010. For the post-event window, both tests consistently identified the comparison between EBA 2011 and EBA 2014 as significant. In addition, the *t*-test also found a significant difference between EBA 2011 and EBA 2018. This was again consistent with generally rather small differences between the $|\overline{CARs}|$, apart from comparisons with the EBA 2011. Figure 15 visualises theses findings by plotting the $|\overline{CARs}|$ of all EU-wide stress tests and highlighting statistically significant pairs.

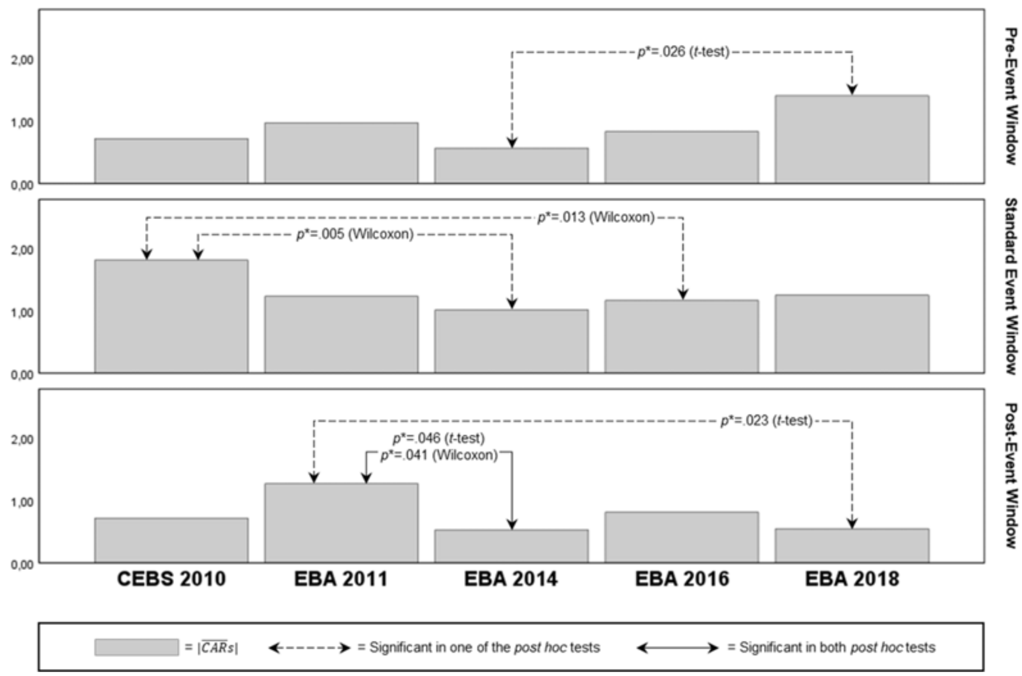


Figure 15. Significant pairs of EU-wide stress tests ($|\overline{CAR}s|$)

Overall, there were no signs of a statistically significant downward trend in the informational value ($|\overline{CAR}|$) for the pre-event window. In contrast, some corresponding evidence was found in the standard and post-event windows. Therefore, for the pre-event window, the null hypothesis was rejected in favour of the alternative hypothesis, while it was retained for the standard and post-event windows. Table 29 provides a summary of the hypothesis test results.

Table 29
Summary of Hypothesis Testing Results at the .05 Significance Level Based on Average Absolute Cumulative Abnormal Returns ($|\overline{CAR}s|$)

Event Window[a]	**Null Hypotesis**	**Alternative Hypothesis**
Pre-Event Window $(-2, 0)$	Rejected	Accepted
Standard Event Window $(-2, +2)$	Accepted	Rejected
Post-Event Window $(+1, n)$	Accepted	Rejected

Note. This table summarises the acceptance and rejection of the relevant hypotheses at the 05 significance level or better for the analyses performed on the average absolute cumulative abnormal returns ($|\overline{CAR}s|$) of the longitudinal sample for each event-window type The parentheses behind the event-window types specify the number of trading days before and after the event date t_0 The term n in the definition of the post-event window $(+1, n)$ refers to a statistical determination of variable event window lengths based on the Ljung-Box (1978) test introduced in this study n = sample size

[a] n = 28

Analogously to the $\overline{CARs}$, the robustness of the results obtained for the $|\overline{CARs}|$ was checked against a different specification of the normal return-generating model. For this purpose, the omnibus and *post hoc* tests were repeated with $|\overline{CARs}|$ calculated based on the Market Model. Table 30 reports the results of the omnibus tests.

Table 30
Robustness Check of the Model Specification for the Calculation of the $|\overline{CARs}|$ Underlying the Omnibus Tests

	One-Way Repeated Measures ANOVA			One-Way Friedman Test		
Event Window[a]	F^b	p	η^2	χ^2_F	p	W
Pre-Event Window $(-2, 0)$	(4, 108) 3.57***	.009	.12	(4) 3.46	.484	.03
Standard Event Window $(-2, +2)$	(4, 108) 3.46**	.011	.11	(4) 9.51**	.049	.09
Post-Event Window $(+1, n)$	(2.89, 77.98) 4.29***	.008	.14	(4) 14.31***	.006	.13

Note. This table shows the results of a robustness check performed to check the robustness of the results from the main analysis against an alternative specification of the normal return-generating model. Specifically, the table shows the results of two omnibus tests (one-way repeated measures ANOVA and one-way Friedman test) for the longitudinal sample of average absolute cumulative abnormal returns ($|\overline{CARs}|$) for each event-window type, with the normal returns used to calculate the underlying absolute cumulative abnormal returns ($|CARs|$) being estimated using the Market Model. The parentheses under the event-window types specify the number of trading days before and after the event date t_0. The term n in the definition of the post-event window $(+1, n)$ refers to a statistical determination of variable event window lengths based on the Ljung-Box (1978) test introduced in this study. F = F-value (numbers in brackets are the degrees of freedom). p = p-value. η^2 = effect size. χ^2_F = χ^2-value (numbers in brackets are the degrees of freedom). W = Kendall's W (effect size). n = sample size.

[a] n = 28. [b] Mauchly's test of sphericity showed that the sphericity assumption was met for the pre-event window ($\chi^2(9)$ = 12.78, p = .174) and the standard event window ($\chi^2(9)$ = 13.59, p = .139), but it was violated for the post-event window ($\chi^2(9)$ = 28.78, p < .001). Therefore, the Greenhouse-Geisser correction (ε = 0.722) was applied to the post-event window.

* $p < .10$. ** $p < .05$. *** $p < .01$.

As with the main analysis (Table 27), the robustness check results in Table 30 provided mixed evidence. Notably, the results of the robustness check and the main analysis were generally consistent in their inconsistency. That is, the results agreed at the .05 significance level or better for all event-window types, except for the ANOVA on the standard event window. This indicates that the results of the main analysis were generally robust. To further evaluate their robustness, the above results were followed up with multiple comparison *post hoc* tests. The results are shown in Table 31.

Table 31

Robustness Check of the Model Specification for the Calculation of the $|\overline{CAR}s|$ Underlying the Multiple Comparison post hoc Tests

Comparison		Pre-Event Window $(-2, 0)$			Standard Event Window $(-2, +2)$			Post-Event Window $(+1, n)$		
			p^*			p^*			p^*	
Sample 1	**Sample 2**	$\Delta\|\overline{CAR}\|$	***t*-test**	**Wilcoxon**	$\Delta\|\overline{CAR}\|$	***t*-test**	**Wilcoxon**	$\Delta\|\overline{CAR}\|$	***t*-test**	**Wilcoxon**
CEBS 2010	EBA 2011	-0.48	1.000	1.000	2.25	**.039**	.225	-0.45	1.000	1.000
	EBA 2014	0.10	1.000	1.000	2.56	.051	**.041**	0.71	1.000	1.000
	EBA 2016	-0.28	1.000	1.000	1.81	.880	.346	0.06	1.000	1.000
	EBA 2018	-1.30	.089	1.000	2.02	.236	1.000	1.30	.081	.088
EBA 2011	CEBS 2010	0.48	1.000	1.000	-2.25	**.039**	.225	0.45	1.000	1.000
	EBA 2014	0.58	.689	1.000	0.31	1.000	1.000	1 16	.173	.759
	EBA 2016	0.20	1.000	1.000	-0.44	1.000	1.000	0 51	1.000	1.000
	EBA 2018	-0.82	.911	1.000	-0.24	1.000	1.000	1.75	**.001**	**.004**
EBA 2014	CEBS 2010	-0.10	1.000	1.000	-2.56	.051	**.041**	-0.71	1.000	1.000
	EBA 2011	-0.58	.689	1.000	-0.31	1.000	1.000	-1.16	.173	.759
	EBA 2016	-0.37	1.000	1.000	-0.75	1.000	1.000	-0.65	1.000	1.000
	EBA 2018	-1.39	.129	1.000	-0.54	1.000	1.000	0.60	.115	.759
EBA 2016	CEBS 2010	0.28	1.000	1.000	-1.81	.880	.346	-0.06	1.000	1.000
	EBA 2011	-0.20	1.000	1.000	0.44	1.000	1.000	-0.51	1.000	1.000
	EBA 2014	0.37	1.000	1.000	0.75	1.000	1.000	0.65	1.000	1.000
	EBA 2018	-1.02	.423	1.000	0.20	1.000	1.000	1.25	.071	.088
EBA 2018	CEBS 2010	1.30	.089	1.000	-2.02	.236	1.000	-1.30	.081	.088
	EBA 2011	0.82	.911	1.000	0.24	1.000	1.000	-1.75	**.001**	**.004**
	EBA 2014	1.39	.129	1.000	0.54	1.000	1.000	-0.60	.115	.759
	EBA 2016	1.02	.423	1.000	-0.20	1.000	1.000	-1.25	.071	.088

Note. This table shows the results of a robustness check performed to check the robustness of the results from the main analysis against an alternative specification of the normal return-generating model. Specifically, the table shows the results of all multiple comparison *post hoc* tests based on the average absolute cumulative abnormal returns ($|\overline{CAR}s|$) of the longitudinal sample for each event-window type, with the normal returns used to calculate the underlying absolute cumulative abnormal returns ($|CARs|$) being estimated using the Market Model. The parentheses behind the event-window types specify the number of trading days before and after the event date t_0. The term n in the definition of the post-event window $(+1, n)$ refers to a statistical determination of variable event window lengths based on the Ljung-Box (1978) test introduced in this study. Values in bold indicate statistical significance at the .05 level or better. Bonferroni-corrected p-values (p^*) can be > 1 due to backward correction used by SPSS; such values are therefore shown as p^* = 1.000. $\Delta|\overline{CAR}|$ = difference in $|\overline{CAR}s|$ (Sample 1 – Sample 2). p^* = Bonferroni-corrected p-value (adjustment for multiple comparisons). CEBS = Committee of European Banking Supervisors. EBA = European Banking Authority.

The *post hoc* test results of the robustness check in Table 31 were found to be qualitatively similar to those of the main analysis (Table 28). This was particularly true for the standard and post-event windows, which showed signs of a downward trend in informational value (as in the main analysis). No trend was observed for the pre-event window. The robustness check thus generally confirmed the results of the main analysis and showed that they were robust to an alternative specification of the normal return-generating model.

6.4 Summary

In this chapter, the empirical results of the study were presented. This was done in a way that reflected the two-step analysis process in which an event study was first conducted and then, building on its results, research-question specific analyses were performed. The empirical results of these two steps are summarised below.

The time-series aggregated results of the event study were reported in the form of descriptive statistics for both the cross-sectional and longitudinal samples. That is, the $CARs$ and $|CARs|$ for the different EU-wide stress tests and event-window types. In most cases, the descriptive statistics indicated that the samples were not normally distributed, justifying the use of non-parametric analysis methods. This result was confirmed by formal normality tests (Appendix H). However, in some cases (especially among the $CARs$ of the longitudinal sample) normality could be assumed.[130] For this reason, both parametric and non-parametric methods were used in the subsequent research-question specific analyses to ensure methodological consistency and comparability of the results.

For two of the further analyses (the *Informational Value Hypothesis* and the *Intertemporal Stability Hypothesis*), the $CARs$ and $|CARs|$ from the event study were aggregated across the samples to form mean values, *i.e.* $\overline{CAR}s$ and $|\overline{CAR}s|$. The results of the *Informational Value Hypothesis* (Research Question 1) generally showed statistically significant $\overline{CAR}s$ and $|\overline{CAR}s|$; except for the post-event window, where most of the $|\overline{CAR}s|$ were not statistically significant. This suggested that the disclosure of

[130] It should be emphasized again that many parametric methods (especially *t*-test and ANOVA) are quite robust to deviations from the assumption of normality and can therefore often be used even in the absence of a normally distributed sample, see, for example, Boneau (1960), Cicchitelli (1989), Glass *et al.* (1972), Harwell *et al.* (1992), Posten (1979), Schmider *et al.* (2010).

EU-wide stress test results conveyed valuable new information to bank stock investors in most cases. The robustness checks yielded results that were qualitatively similar to the main analysis, with some limitations in the post-event window. This confirmed that the results of the main analysis were generally robust to an alternative specification of the normal return-generating model.

Analysis of the *Functional Relationship Hypothesis* (Research Question 2) suggested that the relationship between EU-wide stress test results and banks' corresponding abnormal stock returns tended to be quadratic rather than linear. The R^2 of the quadratic fits were consistently higher than those of the linear fits (except for the 2010 EU-wide stress test). Specifically, the quadratic fit has been shown to explain up to 34.4 percent of the total variance in the $CARs$ (compared to 24.5 percent for the linear fit). The results of the regression analysis were confirmed by *MSE* and *RMSE* values. The robustness checks provided moderate but consistent support for the results of the main analysis, regardless of the risk measure used.

The results for the *Intertemporal Stability Hypothesis* (Research Question 3) showed that the $\overline{CAR}s$ observed in response to EU-wide stress test results experienced statistically significant changes over time (omnibus tests), regardless of the event window. The subsequent *post-hoc* tests identified the exact EU-wide stress tests that caused these changes and showed a statistically significant downward trend in the informational value for the post-event window. No such trend could be observed for the pre-event and standard event windows. For the $|\overline{CAR}s|$, the omnibus test results were less consistent, but their collective evidence still indicated statistically significant changes over time for all event-window types. The corresponding *post hoc* tests found some evidence of a statistically significant downward trend in informational value for the standard and post-event windows, while no such evidence could be found for the pre-event window. Overall, this provided mixed evidence for the intertemporal stability of the informational value of EU-wide stress test results. The robustness checks for both $\overline{CAR}s$ and $|\overline{CAR}s|$ yielded results that were qualitatively similar to those of the main analysis, confirming that that they were robust to an alternative specification of the normal return-generating model.

Chapter 7
Discussion

7.1 Introduction

This chapter discusses the empirical results of the study. The discussion starts with the interpretation of the obtained results (Section 7.2). This includes interpreting the meaning of the results and synthesising them with theory and evidence from previous studies. The interpretation is organised to reflect the three research questions of this study, *i.e.* the *Informational Value Hypothesis*, the *Functional Relationship Hypothesis*, and the *Intertemporal Stability Hypothesis*. The interpretation of the results is followed by a discussion of their implications (Section 7.3); this study has important theoretical and practical implications for research, supervisory policy, and investment practice. To complement this chapter, the limitations of the study are discussed (Section 7.4). Finally, this chapter is summarised and concluded (Section 7.5).

7.2 Interpretation of the Results

This section interprets the empirical results of the study presented in Chapter 6. Each of the three research questions of the study is discussed in a separate section. At the beginning of each section, the purpose of the research question and the research approach are outlined to facilitate readability. This is followed by a summary of the results obtained. Furthermore, the meaning of the results is interpreted and synthesised with theory and findings from previous studies. This includes discussing the support or contradiction of the results of this study in relation to the theories tested and the evidence provided by the existing empirical literature. Finally, the main conclusions from each discussion are summarised.

7.2.1 The Informational Value Hypothesis

The purpose of the *Informational Value Hypothesis* (Research Question 1) was to quantify the average informational value of EU-wide stress test results. This implied testing whether the disclosures of EU-wide stress test results have caused statistically significant average (absolute) cumulative abnormal returns ($\overline{CAR}s$ and $|\overline{CAR}s|$) in the cross-section of the sample banks' stocks. The analysis therefore represented a classic test of semi-strong form efficiency.

Accordingly, the null hypothesis for the directional $\overline{CAR}s$ assumed zero abnormal returns, while the alternative hypothesis assumed non-zero abnormal returns. Similarly, the null hypothesis for the non-directional $|\overline{CAR}s|$ assumed abnormal returns equal to the average absolute estimation error of the normal return-generating model used ($|\gamma|$), while the alternative hypothesis assumed abnormal returns unequal to $|\gamma|$. The asset pricing models used to estimate normal (expected) returns were the Fama and French (1993) Three-Factor Model for the main analysis and the Market Model for the robustness checks. The results for the $\overline{CAR}s$ and $|\overline{CAR}s|$ are discussed separately below.

Average Cumulative Abnormal Returns ($\overline{CAR}s$)

The results for the $\overline{CAR}s$ in Table 13 showed statistically significant abnormal returns at the .05 level or better in 11 out of a total of 15 cases (*i.e.* combinations of EU-wide stress tests and event-window types). This suggests that EU-wide stress test results have conveyed valuable new information to investors and have caused affected banks' stock prices to adjust, thus providing support for semi-strong form efficiency. However, this interpretation deserves a closer look and requires some qualifications.

First, the above causal relationship requires, at least for the pre-event window (where four of the five cases were significant), that result information has been leaked to the public prior to official disclosure. A news search for information leaks revealed that while some information on the scenarios and methodology of the EBA 2011 was leaked (Financial Times 2011), no information leaks on EU-wide stress test results could be identified. An alternative explanation might be speculative pre-disclosure positioning by investors.[131] However, the lack of a consistent pattern in the distribution

[131] It has been shown that the forthcoming public disclosure of information can increase incentives for investors to acquire private information prior to disclosure and to position themselves accordingly

of positive and negative $\overline{CAR}s$ does not support this explanation (see also the discussion of $|\overline{CAR}s|$ below). Therefore, the question remains why most $\overline{CAR}s$ in the pre-event window (and arguably also in the standard event window) were statistically significant. As discussed below, the answer to this question is likely to be found in the properties of the $\overline{CAR}s$ measure.

Second, it is difficult to derive any deeper meaning from the results obtained for the $\overline{CAR}s$ across the different event-window types. They showed no discernible pattern in the distribution of significant and non-significant cases, while positive and negative $\overline{CAR}s$ were almost evenly distributed and ranged from -1.50% to +2.54%. In short, the results overall were quite ambiguous and inconclusive (although the robustness checks confirmed that they were robust). This inconclusiveness was also consistent with the collective evidence from previous studies (Table 3 and Table 4), which was similarly ambiguous. For example, for the standard event window – which is best suited for cross-study comparisons – Candelon and Sy (2015) and Georgoutsos and Moratis (2021) have found significant $\overline{CAR}s$, while Cardinali and Nordmark (2011) and Petrella and Resti (2013) have found non-significant $\overline{CAR}s$, for various EU-wide stress tests.

Third, the inconclusiveness of the $\overline{CAR}s$ is in stark contrast to the $|\overline{CAR}s|$, which produced clear and meaningful patterns (discussed below). This provides support for Flannery *et al.* (2017) who argued that the conventional $\overline{CAR}$ measure is inadequate to capture disparate stock price responses because it cannot distinguish between positive and negative information effects. This is also illustrated in the following discussion of the results of the $|\overline{CAR}s|$.

in the market (Demski and Feltham 1994, Kim and Verrecchia 1991, McNichols and Trueman 1994).

Average Absolute Cumulative Abnormal Returns ($|\overline{CARs}|$)

Similar to the $\overline{CAR}s$, the results obtained for the $|\overline{CAR}s|$ showed statistically significant abnormal returns in 10 of the 15 total cases (Table 16). At first glance, this suggests support for semi-strong form efficiency, but again, important caveats need to be made. These are discussed below in connection with the overall results.

In contrast to the $\overline{CAR}s$, the $|\overline{CAR}s|$ showed a very clear pattern in the distribution of significant and non-significant cases. For the pre-event and standard event window, almost all EU-wide stress tests were statistically significant, while almost none of them were statistically significant for the post-event window. The only two exceptions were the EBA 2018 (which was not significant in the pre-event window) and the EBA 2011 (which was significant in the post-event window). Notably, all of the EU-wide stress tests found to be significant in the pre-event and standard event windows were significant at the .01 level in at least one of the two significance tests used (*t*-test and Wilcoxon signed-rank test).

The observed pattern of significant pre-event and standard event windows and non-significant post-event windows is not intuitively explainable. Assuming that the results of EU-wide stress tests have informational value for investors, the opposite would have been expected. Two reasons for the observed pattern are conceivable: information leaks and speculative pre-disclosure positioning by investors. As discussed above, no evidence of information leakage on the results of EU-wide stress tests could be found. Speculative pre-disclosure positioning (which could have led to increased trading activity and thus abnormal returns) could be ruled out by comparing the levels of the $|\overline{CAR}s|$ and the corresponding $|\gamma|$ in Table 16. The comparisons showed for the pre-event and standard event windows that the $|\overline{CAR}s|$ were generally much smaller than the $|\gamma|$, *i.e.* the $|\overline{CAR}s|$ were smaller than the typical absolute estimation errors of the Fama and French (1993) Three-Factor Model. This means that the statistically significant $|\overline{CAR}s|$ in the pre-event and standard event windows indicate *lower* than normal returns. In contrast, $|\overline{CAR}s|$ and $|\gamma|$ were much more similar in the post-event window, indicating that returns were about normal after EU-wide stress test results were disclosed. This could be attributable to a number of different reasons, such as overly lenient stress scenarios (Acharya *et al.* 2014), failure to adequately account for banks' sovereign debt exposures during the European sovereign debt crisis (Blundell-Wignall and Slovik 2010), or an overall lack of credibility of the stress tests (Ong and

Pazarbasioglu 2014). This pattern of significant pre-event and standard event windows and non-significant post-event windows is a surprising finding. Put simply, this finding suggests that investors "held their breath" ahead of the disclosure of EU-wide stress test results, reduced their trading activity below normal levels and resumed normal trading after the results were disclosed.

In summary, the significant $|\overline{CAR}s|$ in the pre-event and standard event windows do not support semi-strong form efficiency (because they are not based on new information) and the generally non-significant $|\overline{CAR}s|$ in the post-event window suggest that the results of EU-wide stress tests were, on average, not particularly informative for investors. These findings confirm the results of Flannery *et al.* (2017) and Georgoutsos and Moratis (2021), who have shown for the US and the EU, respectively, that non-directional $|\overline{CAR}s|$ are better suited than directional $\overline{CAR}s$ to capture disparate stock price responses to supervisory stress test results and thus to produce meaningful results. To the best of the author's knowledge, Georgoutsos and Moratis (2021) is the only other study to date that has used $|\overline{CAR}s|$ in the European context, albeit only for the 2016 and 2018 EU-wide stress tests. Remarkably, they came to very similar conclusions, noting that "the evidence is overwhelmingly in favor of the null hypothesis that the stress tests had no impact" (Georgoutsos and Moratis 2021, p. 993).

Four important conclusions can be drawn from the above discussion. First, this study confirmed the findings of previous studies that the magnitude, direction, and statistical significance of the $\overline{CAR}s$ observed in response to EU-wide stress test results were erratic. Second, the study showed that this was the case even when a systematic and consistent methodology was applied across all EU-wide stress tests available for research. Third, this confirmed the diagnosis of Flannery *et al.* (2017) that the conventional $\overline{CAR}$ measure is inadequate to explain the average price response of a set of stocks to a given event. Fourth, the study revealed meaningful and unexpected patterns based on the analysis of $|\overline{CAR}s|$, indicating that the results of EU-wide stress tests were, on average, not particularly informative for investors, confirming the results of Georgoutsos and Moratis (2021).

It should be noted, with regard to the following discussion of the *Functional Relationship Hypothesis*, that the results above are based on averages across the samples and do not prejudice the bank-level results below.

7.2.2 The Functional Relationship Hypothesis

The *Functional Relationship Hypothesis* (Research Question 2) aimed to determine the shape of the relationship curve between EU-wide stress test results and corresponding abnormal stock returns at bank level. This involved fitting polynomial curves to the empirically observed data, *i.e.* stress test results (units of risk) and abnormal stock returns (units of return), and constituted a test of the mediating effect of rational choice theory and the risk-return tradeoff of investments on the above relationship. The terms "stress test results" and "abnormal stock returns" above represent the variables that were specifically related in the analysis, namely (1) the $CARs$ of the sample banks over the post-event window, and (2) the corresponding capital ratio differences (ΔCRs) between the stressed and actual capital ratios at the end of the fiscal year before the respective stress test (Section 5.3.2.3.1).

More precisely, it was tested whether the stock prices of the affected banks responded proportionally or disproportionately to their stress test results; that is, whether the functional relationship for a given EU-wide stress test can best be described as linear or non-linear. The analysis was constrained to first- and second-degree polynomials to avoid unstable oscillation and to keep the relationship economically interpretable. The null hypothesis assumed a linear relationship between stress test results and abnormal stock returns, while the alternative hypothesis assumed a non-linear relationship. If the null hypothesis was rejected in favour of the alternative hypothesis (*i.e.* if a non-linear quadratic relationship was discovered), further analysis was carried out to determine whether the parabola found opened upwards or downwards. It is important to note, however, that the fitted linear and quadratic functions were not tested against each other, but independently against the actual, empirically observed data. That is, no test for differences between the models was carried out.

The results in Table 19 showed almost consistently for all EU-wide stress tests that the functional relationship between stress test results and corresponding abnormal stock returns can best be described as quadratic rather than linear. More specifically, the relationsip was found to be quadratic at the .05 significance level or better for all EU-wide stress tests except CEBS 2010, which was not significant (however, even for CEBS 2010, R^2 and *MSE* suggest that the relationship was quadratic rather than linear). The discovered non-linear quadratic relationship shows that stress test results and abnormal stock returns have not been proportional to each other, suggesting a mediating effect based on rational choice theory and the risk-return tradeoff of investments. The robustness checks provided moderate but consistent support for the main analysis results, confirming that they were robust to alternative risk measures.

To the best of the author's knowledge, this was the first study ever to analyse the relationship between stress test results and abnormal stock returns at the bank level. Previous studies have typically used simple dichotomous "pass vs. fail" comparisons to describe how the stocks of specific groups of banks responded to their stress test results. Many of them point to a linear proportional relationship. Ahnert *et al.* (2020), for example, showed that passing banks experienced, on average, significantly positive abnormal stock returns of 0.50%, while failing banks experienced significantly negative abnormal stock returns of -1.74% on the results disclosure date. Similarly, Alves *et al.* (2015) found that banks that clearly passed an EU-wide stress test saw stronger positive $CARs$ in their stocks, while banks that narrowly passed the same stress test saw weaker positive $CARs$ in their stocks. For US stress tests, Morgan *et al.* (2014) showed that banks that were found to have larger capital shortfalls experienced more negative abnormal returns. Similarly, Fernandes *et al.* (2020) found that the direction of capital market reactions tended to depend on the nature of the stress test information disclosed (*e.g.* whether banks passed or failed a stress test, or whether announced stress scenarios were more or less severe than expected by the market).

On the other hand, there are several studies that have provided evidence against a linear proportional relationship. Sahin *et al.* (2020) found for the US that some banks' stock prices indreased in response to the disclosure of SCAP results, independent of their stress test result. They also found that the stock prices of some banks that passed the 2014 CCAR decreased in response to the results disclosure. Notably, with respect to the 2011 CCAR – for which no bank-level results were disclosed – no abnormal stock returns were observed, suggesting that there is some stock price formation process associated with disclosures of supervisory stress test results. In the European context, Georgescu *et al.* (2017) showed that abnormal stock returns in response to the 2016 EU-wide stress test results were stronger for banks with weaker stress test results, indicating a non-linear disproportional relationship. Using quantile regressions for the 2016 and 2018 EU-wide stress tests, Georgoutsos and Moratis (2021) showed that Common Equity Tier 1, leverage, and profitability ratios were important determinants of abnormal stock returns and that there was a non-linear relationship between them and the abnormal stock returns observed for a particular subset of banks.

Given the evidence obtained for a non-linear quadratic relationship, the analysis was continued to determine whether the parabolas found opened upwards or downwards. This was important for the economic interpretation of the functional relationship curve, since upward and downward opening parabolas have very different economic consequences. The results of the analysis suggested that the parabolas opened upwards in all five EU-wide stress tests examined (Figure 13), although a possible influence of outliers cannot be completely ruled out.

This U-shaped relationship curve revealed a counterintuitive effect as it implied that banks with *negative* stress test results tended to experience *positive* abnormal stock returns. For example, the latest 2018 EU-wide stress test found that the CET 1 ratios of Allied Irish Banks, Banco de Sabadell, and Bank of Ireland decreased by 6.00, 5.04, and 4.67 percentage points, respectively, under the adverse scenario. However, in response to the results disclosure, the stock prices of Allied Irish Banks, Banco de Sabadell, and Bank of Ireland increased by 1.14%, 0.50%, and 0.96%, respectively. This effect could be explained by the revision of investors' prior risk-return expectations in the light of EU-wide stress test results and suggests that the results for this particular group of banks, while negative, have been better than expected. However, this surprising finding deserves further investigation, possibly including behavioural aspects. It is therefore recommended as an area for future research (Section 8.5).

At the other end of the stress test result spectrum, the U-shaped relationship implied that some banks' capital ratios have *increased* under stress and that their stock prices have risen disproportionally. This was the case, for example, for OTP Bank and PKO Bank Polski, whose Tier 1 ratios increased by 3.00 and 2.40 percentage points, respectively, under the adverse scenario of the 2010 EU-wide stress test; in response, their stock prices rose 2.96% and 3.15%, respectively. The concept of antifragility proposed by Taleb and Douady (2013) offers a possible explanation for why some banks' capital ratios – and thus their stock prices – have risen under stress. Antifragility has already been discussed in connection with supervisory stress testing (*e.g.* Montesi and Papiro 2018, Taleb *et al.* 2012, Taleb and Douady 2013) but certainly requires further investigation. Finally, in line with semi-strong form efficiency expectations, stocks of banks whose results ranked in the middle of the stress test result spectrum (and thus contained the least new information) tended to be the least responsive to the results disclosures.

Five important conclusions emerge from the above discussion. First, this study showed at the bank level that the relationship between EU-wide stress test results and banks' corresponding abnormal stock returns tended to be quadratic rather than linear. Second, this suggests a mediating effect of rational choice theory and the risk-return tradeoff of investments on the above relationship. Third, this implies that disclosure of EU-wide stress test results has improved investors' ability to price-discriminate between banks. Fourth, the obtained parabolas for all five EU-wide stress tests examined opened upwards, representing a U-shaped functional relationship curve (although a possible influence of outliers cannot be ruled out). Fifth, this revealed a counterintuitive effect at the negative end of the stress test result spectrum, implying that banks with negative stress test results tended to experience positive abnormal stock returns.

7.2.3 The Intertemporal Stability Hypothesis

The purpose of the *Intertemporal Stability Hypothesis* (Research Question 3) was to determine the dynamics of the informational value of EU-wide stress test results over time. This was done through a longitudinal analysis of the five EU-wide stress tests available for research and implied a test for the presence of moderating effects due to Goodhart's law. Specifically, omnibus tests were first carried out to analyse whether the informational value ($\overline{CAR}s$ and $|\overline{CAR}s|$) was subject to overall changes across the five EU-wide stress tests. This was followed by multiple comparison *post hoc* tests to identify the exact EU-wide stress tests that caused significant changes.

The null hypothesis assumed that an EU-wide stress test that succeeded another EU-wide stress test in time had a comparatively lower informational value. Accordingly, the alternative hypothesis assumed that an EU-wide stress test that succeeded in time to another EU-wide stress test had an identical or comparatively higher informational value. The results of the multiple comparison *post hoc* tests allowed testing of the hypotheses and deducing whether the informational value has been subject to a statistically significant downward trend. The results for the $\overline{CAR}s$ and $|\overline{CAR}s|$ are discussed separately below.

Average Cumulative Abnormal Returns ($\overline{CARs}$)

The omnibus test results for the $\overline{CARs}$ in Table 22 were statistically significant at the .01 level across all event-window types, indicating that there were significant changes in informational value across the five EU-wide stress tests. The results of the robustness checks were very similar, confirming that the results of the main analysis were robust to an alternative specification of the normal return-generating model. The subsequent *post hoc* tests identified exactly the EU-wide stress tests that caused the significant changes (Table 23). Based on this information, no statistically significant downward trend could be determined for the pre-event and standard event windows.[132] In contrast, a statistically significant decrease in the informational value was found for the post-event window (Figure 14). This is also causally conclusive, since the results of EU-wide stress tests can only have their (full) effect over the post-event window, because the pre-event window and – in parts – also the standard event window cover times before the results are disclosed.

However, the downward trend observed for the post-event window should be interpreted with caution as it is based solely on informational value differences between the CEBS 2010 and subsequent EU-wide stress tests. It is not supported by a "chain" of continuously decreasing informational values from one stress test to the next. Instead, EU-wide stress tests following the CEBS 2010 showed some variation in their $\overline{CARs}$ which, while not statistically significant, is not indicative of a downward trend in informational value.

In addition, a general caveat about the results of the $\overline{CARs}$ is that they are difficult to interpret due to their numerical properties. This is because $\overline{CARs}$ can take positive or negative values, but their sign is not an indication of their informational value, since both large (small) positive and negative $\overline{CARs}$ signify high (low) informational value. The only difference is the direction in which stock prices have moved in response to EU-wide stress test results. This problem is illustrated, for example, in the comparison between the $\overline{CARs}$ of CEBS 2010 and EBA 2011 in the standard event window (see Figure 14). Therefore, additional judgment is required when interpreting the results of the $\overline{CARs}$. This numerical issue does not affect the general accuracy of

[132] While the pre-event window indicated more of an upward trend, in the standard event window the decrease in informational value from CEBS 2010 to EBA 2011 was offset by increases in informational value from EBA 2011 to EBA 2018 and from EBA 2016 to EBA 2018, so that overall no downward trend could be assumed (Figure 14).

the above interpretations. For the $|\overline{CARs}|$, this problem does not exist since they can only be positive due to their absolute values.

Average Absolute Cumulative Abnormal Returns ($|\overline{CARs}|$)

For the $|\overline{CARs}|$, the omnibus test results in Table 27 were less consistent overall, but the collective evidence of the two tests used (one-way repeated measures ANOVAs and one-way Friedman Tests) still showed statistically significant changes across all event-window types. The robustness checks yielded qualitatively similar results, suggesting that the results of the main analysis were generally robust to an alternative specification of the normal return-generating model. The *post hoc* tests did not indicate a statistically significant decrease in the informational value for the pre-event window. However, they found evidence of a statistically significant downward trend in the informational value for the standard and post-event windows (Figure 15). Remarkably, compared to the $\overline{CARs}$, in the *post hoc* analysis of the $|\overline{CARs}|$, there was significantly less agreement between the two tests used (paired-sample *t*-test and paired-sample Wilcoxon signed-rank test). This could indicate a lack of robustness between parametric and non-parametric approaches.

Similar to the $\overline{CARs}$, the downward trend identified for the standard and the post-event window based on the $|\overline{CARs}|$ should be interpreted with caution. Again, the evidence is based solely on informational value differences between the CEBS 2010 or EBA 2011, respectively, and subsequent EU-wide stress tests. Although a decreasing trend can be seen in the standard event window across CEBS 2010, EBA 2011, and EBA 2014, this sequential change was found not to be continuously statistically significant; in addition, the informational value increased again in the two subsequent EU-wide stress tests. For the post-event window, on the other hand, the informational value alternated between increases and decreases across the five EU-wide stress tests, suggesting that the significant differences found have no deeper meaning.

Overall, longitudinal analysis of both $\overline{CARs}$ and $|\overline{CARs}|$ found mixed evidence. The results indicate that the informational value for the pre-event window has been intertemporally stable, while some evidence of a downward trend was found for the standard and post-event windows. However, the evidence was rather weak and should be treated with caution. As a result, support for a moderating effect of Goodhart's law on the relationship between EU-wide stress test results and abnormal stock returns is limited.

To the best of the author's knowledge, this is the first-ever longitudinal analysis of EU-wide stress tests and can therefore only be compared with studies in other locations. The results of this study are broadly consistent with those of Flannery *et al.* (2017), who examined US stress tests from 2009 to 2015 and found that the informational value of their results has been relatively stable. In contrast, several other studies have argued that the informational value of US stress tests has decreased over time (*e.g.* Candelon and Sy 2015, Fernandes *et al.* 2020, Sahin *et al.* 2020). However, any cross-study comparison between US and EU-wide stress tests must consider the different stress test designs, institutional frameworks, economic environments, and the significantly smaller but more stable group of banks subjected to US stress tests.

Four important conclusions can be drawn from the above discussion. First, this study has shown that the methodological disadvantage of $\overline{CARs}$ (compared to $|\overline{CARs}|$) is also relevant to longitudinal analyses. Second, the study found that the informational value was intertemporally stable in the pre-event window (which is causally consistent as the disclosure of EU-wide stress test results is not covered by this event window). Third, evidence of statistically significant downward trends in the informational value of EU-wide stress test results was rather weak. Fourth, this suggests that support for a moderating effect of Goodhart's law on the informational value of EU-wide stress test results is limited.

7.3 Implications of the Results

This study examined three research questions related to banks' abnormal stock returns in response to disclosure of EU-wide stress test results. The theoretical and practical implications of the findings of each of these research questions are discussed below. Specifically, this study has important implications for research, supervisory policy, and investment practice (see Section 1.5). The discussion of implications is organised to reflect the research questions of this study.

The *Informational Value Hypothesis* (Research Question 1) assumed significant (absolute) abnormal returns across three event-window types. Interestingly, it was found that the actual results disclosure (post-event window) was not particularly informative for investors, while the pre-event and standard event windows showed statistically significant abnormal returns. Notably, however, these abnormal returns were significant because they were abnormally *low*, suggesting that investors "held their breath" before the results were disclosed.

The underlying event study emphasised the importance of a systematic model selection procedure (for the normal return-generating model) and the choice of an appropriate abnormal return measure. Regarding the latter, this study supports the few existing studies (Flannery *et al.* 2017, Georgoutsos and Moratis 2021) that compared the performance of conventional $\overline{CAR}s$ and innovative $|\overline{CAR}s|$. Like the previous studies, this study found that $|\overline{CAR}s|$ are better suited than $\overline{CAR}s$ to capture disparate stock price responses, leading to more meaningful results and an overall better understanding of the phenomena under study. Therefore, future researchers should consider implementing an appropriate model selection procedure and making an informed choice of the abnormal return measure used when capturing positive and negative information effects at the same time.

A more practical implication of the findings of this research question concerns EBA's stress test results disclosure policy. The results for this research question have shown that prior to the disclosure of EU-wide stress test results, stock market tensions have led to abnormally low trading activity and thus potentially reduced liquidity in bank stocks. To counteract this effect, the EBA should complement its considerable post-disclosure transparency efforts (*e.g.* bank-level stress test reports, press releases, analyst presentations, databases, and interactive tools) with means to signal to the market the overall tendency of the stress test results, possibly at an aggregated level across all affected banks (so as not to pre-empt bank-level disclosure). This approach would also be consistent with previous recommendations for increased use of disclosure of aggregated result information (*e.g.* Goldstein and Sapra 2014, Schuermann 2014).

The *Functional Relationship Hypothesis* (Research Question 2) assumed a non-linear relationship between EU-wide stress test results and abnormal stock returns of affected banks. This hypothesis was confirmed, revealing a counterintuitive U-shaped functional relationship curve, implying that the stocks of banks whose stress test results were at either end of the stress test result spectrum tended to experience disproportionately positive abnormal returns.

This bank-level evidence showed that there is merit in going beyond the commonly used dichotomous "pass vs. fail" approach when disaggregating the average intervention effect. The findings suggest that there is a mediating effect of rational choice theory and the risk-return tradeoff of investments, causing investors to revise their prior risk-return expectations about banks such that the characteristic U-shaped relationship curve is created. Building on this, future research should therefore further examine the stock price formation processes occurring at the extreme ends of the stress test result spectrum. Consequently, this is one of the areas recommended for future research (Section 8.5), possibly considering behavioural aspects.

From a more practical perspective, the EBA and national competent authorities should be aware of the U-shaped functional relationship between EU-wide stress test results and the corresponding abnormal stock returns. First, because it demonstrates that stress test results have helped investors better price-discriminate between banks. In other words, it shows that the EU-wide stress tests examined have met their market discipline objective. Second, it implies that the EBA and national competent authorities had no reason to be unduly concerned about negative abnormal stock returns in response to the disclosure of EU-wide stress test results. This also includes possible consequences that could follow from negative stock price reactions, such as short selling or bank runs. Notwithstanding, the EBA and national competent authorities should remain vigilant in this regard.

The U-shaped functional relationship also presents an opportunity for investors to develop opportunistic or event-driven investment strategies. This could involve, for example, buying or overweighting bank stocks expected to experience stress test results on the positive or negative end of the stress test result spectrum, and selling or underweight bank stocks expected to experience stress test results in the middle of the stress test result spectrum.

The *Intertemporal Stability Hypothesis* (Research Question 3) assumed that the informational value of EU-wide stress test results was intertemporally stable across the five EU-wide stress tests available for research. For the pre-event window, the informational value was confirmed to be stable over time, while for the standard event window and the post-event window, weak evidence of a statistically significant downward trend was found.

These findings suggest that there has been some, but limited, evidence of an undesirable moderating effect on the relationship between EU-wide stress test results and banks' corresponding abnormal stock returns due to Goodhart's law. In contrast to US stress tests, for which most studies have found a downward trend in informational value (Candelon and Sy 2015, Fernandes *et al.* 2020, Sahin *et al.* 2020), this means that EU-wide stress tests have consistently provided relatively stable and reliable results. This finding is reassuring for investors who have relied on and acted on the results of EU-wide stress tests over the past decade.

This finding is also important information for the EBA as it suggests that the Quality Assurance Process (QAP) and overall methodology of EU-wide stress tests has worked reasonably well in avoiding perverse incentives for banks to "to game the system". Nevertheless, the evidence found gives reason for the EBA to remain alert. Developments in future EU-wide stress tests should be followed and critically examined by research.

7.4 Limitations of the Study

Despite considerable efforts to minimise threats to the validity and reliability of this study (*e.g.* research design, model selection, confounding control) and appropriate robustness checks, the results are subject to some limitations that should be considered. There are three main limitations to this study. They relate to the research strategy, the data set, and the potential influence of extraneous variables and market anomalies that could not be controlled.

First, the nature of EU-wide stress tests as sovereign public interventions justified the use of a quasi-natural experimental research strategy. An inherent disadvantage of this strategy is that the intervention is not randomly assigned and controlled by a force other than the researcher. In the context of this study, this meant that the CEBS and the EBA applied a size-based selection rule to select banks into EU-wide stress tests. This introduced a size bias in the research population, which was mitigated by using the Fama and French (1993) Three-Factor Model, which explicitly accounts for firm size as one of the asset pricing factors. Furthermore, selection by the CEBS and the EBA meant that the research population was fixed and finite. It was also relatively small, ranging from $N = 48$ to $N = 123$. Therefore, a purposive non-probability sampling method (census sampling) was used, which limits the representativeness of the samples and thus the external validity of the study. A general limitation of the study is therefore that the results cannot be generalised beyond the samples.

Second, since the interventions were controlled by the CEBS and the EBA, the validity of the data (stressed capital ratios representing the stress test results) could not be verified. Any data quality issues would limit the validity of the results of this study. However, since the stress test results are subject to multiple checks as part of the quality assurance process (QAP) prior to disclosure (Figure 4), it seems reasonable to assume that the data was not subject to any major quality issues. In addition to data quality, several unavoidable effects threaten the external validity and comparability of the results obtained. Specifically, these are attrition, maturation, and instrumentation effects, which are explained in more detail below.

Attrition: Over the course of the study period (2010 to 2018), several banks had to be removed from the samples due to delistings (mergers, takeovers, and nationalisations). This was particularly pronounced during the study period as a result of the recent global financial crisis. The delistings created a survivorship bias in the remaining sample banks, which must be taken into account when comparing the results for different EU-wide stress tests. However, this limitation only affects the results of the cross-sectional analyses (Research Questions 1 and 2) and not those of the longitudinal analysis (Research Question 3), which was based on panel data.

Maturation: Furthermore, changes in banks' capitalisation and asset structure over time also pose a threat to the comparability of study results between various EU-wide stress tests. These dynamics could not be controlled as banking operations continued between the different EU-wide stress tests. This limitation affects the results of both cross-sectional and longitudinal analyses.

Instrumentation: Similarly, variations in the scope and severity of stress scenarios and other details of EU-wide stress tests (*e.g.* different capital ratio threshold levels) affect comparability. Although the various stress tests are generally equivalent, they are not exactly the same. Therefore, the observed abnormal stock returns may reflect differences in the design of the different EU-wide stress tests. With regard to Research Question 2 (*The Functional Relationship Hypothesis*), it should also be mentioned that the analyses performed were constrained to first- and second-degree polynomials, to define the bounds of the permissible results.

Third, despite extensive controls, some extraneous variables and market anomalies could not be controlled for, thus possibly affecting the results of this study. These include factors related to the global financial crisis (*e.g.* government bank bailouts), subsequent macroeconomic events (*e.g.* European sovereign debt crisis), and known market anomalies. Among the market anomalies, two effects might be particularly important: the weekend effect and the day-of-the week effect. This is because the results of EU-wide stress tests (with the exception of EBA 2014) have always been disclosed on the last day of a trading week (Table 7).[133] These effects could therefore influence the results of this study and thus threat its internal validity. Analysing the impact of these effects was beyond the scope of this study and is left to future research.

Finally, it should be mentioned that, in principle, an alternative research paradigm could have been used to conduct this study, namely interpretivism. This could potentially have led to deeper insights into the meaning and motivation of investor behaviour and to higher content validity of the findings. However, this would have fundamentally changed the character of the study and might have led to less reliable results due to the validity-reliability tradeoff.

Overall, the systematic design and thorough conduct of this study give confidence in its results. However, future research should address the above limitations wherever possible. This could include, for example, extensions of the existing data set and cross-study comparisons with the results of this study. Specific recommendations for areas of future research are provided in Section 8.5.

7.5 Summary

In this chapter, the empirical results of the study were discussed. This involved interpreting the meaning of the results and synthesising them with theory and evidence from previous studies. The analysis of the *Informational Value Hypothesis* yielded mixed results that differed depending on the abnormal return measure used. Consistent with previous studies, the conventional (directional) $\overline{CAR}$ measure found inconclusive results across all event-window types. In contrast, the innovative (non-directional) $|\overline{CAR}|$ measure identified clear and meaningful patterns for each event-window type,

[133] Chen and Singal (2003), for example, found a systematic positive abnormal return on Fridays.

suggesting that investors were under tension in the pre-event and standard event windows, while actual disclosure of EU-wide stress test results was not particularly informative for investors (post-event window). This was consistent with the findings of Georgoutsos and Moratis (2021), the only other study to date that has used $|\overline{CARs}|$ in the European context.

The *Functional Relationship Hypothesis* suggested that the relationship between EU-wide stress test results and the bank's corresponding abnormal stock returns was quadratic rather than linear, revealing a counterintuitive U-shaped relationship. This finding contradicted several previous studies that have suggested a linear proportional relationship between stress test results and abnormal returns. However, it provided support for several other studies whose results have indicated a more non-linear relationship. The evidence of this study suggests a mediating effect based on rational choice theory and the risk-return tradeoff of investments.

The *Intertemporal Stability Hypothesis* found weak evidence for a downward trend in the informational value of EU-wide stress test results for the standard and post-event windows, but not for the pre-event window. These findings broadly agreed with those of Flannery *et al.* (2017) for US stress tests. However, they contrasted with several other studies that have found a decreasing trend in the informational value of US stress test results. The evidence presented in this study provided some, but limited, support for the presence of a moderating effect of Goodhart's law on the informational value of EU-wide stress test results.

The above interpretations of the results have been complemented by a discussion of their theoretical and practical implications for research, supervisory policy, and investment practice. Finally, the limitations of the study were discussed. They mainly relate to the research strategy, the data set, and the potential influence of extraneous variables and market anomalies that could not be controlled.

Chapter 8
Summary and Conclusions

8.1 Introduction

This chapter summarises and reflects on the research process of this study. It starts with a summary of the main research steps (Section 8.2) and continues with the key findings of the study (Section 8.3). In addition to novel and unexpected insights, this also includes a clear and definitive answer to each of the research questions. This is followed by a detailed presentation of the contributions of this study (Section 8.4). The study made several contributions, including contributions to theory, methodology and methods, and supervisory policy and investment practice. Finally, recommendations for future research are made in four key areas (Section 8.5). This includes extensions of the existing data set, analyses of behavioural aspects, further investigations of the observed U-shaped mediating effect, and applications of the theoretical and methodological contributions of this study, including the theoretical framework.

8.2 Summary of the Research

The purpose of this study was to improve the understanding of banks' abnormal stock returns in response to the results of the five EU-wide stress tests conducted by the CEBS and the EBA between 2010 and 2018. Three research questions were formulated to address significant gaps in the existing literature. They focused on (1) quantifying the informational value of EU-wide stress test results for bank stock prices, (2) determining the functional relationship between stress test results and the corresponding abnormal stock returns, and (3) examining the intertemporal stability of the informational value of EU-wide stress test results across the various exercises.

Based on the literature review, a dedicated theoretical framework was developed, building on and connecting the following theoretical elements: bank opacity, information uncertainty, the risk-return tradeoff of investments, rational choice theory, Goodhart's law, and the efficient market hypothesis. This framework provided the theoretical basis for answering the research questions and facilitated the formulation of empirically testable hypotheses.

The study was conducted on five cross-sectional samples ($n = 33$ to $n = 59$) and one longitudinal sample ($n = 28$) of banks subjected to EU-wide stress tests. In total, the cross-sectional samples consisted of 227 bank-year observations and the longitudinal sample consisted of 140 bank-year observations. The required stock prices and capital ratios of the sample banks were collected from Bloomberg and from the official bank-level result reports published by the CEBS and the EBA using structured direct observation.

Methodologically, the study took an objectivist ontological and an empirical-positivist epistemological view and pursued a quasi-natural experimental strategy. The strategy was operationalised in a two-step process. First, an event study was performed to obtain the sample banks' (absolute) cumulative abnormal returns ($CARs$ or $|CARs|$) over three different event-window types. This step involved several advancements and extensions to the widely used standard event study approach of Campbell *et al.* (1997) and MacKinlay (1997). Second, research-question specific analyses were performed on the $CARs$ and $|CARs|$ to test the hypotheses and answer the research questions using a variety of statistical methods.

The obtained empirical results were thoroughly discussed. This involved interpreting their meaning and synthesising them with theory and evidence from previous studies, yielding important and unexpected new insights. The key findings of this study are summarised in the next section. In addition, the implications and limitations of this study were discussed. The findings of the study have important implications for research, supervisory policy, and investment practice. The main limitation is that the results cannot be generalised beyond the samples due to the purposive non-probability sampling method used. Finally, the contributions of this study were presented and four key areas for future research were recommended (Sections 8.4 and 8.5).

8.3 Summary of the Key Findings

The aim of this study was to examine banks' abnormal stock returns in response to the results of the five EU-wide stress tests conducted by the CEBS and the EBA between 2010 and 2018. Based on gaps in the existing literature, the study addressed three complementary research questions on (1) the average intervention effect, (2) the relationship between stress test results and abnormal stock returns at the bank level, and (3) the dynamics or intertemporal stability of the average intervention effect across the five EU-wide stress tests examined. In the following, the research questions are restated to facilitate a more coherent reading. A clear and definitive answer is given for each of the research questions.

Research Question 1: *What is the average value of the information contained in the results of EU-wide stress tests measured in terms of abnormal stock returns?*

Based on the results of an event study and subsequent quantitative analyses, it can be concluded that the average informational value of EU-wide stress test results ranges between -1.50% and +2.54% ($\overline{CARs}$) and between +1.06% and +4.36% ($|\overline{CARs}|$), across all event-window types.[134] Thus, the disclosure of EU-wide stress test results generally provided investors with valuable new information, which in most cases was economically significant.

However, in terms of statistical significance, the evidence was less clear. Consistent with previous studies, the analysis of the $\overline{CARs}$ yielded mixed evidence for all event-window types, making it difficult to draw any meaningful conclusions. In contrast, the results for the $|\overline{CARs}|$ revealed clear and meaningful patterns for each event-window type. For the pre-event window and the standard event window, highly significant $|\overline{CARs}|$ were found almost consistently. However, closer analysis showed that all significant $|\overline{CARs}|$ were significant because they were abnormally small. This suggests that prior to the disclosure of EU-wide stress test results, investors "held their breath" by reducing their normal trading activity. For the post-event window, however, almost no statistically significant $|\overline{CARs}|$ could be detected, indicating that the results of EU-wide stress tests were not particularly informative for investors in the cross-section of the samples.

[134] For details, see Tables 14 and 17.

Research Question 2: *What is the functional relationship between new information from EU-wide stress test results and corresponding abnormal stock returns?*

Based on the results of a curve-fitting procedure and subsequent quantitative analyses, it can be concluded that the relationship between EU-wide stress test results and the corresponding abnormal stock returns tended to be quadratic rather than linear. The evidence indicated non-linear quadratic relationships for all EU-wide stress tests examined at the .05 level of significance (except for CEBS 2010). Notably, the parabolas found for all EU-wide stress tests (including CEBS 2010) opened upwards, revealing a counterintuitive U-shaped relationship curve. This implied that the stock prices of banks whose stress test results were at either end of the stress test result spectrum experienced disproportionally high positive abnormal returns. This means that both banks with particularly positive *and* negative stress test results tended to experience particularly positive abnormal stock returns.

However, it is important to note that the linear and quadratic fitted functions were not tested against each other, but independently against the actual, empirically observed data to evaluate the quality of the approximations. That is, no test for differences between the models was carried out. In addition, with regard to the upward-opening parabolas found, it should be noted that a possible influence of outliers cannot be completely ruled out.

Research Question 3: *How has the informational value of EU-wide stress test results, measured in abnormal stock returns, changed over time?*

Based on the results of a longitudinal analysis, it can be concluded that the informational value of EU-wide stress test results has remained relatively stable over time. However, some differences could be observed for the two measures of informational value used ($\overline{CARs}$ and $|\overline{CARs}|$), as well as for the different event-window types.

The analysis of the $\overline{CARs}$ revealed weak evidence of a statistically significant downward trend in the informational value for the post-event window. No such evidence could be found for the pre-event and standard event windows. Similarly, the $|\overline{CARs}|$ showed no signs of a downward trend for the pre-event window, while some evidence of a decrease in informational value was found for the standard and the post-event windows. Overall, the evidence above should be interpreted with caution as it is not supported by a "chain" of continuously decreasing informational values from one stress test to the next, but rather represents informational value differences between individual EU-wide stress tests.

8.4 Contribution of the Study

This study's original contribution to knowledge are more comprehensive insights into the informational value of EU-wide stress test results for bank stock pricing. While most of its contributions are empirical in nature, the study also contributes to the body of knowledge in a variety of other ways. These include contributions to theory (Section 8.4.1), methodology and methods (Section 8.4.2), and supervisory and investment practice (Section 8.4.3). The empirical contributions have been presented and discussed in detail in Chapter 6 and Chapter 7 (see also the summary of key findings above) and are not repeated here for the sake of brevity. However, some of the main empirical findings are highlighted in the following sections where appropriate.

8.4.1 Contribution to Theory

This study addressed several gaps in the existing literature on bank's abnormal stock returns in response to EU-wide stress test results. As previous studies in this field have been fairly silent on the theory on which their research is based, this study was able to make several important theoretical contributions.

First, due to the lack of theoretical foundations in the existing literature, this study was able to contribute an inventory of theories, constructs, and debates relevant to the field. This involved identifying, reviewing, and synthesising the theoretical elements that might help explain the boundary conditions and process by which the disclosure of supervisory information affects banks' stock prices (Chapter 3 and Chapter 4). Besides the efficient market hypothesis, which has been implicitly tested by all previous studies without reference to it,[135] the final inventory also includes: bank opacity, information uncertainty, rational choice theory, the risk-return tradeoff of investments, and Goodhart's law.

[135] It should be noted that Ahnert *et al.* (2020) and Alves *et al.* (2015) have briefly mentioned the efficient market hypothesis but did not elaborate on it further.

Second, building on the above inventory, this study developed a dedicated theoretical framework for studying abnormal bank stock returns in response to supervisory transparency measures (Section 4.2). The framework integrates the relevant theoretical elements into the efficient market hypothesis as the key formal theory. More specifically, bank opacity and information uncertainty were introduced as important antecedents of informationally efficient stock prices to explain the impact of supervisory information disclosures (*e.g.* EU-wide stress test results) on bank stock prices. In addition, rational choice theory and the risk-return tradeoff of investments have been established as mediators to explain the process by which new supervisory information affects bank stock prices. Similarly, Goodhart's law was introduced as a moderator that could affect the relationship between recurring disclosures of supervisory information and corresponding abnormal bank stock returns over time; thereby responding to calls by Kok *et al.* (2019) and Quagliariello (2020) for more research on the effects of perverse incentives. Although the framework was developed for the specific needs of this study, it was deliberately designed to be applicable and extensible for future research. This study thus significantly extends and advances the theoretical concepts available to study abnormal bank stock returns in response to supervisory transparency measures.

Third, to the best of the author's knowledge, this is the first study ever to theoretically link and empirically test the mediating and moderating effects proposed in the theoretical framework. Therefore, this study extends the existing literature, which has typically relied solely on classic semi-strong form efficiency tests, to examine and understand the relationship between disclosure of supervisory information and abnormal returns in bank stocks. An important theoretical and empirical contribution of this study is therefore to improve the understanding of the process by which the disclosure of supervisory information affects banks' stock prices. The results of the empirical analysis revealed a counterintuitive U-shaped mediating effect of rational choice theory and the risk-return tradeoff of investments on the relationship between EU-wide stress test results and banks' corresponding abnormal stock returns. Equally revealing was the fact that only weak evidence of a moderating effect of Goodhart's law could be found, suggesting that the informational value of EU-wide stress test results has remained relatively stable over time.

8.4.2 Contribution to Methodology and Methods

Using the efficient market hypothesis as the key formal theory justified conducting the empirical analysis based on event study methods. However, event studies are associated with a number of problems (see also Section 3.5.3) that have been largely ignored in previous studies. By addressing these problems, this study made several significant contributions to methodology and methods.

First, the main problem with event studies is that market efficiency *per se* is not testable and must always be tested jointly with an asset pricing model (joint-hypothesis problem, also known as “bad-model problem”). Consequently, the asset pricing model used to estimate normal (expected) returns is critical to the internal validity of any event study. Despite this, previous studies have typically chosen their asset pricing models arbitrarily or resorted to the convenient Market Model (Table 3). In contrast, this study addressed the joint-hypothesis problem by introducing a systematic model selection procedure based on the goodness-of-fit of a set of candidate models. This approach provides a statistical basis for identifying and selecting the most appropriate asset pricing model, helping researchers to make informed methodological choices. This study thus contributes to the advancement and extension of the widely used standard event study approach of Campbell *et al.* (1997) and MacKinlay (1997), which does not provide for any model selection. Based on the model selection results, this study used the Fama and French (1993) Three-Factor Model to estimate normal (expected) returns. To the best of the author’s knowledge, this was the first time ever that this particular model was used in the context of EU-wide stress tests, thus counteracting the prevailing model monoculture.

Second, the validity of event study results can be affected by factors other than those being studied. This study conceptualised and implemented a coherent approach to control for such confounding factors, a step often omitted by previous studies. The confounding control proposed in this study operates at multiple levels and includes design, measurement, and logic controls (Section 5.3.1.3). The design and measurement controls are closely related to the implementation of the systematic model selection and the construction of the candidate models, while the logic control aims to eliminate known extraneous factors. These controls helped reduce the risk of alternative explanations for the results obtained, thereby increasing their internal validity. The proposed approach is easily transferable to other research projects. This study thus makes a contribution to the research design of event studies and other quasi-experimental designs.

Third, another problem area of event studies is determining the appropriate length of event windows. This is important to find the right balance between capturing enough signal and sampling too much noise. The existing event study literature is dominated by fixed event windows that are set at the discretion of the researcher and are therefore prone to subjectivity and bias. This study provides a new method to statistically determine the appropriate event window length based on serial correlation and a recontextualisation of the Ljung-Box (1978) test (Section 5.3.2.1.1). Starting from the event date, this method determines the end of an event window as the trading day when the serial correlation is no longer significant (*i.e.* the day when the stock price has fully incorporated the event's initial price signal). This is an objective approach that also allows the definition of variable event window lengths for each individual observation. The method proposed in this study thus helps to reduce subjectivity and bias and complements the few other methods available.[136] A disadvantage of this method is, however, that it cannot be used to determine the length of pre-event windows (since it relies on the event's initial price signal).

Fourth, this is one of the first studies to apply the $|\overline{CARs}|$ measure proposed by Flannery *et al.* (2017) in a European context. More specifically, to the best of the author's knowledge, there is only one other study (Georgoutsos and Moratis 2021) that has used this measure to examine abnormal bank stock returns in response to EU-wide stress tests. While Georgoutsos and Moratis' (2021) study was limited to just two stress tests, this study determined the $|\overline{CARs}|$ (and conventional $\overline{CARs}$) for all five EU-wide stress tests available for research. Flannery *et al.* (2017) argued that $|\overline{CARs}|$ should be preferred over $\overline{CARs}$ because conventional event studies are conceptually flawed and thus cannot distinguish between positive and negative information effects. This study contributes to the methodological literature by providing a test and comparison of these two measures in the European context, based on a reasonably large number of events. The empirical results confirmed that $|\overline{CARs}|$ are indeed better than $\overline{CARs}$ at capturing disparate stock price responses, allowing for deeper insights and better understanding of the observed phenomena.

[136] For other methods, see De Franco *et al.* (2007), Krivin *et al.* (1997), and Lins *et al.* (2013).

Fifth, based on the event study results, this study developed two distinct methods for investigating the mediating and moderating effects proposed in the theoretical framework. In addition to the corresponding theoretical contributions (Section 8.4.1), the operationalisation of these problems contributes to the development of methods. More specifically, to the best of the author's knowledge, this is the first study ever to conduct a longitudinal analysis of EU-wide stress tests. Repeated measures ANOVAs and Friedman tests were used to examine the intertemporal stability of the informational value of EU-wide stress test results (moderator analysis). This fills a gap in the existing literature and adds a new perspective to the common cross-sectional approach. Similarly, previous studies have typically assessed the relationship between EU-wide stress test results and banks' corresponding abnormal stock returns based on two-group ("pass vs. fail") comparisons. This study went beyond this simple dichotomous approach by introducing a bank-level functional relationship analysis based on polynomial curve fitting (mediator analysis). However, it should be noted that the fitted functions were not tested against each other, but independently against the actual, empirically observed data. The empirical results nonetheless provided more nuanced insights that contribute to the literature and have important practical implications for supervisors and investors.

8.4.3 Contribution to Supervisory Policy and Investment Practice

The empirical findings of this study contribute to the development of supervisory policy and investment practice in several ways. These practical contributions relate to three broad areas: the functioning of EU-wide stress tests, EBA's stress test results disclosure policy, and investment opportunities in the context of EU-wide stress tests.

First, to the best of the author's knowledge, this study provided the first-ever longitudinal analysis of EU-wide stress tests. Its findings provide the EBA and national competent authorities with general reassurance that the methodology and Quality Assurance Process (QAP) of EU-wide stress tests worked reasonably well. Despite some weak evidence of decreasing informational value over time, no concrete evidence of bank manipulation attempts or perverse incentives could be found in EU-wide stress tests. However, the evidence uncovered in this study, combined with more worrying findings from previous studies of US stress tests (Candelon and Sy 2015, Fernandes *et al.* 2020, Sahin *et al.* 2020), still gives the EBA reason to remain alert. This study thus contributes to assessing the functioning of EU-wide stress tests and to better understanding the reliability of their results.

Second, the study provides some impetus for the further development and refinement of EBA's stress test results disclosure policy. It thus contributes to the ongoing debate on bank supervisors' disclosure policies in relation to supervisory stress test results (see, for example, Faria-e-Castro *et al.* 2017, Goldstein and Leitner 2018, and Pacicco *et al.* 2020). The impetus of this study is based on several considerations, which are explained separately below. On the one hand, this study did not find any unintended negative effects of disclosures of EU-wide stress test results from a stock pricing perspective. This should give the EBA confidence to continue its established practice of detailed disclosure (despite recent proposals to reduce supervisory information disclosure; see Goldstein and Leitner 2018 and Goldstein and Yang 2019).

On the other hand, the results of the $|\overline{CAR}s|$ for the post-event window – which are consistent with those of Georgoutsos and Moratis (2021) – raise the question of why the absolute abnormal stock returns of the affected banks were not more significant. If the results of EU-wide stress tests have not been particularly informative for investors, as the evidence suggests, then there may be structural reasons. Possible reasons include overly lenient stress scenarios (Acharya *et al.* 2014), potential credibility issues (Ong and Pazarbasioglu 2014), or the EBA's timing of results disclosure at the end of a trading week after close of trading (Table 7). This question should be analysed by the EBA and addressed accordingly; it also represents a possible avenue for future research.

Finally, the finding from the analysis of the $|\overline{CAR}s|$ that investors reduced their trading activity in the pre-event and standard event windows provided new insights with clear implications for EBA's stress test results disclosure policy.[137] To mitigate potential negative consequences for the liquidity of bank stocks during these periods, the EBA should consider counteracting the effect of investors' "holding their breath" by signalling the overall stress test results to the market. A possible solution to this

[137] Remarkably, Morris and Shin (2002, p. 1532) noted in this regard that "[t]he challenge for central banks and other public organizations is to strike the right balance between providing timely and frequent information to the private sector so as to allow it to pursue its goals, but to recognize the inherent limitations in any disclosure and to guard against the potential damage done by noise. This is a difficult balancing act at the best of times, but this task is likely to become even harder. As central banks' activities impinge more and more on the actions of market participants, the latter have reciprocated by stepping up their surveillance of central banks' activities and pronouncements."

problem would be advance disclosure of EU-wide stress test results at an aggregated level for all affected banks. This would address the problem without pre-empting the disclosure of results at the bank level, and would also be in line with previous recommendations for increased disclosure of aggregated stress test result information (see, for example, Goldstein and Sapra 2014 and Schuermann 2014).

Third, the bank-level functional relationship analysis of this study provided a more sophisticated understanding of the relationship between EU-wide stress test results and banks' corresponding abnormal stock returns. The EBA and national competent authorities should take note of the discovered U-shaped relationship as it suggests that the results of EU-wide stress tests have helped investors to revise their prior risk-return expectations and better price-discriminate between banks. It thus indicates that the EBA has achieved the market discipline objective inherent in EU-wide stress tests. The functional relationship curve found in this study also implies that the EBA and national competent authorities had no reason to be unduly concerned about negative abnormal stock returns in response to EU-wide stress test results. However, supervisory authorities should remain vigilant in this regard. This study thus contributes to a more nuanced understanding of how investors use new, formerly confidential, supervisory information to price bank stocks.

The discovered U-shaped functional relationship also offers investors the opportunity to learn and benefit from potentially recurring stock price patterns in response to the disclosure of EU-wide stress test results. For example, they could develop opportunistic or event-driven investment strategies specifically targeting the results disclosure events of EU-wide stress tests. This could involve buying or overweighting bank stocks that are expected to experience stress test results on the positive or negative end of the stress test result spectrum and selling or underweighting bank stocks that are expected to experience stress test results in the middle of the stress test result spectrum. This study thus also contributes to investment practice by showing the dynamics of stock markets and the resulting investment opportunities.

8.5 Recommendations for Future Research

This study has provided new insights into banks' abnormal stock returns in response to EU-wide stress test results. The insights gained are comprehensive in that they include cross-sectional, longitudinal, and bank-level perspectives. Especially the latter two perspectives have provided novel and unprecedented insights. However, this study also has some limitations (Section 7.4) that should be addressed by future research. At the same time, the study lends itself to extensions and opens up new avenues for further investigations. Although a wide range of research is conceivable based on this study, four key areas are recommended for future research.

Almost any empirical study can be extended by adding new data as and when it becomes available. An obvious area for future research is therefore the inclusion of new EU-wide stress tests to challenge the consistency of the results obtained. This could be particularly useful for longitudinal analysis. However, it is recommended to go beyond simply extending the study period, for example by identifying sub-periods characterised by a more stable economic environment. In this way, the potential confounding effects of the European sovereign debt crisis and other factors could be eliminated. The European volatility index (VSTOXX) and the EURIBOR-OIS spread are common market measures that could be used to identify such sub-periods.[138] It might also be instructive to examine the impact of analyst coverage on the magnitude and direction of banks' abnormal stock returns in response to the disclosure of EU-wide stress test results. The I/B/E/S database could be a useful data source in this regard. Future research may also use additional data types such as stock trading volumes or bond and CDS prices to build on and enrich the findings of this study.

Another area for future research could be to extend the analysis to behavioural elements such as investor sentiment and risk perception, which may provide alternative explanations for the abnormal stock returns observed in this study. This would require qualitative research in order to collect and analyse data based on investor surveys or interviews. Alternatively, publicly available data such as the Sentix sentiment index could be used to enhance understanding of how behavioural aspects influence

[138] The EURIBOR-OIS spread is the spread between the Euro Interbank Offered Rate (EURIBOR) and the Euro Overnight Index Average (EONIA) as the corresponding overnight rate swapped to the same maturity, *i.e.* the Overnight Index Swap (OIS). The EURIBOR-OIS spread is considered an important measure of tension in the European interbank market (Heider *et al.* 2015, Taboga 2014, Taylor and Williams 2009).

human behaviour and decision-making in the stock market. Prospect theory (Kahneman and Tversky 1979) and the adaptive markets hypothesis (Lo 2004) can provide useful theoretical foundations for this type of analysis. Future research in this direction could help delve deeper into the underlying reasons for investor behaviour that go beyond the scope of efficient markets and rational expectations.

Given the surprising finding of disproportionately positive abnormal stock returns for banks whose stress test results fall at either end of the spectrum, further research is needed to explain this counterintuitive effect. The concept of antifragility (Taleb and Douady 2013) might explain why some banks' capital ratios – and consequently their stock prices – have increased under stress. More puzzling, however, is the reason why the stock prices of banks that performed poorly in the stress tests have also increased. It seems that at least part of the answer lies in the revision of investors' prior risk-return expectations in light of the stress test results. An interesting avenue for future research would therefore be to try to explain the observed non-linear risk-return relationship with a behavioural approach (possibly based on the suggestions above). Alternatively, more practical research could develop and test opportunistic or event-driven investment strategies that exploit the recurring non-linear return patterns at the extreme ends of the stress test result spectrum.

Finally, future research could build on the theoretical and methodological contributions of this study and apply them to related contexts where supervisory transparency measures are carried out and the results are made public. Examples include the regular EU-wide transparency exercises of the EBA, the Comprehensive Assessments of the ECB, and the insurance stress tests of the European Insurance and Occupational Pensions Authority (EIOPA). Together with the findings of this study, such research could provide a wider perspective that would allow for broader generalisations of the observed phenomena. Furthermore, researchers can take advantage of the theoretical framework developed in this study, which was deliberately designed to be extensible, thus opening up multiple avenues for future research.

References

Acharya, V.V., Engle, R., and Pierret, D., (2014). Testing macroprudential stress tests: The risk of regulatory risk weights. *Journal of Monetary Economics*, 65, pp. 36-53.

Acharya, V.V. and Thakor, A.V. (2016)., The dark side of liquidity creation: Leverage and systemic risk. *Journal of Financial Intermediation*, 28, pp. 4-21.

Adrian, T., Morsink, J., and Schumacher, L., (2020). *Stress Testing at the IMF*. IMF Departmental Paper Series Monetary and Capital Markets Department (No. 20/04), https://doi.org/10.5089/9781513520742.087

Adrian, T. and Rosenberg, J., (2008). Stock returns and volatility: Pricing the short-run and long-run components of market risk. *The Journal of Finance*, 63(6), pp. 2997-3030.

Aggarwal, R.K., Krigman, L., and Womack, K.L., (2002). Strategic IPO underpricing, information momentum, and lockup expiration selling. *Journal of Financial Economics*, 66(1), pp. 105-137.

Ahnert, L., Vogt, P., Vonhoff, V., an Weigert, F., (2020). Regulatory stress testing and bank performance. *European Financial Management*, 26(5), pp. 1449-1488.

Ajlouni, M.M. and Toms, S., (2008). Signalling Characteristics and Information Content of Directors' Dealings on the London Stock Exchange. *Journal of Risk and Governance*, 1(1), pp. 1-24.

Akaike, H., (1973). Information theory and an extension of the maximum likelihood principle. In B.N. Petrov and F. Csáki (Eds.), *Second International Symposium on Information Theory* (pp. 267-281). Akadémiai Kiadó.

Akaike, H., (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, 19(6), pp. 716-723.

Akerlof, G.A., (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*, 84(3), pp. 488-500.

Alessandri, P., Gai, P., Kapadia, S., Mora, N., and Puhr, C., (2009). Towards a Framework for Quantifying Systemic Stability. *International Journal of Central Banking*, 5(3), pp. 47-81.

Alexander, K., (2006). Corporate governance and banks: The role of regulation in reducing the principal-agent problem. *Journal of Banking Regulation*, 7(1), pp. 17-40.

Allenspach, N., (2009). *Banking and transparency: Is more information always better?* SNB Working Paper (No. 2009-11).

Alston, W.P., (1989). *Epistemic Justification: Essays in the Theory of Knowledge*. Cornell University Press.

Alvarez, F. and Barlevy, G., (2021). Mandatory disclosure and financial contagion. *Journal of Economic Theory*, 194, pp. 105237.

Alves, C., Mendes, V., and Pereira da Silva, P., (2015). Do Stress Tests Matter ? A Study on the Impact of the Disclosure of Stress Test Results on European Financial Stocks and CDS Markets. *Applied Economics*, 47(12), pp. 1213-1229.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets*, 5(1), pp. 31-56.

Anderson, R., Daníelsson, J., Baba, C., Das, U., Kang, H., and Basurto, M.S., (2018). *Macroprudential Stress Tests and Policies: Searching for Robust and Implementable Frameworks*. IMF Working Paper (No. 18/197).

Anolli, M., Beccalli, E., and Molyneux, P., (2014). Bank earnings forecasts, risk and the crisis. *Journal of International Financial Markets, Institutions and Money*, 29, pp. 309-335.

Ardalan, K., (2003). Theories and controversies in finance: a paradigmatic overview. *International Journal of Social Economics*, 30(1-2), pp. 199-208.

Armitage, S., (1995). Event study methods and evidence on their performance. *Journal of Economic Surveys*, 9(1), pp. 25-52.

Ashley, J.W., (1962). Stock Prices and Changes in Earnings and Dividends: Some Empirical Results. *Journal of Political Economy*, 70(1), pp. 82-85.

Ashton, D.J., (1996). The power of tests of fund manager performance. *Journal of Business Finance and Accounting*, 23(1), pp. 1-11.

Ashton, D.J. and Trinh, C., (2018). Evaluating the information content of earnings forecasts. *Accounting and Business Research*, 48(6), pp. 674-699.

Ashton, P. and Christophers, B., (2015). On arbitration, arbitrage and arbitrariness in financial markets and their governance: Unpacking LIBOR and the LIBOR scandal. *Economy and Society*, 44(2), pp. 188-217.

Atanasov, V. and Black, B., (2016). Shock-based causal inference in corporate finance and accounting research. *Critical Finance Review*, 5, pp. 207-304.

Audi, R., (2011). *Epistemology: A Contemporary Introduction to the Theory of Knowledge* (3rd Ed.). Routledge.

Bachelier, L., (1900). *Théorie de la Spéculation*. Reprinted in English in P.H. Cootner (Ed.), (1967). The Random Character of Stock Market Prices (pp. 17-78). MIT Press.

Baillie, R.T. and DeGennaro, R.P., (1990). Stock returns and volatility. *Journal of Financial and Quantitative Analysis*, 25(2), pp. 203-214.

Baker, A.C., (2016). Single-Firm Event Studies, Securities Fraud, and Financial Crisis: Problems of Inference. *Stanford Law Review*, 68(5), pp 1207-1261.

Baker, H.K. and Edelman, R.B., (1990). OTC market switching and stock returns: some empirical evidence. *Journal of Financial Research*, 13(4), pp. 325-338.

Baker, H.K. and Edelman, R.B., (1992). AMEX-to-NYSE transfers, market microstructure, and shareholder wealth. *Financial Management*, 21(4), pp. 60-72.

Baker, H.K., Khan, W.A., and Edelman, R.B., (1994). The post-dual listing anomaly. *Journal of Economics and Business*, 46(4), pp. 287-297.

Baker, H.K., Nofsinger, J.R., and Weaver, D.G., (2002). International cross-listing and visibility. *Journal of Financial and Quantitative Analysis*, 37(3), pp. 495-521.

Bali, T.G., (2008). The intertemporal relation between expected returns and risk. *Journal of Financial Economics*, 87(1), pp. 101-131.

Ball, R. and Brown, P., (1968). An Empirical Evaluation of Accounting Income Numbers. *Journal of Accounting Research*, 6(2), pp. 159-178.

Ball, R. and Kothari, S.P., (1989). Nonstationary expected returns: Implications for tests of market efficiency and serial correlation in returns. *Journal of Financial Economics*, 25(1), pp. 51-74.

Bannier, C.E., Behr, P., and Güttler, A., (2010). Rating opaque borrowers: why are unsolicited ratings lower?. *Review of Finance*, 14(2), pp. 263-294.

Bank, M. and Baumann, R.H., (2016). Price formation, market quality and the effects of reduced latency in the very short run. *Research in International Business and Finance*, 37, pp. 629-645.

Banz, R.W., (1981). The Relationship between Return and Market Value of Common Stocks. *Journal of Financial Economics*, 9(1), pp. 3-18.

Barber, B.M. and Lyon, J.D., (1997a). Firm size, book-to-market ratio, and security returns: A holdout sample of financial firms. *The Journal of Finance*, 52(2), pp. 875-883.

Barber, B.M. and Lyon, J.D., (1997b). Detecting long-run abnormal stock returns: The empirical power and specification of test statistics. *Journal of Financial Economics*, 43(3), pp. 341-372.

Barberis, N., (2018). Psychology-Based Models of Asset Prices and Trading Volume. In B.D. Bernheim, S. DellaVigna, and D. Laibson (Eds.), *Handbook of Behavioral Economics – Foundations and Applications 1* (pp. 79-175). North-Holland.

Barberis, N. and Thaler, R.H., (2003). *A Survey of Behavioral Finance*. In G.M. Constantinides, M. Harris, and R.M. Stulz (Eds.), Handbook of the Economics of Finance (Vol. 1) (pp. 1053-1128). Elsevier.

Barclay, M.J., (1987). Dividends, taxes, and common stock prices: The ex-dividend day behavior of common stock prices before the income tax. *Journal of Financial Economics*, 19(1), pp. 31-44.

Barker, C.A., (1956). Effective Stock Splits. *Harvard Business Review*, 34(1), pp. 101-106.

Barker, C.A., (1957). Stock Splits in a Bull Market. *Harvard Business Review*, 35(3), pp. 72-79.

Barker, C.A., (1958). Evaluation of Stock Dividends. *Harvard Business Review*, 36(4), pp. 99-114.

Barker, C.A., (1959). Price changes of stock-dividend shares at ex-dividend dates. *The Journal of Finance*, 14(3), pp. 373-378.

Baron, D.P. (1982). A model of the demand for investment banking advising and distribution services for new issues. *The Journal of Finance*, 37(4), pp. 955-976.

Bartlett, R.P., (2012). Making banks transparent. *Vanderbilt Law Review*, 65(2), pp. 293-386.

Bar-Yosef, S. and Brown, L.D., (1977). A reexamination of stock splits using moving betas. *The Journal of Finance*, 32(4), pp. 1069-1080.

Basel Committee on Banking Supervision (BCBS), (1988). International Convergence of Capital Measurement and Capital Standards. Available at: https://www.bis.org/publ/bcbs04a.pdf

Basel Committee on Banking Supervision (BCBS), (1996). Amendment to the Capital Accord to Incorporate Market Risk. Available at: https://www.bis.org/publ/bcbs24.pdf

Basel Committee on Banking Supervision (BCBS), (2004). International Convergence of Capital Measurement and Capital Standards: A Revised Framework. Available at: https://www.bis.org/publ/bcbs107.pdf

Basel Committee on Banking Supervision (BCBS), (2006). International Convergence of Capital Measurement and Capital Standards: A Revised Framework, Bank for International Settlements. Available at: https://www.bis.org/publ/bcbs128.pdf

Basel Committee on Banking Supervision (BCBS), (2009). Principles for Sound Stress Testing Practices and Supervision. Bank for International Settlements (BIS). Available at: https://www.bis.org/publ/bcbs155.pdf

Basel Committee on Banking Supervision (BCBS), (2010). Basel III: A global regulatory framework for more resilient banks and banking systems. Available at: https://www.bis.org/publ/bcbs189_dec2010.pdf

Basel Committee on Banking Supervision (BCBS), (2017a). Basel III: Finalising Post-Crisis Reforms. Available at: https://www.bis.org/bcbs/publ/d424.pdf

Basel Committee on Banking Supervision (BCBS), (2017b). Supervisory and Bank Stress Testing: Range of Practices. Available at: https://www.bis.org/bcbs/publ/d427.pdf

Basu, S., (1977). Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *The Journal of Finance*, 32(3), pp. 663-682.

Baudino, P., (2009). Use of Macro Stress Tests in Policy-Making. In M. Quagliariello (Ed.), *Stress-Testing the Banking System: Methodologies and Applications* (pp. 117-129). Cambridge University Press.

Baudino, P., Goetschmann, R., Henry, J., Taniguchi, K., and Zhu, W., (2018). Stress-Testing Banks: A Comparative Analysis. *FSI Insights on Policy Implementation*, (12). Available at: https://www.bis.org/fsi/publ/insights12.pdf

Baule, R. and Tallau, C., (2016). Stock Returns Following Large Price Changes and News Releases – Evidence from Germany. *Credit and Capital Markets*, 49(1), pp. 57-91.

Beatty, R.P. and Ritter, J.R., (1986). Investment banking, reputation, and the underpricing of initial public offerings. *Journal of Financial Economics*, 15(1-2), pp. 213-232.

Beaulac, C. (2022). A moment-matching metric for latent variable generative models. *Journal of Machine Learning Research* (forthcoming).

Beaver, W., (1968). Information Content Of Annual Earnings Announcements. *Journal of Accounting Research*, 6, pp. 67-92.

Berger, A.N. and Davies, S.M., (1998). The information content of bank examinations. *Journal of Financial Services Research*, 14(2), pp. 117-144.

Berlin, M. and Loeys, J., (1988). Bond covenants and delegated monitoring. *Journal of Finance*, 43(2), pp. 397-412.

Berlin, M., (2015). *Disclosure of stress test results*. Working Paper No. 15-31. Federal Reserve Bank of Philadelphia.

Bernanke, B.S., (2009). 'The Supervisory Capital Assessment Program', transcript, *Federal Reserve Bank of Atlanta 2009 Financial Markets Conference*, viewed 5 January 2021, https://www.federalreserve.gov/newsevents/speech/bernanke 20090511a.htm

Bernanke, B.S., (2013). 'Stress testing banks: What have we learned?', transcript, *Federal Reserve Bank of Atlanta 2013 Financial Markets Conference*, viewed 26 January 2021, https://www.federalreserve.gov/newsevents/speech/bernanke 20130408a.htm

Besancenot, D. and Vranceanu, R., (2011). Banks' risk race: A signaling explanation. *International Review of Economics and Finance*, 20(4), pp. 784-791.

Bhojraj, S. and Swaminathan, B., (2006). Macromomentum: returns predictability in international equity indices. *The Journal of Business*, 79(1), pp. 429-451.

Bhushan, R., (1994). An informational efficiency perspective on the post-earnings announcement drift. *Journal of Accounting and Economics*, 18(1), pp. 45-65.

Binder, J.J., (1998). The Event Study Methodology Since 1969. *Review of Quantitative Finance and Accounting*, 11(2), pp. 111-137.

Bird, A., Karolyi, S.A., and Ruchti, T., (2020). Bias and the efficacy of stress test disclosures. Working Paper, 27 March 2020. Available at SSRN: https://ssrn.com/abstract=2626058 or http://dx.doi.org/10.2139/ssrn.2626058

Bischof, J. and Daske, H., (2013). Mandatory disclosure, voluntary disclosure, and stock market liquidity: Evidence from the EU bank stress tests. *Journal of Accounting Research*, 51(5), pp. 997-1029.

Blackwell, D., (1951). 'Comparison of Experiments'. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (pp. 93-102). University of California Press.

Blackwell, D., (1953). Equivalent comparisons of experiments. *Annals of Mathematical Statistics*, 24(2), pp. 265-272.

Blanchard, O., (2009). *The Crisis: Basic Mechanisms and Appropriate Policies*. IMF Working Paper (No. 09/80).

Blaschke, W., Jones, M.T., Majnoni, G., and Martinez Peria, S., (2001). Stress Testing of Financial Systems: An Overview of Issues, Methodologies, and FSAP Experiences, 2001. *IMF Working Paper*, WP/01/88, Available at: http://www.imf.org/external/pubs/ft/wp/2001/wp0188.pdf

Blau, B.M., Brough, T.J., and Griffith, T.G., (2017). Bank opacity and the efficiency of stock prices. *Journal of Banking and Finance*, 76, pp. 32-47.

Blau, B.M., Griffith, T.G., and Whitby, R.J., (2020). Opacity and the comovement in the stock prices of banks. *Accounting and Finance*, 60, pp. 3557-3580.

Blundell-Wignall, A. and Slovik, P., (2010). The EU Stress Test and Sovereign Debt Exposures. *OECD Working Papers on Finance, Insurance and Private Pensions*, (4), pp. 1-13.

Boneau, C.A., (1960). The effects of violations of assumptions underlying the t test. *Psychological Bulletin*, 57(1), pp. 49-64.

Bookstaber, R., Cetina, J., Feldberg, G., Flood, M., and Glasserman, P., (2014). Stress Tests to Promote Financial Stability: Assessing Progress and Looking to the Future. *Journal of Risk Management in Financial Institutions*, 7(1), pp. 16-25.

Borges, M.R., Mendes, J.Z., and Pereira, A., (2019). The Value of Information: The Impact of European Union Bank Stress Tests on Stock Markets. *International Advances in Economic Research*, 25(4), pp. 429-444.

Borio, C., Drehmann, M., and Tsatsaronis, K., (2014). Stress-Testing Macro Stress Testing: Does it Live up to Expectations?. *Journal of Financial Stability*, 12, pp. 3-15.

Boss, M., Krenn, G., Puhr, C., and Summer, M., (2006). Systemic Risk Monitor: A Model for Systemic Risk Analysis and Stress Testing of Banking Systems. *Financial Stability Report*, (11), pp. 83-95.

Bouwman, C.H.S, Hu, S., and Johnson, S.A., (2018). Differential Bank Behaviors around the Dodd–Frank Act Size Thresholds. *Journal of Financial Intermediation*, 34(C), pp. 47-57.

Box, G.E., (1976). Science and statistics. *Journal of the American Statistical Association*, 71(356), pp. 791-799.

Box, G.E., (1979). *Robustness in the Strategy of Scientific Model Building*. In R.L. Launer and G.N. Wilkinson (Eds.), Robustness in Statistics (pp. 201-236). Academic Press.

Box, G.E. and Draper, N.R., (1987). *Empirical Model-Building and Response Surfaces*. Wiley.

Boyd, J.H. and Jagannathan, R., (1994). Ex-dividend price behavior of common stocks. *The Review of Financial Studies*, 7(4), pp. 711-741.

Brav, A. (2000). Inference in long-horizon event studies: a Bayesian approach with application to initial public offerings. *The Journal of Finance*, 55(5), pp. 1979-2016.

Breuer, T. and Csiszár, I., (2013). Systematic stress tests with entropic plausibility constraints. *Journal of Banking and Finance*, 37(5), pp. 1552-1559.

Bris, A., Cantale, S., Hrnjić, E., and Nishiotis, G.P., (2012). The value of information in cross-listing. *Journal of Corporate Finance*, 18(2), pp. 207-220.

Brown, P., Kleidon, A.W., and Marsh, T.A., (1983). New Evidence on the Nature of Size-Related Anomalies in Stock Prices. *Journal of Financial Economics*, 12(1), pp. 33-56.

Brown, S.J. and Warner, J.B., (1980). Measuring security price performance. *Journal of Financial Economics*, 8(3), pp. 205-258.

Brown, S.J. and Warner, J.B., (1985). Using daily stock returns: The case of event studies. *Journal of Financial Economics*, 14(1), pp. 3-31.

Bruner, R.F. and Simms, J.M., (1987). The international debt crisis and bank security returns in 1982. *Journal of Money, Credit and Banking*, 19(1), pp. 46-55.

Bruno, B., Onali, E., and Schaeck, K., (2018). Market Reaction to Bank Liquidity Regulation. *Journal of Financial and Quantitative Analysis*, 53(02), pp. 899-935.

Bunn, P., Cunningham, A., and Drehmann, M., (2005). Stress testing as a tool for assessing systemic risk. *Bank of England Financial Stability Review*, 18(116-26), pp. 19-21.

Burnham, K.P. and Anderson, D.R., (2002), *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (2nd Ed.). Springer.

Burrell, G. and Morgan, G., (1979), *Sociological Paradigms and Organisational Analysis: Elements of the Sociology of Corporate Life*. Heinemann.

Bushee, B.J. and Leuz, C., (2005). Economic Consequences of SEC Disclosure Regulation: Evidence from the OTC Bulletin Board. *Journal of Accounting and Economics*, 39(2), pp. 233-264.

Butler, A.W. and Wan, H., (2010). Stock market liquidity and the long-run stock performance of debt issuers. *The Review of Financial Studies*, 23(11), pp. 3966-3995.

Byun, J. and Rozeff, M.S., (2003). Long-run performance after stock splits: 1927 to 1996. *The Journal of Finance*, 58(3), pp. 1063-1085.

Cable, J. and Holland, K., (1999). Modelling normal returns in event studies: a model-selection approach and pilot study. *The European Journal of Finance*, 5(4), pp. 331-341.

Callaghan, C., (2017). Contemporary insights from social sciences theory: implications for management. *South African Journal of Business Management*, 48(4), pp. 35-45.

Calomiris, C.W. and Mason, J.R., (1997). Contagion and Bank Failures During the Great Depression: The June 1932 Chicago Banking Panic. *The American Economic Review*, 87(5), pp. 863-883.

Campbell, D.T., (1979). Assessing the impact of planned social change. *Evaluation and program planning*, 2(1), pp. 67-90.

Campbell, D.T. and Stanley, J.C., (1963). *Experimental and Quasi-Experimental Designs for Research*. Rand McNally.

Campbell, J.A. and Beranek, W., (1955). Stock price behavior on ex-dividend dates. *The Journal of Finance*, 10(4), pp. 425-429.

Campbell, J.Y., (1987). Stock returns and the term structure. *Journal of Financial Economics*, 18(2), pp. 373-399.

Campbell, J.Y., (2014). Empirical asset pricing: Eugene Fama, Lars Peter Hansen, and Robert Shiller. *The Scandinavian Journal of Economics*, 116(3), pp. 593-634.

Campbell, J.Y. and Hentschel, L., (1992). No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial conomics*, 31(3), pp. 281-318.

Campbell, J.Y., Lo, A.W., and MacKinlay, A.C., (1997). *The Econometrics of Financial Markets*. Princeton University Press, Princeton, NJ, USA.

Campbell, J.Y. and Viceira, L.M., (2005). The term structure of the risk–return tradeoff. *Financial Analysts Journal*, 61(1), pp. 34-44.

Campbell, T.S., (1979). Optimal investment financing decisions and the value of confidentiality. *Journal of Financial and Quantitative Analysis*, pp. 913-924.

Campbell, T.S. and Kracaw, W.A., (1980). Information production, market signalling, and the theory of financial intermediation. *Journal of Finance*, 35(4), pp. 863-882.

Candelon, B., and Sy, A.N.R., (2015). How Did Markets React to Stress Tests ? *IMF Working Paper*, 15/75, International Monetary Fund, pp. 1-20. Available at: http://www.imf.org/external/pubs/cat/longres.aspx?sk=41665.0.

Caprio, G., (2018). *Assessing the FSAP: Quality, Relevance, and Value Added*. IEO Background Paper (No. 18-02/02). International Monetary Fund (IMF), Available at: https://ieo.imf.org/~/media/IEO/Files/evaluations/completed/01-15-2019-financial-surveillance/FISBP180202AssessingtheFSAPQualityRelevanceValue-Added.ashx?la=en.

Carboni, M., Fiordelisi, F., Ricci, O., and Lopes, F.S.S., (2017). Surprised or Not Surprised? The Investors' Reaction to the Comprehensive Assessment Preceding the Launch of the Banking Union. *Journal of Banking and Finance*, 74, pp. 122-132.

Cardinali, A. and Nordmark, J., (2011). How Informative are Bank Stress Tests? Bank Opacity in the European Union. *Lund University*, (Spring 2011), pp. 1-49. Available at: http://lup.lub.lu.se/luur/download?func=downloadFile&recordOId=1974217&fileOId=1974219 (29. September 2020).

Carhill, M., (2009). *Stress-testing US banks using economic-value-of-equity (EVE) models.* In M. Quagliariello (Ed.), Stress-Testing the Banking System: Methodologies and Applications (pp. 18-36). Cambridge University Press.

Carhart, M.M., (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance*, 52(1), pp. 57-82.

Carter, R.B., Dark, F.H., and Singh, A.K., (1998). Underwriter reputation, initial returns, and the long-run performance of IPO stocks. *The Journal of Finance*, 53(1), pp. 285-311.

Chamley, C., Kotlikoff, L.J., and Polemarchakis, H., (2012). Limited-Purpose Banking - Moving from "Trust Me" to "Show Me" Banking. *American Economic Review*, 102(3), pp. 113-19.

Chan, K.C., (1988). On the Contrarian Investment Strategy. *The Journal of Business*, 61(2), pp. 147-163.

Chan, K.C., Gup, B.E., and Pan, M.S., (1997). International stock market efficiency and integration: A study of eighteen nations. *Journal of Business Finance and Accounting*, 24(6), pp. 803-813.

Chan, K.C., Karolyi, G.A., and Stulz, R., (1992). Global financial markets and the risk premium on US equity. *Journal of Financial Economics*, 32(2), pp. 137-167.

Charest, G., (1978). Split information, stock returns and market efficiency-I. *Journal of Financial Economics*, 6(2-3), pp. 265-296.

Chen, C.R., and Chan, A., (1989). Interest Rate Sensitivity, Asymmetry, and the Stock Returns of Financial Institutions. *Financial Review*, 24(3), pp. 457-473.

Chen, H. and Singal, V., (2003). Role of speculative short sales in price formation: The case of the weekend effect. *Journal of Finance*, 58(2), pp. 685-705.

Chen, L.H., Huang, W. and Jiang, G.J., (2017). Herding on earnings news: The role of institutional investors in post-earnings-announcement drift. *Journal of Accounting, Auditing and Finance*, 32(4), pp. 536-560.

Chou, R.Y., (1988). Volatility persistence and stock valuations: Some empirical evidence using GARCH. *Journal of Applied Econometrics*, 3(4), pp. 279-294.

Chowdhury, M., Howe, J.S., and Lin, J.C., (1993). The Relation between Aggregate Insider Transactions and Stock Market Returns. *Journal of Financial and Quantitative Analysis*, 28(3), pp. 431-437.

Choudhry, T., Papadimitriou, F.I., and Shabi, S., (2016). Stock Market Volatility and Business Cycle: Evidence from Linear and Nonlinear Causality Tests. *Journal of Banking and Finance*, 66, pp. 89-101.

Chrystal, A. and Mizen, P., (2003). *Goodhart's Law: Its origins, meaning and implications for monetary policy*. In P. Mizen (Ed.), Central Banking, Monetrary Theory and Practice: Essays in Honour of Charles Goodhart (Vol. 1) (pp. 221-243). Edward Elgar.

Cicchitelli, G., (1989). On the robustness of the one sample t test. *Journal of Statistical Computation and Simulation*, 32(4), pp. 249-258.

Čihák, M., (2004). *Designing Stress Tests for the Czech Banking System* (No. 2004/03). Czech National Bank, Available at: https://www.cnb.cz/export/sites/cnb/en/economic-research/.galleries/research_publications/irpn/download/irpn_3_2004.pdf

Čihák, M., (2007). Introduction to applied stress testing. *IMF Working Papers*, 1-74.

Collis, J. and Hussey, R., (2003). *Business Research: A Practical Guide for Undergraduate and Postgraduate Students* (2nd Ed.). Palgrave Macmillan.

Committee of European Banking Supervisors (CEBS), (2009a). *CEBS' s statement on stress testing exercise*. Press Release, 12 May 2009, https://eba.europa.eu/cebs-s-statement-on-stress-testing-exercise

Committee of European Banking Supervisors (CEBS), (2009b). *CEBS's Press Release on the Results of the EU-Wide Stress Testing Exercise*. Viewed: 23.02.2021. Available at: https://www.eba.europa.eu/sites/default/documents/files/documents/10180/15977/01df9de6-acc8-4b8f-ac72-849d96087795/CEBS-2009-180-Annex-2-%28Press-release-from-CEBS%29.pdf

Committee of European Banking Supervisors (CEBS), (2010a). *CEBS's Statement on Key Features of the Extended EU-Wide Stress Test*. Viewed: 26.05.2021. Available at: https://www.eba.europa.eu/sites/default/documents/files/documents/10180/15977/7450f9a9-31ea-4489-b5e4-00b54b2f5cc9/ST_followupPR.pdf

Committee of European Banking Supervisors (CEBS), (2010b). *CEBS's Press Release on the Results of the 2010 EU-wide Stress Testing Exercise*. Viewed: 15.07.2021. Available at: https://eba.europa.eu/sites/default/documents/files/documents/10180/15938/4a5b185f-43bf-4e4d-b1de-c50c7e656b87/CEBSPressReleasev2.pdf

Committee on the Global Financial System (CGFS), (2000). Stress Testing by Large Financial Institutions: Current Practice and Aggregation Issues, *CGFS Paper No. 14*, Available at: https://www.bis.org/publ/cgfs14.htm

Committee on the Global Financial System (CGFS), (2001). A Survey of Stress Tests and Current Practice at Major Financial Institutions, *CGFS Paper No. 18*, Available at: https://www.bis.org/publ/cgfs18.htm

Committee on the Global Financial System (CGFS), (2005). Stress testing at major financial institutions: survey results and practice, *CGFS Paper No. 24*, Available at: https://www.bis.org/publ/cgfs24.pdf

Conrad, J. and Kaul, G., (1988). Time-Variation in Expected Returns. *The Journal of Business*, 61(4), pp. 409-425.

Copeland, T.E., (1979). Liquidity changes following stock splits. *The Journal of Finance*, 34(1), pp. 115-141.

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.

Corradi, V., Distaso, W., and Mele, A., (2013). Macroeconomic Determinants of Stock Volatility and Volatility Premiums. *Journal of Monetary Economics*, 60(2), pp. 203-220.

Dahiya, S., Iannotta, G., and Navone, M., (2017). Firm opacity lies in the eye of the beholder. *Financial Management*, 46(3), pp. 553-592.

Dang, T.V., Gorton, G., and Holmström, B., (2015). The Information Sensitivity of a Security. *Working Paper, Columbia University*.

Daniel, K. and Titman, S., (1999). Market efficiency in an irrational world. *Financial Analysts Journal*, 55(6), pp. 28-40.

Daniel, K., Hirshleifer, D., and Subrahmanyam, A., (1998). Investor Psychology and Security Market Under- and Overreactions. *The Journal of Finance*, 53(6), pp. 1839-1885.

Daniel, K., Hirshleifer, D., and Teoh, S.H., (2002). Investor psychology in capital markets: Evidence and policy implications. *Journal of Monetary Economics*, 49(1), pp. 139-209.

Daníelsson, J., (2002). The emperor has no clothes: Limits to risk modelling. *Journal of Banking and Finance*, 26(7), pp. 1273-1296.

Daníelsson, J., (2003). On the feasibility of risk based regulation. *CESifo Economic Studies*, 49(2), pp. 157-179.

Davies, P.L. and Canes, M., (1978). Stock prices and the publication of second-hand information. *Journal of Business*, 51(1), pp. 43-56.

De Bondt, W.F.M. and Thaler, R.H., (1985). Does the stock market overreact?. *The Journal of Finance*, 40(3), pp. 793-805.

De Bondt, W.F.M. and Thaler, R.H., (1987). Further evidence on investor overreaction and stock market seasonality. *The Journal of Finance*, 42(3), pp. 557-581.

De Ceuster, M.J. and Masschelein, N., (2003). Regulating banks through market discipline: A survey of the issues. *Journal of Economic Surveys*, 17(5), pp. 749-766.

Dees, S., Henry, J., and Martin, R. (Eds.), (2017). *STAMP€: Stress-Test Analytics for Macroprudential Purposes in the Euro Area*. European Central Bank, available at: https://www.ecb.europa.eu/pub/pdf/other/stampe201702.en.pdf

De Franco, G., Lu, H., and Vasvari, F.P., (2007). Wealth transfer effects of analysts' misleading behavior. *Journal of Accounting Research*, 45(1), pp. 71-110.

Del Brio, E.B. and De Miguel, A., (2010). Dividends and market signalling: An analysis of corporate insider trading. *European Financial Management*, 16(3), pp. 480-497.

Dell'Ariccia, G., (2001). Asymmetric information and the structure of the banking industry. *European Economic Review*, 45(10), 1957-1980.

Demski, J.S. and Feltham, G.A., (1994). Market response to financial reports. *Journal of Accounting and Economics*, 17(1-2), pp. 3-40.

Dent, K., Westwood, B., and Segoviano, M., (2016). Stress Testing of Banks: An Introduction. *Bank of England Quarterly Bulletin*, Q3. Available at: https://www.bankofengland.co.uk/-/media/boe/files/quarterly-bulletin/2016/stress-testing-of-banks-an-introduction.pdf

Derrien, F. and Womack, K.L., (2003). Auctions vs. bookbuilding and the control of underpricing in hot IPO markets. *The Review of Financial Studies*, 16(1), pp. 31-61.

Desai, H. and Jain, P.C., (1997). Long-run common stock returns following stock splits and reverse splits. *The Journal of Business*, 70(3), pp. 409-433.

Dewally, M. and Shao, Y., (2013). Financial derivatives, opacity, and crash risk: Evidence from large US banks. *Journal of Financial Stability*, 9(4), pp. 565-577.

Dharan, B.G. and Ikenberry, D.L., (1995). The long-run negative drift of post-listing stock returns. *The Journal of Finance*, 50(5), pp. 1547-1574.

Diamond, D.W., (1989). Reputation acquisition in debt markets. *Journal of Political Economy*, 97(4), pp. 828-862.

Diamond, D.W., (1991). Monitoring and reputation: The choice between bank loans and directly placed debt. *Journal of Political Economy*, 99(4), pp. 689-721.

DiNardo, J., (2008). Natural Experiments and Quasi-Natural Experiments. In: Durlauf, S.N. and Blume, L.E. (Eds.). *The New Palgrave Dictionary of Economics* (2nd ed., Vol. 8). Palgrave Macmillan, London, UK. pp. 856–864.

Dinenis, E. and Staikouras, S.K., (1998). Interest Rate Changes and Common Stock Returns of Financial Institutions: Evidence from the UK. *The European Journal of Finance*, 4(2), pp. 113-127.

Dolley, J.C., (1933). Characteristics and Procedure of Common Stock Split-Ups. *Harvard Business Review*, 11(3), pp. 316-326.

Dunning, T., (2012). *Natural Experiments in the Social Sciences: A Design-Based Approach*. Cambridge University Press, Cambridge, UK.

Eades, K.M., Hess, P.J., and Kim, E.H., (1984). On interpreting security returns during the ex-dividend period. *Journal of Financial Economics*, 13(1), pp. 3-34.

Ebner, A., (2018). The Financial Stability Aspects of the EU-wide Stress Test. *Journal of Financial Regulation*, 4(2), pp. 326-336.

Economic and Financial Affairs Council (ECOFIN), (2009). *Economic and Financial Affairs*. Press Release (No. 352), 2 December 2009, available at: https://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/ecofin/111706.pdf

Edelman, R.B. and Baker, H.K., (1990). Liquidity and stock exchange listing. *Financial Review*, 25(2), pp. 231-249.

Edelman, R.B. and Baker, H.K., (1993). The impact of company pre-listing attributes on the market reaction to NYSE listings. *Financial Review*, 28(3), pp. 431-448.

Edelman, R.B. and Baker, H.K., (1994). The postlisting returns anomaly revisited. *Quarterly Journal of Business and Economics*, 33(2), 54-68.

Edwards, R.D., Magee, J., and Bassetti, W.H.C., (2018). *Technical Analysis of Stock Trends* (11 ed.). Routledge.

Eisdorfer, A. and Giaccotto, C., (2014). Pricing assets with stochastic cash-flow growth. *Quantitative Finance*, 14(6), pp. 1005-1017.

Ellahie, A., (2012). 'Capital Market Consequences of EU Bank Stress Tests', paper presented at the *Joint Conference of the Federal Reserve Bank of New York and the Journal of Accounting Research*, 20-21 September, New York, viewed 10 June 2021, https://www.newyorkfed.org/medialibrary/media/research/conference/2012/FinancialServices2012/Ellahie.pdf

Elton, E.J. and Gruber, M.J., (1970). Marginal Stockholder Tax Rates and the Clientele Effect. *The Review of Economics and Statistics*, 52(1), pp. 68-74.

Elyasiani, E., Hauser, S., and Lauterbach, B., (2000). Market response to liquidity improvements: Evidence from exchange listings. *Financial Review*, 35(1), pp. 1-14.

Enria, A., (2018). 'What we have learnt from EU-wide stress tests', transcript, *Speech delivered at the National Bank of Romania*, 15 November 2018, viewed 24 February 2021, https://www.eba.europa.eu/calendar%3Fp_p_id%3D8%26_8_struts_action%3D%252Fcalendar%252Fview_event%26_8_eventId%3D2453933

Espenlaub, S., Gregory, A., and Tonks, I., (2000). Re-assessing the long-term underperformance of UK Initial Public Offerings. *European Financial Management*, 6(3), pp. 319-342.

European Banking Authority (EBA), (2011a). '2011 EU-Wide Stress Test: Objectives, outcome and recommendations', viewed 24 February 2021, https://www.eba.europa.eu/sites/default/documents/files/documents/10180/15935/fef35b7a-28aa-40ad-8837-78e6998e9664/Presentation%20to%20Analysts.pdf?retry=1

European Banking Authority (EBA), (2011b). *Results of the 2011 EU-Wide Stress Test*. Viewed: 25 February 2021. Available at: https://www.eba.europa.eu/sites/default/documents/files/documents/10180/15935/b8211d3b-562e-40d4-80f8-0b5736c20345/2011%20EU-wide%20stress%20test%20results%20-%20press%20release%20-%20FINAL.pdf?retry=1

European Banking Authority (EBA), (2011c). *2011 EU-Wide Stress Test: Methodological Note*, Version 1.1. Viewed: 26.05.2021. Available at: https://www.eba.europa.eu/sites/default/documents/files/documents/10180/15932/1cf7d810-5ac4-4771-8409-17f038c2bb44/EBA-ST-2011-004-Detailed-Methodological-Note_1.pdf

European Banking Authority (EBA), (2011d). *The EBA Announces Stress Test Publication Date*. Viewed: 15.07.2021. Available at: https://eba.europa.eu/the-eba-announces-stress-test-publication-date

European Banking Authority (EBA), (2014a). *Methodological Note EU-Wide Stress Test 2014*, Version 2.0. Viewed: 26.05.2021. Available at: https://www.eba.europa.eu/sites/default/documents/files/documents/10180/669262/0322dcfa-9f84-4430-b275-98388906676e/Methodological%20Note.pdf

European Banking Authority (EBA), (2014b). *EBA Announces 2014 EU-Wide Stress Test Publication Date*. Viewed: 15.07.2021. Available at: https://www.eba.europa.eu/eba-announces-2014-eu-wide-stress-test-publication-date

European Banking Authority (EBA), (2016a). '2016 EU-Wide Stress Test: Presentation to Analysts', viewed 24 February 2021, https://www.eba.europa.eu/sites/default/documents/files/documents/10180/1532819/7db99f99-e3c1-41f6-978e-19907085dfda/2016-EU-wide-stress-test-Presentation-to-analysts.pdf?retry=1

European Banking Authority (EBA), (2016b). *2016 EU-Wide Stress Test: Methodological Note*. Viewed: 26.05.2021. Available at: https://www.eba.europa.eu/sites/default/documents/files/documents/10180/1259315/e077989b-c5a2-4f1f-a683-da9a53f70704/2016%20EU-wide%20stress%20test-Methodological%20note.pdf

European Banking Authority (EBA), (2016c). *EBA Announces Timing for Publication of 2016 EU-wide Stress Test Results*. Viewed: 15.07.2021. Available at: https://eba.europa.eu/eba-announces-timing-for-publication-of-2016-eu-wide-stress-test-results

European Banking Authority (EBA), (2018a). '2018 EU-Wide Stress Test Results', viewed 26 May 2021, https://www.eba.europa.eu/documents/10180/2419200/2018-EU-wide-stress-test-Presentation-to-analysts.pdf

European Banking Authority (EBA), (2018b). *2018 EU-Wide Stress Test: Methodological Note*. Viewed: 26.05.2021. Available at: https://www.eba.europa.eu/sites/default/documents/files/documents/10180/2106643/a72411ca-3d95-44d3-9c6a-2c36de7d482f/2018%20EU-wide%20stress%20test%20-%20Methodological%20Note.pdf

European Banking Authority (EBA), (2018c). *All You Need to Know About the 2018 EU-wide Stress Test*. Viewed: 15.07.2021. Available at: https://eba.europa.eu/risk-analysis-and-data/eu-wide-stress-testing/2018

European Banking Authority (EBA), (2020a). *EU-Wide Stress Testing*, viewed 4 October 2020, https://eba.europa.eu/risk-analysis-and-data/eu-wide-stress-testing

European Banking Authority (EBA), (2020b). *EU-wide stress testing 2014*, viewed 24 February 2021, https://www.eba.europa.eu/risk-analysis-and-data/eu-wide-stress-testing/2014

European Banking Authority (EBA), (2020c). *Discussion Paper on the Future Changes to the EU-Wwide Stress Test*. Viewed: 25. February 2021. Available at: https://www.eba.europa.eu/sites/default/documents/files/document_library/Calendar/EBA%20Official%20Meetings/2020/Discussion%20Paper%20on%20the%20future%20changes%20to%20the%20EU-wide%20stress%20test/Discussion%20Paper%20on%20the%20future%20changes%20to%20the%20EU-wide%20stress%20test%20-%20FINAL%20-.pdf

European Banking Authority (EBA), (2022). *EU Capital Exercise*, viewed 11 March 2022, https://www.eba.europa.eu/risk-analysis-and-data/eu-capital-exercise

European Central Bank (ECB), (2022). *Comprehensive Assessment*, viewed 11 March 2022, https://www.bankingsupervision.europa.eu/banking/tasks/comprehensive_assessment/html/index.en.html

Evans, P., (1985). Money, Output and Goodhart's Law: The US Experience. *Review of Economics and Statistics*, 67(1), pp. 1-8.

Fama, E.F., (1965a). The behavior of stock-market prices. *The Journal of Business*, 38(1), pp. 34-105.

Fama, E.F., (1965b). Random Walks in Stock Market Prices. *Financial Analysts Journal*, 21(5), pp. 55-59.

Fama, E.F., (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2), pp. 383-417.

Fama, E.F., (1991). Efficient Capital Markets: II. *The Journal of Finance*, 46(5), pp. 1575-1617.

Fama, E.F., (1998). Market Efficiency, Long-Term Returns, and Behavioral Finance. *Journal of Financial Economics*, 49(3), pp. 283-306.

Fama, E.F., (2014). Two pillars of asset pricing. *American Economic Review*, 104(6), pp. 1467-1485.

Fama, E.F. and French, K.R., (1988). Permanent and temporary components of stock prices. *Journal of Political Economy*, 96(2), pp. 246-273.

Fama, E.F. and French, K.R., (1992). The Cross-Section of Expected Stock Returns. *Journal of Finance*, 47(2), pp. 427-465.

Fama, E.F. and French, K.R., (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*, 33, pp. 3-56.

Fama, E.F. and French, K.R., (1996). Multifactor explanations of asset pricing anomalies. *The Journal of Finance*, 51(1), pp. 55-84.

Fama, E.F. and French, K.R., (2006). Profitability, investment and average returns. *Journal of Financial Economics*, 82(3), pp. 491-518.

Fama, E.F. and French, K.R., (2008). Dissecting anomalies. *The Journal of Finance*, 63(4), pp. 1653-1678.

Fama, E.F. and French, K.R., (2012). Size, Value, and Momentum in International Stock Returns. *Journal of Financial Economics*, 105(3), pp. 457-472.

Fama, E.F. and French, K.R., (2015). A Five-Factor Asset Pricing Model. *Journal of Financial Economics*, 116(1), pp. 1-22.

Fama, E.F. and French, K.R., (2016). Dissecting anomalies with a five-factor model. *The Review of Financial Studies*, 29(1), pp. 69-103.

Fama, E.F., Fisher, L., Jensen, M.C., and Roll, R., (1969). The Adjustment of Stock Prices to New Information. *International Economic Review*, *10*(1), pp. 1-21.

Faria-e-Castro, M., Martinez, J., and Philippon, T., (2017). Runs versus lemons: information disclosure and fiscal capacity. *The Review of Economic Studies*, 84(4), pp. 1683-1707.

Federal Reserve System (Fed), (2009a). *The Supervisory Capital Assessment Program: Design and Implementation*. https://www.federalreserve.gov/newsevents/pressreleases/files/bcreg20090424a1.pdf

Federal Reserve System (Fed), (2009b). *The Supervisory Capital Assessment Program: Overview of Results*. https://www.federalreserve.gov/newsevents/files/bcreg20090507a1.pdf

Federal Reserve System (Fed), (2020a). *Stress Tests and Capital Planning: Comprehensive Capital Analysis and Review*. https://www.federalreserve.gov/supervisionreg/ccar.htm

Federal Reserve System (Fed), (2020b). *Stress Tests and Capital Planning: Dodd-Frank Act Stress Tests.* https://www.federalreserve.gov/supervisionreg/dfa-stress-tests.htm

Felixson, K. and Liljeblom, E., (2008). Evidence of ex-dividend trading by investor tax category. *The European Journal of Finance*, 14(1), pp. 1-21.

Fender, I., Gibson, M.S., and Mosser, P.C., (2001). An international survey of stress tests. *Current Issues in Economics and Finance*, 7(10).

Feng, L., Lan, W., Liu, B., and Ma, Y., (2021). High-dimensional test for alpha in linear factor pricing models with sparse alternatives. *Journal of Econometrics*.

Fernandes, M., Igan, D., and Pinheiro, M., (2020). March Madness in Wall Street: (What) Does the Market Learn from Stress Tests?. *Journal of Banking and Finance*, 112, pp. 105250.

Ferson, W.E. and Schadt, R.W., (1996). Measuring fund strategy and performance in changing economic conditions. *The Journal of Finance*, 51(2), pp. 425-461.

Fidrmuc, J.P., Goergen, M., and Renneboog, L., (2006). Insider trading, news releases, and ownership concentration. *The Journal of Finance*, 61(6), pp. 2931-2973.

Financial Times, (2011, March 8). Concerns over latest EU bank stress tests. *Financial Times*. Available at: https://www.ft.com/content/1fdfeede-4971-11e0-b051-00144feab49a [Accessed 23 June 2021].

Fiordelisi, F., Minnucci, F., Previati, D., and Ricci, O., (2020). Bail-in regulation and stock market reaction. *Economics Letters*, 186, pp. 108801.

Fisher, L., (1966). Some new stock-market indexes. *The Journal of Business*, 39(1), pp. 191-225.

Flannery, M.J. and Houston, J.F., (1999). The value of a government monitor for US banking firms. *Journal of Money, Credit and Banking*, 31(1), pp. 14-34.

Flannery, M.J. and James, C.M., (1984). The Effect of Interest Rate Changes on the Common Stock Returns of Financial Institutions. *The Journal of Finance*, 39(4), pp. 1141-1153.

Flannery, M.J., Hirtle, B., and Kovner, A., (2017). Evaluating the Information in the Federal Reserve Stress Tests. *Journal of Financial Intermediation*, 29, pp. 1-18.

Flannery, M.J., Kwan, S.H., and Nimalendran, M., (2004). Market evidence on the opaqueness of banking firms' assets. *Journal of Financial Economics*, 71(3), pp. 419-460.

Flannery, M.J., Kwan, S.H., and Nimalendran, M., (2013). The 2007–2009 financial crisis and bank opaqueness. *Journal of Financial Intermediation*, 22(1), pp. 55-84.

Flood, M.D. and Korenko, G.G., (2015). Systematic scenario selection: stress testing and the nature of uncertainty. *Quantitative Finance*, 15(1), pp. 43-59.

Fontana, G., Realfonzo, R., and Passarella, M.V., (2020). Monetary Economics After the Global Financial Crisis: What has Happened to the Endogenous Money Theory?. *European Journal of Economics and Economic Policies*, 17(3), pp. 339-355.

Fosu, S., Ntim, C.G., Coffie, W., and Murinde, V., (2017). Bank opacity and risk-taking: Evidence from analysts' forecasts. *Journal of Financial Stability*, 33, pp. 81-95.

Fosu, S., Danso, A., Agyei-Boapeah, H., Ntim, C.G., and Murinde, V., (2018). How does banking market power affect bank opacity? Evidence from analysts' forecasts. *International Review of Financial Analysis*, 60, pp. 38-52.

Francis, B.B., Hasan, I., Song, L., and Yeung, B., (2015). What Determines Bank-Specific Variations in Bank Stock Returns? Global Evidence. *Journal of Financial Intermediation*, 24(3), pp. 312-324.

Frank, M.Z. and Shen, T., (2016). Investment and the weighted average cost of capital. *Journal of Financial Economics*, 119(2), pp. 300-315.

Freeman, C. and Soete, L., (2009). Developing science, technology and innovation indicators: What we can learn from the past. *Research Policy*, 38(4), pp. 583-589.

French, K.R., Schwert, G.W., and Stambaugh, R.F., (1987). Expected stock returns and volatility. *Journal of Financial Economics*, 19(1), pp. 3-29.

Friend, I. and Vickers, D., (1965). Portfolio selection and investment performance. *The Journal of Finance*, 20(3), pp. 391-415.

Fung, S. and Loveland, R., (2020). When Do Informed Traders Acquire and Trade on Informational Advantage? Evidence from Federal Reserve Stress Tests. *Journal of Futures Markets*, 40(10), pp. 1459-1485.

Furlong, F., (2011). Stress testing and bank capital supervision. *FRBSF Economic Letter*, 2011, 20, Available at: https://www.frbsf.org/economic-research/files/el2011-20.pdf.

Garcia, L., Lewrick, U., and Sečnik, T., (2021). Window Dressing Systemic Importance: Evidence from EU Banks and the G-SIB Framework. EBA Staff Paper (No. 12).

García, R.E. and Steele, S., (2022). Stress testing and bank business patterns: A regression discontinuity study. *Journal of Banking and Finance*, (135), pp. 105964.

Gelman, A., (2008). Objections to Bayesian statistics. *Bayesian Analysis*, 3(3), pp. 445-449.

Georgescu, O.-M., Gross, M., Kapp, D., and Kok, C., (2017). Do Stress Tests Matter? Evidence from the 2014 and 2016 Stress Tests. *ECB Working Paper*, No. 2054, European Central Bank (ECB), Available at: http://dx.doi.org/10.2866/622534

Georgoutsos, D. and Moratis, G., (2021). On the informative value of the EU-wide stress tests and the determinants of banks' stock return reactions. *Empirica*, 48(4), pp. 997-1008.

Geithner, T.F., (2014). *Stress Test: Reflections on Financial Crises*. Random House.

Gendron, Y. and Smith-Lacroix, J.H., (2015). The global financial crisis: Essay on the possibility of substantive change in the discipline of finance. *Critical Perspectives on Accounting*, 30, pp. 83-101.

Gerhardt, M., and Vander Vennet, R., (2017). European Bank Stress Test and Sovereign Exposures. *Applied Economics Letters*, 24(14), pp. 972-976.

Gettier, E.L., (1963). Is Justified True Belief Knowledge?, *Analysis*, 23(6), pp. 121-123.

Ghysels, E., Santa-Clara, P., and Valkanov, R., (2005). There is a risk-return trade-off after all. *Journal of Financial Economics*, 76(3), pp. 509-548.

Gick, W. and Pausch, T., (2012). *Persuasion by stress testing: Optimal disclosure of supervisory information in the banking sector.* Discussion Paper No. 32/2012. Deutsche Bundesbank.

Gill, J. and Johnson, P., (2010). *Research Methods for Managers* (4th Ed.). Sage Publications.

Gilson, R.J. and Kraakman, R.H., (1984). The mechanisms of market efficiency. *Virginia Law Review*, 70, pp. 549-644.

Gilson, R.J. and Kraakman, R.H., (2003). The Mechanisms of Market Efficiency Twenty Years Later: The Hindsight Bias. *The Journal of Corporation Law*, 28(4), pp. 715-742.

Glass, G.V., Peckham, P.D., and Sanders, J.R., (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research*, 42(3), pp. 237-288.

Glasserman, P. and Tangirala, G., (2016). Are the Federal Reserve's Stress Test Results Predictable?. *The Journal of Alternative Investments*, 18(4), pp. 82-97.

Glasserman, P., Kang, C., and Kang, W., (2015). Stress scenario selection by empirical likelihood. *Quantitative Finance*, 15(1), pp. 25-41.

Glosten, L.R., Jagannathan, R., and Runkle, D.E., (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48(5), pp. 1779-1801.

Godfrey, M.D., Granger, C.W., and Morgenstern, O., (1964). The random-walk hypothesis of stock market behavior, *Kyklos*, 17(1), pp. 1-30.

Goff, D., Hulburt, H., Keasler, T., and Walsh, J., (2008). Isolating the information content of equity analysts' recommendation changes, Post Reg FD. *Financial Review*, 43(2), pp. 303-321.

Goodhart, C.A., (1975). *Problems of Monetary Management: The U.K. Experience*. Papers in Monetary Economics (Vol. 1), Reserve Bank of Australia.

Goodhart, C.A., (2016). *In Praise of Stress Tests*. In R.W. Anderson (Ed.), Stress Testing and Macroprudential Regulation: A Transatlantic Assessment (pp. 141-153). CEPR Press.

Goldman, A.I,. (1986). *Epistemology and Cognition*. Harvard University Press.

Goldstein, I. and Leitner, Y., (2018). Stress tests and information disclosure. *Journal of Economic Theory*, 177, pp. 34-69.

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), pp. 1-54.

Goldstein, I. and Yang, L., (2019). Good disclosure, bad disclosure. *Journal of Financial Economics*, 131(1), pp. 118-138.

Goles, T. and Hirschheim, R., (2000). The Paradigm is Dead, the Paradigm is Dead… Long Live the Paradigm: The Legacy of Burrell and Morgan. *Omega*, 28(3), pp. 249-268.

Gompers, P.A. and Lerner, J., (2003). The really long-run performance of initial public offerings: The pre-Nasdaq evidence. *The Journal of Finance*, 58(4), pp. 1355-1392.

Goncharenko, R., Hledik, J., and Pinto, R., (2018). The dark side of stress tests: Negative effects of information disclosure. *Journal of Financial Stability*, 37, pp. 49-59.

Grant, C. and Osanloo, A., (2014). Understanding, Selecting, and Integrating a Theoretical Framework in Dissertation Research: Creating the Blueprint for Your "House". *Administrative Issues Journal: Connecting Education. Practice, and Research*, 4(2), pp. 12-26.

Greenspan, A., (1996). 'Remarks', transcript, *Financial Markets Conference of the Federal Reserve Bank of Atlanta*, 23 February 1996, viewed 6 January 2021, https://fraser.stlouisfed.org/title/statements-speeches-alan-greenspan-452/remarks-financial-markets-conference-federal-reserve-bank-atlanta-coral-gables-florida-8561

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

Griffin, J.M., Ji, X., and Martin, J.S., (2003). Momentum investing and business cycle risk: Evidence from pole to pole. *The Journal of Finance*, 58(6), pp. 2515-2547.

Grinblatt, M.S., Masulis, R.W. and Titman, S., (1984). The valuation effects of stock splits and stock dividends. *Journal of Financial Economics*, 13(4), pp. 461-490.

Gropp, R., Mosk, T., Ongena, S., and Wix, C., (2019). Banks Response to Higher Capital Requirements: Evidence from a Quasi-Natural Experiment. *The Review of Financial Studies*, 32(1), pp. 266-299.

Grossman, S.J. and Stiglitz, J.E., (1980). On the impossibility of informationally efficient markets. *The American Economic Review*, 70(3), pp. 393-408.

Guo, H. and Whitelaw, R.F., (2006). Uncovering the risk-return relation in the stock market. *Journal of Finance*, 61(3), pp. 1433-1463.

Guo, H., Savickas, R., Wang, Z., and Yang, J., (2009). Is the value premium a proxy for time-varying investment opportunities? Some time-series evidence. *Journal of Financial and Quantitative Analysis*, 44(1), pp. 133-154.

Haggard, K.S. and Howe, J.W., (2012). Are banks opaque? *International Review of Accounting, Banking, and* Finance, 4(1), pp. 51–72.

Haggard, K.S., Martin, X., and Pereira, R., (2008). Does voluntary disclosure improve stock price informativeness?. *Financial Management*, 37(4), pp. 747-768.

Haldane, A.G., (2011). 'Capital discipline', transcript, *American Economic Association*, viewed 19 November 2020, https://www.bankofengland.co.uk/-/media/boe/files/speech/2011/capital-discipline-speech-by-andrew-haldane.pdf

Hamilton, J.D. and Lin, G., (1996). Stock market volatility and the business cycle. *Journal of Applied Econometrics*, 11(5), pp. 573-593.

Hannan, E.J. and Quinn, B.G., (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(2), pp. 190-195.

Harvey, C.R., (1989). Time-varying conditional covariances in tests of asset pricing models. *Journal of Financial Economics*, 24(2), pp. 289-317.

Harwell, M.R., Rubinstein, E.N., Hayes, W.S., and Olds, C.C., (1992). Summarizing Monte Carlo results in methodological research: The one-and two-factor fixed effects ANOVA cases. *Journal of Educational and Behavioral Statistics*, 17(4), pp. 315-339.

Hausman, W.H., West, R.R. and Largay, J.A., (1971). Stock splits, price changes, and trading profits: a synthesis. *The Journal of Business*, 44(1), pp. 69-77.

Heider, F., Hoerova, M., and Holthausen, C., (2015). Liquidity hoarding and interbank market rates: The role of counterparty risk. *Journal of Financial Economics*, 118(2), pp. 336-354.

Henry, J. and Kok, C. (Eds.), (2013). A macro stress testing framework for assessing systemic risks in the banking sector. ECB Occasional Paper (No. 152).

Hillier, D. and Marshall, A.P., (2002). The market evaluation of information in directors' trades. *Journal of Business Finance and Accounting*, 29(1-2), pp. 77-110.

Hirshleifer, D., (2015). Behavioral finance. *Annual Review of Financial Economics*, 7, pp. 133-159.

Hirshleifer, D., and Shumway, T., (2003). Good Day Sunshine: Stock Returns and the Weather. *The Journal of Finance*, 58(3), pp. 1009-1032.

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

Hirtle, B., Schuermann, T., and Stiroh, K.J., (2009). *Macroprudential supervision of financial institutions: lessons from the SCAP*. Federal Reserve Bank of New York Staff Report (No. 409).

Holler, J., (2012). Event-Study Methodology and Statistical Significance. In S. Müller and J. Prokop (Eds.), *Banking, Finance and Accounting Research Series* (pp. 1-399). OlWIR.

Hood, C., and Piotrowska, B., (2021). Goodhart's law and the gaming of UK public spending numbers. *Public Performance and Management Review*, 44(2), pp. 250-271.

Hoque, H., Andriosopoulos, D., Andriosopoulos, K., and Douady, R., (2015). Bank Regulation, Risk and Return: Evidence from the Credit and Sovereign Debt Crises. *Journal of Banking and Finance*, 50, pp. 455-474.

Hou, D. and Skeie, D., (2014). *LIBOR: Origins, economics, crisis, scandal, and reform*. Federal Reserve Bank of New York Staff Report No. 667.

Hou, K. and Moskowitz, T.J., (2005). Market frictions, price delay, and the cross-section of expected returns. *The Review of Financial Studies*, 18(3), pp. 981-1020.

Houpt, J.V. and Embersit, J.A., (1991). *A method for evaluating interest rate risk in U.S. commercial banks*. Federal Reserve Bulletin, 77, pp. 625-637.

Hossain, A.T. and Kryzanowski, L., (2021). Political corruption shielding and corporate acquisitions. *Financial Review*, 56(1), pp. 55-83.

Hu, X., Huang, H., Pan, Z., and Shi, J., (2019). Information Asymmetry and Credit Rating: A Quasi-Natural Experiment from China. *Journal of Banking and Finance*, 106, pp. 132-152.

Hundt, S., Sprungk, B., and Horsch, A., (2017). The information content of credit ratings: evidence from European convertible bond markets. *The European Journal of Finance*, 23(14), pp. 1414-1445.

Hwang, S., Keswani, A., and Shackleton, M.B., (2008). Surprise vs anticipated information announcements: Are prices affected differently? An investigation in the context of stock splits. *Journal of Banking and Finance*, 32(5), pp. 643-653.

Iannotta, G., (2006). Testing for opaqueness in the European banking industry: evidence from bond credit ratings. *Journal of Financial Services Research*, 30(3), pp. 287-309.

Ibbotson, R.G., (1975). Price performance of common stock new issues. *Journal of Financial Economics*, 2(3), pp. 235-272.

Ibbotson, R.G., Sindelar, J.L., and Ritter, J.R., (1994). The market's problems with the pricing of initial public offerings. *Journal of Applied Corporate Finance*, 7(1), pp. 66-74.

Ichikawa, J.J. and Steup, M., (2018). The Analysis of Knowledge. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2018 Ed.), Viewed: 11.4.2021, https://plato.stanford.edu/archives/sum2018/entries/knowledge-analysis/

Ikenberry, D.L. and Ramnath, S., (2002). Underreaction to self-selected news events: The case of stock splits. *The Review of Financial Studies*, 15(2), pp. 489-526.

Imenda, S., (2014). Is there a conceptual difference between theoretical and conceptual frameworks?. *Journal of Social Sciences*, 38(2), pp. 185-195.

International Monetary Fund (IMF), (2000). *Financial Sector Assessment Program (FSAP) – A Review: Lessons from the Pilot and Issues Going Forward*, viewed 9 December 2020, https://www.imf.org/external/np/fsap/2001/review.htm#II_A

International Monetary Fund (IMF), (2010). *IMF Expanding Surveillance to Require Mandatory Financial Stability Assessments of Countries with Systemically Important Financial Sectors*, Press Release (No. 10/357), 27 September 2010, available at: https://www.imf.org/en/News/Articles/2015/09/14/01/49/pr10357

International Monetary Fund (IMF), (2012). *Macrofinancial Stress Testing - Principles and Practices*. IMF Policy Paper. https://doi.org/10.5089/9781498340021.007

International Monetary Fund (IMF), (2014a). Review of the Financial Sector Assessment Program: Further Adaptation to the Post-Crisis Era, *IMF Staff Report*, August 2014, https://doi.org/10.5089/9781498342841.007

International Monetary Fund (IMF), (2014b). *IMF Executive Board Reviews Mandatory Financial Stability Assessments Under the Financial Sector Assessment Program*, Press Release (No. 14/08), 13 January, 2014, available at: https://www.imf.org/en/News/Articles/2015/09/14/01/49/pr1408

International Monetary Fund (IMF), (2014c). *Mandatory Financial Stability Assessments under the FSAP*. Viewed 9 December 2020, https://www.imf.org/external/np/fsap/mandatoryfsap.htm

International Monetary Fund (IMF), (2019). Financial Sector Assessment Program (FSAP), viewed 8 December 2020, https://www.imf.org/en/About/Factsheets/Sheets/2016/08/01/16/14/Financial-Sector-Assessment-Program

International Monetary Fund (IMF), (2020). Financial Sector Assessment Program (FSAP), viewed 8 December 2020, https://www.imf.org/en/Publications/fssa

Issing, O., (1997). Monetary targeting in Germany: The stability of monetary policy and of the monetary system. *Journal of Monetary Economics*, 39(1), pp. 67-79.

Ivković, Z. and Jegadeesh, N., (2004). The timing and value of forecast and recommendation revisions. *Journal of Financial Economics*, 73(3), pp. 433-463.

Jarrow, R.A. and Larsson, M., (2012). The meaning of market efficiency. *Mathematical Finance*, 22(1), pp. 1-30.

Jegadeesh, N. and Titman, S., (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1), pp. 65-91.

Jegadeesh, N. and Titman, S., (2001). Profitability of momentum strategies: An evaluation of alternative explanations. *The Journal of Finance*, 56(2), pp. 699-720.

Jegadeesh, N. and Titman, S., (2011). Momentum. *Annual Review of Financial Economics*, 3(1), pp. 493-509.

Jegadeesh, N., Kim, J., Krische, S.D., and Lee, C.M., (2004). Analyzing the analysts: When do recommendations add value?. *The Journal of Finance*, 59(3), pp. 1083-1124.

Jensen, M.C., (1968). The performance of mutual funds in the period 1945-1964. *The Journal of Finance*, 23(2), pp. 389-416.

Jensen, M.C., (1969). Risk, the pricing of capital assets, and the evaluation of investment portfolios. *The Journal of Business*, 42(2), pp. 167-247.

Jiang, G., Lee, C.M., and Zhang, Y., (2005). Information uncertainty and expected returns. *Review of Accounting Studies*, 10(2-3), pp. 185-221.

Jin, L. and Myers, S.C., (2006). R^2 around the world: new theory and new tests. *Journal of Financial Economics*, 79(2), pp. 257-292.

Jobst, A.A., Ong, L.L., and Schmieder, C., (2013). *A framework for macroprudential bank solvency stress testing: Application to S-25 and other G-20 country FSAPs*. IMF Working Paper (No. 13-68), https://doi.org/10.5089/9781616355074.001

Jones, C.P. and Litzenberger, R.H., (1970). Quarterly earnings reports and intermediate stock price trends. *The Journal of Finance*, 25(1), pp. 143-148.

Jones, J.S., Lee, W.Y., and Yeager, T.J., (2012). Opaque banks, price discovery, and financial instability. *Journal of Financial Intermediation*, 21(3), pp. 383-408.

Jones, J.S., Lee, W.Y., and Yeager, T.J., (2013). Valuation and systemic risk consequences of bank opacity. *Journal of Banking and Finance*, 37(3), pp. 693-706.

Jones, M.T., Hilbers, P., and Slack, G., (2004). Stress Testing Financial Systems: What to Do When the Governor Calls. *IMF Working Paper*, 04/127, https://doi.org/10.5089/9781451855012.001

Joo, B.A. and Durri, K., (2018). Impact of psychological traits on rationality of individual investors. *Theoretical Economics Letters*, 8(11), pp. 1973-1986.

Jordan, J.S., Peek, J., and Rosengren, E.S., (2000). The market reaction to the disclosure of supervisory actions: Implications for bank transparency. *Journal of Financial Intermediation*, 9(3), pp. 298-319.

Joy, O.M., Litzenberger, R.H., and McEnally, R.W., (1977). The adjustment of stock prices to announcements of unanticipated changes in quarterly earnings. *Journal of Accounting Research*, 15(2), pp. 207-225.

Jungherr, J., (2018). Bank opacity and financial crises. *Journal of Banking and Finance*, 97, pp. 157-176.

Kahneman, D. and Tversky, A., (1979). Prospect Theory: An Analysis of Decision Under Risk. *Econometrica*, 47(2), pp. 263-291.

Kajurová, V., and Hvozdenská , J., (2016). Linkages between CDS, bond and stock markets: Evidence from Europe. *MENDELU Working Papers in Business and Economics 63/2016*. Mendel University in Brno.

Kalay, A. (1982). The ex-dividend day behavior of stock prices: a re-examination of the clientele effect. *The Journal of Finance*, 37(4), pp. 1059-1070.

Kalay, A. and Loewenstein, U., (1985). Predictable events and excess returns: The case of dividend announcements. *Journal of Financial Economics*, 14(3), pp. 423-449.

Kalirai, H. and Scheicher, M., (2002). Macroeconomic stress testing: preliminary evidence for Austria. *Financial Stability Report*, (3), pp. 58-74.

Kanas, A., (2014). Uncovering a positive risk-return relation: the role of implied volatility index. *Review of Quantitative Finance and Accounting*, 42(1), pp. 159-170.

Kapinos, P.S., Martin, C., and Mitnik, O.A., (2018). Stress testing banks: Whence and whither?. *Journal of Financial Perspectives*, 5(1), pp. 1-20.

Ke, B. and Ramalingegowda, S., (2005). Do institutional investors exploit the post-earnings announcement drift?. *Journal of Accounting and Economics*, 39(1), pp. 25-53.

Keim, D.B., (1983). Size-related anomalies and stock return seasonality: Further empirical evidence. *Journal of Financial Economics*, 12(1), pp. 13-32.

Kim, O. and Verrecchia, R.E., (1991). Market reaction to anticipated announcements. *Journal of Financial Economics*, 30(2), pp. 273-309.

Kok, C., Müller, C., and Pancaro, C., (2019). The disciplining effect of supervisory scrutiny on banks' risk-taking: evidence from the EU wide stress test. *Macroprudential Bulletin*, 9, pp. 1-19. Available at: https://www.ecb.europa.eu//pub/financial-stability/macroprudential-bulletin/html/ecb.mpbu201910_3~7da43c7c16.en.html

Kolaric, S., Kiesel, F., and Ongena, S., (2021). Market Discipline through Credit Ratings and Too-Big-to-Fail in Banking. *Journal of Money, Credit and Banking*, 53(2-3), pp. 367-400.

Kolodny, N. and Brunero, J., (2020). Instrumental Rationality. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2020 Ed.), Viewed: 8.4.2021, https://plato.stanford.edu/archives/spr2020/entries/rationality-instrumental/

Kon, S.J., (1983). The market-timing performance of mutual fund managers. *Journal of Business*, 56(3), pp. 323-347.

Koski, J.L. and Scruggs, J.T., (1998). Who trades around the ex-dividend day? Evidence from NYSE audit file data. *Financial Management*, 27(3), pp. 58-72.

Kosso, P., (2011). *A Summary of Scientific Method*. Springer.

Kothari, S.P. and Warner, J.B., (1997). Measuring long-horizon security price performance. *Journal of Financial Economics*, 43(3), pp. 301-339.

Kothari, S.P. and Warner, J.B., (2001). Evaluating mutual fund performance. *The Journal of Finance*, 56(5), pp. 1985-2010.

Kothari, S.P. and Warner, J.B., (2007). Econometrics of Event Studies. In B.E. Eckbo (Ed.), *Handbook of Corporate Finance: Empirical Corporate Finance (Vol. 1)* (pp. 3-36). North-Holland.

Krivin, D., Patton, R., Rose, E., and Tabak, D., (1997). Determination of the Appropriate Event Window Length in Individual Stock Event Studies. *Studies in Economics and Finance*, 35, pp. 13-39.

Kuhn, T.S., (1962). *The Structure of Scientific Revolutions*. University of Chicago Press.

Kullback, S. and Leibler, R.A., (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1), pp. 79-86.

Kwan, S.H. and Carleton, W.T., (2010). Financial contracting and the choice between private placement and publicly offered bonds. *Journal of Money, Credit and Banking*, 42(5), pp. 907-929.

Kyle, A.S., (1985). Continuous auctions and insider trading. *Econometrica*, 53(6), pp. 1315-1335.

Lahey, K.E. and Conn, R.L., (1990). Sensitivity of Acquiring Firms' Returns to Alternative Model Specifications and Disaggregation. *Journal of Business Finance and Accounting*, 17(3), pp. 421-439.

Lakonishok, J., and Lee, I., (2001). Are insider trades informative?. *The Review of Financial Studies*, 14(1), pp. 79-111.

Lakonishok, J. and Lev, B., (1987). Stock splits and stock dividends: Why, who, and when. *The Journal of Finance*, 42(4), pp. 913-932.

Lakonishok, J., Shleifer, A., and Vishny, R.W., (1994). Contrarian investment, extrapolation, and risk. *The Journal of Finance*, 49(5), pp. 1541-1578.

Lamoureux, C.G. and Poon, P., (1987). The market reaction to stock splits. *The Journal of Finance*, 42(5), pp. 1347-1370.

Langley, P., (2013). Anticipating uncertainty, reviving risk? On the stress testing of finance in crisis. *Economy and Society*, 42(1), pp. 51-73.

Larsen, G.A. and Resnick, B.G., (1999). A performance comparison between cross-sectional stochastic dominance and traditional event study methodologies. *Review of Quantitative Finance and Accounting*, 12(2), pp. 103-113.

Latané, H.A. and Jones, C.P., (1977). Standardized unexpected earnings – a progress report. *The Journal of Finance*, 32(5), pp. 1457-1465.

Latané, H.A. and Jones, C.P., (1979). Standardized unexpected earnings – 1971-77. *The Journal of Finance*, 34(3), pp. 717-724.

Lau, S.T., Diltz, J.D., and Apilado, V.P., (1994). Valuation effects of international stock exchange listings. *Journal of Banking and Finance*, 18(4), pp. 743-755.

Laughlin, R., (1995). Empirical research in accounting: alternative approaches and a case for "middle-range" thinking. *Accounting, Auditing and Accountability Journal*, 8(1), pp. 63-87.

Lehmann, B.N. and Modest, D.M., (1987). Mutual fund performance evaluation: A comparison of benchmarks and benchmark comparisons. *The Journal of Finance*, 42(2), pp. 233-265.

Leitner, Y. and Williams, B., (2017). *Model Secrecy and Stress Tests*. Federal Reserve Bank of Philadelphia Working Paper (No. 17-41).

Leland, H. and Pyle, D., (1977). Information Asymmetries, Financial Structure and Financial Intermediaries. *The Journal of Finance*, 32(2), pp. 371-387.

Lenkey, S.L., (2014). Advance disclosure of insider trading. *The Review of Financial Studies*, 27(8), pp. 2504-2537.

Lerman, A. and Livnat, J., (2010). The new Form 8-K disclosures. *Review of Accounting Studies*, 15(4), pp. 752-778.

Lester, J., Reynolds, P., Schuermann, T., and Walsh, D. (2012). *Strategic Capital: Defining an Effective Real World View of Capital*. Oliver Wyman Financial Services Report. Available at: http://www.oliverwyman.com/strategic-capital-defining-an-effectivereal-world-view-of-capital.htm.

Lev, B., (1989). On the usefulness of earnings and earnings research: Lessons and directions from two decades of empirical research. *Journal of Accounting Research*, 27, pp. 153-192.

Levis, M., (1993). The Long-Run Performance of Initial Public Offerings: The UK Experience 1980-1988. *Financial Management*, 22(1), pp. 28-41.

Levy, R.A., (1966). Conceptual foundations of technical analysis. *Financial Analysts Journal*, 22(4), pp. 83-89.

Lidén, E.R., (2006). Stock recommendations in Swedish printed media: leading or misleading?. *The European Journal of Finance*, 12(8), pp. 731-748.

Lincoln, Y.S., Lynham, S.A., and Guba, E.G., (2017). Paradigmatic Controversies, Contradictions, and Emerging Confluences, Revisited. In N.K. Denzin and Y.S. Lincoln (Eds.), *The SAGE Handbook of Qualitative Research* (5th Ed.) (pp. 108-150). SAGE Publications.

Lind, J., (1753). *A Treatise of the Scurvy. Of Three Parts Containing an Inquiry into the Nature, Causes and Cure of that Disease.* Sands, Murray and Cochran.

Lins, K.V., Volpin, P., and Wagner, H.F., (2013). Does family control matter? International evidence from the 2008-2009 financial crisis. *The Review of Financial Studies*, 26(10), pp. 2583-2619.

Lintner, J. (1965a). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The Review of Economics and Statistics*, 47(1), pp 13-37.

Lintner, J., (1965b). Security prices, risk, and maximal gains from diversification. *The Journal of Finance*, 20(4), pp. 587-615.

Livingston, M., Naranjo, A., and Zhou, L., (2007). Asset opaqueness and split bond ratings. *Financial Management*, 36(3), pp. 49-62.

Ljung, G.M. and Box, G.E., (1978). On a Measure of Lack of Fit in Time Series Models. *Biometrika*, 65(2), pp. 297-303.

Lo, A.W., (2004). The Adaptive Markets Hypothesis: Market Efficiency from an Evolutionary Perspective. *The Journal of Portfolio Management*, 30(5), pp. 15-29.

Lo, A.W., (2005). Reconciling Efficient Markets with Behavioral Finance: The Adaptive Markets Hypothesis. *Journal of Investment Consulting*, 7(2), pp. 21-44.

Lo, A.W., (2012). Adaptive Markets and the New World Order. *Financial Analysts Journal*, 68(2), pp. 18-29.

Lo, A.W., (2019). *Adaptive Markets: Financial Evolution at the Speed of Thought*. Princeton University Press, Princeton and Oxford.

Lo, A.W. and MacKinlay, A.C., (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *The Review of Financial Studies*, 1(1), pp. 41-66.

Lo, A.W. and MacKinlay, A.C., (2011). *A Non-Random Walk Down Wall Street*. Princeton University Press.

Lopez, J.A., (1999). 'Using CAMELS Ratings to Monitor Bank Conditions', FRBSF Economic Letter (No. 1999-19), Federal Reserve Bank of San Francisco (FRBSF), viewed 22 January 2021, https://www.frbsf.org/economic-research/publications/economic-letter/1999/june/using-camels-ratings-to-monitor-bank-conditions/#subhead1

Lorie, J.H. and Niederhoffer, V., (1968). Predictive and statistical properties of insider trading. *The Journal of Law and Economics*, 11(1), pp. 35-53.

Loughran, T. and Ritter, J.R., (1995). The new issues puzzle. *The Journal of Finance*, 50(1), pp. 23-51.

Loughran, T., Ritter, J.R., and Rydqvist, K., (1994). Initial public offerings: International insights. *Pacific-Basin Finance Journal*, 2(2-3), pp. 165-199.

Lucas, R.E. Jr., (1976). *Econometric Policy Evaluation: A Critique.* In K. Brunner and A.H. Meltzer (Eds.), The Phillips Curve and Labor Markets. Carnegie-Rochester Conference Series on Public Policy (Vol. 1) (pp. 19-46). North Holland.

Lundblad, C., (2007). The risk return tradeoff in the long run: 1836-2003. *Journal of Financial Economics*, 85(1), pp. 123-150.

MacKinlay, A.C., (1997). Event Studies in Economics and Finance. *Journal of Economic Literature*, 35(1), pp. 13-39.

Mandelbrot, B., (1963). The variation of certain speculative prices. *The Journal of Business*, 36(4), pp. 394-419.

Mandelbrot, B., (1966). Forecasts of future prices, unbiased markets, and" martingale" models. *The Journal of Business*, 39(1), pp. 242-255.

Manheim, D. and Garrabrant, S., (2019). Categorizing Variants of Goodhart's Law. *arXiv preprint arXiv*: 1803.04585v4.

Malkiel, B.G., (1989). *Efficient Market Hypothesis*. In J. Eatwell, M. Milgate, and P. Newman (Eds.), Finance (pp. 127-134). Palgrave Macmillan.

Malkiel, B.G., (2003). The efficient market hypothesis and its critics. *Journal of Economic Perspectives*, 17(1), pp. 59-82.

Markowitz, H., (1952). Portfolio Selection. *Journal of Finance*, 7(1), pp. 77-91.

Mawdsley, A., McGuire, M., and O'Donnell, N., (2004). The stress testing of Irish credit institutions. *Financial Stability Report*, 2004, pp. 103-109.

Maynes, E. and Rumsey, J., (1993). Conducting event studies with thinly traded stocks. *Journal of Banking and Finance*, 17(1), pp. 145-157.

McConnell, J.J. and Sanger, G.C. (1984). A trading strategy for new listings on the NYSE. *Financial Analysts Journal*, 40(1), pp. 34-38.

McConnell, P., (2013). Systemic operational risk: the LIBOR manipulation scandal. *Journal of Operational Risk*, 8(3), pp. 59-99.

McMullin, J.L., Miller, B.P., and Twedt, B.J., (2019). Increased mandated disclosure frequency and price formation: Evidence from the 8-K expansion regulation. *Review of Accounting Studies*, 24(1), pp. 1-33.

McNichols, M. and Trueman, B., (1994). Public disclosure, private information collection, and short-term trading. *Journal of Accounting and Economics*, 17(1-2), pp. 69-94.

Metzinger, T., (2011). The No-Self Alternative. In S. Gallagher (Ed.), *The Oxford Handbook of the Self* (pp. 279-296). Oxford University Press.

Meyer, B.D., (1995). Natural and Quasi-Experiments in Economics. *Journal of Business and Economic Statistics*, 13(2), pp. 151-161.

Miller, M.H. and Modigliani, F., (1961). Dividend policy, growth, and the valuation of shares. *The Journal of Business*, 34(4), pp. 411-433.

Miller, R.E. and Reilly, F.K., (1987). An examination of mispricing, returns, and uncertainty for initial public offerings. *Financial Management*, 16(2), pp. 33-38.

Min, B.K. and Kim, T.S., (2012). Are good-news firms riskier than bad-news firms?. *Journal of Banking and Finance*, 36(5), pp. 1528-1535.

Montesi, G. and Papiro, G., (2018). Bank stress testing: A stochastic simulation framework to assess banks' financial fragility. *Risks*, 6(3), pp. 82.

Moodley, N., Muller, C., and Ward, M., (2016). Director dealings as an investment strategy. *Studies in Economics and Econometrics*, 40(2), pp. 105-123.

Moretti, M., Stolz, S., and Swinburne, M., (2008). Stress Testing at the IMF., *IMF Working Paper*, WP/08/206. https://doi.org/10.5089/9781451870640.001

Morgan, D.P., (2002). Rating banks: Risk and uncertainty in an opaque industry. *American Economic Review*, 92(4), pp. 874-888.

Morgan, D.P. and Stiroh, K.J., (2001). Market discipline of banks: The asset test. *Journal of Financial Services Research*, 20(2-3), pp. 195-208.

Morgan, D.P., Peristiani, S., and Savino, V., (2014). The information value of the stress test. *Journal of Money, Credit and Banking*, 46(7), pp. 1479-1500.

Morck, R., Yeung, B., and Yu, W., (2000). The information content of stock markets: why do emerging markets have synchronous stock price movements?. *Journal of Financial Economics*, 58(1-2), pp. 215-260.

Morris, S. and Shin, H.S., (2002). Social value of public information. *American Economic Review*, 92(5), pp. 1521-1534.

Mossin, J., (1966). Equilibrium in a capital asset market. *Econometrica*, 34(4), pp. 768-783.

Mukhtarov, S., Schoute, M., and Wielhouwer, J.L., (2021). The information content of the Solvency II ratio relative to earnings. *Journal of Risk and Insurance*, pp. 1-30.

Muntermann, J., (2005). Automated Mobile Alerting Services-Towards a Level Playing Field in the Financial Community. *Journal of Electronic Commerce Research*, 6(3), pp. 241-250.

Muntermann, J. and Guettler, A., (2007). Intraday stock price effects of ad hoc disclosures: the German case. *Journal of International Financial Markets, Institutions and Money*, 17(1), pp. 1-24.

Muscarella, C.J. and Vetsuypens, M.R., (1989). A simple test of Baron's model of IPO underpricing. *Journal of Financial Economics*, 24(1), pp. 125-135.

Muscarella, C.J. and Vetsuypens, M.R., (1996). Stock splits: Signaling or liquidity? The case of ADR 'solo-splits'. *Journal of Financial Economics*, 42(1), pp. 3-26.

Musumeci, J.J. and Sinkey, J.F., (1990). The international debt crisis, investor contagion, and bank security returns in 1987: The Brazilian experience. *Journal of Money, Credit and Banking*, 22(2), pp. 209-220.

Myers, J.H. and Bakay, A.J., (1948). Influence of Stock Split-Ups on Market Price. *Harvard Business Review*, 26(2), pp. 251-255.

Myers, S.C. and Majluf, N.S., (1984). Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics*, 13(2), pp. 187-221.

Nagel, T., (1989). *The View from Nowhere*. Oxford University Press.

Nayak, S. and Prabhala, N.R., (2001). Disentangling the dividend information in splits: A decomposition using conditional event-study methods. *The Review of Financial Studies*, 14(4), pp. 1083-1116.

Nelson, C.R. and Siegel, A.F., (1987). Parsimonious modeling of yield curves. *The Journal of Business*, 60(4), pp. 473-489.

Nester, M., (1996). An Applied Statistician's Creed. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 45(4), pp. 401-410.

Nichols, W.D. and Brown, S.L., (1981). Assimilating earnings and split information: Is the capital market becoming more efficient?. *Journal of Financial Economics*, 9(3), pp. 309-315.

Nichols, W.D. and McDonald, B., (1983). Stock splits and market anomalies. *Financial Review*, 18(4), pp. 237-256.

Niederhoffer, V. and Osborne, M.F.M., (1966). Market making and reversal on the stock exchange. *Journal of the American Statistical Association*, 61(316), pp. 897-916.

Nier, E. and Baumann, U., (2006). Market discipline, disclosure and moral hazard in banking. *Journal of Financial Intermediation*, 15(3), pp. 332-361.

Norden, L., and Weber, M., (2009). The co-movement of credit default swap, bond and stock markets: An empirical analysis. *European Financial Management*, 15(3), pp. 529-562.

Nymand-Andersen, P., (2018). *Yield curve modelling and a conceptual framework for estimating yield curves: evidence from the European Central Bank's yield curves*. ECB Statistics Paper No. 27.

O'Gorman, K. and MacIntosh, R., (2015). *Research Methods for Business and Management* (2nd Ed.). Goodfellow Publishers.

Olsen, R.A., (1998). Behavioral finance and its implications for stock-price volatility. *Financial Analysts Journal*, 54(2), pp. 10-18.

Ong, L.L. (Ed.), (2014). *A Guide to IMF Stress Testing: Methods and Models*. International Monetary Fund, doi: https://doi.org/10.5089/9781484368589.071

Ong, L.L. and Čihák, M., (2014). Stress testing at the International Monetary Fund: Methods and Models. In L.L. Ong (Ed.), *A Guide to IMF Stress Testing: Methods and Models* (pp. 1-9). International Monetary Fund, doi: https://doi.org/10.5089/9781484368589.071

Ong, L.L. and Pazarbasioglu, C., (2014). Credibility and crisis stress testing. *International Journal of Financial Studies*, 2(1), pp. 15-81.

Pacicco, F., Vena, L., and Venegoni, A., (2020). Communication and financial supervision: How does disclosure affect market stability?. *Journal of Empirical Finance*, 57, pp. 1-15.

Petrella, G. and Resti, A., (2013). Supervisors as Information Producers: Do Stress Tests Reduce Bank Opaqueness? *Journal of Banking and Finance*, 37(12), pp. 5406-5420.

Petrella, G. and Resti, A., (2016). *The interplay between banks and markets: Supervisory stress test results and investor reactions*. In T. Beck and B. Casu (Eds.), The Palgrave Handbook of European Banking (pp. 409-447). Palgrave Macmillan.

Pettit, R.R. and Venkatesh, P.C., (1995). Insider trading and long-run return performance. *Financial Management*, 24(2), pp. 88-103.

Popper, K.R., (2002). *The Logic of Scientific Discovery*. Routledge. (Original work published 1935).

Porter, T.M., (1995). *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton University Press.

Posten, H.O., (1979). The robustness of the one-sample t-test over the Pearson system. *Journal of Statistical Computation and Simulation*, 9(2), pp. 133-149.

Poterba, J.M. and Summers, L.H., (1988). Mean reversion in stock prices: Evidence and implications. *Journal of Financial Economics*, 22(1), pp. 27-59.

Powell, T., (2020). *The Value of Knowledge: The Economics of Enterprise Knowledge and Intelligence*. De Gruyter.

Quagliariello, M., (2020). Are Stress Tests Beauty Contests? (and What We Can Do About it). *Journal of Risk Management in Financial Institutions*, 13(2), pp. 126-134.

Quijano, M., (2014). Information asymmetry in US banks and the 2009 bank stress test. *Economics Letters*, 123(2), pp. 203-205.

Rajgopal, S., Venkatachalam, M., and Kotha, S., (2002). Managerial actions, stock returns, and earnings: the case of business-to-business internet firms. *Journal of Accounting Research*, 40(2), pp. 529-556.

Rao, S., (2019). The Philosophical Paradigm of Financial Market Contagion Research. *International Journal of Management Concepts and Philosophy*, 12(3), pp. 278-295.

Reinganum, M.R., (1981). Misspecification of capital asset pricing: Empirical anomalies based on earnings' yields and market values. *Journal of Financial Economics*, 9(1), pp. 19-46.

Reinganum, M.R., (1983). The anomalous stock market behavior of small firms in January: Empirical tests for tax-loss selling effects. *Journal of Financial Economics*, 12(1), pp. 89-104.

Reiss, J., (2012). The Explanation Paradox. *Journal of Economic Methodology*, 19(1), pp. 43-62.

Reilly, F.K and Brown, K.C., (2012). *Analysis of Investments and Management of Portfolios* (10th Ed.), South-Western Cengage Learning.

Reilly, F.K. and Hatfield, K., (1969). Investor experience with new stock issues. *Financial Analysts Journal*, 25(5), pp. 73-80.

Remenyi, D., Williams, B., Money, A., and Swartz, E., (1998). *Doing Research in Business and Management: An Introduction to Process and Method*. SAGE Publications.

Rendleman Jr, R.J., Jones, C.P., and Latané, H.A., (1987). Further insight into the standardized unexpected earnings anomaly: Size and serial correlation effects. *Financial Review*, 22(1), pp. 131-144.

Riedel, C. and Wagner, N., (2015). Is risk higher during non-trading periods? The risk trade-off for intraday versus overnight market returns. *Journal of International Financial Markets, Institutions and Money*, 39, pp. 53-64.

Ritter, J.R., (1984). The "Hot Issue" Market of 1980. *The Journal of Business*, 57(2), pp. 215-240.

Ritter, J.R., (1991). The long-run performance of initial public offerings. *The Journal of Finance*, 46(1), pp. 3-27.

Rock, K., (1986). Why new issues are underpriced. *Journal of Financial Economics*, 15(1-2), pp. 187-212.

Roosenboom, P. and Van Dijk, M.A., (2009). The market reaction to cross-listings: Does the destination market matter?. *Journal of Banking and Finance*, 33(10), pp. 1898-1908.

Rosen, G., (2020). Abstract Objects. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2020 Ed.), Viewed: 8.4.2021, https://plato.stanford.edu/archives/spr2020/entries/abstract-objects/

Ross, S.A., (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3), pp. 341-60.

Roussanov, N., Ruan, H., and Wei, Y., (2021). Marketing mutual funds. *The Review of Financial Studies*, 34(6), pp. 3045-3094.

Rouwenhorst, K.G., (1998). International momentum strategies. *The Journal of Finance*, 53(1), pp. 267-284.

Rudd, A. and Rosenberg, B., (1980). The "market model" in investment management. *The Journal of Finance*, 35(2), pp. 597-607.

Sahin, C. and de Haan, J., (2016). Market Reactions to the ECB's Comprehensive Assessment. *Economics Letters*, 140, pp. 1-5.

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

Samuelson, P.A., (1965). Proof That Properly Anticipated Prices Fluctuate Randomly. *Management Review*, 6(2), pp. 41-49.

Sanger, G.C. and McConnell, J.J., (1986). Stock exchange listings, firm value, and security market efficiency: The impact of NASDAQ. *Journal of Financial and Quantitative Analysis*, 21(1), pp. 1-25.

Saunders, M., Lewis, P., and Thornhill, A., (2009a). Understanding Research Philosophies and Approaches. In M. Saunders, P. Lewis, and A. Thornhill (Eds.), *Research Methods for Business Students* (5^{th} Ed.) (pp. 106-135). Prentice Hall.

Saunders, M., Lewis, P., and Thornhill, A., (2009b). Formulating the Research Design. In M. Saunders, P. Lewis, and A. Thornhill (Eds.), *Research Methods for Business Students* (5^{th} Ed) (pp. 136-167). Prentice Hall.

Schechter, J., (2013). Deductive Reasoning. In H. Pashler (Ed.), *Encyclopedia of the Mind* (pp. 226-230). Sage Publications.

Schmider, E., Ziegler, M., Danay, E., Beyer, L., and Bühner, M., (2010). Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution. *European Research Journal of Methods for the Behavioral and Social Sciences*, 6(4), pp. 147-151.

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

Schuermann, T., (2016). Stress Testing in Wartime and in Peacetime. In R.W. Anderson (Ed.), *Stress Testing and Macroprudential Regulation: A Transatlantic Assessment* (pp. 125-139). CEPR Press.

Schuermann, T., (2020). Capital Adequacy Pre-and Postcrisis and the Role of Stress Testing. *Journal of Money, Credit and Banking*, 52(S1), pp. 87-105.

Schultz, P., (2003). Pseudo market timing and the long-run underperformance of IPOs. *The Journal of Finance*, 58(2), pp. 483-517.

Schultz, M. and Hatch, M.J., (1996). Living with multiple paradigms the case of paradigm interplay in organizational culture studies. *Academy of Management Review*, 21(2), pp. 529-557.

Schwarz, G., (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), pp. 461-464.

Schwert, G.W., (1983). Size and Stock Returns, and other Empirical Regularities. *Journal of Financial Economics*, 12(1), pp. 3-12.

Seyhun, H.N., (1986). Insiders' profits, costs of trading, and market efficiency. *Journal of Financial Economics*, 16(2), pp. 189-212.

Shadish, W.R., Cook, T.D., and Campbell, D.T., (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin.

Sharpe, W.F., (1963). A simplified model for portfolio analysis. *Management Science*, 9(2), pp. 277-293.

Sharpe, W.F., (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *The Journal of Finance*, 19(3), pp. 425-442.

Sharpe, W.F., (1966). Mutual Fund Performance. *The Journal of Business*, 39(1), pp. 119-138.

Shefrin, H. and Statman, M., (1985). The disposition to sell winners too early and ride losers too long: Theory and evidence. *The Journal of Finance*, 40(3), pp. 777-790.

Shefrin, H. and Statman, M., (1994). Behavioral capital asset pricing theory. *Journal of Financial and Quantitative Analysis*, 29(3), pp. 323-349.

Sheng, A. and Looi, T.G., (2003). *Is there a Goodhart's Law in Financial Regulation?*. In P. Mizen (Ed.), Monetary History, Exchange Rates and Financial Markets: Essays in Honour of Charles Goodhart (Vol. 2) (pp. 234-249). Edward Elgar.

Shope, R.K., (2017). *The Analysis of Knowing: A Decade of Research*. Princeton University Press.

Shiller, R.J., (1984). Stock Prices and Social Dynamics. *Brookings Papers on Economic Activity*, 15(2), pp. 457-510.

Shiller, R.J., (1990). Market volatility and investor behavior. *The American Economic Review*, 80(2), pp. 58-62.

Shiller, R.J., (1999). *Human Behavior and the Efficiency of the Financial System*. In J.B. Taylor and M. Woodford (Eds.), Handbook of Macroeconomics (Vol. 1C). (pp. 1305-1340). North-Holland.

Shiller, R.J., (2003). From efficient markets theory to behavioral finance. *Journal of Economic Perspectives*, 17(1), pp. 83-104.

Sierra, G.E. and Yeager, T.J., (2004). *What Does the Federal Reserve's Economic Value Model Tell Us About Interest Rate Risk at US Community Banks?*. Federal Reserve Bank of St. Louis Review, 86(6), pp. 45-60.

Smirlock, M. and Kaufold, H., (1987). Bank foreign lending, mandatory disclosure rules, and the reaction of bank stock prices to the Mexican debt crisis. *Journal of Business*, pp. 347-364.

Smirlock, M. and Starks, L., (1986). Day-of-the-week and intraday effects in stock returns. *Journal of Financial Economics*, 17(1), pp. 197-210.

Solnik, B.H., (1973). Note on the validity of the random walk for European stock prices. *The Journal of Finance*, 28(5), pp. 1151-1159.

Sorge, M., (2004). Stress-Testing Financial Systems: An Overview of Current Methodologies. *BIS Working Paper No. 165*. Available online at: https://www.bis.org/publ/work165.htm

Sorge, M. and Virolainen, K., (2006). A comparative analysis of macro stress-testing methodologies with application to Finland. *Journal of Financial Stability*, 2(2), pp. 113-151.

Spargoli, F. (2013). Bank Recapitalization and the Information Value of a Stress Test in a Crisis.

Spargoli, F. and Upper, C., (2018). *Are Banks Opaque? Evidence from Insider Trading.* BIS Working Paper (No. 697).

Stambaugh, R.F., (1982). On the exclusion of assets from tests of the two-parameter model: A sensitivity analysis. *Journal of Financial Economics*, 10(3), pp. 237-268.

Stapleton, R.C. and Subrahmanyam, M.G., (1983). The market model and capital asset pricing theory: a note. *The Journal of Finance*, 38(5), pp. 1637-1642.

Stehle, R., Ehrhardt, O., and Przyborowsky, R., (2000). Long-run stock performance of German initial public offerings and seasoned equity issues. *European Financial Management*, 6(2), pp. 173-196.

Steup, M. and Neta, R., (2020). Epistemology. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2020 Ed.), Viewed: 11.4.2021, https://plato.stanford.edu/archives/spr2020/entries/epistemology/

Strathern, M., (1997). 'Improving Ratings': Audit in the British University System. European Review, 5(3), pp. 305-321.

Summers, L.H., (1986). Does the stock market rationally reflect fundamental values?. *The Journal of Finance*, 41(3), pp. 591-601.

Svensson, L.E.O., (1994). Estimating and Interpreting Forward Interest Rates: Sweden 1992-1994, NBER Working Paper No. 4871.

Taboga, M., (2014). What is a Prime Bank? A EURIBOR-OIS Spread Perspective. International Finance, 17(1), pp. 51-75.

Taleb, N.N., and Douady, R., (2013). Mathematical definition, mapping, and detection of (anti)fragility. *Quantitative Finance*, 13(11), pp. 1677-1689.

Taleb, N.N., Canetti, E., Kinda, T., Loukoianova, E., and Schmieder, C., (2012). *A New Heuristic Measure of Fragility and Tail Risks: Application to Stress Testing.* IMF Working Paper (No. 12/216).

Tanzi, V., (2013). *Dollars, Euros, and Debt: How we got Into Fiscal Drisis, and how we get out of it*. Palgrave Macmillan.

Taylor, J.B. and Williams, J.C., (2009). A black swan in the money market. *American Economic Journal: Macroeconomics*, 1(1), pp. 58-83.

Teney, D., Abbasnejad, E., Kafle, K., Shrestha, R., Kanan, C., and Van Den Hengel, A. (2020). On the value of out-of-distribution testing: An example of Goodhart's law. *Advances in Neural Information Processing Systems*, 33, pp. 407-417.

Teoh, S.H., Welch, I., and Wong, T.J., (1998). Earnings management and the long-run market performance of initial public offerings. *The Journal of Finance*, 53(6), pp. 1935-1974.

Tetlock, P.C., Saar-Tsechansky, M., and Macskassy, S., (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, 63(3), pp. 1437-1467.

Thaler, R.H., (1987). Anomalies: the January effect. *Journal of Economic Perspectives*, 1(1), pp. 197-201.

Thaler, R.H., (1999). The end of behavioral finance. *Financial Analysts Journal*, 55(6), pp. 12-17.

Thaler, R.H. (Ed.), (2005). *Advances in Behavioral Finance* (Vol. 2). Princeton University Press.

Thicke, M., (2018). Market epistemology. *Synthese*, 195(12), pp. 5571-5594.

Thompson, R., (1995). *Empirical Methods of Event Studies in Corporate Finance*. In R.A. Jarrow, V. Maksimovic, and W.T. Ziemba (Eds.), Handbooks in Operations Research and Management Science,Vol. 9 (pp. 963-992). North-Holland.

Thornton, D.L., (2008). *The unusual behavior of the federal funds and 10-year Treasury rates: A conundrum or Goodhart's Law?*. Federal Reserve Bank of St. Louis, Working Paper (No. 2007-039).

Torchio, F., (2009). Proper Event Study Analysis in Securities Litigation. *The Journal of Corporation Law*, 35, pp. 159-168.

Truong, C., (2011). Post-earnings announcement abnormal return in the Chinese equity market. *Journal of International Financial Markets, Institutions and Money*, 21(5), pp. 637-661.

Van Dijk, M.A., (2011). Is Size Dead? A Review of the Size Effect in Equity Returns. *Journal of Banking and Finance*, 35(12), pp. 3263-3274.

Van Roy, P., (2013). Is there a difference between solicited and unsolicited bank ratings and, if so, why?. *Journal of Financial Services Research*, 44(1), pp. 53-86.

Veldkamp, L.L., (2006). Information markets and the comovement of asset prices. *The Review of Economic Studies*, 73(3), pp. 823-845.

Wagner, W., (2007). Financial development and the opacity of banks. *Economics Letters*, 97(1), pp. 6-10.

Wall, L.D., (2014a). The adoption of stress testing: Why the Basel capital measures were not enough. *Journal of Banking Regulation*, 15(3-4), pp. 266-276.

Wall, L.D., (2014b). Measuring capital adequacy: supervisory stress-tests in a Basel world. *Journal of Financial Perspectives*, 2(1), pp. 85-94.

Wang, Y.C. and Chou, R.K., (2018). The Impact of Share Pledging Regulations on Stock Trading and Firm Valuation. *Journal of Banking and Finance*, 89, pp. 1-13.

Watts, R.L., (1978). Systematic 'abnormal' returns after quarterly earnings announcements. *Journal of Financial Economics*, 6(2-3), pp. 127-150.

Wellink, A.H.E.M., (1996). *Budgetary Control: Goodhart's Law in Government Finances?*. In C. Kool, J. Muysken, and T. van Veen (Eds.), Essays on Money, Banking, and Regulation (pp. 69-90. Kluwer Academic Publishers.

White, W.R., (2017). Conducting monetary policy in a complex, adaptive economy: Past mistakes and future possibilities. *Credit and Capital Markets*, 50(2), pp. 213-235.

Williams, B., (2011). *Ethics and the Limits of Philosophy*. Routledge.

Williamson, J.P., (1972). Measurement and forecasting of mutual fund performance: choosing an investment strategy. *Financial Analysts Journal*, 28(6), pp. 78-84.

Williamson, T., (2002), *Knowledge and Its Limits*. Oxford University Press.

Willmott, C.J., (2009). On the Evaluation of Model Performance in Physical Geography. In G.L. Gaile and C.J. Willmott (Eds.), *Spatial Statistics and Models* (pp. 443-460). Springer.

Womack, K.L., (1996). Do brokerage analysts' recommendations have investment value?. *The Journal of Finance*, 51(1), pp. 137-167.

World Bank, (2020). Financial Sector Assessment Program (FSAP): Overview, viewed 8 December 2020, https://www.worldbank.org/en/programs/financial-sector-assessment-program

Yazici, B. and Muradoğlu, G., (2002). Dissemination of stock recommendations and small investors: who benefits?. *Multinational Finance Journal*, 6(1), pp. 29-42.

Ying, L.K., Lewellen, W.G., Schlarbaum, G.G., and Lease, R.C., (1977). Stock exchange listings and securities returns. *Journal of Financial and Quantitative Analysis*, 12(3), pp. 415-432.

Zarowin, P., (1989). Does the stock market overreact to corporate earnings information?. The Journal of Finance, 44(5), pp. 1385-1399.

List of Appendices

Appendix A – Standard Event Window Determination

The table below shows descriptive statistics on the length and distribution of event windows around the event date t_0 used in previous studies of US and EU-wide stress tests. The results from the descriptive statistics were used to determine the overall length and distribution of the standard event window used in this study to ensure the best possible cross-study comparability.

	Pre-Event Days ($\leq t_0$) of the Event Windows[a]				**Post-Event Days ($> t_0$) of the Event Windows**[a]				**Total Days of the Event Windows (Entire Event Window Length)**[a]			
Group of Previous Studies	***Mdn***	***Min***	***Max***	**Range**	***Mdn***	***Min***	***Max***	**Range**	***Mdn***	***Min***	***Max***	**Range**
Studies of EU-wide stress tests[b] (*n* = 9)	2	1	23	22	2	0	22	22	5	1	45	44
Studies of US stress tests[c] (*n* = 8)	2	1	11	10	1	0	10	10	3	1	21	20
Studies of EU-wide and US stress tests (*n* = 15)[d]	2	1	23	22	1	0	22	22	5	1	45	44

Note. t_0 = event date. *Mdn* = median. *Min* = minimum. *Max* = maximum. EU = European Union. US = United States. *n* = sample size (number of studies).

[a] In trading days. [b] Ahnert *et al.* (2020), Alves *et al.* (2015), Borges *et al.* (2019), Candelon and Sy (2015), Cardinali and Nordmark (2011), Georgescu *et al.* (2017), Georgoutsos and Moratis (2021), Gerhardt and Vander Vennet (2017), and Petrella and Resti (2013). [c] Ahnert *et al.* (2020), Candelon and Sy (2015), Fernandes *et al.* (2020), Flannery *et al.* (2017), Fung and Loveland (2020), Morgan *et al.* (2014), Quijano (2014), and Sahin *et al.* (2020). [d] This group of studies includes the studies of US and EU-wide stress tests stress tests listed above; it should be noted, however, that the studies by Ahnert *et al.* (2020) and Candelon and Sy (2015) have examined both US and EU-wide stress tests and are therefore included in both of the above groups, but only once in this common group.

Appendix B – Post-Event Window Determination

B.1 Ljung-Box Tests – Detailed Results

The table below shows the detailed results of the Ljung-Box (1978) tests performed for each sample bank to determine the individual post-event window lengths $(+1, n)$. That is, to determine the numerical value of the term n included in the above definition of the post-event window. This numerical value can be found in the "Length" columns of the table for each individual sample bank; the values given there represent trading days. The results reported in the table are based on a time series length of $T = 25$ and a number of lags tested $L = 3$. However, the results were robust to variations in both parameters (not reported).

Country	Bank	Post-Event Windows (T = 25, L = 3)														
		CEBS 2010 (n = 50)			EBA 2011 (n = 51)			EBA 2014 (n = 59)			EBA 2016 (n = 34)			EBA 2018 (n = 33)		
		Start	End	Length[a]	Start	End	Length[a]	Start	End	Length[a]	Start	End	Length[a]	Start	End	Length[a]
AT	Erste Group Bank	26-Jul	26-Jul	1	18-Jul	20-Jul	3	27 Oct.	27 Oct.	1	1 Aug.	1 Aug.	1	5 Nov.	5 Nov.	1
BE	KBC	26-Jul	26-Jul	1	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	1 Aug.	1 Aug.	1	5 Nov.	5 Nov.	1
DE	Commerzbank	26-Jul	26-Jul	1	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	1 Aug.	2 Aug.	2	5 Nov.	5 Nov.	1
DE	Deutsche Bank	26-Jul	27-Jul	2	18-Jul	19-Jul	2	27 Oct.	27 Oct.	1	1 Aug.	1 Aug.	1	5 Nov.	5 Nov.	1
DK	Danske Bank	26-Jul	26-Jul	1	18-Jul	21-Jul	4	27 Oct.	27 Oct.	1	1 Aug.	1 Aug.	1	5 Nov.	5 Nov.	1
DK	Jyske Bank	26-Jul	29-Jul	4	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	1 Aug.	1 Aug.	1	5 Nov.	6 Nov.	2
ES	Banco Bilbao Vizcaya Argentaria	26-Jul	26-Jul	1	18-Jul	18-Jul	1	27 Oct.	29 Oct.	3	1 Aug.	1 Aug.	1	5 Nov.	5 Nov.	1
ES	Banco de Sabadell	26-Jul	26-Jul	1	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	1 Aug.	4 Aug.	4	5 Nov.	5 Nov.	1

ES	Banco Santander	26-Jul	26-Jul	1	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	1 Aug.	4 Aug.	4	5 Nov.	5 Nov.	1
ES	Caixabank	26-Jul	26-Jul	1	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	1 Aug.	2 Aug.	2	5 Nov.	5 Nov.	1
FR	BNP Paribas	26-Jul	26-Jul	1	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	1 Aug.	2 Aug.	2	5 Nov.	5 Nov.	1
FR	Crédit Agricole	26-Jul	26-Jul	1	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	1 Aug.	3 Aug.	3	5 Nov.	5 Nov.	1
FR	Société Générale	26-Jul	26-Jul	1	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	1 Aug.	1 Aug.	1	5 Nov.	5 Nov.	1
GB	Barclays Bank	26-Jul	28-Jul	3	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	1 Aug.	1 Aug.	1	5 Nov.	5 Nov.	1
GB	HSBC	26-Jul	26-Jul	1	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	1 Aug.	1 Aug.	1	5 Nov.	5 Nov.	1
GB	Lloyds Banking Group	26-Jul	26-Jul	1	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	1 Aug.	1 Aug.	1	5 Nov.	5 Nov.	1
GB	Royal Bank of Scotland	26-Jul	26-Jul	1	18-Jul	20-Jul	3	27 Oct.	27 Oct.	1	1 Aug.	1 Aug.	1	5 Nov.	5 Nov.	1
HU	OTP Bank	26-Jul	26-Jul	1	18-Jul	21-Jul	4	27 Oct.	27 Oct.	1	1 Aug.	3 Aug.	3	5 Nov.	5 Nov.	1
IE	Allied Irish Banks	26-Jul	26-Jul	1	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	1 Aug.	2 Aug.	2	5 Nov.	5 Nov.	1
IE	Bank of Ireland	26-Jul	26-Jul	1	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	1 Aug.	1 Aug.	1	5 Nov.	7 Nov.	3
IT	Intesa Sanpaolo	26-Jul	26-Jul	1	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	1 Aug.	1 Aug.	1	5 Nov.	5 Nov.	1
IT	UBI Banca	26-Jul	26-Jul	1	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	1 Aug.	1 Aug.	1	5 Nov.	8 Nov.	4
IT	Unicredit	26-Jul	26-Jul	1	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	1 Aug.	4 Aug.	4	5 Nov.	5 Nov.	1
NL	ING Bank	26-Jul	26-Jul	1	18-Jul	18-Jul	1	27 Oct.	28 Oct.	2	1 Aug.	1 Aug.	1	5 Nov.	5 Nov.	1
PL	Powszechna Kasa Oszczędności Bank Polski	26-Jul	26-Jul	1	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	1 Aug.	1 Aug.	1	5 Nov.	5 Nov.	1
SE	Skandinaviska Enskilda Banken	26-Jul	26-Jul	1	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	1 Aug.	1 Aug.	1	5 Nov.	5 Nov.	1
SE	Svenska Handelsbanken	26-Jul	26-Jul	1	18-Jul	21-Jul	4	27 Oct.	27 Oct.	1	1 Aug.	1 Aug.	1	5 Nov.	5 Nov.	1
SE	Swedbank	26-Jul	26-Jul	1	18-Jul	21-Jul	4	27 Oct.	27 Oct.	1	1 Aug.	1 Aug.	1	5 Nov.	5 Nov.	1
ES	Banco Popular Español	26-Jul	26-Jul	1	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	1 Aug.	1 Aug.	1	NA	NA	NA
IT	Banca Monte dei Paschi di Siena	26-Jul	26-Jul	1	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	1 Aug.	1 Aug.	1	NA	NA	NA
IT	Banco Popolare	26-Jul	26-Jul	1	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	1 Aug.	1 Aug.	1	NA	NA	NA
SE	Nordea Bank	26-Jul	26-Jul	1	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	1 Aug.	2 Aug.	2	NA	NA	NA
BE	Dexia	26-Jul	26-Jul	1	18-Jul	18-Jul	1	27 Oct.	29 Oct.	3	NA	NA	NA	NA	NA	NA
DK	Sydbank	26-Jul	28-Jul	3	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	NA	NA	NA	NA	NA	NA
ES	Bankinter	26-Jul	26-Jul	1	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	NA	NA	NA	NA	NA	NA
GR	Alpha Bank	26-Jul	26-Jul	1	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	NA	NA	NA	NA	NA	NA
GR	Eurobank Ergasias	26-Jul	26-Jul	1	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	NA	NA	NA	NA	NA	NA
GR	National Bank of Greece	26-Jul	26-Jul	1	18-Jul	19-Jul	2	27 Oct.	27 Oct.	1	NA	NA	NA	NA	NA	NA
GR	Piraeus Bank	26-Jul	26-Jul	1	18-Jul	19-Jul	2	27 Oct.	27 Oct.	1	NA	NA	NA	NA	NA	NA
MT	Bank of Valletta	26-Jul	26-Jul	1	18-Jul	19-Jul	2	27 Oct.	27 Oct.	1	NA	NA	NA	NA	NA	NA

PT	Banco BPI	26-Jul	26-Jul	1	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	NA	NA	NA	NA	NA	NA
PT	Banco Comercial Portugues	26-Jul	26-Jul	1	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	NA	NA	NA	NA	NA	NA
CY	Cyprus Popular Bank	26-Jul	26-Jul	1	18-Jul	19-Jul	2	NA	NA	NA	NA	NA	NA	NA	NA	NA
DE	Landesbank Berlin	26-Jul	27-Jul	2	18-Jul	18-Jul	1	NA	NA	NA	NA	NA	NA	NA	NA	NA
ES	Banco Pastor	26-Jul	26-Jul	1	18-Jul	18-Jul	1	NA	NA	NA	NA	NA	NA	NA	NA	NA
GR	Agricultural Bank of Greece	26-Jul	26-Jul	1	18-Jul	18-Jul	1	NA	NA	NA	NA	NA	NA	NA	NA	NA
GR	TT Hellenic Postbank	26-Jul	26-Jul	1	18-Jul	18-Jul	1	NA	NA	NA	NA	NA	NA	NA	NA	NA
DE	Deutsche Postbank	26-Jul	26-Jul	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
ES	Banco Guipuzcoano	26-Jul	26-Jul	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
HU	Takarék Jelzálogbank	26-Jul	26-Jul	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
NO	DNB Bank	NA	NA	NA	18-Jul	18-Jul	1	27 Oct.	27 Oct.	1	1 Aug.	1 Aug.	1	5 Nov.	5 Nov.	1
IE	Permanent TSB	NA	NA	NA	18-Jul	18-Jul	1	27 Oct.	30 Oct.	4	NA	NA	NA	NA	NA	NA
AT	Raiffeisen Bank International	NA	NA	NA	18-Jul	19-Jul	2	NA	NA	NA	NA	NA	NA	5 Nov.	5 Nov.	1
SI	Nova Kreditna Banka Maribor	NA	NA	NA	18-Jul	21-Jul	4	NA	NA	NA	NA	NA	NA	NA	NA	NA
DE	Aareal Bank	NA	NA	NA	NA	NA	NA	27 Oct.	29 Oct.	3	NA	NA	NA	NA	NA	NA
DE	IKB Deutsche Industriebank	NA	NA	NA	NA	NA	NA	27 Oct.	30 Oct.	4	NA	NA	NA	NA	NA	NA
ES	Liberbank	NA	NA	NA	NA	NA	NA	27 Oct.	27 Oct.	1	NA	NA	NA	NA	NA	NA
IT	Banca Carige	NA	NA	NA	NA	NA	NA	27 Oct.	27 Oct.	1	NA	NA	NA	NA	NA	NA
IT	Banca Piccolo Credito Valtellinese	NA	NA	NA	NA	NA	NA	27 Oct.	27 Oct.	1	NA	NA	NA	NA	NA	NA
IT	Banca Popolare di Milano	NA	NA	NA	NA	NA	NA	27 Oct.	27 Oct.	1	NA	NA	NA	NA	NA	NA
IT	Banca Popolare di Sondrio	NA	NA	NA	NA	NA	NA	27 Oct.	27 Oct.	1	NA	NA	NA	NA	NA	NA
IT	BPER Banca	NA	NA	NA	NA	NA	NA	27 Oct.	27 Oct.	1	NA	NA	NA	NA	NA	NA
IT	Credito Emiliano	NA	NA	NA	NA	NA	NA	27 Oct.	30 Oct.	4	NA	NA	NA	NA	NA	NA
IT	Mediobanca	NA	NA	NA	NA	NA	NA	27 Oct.	30 Oct.	4	NA	NA	NA	NA	NA	NA
PL	Alior Bank	NA	NA	NA	NA	NA	NA	27 Oct.	27 Oct.	1	NA	NA	NA	NA	NA	NA
PL	Bank BPH	NA	NA	NA	NA	NA	NA	27 Oct.	27 Oct.	1	NA	NA	NA	NA	NA	NA
PL	Bank Handlowy w Warszawie	NA	NA	NA	NA	NA	NA	27 Oct.	27 Oct.	1	NA	NA	NA	NA	NA	NA
PL	Bank Ochrony Środowiska	NA	NA	NA	NA	NA	NA	27 Oct.	27 Oct.	1	NA	NA	NA	NA	NA	NA
PL	Getin Noble Bank	NA	NA	NA	NA	NA	NA	27 Oct.	27 Oct.	1	NA	NA	NA	NA	NA	NA
NL	ABN AMRO	NA	NA	NA	NA	NA	NA	NA	NA	NA	1 Aug.	1 Aug.	1	5 Nov.	5 Nov.	1
IT	Banco BPM	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	5 Nov.	5 Nov.	1

PL	Bank Polska Kasa Opieki	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	5 Nov.	5 Nov.	1

Note. T = length of the time series used for the Ljung-Box (1978) test. L = number of lags tested in the Ljung-Box (1978) test. CEBS = Committee of European Banking Supervisors. EBA = European Banking Authority. n = sample size. AT = Austria. BE = Belgium. CY = Cyprus. DE = Germany. DK = Denmark. ES = Spain. FR = France. GB = United Kingdom of Great Britain and Northern Ireland. GR = Greece. HU = Hungary. IE = Ireland. IT = Italy. MT = Malta. NL = Netherlands. NO = Norway. PL = Poland. PT = Portugal. SE = Sweden. SI = Slovenia. NA = not applicable (bank was not subjected to the corresponding EU-wide stress test).

[a] In trading days.

B.2 Ljung-Box Tests – Descriptive Statistics

The table below shows descriptive statistics of the detailed results of the Ljung-Box (1978) tests performed for each sample bank to determine the individual length n of the post-event window $(+1, n)$. The values in the table represent trading days, except for the relative frequency in numbers and percentages (Freq.). The detailed results of the Ljung-Box (1978) tests are reported in Appendix B.1 above.

	EU-Wide Stress Tests					
	CEBS 2010 (*n* = 50)	**EBA 2011 (*n* = 51)**	**EBA 2014 (*n* = 59)**	**EBA 2016 (*n* = 34)**	**EBA 2018 (*n* = 33)**	**Total (*n* = 227)**
M	1.180	1.490	1.322	1.529	1.182	1.339
Mdn	1	1	1	1	1	1
Mode	1	1	1	1	1	1
Min	1	1	1	1	1	1
Max	4	4	4	4	4	4
Freq. *L* = 1 (%)	45 (0.90)	38 (0.75)	51 (0.86)	24 (0.71)	30 (0 91)	188 (0.83)
Freq. *L* = 2 (%)	2 (0.04)	6 (0.12)	1 (0.02)	5 (0.15)	1 (0.03)	15 (0.07)
Freq. *L* = 3 (%)	2 (0.04)	2 (0.04)	3 (0.05)	2 (0.06)	1 (0.03)	10 (0.04)
Freq. *L* = 4 (%)	1 (0.02)	5 (0.10)	4 (0.07)	3 (0.09)	1 (0.03)	14 (0.06)

Note. CEBS = Committee of European Banking Supervisors. EBA = European Banking Authority. *n* = sample size. *M* = mean. *Mdn* = median. *Min* = minimum. *Max* = maximum. Freq. = relative frequency in numbers and percentages. *L* = length of the post-event window (in trading days), where *L* = 1 is the event date (t_0).

Appendix C – List of the Population

The table below lists all banks that have been subjected to the five EU-wide stress tests examined. It is therefore a complete list of the population of this study. The population size ranged from $N = 48$ to $N = 123$ banks. The dichotomous variables "Yes" and "No" shown in the table indicate whether or not a bank has been subjected to the respective EU-wide stress test.

		EU-Wide Stress Tests				
Country	**Bank**	**CEBS 2010 ($N = 91$)**	**EBA 2011 ($N = 90$)**	**EBA 2014 ($N = 123$)**	**EBA 2016 ($N = 51$)**	**EBA 2018 ($N = 48$)**
AT	Erste Group Bank	Yes	Yes	Yes	Yes	Yes
BE	KBC	Yes	Yes	Yes	Yes	Yes
DE	Bayerische Landesbank	Yes	Yes	Yes	Yes	Yes
DE	Commerzbank	Yes	Yes	Yes	Yes	Yes
DE	Deutsche Bank	Yes	Yes	Yes	Yes	Yes
DE	Landesbank Baden-Württemberg	Yes	Yes	Yes	Yes	Yes
DE	Norddeutsche Landesbank	Yes	Yes	Yes	Yes	Yes
DK	Danske Bank	Yes	Yes	Yes	Yes	Yes
DK	Jyske Bank	Yes	Yes	Yes	Yes	Yes
ES	Banco Bilbao Vizcaya Argentaria	Yes	Yes	Yes	Yes	Yes
ES	Banco de Sabadell	Yes	Yes	Yes	Yes	Yes
ES	Banco Santander	Yes	Yes	Yes	Yes	Yes
ES	Caixabank	Yes	Yes	Yes	Yes	Yes
FR	BNP Paribas	Yes	Yes	Yes	Yes	Yes
FR	BPCE Group	Yes	Yes	Yes	Yes	Yes
FR	Crédit Agricole	Yes	Yes	Yes	Yes	Yes
FR	Société Générale	Yes	Yes	Yes	Yes	Yes
GB	Barclays Bank	Yes	Yes	Yes	Yes	Yes

GB	HSBC	Yes	Yes	Yes	Yes	Yes
GB	Lloyds Banking Group	Yes	Yes	Yes	Yes	Yes
GB	Royal Bank of Scotland	Yes	Yes	Yes	Yes	Yes
HU	OTP Bank	Yes	Yes	Yes	Yes	Yes
IE	Allied Irish Banks	Yes	Yes	Yes	Yes	Yes
IE	Bank of Ireland	Yes	Yes	Yes	Yes	Yes
IT	Intesa Sanpaolo	Yes	Yes	Yes	Yes	Yes
IT	UBI Banca	Yes	Yes	Yes	Yes	Yes
IT	Unicredit	Yes	Yes	Yes	Yes	Yes
NL	ING Bank	Yes	Yes	Yes	Yes	Yes
PL	Powszechna Kasa Oszczędności Bank Polski	Yes	Yes	Yes	Yes	Yes
SE	Nordea Bank	Yes	Yes	Yes	Yes	Yes
SE	Skandinaviska Enskilda Banken	Yes	Yes	Yes	Yes	Yes
SE	Svenska Handelsbanken	Yes	Yes	Yes	Yes	Yes
SE	Swedbank	Yes	Yes	Yes	Yes	Yes
DE	Dekabank	Yes	Yes	Yes	Yes	No
ES	Banco Popular Español	Yes	Yes	Yes	Yes	No
IT	Banca Monte dei Paschi di Siena	Yes	Yes	Yes	Yes	No
IT	Banco Popolare	Yes	Yes	Yes	Yes	No
DE	DZ Bank	Yes	Yes	Yes	No	Yes
BE	Dexia	Yes	Yes	Yes	No	No
CY	Bank of Cyprus	Yes	Yes	Yes	No	No
DE	HSH Nordbank	Yes	Yes	Yes	No	No
DE	Hypo Real Estate	Yes	Yes	Yes	No	No
DE	Landesbank Berlin	Yes	Yes	Yes	No	No
DE	WGZ Bank	Yes	Yes	Yes	No	No
DK	Sydbank	Yes	Yes	Yes	No	No
ES	Banco Mare Nostrum[a]	Yes	Yes	Yes	No	No
ES	Bankinter	Yes	Yes	Yes	No	No
ES	Ibercaja[b]	Yes	Yes	Yes	No	No
ES	NCG Banco[c]	Yes	Yes	Yes	No	No
ES	Unicaja[d]	Yes	Yes	Yes	No	No

FI	OP-Pohjola Group	Yes	Yes	Yes	No	No
GR	Alpha Bank	Yes	Yes	Yes	No	No
GR	Eurobank Ergasias	Yes	Yes	Yes	No	No
GR	National Bank of Greece	Yes	Yes	Yes	No	No
GR	Piraeus Bank	Yes	Yes	Yes	No	No
LU	Banque et Caisse d'Epargne de l'Etat	Yes	Yes	Yes	No	No
MT	Bank of Valletta	Yes	Yes	Yes	No	No
NL	SNS Bank	Yes	Yes	Yes	No	No
PT	Banco BPI	Yes	Yes	Yes	No	No
PT	Banco Comercial Portugues	Yes	Yes	Yes	No	No
PT	Caixa Geral de Depósitos	Yes	Yes	Yes	No	No
SI	Nova Ljubljanska Banka	Yes	Yes	Yes	No	No
CY	Cyprus Popular Bank	Yes	Yes	No	No	No
DE	WestLB	Yes	Yes	No	No	No
ES	Banca Civica[e]	Yes	Yes	No	No	No
ES	Banca March	Yes	Yes	No	No	No
ES	Banco Pastor	Yes	Yes	No	No	No
ES	Base[f]	Yes	Yes	No	No	No
ES	Caja de Ahorros y Monte de Piedad de Ontinyent	Yes	Yes	No	No	No
ES	Caja Vital Kuxta[g]	Yes	Yes	No	No	No
ES	Caja3[h]	Yes	Yes	No	No	No
ES	Colonya[i]	Yes	Yes	No	No	No
ES	Diada[j]	Yes	Yes	No	No	No
ES	Espiga[k]	Yes	Yes	No	No	No
ES	Grupo BBK[l]	Yes	Yes	No	No	No
ES	Kuxta[m]	Yes	Yes	No	No	No
ES	Unnim[n]	Yes	Yes	No	No	No
GR	Agricultural Bank of Greece	Yes	Yes	No	No	No
GR	TT Hellenic Postbank	Yes	Yes	No	No	No
NL	Rabobank	Yes	Yes	No	No	No
PT	Espírito Santo Financial Group	Yes	Yes	No	No	No
DE	Landesbank Hessen-Thüringen	Yes	No	Yes	Yes	Yes

AT	Raiffeisen Zentralbank Oesterreich	Yes	No	Yes	No	No
DE	Deutsche Postbank	Yes	No	No	No	No
ES	Banco Guipuzcoano	Yes	No	No	No	No
ES	Caja Sol[o]	Yes	No	No	No	No
ES	Cajasur[p]	Yes	No	No	No	No
ES	Jupiter[q]	Yes	No	No	No	No
HU	Takarék Jelzálogbank	Yes	No	No	No	No
LU	Banque Raiffeisen	Yes	No	No	No	No
NL	ABN/Fortis Bank Nederland	Yes	No	No	No	No
DK	Nykredit	No	Yes	Yes	Yes	Yes
NL	ABN AMRO	No	Yes	Yes	Yes	Yes
NO	DNB Bank	No	Yes	Yes	Yes	Yes
ES	BFA Tenedora de Acciones	No	Yes	Yes	Yes	No
AT	Oesterreichische Volksbank	No	Yes	Yes	No	No
IE	Permanent TSB	No	Yes	Yes	No	No
SI	Nova Kreditna Banka Maribor	No	Yes	Yes	No	No
AT	Raiffeisen Bank International	No	Yes	No	No	Yes
ES	Effibank	No	Yes	No	No	No
BE	Belfius Banque	No	No	Yes	Yes	Yes
DE	NRW Bank	No	No	Yes	Yes	Yes
FR	Crédit Mutuel	No	No	Yes	Yes	Yes
FR	La Banque Postale	No	No	Yes	Yes	Yes
NL	Bank Nederlandse Gemeenten	No	No	Yes	Yes	Yes
NL	Coöperatieve Centrale Raiffeisen-Boerenleenbank	No	No	Yes	Yes	Yes
DE	Volkswagen Financial Services	No	No	Yes	Yes	No
AT	BAWAG PSK	No	No	Yes	No	No
AT	Raiffeisenlandesbank Niederösterreich-Wien	No	No	Yes	No	No
AT	Raiffeisenlandesbank Oberösterreich	No	No	Yes	No	No
BE	AXA Bank Europe	No	No	Yes	No	No
BE	Investar	No	No	Yes	No	No
CY	Co-operative Central Bank	No	No	Yes	No	No
CY	Hellenic Bank	No	No	Yes	No	No

DE	Aareal Bank	No	No	Yes	No	No
DE	Deutsche Apotheker- und Ärztebank	No	No	Yes	No	No
DE	HASPA Finanzholding	No	No	Yes	No	No
DE	IKB Deutsche Industriebank	No	No	Yes	No	No
DE	KfW IPEX-Bank	No	No	Yes	No	No
DE	Landeskreditbank Baden-Württemberg Förderbank	No	No	Yes	No	No
DE	Landwirtschaftliche Rentenbank	No	No	Yes	No	No
DE	Münchener Hypothekenbank	No	No	Yes	No	No
DE	Wüstenrot Bausparkasse	No	No	Yes	No	No
DE	Wüstenrot Pfandbriefbank	No	No	Yes	No	No
ES	Cajas Rurales Unidas	No	No	Yes	No	No
ES	Catalunya Banc	No	No	Yes	No	No
ES	Kutxabank	No	No	Yes	No	No
ES	Liberbank	No	No	Yes	No	No
FR	Banque PSA Finance	No	No	Yes	No	No
FR	BPI France	No	No	Yes	No	No
FR	Caisse de Refinancement de l'Habitat	No	No	Yes	No	No
FR	RCI Banque	No	No	Yes	No	No
FR	Société de Financement Local	No	No	Yes	No	No
IT	Banca Carige	No	No	Yes	No	No
IT	Banca Piccolo Credito Valtellinese	No	No	Yes	No	No
IT	Banca Popolare di Milano	No	No	Yes	No	No
IT	Banca Popolare di Sondrio	No	No	Yes	No	No
IT	Banca Popolare di Vicenza	No	No	Yes	No	No
IT	BPER Banca	No	No	Yes	No	No
IT	Credito Emiliano	No	No	Yes	No	No
IT	Iccrea	No	No	Yes	No	No
IT	Mediobanca	No	No	Yes	No	No
IT	Veneto Banca	No	No	Yes	No	No
LU	Precision Capital	No	No	Yes	No	No
LV	ABLV Bank	No	No	Yes	No	No
NL	Nederlandse Waterschapsbank	No	No	Yes	No	No

PL	Alior Bank	No	No	Yes	No	No
PL	Bank BPH	No	No	Yes	No	No
PL	Bank Handlowy w Warszawie	No	No	Yes	No	No
PL	Bank Ochrony Środowiska	No	No	Yes	No	No
PL	Getin Noble Bank	No	No	Yes	No	No
SI	Slovenska Izvozna in Razvojna Banka	No	No	Yes	No	No
FI	OP Osuuskunta	No	No	No	Yes	Yes
AT	Raiffeisen Landesbanken Holding	No	No	No	Yes	No
IT	Banco BPM	No	No	No	No	Yes
PL	Bank Polska Kasa Opieki	No	No	No	No	Yes

Note. CEBS = Committee of European Banking Supervisors. EBA = European Banking Authority. *N* = population size. AT = Austria. BE = Belgium. CY = Cyprus. DE = Germany. DK = Denmark. ES = Spain. FI = Finland. FR = France. GB = United Kingdom of Great Britain and Northern Ireland. GR = Greece. HU = Hungary. IE = Ireland. IT = Italy. LU = Luxembourg. LV = Latvia. MT = Malta. NL = Netherlands. NO = Norway. PL = Poland. PT = Portugal. SE = Sweden. SI = Slovenia.

[a] Caja de Ahorros de Murcia; Caixa Déstalvis del Penedes; Caja de Ahorros y Monte de Piedad de las Baleares (SA Nostra); Caja General de Ahorros de Granada. [b] Caja de Ahorros y Monte Piedad de Zaragoza, Aragon y Rioja. [c] Caja de Ahorros de Galicia; Caixa de Aforros de Vigo, Ourense e Pontevedera (Caixanova). [d] Monte de Piedad y Caja de Ahorros de Ronda, Cadiz, Almeria, Malaga, Antequera y Jaen. [e] Caja de Ahorros y Monte de Piedad de Navarra, Caja de Ahorros Municipal de Burgos y Caja General de Ahorros de Canarias. [f] Caja de Ahorros del Mediterráneo (CAM); Caja de Ahorros de Asturias; Caja de Ahorros de Santander y Cantabria; Caja de Ahorros y Monte de Piedad de Extremadura. [g] Caja de Ahorros de Vitoria y Alava. [h] Caja de Ahorros y Monte Piedad del Círculo Católico de Obreos de Burgos (Caja Círculo); Monte de Piedad y Caja General de Ahorros de Badajoz; Caja de Ahorros de la Inmaculada de Aragón. [i] Caixa d'Estalvis de Pollensa. [j] Caixa Déstalvis de Catalunya; Caixa Déstalvis de Tarragona; Caixa Déstalvis de Manresa. [k] Caja de Ahorros de Salamanca y Soria (Caja Duero); Caja de Espana de Inversiones Caja de Ahorros y Monte de Piedad (Caja Espana). [l] Bilbao Bizkaia Kutxa, Aurrezki Kuxta eta Bahitetxea. [m] Caja de Ahorros y Monte Piedad de Gipuzkoa y San Sebastian. [n] Caixa Déstalvis de Sabadell; Caixa Déstalvis de Terrassa; Caixa Déstalvis Comarcal de Manlleu. [o] Monte de Piedad y Caja de Ahorros San Fernando de Huelva, Jerez y Sevilla (Caja Sol); Caja de Ahorro Provincial de Guadalajara. [p] Caja de Ahorros y Monte de Piedad de Cordoba. [q] Caja de Ahorros y Monte de Piedad de Madrid (Caja Madrid); Caja de Ahorros de Valencia, Castellón y Alicante (Bancaja); Caixa Déstalvis Laietana; Caja Insular de Ahorros de Canarias; Caja de Ahorros y Monte de Piedad de Avila; Caja de Ahorros y Monte de Piedad de Segovia; Caja de Ahorros de la Rioja.

Appendix D – Elements of the Cross-Sectional Samples

The table below lists all banks included in the five cross-sectional samples, *i.e.* the elements of the samples. The dichotomous variables "Yes" and "No" shown in the table indicate whether a bank was included in the respective sample or not. The size of the cross-sectional samples ranged from $n = 33$ to $n = 59$ sample banks.

				EU-Wide Stress Tests (Cross-Sectional Samples)				
Country	**Bank**	**LEI**	**ISIN**	**CEBS 2010 ($n = 50$)**	**EBA 2011 ($n = 51$)**	**EBA 2014 ($n = 59$)**	**EBA 2016 ($n = 34$)**	**EBA 2018 ($n = 33$)**
AT	Erste Group Bank	PQOH26KWDF7CG10L6792	AT0000652011	Yes	Yes	Yes	Yes	Yes
BE	KBC	213800X3Q9LSAKRUWY91	BE0003565737	Yes	Yes	Yes	Yes	Yes
DE	Deutsche Bank	7LTWFZYICNSX8D621K86	DE0005140008	Yes	Yes	Yes	Yes	Yes
DE	Commerzbank	851WYGNLUQLFZBSYGB56	DE000CBK1001	Yes	Yes	Yes	Yes	Yes
DK	Danske Bank	MAES062Z21O4RZ2U7M96	DK0010274414	Yes	Yes	Yes	Yes	Yes
DK	Jyske Bank	3M5E1GQGKL17HI6CPN30	DK0010307958	Yes	Yes	Yes	Yes	Yes
ES	Banco Bilbao Vizcaya Argentaria	K8MS7FD7N5Z2WQ51AZ71	ES0113211835	Yes	Yes	Yes	Yes	Yes
ES	Banco de Sabadell	SI5RG2M0WQQLZCXKRM20	ES0113860A34	Yes	Yes	Yes	Yes	Yes
ES	Banco Santander	5493006QMFDDMYWIAM13	ES0113900J37	Yes	Yes	Yes	Yes	Yes
ES	Caixabank	7CUNS533WID6K7DGFI87	ES0140609019	Yes	Yes	Yes	Yes	Yes
FR	Crédit Agricole	969500TJ5KRTCJQWXH05	FR0000045072	Yes	Yes	Yes	Yes	Yes
FR	Société Générale	O2RNE8IBXP4R0TD8PU41	FR0000130809	Yes	Yes	Yes	Yes	Yes
FR	BNP Paribas	R0MUWSFPU8MPRO8K5P83	FR0000131104	Yes	Yes	Yes	Yes	Yes
GB	HSBC	MLU0ZO3ML4LN2LL2TL39	GB0005405286	Yes	Yes	Yes	Yes	Yes
GB	Lloyds Banking Group	549300PPXHEU2JF0AM85	GB0008706128	Yes	Yes	Yes	Yes	Yes
GB	Barclays Bank	G5GSEF7VJP5I7OUK5573	GB0031348658	Yes	Yes	Yes	Yes	Yes
GB	Royal Bank of Scotland	2138005O9XJIJN4JPN90	GB00B7T77214	Yes	Yes	Yes	Yes	Yes

HU	OTP Bank	529900W3MOO00A18X956	HU0000061726	Yes	Yes	Yes	Yes	Yes
IE	Bank of Ireland	Q2GQA2KF6XJ24W42G291	IE00BD1RP616	Yes	Yes	Yes	Yes	Yes
IE	Allied Irish Banks	3U8WV1YX2VMUHH7Z1Q21	IE00BF0L3536	Yes	Yes	Yes	Yes	Yes
IT	Intesa Sanpaolo	2W8N8UU78PMDQKZENC08	IT0000072618	Yes	Yes	Yes	Yes	Yes
IT	UBI Banca	81560097964CBDAED282	IT0003487029	Yes	Yes	Yes	Yes	Yes
IT	Unicredit	549300TRUWO2CD2G5692	IT0005239360	Yes	Yes	Yes	Yes	Yes
NL	ING Bank	3TK20IVIUJ8J3ZU0QE75	NL0011821202	Yes	Yes	Yes	Yes	Yes
PL	Powszechna Kasa Oszczędności Bank Polski	P4GTT6GF1W40CVIMFR43	PLPKO0000016	Yes	Yes	Yes	Yes	Yes
SE	Skandinaviska Enskilda Banken	F3JS33DEI6XQ4ZBPTN86	SE0000148884	Yes	Yes	Yes	Yes	Yes
SE	Swedbank	M312WZV08Y7LYUC71685	SE0000242455	Yes	Yes	Yes	Yes	Yes
SE	Svenska Handelsbanken	NHBDILHZTYCNBV5UYZ31	SE0007100599	Yes	Yes	Yes	Yes	Yes
BE	Dexia	D3K6HXMBBB6SK9OXH394	BE0974290224	Yes	Yes	Yes	No	No
CY	Cyprus Popular Bank	549300P8PCUCMISDC956	CY0000200119	Yes	Yes	No	No	No
DE	Landesbank Berlin	GTQYZJON3I7SXRNJTT73	DE0008023227	Yes	Yes	No	No	No
DK	Sydbank	GP5DT10VX1QRQUKVBK64	DK0010311471	Yes	Yes	Yes	No	No
ES	Bankinter	VWMYAEQSTOPNV0SUGU82	ES0113679I37	Yes	Yes	Yes	No	No
ES	Banco Pastor	549300UFE7EDE4N17L58	ES0113770434	Yes	Yes	No	No	No
ES	Banco Popular Español	80H66LPTVDLM0P28XF25	ES0113790226	Yes	Yes	Yes	Yes	No
GR	National Bank of Greece	5UMCZOEYKCVFAW8ZLO05	GRS003003027	Yes	Yes	Yes	No	No
GR	Piraeus Bank	M6AD1Y1KW32H8THQ6F76	GRS014003024	Yes	Yes	Yes	No	No
GR	Alpha Bank	5299009N55YRQC69CN08	GRS015003007	Yes	Yes	Yes	No	No
GR	Eurobank Ergasias	JEUVK5RWVJEN8W0C9M24	GRS323003012	Yes	Yes	Yes	No	No
GR	Agricultural Bank of Greece	NA	GRS414003004	Yes	Yes	No	No	No
GR	TT Hellenic Postbank	2138008SAAQBO2AA3G77	GRS492003009	Yes	Yes	No	No	No
IT	Banco Popolare	5493006P8PDBI8LC0O96	IT0005002883	Yes	Yes	Yes	Yes	No
IT	Banca Monte dei Paschi di Siena	J4CP7MHCXR8DAQMKIL78	IT0005218752	Yes	Yes	Yes	Yes	No
MT	Bank of Valletta	529900RWC8ZYB066JF16	MT0000020116	Yes	Yes	Yes	No	No
PT	Banco Comercial Portugues	JU1U6S0DG9YLT7N8ZV32	PTBCP0AM0015	Yes	Yes	Yes	No	No
PT	Banco BPI	3DM5DPGI3W6OU6GJ4N92	PTBPI0AM0004	Yes	Yes	Yes	No	No
SE	Nordea Bank	6SCPQ280AIY8EP3XFW53	SE0000427361	Yes	Yes	Yes	Yes	No
DE	Deutsche Postbank	QPA2KT0GZRLD6DKRHZ40	DE0008001009	Yes	No	No	No	No
ES	Banco Guipuzcoano	NA	ES0113580C31	Yes	No	No	No	No

HU	Takarék Jelzálogbank	5299007F4BUUY6S14E44	HU0000078175	Yes	No	No	No	No
AT	Raiffeisen Bank International	9ZHRYM6F437SQJ6OUG95	AT0000606306	No	Yes	No	No	Yes
IE	Permanent TSB	635400DTNHVYGZODKQ93	IE00BWB8X525	No	Yes	Yes	No	No
NO	DNB Bank	549300GKFG0RYRRQ1414	NO0010031479	No	Yes	Yes	Yes	Yes
SI	Nova Kreditna Banka Maribor	549300J0GSZ83GTKBZ89	SI0021104052	No	Yes	No	No	No
DE	Aareal Bank	EZKODONU5TYHW4PP1R34	DE0005408116	No	No	Yes	No	No
DE	IKB Deutsche Industriebank	PWEFG14QWWESISQ84C69	DE0008063306	No	No	Yes	No	No
ES	Liberbank	635400XT3V7WHLSFYY25	ES0168675090	No	No	Yes	No	No
IT	Mediobanca	PSNL19R2RXX5U3QWHI44	IT0000062957	No	No	Yes	No	No
IT	Banca Popolare di Milano	8156009BC82130E7FC43	IT0000064482	No	No	Yes	No	No
IT	BPER Banca	N747OI7JINV7RUUH6190	IT0000066123	No	No	Yes	No	No
IT	Banca Popolare di Sondrio	J48C8PCSJVUBR8KCW529	IT0000784196	No	No	Yes	No	No
IT	Credito Emiliano	8156004B244AA70DE787	IT0003121677	No	No	Yes	No	No
IT	Banca Carige	F1T87K3OQ2OV1UORLH26	IT0005108763	No	No	Yes	No	No
IT	Banca Piccolo Credito Valtellinese	549300BDV4C410CYAQ76	IT0005319444	No	No	Yes	No	No
PL	Alior Bank	259400QHDOZWMJ103294	PLALIOR00045	No	No	Yes	No	No
PL	Bank Handlowy w Warszawie	XLEZHWWOI4HFQDGL4793	PLBH00000012	No	No	Yes	No	No
PL	Bank Ochrony Środowiska	MKP1B7E76TN04CD85Z79	PLBOS0000019	No	No	Yes	No	No
PL	Bank BPH	H8MIG1MMPR6JXUQVBM04	PLBPH0000019	No	No	Yes	No	No
PL	Getin Noble Bank	2594000SEGUR418W2G08	PLGETBK00012	No	No	Yes	No	No
NL	ABN AMRO	724500DWE10NNL1AXZ52	NL0011540547	No	No	No	Yes	Yes
IT	Banco BPM	815600E4E6DCD2D25E30	IT0005218380	No	No	No	No	Yes
PL	Bank Polska Kasa Opieki	5493000LKS7B3UTF7H35	PLPEKAO00016	No	No	No	No	Yes

Note. LEI = Legal Entity Identifier. ISIN = International Securities Identification Number. CEBS = Committee of European Banking Supervisors. EBA = European Banking Authority. *n* = sample size. NA = not applicable. AT = Austria. BE = Belgium. CY = Cyprus. DE = Germany. DK = Denmark. ES = Spain. FR = France. GB = United Kingdom of Great Britain and Northern Ireland. GR = Greece. HU = Hungary. IE = Ireland. IT = Italy. MT = Malta. NL = Netherlands. NO = Norway. PL = Poland. PT = Portugal. SE = Sweden. SI = Slovenia.

Appendix E – Elements of the Longitudinal Sample

The table below lists all banks included in the longitudinal sample, *i.e.* the elements of the samples. The size of the longitudinal sample was $n = 28$ sample banks.

				EU-Wide Stress Tests				
Country	**Bank**	**LEI**	**ISIN**	**CEBS 2010 ($n = 28$)**	**EBA 2011 ($n = 28$)**	**EBA 2014 ($n = 28$)**	**EBA 2016 ($n = 28$)**	**EBA 2018 ($n = 28$)**
AT	Erste Group Bank	PQOH26KWDF7CG10L6792	AT0000652011	Yes	Yes	Yes	Yes	Yes
BE	KBC	213800X3Q9LSAKRUWY91	BE0003565737	Yes	Yes	Yes	Yes	Yes
DE	Deutsche Bank	7LTWFZYICNSX8D621K86	DE0005140008	Yes	Yes	Yes	Yes	Yes
DE	Commerzbank	851WYGNLUQLFZBSYGB56	DE000CBK1001	Yes	Yes	Yes	Yes	Yes
DK	Danske Bank	MAES062Z21O4RZ2U7M96	DK0010274414	Yes	Yes	Yes	Yes	Yes
DK	Jyske Bank	3M5E1GQGKL17HI6CPN30	DK0010307958	Yes	Yes	Yes	Yes	Yes
ES	Banco Bilbao Vizcaya Argentaria	K8MS7FD7N5Z2WQ51AZ71	ES0113211835	Yes	Yes	Yes	Yes	Yes
ES	Banco de Sabadell	SI5RG2M0WQQLZCXKRM20	ES0113860A34	Yes	Yes	Yes	Yes	Yes
ES	Banco Santander	5493006QMFDDMYWIAM13	ES0113900J37	Yes	Yes	Yes	Yes	Yes
ES	Caixabank	7CUNS533WID6K7DGFI87	ES0140609019	Yes	Yes	Yes	Yes	Yes
FR	Crédit Agricole	969500TJ5KRTCJQWXH05	FR0000045072	Yes	Yes	Yes	Yes	Yes
FR	Société Générale	O2RNE8IBXP4R0TD8PU41	FR0000130809	Yes	Yes	Yes	Yes	Yes
FR	BNP Paribas	R0MUWSFPU8MPRO8K5P83	FR0000131104	Yes	Yes	Yes	Yes	Yes
GB	HSBC	MLU0ZO3ML4LN2LL2TL39	GB0005405286	Yes	Yes	Yes	Yes	Yes
GB	Lloyds Banking Group	549300PPXHEU2JF0AM85	GB0008706128	Yes	Yes	Yes	Yes	Yes
GB	Barclays Bank	G5GSEF7VJP5I7OUK5573	GB0031348658	Yes	Yes	Yes	Yes	Yes
GB	Royal Bank of Scotland	2138005O9XJIJN4JPN90	GB00B7T77214	Yes	Yes	Yes	Yes	Yes
HU	OTP Bank	529900W3MOO00A18X956	HU0000061726	Yes	Yes	Yes	Yes	Yes
IE	Bank of Ireland	Q2GQA2KF6XJ24W42G291	IE00BD1RP616	Yes	Yes	Yes	Yes	Yes

IE	Allied Irish Banks	3U8WV1YX2VMUHH7Z1Q21	IE00BF0L3536	Yes	Yes	Yes	Yes	Yes
IT	Intesa Sanpaolo	2W8N8UU78PMDQKZENC08	IT0000072618	Yes	Yes	Yes	Yes	Yes
IT	UBI Banca	81560097964CBDAED282	IT0003487029	Yes	Yes	Yes	Yes	Yes
IT	Unicredit	549300TRUWO2CD2G5692	IT0005239360	Yes	Yes	Yes	Yes	Yes
NL	ING Bank	3TK20IVIUJ8J3ZU0QE75	NL0011821202	Yes	Yes	Yes	Yes	Yes
PL	Powszechna Kasa Oszczędności Bank Polski	P4GTT6GF1W40CVIMFR43	PLPKO0000016	Yes	Yes	Yes	Yes	Yes
SE	Skandinaviska Enskilda Banken	F3JS33DEI6XQ4ZBPTN86	SE0000148884	Yes	Yes	Yes	Yes	Yes
SE	Swedbank	M312WZV08Y7LYUC71685	SE0000242455	Yes	Yes	Yes	Yes	Yes
SE	Svenska Handelsbanken	NHBDILHZTYCNBV5UYZ31	SE0007100599	Yes	Yes	Yes	Yes	Yes

Note. LEI = Legal Entity Identifier. ISIN = International Securities Identification Number. CEBS = Committee of European Banking Supervisors. EBA = European Banking Authority. n = sample size. AT = Austria. BE = Belgium. DE = Germany. DK = Denmark. ES = Spain. FR = France. GB = United Kingdom of Great Britain and Northern Ireland. HU = Hungary. IE = Ireland. IT = Italy. NL = Netherlands. PL = Poland. SE = Sweden.

Appendix F – Regression-Based Goodness-of-Fit Tests

The table below shows the results of the regression-based goodness-of-fit tests of the $R = 6$ candidate asset pricing models. All candidate models were run on the same sample of n = 227 bank-year observations and computed separately for each event-window type. The parentheses behind the event-window types specify the number of trading days before and after the event date t_0. The term n in the definition of the post-event window $(+1, n)$ refers to a statistical determination of variable event window lengths based on the Ljung-Box (1978) test introduced in this study. To perform the goodness-of-fit tests, the actual (observed) returns were regressed on the normal (expected) returns estimated by each candidate model.

	MAR (n = 227) (k = 1)	MAM (n = 227) (k = 1)	MM (n = 227) (k = 1)	CAPM (n = 227) (k = 2)	FF3F (n = 227) (k = 4)	FF5F (n = 227) (k = 6)
Panel A: Pre-Event Window $(-\mathbf{2}, \mathbf{0})$						
R^2	0.019	0.228	0.243	0.239	0.295	0.316
R^2_{adj}	0.015	0.225	0.240	0.232	0.283	0.297
$F(1, 225)$	4.38**	66.46***	72.28***	70.64***	94.27***	103.94***
SSE	0.336	0.265	0.259	0.261	0.241	0.234
α	0.016	-0.004	0.002	-0.002	-0.001	-0.001
β	1.327	1.192	0.871	0.933	1.036	0.983
Panel B: Standard Event Window $(-\mathbf{2}, +\mathbf{2})$						
R^2	0.019	0.527	0.557	0.616	0.611	0.593
R^2_{adj}	0.015	0.525	0.555	0.613	0.604	0.582
$F(1, 225)$	4.43**	250.51***	282.71***	361.06***	352.88***	327.31***
SSE	1.371	0.662	0.620	0.537	0.544	0.570
α	0.027	-0.008	0.006	-0.005	0.003	0.004
β	1.601	2.216	1.565	1.783	1.255	1.037
Panel C: Post-Event Window $(+\mathbf{1}, \boldsymbol{n})$						
R^2	0.009	0.300	0.417	0.440	0.374	0.316
R^2_{adj}	0.004	0.297	0.414	0.435	0.363	0.298
$F(1, 225)$	1 94	96.37***	160.65***	176.61***	134.42***	104.08***
SSE	0.438	0.309	0.258	0.247	0.276	0.302
α	-0.002	0.006	0.010	0.007	0.005	0.003
β	1.749	2.155	2.043	2.189	1.232	0.965

Note. MAR = Mean-Adjusted Return Model. MAM = Market-Adjusted Model. MM = Market Model. CAPM = Capital Asset Pricing Model. FF3F = Fama and French (1993) Three-Factor Model. FF5F = Fama and French (2015) Five-Factor Model. n = sample size. k = number of model parameters estimated. R^2 = coefficient of determination. R^2_{adj} = adjusted R^2. F = F-value. *SSE* = sum of squared errors. α = constant (intercept). β = regression coefficient (slope).

* $p < .10$. ** $p < .05$. *** $p < .01$

Appendix G – Information Criteria-Based Goodness-of-Fit Tests

The tables below show the results of information criteria-based goodness-of-fit tests of the $R = 6$ candidate asset pricing models. All candidate models were run on the same sample of n = 227 bank-year observations and computed separately for each event-window type (Panels A to C). The parentheses behind the event-window types specify the number of trading days before and after the event date t_0. The term n in the definition of the post-event window $(+1, n)$ refers to a statistical determination of variable event window lengths based on the Ljung-Box (1978) test introduced in this study. To perform the goodness-of-fit tests in the ordinary least squares (OLS) framework, the information criteria were calculated using the sum of squared errors (*SSE*) of each candidate model. The specific information criteria calculated were the Akaike Information Criterion (*AIC*), the Schwarz Information Criterion (*SIC*), and the Hannan-Quinn Information Criterion (*HQIC*).

G.1 Akaike Information Criterion (*AIC*)

	MAR (n = 227) (k = 1)	**MAM** (n = 227) (k = 1)	**MM** (n = 227) (k = 1)	**CAPM** (n = 227) (k = 2)	**FF3F** (n = 227) (k = 4)	**FF5F** (n = 227) (k = 6)
Panel A: Pre-Event Window $(\mathbf{-2, 0})$						
AIC	-1,445.23	-1,529.24	-1,532.20	-1,530.23	-1,545.76	**-1,548.43**
ΔAIC	103.19	19.19	16.22	18.19	2.67	**0.00**
L	< .001	< .001	< .001	< .001	.264	**1.000**
w	< .001	< .001	< .001	< .001	.209	**.791**
Panel B: Standard Event Window $(\mathbf{-2, +2})$						
AIC	-1,138.25	-1,255.62	-1,294.37	-1,305.97	-1,344.45	**-1,345.07**
ΔAIC	206.83	89.45	50.70	39.10	0.62	**0.00**
L	< .001	< .001	< .001	< .001	.733	**1.000**
w	< .001	< .001	< .001	< .001	.423	**.577**
Panel C: Post-Event Window $(\mathbf{+1, n})$						
AIC	-1,415.53	-1,469.62	-1,495.68	-1,497.18	**-1,508.77**	-1,488.89
ΔAIC	93.23	39.15	13.09	11.59	**0.00**	19.88
L	< .001	< .001	.001	.003	**1.000**	< .001
w	< .001	< .001	.001	.003	**.995**	< .001

Note. For ease of interpretation, the best *AIC* values and metrics for each panel are in bold. MAR = Mean-Adjusted Return Model. MAM = Market-Adjusted Model. MM = Market Model. CAPM = Capital Asset Pricing Model. FF3F = Fama and French (1993) Three-Factor Model. FF5F = Fama and French (2015) Five-Factor Model. n = sample size. k = number of model parameters estimated. AIC = *AIC* value. ΔAIC = *AIC* difference. L = relative likelihood. w = Akaike weight.

G.2 Schwarz Information Criterion (SIC)

	MAR (n = 227) (k = 1)	MAM (n = 227) (k = 1)	MM (n = 227) (k = 1)	CAPM (n = 227) (k = 2)	FF3F (n = 227) (k = 4)	FF5F (n = 227) (k = 6)
Panel A: Pre-Event Window $(-2, 0)$						
SIC	-1,441.81	-1,525.81	-1,525.35	-1,523.38	**-1,532.06**	-1,527.88
ΔSIC	90.25	6.25	6.71	8.68	**0.00**	4.18
L	< .001	.044	.035	.013	**1.000**	.123
w	< .001	.036	.029	.011	**.823**	.102
Panel B: Standard Event Window $(-2, +2)$						
SIC	-1,134.82	-1,252.20	-1,287.52	-1,299.12	**-1,330.75**	-1,324.52
ΔSIC	195.93	78.56	43.23	31.63	**0.00**	6.23
L	< .001	< .001	< .001	< .001	**1.000**	.044
w	< .001	< .001	< .001	< .001	**.957**	.043
Panel C: Post-Event Window $(+1, n)$						
SIC	-1,412.11	-1,466.20	-1,488.83	-1,490.33	**-1,495.07**	-1,468.34
ΔSIC	82.96	28.87	6.24	4.74	**0.00**	26.73
L	< .001	< .001	.044	.094	**1.000**	< .001
w	< .001	< .001	.039	.082	**.879**	< .001

Note. For ease of interpretation, the best *SIC* values and metrics for each panel are in bold. MAR = Mean-Adjusted Return Model. MAM = Market-Adjusted Model. MM = Market Model. CAPM = Capital Asset Pricing Model. FF3F = Fama and French (1993) Three-Factor Model. FF5F = Fama and French (2015) Five-Factor Model. n = sample size. k = number of model parameters estimated. SIC = *SIC* value. ΔSIC = *SIC* difference. L = relative likelihood. w = Akaike weight.

G.3 Hannan-Quinn Information Criterion (*HQIC*)

	MAR (n = 227) (k = 1)	MAM (n = 227) (k = 1)	MM (n = 227) (k = 1)	CAPM (n = 227) (k = 2)	FF3F (n = 227) (k = 4)	FF5F (n = 227) (k = 6)
Panel A: Pre-Event Window $(-2, 0)$						
$HQIC$	-1,443.85	-1,527.85	-1,529.44	-1,527.47	**-1,540.23**	-1,540.13
$\Delta HQIC$	96.38	12.38	10.79	12.76	**0.00**	0.10
L	< .001	.002	.005	.002	**1.000**	.952
w	< .001	.001	.002	.001	**.510**	.486
Panel B: Standard Event Window $(-2, +2)$						
$HQIC$	-1,136.86	-1,254.24	-1,291.61	-1,303.21	**-1,338.92**	-1,336.78
$\Delta HQIC$	202.06	84.69	47.32	35.72	**0.00**	2.14
L	< .001	< .001	< .001	< .001	**1.000**	.342
w	< .001	< .001	< .001	< .001	**.745**	.255
Panel C: Post-Event Window $(+1, n)$						
$HQIC$	-1,414.15	-1,468.24	-1,492.92	-1,494.42	**-1,503.24**	-1,480.60
$\Delta HQIC$	89.09	35.00	10.32	8.82	**0.00**	22.64
L	< .001	< .001	.006	.012	**1.000**	< .001
w	< .001	< .001	.006	.012	**.982**	< .001

Note. For ease of interpretation, the best *HQIC* values and metrics for each panel are in bold. MAR = Mean-Adjusted Return Model. MAM = Market-Adjusted Model. MM = Market Model. CAPM = Capital Asset Pricing Model. FF3F = Fama and French (1993) Three-Factor Model. FF5F = Fama and French (2015) Five-Factor Model. n = sample size. k = number of model parameters estimated. $HQIC$ = *HQIC* value. $\Delta HQIC$ = *HQIC* difference. L = relative likelihood. w = Akaike weight.

Appendix H – Normality Tests

The tables below show the results of two formal normality tests (Lilliefors test and Shapiro-Wilk test) performed to test whether the cumulative abnormal returns ($CARs$) and the absolute cumulative abnormal returns ($|CARs|$) of the cross-sectional samples and of the longitudinal sample were normally distributed. The tests were performed for each event-window type (Panels A to C). The parentheses behind the event-window types specify the number of trading days before and after the event date t_0. The term n in the definition of the post-event window $(+1, n)$ refers to a statistical determination of variable event window lengths based on the Ljung-Box (1978) test introduced in this study.

H.1 Cumulative Abnormal Returns ($CARs$) Based on the Cross-Sectional Samples

		Lilliefors Test[a]		Shapiro-Wilk Test[a]	
Panel A: Pre-Event Window $(\mathbf{-2, 0})$	***df***	***D***	***p***	***W***	***p***
CEBS 2010	50	.113	.150	.932***	.007
EBA 2011	51	.116*	.084	.812***	< .001
EBA 2014	59	202***	< .001	.738***	< .001
EBA 2016	34	.139*	.095	.929**	.029
EBA 2018	33	.113	200	959	.236
Panel B: Standard Event Window $(\mathbf{-2, +2})$					
CEBS 2010	50	.158***	.003	.924***	.003
EBA 2011	51	.198***	< .001	.745***	< .001
EBA 2014	59	135***	.009	.907***	< .001
EBA 2016	34	176***	.009	.850***	< .001
EBA 2018	33	.071	200	974	.583
Panel C: Post-Event Window $(\mathbf{+1, n})$					
CEBS 2010	50	.224***	< .001	.731***	< .001
EBA 2011	51	137**	.018	.857***	< .001
EBA 2014	59	223***	< .001	.767***	< .001
EBA 2016	34	.137	.105	.878***	.001
EBA 2018	33	.204***	.001	.873***	.001

Note. The normal returns used to calculate the cumulative abnormal returns ($CARs$) were estimated using the Fama and French (1993) Three-Factor Model. *df* = degrees of freedom. D = Lilliefors test statistic. W = Shapiro-Wilk test statistic. p = p value. *n* = sample size. CEBS = Committee of European Banking Supervisors. EBA = European Banking Authority.

[a] The sample size varied depending on the sample: for CEBS 2010 *n* = 50, EBA 2011 *n* = 51, EBA 2014 *n* = 59, EBA 2016 *n* = 34, and EBA 2018 *n* = 33.

* $p < .10$. ** $p < .05$. *** $p < .01$.

H.2 Absolute Cumulative Abnormal Returns ($|CARs|$) Based on the Cross-Sectional Samples

		Lilliefors Test[a]		Shapiro-Wilk Test[a]	
Panel A: Pre-Event Window $(-2, 0)$	***df***	D	p	W	p
CEBS 2010	50	208***	< .001	.798***	< .001
EBA 2011	51	189***	< .001	.623***	< .001
EBA 2014	59	255***	< .001	.633***	< .001
EBA 2016	34	218***	< .001	.791***	< .001
EBA 2018	33	.129	.178	.923**	.023
Panel B: Standard Event Window $(-2, +2)$					
CEBS 2010	50	217***	< .001	.781***	< .001
EBA 2011	51	238***	< .001	.586***	< .001
EBA 2014	59	214***	< .001	.771***	< .001
EBA 2016	34	264***	< .001	.692***	< .001
EBA 2018	33	.156**	.039	.906***	.008
Panel C: Post-Event Window $(+1, n)$					
CEBS 2010	50	262***	< .001	.620***	< .001
EBA 2011	51	181***	< .001	.696***	< .001
EBA 2014	59	252***	< .001	.598***	< .001
EBA 2016	34	224***	< .001	.719***	< .001
EBA 2018	33	226***	< .001	.723***	< .001

Note. The normal returns used to calculate the absolute cumulative abnormal returns ($|CARs|$) were estimated using the Fama and French (1993) Three-Factor Model. *df* = degrees of freedom. D = Lilliefors test statistic. W = Shapiro-Wilk test statistic. p = p value. *n* = sample size. CEBS = Committee of European Banking Supervisors. EBA = European Banking Authority.

[a] The sample size varied depending on the sample: for CEBS 2010 *n* = 50, EBA 2011 *n* = 51, EBA 2014 *n* = 59, EBA 2016 *n* = 34, and EBA 2018 *n* = 33.

* $p < .10$. ** $p < .05$. *** $p < .01$.

H.3 Cumulative Abnormal Returns ($CARs$) Based on the Longitudinal Sample

		Lilliefors Test[a]		Shapiro-Wilk Test[a]	
Panel A: Pre-Event Window $(-2, 0)$	***df***	***D***	***p***	***W***	***p***
CEBS 2010	28	.079	.200	967	.492
EBA 2011	28	.097	200	993	.999
EBA 2014	28	.106	.200	969	.560
EBA 2016	28	.148	.119	.930*	.061
EBA 2018	28	.114	200	958	.310
Panel B: Standard Event Window $(-2, +2)$					
CEBS 2010	28	.095	200	978	.793
EBA 2011	28	.142	154	954	.248
EBA 2014	28	.189**	.011	.917**	.030
EBA 2016	28	220***	.001	.832***	< .001
EBA 2018	28	.085	200	969	.559
Panel C: Post-Event Window $(+1, n)$					
CEBS 2010	28	.154*	.085	.892***	.007
EBA 2011	28	.139	174	.934*	.078
EBA 2014	28	.184**	.016	.894***	.008
EBA 2016	28	.176**	.026	.822***	< .001
EBA 2018	28	.224***	.001	.841***	.001

Note. The normal returns used to calculate the cumulative abnormal returns ($CARs$) were estimated using the Fama and French (1993) Three-Factor Model. df = degrees of freedom. D = Lilliefors test statistic. W = Shapiro-Wilk test statistic. p = p value. n = sample size. CEBS = Committee of European Banking Supervisors. EBA = European Banking Authority.

[a] n = 28.

* $p < .10$. ** $p < .05$. *** $p < .01$.

H.4 Absolute Cumulative Abnormal Returns ($|CARs|$) Based on the Longitudinal Sample

		Lilliefors Test[a]		Shapiro-Wilk Test[a]	
Panel A: Pre-Event Window $(-2, 0)$	***df***	D	p	W	p
CEBS 2010	28	.085	.200	945	.146
EBA 2011	28	180**	.020	.912**	.022
EBA 2014	28	213***	.002	.868***	.002
EBA 2016	28	238***	< .001	.776***	< .001
EBA 2018	28	.153*	.091	.924**	.045
Panel B: Standard Event Window $(-2, +2)$					
CEBS 2010	28	.159*	.068	.902**	.013
EBA 2011	28	.207***	.003	.828***	< .001
EBA 2014	28	270***	< .001	.715***	< .001
EBA 2016	28	329***	< .001	.663***	< .001
EBA 2018	28	.161*	.062	.885***	.005
Panel C: Post-Event Window $(+1, n)$					
CEBS 2010	28	184**	.017	.761***	< .001
EBA 2011	28	.107	200	939	.104
EBA 2014	28	268***	< .001	.710***	< .001
EBA 2016	28	245***	< .001	.677***	< .001
EBA 2018	28	.252***	< .001	.703***	< .001

Note. The normal returns used to calculate the absolute cumulative abnormal returns ($|CARs|$) were estimated using the Fama and French (1993) Three-Factor Model. *df* = degrees of freedom. D = Lilliefors test statistic. W = Shapiro-Wilk test statistic. p = p value. n = sample size. CEBS = Committee of European Banking Supervisors. EBA = European Banking Authority.

[a] n = 28.

* $p < .10$. ** $p < .05$. *** $p < .01$.