NBER WORKING PAPER SERIES

SYSTEMIC RISK MEASURES: FROM THE PANIC OF 1907 TO THE BANKING STRESS OF 2023

Viral V. Acharya
Markus K. Brunnermeier
Diane Pierret

Working Paper 33211
http://www.nber.org/papers/w33211

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2024, Revised May 2025

The authors would like to thank Xinlin Yuan for excellent research assistance, and the authors of the paper “Failing Banks” —Sergio Correia, Stephan Luck, and Emil Verner— for sharing the data they collected on historical bank balance sheets. The authors also thank Richard Berner, Robert Engle, Peter Koudijs, Dmitry Kuvshinov, Matthew Richardson, Tony Saunders, Dick Sylla, Bruce Tuckman, and members of the Volatility and Risk Institute at NYU Stern School of Business, and an anonymous referee for their comments. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed additional relationships of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w33211

Systemic Risk Measures: From the Panic of 1907 to the Banking Stress of 2023
Viral V. Acharya, Markus K. Brunnermeier, and Diane Pierret
NBER Working Paper No. 33211
November 2024, Revised May 2025
JEL No. G01, G20, G21, G23, G28

ABSTRACT

We assess the efficacy of market-based systemic risk measures that rely on U.S. financial firms' stock return co-movements with market- or sector-wide returns under stress from 1895 to 2023. Stress episodes are identified using corporate bond spread widening and narrative dating, spanning from the Panic of 1907 to the Banking Stress of 2023. Measures observed prior to the onset of stress episodes predict market outcomes (realized volatility and returns), balance sheet outcomes (lending, profitability, and run risk), and bank failures. Specifically, the measures are: (i) particularly effective in capturing the cross-sectional ranking of institutions conditional on a stress episode, rather than aggregate outcomes;(ii) more informative when stress episodes are severe; and (iii) relevant for both banks and non-bank financial institutions, although measures incorporating market leverage are especially informative for banks. A comparative analysis shows that market-based indicators offer information that is distinct from, and complementary to, traditional balance sheet metrics used in supervisory and macroprudential risk assessment.

Viral V. Acharya
New York University
Leonard N. Stern School of Business
and CEPR
and also NBER
vacharya@stern.nyu.edu

Markus K. Brunnermeier
Princeton University
Department of Economics
and CEPR
and also NBER
markus@princeton.edu

Diane Pierret
University of Luxembourg
diane.pierret@uni.lu

1 Introduction

More than 15 years after the 2007/8 global financial crisis (GFC), when banks and non-bank financial intermediaries failed *en masse*, it is time to evaluate selected key systemic risk measures that emerged in its wake and have been widely used since. Financial regulation underwent a fundamental shift since the GFC. Instead of focusing solely on *microprudential* regulation, which emphasizes the stress of individual institutions in isolation, *macroprudential* regulation, which addresses systemic risk, i.e., the risk that the entire financial system is under stress, became paramount.[1] General equilibrium channels, such as fire sales and liquidity spirals and their real-sector consequences such as credit crunch and lack of market intermediation, gained primary importance. Researchers, therefore, looked for empirical measures of financial vulnerability that had explanatory power, especially in the cross-section of financial firms, for where such channels might be the most powerful at work.

Some of the systemic risk measures that emerged capture how much the risk of a specific financial institution spills over to the rest of the financial sector, while others measure how much an individual institution is exposed to a system-wide financial crisis. Some others combine such information with that on non-equity liabilities to estimate under-capitalization of financial firms under stress. Ideally, these measures should also determine whether a financial crisis is a temporary phenomenon, indicating economic resilience, or a more prolonged, permanent, issue (Brunnermeier, 2024). In practice, resilience is shaped ex post by the choice of specific public policies which systemic risk measures may reflect ex ante (at least to some extent). In a nutshell, systemic risk measures aim to capture which financial firms are affected by stress-time amplifiers, are potentially at risk of failure, and are likely to withdraw provision of intermediation with adverse consequences to the real economy.

This paper takes a set of widely used market-based indicators of systemic risk and evaluates their empirical performance across more than a century of U.S. financial history. Specifically, it examines the contribution and exposure versions of CoVaR (Adrian & Brunnermeier, 2016), the Marginal Expected Shortfall or the MES (Acharya et al., 2017), and SRISK (Acharya, Engle & Richardson, 2012). All four measures are depicted in Figure 1. The paper identifies stress episodes shown as grey bars in Figure 1, using a two-step procedure. First, episodes are identified based on a narrative analysis and previous studies that date financial crises or stress periods. Second, the start and end points of each episode are specified based on the elevation of the Gilchrist & Zakrajšek (2012) credit spread, for the period from 1959 to 2023, and for the earlier period from 1895 to 1958, alternative credit spread measures are used due to sparser data availability.

The aim of this paper is not to provide a comprehensive survey of systemic risk measures or a literature review in the traditional sense.[2] Nor is the goal to develop or test early warning models of systemic risk in the aggregate or time-series sense (see, e.g., Greenwood et al., 2022). Rather, the analysis is conditional on the occurrence of a financial stress episode and focuses on the ability of market-based systemic risk measures to predict the cross-sectional distribution of institutional fragility during such episodes.[3] Although not designed to forecast crises, this approach

[1] Note that Acharya (2009), Crockett (2000), and Borio (2003) were important precursors of this shift.

[2] See Bisias et al. (2012) and Benoit et al. (2017) for comprehensive surveys of systemic risk analytics.

[3] In that sense, our work is closest in spirit to Brownlees et al. (2020) who evaluate the performance of CoVaR and SRISK in identifying vulnerable institutions during U.S. banking panics between 1866 and 1933. Like us, they find that these measures are more effective at explaining the cross-section of institutional distress than at signaling aggregate conditions. We extend this line of research by examining a broader set of outcomes and systemic risk indicators over a longer historical period using a unified predictive framework.

remains policy-relevant: once stress emerges, understanding which institutions are most vulnerable is essential for guiding supervisory attention and determining an adequate policy response.

Unlike most of the existing literature, which focuses on the post-2008 period or the GFC alone, this paper extends the time horizon substantially further back in history. To do so, we construct a novel dataset by combining CRSP data with equity return data from Global Financial Data (GFD) for U.S. financial institutions. The GFD series allows us to extend coverage back to 1895 and includes 5,360 financial institutions—especially banks traded over the counter—that are not present in CRSP. This results in, to our knowledge, the most comprehensive historical dataset used to compute systemic risk measures across such a wide range of stress episodes, from the Panic of 1907 to the Banking Stress of 2023.

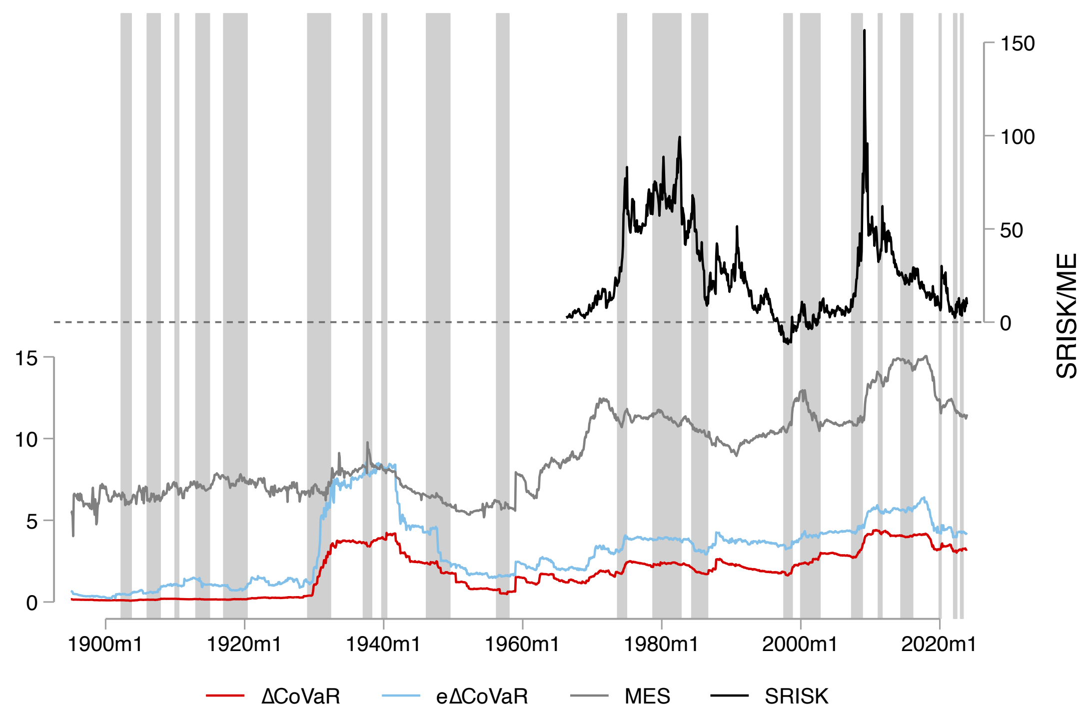

Figure 1: Market-capitalization weighted average ΔCoVaR, eΔCoVaR, MES and SRISK divided by the market capitalization (ME) from January 1895 until December 2023. The grey vertical bars represent the credit spread peak-to-trough months.

Our empirical analysis shows that market-based systemic risk measures exhibit substantial explanatory power for the cross-section of institutional outcomes during stress episodes. Institutions with higher ex-ante risk measures are more likely to experience greater volatility and lower returns, and these measures also help predict bank failures and deteriorations in balance sheet fundamentals, including loan growth, profitability (return on assets), and run risk (proxied by the growth in the ratio of uninsured deposits to total deposits). These measures tend to be more informative in three cases: First, they are especially effective in capturing the cross-sectional ranking of outcomes conditional on a stress episode, rather than the aggregate time series. Second, they are more informative when stress episodes are severe. Third, they are relevant for both banks and non-bank financial institutions, though measures incorporating market leverage are especially informative for banks. In addition, we conduct a comprehensive comparative exercise based on nearly 500 predictive regressions, evaluating market-based indicators both on their own and in combination with balance sheet metrics drawn from Correia, Luck & Verner (2024). While balance sheet indicators often perform well—especially over longer horizons—market-based measures consistently add predictive value, and the combination of the two delivers the strongest results across outcomes.

In summary, market-based systemic risk measures pass the efficacy test across a wide range of U.S. financial crises and over a long period of time, helping to explain both market-based and real-sector stress. These findings highlight, therefore, that market-based systemic risk measures should be a crucial part of the toolbox for macroprudential and supervisory assessments of the financial sector. Unlike traditional regulatory accounting measures, our systemic risk measures do not suffer from the backward-looking bias, since these measures are based on stock market prices. For example, prior to the GFC, accounting and regulatory ratios (such as the Basel risk-weight based capitalization ratios) created the misleading illusion that banks had sufficiently large equity cushions. While regulatory stress tests adopted since the GFC are also useful remedies, they continue to rely on regulatory risk weights (Acharya, Engle & Pierret, 2014), depend on specific stress scenarios that do not necessarily evolve with the evolving nature of financial risks, and are less robust to manipulation ("regulatory arbitrage") by the financial sector. At a minimum, market-based systemic risk measures analyzed in this article offer substantial promise as valuable complementary tools to regulatory stress tests.

The structure of this article is as follows: Section 2 describes the data sources, the stress episodes, and the systemic risk measures. Section 3 documents the predictive power of systemic risk measures in predicting subsequent stress volatility and returns, bank failures, and real variables. Section 4 compares the predictive performance of market-based systemic risk measures and balance sheet indicators from the literature. Section 5 concludes with some directions for further research.

2 Data and Methodology

2.1 Data and Sample

We collect equity price data for U.S. financial institutions from two main sources. From the Center for Research in Security Prices (CRSP) database, we obtain daily stock prices for 5,417 financial institutions from December 1925 to December 2023.[4] The CRSP dataset is merged with Compustat to retrieve total bank liabilities that are available starting in 1965. CRSP coverage of banks is particularly limited in the early decades of the 20th century, as many bank stocks were traded over the counter and are therefore missing from CRSP.

To extend equity return coverage prior to 1959, we supplement CRSP with historical stock return data from Global Financial Data (GFD), which includes 5,360 financial institutions from 1791 to 1958—of which 3,221 are banks, 1,183 are trust and thrift companies, and 806 are insurance companies. Listings span major U.S. exchanges, but the majority of securities (4,014) are traded over the counter. To estimate systemic risk measures, we use the GFD sample from 1895, when coverage becomes sufficiently stable to construct a reliable financial sector index.

The financial sector index is a market-capitalization weighted average of stock prices of the most liquid stocks in GFD, identified as those with at least 90% non-zero return observations. From 1959 onward, we rely on CRSP because the CRSP-FRB linking table provides a reliable and standardized way to match stock return data with regulatory bank balance sheets from Call

[4]"Financial institutions" are defined based on the same SIC codes used in Adrian & Brunnermeier (2016).

Reports. For this period, the financial index is constructed as a market-cap-weighted average of all financial institutions in the CRSP sample.

The balance sheet and income statement variables are from the Federal Financial Institutions Examination Council (FFIEC) Consolidated Reports of Condition and Income ("Call Reports"), available for the sample of commercial banks and collected from 1959 onwards by Correia, Luck & Verner (2024). We use the CRSP-FRB linking table of the Federal Reserve of New York to match the CRSP database to Call Reports from 1986 to 2023. The rest of the sample (1959-1986) is matched manually based on the name and location of the bank. Call Reports are aggregated at the parent bank level using the FFIEC relationship table, where the parent is the bank with publicly listed stocks in the CRSP database.

We also use the list of bank failures and assistance transactions from the Federal Deposit Insurance Corporation (FDIC) from 1934 to the present (similarly to Correia, Luck & Verner, 2024). Finally, the macroeconomic variables and financial indices we use in this paper are detailed in the next section.

2.2 Definition of Stress Episodes

Our methodology for dating stress episodes uses several sources, including data on corporate bond spreads, previous banking studies, and a narrative analysis of stress events. We first identify broad event "windows" from multiple sources. For the "early" sample spanning 1895 to 1958, the years of the event window are selected from stock market crashes and banking crisis years identified for the U.S. in Reinhart & Rogoff (2009, 2011) ("RR years" hereafter). For the "modern" sample spanning 1959 to 2023, we rely on a narrative analysis of the most recent stress episodes to identify window years in addition to RR years. The narrative of stress episodes, which draws on several sources including Markham (2022), as well as a detailed description of our methodology, are provided in the Appendix. Although our methodology differs from Reinhart & Rogoff (2009),[5] the selection of event windows based on the narrative complements their database, which ends in 2010.

Second, we search for the trough and peak values of a credit spread for each window to determine the start and end months of the episode, respectively. As the credit spread, we use the Gilchrist & Zakrajšek (2012) corporate bond spread index, commonly referred to as the GZ spread, which is available monthly starting in 1973. The GZ spread is constructed as an unweighted average of credit spreads from senior unsecured bonds issued by U.S. non-financial firms. It reflects the average credit risk premium demanded by investors for holding corporate bonds over risk-free securities. Between 1922 and 1973, we use the difference between Moody's Seasoned Baa and Aaa Corporate Bond Yield indices. Moody's corporate bond yields are available monthly starting in January 1919 from the Federal Reserve Bank of St. Louis database FRED. Finally for the 1895-1921 period, we rely on the average yield on U.S. railroad bonds having at least ten years maturity from Macaulay (1938). This index is selected among the indices available from the NBER Macrohistory database because it has the highest correlation (0.77) with the Moody's Baa–Aaa spread over the period when the two series overlap (1919–1934). The railroad bond yield is available at a monthly frequency from January 1857 through December 1934.

[INSERT TABLE 1 HERE]

All stress episodes identified with our methodology are listed in Table 1, together with the start date, end date, length in months, the corresponding percentage point change in the credit spread,

[5] The identification of stock market crashes in Reinhart & Rogoff (2009) is based on real equity prices following the methodology of Barro & Ursúa (2017), while our methodology relies on nominal corporate bond spreads.

and the returns on a stock index and a financial index over the episode. The stock index is the S&P500 starting in 1959 and is the Dow Jones Industrial Average (DJI) index before, since the S&P500 index is not available for the full sample. The financial index is the GFD financial index for the early period, and the CRSP financial index starting in 1959. We additionally report the maximum drawdown ("dd") on the indices. The maximum drawdown is the percentage difference between the minimum index value and its prior maximum value within an episode.[6]

We note that the Great Depression was the most severe among early episodes, with a credit spread increase of almost 5 p.p., and negative corrections to the DJI and CRSP financial stock indices of 86 and 92 percent, respectively, over 39 months from February 1929 to May 1932. Other than the Great Depression, in the early sample, the Panic of 1907 stands out as the next most severe RR banking crisis with an increase in the railroad bond yield index of 0.7 p.p., and negative corrections to the DJI and CRSP financial stock indices of 43 and 30 percent, respectively, over 22 months. Among modern stress episodes, the GFC was the most severe with a 6.4 percentage points (p.p.) increase in the GZ spread and negative corrections to the S&P500 and CRSP financial stock indices of 41 and 52 percent, respectively, over 18 months from May 2007 to November 2008. The largest increase in the GZ spread during the GFC is followed by the Covid19 Pandemic (2.37 p.p. increase over 3 months), and the Dot.com Bubble (2.36 p.p. increase over 33 months).

We evaluate the robustness of our results to an alternative episode dating methodology, namely the cycle dating algorithm of Bry & Boschan (1971) that is independent from the narrative analysis and RR years. Comparing our approach to the Bry-Boschan (BB) algorithm in the Appendix, we find that the majority of episodes are matched across both methods. When differences arise, the BB algorithm typically identifies the same peak but assigns an earlier start to the episode. Another discrepancy is the LTCM crisis and the Dot.com bubble, which are treated as two distinct events in our chronology but are grouped into a single extended episode by BB. Additionally, the BB algorithm identifies a few extra episodes that exhibit only mild fluctuations in credit spreads and negligible market drawdowns. These differences reflect the fact that the BB algorithm is structured around parameters governing the length of cycles and phases, but it does not include any criterion related to the severity of the episode.[7]

Overall, our methodology yields more episodes than those identified in the literature on crises (Baron, Verner & Xiong, 2021; Bordo et al., 2001; Jamilov et al., 2024; Jordà, Schularick & Taylor, 2015; Laeven & Valencia, 2013, 2020; Reinhart & Rogoff, 2011). For example, the Jordà-Schularick-Taylor Macrohistory Database lists 1907, 1930, 1984, and 2007 as the first years of a systemic banking crisis.[8] In contrast, the episodes in Table 1 also include periods that could be described as episodes of stressed conditions in credit and stock markets.

[6]All reported estimates are from nominal spreads and indices, contrasting with the RR approach based on real equity prices. For example, during the 1977–82 stock market crash episode, the S&P500 index and the CRSP financial index showed positive nominal returns of 44 p.p. and 43 p.p., respectively, over the episode, since the indices are not inflation-adjusted. However, the episode also features maximum drawdowns in nominal terms of -24 p.p. and -34 p.p. on the S&P500 index and the CRSP financial index, respectively.

[7]We further test the robustness of the results of Section 3.1 to the BB dating methodology, and report the results in the Appendix.

[8]In addition, while Reinhart & Rogoff (2009) define the S&L crisis as spanning 1984 to 1991, Baron, Verner & Xiong (2021) identify two distinct episodes beginning in 1984 and 1990, respectively, and classify the years 1990–1992 as a banking crisis without panic. The years 1991 and 1992 are also flagged as episodes of bank runs in the methodology of Jamilov et al. (2024). Our alternative episode dating approach, based on the Bry-Boschan algorithm, similarly detects an episode in the early 1990s. However, this period is characterized by positive stock market performance and no drawdown in either the S&P 500 or financial sector indices.

2.3 Systemic Risk Measures

While a range of systemic risk measures have been proposed in the literature,[9] we focus on two sets that rely on the co-movement of a financial firm's market return with that on aggregate market indices under stress. Empirically, it is a tall order to tease out whether such co-movement is due to an underlying correlation of risks or an outcome of contagion (due to interconnectedness, information linkages, or general equilibrium effects) following a manifestation of risk at one or more financial firms. Systemic risk measures often bypass this issue of identifying the mechanisms at work. They are designed instead to capture in a reduced-form co-movement sense which financial firms are affected by stress-time amplifiers, at risk of failure, and likely to withdraw provision of intermediation to the real economy.

In particular, we focus on △CoVaR and the exposure △CoVaR (e△CoVaR) of Adrian & Brunnermeier (2016), and MES and SRISK of Acharya, Engle & Richardson (2012), Acharya et al. (2017), and Brownlees & Engle (2017). These measures share three common features: (i) they are primarily based on stock price data, (ii) they are derived from bivariate models linking firm-level and market returns, with a focus on tail dependence, and (iii) they are straightforward to implement and replicate using standard CRSP and Compustat data.

2.3.1 Definitions

△CoVaR (Contribution CoVaR). △CoVaR captures the contribution of an institution to system-wide stress via spillover effects: it reflects how much worse the system performs when a specific firm experiences a loss. Formally, it is the change in the value-at-risk (VaR) at 95% of the financial system portfolio conditional on a firm being under "distress" ($X^i = VaR^i$) relative to its median "state" ($X^i = med^i$):

$$\Delta CoVaR^{system|i} = CoVaR^{system|X^i=VaR^i} - CoVaR^{system|X^i=med^i}, \tag{1}$$

where $\Pr(X^{system}|X^i = VaR^i \leq CoVaR^{system|X^i=VaR^i}) = 0.95$, X^i is a "return loss" for firm i, $X^{system} = -r^{system}$ is the net return loss for a financial index.

e△CoVaR (Exposure CoVaR). e△CoVaR is a measure of the firm's exposure to system-wide stress and captures how vulnerable it is to aggregate shocks. The exposure CoVaR is formally defined as the change in the VaR at 95% of the firm conditional on the financial index being under distress relative to its median state:

$$e\Delta CoVaR^{i|system} = eCoVaR^{i|X^{syst}=VaR^{syst}} - eCoVaR^{i|X^{syst}=med^{syst}}. \tag{2}$$

Both CoVaR variants are derived from quantile regressions and reflect co-movement in the tails of the return distribution. The estimation of $\Delta CoVaR^i_t$ for firm i at time t requires the estimation of $\beta^i_{.95,t}$ from quantile regressions, and the estimation of the median and 95% quantile of the firm's return loss distribution: $\Delta CoVaR^i_t = \hat{\beta}^i_{.95,t}(VaR^i_t - med^i_t)$. Similarly, $e\Delta CoVaR^i_t$ requires the estimation of $\beta^{i,e}_{.95,t}$ from quantile regressions, and the estimation of the median and 95% quantile of the system return loss distribution: $e\Delta CoVaR^i_t = \hat{\beta}^{i,e}_{.95,t}(VaR^{syst}_t - med^{syst}_t)$.

[9] For example, Billio et al., 2012; Huang, Zhou & Zhu, 2009; Huang, Zhou & Haibin, 2012 propose alternative risk measures based on market data. See Bisias et al. (2012) and Benoit et al. (2017) for comprehensive surveys of systemic risk analytics.

MES (Marginal Expected Shortfall). The Marginal Expected Shortfall (MES) is the expected weekly return loss of the firm conditional on a $-c$ loss on the market index during a week:

$$MES_{it} = -\mathrm{E}_t(R_{it+1}|R_{mt+1} < c). \tag{3}$$

This is another exposure measure: it reflects how much a firm tends to lose during systemic events, based on co-movement with market returns (R_{mt}) in the left tail. The MES can be approximated by $MES_{it}^{stat} = -\beta_i \mathrm{E}\left(r_{mt+1}|r_{mt+1} < c\right)$, where $r_{mt} = \log(1 + R_{mt})$ is the market logarithmic return, and β_i is the market beta of firm i. In addition, $\mathrm{E}\left(r_{mt+1}|r_{mt+1} < c\right)$ is a function of market volatility σ_m, as described in the Appendix. Using this approximation, we can derive the long-run version of MES, the Long-Run Marginal Expected Shortfall (LRMES), as a simple function of the MES: $LRMES_{it} = \sqrt{h}MES_{it}$. The LRMES is the six-month return loss of a firm conditional on a 40% loss ($h = 24, C = -0.4$) on the market index, and $c = \log(1 + C)/\sqrt{h}$ (Brownlees & Engle, 2017).

SRISK and Market Leverage. SRISK is the expected capital shortfall (in U.S. dollars) of the firm in the aggregate stress scenario of a 40% loss on the market index over six months:

$$SRISK_{it} = k * (ME_{it}(1 - LRMES_{it}) + D_{it}) - ME_{it}(1 - LRMES_{it}), \tag{4}$$

where ME_{it} is the market value of equity of institution i, D_{it} are its total non-equity liabilities, and $k \leq ME_{it}/A_{it}$. We also use a measure of market leverage $Lvg_{it} = A_{it}/ME_{it}$, where quasi-market assets $A_{it} = ME_{it} + D_{it}$. Because SRISK incorporates both expected return loss (via LRMES, derived from MES) and leverage, it reflects a firm's contribution to system-wide risk through its inability to absorb losses. It is particularly useful for identifying institutions whose failure could impose large externalities on the financial system.

2.3.2 Estimation

Measures are derived at the end of each month based on a rolling window of ten years of data available up to that month. We derive △CoVaR, e△CoVaR, LRMES, and MES from CRSP weekly returns from January 1959 until December 2023, and from GFD monthly returns from January 1895 until December 1958.[10] SRISK is derived for a shorter time period starting in the 1960s due to the limited availability of liabilities from Compustat. We additionally require at least three years of available returns for the estimation of systemic risk measures.

The system return for △CoVaR and e△CoVaR is the return on the CRSP financial index. The market index for MES is the S&P500 index for the modern sample, and the GFD financial index for the early sample, due to the unavailability of the S&P500 index for early dates. We employ $k = 0.08$ (8%) consistent with the choices made at NYU Stern VLAB (vlab.stern.nyu.edu/srisk).

2.3.3 Systemic risk from 1895 to 2023

We show the monthly time series of market-capitalization weighted averages of △CoVaR, e△CoVaR, MES, and the ratio of SRISK divided by the firm's market capitalization (ME) in Figure 1 for the entire sample period. △CoVaR and e△CoVaR increased substantially during the Great Depression, while △CoVaR was the lowest before the Great Depression due to low quantile beta

[10] △CoVaR, e△CoVaR, LRMES, and MES are scaled by $\sqrt{4}$ (assuming an average of four weeks per month) in the GFD sample to make them comparable to the measures derived from weekly returns in the CRSP sample.

estimates ($\hat{\beta}^i_{.95,t}$) and low firm VaR.[11] All measures also increase during the GFC, and MES and e$\triangle$CoVaR continue to rise between the European Sovereign Debt Crisis and the Covid19 Pandemic.[12] SRISK/ME is only available starting in the 1960s, and reached a maximum value of 156% of the financial sector market capitalization due to extreme market leverage levels during the GFC.[13] SRISK can take negative values when a financial institution has a capital surplus, as was the case in the late 1990s and early 2000s as a consequence of low market leverage. In the Appendix, we decompose SRISK/ME as a linear function of MES and market leverage. We therefore also report the results for MES and market leverage separately throughout the paper. A summary table of the measures is also available in the Appendix.

3 Predictive Regressions

In this section, we assess the ability of systemic risk measures to predict the cross-section of (i) financial firm outcomes, viz., realized volatility and market returns of banks and non-banks (Section 3.1); (ii) bank failures (Section 3.2); and, (iii) bank balance-sheet outcomes (Section 3.3). We examine what market-based systemic risk measures can reveal across a long history of financial stress episodes—when they are most informative, for which types of institutions, and for which outcomes. Understanding which institutions are most exposed during periods of financial stress can help guide supervisory attention and inform the design of more effective policy responses.

3.1 Market Outcomes

3.1.1 Banks vs. non-banks

We estimate the following specification to predict market outcomes of banks and non-banks during a stress episode:

$$\begin{aligned} y_{ie} &= \beta_1 Measure_{ie} \times bank_i + \beta_2 Measure_{ie} \times (1 - bank_i) \\ &\quad + (\delta_1 bank_i + \delta_2 (1 - bank_i)) \times control_{ie} + \alpha\, bank_i + \alpha_e + \epsilon_{ie} \end{aligned} \tag{5}$$

where y_{ie} is the market outcome of firm i during stress episode e (21 episodes), $bank_i$ is an indicator variable taking the value of one for banks and zero for non-banks, and α_e are episode fixed effects.[14] The systemic risk measure, $Measure_{ie}$ and the variable $control_{ie}$ controlling for firm size are all

[11] This pattern of a low $\triangle$CoVaR before the Great Depression is confirmed in Brownlees et al. (2020) who estimate $\triangle$CoVaR for the New York banking system before 1933.

[12] We also note an increase in all measures in 1959, which is due to the different samples from GFD prior to 1959 and CRSP afterwards. The transition to CRSP introduces higher-frequency, exchange-traded stocks with more accurate pricing, which leads to higher estimated comovements and volatilities. In contrast, the GFD sample includes many illiquid OTC stocks with stale prices, which tend to understate systemic risk. We report robustness tests restricting the GFD sample to only the most liquid stocks in the Appendix; the results of the predictive regressions remain qualitatively unchanged.

[13] During the GFC, some large financial institutions had capital shortfalls according to SRISK that amounted to several times their market capitalizations, explaining the SRISK/ME peak above 100% of the sector market capitalization during that episode. Starting in the mid-1970s and into the 1980s, large values of the SRISK/ME ratio came from high market leverage as many institutions' market capitalization fell while their total liabilities continued to increase. This period was challenging for financial institutions due to unfavorable economic conditions, and in particular the ongoing effects of the 1973 oil crisis, coupled with stagflation, and high interest rates.

[14] Banks in the CRSP sample correspond to institutions for which we could identify a unique identifier (RSSD ID) assigned by the Federal Reserve in its National Information Center (NIC) database. We use GFD classification to define banks in the GFD sample.

measured *the quarter before* the episode starts. The size of a financial institution (measured by its market capitalization) is a clear confounding variable of systemic risk measures in this regression, given its correlation with both the systemic risk measures and the market outcomes.

[INSERT TABLE 2 HERE]

Table 2 presents predictive regressions of realized volatility (Panel A) and realized returns (Panel B) during stress episodes, using systemic risk measures constructed using only information available up to the quarter before the start of each episode. We note that the samples underlying Columns (1)–(6) and Columns (7)–(12) are different, in both Panel A and Panel B. Columns (1)–(6) include both early and modern episodes, spanning from 1903 to 2023, while Columns (7)–(12) use measures based on leverage and therefore cover only the modern episodes, starting with the oil price shock of 1973. For the measures presented in Columns (1)–(6), we also provide separate tables reporting results for the early and modern samples individually in the Appendix. Overall, the results confirm the expected signs: higher systemic risk measures are generally associated with higher realized volatility and lower realized returns during stress episodes.

In Panel A, the systemic risk measures become notably more informative about the cross-section of realized volatility once episode fixed effects are included. Coefficient estimates increase in both economic magnitude and statistical significance for banks and non-banks,[15] indicating that these measures are more informative about the cross-section of financial firm vulnerability once aggregate variation across episodes is accounted for. This is consistent with the idea that systemic risk measures tend to be low prior to stress episodes—when volatility is still subdued—a pattern related to the volatility paradox (Brunnermeier & Sannikov, 2014). Without fixed effects, the measures are less effective because they fail to capture the level of realized volatility during the episode. However, conditional on being in a stress episode, they help explain which institutions are most exposed.

This pattern does not hold for leverage or leverage-based measures such as SRISK, since leverage tends to predict future aggregate volatility. In particular, SRISK and market leverage (Lvg) are more predictive for banks than for non-banks, consistent with theories emphasizing the role of balance sheet fragility and funding risk in banking crises. Comparing the adjusted R-squared values in Columns (11)–(12), we also note that the regression fit systematically improves when we decompose SRISK/ME and estimate separate parameters for its components MES and Lvg.

Panel B presents similar patterns for realized returns. Systemic risk measures generally become more significant and economically meaningful when episode fixed effects are included. This further suggests that the systemic risk measures are more effective in explaining relative rather than absolute performance during stress events. As in Panel A, SRISK and leverage significantly predict returns for banks but not for non-banks. For non-banks, the signs on these variables are often inconsistent, potentially due to crisis-related spillovers or amplification mechanisms that disproportionately affect banks.

3.1.2 Severe crises, bank crises, and milder stress episodes

To assess whether the predictive power of systemic risk measures in the cross-section varies with the severity of financial stress, we estimate regressions that interact each risk measure with an episode-specific variable capturing the intensity of market stress. Specifically, we define severity as the drawdown in a credit spread over the course of the episode, measured as the cumulative

[15] The coefficient on ΔCoVaR for non-banks reverses sign—from negative to positive—when fixed effects are included,

increase in percentage points. This credit spread proxy varies over time: we use the Gilchrist & Zakrajšek (2012) spread (GZ spread) for the period 1973–2023, the spread between Moody's Baa and Aaa corporate bond yields for 1922–1972, and the average yield on U.S. railroad bonds from Macaulay (1938) for 1895–1921. As an alternative measure, we also use an indicator variable identifying the most severe banking crises in our sample, namely the Great Depression and the GFC. The regressions reported in Table 3 allow us to test whether systemic risk measures are more informative when stress episodes are more severe. All regressions include episode fixed effects, so that the analysis focuses on explaining the cross-sectional differences in market outcomes among financial firms during stress episodes.

[INSERT TABLE 3 HERE]

In Panel A, which focuses on realized volatility, the interaction terms between each systemic risk measure and the episode-level drawdown confirm that systemic risk measures are especially informative during more severe episodes. For example, the interaction term for ΔCoVaR is large and highly significant (Column 2), indicating that institutions with high pre-crisis ΔCoVaR experience significantly greater realized volatility when credit conditions deteriorate sharply. Similar patterns are observed for eΔCoVaR and MES, both of which capture co-movements with market returns and tend to amplify during systemic episodes. In contrast, leverage-based measures such as SRISK and Lvg do not display significant interactions with severity. Their coefficients remain positive even during milder episodes, suggesting that leverage consistently predicts higher volatility regardless of broader market stress.

The distinction between severe and milder episodes is particularly relevant for ΔCoVaR, consistent with the interpretation that firm-level spillovers intensify during periods of heightened stress. To illustrate, consider the GFC, during which the drawdown variable reaches 6.4 p.p. The total marginal effect of ΔCoVaR on realized volatility in such a case combines the baseline coefficient and the interaction term $-0.45+0.52\times 6.4$p.p., resulting in a significantly larger marginal effect of 2.9 p.p., compared to a counterfactual episode with no credit spread widening, where the marginal effect would be negative (-0.45 p.p.). Similarly, in Column (3), the marginal effect of ΔCoVaR rises to 2.95 p.p. (0.14 + 2.81) during the most severe banking crises (the Great Depression and the GFC), while it remains only 0.14 p.p. and is not statistically significant during milder episodes. We find similar results for eΔCoVaR and MES, which are also more informative during severe banking crises, but unlike ΔCoVaR, they remain statistically significant predictors of realized volatility even during milder stress episodes.

Leverage-based measures like SRISK/ME show weaker sensitivity to the continuous drawdown variable but display a notable difference during the GFC. For example, in Column (12), the marginal effect of SRISK/ME is 2.98 p.p. during the GFC, compared to only 0.43 p.p. during other stress episodes. This pattern reflects the more pronounced role of market leverage during the GFC, as evidenced in Column (15).

Panel B presents the regressions for realized returns. Here, the interaction terms between systemic risk measures and episode severity are generally small and statistically insignificant. However, SRISK/ME and market leverage significantly predict realized returns only during the GFC, as reported in Columns (12), (15), and (18). In other words, institutions with higher systemic risk tend to experience lower returns even during milder stress episodes, but market leverage becomes an important predictor of lower returns only during a major banking crisis.

Overall, the results in Table 2 emphasize two complementary roles of systemic risk measures: they are highly informative about *relative* volatility outcomes when stress intensifies, and they consistently predict *relative* performance losses across institutions, even in less severe episodes.

3.2 Bank Failures

We define a failure at the parent bank level — where the parent bank is the entity with publicly listed stock — if either the parent itself or any of its subsidiary banks, as identified from the FFIEC relationship table, failed or received assistance from the FDIC. We count 51 bank failures out of 1,131 institutions in the modern sample (1959-2023). The most recent failures are Silicon Valley Bank (March 7, 2023), Signature Bank (March 12, 2023), and First Republic Bank (May 1, 2023).

While these bank failures correspond to a stress episode according to our definition in Table 1, bank failure dates do not always coincide with a stress episode and are in fact often delayed (notably, following the GFC). To retain a maximum number of bank failures in the analysis, we do not condition our sample on stress episodes in this section.

3.2.1 Systemic risk measures trends before failure

The date of a bank failure is defined in our sample as the date of the last available Call Report for the bank that failed or received assistance according to the FDIC. We follow Correia, Luck & Verner (2024), who document gradually deteriorating trends in balance sheet indicators during the ten years before a bank's failure. Using the same methodology, we study the dynamics of systemic risk indicators for failing banks. Specifically, we estimate: $y_{it} = \alpha_i + \sum_{j=-40}^{-1} \beta_j \times \mathbf{1}_{j=t} + \epsilon_{it}$, where y_{it} is a systemic risk indicator or a capitalization measure of failing bank i, j indexes the number of quarters before failure, and α_i are bank fixed effects. The sample is restricted to banks that failed from 1959 through 2023, and the ten years before they failed. The trends in systemic risk indicators before failure are captured by the sequence $\{\beta_j\}$, which is presented in Figure 2, with $j = -40$ (ten years before failure) serving as the benchmark period.

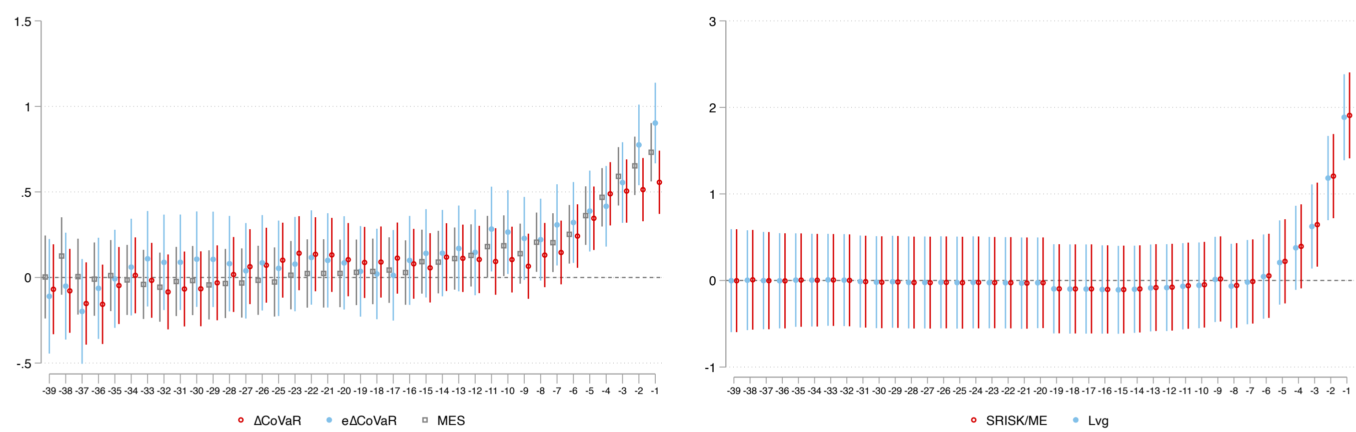

Figure 2: The figure presents the sequence of coefficients from estimating $y_{it} = \alpha_i + \sum_{j=-40}^{-1} \beta_j \times \mathbf{1}_{j=t} + \epsilon_{it}$ (see Correia, Luck & Verner, 2024), where the dependent variable is a quarterly market-based measure of systemic risk for failing bank i, as indicated in the figure legend. The sample is restricted to banks that failed from 1959 through 2023, and the ten years before they fail. All measures are scaled by their standard deviations.

From Figure 2, we note that all systemic risk indicators increase in the year before failure. $\triangle$CoVaR, e$\triangle$CoVaR and MES respectively increase six, seven, and eight quarters before the last Call Report is filed by the failing bank. In the three quarters before the bank files its last Call report, SRISK/ME and market leverage are significantly larger than ten years before failure. For example, SRISK/ME increases by almost two standard deviations the quarter before the last Call Report relative to ten years before failure. Measures that capture co-movements with a market index increase by less than a standard deviation, but start increasing earlier.

3.2.2 Predicting bank failures

Next, we compare failing banks with their non-failing counterparts. We estimate the following specification to predict a bank failure in the next year:

$$Failure_{i,t+1\to t+h} \quad = \quad \alpha + \beta_1 Measure_{it} + (\beta_2 Measure_{it} + \beta_3) \times gz_t + \epsilon_{i,t+1\to t+h} \tag{6}$$

where $Failure_{i,t+1\to t+h} = 1$ if a bank fails in the next $h = 4$ quarters and 0 otherwise, and gz_t denotes the GZ spread (Gilchrist & Zakrajšek, 2012) in quarter t. To understand how stress conditions influence the predictive model, we use the GZ spread as an indicator of aggregate economic conditions and interact it with bank-level systemic risk measures. The choice of this indicator is motivated by the observation that bank failures often occur with a delay relative to systemic stress episodes, typically at the end or after a stress episode—that is, when the GZ spread reaches or approaches its peak.

We present the OLS estimates of this regression in Table 4. In Panel A, we report the results of a restricted regression, imposing $\beta_2 = \beta_3 = 0$, while Panel B reports the results of the full unrestricted specification of eq. (6). We also report the adjusted R-squared of each linear probability model, and the Pseudo R-squared from a corresponding logit regression. To quantify the predictive power of each measure, we construct receiver operating characteristic (ROC) curves and compare the area under the curve (AUC) across models. The ROC curve shows the true positive rate as a function of the false positive rate for any classification threshold. We display the ROC curves in Figure 3 and report the AUC of each model in Table 4. A model with an AUC of 0.5 is uninformative and corresponds to the green reference line in Figure 3. Models with AUC above 0.5 are more likely to predict true failures than false ones.

Table 4 shows that systemic risk measures tend to predict a higher probability of bank failure in the next year. Additionally, banks with lower market leverage are less likely to fail. The capital structure of the bank appears to be a strong predictor of bank failure. In Panel A, the AUC is the largest for SRISK/ME (0.9096) that includes both MES and market leverage, followed by market leverage (0.8801), and by MES combined with market leverage (0.8655). From the left panel of Figure 3, we observe the same pattern where measures that are a function of the capital structure of the bank (SRISK/ME, market leverage) are stronger predictors of failures than measures capturing co-movements with the market index under stress conditions (△CoVaR, e△CoVaR, MES).

[INSERT TABLE 4 HERE]

In Panel B, reporting the OLS estimates of the full regression specification of eq. (6), the β_1 parameters capture the marginal effect of each indicator under a counterfactual GZ spread of zero.[16] In contrast, the β_2 parameters capture the differential effect of each indicator in predicting bank failure when the GZ spread is wider by one percentage point. For example, to gauge the differential effect of the indicators during the GFC where the GZ spread reached a peak of almost 8 percentage points, the β_2 estimates should be multiplied by a factor of 8.[17] Thus, the marginal effect of a systemic risk measure is $\beta_1 + 8\beta_2$ during the worst months of the GFC.

[16] Similarly, the β_3 parameters capture the effect of the GZ spread for a bank with a systemic risk measure equal to zero.

[17] See Figure 1 (Panel A) in Favara et al. (2016).

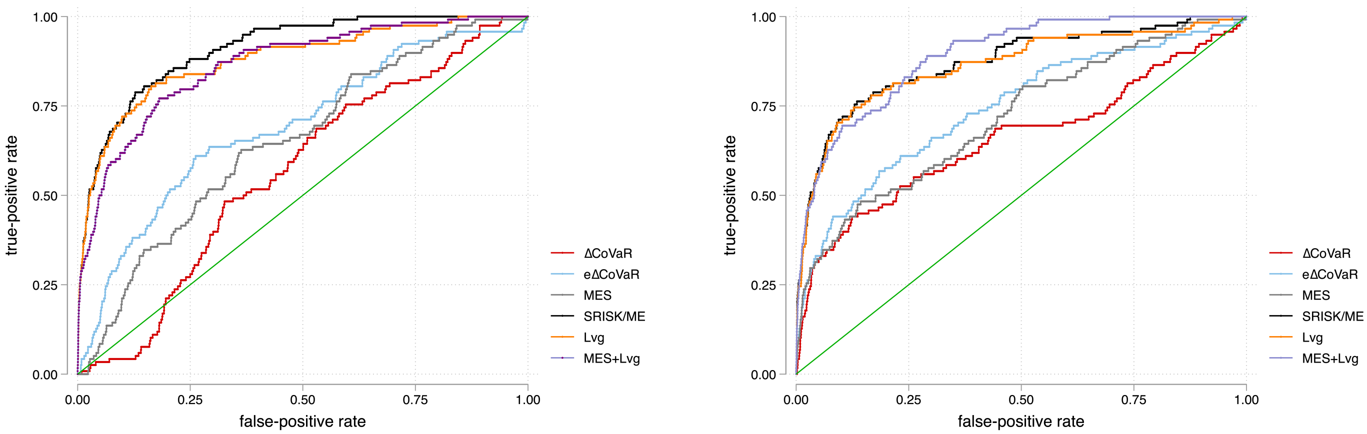

Figure 3: Receiver operating characteristic curves (ROC) from predicting bank failures next year using market-based indicators of systemic risk. The left panel displays the ROC curves relative to the classification model of eq. (6), imposing $\beta_2 = \beta_3 = 0$. The right panel displays ROC curves relative to the unrestricted model of eq. (6), where measures are interacted with the GZ spread.

Comparing Panel A and Panel B, we find that the predictive accuracy of $\triangle$CoVaR, e$\triangle$CoVaR, and MES improves after accounting for interactions with the GZ spread. In contrast, the AUCs of SRISK/ME and market leverage slightly decrease under the full specification, despite higher Pseudo R-squared values and statistically significant coefficients. While capital structure consistently predicts bank failure, the predictive power of systemic risk measures—$\triangle$CoVaR, e$\triangle$CoVaR, and MES—depends strongly on economic conditions. For example, at the lowest observed value of the GZ spread (0.6 percentage points in 1978), a one percentage point increase in $\triangle$CoVaR, e$\triangle$CoVaR, and MES is associated with an 8%, 10%, and 5% higher failure rate, respectively. By contrast, during the GFC, these marginal effects rise to 116%, 138%, and 63%, respectively. The ROC curves corresponding to the regression results in Panel B are shown in the right panel of Figure 3. Compared to the left panel, the ROC curves for $\triangle$CoVaR, e$\triangle$CoVaR, and MES shift closer to those of the capital structure-based measures, reflecting the improvement in predictive performance when accounting for aggregate stress conditions.

3.3 Balance Sheet Outcomes

To assess how financial fragility, proxied by our systemic risk measures, might affect the real economy, we analyze their impact on bank balance sheets during stress episodes. We follow a methodology similar to the one adopted in Section 3.1. We focus on outcomes reported by banks in Call Reports during stress episodes, and measure systemic risk the quarter before the starting date of the episode as identified in Table 1. We estimate the following specification to predict bank balance sheet outcomes during a stress episode:

$$y_{ie} \quad = \quad \beta Measure_{ie} + \delta controls_{ie} + \alpha_e + \epsilon_{ie} \tag{7}$$

where y_{ie} is the average balance sheet outcome of bank i during episode e, and α_e are episode fixed effects. The variables $controls_{ie}$ include the size of the bank and, as a pre-trend, the average balance sheet outcome during the year prior to the episode.

[INSERT TABLE 5 HERE]

The estimates of this regression are presented in Table 5, where the dependent variable is, successively, loan growth, the return on assets (to measure profitability), and the growth in the ratio of uninsured deposits to total deposits (to proxy for run risks). Growth variables are quarterly percentage growth rates, and the return on assets is the quarterly net income divided by lagged total assets. Dependent variables are then averaged over the quarters of an episode for each bank to construct y_{ie}.

Table 5 shows that systemic risk measures predict reduced loan growth, lower bank profitability, and less reliance on uninsured deposits during stress episodes. For example, a one p.p. increase in $\triangle$CoVaR is associated with a -1.14 p.p. average loan growth rate during a stress episode. We report the results separately for the two main categories of C&I and real estate loans in the Appendix and find that systemic risk measures mostly predict C&I loans, although $\triangle$CoVaR and e$\triangle$CoVaR also predict negative real estate loan growth. Systemic risk indicators also predict lower bank profitability in a stress episode. For example, a one p.p. increase in e$\triangle$CoVaR is associated with a -0.02 percent return on assets, which corresponds to 2.5 times its standard deviation. Similarly, a one p.p. increase in MES is associated with a -0.01 percent return on assets. Finally, not all systemic risk measures have statistically significant estimates for predicting uninsured deposits dynamics, but the estimates of $\triangle$CoVaR, e$\triangle$CoVaR, and MES are statistically significant at the 5% level. The measures predict a loss of uninsured deposits, as opposed to insured ones, during a stress episode, consistent with the interpretation of runs by uninsured depositors from banks with weaker fundamentals.

Overall, these results are consistent with banking fragility leaving the economy less resilient in terms of bank credit and funding outcomes; crucially, the fragility can be predicted by market-based measures of bank systemic risk.

4 Comparative Analysis of Market-Based and Balance Sheet Indicators of Risk

In this section, we assess the predictive power of market-based systemic risk measures relative to, and in combination with, balance sheet-based indicators across the three exercises we have focused on, viz., (i) predicting market outcomes during stress episodes, (ii) predicting balance sheet outcomes during stress episodes, and (iii) forecasting bank failure over one-, three-, and five-year horizons. In each case, we compare models that include only market-based indicators, only balance sheet variables, and both together. The objective is to evaluate the relative and joint explanatory power of these measures and to identify where they offer complementary insights.

There are at least two important reasons to undertake this exercise. One, balance-sheet based measures if used for direct supervisory or regulatory purposes may be easier to manipulate ("regulatory arbitrage") relative to market-based measures, rendering them less useful post adoption in policy (a version of Goodhart's Law, see Goodhart (1984)). Hence, if market-based measures can perform as good or better as balance-sheet based measures, then there are some advantages in terms of robustness. Secondly, if market-based measures and balance-sheet measures have complementary information, then they can be used jointly by supervisors and regulators, as a sort of "belts and suspenders" approach, so as to avoid having blind spots through which financial fragility can unexpectedly emerge.

As balance sheet indicators, we use the variables proposed by Correia, Luck & Verner (2024) (hereafter, "CLV") to predict bank failures. Correia, Luck & Verner (2024) show that bank failures are surprisingly predictable using simple accounting metrics from banks' financial statements. In this section, we replicate their variables at the consolidated parent bank level. The CLV framework includes four variables: solvency (net income over total assets), which captures a lower risk of insolvency; funding (time deposits over total deposits), which reflects reliance on less stable, non-core funding; their interaction; and asset growth (the change in log total assets over the prior three years), which signals potential risk-taking through rapid expansion (Fahlenbrach, Prilmeier & Stulz, 2018). Market-based systemic risk measures include ΔCoVaR, eΔCoVaR, MES, SRISK/ME, and market leverage (Lvg), which may enter directly or through interactions with the risk measure.

[INSERT TABLE 6 HERE]

Table 6 summarizes the comparative explanatory power of these indicators, based on the results of 480 regressions. Panel A presents the best-performing specification for each outcome when comparing market-based measures to the baseline CLV model. Panel B reports results from models that include both market-based and CLV variables. Market-based measures are combined with leverage either additively ("+Lvg") or interactively ("×Lvg"). Each row indicates whether the sample includes non-banks and whether episode fixed effects are included. Panel C focuses on predicting bank failures over one-, three-, and five-year horizons. Explanatory power is measured by R-squared for continuous outcomes (Panels A and B) and by the area under the ROC curve (AUC) for binary outcomes (Panel C). R-squared values are reported as adjusted R-squared when episode fixed effects are excluded and as within R-squared when they are included. Each cell shows the highest-performing specification in terms of percentage gain relative to the CLV baseline in parentheses. Methodological details and the full set of results are reported in the Appendix.

Results from Panel A show that market-based measures perform particularly well when episode fixed effects are included, consistent with their strength in explaining the cross-sectional ranking of risk within stress episodes rather than the unconditional level of outcomes across episodes. The best-performing model for realized volatility includes MES × Lvg (R-squared = 0.043), improving on the CLV model by 19%. For realized returns, eΔCoVaR × Lvg yields an R-squared of 0.010, outperforming CLV by 25%. Profitability is best predicted by specifications that interact market-based measures (ΔCoVaR, eΔCoVaR, MES) with leverage, reaching an R-squared of 0.215—7% above CLV. ΔCoVaR consistently performs best for predicting loan growth, achieving a 20% gain relative to the CLV benchmark. The only exception is run risk, where the CLV model remains dominant, even with fixed effects included.

In Panel B, predictive performance improves markedly when market-based and balance sheet indicators are combined. For several outcomes, such as volatility and returns, the gains are large—up to 260% and 125%, respectively. Even for run risk, where CLV alone performs well, performance improves when ΔCoVaR × Lvg is added.

Expanding the sample to include non-bank financial institutions confirms these patterns. In this case, we use Compustat to replicate the CLV variables, omitting the funding ratio, which cannot be constructed due to the lack of deposit decomposition. The CLV model in this sample is thus based solely on solvency and asset growth. Market-based measures remain strong predictors of cross-sectional risk, with MES × Lvg delivering the highest explanatory power for both volatility and returns. These results reinforce the conclusion that market-based indicators are particularly effective at ranking risk within episodes, while balance sheet fundamentals better reflect the overall level of risk over time.

Panel C turns to bank failure prediction. At the one-year horizon, market-based indicators—especially SRISK/ME—outperform the CLV model, with an AUC of 0.9096 compared to 0.8455.

Over three- and five-year horizons, however, the CLV model performs best, with AUCs of 0.8098 and 0.7710, respectively. We report the coefficient estimates of the CLV model in Appendix Table 13. Our estimates have the same signs as those in Correia, Luck & Verner (2024), despite the different sample used in our analysis.[18] The strong predictive power of CLV is consistent with the regulatory framework governing bank failures, which relies heavily on balance sheet measures. In particular, U.S. banking laws define critical undercapitalization based on the tangible equity ratio, a book value metric.[19] The FDIC Resolution Handbook also considers balance sheet indicators, such as the asset and liability composition of the bank, to determine the resolution structure. Profitability, a key component of CLV, predicts failure because persistent losses erode book capital, bringing banks closer to closure thresholds. CLV variables are thus particularly suited to predicting FDIC resolutions and assistance transactions.

Across all exercises, models that combine market-based and balance sheet indicators deliver the highest predictive power. In the failure prediction context, the combination of ΔCoVaR $\times$ Lvg and CLV increases predictive performance by 10% and 9% for the one- and three-year horizons, respectively. Even for predicting failures at the five-year horizon, the best results are obtained by combining a measure of co-movement with the market index under stress conditions (ΔCoVaR, eΔCoVaR, or MES), market leverage, and CLV variables, leading to an 8% gain in predictive accuracy. These findings underscore the complementarity of the two approaches. A robust monitoring framework should therefore integrate forward-looking market signals with structural balance sheet fundamentals to capture distinct but reinforcing dimensions of systemic risk.

5 Conclusion

In this paper, we documented robust cross-sectional predictive power in stock market based systemic risk measures for financial fragility —adverse stock market returns for financial firms, bank failures, and deterioration in the growth of bank deposits, credit and profitability— witnessed over the period 1895 to 2023. These findings also point to several avenues for future research on the drivers and underlying mechanisms of systemic risk. First, both non-bank financial intermediaries (NBFIs) and deposit-like wholesale finance claims have evolved significantly over the sample period we studied. How has the systemic risk of NBFIs evolved over time and what role has reliance on wholesale liabilities played in this evolution? Second, banks and NBFIs have become significantly interconnected over time via both term lending and provision of credit lines by banks. How does this interconnection show up in systemic risk measures? Third, an advantage of using stock prices of financial firms in systemic risk measurement is that they are least vulnerable to a bias arising from public backstops. Nevertheless, the role of these backstops has continued to rise, for banks as well as NBFIs. Has this affected the efficacy of systemic risk measures over time? Clearly, there is much scope for further research.

[18] In contrast to their methodology, we estimate the probability of failure at the parent bank level, and our sample is limited to publicly traded banks. As a result, the AUCs we obtain in our analyses should not to be compared to the AUC reported in Correia, Luck & Verner (2024) since the sample is different.

[19] U.S. Prompt Corrective Action rules (12 U.S.C. §1831o) require regulators to close an insured depository institution within 90 days of becoming critically undercapitalized—typically defined as a tangible equity ratio below two percent.

References

Acharya VV. 2009. A theory of systemic risk and design of prudential bank regulation. *Journal of financial stability* 5(3):224–255

Acharya VV, Engle RF, Pierret D. 2014. Testing macroprudential stress tests: The risk of regulatory risk weights. *Journal of Monetary Economics* 65:36–53

Acharya VV, Engle RF, Richardson M. 2012. Capital shortfall: A new approach to ranking and regulating systemic risks. *The American Economic Review* 102(3):59–64

Acharya VV, Pedersen LH, Philippon T, Richardson M. 2017. Measuring systemic risk. *The Review of Financial Studies* 30(1):2–47

Adrian T, Brunnermeier MK. 2016. CoVaR. *The American Economic Review* 106(7):1705–1741

Baron M, Verner E, Xiong W. 2021. Banking crises without panics. *The Quarterly Journal of Economics* 136(1):51–113

Barro RJ, Ursúa JF. 2017. Stock-market crashes and depressions. *Research in Economics* 71(3):384–398

Benoit S, Colliard JE, Hurlin C, Pérignon C. 2017. Where the risks lie: A survey on systemic risk. *Review of Finance* 21(1):109–152

Billio M, Getmansky M, Lo AW, Pelizzon L. 2012. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics* 104(3):535–559

Bisias D, Flood M, Lo AW, Valavanis S. 2012. A survey of systemic risk analytics. *Annual Review of Financial Economics* 4(1):255–296

Bordo M, Eichengreen B, Klingebiel D, Martinez-Peria MS. 2001. Is the crisis problem growing more severe? *Economic policy* 16(32):52–82

Borio C. 2003. Towards a macroprudential framework for financial supervision and regulation? *CESifo Economic Studies* 49(2):181–215

Brownlees C, Chabot B, Ghysels E, Kurz C. 2020. Back to the future: Backtesting systemic risk measures during historical bank runs and the great depression. *Journal of Banking & Finance* 113:105736

Brownlees C, Engle RF. 2017. SRISK: A conditional capital shortfall measure of systemic risk. *The Review of Financial Studies* 30(1):48–79

Brunnermeier MK. 2024. Presidential address: Macrofinance and resilience. *The Journal of Finance*

Brunnermeier MK, Sannikov Y. 2014. A macroeconomic model with a financial sector. *American Economic Review* 104(2):379–421

Bry G, Boschan C. 1971. Cyclical analysis of times series: selected procedures and computer programs. National Bureau of Economic Research

Correia SA, Luck S, Verner E. 2024. Failing banks. National Bureau of Economic Research Working Paper 32907

Crockett A. 2000. Marrying the micro-and macro-prudential dimensions of financial stability. *BIS speeches* 21

Fahlenbrach R, Prilmeier R, Stulz RM. 2018. Why does fast loan growth predict poor performance for banks? *The Review of Financial Studies* 31(3):1014–1063

Favara G, Gilchrist S, Lewis KF, Zakrajšek E. 2016. Updating the recession risk and the excess bond premium. FEDS Notes. Board of Governors of the Federal Reserve System

Gilchrist S, Zakrajšek E. 2012. Credit spreads and business cycle fluctuations. *The American Economic Review* 102(4):1692–1720

Goodhart CAE. 1984. Problems of monetary management: The U.K. experience. In *Monetary Theory and Practice: The UK Experience*, ed. AS Courakis. Palgrave Macmillan, London, 91–121

Greenwood R, Hanson SG, Shleifer A, Sørensen JA. 2022. Predictable financial crises. *The Journal of Finance* 77(2):863–921

Huang X, Zhou H, Haibin Z. 2012. Systemic risk contributions. *Journal of Financial Services Research* 42:55–83

Huang X, Zhou H, Zhu H. 2009. A framework for assessing the systemic risk of major financial institutions. *Journal of Banking & Finance* 33:11:2036–2049

Jamilov R, König T, Müller K, Saidi F. 2024. Two centuries of systemic bank runs. Tech. rep., ECONtribute Discussion Paper

Jordà Ò, Schularick M, Taylor AM. 2015. Leveraged bubbles. *Journal of Monetary Economics* 76:S1–S20

Laeven L, Valencia F. 2013. Systemic banking crises database. *IMF Economic Review* 61(2):225–270

Laeven L, Valencia F. 2020. Systemic banking crises database II. *IMF Economic Review* 68:307–361

Macaulay FR. 1938. The movement of interest rates, bond yields and stock prices in the United States since 1856. National Bureau of Economic Research

Markham JW. 2022. From the age of derivatives into the new millennium: A financial history of the United States 1970–2001. Routledge

Reinhart CM, Rogoff KS. 2009. This time is different: Eight centuries of financial folly. Princeton, NJ: Princeton University Press

Reinhart CM, Rogoff KS. 2011. From financial crash to debt crisis. *The American Economic Review* 101(5):1676–1706

episode	start	end	months	credit	stock index		fin. index	
				sprd (%)	ret (%)	dd (%)	ret (%)	dd (%)
1903-04 stock market crash	4-1902	9-1903	16	0.33	-31.65	-31.65	-36.65	-46.38
Panic of 1907	1-1906	11-1907	22	0.68	-42.88	-42.88	-30.49	-30.49
1910-11 stock market crash	1-1910	7-1910	6	0.12	-15.60	-15.60	7.88	-6.00
1914 banking crisis	1-1913	12-1914	23	0.38	-34.81	-34.81	-18.56	-23.99
1916-21 stock market crash	1-1917	5-1920	40	1.79	-3.79	-23.33	3.92	-30.29
The Great Depression	2-1929	5-1932	39	4.64	-85.90	-88.24	-91.95	-95.05
Recession of 1937–38	2-1937	4-1938	14	1.86	-40.34	-47.13	-16.30	-30.27
1939-41 stock market crash	10-1939	6-1940	8	0.42	-19.63	-23.48	-15.13	-22.68
Post-World War II Recession	1-1946	7-1949	42	0.32	-14.05	-21.42	-5.07	-19.40
Recession of 1958	4-1956	1-1958	21	0.79	-12.81	-15.86	-22.86	-30.29
High inflation in the U.S.	9-1973	12-1974	15	1.75	-36.77	-41.40	-35.73	-46.82
1977-82 stock market crash	10-1978	10-1982	48	1.21	43.54	-23.79	43.44	-34.12
S&L Crisis	5-1984	8-1986	27	1.58	68.00	0	54.24	-11.14
LTCM hedge fund failure	8-1997	10-1998	14	1.12	22.15	-15.57	-10.08	-36.40
Dot.com Bubble	1-2000	10-2002	33	2.36	-36.48	-46.28	-24.11	-36.84
Global Financial Crisis	5-2007	11-2008	18	6.44	-41.45	-42.16	-52.09	-52.09
European Sovereign Debt Crisis	3-2011	9-2011	6	1.16	-14.66	-17.03	-18.99	-20.15
2014-16 oil price shock	6-2014	2-2016	20	1.52	-1.43	-8.89	0.90	-9.81
Covid19 Pandemic	12-2019	3-2020	3	2.37	-20.00	-20.00	-18.47	-18.47
Ukraine war/energy crisis	12-2021	6-2022	6	0.70	-20.58	-20.58	-8.77	-8.77
Banking Stress of 2023	1-2023	5-2023	4	0.15	2.53	0	-2.96	-5.97

Table 1: Stress episodes description. "start" and "end" respectively indicate the start date and the end date of the episode, and "months" is the length in months of the episode. "credit sprd" is the change in the credit spread in percentage points. The credit spread is proxied by the GZ spread (Gilchrist & Zakrajšek, 2012) for the 1973-2023 period, the spread between Moody's Baa and Aaa Corporate Bond Yields for 1922-1972, and the average yield on U.S. railroad bonds (Macaulay, 1938) for 1895-1921. "stock index" is the S&P500 index for 1959-2023, and the Dow Jones Industrial Average index before 1959. "fin. index" is the CRSP financial index for 1959-2023, and the GFD financial index before 1959. "ret" denotes the index return in percentage points over the episode. "dd" is the maximum drawdown on an index, defined as the percentage difference between its minimum and the prior maximum value within an episode. There is no drawdown when the minimum index value is reached at the beginning of the episode.

Panel A: Realized volatility during stress episodes												
Measure:	ΔCoVaR		eΔCoVaR		MES		SRISK/ME		Lvg		MES (+Lvg)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Measure×bank	-0.02	0.25**	0.19***	0.32***	-0.02	0.31***	1.46***	1.46***	0.12***	0.08***	0.03	0.41***
	(-0.25)	(2.11)	(2.73)	(4.10)	(-0.36)	(7.37)	(4.79)	(4.51)	(4.76)	(3.75)	(0.30)	(5.05)
Measure×nonbank	-0.78***	0.28**	0.47***	0.61***	0.32***	0.47***	0.35***	0.33*	0.02***	0.02*	0.37***	0.52***
	(-6.36)	(2.08)	(6.76)	(8.83)	(6.27)	(10.58)	(2.70)	(1.93)	(2.73)	(1.94)	(6.02)	(8.34)
Lvg×bank											0.12***	0.09***
											(4.63)	(3.72)
Lvg×nonbank											0.02**	0.02
											(2.30)	(1.47)
Adj. R-squared	0.186	0.398	0.189	0.409	0.191	0.421	0.023	0.238	0.022	0.233	0.036	0.265
Panel B: Realized returns during stress episodes												
Measure:	ΔCoVaR		eΔCoVaR		MES		SRISK/ME		Lvg		MES (+Lvg)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Measure×bank	-4.51***	-5.38***	-1.14*	-1.23*	-0.98***	-1.25***	-2.38***	-2.73***	-0.19***	-0.11***	-0.38	-1.15***
	(-10.30)	(-6.63)	(-1.76)	(-1.82)	(-5.42)	(-7.05)	(-3.46)	(-4.35)	(-3.34)	(-2.60)	(-1.60)	(-5.18)
Measure×nonbank	-1.40***	-3.13***	-0.79***	-1.58***	0.23	-0.83***	0.09	0.39***	-0.01	0.02**	0.73**	-0.39*
	(-3.20)	(-4.62)	(-2.93)	(-5.90)	(1.05)	(-4.83)	(0.46)	(2.59)	(-0.32)	(2.00)	(2.54)	(-1.74)
Lvg×bank											-0.17***	-0.10**
											(-2.86)	(-2.18)
Lvg×nonbank											-0.01	0.03**
											(-0.84)	(2.53)
Adj. R-squared	0.006	0.180	0.004	0.179	0.004	0.180	0.003	0.246	0.003	0.245	0.006	0.247
Observations	11,636	11,636	11,636	11,636	11,636	11,636	6,316	6,316	6,316	6,316	6,316	6,316
Episode FE	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y

Table 2: Predictive regression of realized volatility (Panel A) and realized returns (Panel B) during stress episodes (1895-2023). The dependent variable is the market outcome of a financial institution during stress episodes. The systemic risk measures and the control variable for the firm size are all measured the quarter before the episode starts. Stress episodes are defined in Table 1. Coefficient estimates on SRISK/ME and Lvg are multiplied by 100. The even Columns (2) to (12) include episode fixed effects. t-statistics based on standard errors clustered at the firm level are reported in parentheses. Columns (1)-(6) cover the 1895-2023 period with a sample of 4,054 financial institutions, including 1,953 banks. Columns (7)-(12) cover the 1959-2023 period with 2,179 financial institutions, including 758 banks.

Panel A: Realized volatility during stress episodes																		
Measure:		ΔCoVaR			eΔCoVaR			MES			SRISK/ME			Lvg			MES (+Lvg)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Measure	0.25**	-0.45***	0.14	0.51***	0.30***	0.46***	0.43***	0.32***	0.39***	0.48**	0.57	0.43**	0.03**	0.05	0.03**	0.49***	0.41***	0.46***
	(2.28)	(-2.76)	(1.27)	(8.39)	(3.76)	(7.68)	(11.95)	(6.55)	(10.59)	(2.16)	(1.08)	(2.08)	(2.16)	(1.06)	(2.05)	(9.46)	(5.27)	(8.42)
Measure×dd		0.52***			0.15***			0.07***			-0.05			-0.01			0.04*	
		(6.11)			(4.29)			(3.82)			(-0.21)			(-0.44)			(1.89)	
Measure×crisis			2.81***			0.78***			0.39***			2.55**			0.12***			0.37***
			(6.30)			(3.94)			(4.45)			(2.05)			(2.64)			(3.69)
Lvg																0.03*	0.04	0.02*
																(1.82)	(0.84)	(1.69)
Lvg×dd																	-0.01	
																	(-0.33)	
Lvg×crisis																		0.14**
																		(2.47)
Within R-squared	0.018	0.025	0.024	0.034	0.037	0.037	0.055	0.058	0.058	0.029	0.029	0.033	0.026	0.027	0.028	0.068	0.069	0.072
Panel B: Realized returns during stress episodes																		
Measure:		ΔCoVaR			eΔCoVaR			MES			SRISK/ME			Lvg			MES (+Lvg)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Measure	-3.99***	-4.35***	-3.93***	-1.46***	-1.77***	-1.50***	-0.97***	-0.97***	-0.96***	-0.02	0.37	0.07	0.01	0.04	0.01	-0.58***	-0.55**	-0.59***
	(-5.86)	(-6.24)	(-5.69)	(-4.98)	(-5.22)	(-4.95)	(-6.81)	(-5.99)	(-6.47)	(-0.09)	(0.66)	(0.30)	(0.42)	(0.94)	(0.85)	(-3.10)	(-2.37)	(-3.03)
Measure×dd		0.27			0.22			0.01			-0.21			-0.02			-0.02	
		(0.96)			(1.62)			(0.01)			(-0.77)			(-0.86)			(-0.31)	
Measure×crisis			-1.66			0.73			-0.18			-5.68**			-0.45**			0.08
			(-1.04)			(0.99)			(-0.53)			(-2.04)			(-2.01)			(0.21)
Lvg																0.01	0.05	0.02
																(0.76)	(1.20)	(1.19)
Lvg×dd																	-0.02	
																	(-1.03)	
Lvg×crisis																		-0.45**
																		(-1.99)
Within R-squared	0.007	0.007	0.007	0.006	0.007	0.006	0.008	0.008	0.008	0.003	0.003	0.004	0.003	0.003	0.004	0.006	0.006	0.006
Observations	11,636	11,636	11,636	11,636	11,636	11,636	11,636	11,636	11,636	6,316	6,316	6,316	6,316	6,316	6,316	6,316	6,316	6,316
Episode FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Table 3: Predictive regression of realized volatility (Panel A) and realized returns (Panel B) during stress episodes (1895-2023). The dependent variable is the market outcome of a financial institution during stress episodes. The systemic risk measures and the control variable for the firm size are all measured the quarter before the episode starts. The drawdown variable "dd" is the change in the credit spread over the episode. The variable "crisis" indicates the Great Depression and the GFC. Stress episodes and the credit spread are defined in Table 1. Coefficient estimates on SRISK/ME and Lvg are multiplied by 100. All columns include episode fixed effects. t-statistics based on standard errors clustered at the firm level are reported in parentheses. Columns (1)-(9) cover the 1895-2023 period with a sample of 4,054 financial institutions, including 1,953 banks. Columns (10)-(18) cover the 1959-2023 period with 2,179 financial institutions, including 758 banks.

Panel A: Failure probability (fail in next year)						
Measure:	ΔCoVaR	eΔCoVaR	MES	SRISK/ME	Lvg	MES (+Lvg)
	(1)	(2)	(3)	(4)	(5)	(6)
Measure	0.05	0.13***	0.04***	0.82***	0.07***	0.04***
	(1.63)	(7.39)	(5.45)	(40.81)	(40.48)	(5.03)
Lvg						0.07***
						(40.42)
Observations	30,889	30,889	30,889	30,889	30,889	30,889
Adj. R-squared	0.001	0.002	0.001	0.051	0.050	0.051
Pseudo R2 (logit)	0.002	0.031	0.018	0.111	0.108	0.122
AUC	0.5639	0.6915	0.6500	0.9096	0.8801	0.8655

Panel B: Failure probability (fail in next year), stress episodes						
Measure:	ΔCoVaR	eΔCoVaR	MES	SRISK/ME	Lvg	MES (+Lvg)
	(1)	(2)	(3)	(4)	(5)	(6)
Measure	-0.28***	-0.24***	-0.12***	0.49***	0.04***	-0.12***
	(-4.02)	(-5.64)	(-6.38)	(10.57)	(11.12)	(-6.28)
Measure×gz	14.53***	17.32***	7.88***	15.16***	1.09***	7.41***
	(5.03)	(9.70)	(9.73)	(8.07)	(7.15)	(9.36)
Lvg						0.04***
						(11.82)
Lvg×gz						0.96***
						(6.29)
gz	0.11*	-0.06	-0.09	0.28***	0.18***	-0.25***
	(1.67)	(-1.03)	(-1.44)	(8.19)	(4.36)	(-3.97)
Observations	30,889	30,889	30,889	30,889	30,889	30,889
Adj. R-squared	0.005	0.009	0.008	0.057	0.055	0.059
Pseudo R2 (logit)	0.051	0.085	0.075	0.157	0.154	0.173
AUC	0.6580	0.7443	0.7174	0.8702	0.8619	0.8905

Table 4: Predicting bank failures: systemic risk and capitalization measures. The dependent variable is equal to one if the bank fails in the next year, and zero otherwise. Measure is △CoVaR in Column (1), e△CoVaR in Column (2), MES in Column (3), SRISK/ME in Column (4), market leverage in Column (5), and MES controlling for market leverage in Column (6). OLS estimates and adjusted R-squared refer to the linear probability model described in eq. (6). Coefficient estimates on SRISK/ME and Lvg are multiplied by 100. Pseudo R-squared are obtained from corresponding logit regressions. AUC is the area under the receiver operating characteristic (ROC) curve. Sample: 819 banks.

Measure:	ΔCoVaR		eΔCoVaR		MES		SRISK/ME		Lvg		MES (+Lvg)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
loan growth												
Measure	-1.36***	-1.14**	-0.49***	-0.31**	-0.23***	-0.14	-0.70***	-0.46*	-0.06***	-0.04*	-0.19***	-0.13
	(-3.08)	(-2.09)	(-2.99)	(-2.01)	(-3.22)	(-1.52)	(-4.14)	(-1.78)	(-4.33)	(-1.67)	(-2.58)	(-1.40)
Lvg											-0.05***	-0.03
											(-3.12)	(-1.51)
Observations	2,222	2,222	2,222	2,222	2,222	2,222	2,222	2,222	2,222	2,222	2,222	2,222
Adj. R-squared	0.006	0.005	0.004	0.004	0.003	0.004	0.003	0.004	0.003	0.004	0.004	0.004
return on assets												
Measure	-0.01	0.01	-0.03***	-0.02**	-0.01**	-0.01**	-0.06**	-0.07**	-0.01***	-0.01***	-0.01	-0.01
	(-0.84)	(0.82)	(-3.14)	(-2.27)	(-2.20)	(-2.17)	(-2.22)	(-2.02)	(-3.25)	(-3.24)	(-1.41)	(-1.56)
Lvg											-0.01***	-0.01***
											(-3.07)	(-3.15)
Observations	2,227	2,227	2,227	2,227	2,227	2,227	2,227	2,227	2,227	2,227	2,227	2,227
Adj. R-squared	0.176	0.297	0.181	0.299	0.179	0.299	0.186	0.307	0.189	0.311	0.189	0.312
uninsured deposit growth												
Measure	-1.07***	-0.36	-0.58**	-0.33	-0.27**	-0.21	-1.65	-1.41	-0.13	-0.11	-0.21**	-0.18
	(-3.55)	(-0.79)	(-2.23)	(-1.25)	(-2.21)	(-1.62)	(-1.34)	(-1.00)	(-1.22)	(-0.93)	(-1.98)	(-1.51)
Lvg											-0.11	-0.10
											(-1.04)	(-0.87)
Observations	1,726	1,726	1,726	1,726	1,726	1,726	1,726	1,726	1,726	1,726	1,726	1,726
Adj. R-squared	0.013	0.028	0.012	0.029	0.012	0.029	0.015	0.031	0.014	0.031	0.015	0.031
Episode FE	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y

Table 5: Predictive regression of balance sheet outcomes during stress episodes (1959-2023). The dependent variable is an average balance sheet outcome of a bank during a stress episode. The systemic risk measure (Measure) and the control variables are all measured the quarter before the episode starts. Control variables include the firm size and the average bank outcome the year before the episode starts. Stress episodes are defined in Table 1. Measure is indicated in the first line of the table. Coefficient estimates on SRISK/ME and Lvg are multiplied by 100. The even Columns (2) to (12) include episode fixed effects. t-statistics based on standard errors clustered at the bank level are reported in parentheses. Sample: 723 banks (loan growth), 724 banks (return on assets), 602 banks (uninsured deposit growth).

Panel A: Market-based vs. CLV balance sheet variables (R^2)						
	Incl.	Market outcomes		Balance sheet outcomes		
Ep. FE	nonbanks	volatility	returns	loan growth	profitability	run risk
N	N	MES+Lvg 0.011 (+120%)	CLV 0.065	ΔCoVaR 0.006 (+20%)	CLV 0.218	CLV 0.042
Y	N	MES×Lvg 0.043 (+19%)	eΔCoVaR×Lvg 0.010 (+25%)	ΔCoVaR 0.006 (+20%)	e,ΔCoVaR,MES×Lvg 0.215 (+7%)	CLV 0.036
N	Y	CLV 0.032	CLV 0.007			
Y	Y	MES×Lvg 0.069 (+47%)	MES×Lvg 0.012 (+140%)			

Panel B: Market-based + CLV balance sheet variables (R^2)						
	Incl.	Market outcomes		Balance sheet outcomes		
Ep. FE	nonbanks	volatility	returns	loan growth	profitability	run risk
N	N	MES+Lvg 0.018 (+260%)	ΔCoVaR×Lvg 0.085 (+31%)	ΔCoVaR 0.007 (+40%)	e,ΔCoVaR,MES+Lvg 0.232 (+6%)	ΔCoVaR×Lvg 0.051 (+21%)
Y	N	MES×Lvg 0.067 (+86%)	eΔCoVaR×Lvg 0.018 (+125%)	ΔCoVaR 0.007 (+40%)	e,ΔCoVaR,MES×Lvg 0.228 (+14%)	ΔCoVaR×Lvg 0.045 (+25%)
N	Y	eΔCoVaR×Lvg 0.055 (+72%)	eΔCoVaR×Lvg 0.010 (+43%)			
Y	Y	MES×Lvg 0.089 (+89%)	MES×Lvg 0.013 (+160%)			

Panel C: Predicting bank failures (AUC)							
		Market-based vs. CLV			Market-based + CLV		
	Incl.	fail in x years			fail in x years		
Ep. FE	nonbanks	x = 1	x = 3	x = 5	x = 1	x = 3	x = 5
N	N	SRISK/ME 0.9096 (+8%)	CLV 0.8098	CLV 0.7710	ΔCoVaR×Lvg 0.9275 (+10%)	ΔCoVaR×Lvg 0.8802 (+9%)	e,ΔCoVaR,MES×Lvg 0.8350 (+8%)

Table 6: This table compares the explanatory power (measured by R^2 or AUC) of market-based and balance sheet indicators. R^2 is the adjusted R-squared for regressions without episode fixed effects and the within R-squared for regressions with fixed effects. AUC is the area under the receiver operating characteristic (ROC) curve. Each cell reports the specification that yields the largest percentage gain in explanatory power (R-squared in Panels A and B; AUC in Panel C) relative to the baseline model based on the balance sheet variables from Correia et al. (2024), denoted "CLV." The cell displays the name of the best-performing specification, followed by the maximum R-squared or AUC and the percentage gain in parentheses. Panel A reports the predictive performance of market-based measures relative to CLV, while Panel B presents results when both market-based and CLV variables are included in the model. Panel C evaluates the ability of these measures to predict bank failure over 1-, 3-, and 5-year horizons, comparing market-based indicators to CLV alone ("vs. CLV") and in combination with CLV ("+CLV"). Episode fixed effects (Ep. FE) and whether non-banks are included in the sample are indicated. Market-based measures include MES, ΔCoVaR, eΔCoVaR, SRISK/ME, and versions combined with leverage (Lvg). "+Lvg" indicates a model including the measure and Lvg, while "×Lvg" denotes a model including the measure, Lvg, and their interaction. e,ΔCoVaR,MES×Lvg indicates that eΔCoVaR×Lvg, ΔCoVaR×Lvg, and MES×Lvg have the same predictive performance.

Systemic Risk Measures:
From the Panic of 1907 to the Banking Stress of 2023
Online Appendix

Viral V. Acharya, Markus K. Brunnermeier, and Diane Pierret*

May 3, 2025

*Acharya: New York University Stern School of Business, CEPR, ECGI, and NBER (e-mail: vva1@stern.nyu.edu); Brunnermeier: Princeton University, CEPR, CESIfo, and NBER (e-mail: markus@princeton.edu); Pierret: University of Luxembourg and CEPR (e-mail: diane.pierret@uni.lu).

This appendix provides details about the sample construction in Section 1, the definition of stress episodes in Section 2, additional results and robustness tests in Section 3, a comparative analysis of market-based and balance sheet indicators in Section 4, the results of an alternative episode dating procedure in Section 5, and the definition and estimation of systemic risk measures in Section 6.

1 Data and Sample Description

1.1 CRSP/Compustat Sample

Our initial sample comprises daily equity data from the Center for Research in Security Prices (CRSP) Database for all U.S. financial institutions with two-digit Compustat SIC codes between 60 and 67, as in Adrian & Brunnermeier (2016). Following their methodology, we retain only ordinary common shares and exclude daily equity observations with missing or negative prices. This yields a sample of 5,417 financial institutions with unique identifying PERMCO codes in CRSP, covering the period from December 31, 1925, to December 29, 2023. Market capitalization is calculated by multiplying the number of shares outstanding by the stock price. We derive logarithmic returns and aggregate the data at the weekly frequency to estimate systemic risk measures, and to derive realized returns and realized volatility.

We detail the estimation procedure for monthly systemic risk measures that capture co-movements of financial firms' returns with the returns on an index ($\triangle$CoVaR, e$\triangle$CoVaR, MES, LRMES) in Section 6. Additionally, the systemic risk measure SRISK is a function of LRMES, firm market capitalization, and total non-equity liabilities, as defined in eq. (4) of the paper. We obtain the total liabilities of financial institutions from Compustat, merging the Compustat dataset with our stock returns and systemic risk measures aggregated at the quarterly frequency, resulting in a sample of 4,452 financial firms with unique identifying PERMCO codes. While total asset data are available starting in the first quarter of 1962, total liabilities data are only available for a sample of 3,762 financial firms beginning in the first quarter of 1965 and continuing until the last quarter of 2023.

1.2 Historical Equity Data from Global Financial Data

To extend the coverage of equity prices to the period before 1959—when U.S. banks were not yet required to file Call Reports—we supplement the CRSP dataset with historical stock return data from Global Financial Data (GFD). CRSP coverage of banks is particularly limited in the early decades of the 20th century, as many bank stocks were traded over the counter and are therefore missing from the database. Among the 85 financial institutions in CRSP before 1959, only 29 are identified as banks based on their SIC codes, and 13 of these were delisted before 1930.

The GFD dataset offers broader coverage of the financial sector from 1895 to 1958, with a temporary gap from August to November 1914, when markets were closed due to World War I. We collect stock price and market capitalization data for all 5,360 financial institutions available in GFD from 1791 to 1958. The sample is restricted to common stocks in the finance sector and to firms incorporated in the U.S. Our dataset includes 3,221 banks, 1,183 trust and thrift companies, 806 insurance companies, and 2,258 other non-bank financial institutions, such as consumer finance, corporate finance, and general finance companies. Listings span major U.S. exchanges, but the majority of securities (4,014) are traded over the counter. The stock prices are split-adjusted closing prices, though the frequency of available data varies over time (daily, weekly, or monthly).

We collapse all return series to monthly frequency, excluding securities available only at quarterly or annual intervals. We estimate systemic risk measures starting in 1895, when the sample coverage becomes sufficiently stable to construct a reliable index of financial institution stock returns. To construct this index, we compute a market-cap-weighted average based on the most liquid stocks, identified using the fraction of zero-return days. At each point in time, index constituents must have at least 90% non-zero return observations. While GFD data extend beyond 1959, we rely on CRSP from that point onward because only CRSP identifiers can be dynamically linked to regulatory data from Call Reports via the NY Fed's CRSP-FRB linking table, as detailed in the next section.

1.3 Call Reports and FDIC Bank Failures & Assistance Data

In addition to Compustat data, we use balance sheet and income statement variables from the Federal Financial Institutions Examination Council (FFIEC) Consolidated Reports of Condition and Income ("Call Reports"). Call Report data are available for the sample of commercial banks, collected from 1959 onwards by Correia, Luck & Verner (2024). Merging Call Reports with the CRSP dataset of stock data and systemic risk measures involves two steps: (1) assigning regulatory identification numbers (RSSD ID) to banks reporting in CRSP, and (2) reconstructing the structure of bank ownership to aggregate the Call Reports at the parent bank level corresponding to the institution reporting in CRSP.

In the first step, we use the NY Fed CRSP-FRB Linking Table.[1] This table dynamically links the permanent company number (PERMCO) used in CRSP to the unique regulatory identification numbers used in Call Reports (RSSD ID), including 1,471 PERMCO-RSSD links from June 30, 1986, to September 29, 2023. For the period from the first quarter of 1959 to the first quarter of 1986, we extend the links of banks present in the CRSP dataset (after June 1986) backward in time for our list of PERMCOs. For the 384 PERMCOs that appear only before June 1986 in the CRSP dataset, we manually matched 159 of them with RSSD identifiers based on the bank's name and location (when available from Compustat), using the National Information Center (NIC) website (https://www.ffiec.gov/NPW).

In the second step, we identify the parent bank as the entity in CRSP with an identifying PERMCO, and reconstruct the structure of bank ownership using the FFIEC Relationships Table.[2] We link subsidiary banks to their parents after applying the following filters: we exclude non-controlled relationships; retain only relationships identified after an "initial relationship record" or a "reestablishment of a relationship"; keep only direct relationships; exclude non-equity-based relationships; exclude investments in non-banking companies; exclude unregulated relationships; and exclude relationships if the ownership or control pertains to a non-banking company. The resulting sample includes 9,109 subsidiary banks linked to 1,619 parent banks, ensuring that each subsidiary is linked to only one parent. For the 505 subsidiaries linked to more than one parent, we select the parent bank with the largest percentage of ownership or control in the subsidiary (PCT_EQUITY). However, some of the Call Reports data are not available for the full sample period. This is notably the case of the classification of deposits into (un)insured deposits, which is only available starting in the second quarter of 1982.

[1]NY Fed CRSP-FRB Linking table available at: https://www.newyorkfed.org/research/banking_research/crsp-frb

[2]FFIEC relationships table available at: https://www.ffiec.gov/npw/FinancialReport/DataDownload

Finally, we merge the sample of 9,109 banks with the list of bank failure and assistance transactions from the Federal Deposit Insurance Corporation (FDIC) using the FDIC certificate number (also available to identify institutions in the dataset of Correia, Luck & Verner, 2024).[3] We identify 177 bank failures or assistance transactions in our sample, starting in 1959. The FDIC data are lagged by one quarter to ensure they match the last Call Report data available for the bank before its failure. All data are then aggregated at the parent bank-quarter level for a sample of 1,344 consolidated banks with stock prices available in CRSP.

2 Stress Episodes

We describe our methodology for dating stress episodes in Section 2.1. The narrative analysis of recent stress episodes is presented in Section 2.2.

2.1 Methodology

Our procedure for dating stress episodes involves two steps. First, we identify broad event "windows" from multiple sources, including previous studies and a narrative analysis of stress events. We present the event windows in Table 1. Second, we search for the trough and peak values of a credit spread relative to each window to define the start and end months of the episode. For the credit spread, we use the Gilchrist & Zakrajšek (2012) corporate bond spread index, commonly referred to as the GZ spread and available monthly starting in 1973.[4] Prior to 1973, we use the difference between Moody's Seasoned Baa and Aaa Corporate Bond Yield indices. Moody's corporate bond yields are available monthly starting in January 1919 from the Federal Reserve Bank of St. Louis database FRED. For the early episodes before 1919, we use the average yield on U.S. railroad bonds with at least ten years of maturity of Macaulay (1938), available from NBER Macrohistory database (NBER series 13019A). This index is selected because it has the highest correlation (0.77) with the Moody's Baa–Aaa spread over the period when the two series overlap (1919–1934). One event window spans the years 1916 to 1921; for that episode, we use the railroad bond yield index and begin using the Moody's spread starting in 1922. The three proxies for the credit spread—GZ spread, Moody's spread, and the railroad bond yield—are presented in Figure 1. The figure also presents the annual realized volatility of financial index returns as an indicator of market stress during each episode.

For the "early" sample spanning 1895 to 1958, the years of the event window are selected from stock market crashes and banking crisis years identified for the U.S. by Reinhart & Rogoff (2009, 2011) ("RR years" hereafter). Whenever NBER recession years are consecutive to RR years, we consider the largest window by taking the union of events.[5] For example, our Post-World War II Recession episode starts with an NBER recession in 1945, is followed by RR stock market crashes from 1946-1948, and ends with another NBER recession in 1949. For each episode, we identify the start and end dates as the dates corresponding to the lowest and highest values of the Moody's spread (1922-1972) or railroad bond yield (1895-1921) within the window. If there is no trough before the peak (i.e., the peak date is the first date of the window), there is no increase in the

[3] Bank failures and assistance data available at: https://banks.data.fdic.gov/bankfind-suite/failures

[4] The GZ spread data can be accessed at: https://www.federalreserve.gov/econres/notes/feds-notes/updating-the-recession-risk-and-the-excess-bond-premium-20161006.html

[5] However, we drop isolated NBER recessions that are not directly preceded or followed by RR years. This includes the 1953-54 NBER recession for the early period, and NBER recessions in 1960 and 1970 for the modern sample.

spread within the window, and the episode is dropped. This occurs only for the NBER recession of 1927.

Identification of Event Window Years

	episode	start	end	event type
1	1903-04 stock market crash	1902	1904	1902-03 NBER recession; 1903-04 RR stock market crash.
2	Panic of 1907	1906	1908	1906-08 RR stock market crash; 1907 RR banking crisis; 1907-08 NBER recession.
3	1910-11 stock market crash	1910	1911	1910-11 RR stock market crash; 1910-11 NBER recession.
4	1914 banking crisis	1913	1914	1913-14 RR stock market crash; 1914 RR banking crisis; 1913-14 NBER recession.
5	1916-21 stock market crash	1916	1921	1916-21 RR stock market crash; 1918 NBER recession; 1920-21 NBER recession.
6	The Great Depression	1929	1933	1929-33 NBER recession; 1929-33 RR banking crisis; 1929-32 RR stock market crash.
7	Recession of 1937–38	1937	1938	1937-38 NBER recession; 1937 RR stock market crash
8	1939-41 stock market crash	1939	1941	1939-41 RR stock market crash.
9	Post-World War II Recession	1945	1949	1945 NBER recession; 1946-48 RR stock market crash; 1948-49 NBER recession.
10	Recession of 1958	1956	1958	1957-58 NBER recession; 1956-58 RR stock market crash.
11	High inflation in the U.S.	1973	1975	1973-75 NBER recession; 1973-74 RR stock market crash
12	1977-82 stock market crash	1977	1982	1977-82 RR stock market crash; 1980 NBER recession; 1981-82 NBER recession.
13	S&L crisis	1984	1991	1984-91 RR banking crisis; 1989-91 RR stock market crash; 1990-91 NBER recession.
14	LTCM hedge fund failure	1997	1998	1998 Narrative.
15	Dot.com Bubble	2000	2002	2000-02 RR stock market crash; 2001 NBER recession
16	Global Financial Crisis	2007	2010	2007-10 RR banking crisis; 2008 RR stock market crash; 2008-09 NBER recession.
17	European Sovereign Debt Crisis	2011	2011	2009-11 Narrative.
18	2014-16 oil price shock	2014	2016	2014-16 Narrative.
19	Covid19 Pandemic	2020	2020	2020 Narrative; 2020 NBER recession.
20	Ukraine war/energy crisis	2022	2022	2022 Narrative.
21	Banking Stress of 2023	2023	2023	2023 Narrative.

Table 1: Description of event windows. The table presents the event windows used to search for trough and peak values of a credit spread for the identification of start and end months of stress episodes. RR stock market crash and banking crises years are from Reinhart & Rogoff (2009). The narrative of stress episode events is available in Section 2.2.

For the "modern" sample spanning 1959 to 2023, we rely on a narrative analysis of the most recent stress episodes to identify window years in addition to RR years, since Reinhart & Rogoff (2009) database ends in 2010. The narrative of stress episodes is provided in Section 2.2, and identifies the episode of the LTCM hedge fund failure as well as four episodes after the Global Financial Crisis (GFC), in addition to five episodes based on RR years. The post-GFC narrative episodes include the European sovereign debt crisis in 2011, the 2014-16 oil price shock, the Covid19 pandemic in 2020, the Russian-Ukrainian conflict in 2022, and the failure of regional U.S. banks in 2023. The identification of "windows" based on the narrative comes from the years mentioned in the narrative. We search for a peak value of the GZ spread within the narrative window, and a trough value before the peak date, allowing the trough date to be located slightly (i.e., a maximum of six months) before the start of the narrative window to remain conservative in our predictive regressions. Some narrative windows may overlap with the other windows. For example, the European Sovereign Debt Crisis episode starts when the GFC episode is not over yet. In this case, we start the window after the GFC years in Reinhart & Rogoff (2009).

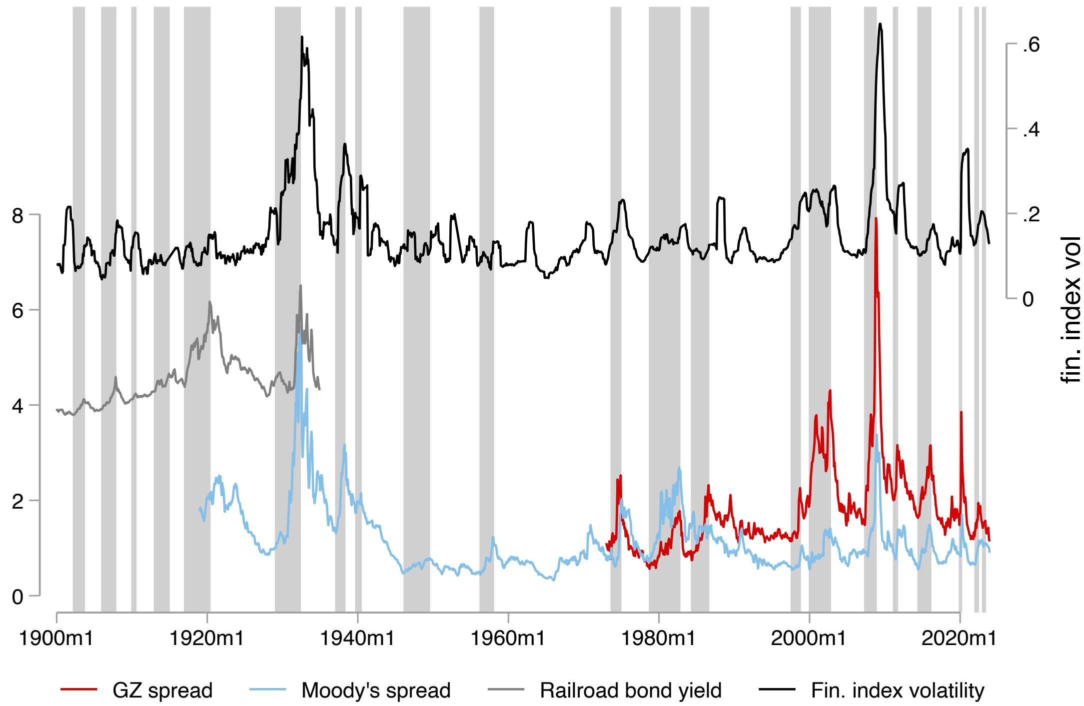

Figure 1: Stress episode dating methodology. "GZ spread" is the GZ spread (Gilchrist & Zakrajšek, 2012) in percentage points. "Moody's spread" is the spread between Moody's Baa and Aaa Corporate Bond Yields in percentage points, and "Railroad bond yield" is the average U.S. railroad bond yield (Macaulay, 1938) in percentage points. The figure also presents the annual realized volatility of financial index returns ("Fin. index volatility") as an indicator of market stress during each episode. Grey vertical bars are the stress episode dates starting with the month of the trough value and ending with the month of the peak value of a credit spread (i.e., GZ spread for 1973-2023, Moody's spread for 1922-1972, and railroad bond yield before 1922).

2.2 Narrative of Stress Episodes

The narrative dating of stress episodes for the modern sample is based on a variety of formal and informal sources, including Markham (2022), and is reproduced in this section.

1970s: High inflation in the U.S.

- Aug 15, 1971: President Nixon announced the suspension of the convertibility of the U.S. dollar into gold, effectively ending the Bretton Woods system.

- 1971-1973: U.S. dollar experienced devaluation.

- Aug 15 - Nov 13, 1971: Nixon imposed a 90-day freeze on wages and prices.
- Nov 14, 1971: Nixon introduced second phase of the economic stabilization program.
- Dec 18, 1971: Smithsonian Agreement introduced limited flexibility in currency exchange rates without fully reinstating the prior fixed system.
- Apr 30, 1973: Nixon formally ended the second phase of wage and price control.
- Oct 17, 1973: OPEC imposed an oil embargo on countries supporting Israel in the Yom Kippur War.
- Jan 1974: Oil production cuts quadrupled the oil price from $2.9 a barrel to $11.65.
- Mar 18, 1974: Most OPEC nations end a 5-month oil embargo against the U.S.
- Aug 8, 1974: President Nixon announces his resignation.
- Dec 6, 1974: The U.S. experienced a severe market crash, with DJIA reaching its lowest point.
- Apr 1, 1979: The reformed Iranian government nationalized its oil industry.
- Jul 25 1979: Paul Volcker became the Chairman of the Fed, adopting tight monetary policies.
- Nov 4, 1979: Iranian militants stormed the U.S. embassy, initiating Iran Hostage Crisis.
- Late 1979: Oil prices surged during this period, reaching historically high levels.
- Jan 1980: Inflation (CPI) reached the peak of around 14%.

1980s: Savings and Loan Crisis, Real Estate Crash & Stock Market Crash in 1987

- 1982: Existing home sales fell nearly 50% from the peak in 1978.
- July 5, 1982: Collapse of Penn Square Bank.
- Oct 15, 1982: The Garn-St.Germain Depository Institutions Act was signed into law, relaxing regulations on savings and loans.
- Sep 19, 1984: Bailout of Continental Illinois National Bank.
- Oct 22, 1986: The Tax Reform Act was signed into law, eliminating tax incentives for real estate.
- Aug 25, 1987: The Dow Jones Industrial Average (DJIA) reached its peak at 2,722.42 points.
- Oct 16, 1987: The stock market experienced a notable decline on Friday.
- Oct 19, 1987: The U.S. stock market experiences a historic crash, with the DJIA dropping 22.6%.
- Oct 20, 1987: Global stock markets experienced significant declines in the wake of Black Monday.

- Aug 2, 1988: First Republic Bank Corporation, based in Texas, collapsed.
- Sep 1988: American Savings and Loan, a major S&L based in California, failed.
- Dec 1988: Gibraltar Savings failed.
- Apr 14, 1989: Lincoln Savings and Loan collapsed in 1989 due to risky and fraudulent activities.
- Feb 2, 1990: CenTrust Bank, an S&L association based in Miami, Florida failed.
- Sep 1, 1990: Citi's stock dropped to below $20.
- Oct 1, 1990: Citi's stock price reached its bottom, around $15.
- Nov 9, 1990: Chase's stock was traded at single-digit of around $3, with rumors that they might fail.

Sep 1998: LTCM hedge fund failure

- Aug 17, 1998: Russia defaulted on its debt obligations. LTCM experienced a huge loss as it had substantial exposure to Russian government bond.
- Sep 23, 1998: LTCM sought a bailout from major Wall Street investment banks to prevent its collapse.
- Sep 23, 1998: On the same day, Alan Greenspan, Chair of the Fed, facilitated a meeting among the banks to arrange a rescue package.
- Late Sep, 1998: A consortium of major financial institutions, including Goldman Sachs, Merrill Lynch, and J.P. Morgan, agrees to inject capital into LTCM to stabilize its positions.
- Oct 1998: LTCM's operations winded down as the fund sells off its remaining positions. o Early 2000: LTCM officially closed.

2000-2002: Dot.com Bubble

- 1995-2000: There was a surge in investments in internet-related companies.
- Mar 20, 2000: NASDAQ reached its all-time high.
- Mar 24, 2000: Dot.com bubble started to burst.
- Mar 2000 - Oct 2002: Stock prices of many internet companies continued to decline.
- Oct 9, 2002: NASDAQ hit bottom.

Global Financial Crisis

- Feb 7, 2007: HSBC announced significant losses in its subprime mortgage division.
- Aug 9, 2007: European Central Bank injected €95bn into the banking system.
- Aug 14, 2007: Fed injected $38bn into the banking system.
- Mar 16, 2008: Bear Stearns was sold to JP Morgan Chase for a fraction of its previous value.
- Jul 11, 2008: IndyMac Bancorp was seized by regulators.
- Sep 7, 2008: U.S. Government seized control of Fannie Mae and Freddie Mac.
- Sep 15, 2008: Lehman Brothers filed for bankruptcy.
- Sep 16, 2008: AIG was bailed out by the U.S. government.
- Oct 29, 2008: The DJIA experienced its largest single-day drop.
- Mar 6, 2009: The DJ hit its lowest point, closing at around 6547 points.

European Sovereign Debt Crisis

- Apr 27, 2009: Greece revealed that its budget deficit is much higher than previously reported.
- Feb 11, 2010: Greece announced austerity measures to address its debt crisis, triggering social unrest.
- Apr 23, 2010: Greece formally requested financial assistance from EU and IMF.
- May 2, 2010: EU and IMF agreed on a €110bn bailout package for Greece.
- May 6, 2010: The DJIA plunged nearly 1000 points in a matter of minutes (flash crash).
- Nov 28, 2010: Ireland requested a €67.5bn bailout package from EU and IMF.
- Aug 4, 2011: S&P downgraded the credit rating of the U.S. from AAA to AA+.
- Aug 8, 2011: European markets experienced significant losses.

Oil Price Shock in 2014-2016

- Mid-2014: Oil prices began declining from $100+ per barrel.
- Nov 27, 2014: OPEC decided not to cut oil production.
- Jun 5, 2015: U.S. crude oil inventories hit their highest level in over 80 years.
- Dec 4, 2015: OPEC maintained its decision to keep production levels unchanged.
- Jan 20, 2016: Oil prices briefly dropped below $30 per barrel.
- Feb 16, 2016: Saudi Arabia, Qatar, and Russia agreed to freeze oil production.

- Apr 17, 2016: OPEC and non-OPEC producers failed to reach an agreement on production freeze.
- Nov 30, 2016: OPEC finalized a deal to cut production by 1.2mn barrels per day.
- Dec 2016: Oil prices began to recover, with crude rising above $50 per barrel.

Repo Market Spike during COVID

- Mar 9, 2020: Stock market plunged worldwide.
- Mar 11, 2020: Repo spread in overnight market spiked.
- Mar 12, 2020: Fed announced plans to inject $1.5tn into the financial system.
- Mar 16, 2020: Fed announced relaunch of QE with $700bn in asset purchases.

Russia-Ukraine Conflict

- Feb 24, 2022: Russia invades Ukraine, leading to sharp declines in global stock markets due to increased risk aversion. Oil prices surge past $100 per barrel for the first time since 2014.
- Late February 2022: Western nations impose severe economic sanctions on Russia, including removing select Russian banks from the SWIFT payment system. The Russian ruble plummets to a record low, and the Moscow Stock Exchange suspends trading.
- Mar 8, 2022: The U.S. bans imports of Russian oil and gas; Brent crude oil prices spike to nearly $130 per barrel, exacerbating global inflation concerns.
- Mar 2022: Prices of wheat, corn, and other commodities surge as Ukraine and Russia account for a significant share of global grain exports, impacting food prices worldwide.
- Jun 15, 2022: In response to soaring inflation partly driven by the conflict, central banks including the U.S. Federal Reserve implement significant interest rate increases. The Fed raises its benchmark rate by 75 basis points, the largest hike since 1994.
- Sep 5, 2022: Russia suspends gas deliveries through the Nord Stream 1 pipeline indefinitely, citing maintenance issues. European natural gas prices hit record highs, deepening the energy crisis and impacting European economies and currencies.

Bank Failures in 2023

- Mar 10, 2023: Silicon Valley Bank closed.
- Mar 12, 2023: Signature Bank closed.
- May 1, 2023: First Republic Bank was acquired by JP Morgan Chase Bank.
- Jul 28, 2023: Heartland Tri-State Bank closed.
- Nov 3, 2023: Citizens Bank at Sac City, IA, failed.

3 Additional Results and Robustness Tests

In Section 3.1, we provide summary statistics for the systemic risk measures described in the paper. In Section 3.2, we assess the sensitivity of our results in Table 3 of the paper to alternative definitions of crisis severity, using the cumulative return over the stress episode for a stock index, a financial index, and a bank index. Section 3.3 reports the results of Table 2 and Table 3 separately for the early sample based on GFD data and the modern sample based on CRSP data. We also report results for the GFD sample restricted to more liquid stocks—defined as securities with fewer than 50% zero-return days in the past—to test the robustness of our early sample findings to potential liquidity issues. Finally, in Section 3.4, we decompose the lending outcome variable from Table 5 of the paper into commercial and industrial (C&I) loans and real estate loans.

3.1 Descriptive Statistics

In Table 2, we report unweighted average systemic risk measures separately for the early and modern samples in Panel A, and for banks and non-banks in Panel B. We compare these measures at three points in time: three years before the stress episode, one year before, and during the episode. While all systemic risk measures tend to peak three years before a crisis, they generally remain at similar levels one year before the episode and decline during the stress period.

We also note that the average values of the measures are lower in the early sample, where we rely on GFD data. This likely reflects stale prices in that dataset, which lead to lower measured volatility. To test this, we report descriptive statistics for the GFD sample restricted to the most liquid stocks—defined as securities with fewer than 50% zero-return days in the past. All systemic risk measures are higher when the GFD sample is limited to these more liquid securities. We assess the robustness of our main results to conditioning the GFD sample on liquid stocks in Section 3.3 of this Appendix and find that the results remain qualitatively unchanged. Additionally, we report in Panel A of Table 2 the number of financial institutions and banks in each sample. The number of institutions typically increases between three and one years before an episode and reaches its maximum one year prior to the start of stress.

Panel B compares banks and non-bank financial institutions (non-banks). Measures capturing co-movement with the market index—ΔCoVaR, eΔCoVaR, and MES—are all higher for non-banks. In contrast, measures that incorporate leverage, such as SRISK/ME, are higher for banks. Similarly, market leverage points to greater undercapitalization among banks relative to non-banks.

Panel A: early vs. modern sample

	GFD sample (1895-1958)						CRSP sample (1959-2023)		
	all stocks			most liquid stocks			all stocks		
	3y before	1y before	stress	3y before	1y before	stress	3y before	1y before	stress
ΔCoVaR	0.45	0.44	0.42	0.92	0.93	0.92	1.81	1.75	1.56
eΔCoVaR	0.86	0.87	0.81	1.95	2.04	1.83	2.68	2.63	2.42
MES	1.71	1.65	1.65	4.16	4.18	4.00	7.37	7.31	6.88
SRISK/ME							20.94	12.16	10.04
Lvg							10.66	9.57	9.64
#fin. institutions	3,649	3,885	3,825	1,137	1,269	1,213	3,214	3,611	3,433
#banks	2,336	2,460	2,426	603	667	628	887	1,012	927

Panel B: banks vs. non-banks

	banks			non-banks		
	3y before	1y before	stress	3y before	1y before	stress
ΔCoVaR	0.81	0.80	0.64	1.48	1.43	1.21
eΔCoVaR	1.20	1.20	0.99	2.30	2.27	2.00
MES	2.46	2.47	2.11	5.64	5.64	5.07
SRISK/ME	26.80	18.96	17.10	17.27	8.05	5.81
Lvg	12.26	11.21	11.43	9.66	8.58	8.56

Table 2: Descriptive statistics. The table presents unweighted average systemic risk measures for months respectively three years and one year before the start date of a stress episode (resp. "3y before" and "1y before"), and during the stress episode ("stress"). Panel A presents descriptive statistics for the early sample (1895-1958) based on GFD data and the modern sample (1959-2023) based on CRSP data, separately. In addition, Panel A reports descriptive statistics for the GFD sample restricted to the most liquid stocks—defined as securities with fewer than 50% zero-return days in the past. Panel B presents descriptive statistics for banks and non-banks separately. SRISK/ME is SRISK scaled by the market capitalization, Lvg is the market leverage. We also report the number of financial institutions (#fin. institutions) and banks (#banks).

3.2 Market Outcomes during Stress Episodes: Drawdown Definitions

episode	credit	stock index		fin. index		bank index	
	sprd	ret (%)	dd (%)	ret (%)	dd (%)	ret (%)	dd (%)
1903-04 stock market crash	0.33	-31.65	-31.65	-36.65	-46.38	-16.31	-24.44
1907 banking crisis	0.68	-42.88	-42.88	-30.49	-30.49	-19.59	-21.90
1910-11 stock market crash	0.12	-15.60	-15.60	7.88	-6.00	-8.77	-12.24
1914 banking crisis	0.38	-34.81	-34.81	-18.56	-23.99	-52.85	-59.96
1916-21 stock market crash	1.79	-3.79	-23.33	3.92	-30.29	9.77	-81.95
The Great Depression	4.64	-85.90	-88.24	-91.95	-95.05	-92.06	-92.06
Recession of 1937–38	1.86	-40.34	-47.13	-16.30	-30.27	-37.72	-53.82
1939-41 stock market crash	0.42	-19.63	-23.48	-15.13	-22.68	-7.99	-17.24
Post-World War II Recession	0.32	-14.05	-21.42	-5.07	-19.40	-16.87	-25.92
Recession of 1958	0.79	-12.81	-15.86	-22.86	-30.29	-47.54	-52.45
High inflation in the U.S.	1.75	-36.77	-41.40	-35.73	-46.82	-45.65	-53.52
1977-82 stock market crash	1.21	43.54	-23.79	43.44	-34.12	42.27	0
S&L Crisis	1.58	68.00	0	54.24	-11.14	71.21	-9.64
LTCM hedge fund failure	1.12	22.15	-15.57	-10.08	-36.40	-3.60	-36.76
Dot.com Bubble	2.36	-36.48	-46.28	-24.11	-36.84	-20.72	-32.77
Global Financial Crisis	6.44	-41.45	-42.16	-52.09	-52.09	-49.47	-49.47
European Sovereign Debt Crisis	1.16	-14.66	-17.03	-18.99	-20.15	-26.05	-27.90
2014-16 oil price shock	1.52	-1.43	-8.89	0.90	-9.81	-24.07	-24.79
Covid19 Pandemic	2.37	-20.00	-20.00	-18.47	-18.47	-36.07	-36.07
Ukraine war/energy crisis	0.70	-20.58	-20.58	-8.77	-8.77	-28.37	-28.37
Bank Failures in 2023	0.15	2.53	0	-2.96	-5.97	-2.83	-8.76

Table 3: Stress episodes: drawdown definitions. The credit spread is proxied by the GZ spread (Gilchrist & Zakrajšek, 2012) for the 1973-2023 period, the spread between Moody's Baa and Aaa Corporate Bond Yields for 1922-1972, and the average yield on U.S. railroad bonds (Macaulay, 1938) for 1895-1921. "stock index" is the S&P500 index for 1959-2023, and the Dow Jones Industrial Average index before 1959. "fin. index" is the CRSP financial index for 1959-2023, and the GFD financial index before 1959. "bank index" is the CRSP market-cap weighted average bank stocks index for 1959-2023, and the GFD market-cap weighted average bank stocks index before 1959. "ret" denotes the index return in percentage points over the episode. "dd" is the maximum drawdown on an index, defined as the percentage difference between its minimum and the prior maximum value within an episode. There is no drawdown when the minimum index value is reached at the beginning of the episode.

Panel A: Realized volatility during stress episodes									
Measure:		ΔCoVaR			eΔCoVaR			MES	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Measure	-0.36**	-0.42***	-0.45**	0.46***	0.36***	0.40***	0.38***	0.37***	0.38***
	(-2.30)	(-2.95)	(-2.27)	(6.01)	(4.56)	(5.05)	(8.90)	(8.68)	(7.97)
Measure×ind_dd	3.04***			0.25			0.39***		
	(6.15)			(1.07)			(3.71)		
Measure×fin_dd		5.17***			1.01***			0.40***	
		(7.23)			(3.22)			(3.39)	
Measure×bank_dd			3.01***			0.51**			0.21**
			(4.92)			(2.26)			(2.10)
Observations	11,636	11,636	11,636	11,636	11,636	11,636	11,636	11,636	11,636
Within R-squared	0.022	0.024	0.021	0.034	0.036	0.035	0.057	0.056	0.055
Episode FE	Y	Y	Y	Y	Y	Y	Y	Y	Y

Measure:		SRISK/ME			Lvg			MES (+Lvg)	
	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Measure	0.51*	0.37	0.63*	0.03*	0.03	0.03	0.46***	0.43***	0.44***
	(1.82)	(1.40)	(1.69)	(1.66)	(1.30)	(1.29)	(8.35)	(7.32)	(6.47)
Measure×ind_dd	-0.19			-0.01			0.53***		
	(-0.19)			(-0.19)			(4.41)		
Measure×fin_dd		0.76			0.04			0.54***	
		(0.67)			(0.48)			(3.46)	
Measure×bank_dd			-0.52			-0.01			0.29*
			(-0.49)			(-0.16)			(1.93)
Lvg							0.03	0.02	0.03
							(1.38)	(1.02)	(1.12)
Lvg×ind_dd							-0.03		
							(-0.39)		
Lvg×fin_dd								0.02	
								(0.27)	
Lvg×bank_dd									-0.02
									(-0.31)
Observations	6,316	6,316	6,316	6,316	6,316	6,316	6,316	6,316	6,316
Within R-squared	0.029	0.030	0.030	0.026	0.027	0.026	0.071	0.070	0.069
Episode FE	Y	Y	Y	Y	Y	Y	Y	Y	Y

Panel B: Realized returns during stress episodes									
Measure:		ΔCoVaR			eΔCoVaR			MES	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Measure	-2.81***	-3.34***	-1.65	-1.38***	-1.67***	-1.06**	-0.77***	-0.82***	-0.53**
	(-3.41)	(-3.85)	(-1.55)	(-3.69)	(-3.64)	(-2.36)	(-4.14)	(-3.78)	(-2.04)
Measure×ind_dd	-5.91			-0.42			-1.50**		
	(-1.54)			(-0.28)			(-2.21)		
Measure×fin_dd		-5.07			1.49			-1.06	
		(-1.16)			(0.70)			(-1.31)	
Measure×bank_dd			-10.18**			-1.83			-2.07***
			(-2.47)			(-1.09)			(-2.71)
Observations	11,636	11,636	11,636	11,636	11,636	11,636	11,636	11,636	11,636
Within R-squared	0.007	0.007	0.008	0.006	0.007	0.007	0.009	0.009	0.010
Episode FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Measure:		SRISK/ME			Lvg			MES (+Lvg)	
	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Measure	1.01	1.14	1.78	0.10*	0.11*	0.15	-0.54**	-0.54*	-0.20
	(1.42)	(1.57)	(1.47)	(1.71)	(1.79)	(1.53)	(-2.48)	(-1.85)	(-0.53)
Measure×ind_dd	-6.60*			-0.59*			0.45		
	(-1.67)			(-1.81)			(-0.52)		
Measure×fin_dd		-8.39*			-0.71*			-0.28	
		(-1.78)			(-1.87)			(-0.21)	
Measure×bank_dd			-6.24			-0.49			-2.07
			(-1.52)			(-1.51)			(-1.64)
Lvg							0.10*	0.11*	0.17*
							(1.77)	(1.84)	(1.70)
Lvg×ind_dd							-0.57*		
							(-1.75)		
Lvg×fin_dd								-0.68*	
								(-1.81)	
Lvg×bank_dd									-0.52
									(-1.62)
Observations	6,316	6,316	6,316	6,316	6,316	6,316	6,316	6,316	6,316
Within R-squared	0.004	0.004	0.004	0.004	0.004	0.004	0.007	0.007	0.008
Episode FE	Y	Y	Y	Y	Y	Y	Y	Y	Y

Table 4: Predictive regression of realized volatility (Panel A) and realized returns (Panel B) during stress episodes (1895-2023). The dependent variable is the market outcome of a financial institution during stress episodes. The systemic risk measures and the control variable for the firm size are all measured the quarter before the episode starts. The drawdown variable "ind_dd" is the loss on the stock index, "fin_dd" is the loss on the financial index, and "bank_dd" is the loss on the bank index over the episode (corresponding to -ret in Table 3). Coefficient estimates on SRISK/ME and Lvg are multiplied by 100. t-statistics based on standard errors clustered at the firm level are reported in parentheses. Sample (Columns (1)-(9)): 4,054 financial institutions, including 1,953 banks. Sample (Columns (10)-(18)): 2,179 financial institutions, including 758 banks.

3.3 Market Outcomes during Stress Episodes: Early vs. Modern Episodes

	Panel A: Realized volatility during early stress episodes					
Measure:	ΔCoVaR		eΔCoVaR		MES	
	(1)	(2)	(3)	(4)	(5)	(6)
Measure×bank	0.62***	-0.03	0.24***	0.10	0.10***	0.16***
	(7.26)	(-0.22)	(3.31)	(1.36)	(2.77)	(4.08)
Measure×nonbank	1.01***	0.73***	0.38***	0.33***	0.30***	0.37***
	(5.95)	(3.87)	(6.19)	(4.63)	(7.54)	(10.89)
Observations	5,320	5,320	5,320	5,320	5,320	5,320
Adj. R-squared	0.049	0.158	0.046	0.161	0.038	0.174
Episode FE	N	Y	N	Y	N	Y
	Panel B: Realized volatility (liquid stocks), early stress episodes					
Measure:	ΔCoVaR		eΔCoVaR		MES	
	(1)	(2)	(3)	(4)	(5)	(6)
Measure×bank	0.91***	0.01	0.57***	0.25**	0.29***	0.24***
	(8.18)	(0.02)	(7.32)	(2.31)	(5.12)	(3.92)
Measure×nonbank	1.01***	0.69**	0.40***	0.35***	0.30***	0.39***
	(4.69)	(2.48)	(5.26)	(3.53)	(5.10)	(10.07)
Observations	2,275	2,275	2,275	2,275	2,275	2,275
Adj. R-squared	0.090	0.203	0.097	0.210	0.070	0.230
Episode FE	N	Y	N	Y	N	Y
	Panel C: Realized volatility during modern stress episodes					
Measure:	ΔCoVaR		eΔCoVaR		MES	
	(1)	(2)	(3)	(4)	(5)	(6)
Measure×bank	-0.17	0.72***	0.55***	0.71***	0.16*	0.45***
	(-0.74)	(3.33)	(3.25)	(4.46)	(1.79)	(5.54)
Measure×nonbank	-1.02***	-0.29	0.88***	0.95***	0.37***	0.51***
	(-6.08)	(-1.46)	(6.52)	(7.60)	(6.13)	(8.25)
Observations	6,316	6,316	6,316	6,316	6,316	6,316
Adj. R-squared	0.015	0.228	0.029	0.247	0.027	0.259
Episode FE	N	Y	N	Y	N	Y

Table 5: Predictive regression of realized volatility during early stress episodes (1895-1958) in Panels A and B, and modern stress episodes (1959-2023) in Panel C. Panel B restricts the GFD sample to the most liquid stocks—defined as securities with fewer than 50% zero-return days in the past.The dependent variable is the realized volatility of a financial institution during stress episodes. The systemic risk measures and the control variable for the firm size are all measured the quarter before the episode starts. t-statistics based on standard errors clustered at the firm level are reported in parentheses. Sample (Panel A): 1,875 financial institutions, including 1,195 banks. Sample (Panel B): 749 financial institutions, including 407 banks. Sample (Panel C): 2,186 financial institutions, including 761 banks.

Panel A: Realized returns during early stress episodes						
Measure:	ΔCoVaR		eΔCoVaR		MES	
	(1)	(2)	(3)	(4)	(5)	(6)
Measure×bank	-4.26***	-5.80***	-0.44	-0.59	-1.12***	-1.25***
	(-7.83)	(-4.46)	(-0.46)	(-0.61)	(-3.58)	(-4.27)
Measure×nonbank	-2.09***	-5.84***	-1.40***	-2.02***	-1.15***	-1.85***
	(-4.68)	(-5.41)	(-4.43)	(-6.33)	(-5.81)	(-8.68)
Observations	5,320	5,320	5,320	5,320	5,320	5,320
Adj. R-squared	0.021	0.073	0.018	0.068	0.020	0.072
Episode FE	N	Y	N	Y	N	Y
Panel B: Realized returns (liquid stocks), early stress episodes						
Measure:	ΔCoVaR		eΔCoVaR		MES	
	(1)	(2)	(3)	(4)	(5)	(6)
Measure×bank	-3.89***	-3.60***	-2.17***	-2.08***	-1.08***	-1.20***
	(-7.08)	(-2.69)	(-7.28)	(-7.12)	(-4.81)	(-5.37)
Measure×nonbank	-1.36***	-3.24***	-1.04***	-1.64***	-0.77***	-1.36***
	(-2.77)	(-2.70)	(-4.46)	(-5.46)	(-3.34)	(-7.31)
Observations	2,275	2,275	2,275	2,275	2,275	2,275
Adj. R-squared	0.045	0.214	0.052	0.225	0.040	0.226
Episode FE	N	Y	N	Y	N	Y
Panel C: Realized returns during modern stress episodes						
Measure:	ΔCoVaR		eΔCoVaR		MES	
	(1)	(2)	(3)	(4)	(5)	(6)
Measure×bank	-4.12***	-4.64***	-1.84***	-2.22***	-0.57**	-1.18***
	(-5.26)	(-5.91)	(-3.66)	(-4.93)	(-2.39)	(-5.43)
Measure×nonbank	0.86	0.36	0.39	-0.75*	0.72**	-0.36
	(1.05)	(0.44)	(-3.66)	(-4.93)	(2.54)	(-1.59)
Observations	6,316	6,316	6,316	6,316	6,316	6,316
Adj. R-squared	0.004	0.246	0.004	0.246	0.005	0.247
Episode FE	N	Y	N	Y	N	Y

Table 6: Predictive regression of realized returns during early stress episodes (1895-1958) in Panels A and B, and modern stress episodes (1959-2023) in Panel C. Panel B restricts the GFD sample to the most liquid stocks—defined as securities with fewer than 50% zero-return days in the past. The dependent variable is the realized returns of a financial institution during stress episodes. The systemic risk measures and the control variable for the firm size are all measured the quarter before the episode starts. t-statistics based on standard errors clustered at the firm level are reported in parentheses. Sample (Panel A): 1,875 financial institutions, including 1,195 banks. Sample (Panel B): 749 financial institutions, including 407 banks. Sample (Panel C): 2,186 financial institutions, including 761 banks.

	Panel A: Realized volatility during early stress episodes								
Measure:		ΔCoVaR			eΔCoVaR			MES	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Measure	0.38***	0.23	0.36***	0.23***	0.12*	0.22***	0.27***	0.16***	0.24***
	(2.75)	(0.98)	(2.60)	(3.86)	(1.89)	(3.72)	(9.78)	(4.64)	(8.90)
Measure×dd		0.16			0.11*			0.11***	
		(1.09)			(1.89)			(3.13)	
Measure×bank_crisis			4.83***			1.03*			0.46**
			(3.46)			(1.87)			(2.47)
Observations	5,320	5,320	5,320	5,320	5,320	5,320	5,320	5,320	5,320
Within R-squared	0.007	0.008	0.011	0.015	0.016	0.017	0.028	0.035	0.033
Episode FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Panel B: Realized volatility (liquid stocks), early stress episodes								
Measure:		ΔCoVaR			eΔCoVaR			MES	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Measure	0.44*	0.34	0.43*	0.33***	0.19*	0.32***	0.35***	0.20***	0.31***
	(1.89)	(0.87)	(1.86)	(3.52)	(1.66)	(3.38)	(8.98)	(4.31)	(8.26)
Measure×dd		0.10			0.13			0.14***	
		(0.46)			(1.39)			(2.95)	
Measure×bank_crisis			0.21			0.50			0.55**
			(0.15)			(0.80)			(2.23)
Observations	2,275	2,275	2,275	2,275	2,275	2,275	2,275	2,275	2,275
Within R-squared	0.006	0.007	0.006	0.023	0.026	0.024	0.046	0.058	0.055
Episode FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
	Panel C: Realized volatility during modern stress episodes								
Measure:		ΔCoVaR			eΔCoVaR			MES	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Measure	0.03	-1.08***	-0.19	0.88***	0.72***	0.83***	0.50***	0.42***	0.47***
	(0.16)	(-4.30)	(-1.08)	(8.61)	(4.38)	(7.65)	(9.63)	(5.39)	(8.60)
Measure×dd		0.64***			0.08*			0.04**	
		(6.44)			(1.66)			(2.01)	
Measure×bank_crisis			2.84***			0.47**			0.36***
			(6.10)			(2.09)			(3.57)
Observations	6,316	6,316	6,316	6,316	6,316	6,316	6,316	6,316	6,316
Within R-squared	0.021	0.032	0.028	0.047	0.048	0.048	0.064	0.065	0.067
Episode FE	Y	Y	Y	Y	Y	Y	Y	Y	Y

Table 7: Predictive regression of realized volatility during early stress episodes (1895-1958) in Panels A and B, and modern stress episodes (1959-2023) in Panel C. Panel B restricts the GFD sample to the most liquid stocks—defined as securities with fewer than 50% zero-return days in the past. The dependent variable is the realized volatility of a financial institution during stress episodes. The systemic risk measures and the control variable for the firm size are all measured the quarter before the episode starts. The drawdown variable "dd" is the change in the credit spread over the episode. The credit spread is the GZ spread for the period 1973-2023, the Moody's spread for 1922-1972, and the average yield on U.S. railroad bonds for 1895-1921. The variable "bank_crisis" indicates the Great Depression and the GFC. t-statistics based on standard errors clustered at the firm level are reported in parentheses. Sample (Panel A): 1,875 financial institutions, including 1,195 banks. Sample (Panel B): 749 financial institutions, including 407 banks. Sample (Panel C): 2,186 financial institutions, including 761 banks.

Panel A: Realized returns during early stress episodes									
Measure:		ΔCoVaR			eΔCoVaR			MES	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Measure	-5.82***	-5.09***	-5.74***	-1.44***	-1.86***	-1.42***	-1.57***	-1.30***	-1.49***
	(-5.07)	(-4.46)	(-4.99)	(-3.27)	(-3.65)	(-3.22)	(-7.54)	(-6.08)	(-7.07)
Measure×dd		-0.74			0.41			-0.27	
		(-0.49)			(0.86)			(-1.53)	
Measure×bank_crisis			-16.99*			-1.81			-1.12
			(-1.85)			(-0.75)			(-1.35)
Observations	5,320	5,320	5,320	5,320	5,320	5,320	5,320	5,320	5,320
Within R-squared	0.021	0.021	0.021	0.013	0.014	0.013	0.020	0.020	0.020
Episode FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Panel B: Realized returns (liquid stocks), early stress episodes									
Measure:		ΔCoVaR			eΔCoVaR			MES	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Measure	-3.35***	-4.01*	-3.44***	-1.74***	-1.10**	-1.70***	-1.29***	-0.93***	-1.18***
	(-2.74)	(-1.84)	(-2.78)	(-6.74)	(-2.45)	(-6.56)	(-8.21)	(-4.71)	(-7.72)
Measure×dd		0.66			-0.62*			-0.36**	
		(0.57)			(-1.91)			(-2.04)	
Measure×bank_crisis			5.28			-1.74			-1.54*
			(0.83)			(-0.75)			(-1.75)
Observations	2,275	2,275	2,275	2,275	2,275	2,275	2,275	2,275	2,275
Within R-squared	0.020	0.020	0.020	0.033	0.036	0.034	0.034	0.037	0.037
Episode FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Panel C: Realized returns during modern stress episodes									
Measure:		ΔCoVaR			eΔCoVaR			MES	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Measure	-1.31**	-1.35*	-1.16*	-1.19***	-1.50***	-1.32***	-0.58***	-0.55**	-0.59***
	(-1.98)	(-1.70)	(-1.72)	(-3.33)	(-3.34)	(-3.49)	(-3.08)	(-2.37)	(-3.03)
Measure×dd		0.03			0.16			-0.02	
		(0.09)			(1.11)			(-0.26)	
Measure×bank_crisis			-1.95			1.09			0.13
			(-1.21)			(1.41)			(0.35)
Observations	6,316	6,316	6,316	6,316	6,316	6,316	6,316	6,316	6,316
Within R-squared	0.003	0.003	0.004	0.005	0.005	0.005	0.006	0.006	0.006
Episode FE	Y	Y	Y	Y	Y	Y	Y	Y	Y

Table 8: Predictive regression of realized returns during early stress episodes (1895-1958) in Panels A and B, and modern stress episodes (1959-2023) in Panel C. Panel B restricts the GFD sample to the most liquid stocks—defined as securities with fewer than 50% zero-return days in the past. The dependent variable is the realized returns of a financial institution during stress episodes. The systemic risk measures and the control variable for the firm size are all measured the quarter before the episode starts. The drawdown variable "dd" is the change in the credit spread over the episode. The credit spread is the GZ spread for the period 1973-2023, the Moody's spread for 1922-1972, and the average yield on U.S. railroad bonds for 1895-1921. The variable "bank_crisis" indicates the Great Depression and the GFC. t-statistics based on standard errors clustered at the firm level are reported in parentheses. Sample (Panel A): 1,875 financial institutions, including 1,195 banks. Sample (Panel B): 749 financial institutions, including 407 banks. Sample (Panel C): 2,186 financial institutions, including 761 banks.

3.4 Lending Outcomes: C&I vs. Real Estate Loans

Measure:	ΔCoVaR		eΔCoVaR		MES		SRISK/ME		Lvg		MES (+Lvg)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: loan growth												
Measure	-1.36***	-1.14**	-0.49***	-0.31**	-0.23***	-0.14	-0.70***	-0.46*	-0.06***	-0.04*	-0.19***	-0.13
	(-3.08)	(-2.09)	(-2.99)	(-2.01)	(-3.22)	(-1.52)	(-4.14)	(-1.78)	(-4.33)	(-1.67)	(-2.58)	(-1.40)
Lvg											-0.05***	-0.03
											(-3.12)	(-1.51)
Observations	2,222	2,222	2,222	2,222	2,222	2,222	2,222	2,222	2,222	2,222	2,222	2,222
Adj. R-squared	0.006	0.005	0.004	0.004	0.003	0.004	0.003	0.004	0.003	0.004	0.004	0.004
Panel B: C&I loan growth												
Measure	-1.22**	-1.19	-0.57**	-0.52*	-0.43***	-0.37***	-1.48***	-0.84**	-0.13***	-0.07*	-0.35***	-0.35***
	(-2.08)	(-1.32)	(-1.97)	(-1.81)	(-3.78)	(-2.79)	(-4.11)	(-2.24)	(-4.36)	(-1.77)	(-2.89)	(-2.67)
Lvg											-0.11***	-0.05
											(-3.28)	(-1.41)
Observations	2,202	2,202	2,202	2,202	2,202	2,202	2,202	2,202	2,202	2,202	2,202	2,202
Adj. R-squared	0.007	0.157	0.006	0.157	0.009	0.158	0.009	0.157	0.008	0.156	0.011	0.158
Panel C: real estate loan growth												
Measure	-1.54**	-1.30*	-0.51***	-0.41***	-0.18	0.01	-1.34***	-0.51*	-0.14***	-0.06***	-0.08	0.03
	(-2.32)	(-1.91)	(-2.94)	(-2.84)	(-1.18)	(0.03)	(-3.65)	(-1.78)	(-4.70)	(-2.58)	(-0.47)	(0.13)
Lvg											-0.13***	-0.06***
											(-4.05)	(-2.83)
Observations	2,210	2,210	2,210	2,210	2,210	2,210	2,210	2,210	2,210	2,210	2,210	2,210
Adj. R-squared	0.007	0.159	0.004	0.158	0.003	0.157	0.007	0.157	0.007	0.158	0.007	0.157
Episode FE	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y

Table 9: Predictive regression of lending outcomes during stress episodes (1959-2023). The dependent variable is an average balance sheet outcome of a bank during a stress episode: we consider the growth in total loans of the bank in Panel A, in Commercial and Industrial (C&I) loans in Panel B, and in real estate loans in Panel C. The systemic risk measure (Measure) and the control variables are all measured the quarter before the episode starts. Control variables include the firm size and the average bank outcome the year before the episode starts. Measure is indicated in the first line of the table. Coefficient estimates on SRISK/ME and Lvg are multiplied by 100. The even Columns (2) to (12) include episode fixed effects. t-statistics based on standard errors clustered at the bank level are reported in parentheses.

4 Comparative Analysis of Market-Based and Balance Sheet Indicators of Risk

4.1 Predicting Market and Balance Sheet Outcomes

We compare regressions using market-based measures to a baseline model based on the balance sheet indicators from Correia, Luck & Verner (2024) (hereafter "CLV") to predict bank outcomes during stress episodes for the 1959-2023 period. The objective is to evaluate the relative and joint predictive performance of market-based systemic risk indicators and balance sheet fundamentals.

We estimate the following regression specification:

$$\begin{aligned} y_{ie} &= \beta_1 Solvency_{ie} + \beta_2 Funding_{ie} + \beta_3 Solvency_{ie} \times Funding_{ie} + \beta_4 Growth_{ie} \\ &\quad + \beta_5 Measure_{ie} + (\beta_6 + \beta_7 Measure_{ie}) \times Lvg_{ie} + \delta control_{ie} + \alpha_e + \epsilon_{ie} \end{aligned} \tag{1}$$

where y_{ie} is the market or balance sheet outcome of bank i during episode e, and α_e captures episode fixed effects. The bank outcomes y_{ie} we consider include realized volatility, realized returns, loan growth, profitability—measured as return on assets (net income over lagged total assets)—and run risk, measured as the change in the ratio of uninsured deposits to total deposits. All variables on the right-hand side of eq. (1) are measured one quarter before the start of the episode. The balance sheet variables are from banks Call Reports and follow Correia, Luck & Verner (2024): $Solvency_{ie}$ is measured as net income over total assets and captures the risk of insolvency; $Funding_{ie}$ is the ratio of time deposits to total deposits and reflects reliance on non-core funding; and $Growth_{ie}$ captures asset expansion, constructed as quintiles of the change in log bank assets over the previous three years. Market-based measures of systemic risk ($Measure_{ie}$) include ΔCoVaR, eΔCoVaR, MES, SRISK/ME, and market leverage (Lvg), which may enter directly or through interactions with the systemic risk measures. Finally, we control for bank size in all specifications, as in the main regression results presented in the paper.

In Table 10, we report R-squared statistics for each outcome, with and without episode fixed effects. The adjusted R-squared is used when fixed effects are not included; the within R-squared is reported when episode fixed effects are present. In Panel A, we compare the predictive performance of market-based indicators with that of the CLV model. Specifically, we contrast a baseline specification that includes only the CLV balance sheet variables—imposing the restriction $\beta_5 = \beta_6 = \beta_7 = 0$—with specifications that include only market-based measures—imposing $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$. In Panel B, we augment the CLV specification with market-based measures to evaluate the extent to which these indicators provide complementary information. The R-squared values that yield the largest percentage gains relative to the CLV benchmark, for each outcome and specification (with and without fixed effects), are marked with an asterisk.

Panel A shows that market-based measures perform relatively well when fixed effects are included, consistent with their role in capturing the cross-sectional ranking of risk within a stress episode rather than the unconditional level of outcomes across episodes. For instance, the highest within R-squared for realized volatility is achieved by MES × Lvg (0.043), outperforming the CLV model (0.036). Similarly, profitability is best predicted by models that include interactions between market-based measures—MES, ΔCoVaR, eΔCoVaR—and leverage, with a maximum within R-squared of 0.215, exceeding by 7% the CLV model (0.200). For loan growth, ΔCoVaR consistently delivers the best performance across specifications, with the highest R-squared values (0.006), 20% above the CLV benchmark. As a result, market-based indicators tend to outperform balance sheet variables when the goal is to rank institutions by relative risk exposure within a given episode. The only outcome for which this pattern does not hold is run risk, where the CLV model dominates, even in specifications with fixed effects.

In Panel B, we find that the models combining CLV variables with market-based measures deliver higher explanatory power than either the CLV-only or market-only specifications. In particular, for realized volatility and profitability, the highest within R-squared values (0.043 and 0.215, respectively) are achieved when market-based indicators are interacted with leverage. For run risk, again, the model combining CLV variables with ΔCoVaR $\times$ Lvg achieves the highest explanatory power, even in the absence of fixed effects. These patterns suggest that market-based and balance sheet-based measures are complementary.

Taken together, the results suggest that market-based and balance sheet indicators capture different dimensions of risk. Market-based measures are well suited to identifying which institutions are most vulnerable within a given episode of stress, while balance sheet indicators—especially those related to funding structure—are more informative about certain types of fundamental fragility, such as exposure to depositor runs. The best results are generally achieved when the two are used together, highlighting their complementarity.

	Panel A: Outcomes during stress episodes (Measure vs. CLV)					
Measure:	ΔCoVaR		eΔCoVaR		MES	
Dependent:						
Realized vol	0.001	0.002	0.005	0.014	0.008	0.018
Realizet ret	0.018	0.001	0.011	0.003	0.004	0.002
Loan growth	0.006*	0.006*	0.005	0.005	0.004	0.005
Profitability	0.170	0.187	0.174	0.187	0.172	0.188
Run risk	0.011	0.003	0.009	0.003	0.008	0.003
	SRISK/ME		Lvg		CLV	
Dependent:						
Realized vol	0.008	0.036	0.006	0.032	0.005	0.036
Realized ret	0.002	0.005	0.002	0.005	0.065*	0.008
Loan growth	0.004	0.005	0.004	0.005	0.005	0.005
Profitability	0.183	0.204	0.189	0.213	0.218*	0.200
Run risk	0.012	0.007	0.011	0.006	0.042*	0.036*
	ΔCoVaR(+Lvg)		eΔCoVaR(+Lvg)		MES(+Lvg)	
Dependent:						
Realized vol	0.006	0.033	0.009	0.041	0.011*	0.042
Realized ret	0.017	0.005	0.011	0.007	0.004	0.006
Loan growth	0.006*	0.006*	0.005	0.005	0.004	0.005
Profitability	0.188	0.213	0.191	0.214*	0.189	0.213
Run risk	0.015	0.007	0.012	0.006	0.012	0.007
	ΔCoVaR(×Lvg)		eΔCoVaR(×Lvg)		MES(×Lvg)	
Dependent:						
Realized vol	0.006	0.037	0.009	0.042	0.011*	0.043*
Realized ret	0.030	0.007	0.019	0.010*	0.008	0.008
Loan growth	0.006*	0.006*	0.004	0.005	0.004	0.005
Profitability	0.188	0.215*	0.191	0.215*	0.189	0.215*
Run risk	0.024	0.018	0.012	0.007	0.012	0.007
Observations:						
Realized vol	1,938	1,938	1,938	1,938	1,938	1,938
Realized ret	1,938	1,938	1,938	1,938	1,938	1,938
Loan growth	1,926	1,926	1,926	1,926	1,926	1,926
Profitability	1,931	1,931	1,931	1,931	1,931	1,931
Run risk	1,501	1,501	1,501	1,501	1,501	1,501
Episode FE	N	Y	N	Y	N	Y

Panel B: Outcomes during stress episodes (Measure+CLV)						
Measure:	ΔCoVaR		eΔCoVaR		MES	
Dependent:						
Realized vol	0.005	0.036	0.012	0.046	0.016	0.050
Realizet ret	0.073	0.008	0.069	0.011	0.065	0.010
Loan growth	0.007*	0.007*	0.006	0.006	0.005	0.006
Profitability	0.218	0.201	0.219	0.202	0.219	0.202
Run risk	0.042	0.036	0.042	0.036	0.042	0.036
	SRISK/ME		Lvg		CLV	
Dependent:						
Realized vol	0.012	0.059	0.011	0.055	0.005	0.036
Realized ret	0.065	0.011	0.065	0.011	0.065	0.008
Loan growth	0.005	0.005	0.005	0.005	0.005	0.005
Profitability	0.227	0.217	0.231	0.225	0.218	0.200
Run risk	0.043	0.037	0.043	0.036	0.042	0.036
	ΔCoVaR(+Lvg)		eΔCoVaR(+Lvg)		MES(+Lvg)	
Dependent:						
Realized vol	0.010	0.056	0.016	0.064	0.018*	0.065
Realized ret	0.072	0.011	0.069	0.014	0.065	0.012
Loan growth	0.006	0.007*	0.005	0.006	0.005	0.006
Profitability	0.231*	0.226	0.232*	0.226	0.231*	0.226
Run risk	0.043	0.037	0.042	0.036	0.042	0.037
	ΔCoVaR(×Lvg)		eΔCoVaR(×Lvg)		MES(×Lvg)	
Dependent:						
Realized vol	0.010	0.058	0.016	0.064	0.018*	0.067*
Realized ret	0.085*	0.014	0.074	0.018*	0.067	0.014
Loan growth	0.006	0.007*	0.005	0.006	0.005	0.006
Profitability	0.231*	0.227*	0.232*	0.228*	0.232*	0.227*
Run risk	0.051*	0.045*	0.041	0.037	0.041	0.037
Observations:						
Realized vol	1,938	1,938	1,938	1,938	1,938	1,938
Realized ret	1,938	1,938	1,938	1,938	1,938	1,938
Loan growth	1,926	1,926	1,926	1,926	1,926	1,926
Profitability	1,931	1,931	1,931	1,931	1,931	1,931
Run risk	1,501	1,501	1,501	1,501	1,501	1,501
Episode FE	N	Y	N	Y	N	Y

Table 10: This table reports the R-squared of predictive regression of market and balance sheet outcomes of banks during stress episodes (1959-2023). The R-squared is the adjusted R-squared for regressions without fixed effects, and the within R-squared for regressions with fixed effects. Panel A compares the predictive performance of market-based measures, the book equity ratio, and the baseline CLV model. Panel B compares the predictive performance of market-based measures and the book equity ratio, in addition to CLV variables. The symbol * indicates the R-squared value that yields the largest percentage gains relative to the CLV benchmark for each dependent variable and regression specification. The dependent variable is the market or balance sheet outcome of a bank during stress episodes. "Profitability" is the return on assets, and "run risk" is the uninsured deposit growth. Sample: 628 banks.

We extend the predictive regressions to a broader sample that includes both banks and non-bank financial institutions. This contrasts with the previous table, which relied on Call Reports and was restricted to banks. In this setting, balance sheet variables are constructed from Compustat. All CLV variables are replicated except for the funding ratio and its interaction with solvency, which require a deposit decomposition that is not available in Compustat. As a result, the CLV specification here includes only solvency and asset growth.

The regression specification and overall design of the exercise remain the same. We aim to assess and compare the predictive performance of market-based systemic risk indicators and balance sheet fundamentals, both individually and jointly. Table 11 reports the R-squared statistics from regressions of realized volatility and realized returns during stress episodes between 1959 and 2023. Adjusted R-squared values are shown for specifications without episode fixed effects, and within R-squared values are reported for models with episode fixed effects. Panel A compares the predictive performance of market-based measures and the CLV model—restricted to solvency and asset growth for consistency with the available Compustat data. Panel B adds market-based measures to this same CLV specification to assess their complementarity.

Panel A shows that market-based indicators generally perform well when episode fixed effects are included, consistent with the idea that these measures are especially informative about the cross-sectional ranking of risk within a stress episode. For realized volatility, MES $\times$ Lvg delivers the best within R-squared (0.069), outperforming all other models, including the restricted CLV model (0.047). Similarly, eΔCoVaR $\times$ Lvg and ΔCoVaR $\times$ Lvg also achieve high explanatory power (within R-squared of 0.059 and 0.035, respectively). For realized returns, while overall R-squared values remain modest, the interaction of MES with leverage again achieves the best within R-squared (0.012), ahead of the restricted CLV model (0.005).

Panel B evaluates the performance of models that combine CLV variables with market-based systemic risk measures. As in the bank-only sample, the joint models deliver higher R-squared than either the market-only or CLV-only specifications. For realized volatility, MES $\times$ Lvg again achieves the highest within R-squared (0.089), confirming that interactions between systemic risk and leverage contribute substantially to explanatory power. For realized returns, the model combining MES $\times$ Lvg with CLV variables also performs best (within R-squared = 0.013), suggesting that even modest improvements can be gained from the joint specification. In all cases, the inclusion of market-based indicators enhances the performance of the restricted CLV model.

Overall, these results reinforce the two main findings from the previous analysis. First, market-based indicators do a better job at predicting the cross-sectional variation of outcomes during stress episodes. Second, market-based and balance-sheet-based indicators are complementary. The best predictive performance is generally achieved when both types of indicators are used together.

	Panel A: Outcomes during stress episodes (Measure vs. CLV)					
Measure:	ΔCoVaR		eΔCoVaR		MES	
Dependent:						
Realized vol	0.007	0.025	0.024	0.050	0.019	0.062
Realizet ret	0.003	0.005	0.003	0.009	0.001	0.011
	SRISK/ME		Lvg		CLV	
Dependent:						
Realized vol	0.012	0.034	0.011	0.031	0.032*	0.047
Realized ret	0.001	0.004	0.001	0.004	0.007*	0.005
	ΔCoVaR(+Lvg)		eΔCoVaR(+Lvg)		MES(+Lvg)	
Dependent:						
Realized vol	0.013	0.031	0.028	0.055	0.023	0.066
Realized ret	0.003	0.005	0.003	0.009	0.001	0.011
	ΔCoVaR(×Lvg)		eΔCoVaR(×Lvg)		MES(×Lvg)	
Dependent:						
Realized vol	0.016	0.035	0.030	0.059	0.025	0.069*
Realized ret	0.003	0.005	0.003	0.009	0.002	0.012*

	Panel B: Outcomes during stress episodes (Measure+CLV)					
Measure:	ΔCoVaR		eΔCoVaR		MES	
Dependent:						
Realized vol	0.034	0.047	0.050	0.071	0.045	0.082
Realizet ret	0.009	0.007	0.009	0.010	0.008	0.012
	SRISK/ME		Lvg		CLV	
Dependent:						
Realized vol	0.039	0.055	0.037	0.053	0.032	0.047
Realized ret	0.007	0.005	0.007	0.005	0.007	0.005
	ΔCoVaR(+Lvg)		eΔCoVaR(+Lvg)		MES(+Lvg)	
Dependent:						
Realized vol	0.040	0.053	0.053	0.075	0.049	0.086
Realized ret	0.009	0.007	0.009	0.010	0.008	0.012
	ΔCoVaR(×Lvg)		eΔCoVaR(×Lvg)		MES(×Lvg)	
Dependent:						
Realized vol	0.042	0.056	0.055*	0.079	0.050	0.089*
Realized ret	0.009	0.007	0.010*	0.011	0.008	0.013*
Observations	5,837	5,837	5,837	5,837	5,837	5,837
Episode FE	N	Y	N	Y	N	Y

Table 11: This table reports the R-squared of predictive regression of realized volatility and realized returns of financial institutions during stress episodes (1959-2023). The R-squared is the adjusted R-squared for regressions without fixed effects, and the within R-squared for regressions with fixed effects. Panel A compares the predictive performance of market-based measures, the book equity ratio, and the baseline CLV model. Panel B compares the predictive performance of market-based measures and the book equity ratio, in addition to CLV variables. The symbol * indicates the R-squared value that yields the largest percentage gains relative to the CLV benchmark for each dependent variable and regression specification. The dependent variable is the market outcome of a financial institution during stress episodes. CLV variables are restricted to solvency and asset growth variables for this sample including non-bank financial institutions. Sample: 1,972 financial institutions, including 720 banks.

4.2 Predicting Bank Failure

To compare the predictive performance of systemic risk measures with the model proposed by Correia, Luck & Verner (2024), we estimate the following specification, which includes their model for predicting a bank failure over the next h quarters:

$$\begin{aligned} Failure_{i,t+1\to t+h} \;=\;& \alpha + \beta_1 Solvency_{it} + \beta_2 Funding_{it} + \beta_3 Solvency_{it} \times Funding_{it} + \beta_4 Growth_{it} \\ & + \beta_5 Measure_{it} + (\beta_6 + \beta_7 Measure_{it}) \times Lvg_{it} + \delta controls_{(i)t} + \epsilon_{i,t+1\to t+h} \end{aligned} \tag{2}$$

where $Solvency_{it}$, $Funding_{it}$, $Growth_{it}$ are the CLV variables that respectively capture insolvency risk, reliance on non-core funding, and bank asset growth, as described in the previous section. Market-based measures of systemic risk ($Measure_{it}$) include ΔCoVaR, eΔCoVaR, MES, SRISK/ME, and market leverage (Lvg), which may enter directly or through interactions with the systemic risk measures. The control variables include bank size (log of market capitalization) for specifications with market-based measures, and three-year real GDP growth in all regressions. This GDP control proxies for aggregate economic conditions and serves a similar purpose to the episode fixed effects used in the cross-sectional outcome regressions, which cannot be included in logit models due to the incidental parameters problem.

Table 12 reports the area under the ROC curve (AUC) for predictive logit regressions of bank failure over one-, three-, and five-year horizons. Panel A compares the performance of market-based measures to the CLV model, while Panel B examines their combined predictive power. An asterisk indicates the AUC value that yields the largest percentage gain relative to the CLV benchmark for each forecasting horizon.

Panel A shows that market-based measures outperform the CLV model when predicting one-year-ahead failures. For this short horizon, several specifications involving MES and eΔCoVaR achieve higher AUC values than the CLV benchmark. Notably, SRISK/ME achieves the highest AUC of 0.9096, compared to 0.8455 for the CLV model. ΔCoVaR × Lvg and MES × Lvg also perform well, with AUCs of 0.8819 and 0.8682, respectively. However, the pattern reverses at longer horizons. For failure predictions over three- and five-year horizons, the CLV model consistently delivers the best performance of 0.8098 and 0.7710, respectively, suggesting that balance sheet fundamentals, particularly past profitability, are more informative about medium-term bank failures.

Panel B confirms that combining market-based indicators with CLV variables enhances predictive performance. The combination of ΔCoVaR × Lvg and CLV yields the top-performing model across all horizons, including AUCs of 0.8802 and 0.8350 for the three- and five-year horizons, respectively. The AUCs of eΔCoVaR × Lvg and MES × Lvg deliver similar predictive accuracy, with AUCs of 0.8749 and 0.8324 in Table 12, respectively, for MES × Lvg at the three- and five-year horizons. These results indicate that market and balance sheet data are complementary, with market-based indicators improving near-term predictive power and balance sheet fundamentals anchoring longer-term predictions.

In summary, the results highlight three main takeaways. First, market-based systemic risk indicators outperform CLV at short horizons, particularly when forecasting failures within the next year. Second, CLV variables are more informative at longer horizons. Third, the best overall performance is achieved when combining both types of indicators, underscoring their complementarity.

Panel A: Fail in x years (Measure vs. CLV)						
Measure:	ΔCoVaR		eΔCoVaR		MES	
x=1	0.5639	0.6969	0.6915	0.7774	0.6500	0.7481
x=3	0.5303	0.6415	0.5955	0.6212	0.5674	0.6059
x=5	0.5130	0.5657	0.5539	0.5355	0.5273	0.5179
	SRISK/ME		Lvg		CLV	
x=1	0.9096*	0.8676	0.8801	0.8665	0.8430	0.8455
x=3	0.7032	0.7051	0.6848	0.7080	0.8098*	0.8091
x=5	0.5722	0.6285	0.5540	0.6339	0.7655	0.7710*
	ΔCoVaR(+Lvg)		eΔCoVaR(+Lvg)		MES(+Lvg)	
x=1	0.8725	0.8616	0.8668	0.8505	0.8655	0.8395
x=3	0.6944	0.7371	0.6980	0.7050	0.6918	0.7112
x=5	0.5685	0.6824	0.6067	0.6444	0.5886	0.6653
	ΔCoVaR(×Lvg)		eΔCoVaR(×Lvg)		MES(×Lvg)	
x=1	0.8819	0.8635	0.8678	0.8506	0.8682	0.8390
x=3	0.6993	0.7422	0.6985	0.7047	0.6946	0.7128
x=5	0.5567	0.6883	0.6047	0.6458	0.5850	0.6681

Panel B: Fail in x years (Measure+CLV)						
Measure:	ΔCoVaR		eΔCoVaR		MES	
x=1	0.8575	0.8853	0.8953	0.8881	0.8925	0.8821
x=3	0.8217	0.8652	0.8391	0.8602	0.8405	0.8634
x=5	0.7755	0.8248	0.7829	0.8217	0.7846	0.8227
	SRISK/ME		Lvg		CLV	
x=1	0.8770	0.9254	0.8731	0.9252	0.8430	0.8455
x=3	0.8205	0.8753	0.8185	0.8755	0.8098	0.8091
x=5	0.7706	0.8303	0.7693	0.8307	0.7655	0.7710
	ΔCoVaR(+Lvg)		eΔCoVaR(+Lvg)		MES(+Lvg)	
x=1	0.8819	0.9260*	0.9089	0.9155	0.9087	0.9187
x=3	0.8291	0.8797*	0.8426	0.8727	0.8443	0.8755
x=5	0.7787	0.8341*	0.7841	0.8307	0.7859	0.8321*
	ΔCoVaR(×Lvg)		eΔCoVaR(×Lvg)		MES(×Lvg)	
x=1	0.8811	0.9275*	0.9085	0.9155	0.9076	0.9187
x=3	0.8282	0.8802*	0.8423	0.8727	0.8433	0.8749
x=5	0.7783	0.8350*	0.7841	0.8312*	0.7854	0.8324*
Observations	30,889	30,889	30,889	30,889	30,889	30,889
Controls	N	Y	N	Y	N	Y

Table 12: This table reports the AUCs of predictive regressions of bank failures. Panel A compares the predictive performance of market-based measures, the book equity ratio, and the baseline CLV model. Panel B compares the predictive performance of market-based measures and the book equity ratio, in addition to CLV variables. The symbol * indicates the AUC value that yields the largest percentage gains relative to the CLV benchmark for each forecasting horizon (x). Controls include asset growth and past 3-year GDP growth for the CLV model, and firm size (log of market cap) and past 3-year GDP growth for the other models. Sample: 819 banks.

In Table 13, we report the results of a linear probability model to predict bank failure within the next five years, together with the corresponding Pseudo R-squared and AUC of the logit regressions. We also report in the table the AUCs of the CLV variables for predicting failure over shorter horizons of one and three years. We note that the estimates we obtain for the CLV model have the same signs as those reported in Correia, Luck & Verner (2024) (Table B9), despite the different sample used in our analysis. In contrast to their methodology, we estimate the probability of failure at the parent bank level, and our sample is limited to publicly traded banks. Banks are more likely to fail in subsequent years if they are less profitable, rely on less stable funding sources such as time deposits, and have experienced rapid asset expansion over the past three years, consistent with the literature showing that fast loan growth predicts poor future bank performance (Fahlenbrach, Prilmeier & Stulz, 2018).

	Failure probability (fail in 5 years)						
solvency	-1.24***			-1.13***	0.66***	0.62***	0.60***
	(-15.62)			(-14.25)	(5.41)	(5.09)	(4.90)
funding		0.08***		0.08***	0.10***	0.11***	0.11***
		(18.65)		(17.51)	(23.02)	(24.77)	(24.53)
solvency×funding					-5.65***	-5.64***	-5.79***
					(-19.12)	(-19.14)	(-19.52)
asset_growth			0.01***			0.01***	0.01***
			(10.73)			(14.19)	(14.64)
agg_cond							0.01***
							(4.17)
Observations	30,889	30,889	30,889	30,889	30,889	30,889	30,889
Adj. R-squared	0.008	0.011	0.004	0.018	0.029	0.035	0.036
Pseudo R-squared	0.016	0.064	0.021	0.079	0.096	0.132	0.135
AUC	0.6051	0.7090	0.6386	0.7152	0.7100	0.7655	0.7710
AUC (1 year)	0.8996	0.7624	0.4982	0.8187	0.8231	0.8430	0.8455
AUC (3 years)	0.7245	0.7387	0.6202	0.7631	0.7575	0.8098	0.8091

Table 13: Predicting bank failures: Correia, Luck & Verner (2024) balance sheet indicators. The dependent variable is equal to one if the bank fails in the next five years, and zero otherwise. OLS estimates and adjusted R-squared refer to the linear probability model of Correia, Luck & Verner (2024) (Table B9). solvency is the ratio of net income to total assets, funding is the ratio of time deposits to total deposits, asset_growth is the change in log total assets over the prior three years, agg_cond is the real GDP growth over the same past three years. Pseudo R-squared are obtained from corresponding logit regressions. AUC is the area under the receiver operating characteristic (ROC) curve. AUC (1 year, resp. 3 years) measures the accuracy of the failure prediction model at one- and three-year horizons, respectively. Sample: 819 banks.

5 Episode Dating Robustness

According to Harding & Pagan (2002), a cycle dating algorithm must perform three key tasks: identify a set of turning points (peaks and troughs) in a time series, ensure that these alternate properly, and apply rules—known as "censoring rules"—to filter the turning points based on predefined criteria for the duration and amplitude of phases and complete cycles.

The most widely used algorithm for this purpose is that of Bry & Boschan (1971), developed for monthly macroeconomic data and adopted by the NBER. It defines a turning point at time t when the value of the series at t is greater (or smaller) than at $t \pm w$, for w from 1 to W, with W typically set to five months. Censoring rules then require that each phase last at least six months and each full cycle at least fifteen. Harding & Pagan (2002) adapted this algorithm to quarterly data (BBQ), such as GDP, using $W = 2$ (six months), a minimum phase duration of two quarters, and a minimum cycle length of five quarters.

However, while the BB and BBQ algorithms are designed for trending macroeconomic series, the credit spread used in our analysis is stationary and exhibits more frequent fluctuations. Applying the standard parameters to such a series results in an excessive number of identified cycles. To address this, we adjust the parameters to capture more economically meaningful episodes by allowing for longer cycles: a window of six months, a minimum phase duration of six months, and a minimum cycle length of 26 months.

The rest of this section proceeds in three steps. In Section 5.1, we apply the BB algorithm to the GZ spread over the 1973–2023 period to identify stress episodes for the modern sample. Next, we extend the analysis to the historical sample by applying the same algorithm to the Moody's Baa-Aaa spread and the railroad bond yield index in Section 5.2. Finally, in Section 5.3, we assess the robustness of our results by evaluating the ability of systemic risk measures to predict market outcomes of financial institutions during stress episodes, using the alternative BB episode dating methodology.

5.1 Modern Episode Dating Robustness

We compare in Figure 2 and Table 14 the episodes of financial stress identified using our methodology (ABP) presented in Section 2.1 and the adapted Bry-Boschan (BB) algorithm. The two approaches yield broadly similar results for the 1959-2023 period: 8 out of the 11 episodes are identified using both ABP and BB procedures.

When the algorithms diverge, the differences tend to be in timing the start of the episode rather than in the recognition of the peak itself. In these cases, the BB algorithm typically identifies the same peak date as ABP but assigns an earlier start to the episode.

A more significant discrepancy arises in the treatment of the LTCM crisis and the dot.com bubble. While Panel A distinguishes these as two separate events, Panel B merges them into a single, extended 62-month episode. In addition, Panel B includes three episodes that do not appear in Panel A: the periods from August 1990 to December 1991, from February 1994 to December 1995, and from December 2004 to September 2006. These episodes are associated with only mild fluctuations in credit spreads and show little or no evidence of stock market or financial sector drawdowns. Their appearance in the BB-based chronology reflects a key feature of the algorithm: it does not impose any explicit threshold for the magnitude of the shock, and therefore may capture cyclical patterns that are not economically meaningful in terms of market stress.

Panel A: Modern Sample (1959-2023) - ABP episodes								
episode	start	end	months	GZ	SP500		fin. index	
				sprd (%)	ret (%)	dd (%)	ret (%)	dd (%)
High inflation in the U.S.	9-1973	12-1974	15	1.75	-36.77	-41.40	-35.73	-46.82
1977-82 stock market crash	10-1978	10-1982	48	1.21	43.54	-23.79	43.44	-34.12
S&L Crisis	5-1984	8-1986	27	1.58	68.00	0	54.24	-11.14
LTCM hedge fund failure	8-1997	10-1998	14	1.12	22.15	-15.57	-10.08	-36.40
Dot.com Bubble	1-2000	10-2002	33	2.36	-36.48	-46.28	-24.11	-36.84
Global Financial Crisis	5-2007	11-2008	18	6.44	-41.45	-42.16	-52.09	-52.09
European Sovereign Debt Crisis	3-2011	9-2011	6	1.16	-14.66	-17.03	-18.99	-20.15
2014-16 oil price shock	6-2014	2-2016	20	1.52	-1.43	-8.89	0.90	-9.81
Covid19 Pandemic	12-2019	3-2020	3	2.37	-20.00	-20.00	-18.47	-18.47
Ukraine war/energy crisis	12-2021	6-2022	6	0.70	-20.58	-20.58	-8.77	-8.77
Bank Failures in 2023	1-2023	5-2023	4	0.15	2.53	0	-2.96	-5.97

Panel B: Modern Sample (1959-2023) - BB episodes (w=8, p=6, c=26)								
episode	start	end	months	GZ	SP500		fin. index	
				sprd (%)	ret (%)	dd (%)	ret (%)	dd (%)
High inflation in the U.S.	9-1973	12-1974	15	1.75	-36.77	-41.40	-35.73	-46.82
1977-82 stock market crash	10-1978	10-1982	48	1.21	43.54	-23.79	43.44	-34.12
S&L Crisis	5-1984	8-1986	27	1.58	68.00	0	54.24	-11.14
New BB episode 1	8-1990*	12-1991*	16	0.42	29.31	0	40.55	0
New BB episode 2	2-1994*	12-1995*	22	0.28	31.85	0	28.87	0
LTCM hedge fund failure	8-1997							
+ Dot.com Bubble		10-2002	62	3.17	-1.52	-46.28	-35.07	-51.73
New BB episode 3	12-2004*	9-2006*	22	0.35	10.23	0	6.78	-8.89
Global Financial Crisis	5-2007	11-2008	18	6.44	-41.45	-42.16	-52.09	-52.09
European Sovereign Debt Crisis	3-2011	9-2011	6	1.16	-14.66	-17.03	-18.99	-20.15
2014-16 oil price shock	6-2014	2-2016	20	1.52	-1.43	-8.89	0.90	-9.81
Covid19 Pandemic	1-2018*	3-2020	26	2.58	-8.47	-20.00	-5.93	-18.47
Ukraine war/energy crisis	10-2021*	6-2022	8	0.76	-17.81	-20.58	-7.21	-8.77
Bank Failures in 2023	-	-	-	-	-	-	-	-

Table 14: Comparison of modern stress episode dating. Episode dates are based on our methodology presented in Section 2.1 (ABP) in Panel A, and on the Bry & Boschan (1971) cycle dating algorithm in Panel B. The symbol * indicates a different date identified by the BB algorithm. “start” and “end” respectively indicate the start date and the end date of the episode, and “months” is the length in months of the episode. “GZ sprd” is the change in the GZ spread in percentage points. “SP500” is the S&P500 index, “fin. index” is the CRSP financial index. “ret” denotes the index return in percentage points over the episode. “dd” is the maximum drawdown on an index, defined as the percentage difference between its minimum and the prior maximum value within an episode. There is no drawdown when the minimum index value is reached at the beginning of the episode.

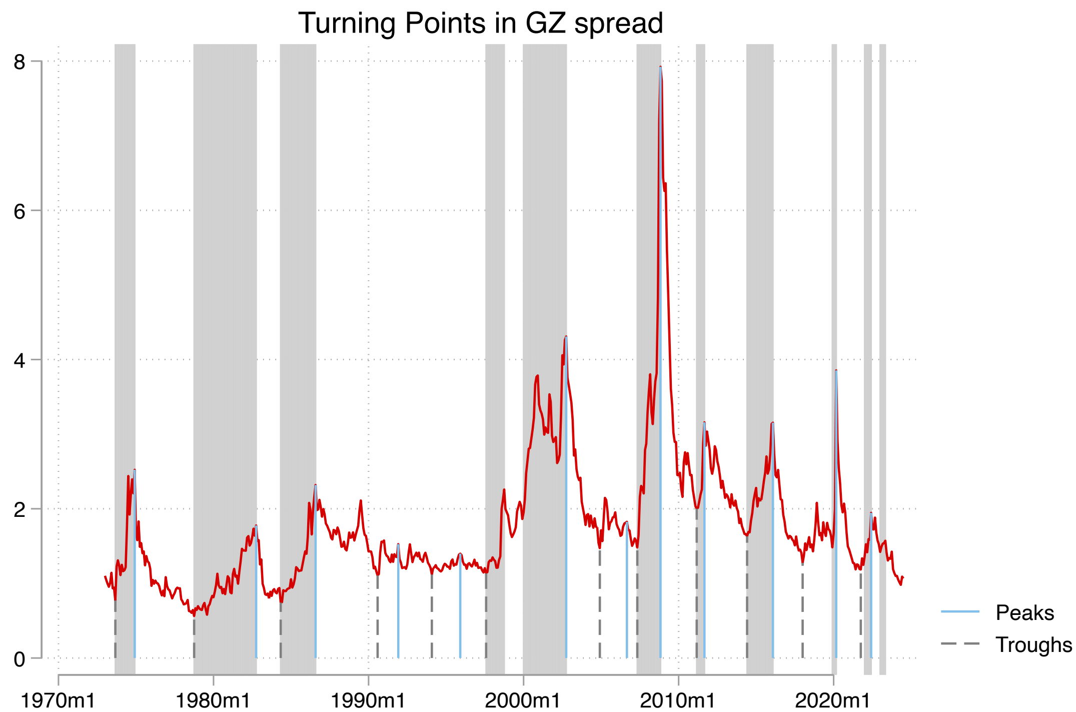

Figure 2: GZ turning points identification: Bry & Boschan (1971) algorithm. "Peaks" and "troughs" dates are identified using the BB algorithm. Grey shaded areas correspond to the stress episodes identified using our methodology presented in Section 2.1. BB parameters: window=8 months, phase=6 months, cycle=26 months.

5.2 Early Episode Dating Based Robustness

Panels A and B of Table 15 and Figure 3 present a comparison of stress episodes in the early sample (1895–1958) identified using the ABP methodology and the BB algorithm, respectively. There is a substantial overlap between the two approaches: 9 out of the 10 episodes in Panel A also appear in Panel B, indicating that both methods capture the main stress episodes during this period.

As in the modern sample, the BB algorithm tends to assign an earlier start to many of the episodes. In most cases, the peak date remains consistent between the two panels, but the BB algorithm typically identifies a trough several months earlier.

A key divergence is found in the timing and length of the Great Depression. Panel A dates the episode from February 1929 to May 1932, while the BB methodology (Panel B) begins it earlier, in March 1928, and extends the episode to a total of 50 months. The BB algorithm also identifies an additional episode from April 1953 to February 1954, which does not appear in the ABP chronology. This episode coincides with a period of mild spread elevation but shows no significant drawdown in equity or financial sector returns.

Panel A: Early Sample (1895-1958) - ABP episodes								
episode	start	end	months	credit	SP500		fin. index	
				sprd (%)	ret (%)	dd (%)	ret (%)	dd (%)
1903-04 stock market crash	4-1902	9-1903	16	0.33	-31.65	-31.65	-36.65	-46.38
1907 banking crisis	1-1906	11-1907	22	0.68	-42.88	-42.88	-30.49	-30.49
1910-11 stock market crash	1-1910	7-1910	6	0.12	-15.60	-15.60	7.88	-6.00
1914 banking crisis	1-1913	12-1914	23	0.38	-34.81	-34.81	-18.56	-23.99
1916-21 stock market crash	1-1917	5-1920	40	1.79	-3.79	-23.33	3.92	-30.29
The Great Depression	2-1929	5-1932	39	4.64	-85.90	-88.24	-91.95	-95.05
Recession of 1937–38	2-1937	4-1938	14	1.86	-40.34	-47.13	-16.30	-30.27
1939-41 stock market crash	10-1939	6-1940	8	0.42	-19.63	-23.48	-15.13	-22.68
Post-World War II Recession	1-1946	7-1949	42	0.32	-14.05	-21.42	-5.07	-19.40
Recession of 1958	4-1956	1-1958	21	0.79	-12.81	-15.86	-22.86	-30.29

Panel B: Early Sample (1895-1958) - BB episodes (w=8, p=6, c=26)								
episode	start	end	months	credit	SP500		fin. index	
				sprd (%)	ret (%)	dd (%)	ret (%)	dd (%)
1903-04 stock market crash	3-1901*	9-1903	23	0.32	-34.03	-40.11	-30.45	-50.66
1907 banking crisis	2-1905*	11-1907	33	0.72	-23.47	-42.88	-40.39	-48.57
1910-11 stock market crash	2-1909*	7-1910	17	0.21	-4.63	-22.08	3.67	-6.00
1914 banking crisis	2-1914*	9-1915*	19	0.37	10.46	0	-15.73	-22.22
1916-21 stock market crash	1-1917	5-1920	40	1.79	-3.79	-23.33	3.92	-30.29
The Great Depression	3-1928*	5-1932	50	4.78	-79.14	-88.24	-95.59	-97.65
Recession of 1937–38	2-1937	4-1938	14	1.86	-40.34	-47.13	-49.50	-60.00
1939-41 stock market crash	10-1939	6-1940	8	0.42	-19.63	-23.48	-19.42	-31.00
Post-World War II Recession	3-1946	7-1949	42	0.32	-11.85	-21.42	5.22	-15.25
New BB episode	4-1953*	2-1954*	10	0.24	7.20	0	0.26	0
Recession of 1958	4-1956	1-1958	21	0.79	-12.81	-15.86	-16.21	-23.39

Table 15: Comparison of early stress episode dating. Episode dates are based on our methodology presented in Section 2.1 (ABP) in Panel A, and on the Bry & Boschan (1971) cycle dating algorithm in Panel B. The symbol * indicates a different date identified by the BB algorithm. "start" and "end" respectively indicate the start date and the end date of the episode, and "months" is the length in months of the episode. "credit sprd" is the change in the spread between Moody's Baa and Aaa Corporate Bond Yields in percentage points for 1922-1972, and the change in the average yield of U.S. railroad bonds (Macaulay, 1938) for 1895-1921. "DJI" is the Dow Jones Industrial Average index, "fin. index" is the GFD financial index. "ret" denotes the index return in percentage points over the episode. "dd" is the maximum drawdown on an index, defined as the percentage difference between its minimum and the prior maximum value within an episode. There is no drawdown when the minimum index value is reached at the beginning of the episode.

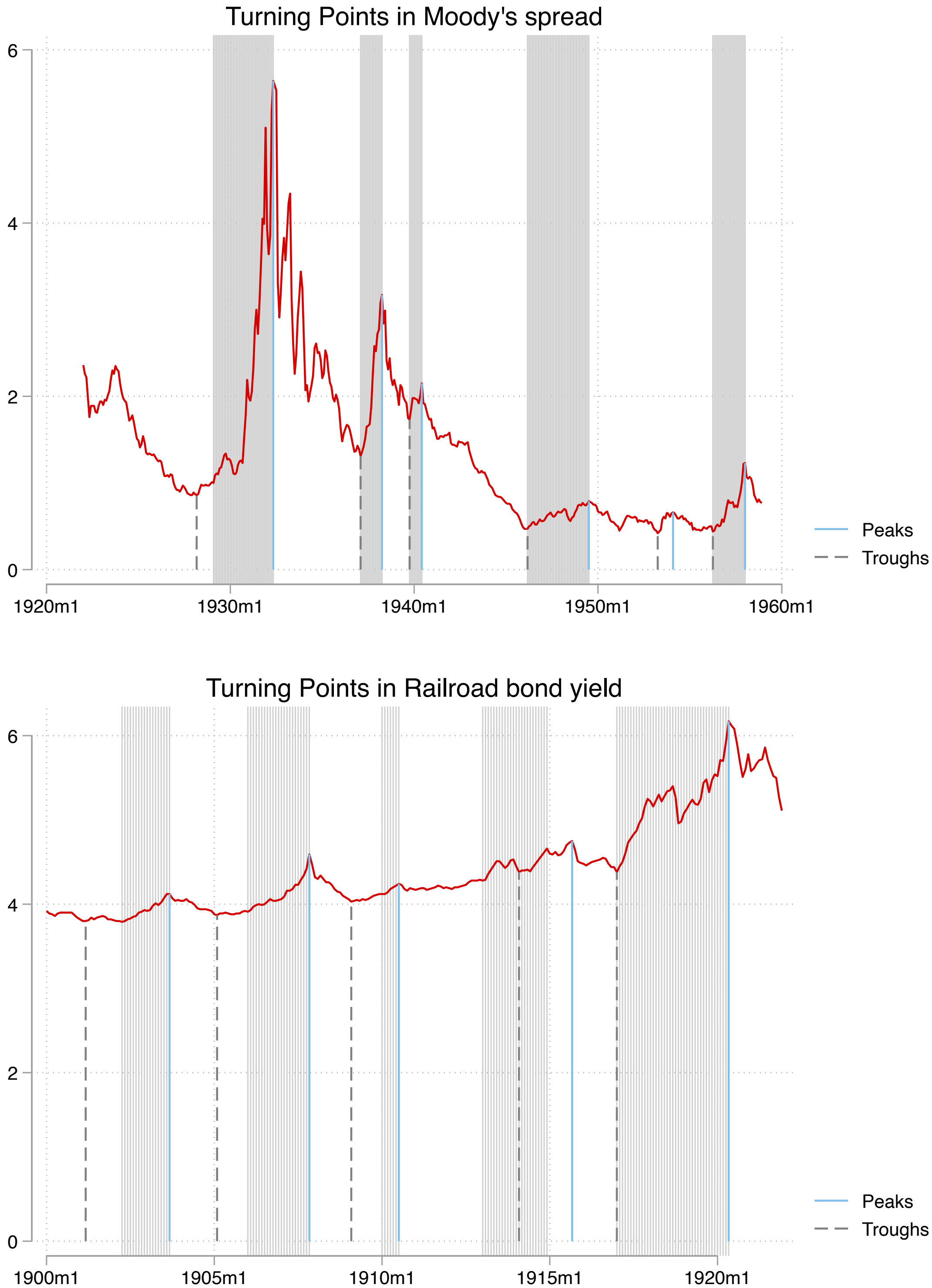

Figure 3: Moody's spread and Railroad bond yield turning points identification: Bry & Boschan (1971) algorithm. "Peaks" and "troughs" dates are identified using the BB algorithm. Grey shaded areas correspond to the stress episodes using our methodology presented in Section 2.1. BB parameters: window=8 months, phase=6 months, cycle=26 months.

5.3 Market Outcomes: Robustness to BB Episode Dating

	Panel A: Realized volatility during stress episodes					
Measure:	ΔCoVaR		eΔCoVaR		MES	
	(1)	(2)	(3)	(4)	(5)	(6)
Measure×bank	-0.03	0.24**	0.07	0.27***	-0.05	0.23***
	(-0.31)	(2.30)	(1.10)	(3.55)	(-1.41)	(7.42)
Measure×nonbank	-0.39***	0.29**	0.39***	0.47***	0.23***	0.33***
	(-3.83)	(2.34)	(7.94)	(9.23)	(8.00)	(13.20)
Observations	12,370	12,370	12,370	12,370	12,370	12,370
Adj. R-squared	0.126	0.332	0.133	0.344	0.134	0.352
Episode FE	N	Y	N	Y	N	Y
Measure:	SRISK/ME		Lvg		MES (+Lvg)	
	(7)	(8)	(9)	(10)	(11)	(12)
Measure×bank	1.08***	0.98***	0.09***	0.07***	0.06	0.28***
	(3.78)	(3.77)	(3.65)	(4.34)	(0.99)	(5.01)
Measure×nonbank	0.69***	0.67***	0.05***	0.05***	0.22***	0.31***
	(5.53)	(4.75)	(5.61)	(4.80)	(7.09)	(9.18)
Lvg×bank					0.09***	0.07***
					(3.62)	(4.14)
Lvg×nonbank					0.05***	0.05***
					(5.58)	(4.62)
Observations	6,584	6,584	6,584	6,584	6,584	6,584
Adj. R-squared	0.108	0.237	0.105	0.232	0.116	0.255
Episode FE	N	Y	N	Y	N	Y

Panel B: Realized returns during stress episodes						
Measure:	ΔCoVaR		eΔCoVaR		MES	
	(1)	(2)	(3)	(4)	(5)	(6)
Measure×bank	-5.33***	-5.29***	-1.16*	-1.23*	0.02	-0.88***
	(-12.35)	(-6.80)	(-1.85)	(-1.87)	(0.07)	(-4.61)
Measure×nonbank	-2.80***	-3.42***	-1.22***	-1.87***	-0.70***	-0.96***
	(-5.26)	(-4.89)	(-4.07)	(-6.26)	(-3.60)	(-4.79)
Observations	12,370	12,370	12,370	12,370	12,370	12,370
Adj. R-squared	0.007	0.140	0.004	0.140	0.002	0.139
Episode FE	N	Y	N	Y	N	Y
Measure:	SRISK/ME		Lvg		MES (+Lvg)	
	(7)	(8)	(9)	(10)	(11)	(12)
Measure×bank	-0.98**	-1.36***	-0.07**	-0.07***	-0.70***	-0.90***
	(-2.41)	(-3.08)	(-2.34)	(-3.07)	(-2.83)	(-3.77)
Measure×nonbank	-0.82***	-0.72***	-0.07***	-0.06***	0.37	-0.55**
	(-3.96)	(-2.94)	(-4.00)	(-3.11)	(1.27)	(-2.14)
Lvg×bank					-0.05**	-0.07***
					(-2.01)	(-2.87)
Lvg×nonbank					-0.07***	-0.05***
					(-3.98)	(-2.72)
Observations	6,584	6,584	6,584	6,584	6,584	6,584
Adj. R-squared	0.001	0.193	0.001	0.193	0.002	0.195
Episode FE	N	Y	N	Y	N	Y

Table 16: Predictive regression of realized volatility (Panel A) and realized returns (Panel B) during 23 BB stress episodes (1895-2023). The dependent variable is the market outcome of a financial institution during stress episodes. The systemic risk measures and the control variable for the firm size are all measured the quarter before the episode starts. Stress episodes are defined in Tables 14 (Panel B) and 15 (Panel B). Coefficient estimates on SRISK/ME and Lvg are multiplied by 100. t-statistics based on standard errors clustered at the firm level are reported in parentheses. Sample (Columns (1)-(6)): 4,161 financial institutions, including 2,037 banks. Sample (Columns (7)-(12)): 2,261 financial institutions, including 826 banks.

Panel A: Realized volatility during stress episodes									
Measure:		ΔCoVaR			eΔCoVaR			MES	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Measure	0.25***	-0.28**	0.28***	0.41***	0.22***	0.37***	0.31***	0.21***	0.30***
	(2.61)	(-2.10)	(2.76)	(8.25)	(3.57)	(7.60)	(14.76)	(8.28)	(14.09)
Measure×dd		0.39***			0.12***			0.06***	
		(5.94)			(4.54)			(4.57)	
Measure×bank_crisis			-0.18			0.50***			0.05
			(-0.84)			(2.73)			(0.83)
Observations	12,370	12,370	12,370	12,370	12,370	12,370	12,370	12,370	12,370
Within R-squared	0.029	0.035	0.029	0.045	0.049	0.048	0.060	0.064	0.060
Episode FE	Y	Y	Y	Y	Y	Y	Y	Y	Y

Measure:		SRISK/ME			Lvg			MES (+Lvg)	
	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Measure	0.73***	0.60***	0.70***	0.06***	0.05***	0.05***	0.31***	0.20***	0.31***
	(5.45)	(4.64)	(5.65)	(5.64)	(4.70)	(5.95)	(10.72)	(5.32)	(10.53)
Measure×dd		0.15			0.01			0.06***	
		(1.37)			(0.68)			(3.58)	
Measure×bank_crisis			0.12			0.01			0.01
			(0.41)			(0.45)			(0.14)
Lvg							0.05***	0.05***	0.05***
							(5.43)	(4.79)	(5.91)
Lvg×dd								0.01	
								(0.52)	
Lvg×bank_crisis									0.01
									(0.33)
Observations	6,584	6,584	6,584	6,584	6,584	6,584	6,584	6,584	6,584
Within R-squared	0.100	0.101	0.100	0.094	0.095	0.095	0.125	0.129	0.125
Episode FE	Y	Y	Y	Y	Y	Y	Y	Y	Y

Panel B: Realized returns during stress episodes									
Measure:		ΔCoVaR			eΔCoVaR			MES	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Measure	-4.16***	-3.51***	-4.15***	-1.67***	-2.22***	-1.55***	-0.94***	-1.01***	-0.87***
	(-6.17)	(-4.58)	(-5.61)	(-5.49)	(-5.54)	(-4.85)	(-5.95)	(-5.09)	(-5.01)
Measure×dd		-0.48			0.36**			0.04	
		(-1.47)			(2.37)			(0.60)	
Measure×bank_crisis			-0.07			-1.56***			-0.54**
			(-0.08)			(-3.35)			(-2.30)
Observations	12,370	12,370	12,370	12,370	12,370	12,370	12,370	12,370	12,370
Within R-squared	0.010	0.011	0.010	0.011	0.011	0.011	0.011	0.011	0.011
Episode FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Measure:		SRISK/ME			Lvg			MES (+Lvg)	
	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Measure	-0.84***	-0.82***	-0.87***	-0.06***	-0.07***	-0.06***	-0.69***	-0.80***	-0.58**
	(-3.77)	(-3.42)	(-4.06)	(-3.65)	(-3.51)	(-4.08)	(-3.29)	(-2.70)	(-2.47)
Measure×dd		-0.02			0.01			0.05	
		(-0.12)			(0.49)			(0.60)	
Measure×bank_crisis			0.19			0.02			-0.61**
			(0.47)			(0.72)			(-2.55)
Lvg							-0.05***	-0.06***	-0.06***
							(-3.19)	(-3.37)	(-3.80)
Lvg×dd								0.01	
								(0.72)	
Lvg×bank_crisis									0.02
									(0.75)
Observations	6,584	6,584	6,584	6,584	6,584	6,584	6,584	6,584	6,584
Within R-squared	0.012	0.012	0.012	0.011	0.011	0.011	0.014	0.014	0.015
Episode FE	Y	Y	Y	Y	Y	Y	Y	Y	Y

Table 17: Predictive regression of realized volatility (Panel A) and realized returns (Panel B) during 23 BB stress episodes (1895-2023). The dependent variable is the market outcome of a financial institution during stress episodes. The systemic risk measures and the control variable for the firm size are all measured the quarter before the episode starts. Stress episodes are defined in Tables 14 (Panel B) and 15 (Panel B). The drawdown variable "dd" is the change in the credit spread over the episode. The credit spread is the GZ spread for the period 1973-2023, the Moody's spread for 1922-1972, and the average yield on U.S. railroad bonds for 1895-1921. The variable "bank_crisis" indicates the Great Depression and the GFC. Coefficient estimates on SRISK/ME and Lvg are multiplied by 100. t-statistics based on standard errors clustered at the firm level are reported in parentheses. Sample (Columns (1)-(9)): 4,161 financial institutions, including 2,037 banks. Sample (Columns (10)-(18)): 2,261 financial institutions, including 826 banks.

6 Systemic Risk Measures: Definitions and Estimation

In this section, we provide additional details on the derivation and estimation of the △CoVaR, the exposure △CoVaR (e△CoVaR), the Marginal Expected Shortfall (MES), and the Long-Run Marginal Expected Shortfall (LRMES). We also decompose the ratio of SRISK to the firm market capitalization into a linear function of LRMES and market leverage. All systemic risk measures are defined in Section 2.3 of the paper.

6.1 △CoVaR and e△CoVaR

The "return loss" for firm i and the return loss for a financial index are respectively denoted by X^i and X^{system}. The predicted value from the quantile regression of financial system portfolio return losses on the losses of firm i gives the value-at-risk (VaR) of the financial system conditional on X^i:

$$CoVaR_q^{system|X^i} = \hat{X}_q^{system|X^i} = \hat{\alpha}_q^i + \hat{\beta}_q^i X^i, \tag{3}$$

where $\hat{X}^{system|X^i}$ denotes the predicted value for a q%-quantile of the financial system conditional on a return realization X^i of institution i.

Using the predicted value of $X^i = VaR_q^i$ yields the $CoVaR_q^i$ measure (precisely, $CoVaR_q^{system|X^i=VaR_q^i}$)

$$CoVaR_q^i = VaR^{system|X^i=VaR_q^i} = \hat{\alpha}_q^i + \hat{\beta}_q^i VaR_q^i \tag{4}$$

and

$$\Delta CoVaR_q^i = CoVaR_q^i - CoVaR_q^{system|VaR_{50}^i} = \hat{\beta}_q^i (VaR_q^i - VaR_{50}^i). \tag{5}$$

Similarly, the $eCoVaR_q^i$ measure is obtained from:

$$eCoVaR_q^i = VaR^{i|X^{syst}=VaR_q^{syst}} = \hat{\alpha}_q^{i,e} + \hat{\beta}_q^{i,e} VaR_q^{syst} \tag{6}$$

and

$$e\Delta CoVaR_q^i = eCoVaR_q^i - eCoVaR_q^{i|VaR_{50}^{syst}} = \hat{\beta}_q^{i,e}(VaR_q^{syst} - VaR_{50}^{syst}). \tag{7}$$

The estimation of $\Delta CoVaR_q^i$ requires the estimation of $\hat{\beta}_q^i$ from quantile regressions of a financial index return losses on the firm return losses, and the estimation of the VaR at q% (VaR_q^i) and the median (VaR_{50}^i) of the firm return loss. The estimation of $e\Delta CoVaR_q^i$ requires the estimation of $\hat{\beta}_q^{i,e}$ from quantile regressions of a firm return losses on the financial index return losses, and the estimation of the VaR at q% (VaR_q^{syst}) and the median (VaR_{50}^{syst}) of the financial index return loss. We estimate $\hat{\beta}_q^i$, $\hat{\beta}_q^{i,e}$, VaR_q^i, VaR_q^{syst}, VaR_{50}^i, and VaR_{50}^{syst} at the end of each month based using a sample of ten years of weekly (for CRSP) or monthly (for GFD) logarithmic return losses on firm stock prices and a financial index, available up to each month. Measures estimated using monthly GFD returns are then rescaled to a weekly frequency by dividing them by $\sqrt{4}$, assuming an average of four weeks per month. We use $q = 95\%$ and, for the financial index, we use the CRSP financial index for the modern sample and the GFD financial index for the early sample.

6.2 LRMES and MES

The firm and market logarithmic returns are respectively denoted by $r_{it} = -X^i$ and r_{mt}, and are assumed to be i.i.d. from a bivariate normal distribution with zero mean. The market volatility, firm volatility, and correlation parameters are denoted respectively as σ_m, σ_i, and ρ_i. MES can be approximated (Brownlees & Engle, 2017, p. 55) by

$$MES_{it} = -\beta_i \mathrm{E}\left(r_{mt+1} | r_{mt+1} < c\right) \tag{8}$$

where $\beta_i = \rho_i \frac{\sigma_i}{\sigma_m}$ (market beta of institution i), and

$$\mathrm{E}\left(r_{mt+1} | r_{mt+1} < c\right) = -\sigma_m \frac{\phi(c/\sigma_m)}{\Phi(c/\sigma_m)}, \tag{9}$$

with $\phi(.)$ and $\Phi(.)$ denoting, respectively, the pdf and cdf of a standard normal distribution. From that definition, LRMES is obtained from $LRMES_{it} = \sqrt{h} MES_{it}$, with h the forecast horizon, and $c = \log(1 + C)/\sqrt{h}$.

The estimation of MES and LRMES requires the estimation of the firm's market beta β_i and the market volatility σ_m. The firm's market beta β_i and the market volatility σ_m are estimated at the end of each month using a sample of ten years of weekly (for CRSP) or monthly (for GFD) logarithmic return losses on firm stock prices and a financial index, available up to each month. MES estimated using monthly GFD returns is then rescaled to a weekly frequency by diving it by $\sqrt{4}$, assuming an average of four weeks per month. We use the S&P500 index as the market index for the modern sample, and the GFD financial index for the early sample due to the unavailability of the S&P500 index for that period. In addition, we set $C = -0.4$ (-40%), and $h = 24$ weeks (6 months) in accordance with the choices made at NYU Stern VLAB (vlab.stern.nyu.edu/srisk).

6.3 SRISK Decomposition

SRISK is the expected capital shortfall (in U.S. dollars) of the firm in the scenario of a 40% loss on the market index over six months. It is a function of LRMES, the market capitalization of the firm (ME_{it}) and its total non-equity liabilities (D_{it}):

$$SRISK_{it} = (k - 1) * ME_{it}(1 - LRMES_{it}) + kD_{it} \tag{10}$$

where k is a prudential capital ratio such that $k \leq ME_{it}/\left(ME_{it} + D_{it}\right)$.

In the paper, we use the ratio of SRISK scaled by the market capitalization of the firm (ME_{it}):

$$\frac{SRISK_{it}}{ME_{it}} = (k - 1) * (1 - LRMES_{it}) + k\frac{D_{it}}{ME_{it}} \tag{11}$$

which can be decomposed into

$$\frac{SRISK_{it}}{ME_{it}} = (k - 1) * (1 - LRMES_{it}) + k\left(Lvg_{it} - 1\right) \tag{12}$$

where $Lvg_{it} = \left(ME_{it} + D_{it}\right)/ME_{it}$. From eq. (12), the ratio of SRISK to market capitalization is a linear function of LRMES and market leverage (Lvg_{it}). In addition, LRMES is a function of MES, given by $LRMES_{it} = \sqrt{h} MES_{it}$. Regression specifications that include (LR)MES and market leverage as independent variables instead of SRISK alone allow for more flexibility in the weighting of the two components in eq. (12).

References

Adrian T, Brunnermeier MK. 2016. CoVaR. *The American Economic Review* 106(7):1705–1741

Brownlees C, Engle RF. 2017. SRISK: A conditional capital shortfall measure of systemic risk. *The Review of Financial Studies* 30(1):48–79

Bry G, Boschan C. 1971. Cyclical analysis of times series: selected procedures and computer programs. National Bureau of Economic Research

Correia SA, Luck S, Verner E. 2024. Failing banks. National Bureau of Economic Research Working Paper 32907

Fahlenbrach R, Prilmeier R, Stulz RM. 2018. Why does fast loan growth predict poor performance for banks? *The Review of Financial Studies* 31(3):1014–1063

Gilchrist S, Zakrajšek E. 2012. Credit spreads and business cycle fluctuations. *The American Economic Review* 102(4):1692–1720

Harding D, Pagan A. 2002. Dissecting the cycle: a methodological investigation. *Journal of monetary economics* 49(2):365–381

Macaulay FR. 1938. The movement of interest rates, bond yields and stock prices in the United States since 1856. National Bureau of Economic Research

Markham JW. 2022. From the age of derivatives into the new millennium: A financial history of the United States 1970–2001. Routledge

Reinhart CM, Rogoff KS. 2009. This time is different: Eight centuries of financial folly. Princeton, NJ: Princeton University Press

Reinhart CM, Rogoff KS. 2011. From financial crash to debt crisis. *The American Economic Review* 101(5):1676–1706