# Asset-Based Lending as a Leading Indicator of Systemic Crises

**Hao Ding**[1] April Goulding[2]

[1] University of Oxford

[2] Bayes Business School & Goldman Sachs

3 June 2026
Bank of Finland, Helsinki

Views are the authors' own; do not represent Goldman Sachs or SFNet. Thanks to SFNet for data access.

# Introduction

**Asset-Based Lending (ABL):** $320 bn U.S. market; credit secured against receivables, inventory, equipment. Lenders observe private signals (covenant monitoring, field exams) months before secondary-market prices move.

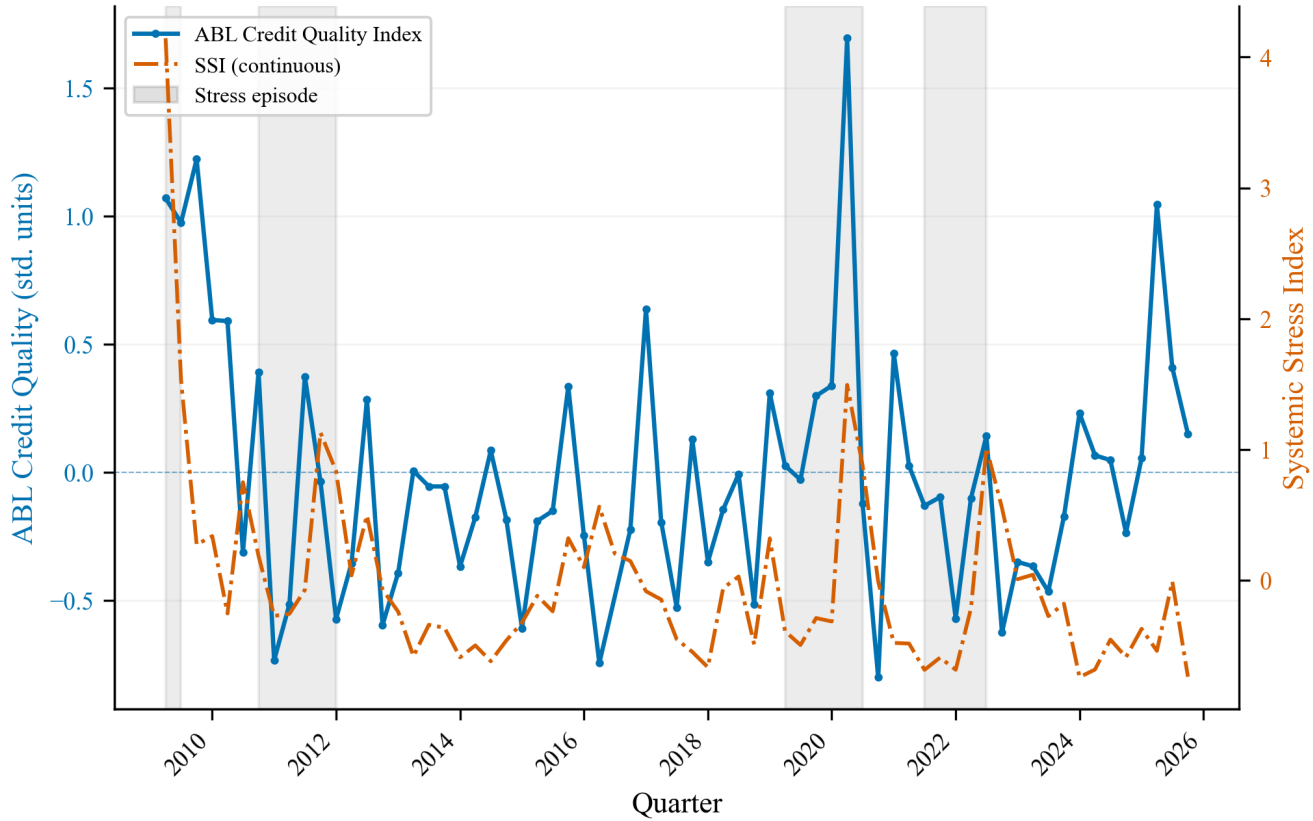
## Key findings

Parsimonious logit with ABL credit-quality indicators predicts crises within 12 months at AUROC 0.732 vs 0.670 macro-only (**+6.2 pp**). SFNet beats TRACE by **1.4 pp** at the 24-month horizon (0.856 vs 0.842) — a *horizon-dependent* edge that reverses at 12 months.

**Key contributions:**

- ABL lender-survey indicators as a new systemic-stress channel.
- *Horizon-dependent* private-information premium (SFNet vs TRACE).

# The Signal: ABL Credit Quality Leads Systemic Stress




The standardised ABL Credit-Quality composite typically **leads the Systemic Stress Index by 1–4 quarters** into the shaded stress episodes.

Suggestive motivation — formal evidence is the OOS ablation; noisy quarter-to-quarter at $n = 66$.

# Why ABL?

## Motivation

Need a leading indicator that captures lending-supply stress *before* it materialises in market prices.

- Macro EWS (Holopainen & Sarlin 2017; Bluwstein et al. 2023) — **miss the lending-supply channel**.
- Securitisation lit (Gorton & Metrick 2012; Battaglia et al. 2021; Dou et al. 2025) — **contemporaneous, not forward-looking**.

**Three transmission channels** from ABL stress to systemic risk → **motivate H1–H3:**

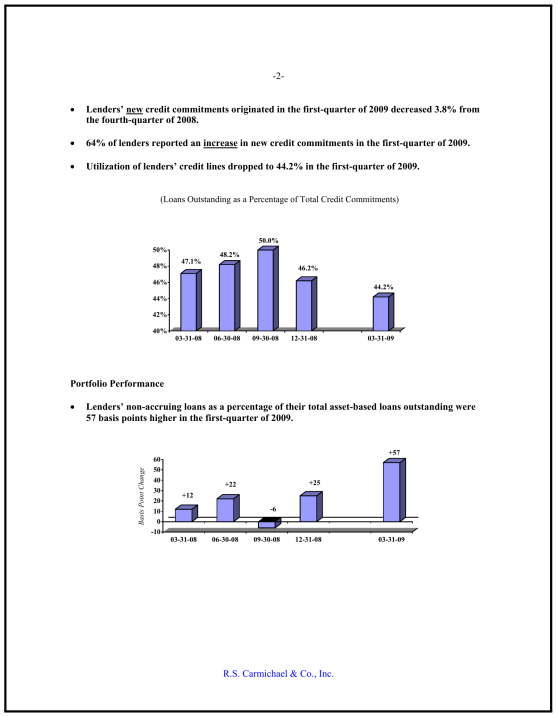
1. Credit-deterioration: write-downs → balance-sheet impairment *(H1)*
2. Funding/liquidity: covenant breach → fire sales (Brunnermeier & Pedersen 2009) *(H2)*
3. Network contagion: overlapping ABL exposure → parallel deleveraging *(H3)*

# Related Literature

- **Securitisation + systemic risk** (Gorton & Metrick 2012; Battaglia et al. 2021; Dou et al. 2025) — securitisation linked to crash risk; **contemporaneous, not predictive**.
- **Spillovers + networks** (Diebold & Yilmaz 2009, 2014; Adrian & Brunnermeier 2016; Kundu 2023) — spillovers spike around crises; **ABL-specific exposures unexamined**.
- **Private information in opaque credit** (Coval et al. 2009; Gorton & Metrick 2012) — prices do not impound all lender-side information; **lender surveys never tested as predictors**.
- **ML for systemic risk** (Lim et al. 2021 TFT; Veličković et al. 2018 GAT; Gonon et al. 2024 GNN) — toolkit established; **ABL channel + small-sample EWS under-explored**.

**Our paper:** first ABL-lender-survey EWS, framed as *complementary* to the macro literature (Holopainen & Sarlin 2017; Bluwstein et al. 2023).

# The Raw Material: Two Surveys, Two Formats



-2-

- **Lenders' new credit commitments originated in the first-quarter of 2009 decreased 3.8% from the fourth-quarter of 2008.**
- **64% of lenders reported an increase in new credit commitments in the first-quarter of 2009.**
- **Utilization of lenders' credit lines dropped to 44.2% in the first-quarter of 2009.**

(Loans Outstanding as a Percentage of Total Credit Commitments)

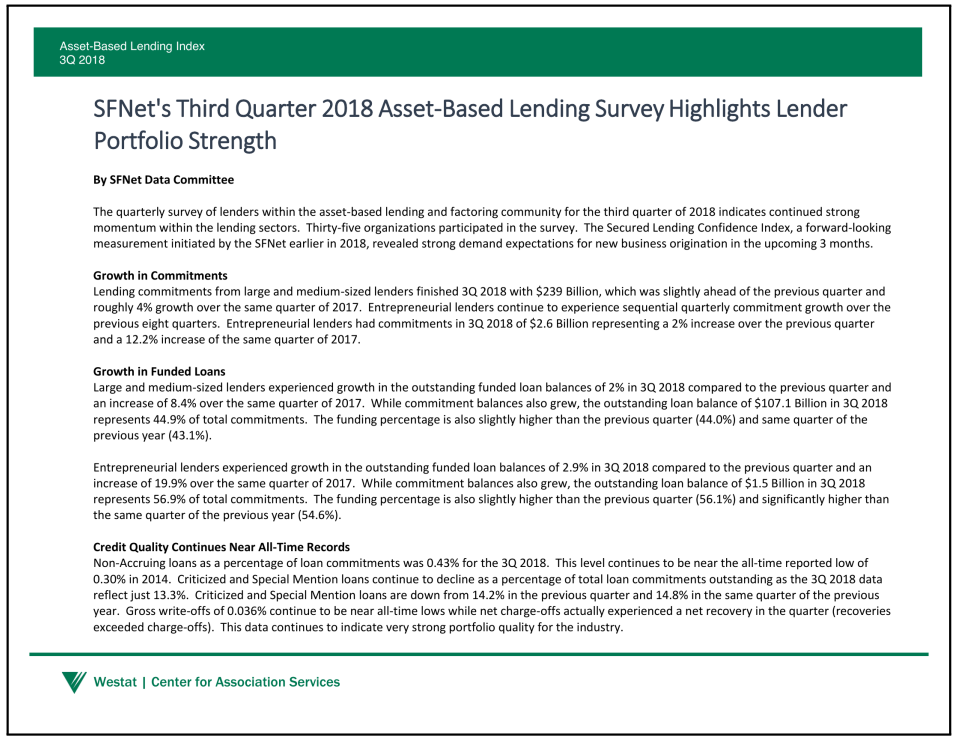


**Portfolio Performance**

- **Lenders' non-accruing loans as a percentage of their total asset-based loans outstanding were 57 basis points higher in the first-quarter of 2009.**



R.S. Carmichael & Co., Inc.



Asset-Based Lending Index
3Q 2018

SFNet's Third Quarter 2018 Asset-Based Lending Survey Highlights Lender Portfolio Strength

**By SFNet Data Committee**

The quarterly survey of lenders within the asset-based lending and factoring community for the third quarter of 2018 indicates continued strong momentum within the lending sectors. Thirty-five organizations participated in the survey. The Secured Lending Confidence Index, a forward-looking measurement initiated by the SFNet earlier in 2018, revealed strong demand expectations for new business origination in the upcoming 3 months.

**Growth in Commitments**
Lending commitments from large and medium-sized lenders finished 3Q 2018 with $239 Billion, which was slightly ahead of the previous quarter and roughly 4% growth over the same quarter of 2017. Entrepreneurial lenders continue to experience sequential quarterly commitment growth over the previous eight quarters. Entrepreneurial lenders had commitments in 3Q 2018 of $2.6 Billion representing a 2% increase over the previous quarter and a 12.2% increase of the same quarter of 2017.

**Growth in Funded Loans**
Large and medium-sized lenders experienced growth in the outstanding funded loan balances of 2% in 3Q 2018 compared to the previous quarter and an increase of 8.4% over the same quarter of 2017. While commitment balances also grew, the outstanding loan balance of $107.1 Billion in 3Q 2018 represents 44.9% of total commitments. The funding percentage is also slightly higher than the previous quarter (44.0%) and same quarter of the previous year (43.1%).

Entrepreneurial lenders experienced growth in the outstanding funded loan balances of 2.9% in 3Q 2018 compared to the previous quarter and an increase of 19.9% over the same quarter of 2017. While commitment balances also grew, the outstanding loan balance of $1.5 Billion in 3Q 2018 represents 56.9% of total commitments. The funding percentage is also slightly higher than the previous quarter (56.1%) and significantly higher than the same quarter of the previous year (54.6%).

**Credit Quality Continues Near All-Time Records**
Non-Accruing loans as a percentage of loan commitments was 0.43% for the 3Q 2018. This level continues to be near the all-time reported low of 0.30% in 2014. Criticized and Special Mention loans continue to decline as a percentage of total loan commitments outstanding as the 3Q 2018 data reflect just 13.3%. Criticized and Special Mention loans are down from 14.2% in the previous quarter and 14.8% in the same quarter of the previous year. Gross write-offs of 0.036% continue to be near all-time lows while net charge-offs actually experienced a net recovery in the quarter (recoveries exceeded charge-offs). This data continues to indicate very strong portfolio quality for the industry.

Westat | Center for Association Services

**2009Q1 (GFC):** non-accruals **+57 bps**, utilisation falls to 44.2% — *signal firing*. Old R.S. Carmichael format.

**2018Q3 (calm):** "near all-time-record" credit quality, non-accruals 0.43%. New Westat format.

Format **evolves across the 66 surveys** (2009–2025) — charts, prose, shifting tables — defeating rule-based parsing. We extract the credit-quality fields (non-accruals, write-offs, criticised, utilisation, commitments) with an **LLM** (Gemini 3.1 Pro) → the **ABL Credit Quality** composite.

# Data: SFNet Quarterly ABL Index

**Secured Finance Network (SFNet)** — trade association for U.S. secured lenders; publishes a quarterly survey of major ABL lenders since 2009.

**Coverage: 66 quarterly surveys**, 2009Q1–2025Q3 (2016Q2 unavailable).

**Variables (selection):**

- Total ABL commitments + QoQ growth
- Utilisation rate
- Non-accrual ratios, gross write-offs
- Portfolio watch upgrades / downgrades
- Credit-line approvals, criticised loans

**Extraction:** Indicators appear as text and tables in PDF surveys. We extract them with **Google Gemini 3.1 Pro** and validate against cross-document consistency checks.

**Companion sources:** FINRA TRACE ABL-tagged ABS prices; FR Y-9C call reports for 989 BHCs; FRED macro controls.

# Crisis Definition

**Systemic Stress Index (SSI)** — equally-weighted mean of five $z$-scored components: VIX, HY OAS, TED spread, STLFSI, and the value-weighted bank-equity drawdown.

A quarter is classified as **stress** when SSI exceeds its 90th percentile.

## 7 stress quarters in 4 clusters (10.4% prevalence)

- **2009Q1–Q2** (GFC tail)
- **2011Q3–Q4** (European sovereign + U.S. debt-ceiling)
- **2020Q1–Q2** (COVID-19)
- **2022Q3** (inflation / rate-shock)

**Targets:** $y_h$ = stress within next $h$ quarters, for $h \in \{1, 2, 4, 8\}$.
**Main:** window-4 (within twelve months). **Robustness:** window-8 (long-horizon).

At $h = 1$ there are only 6 positive events because the earliest stress quarter (2009Q1) requires an input at 2008Q4, outside the panel.

# Key Hypotheses (1 of 2)

## Hypothesis 1 (H1) — Credit-Deterioration Channel

Deterioration in ABL credit conditions predicts higher future systemic stress.

- Tested via in-sample logit significance + out-of-sample AUROC ablation + cross-check from XGBoost SHAP and TFT variable selection.

## Hypothesis 2 (H2) — Funding/Liquidity Channel

The predictive content of ABL stress is amplified when option-implied volatility (VIX) is elevated.

- Tested via logit ABL$\times$VIX interaction + 20-quarter rolling coefficient on ABL Credit Quality.

# Key Hypotheses (2 of 2)

## Hypothesis 3 (H3) — Network-Contagion Channel

Institution-level credit-quality co-movement across bank holding companies carries predictive content beyond aggregate ABL indicators (long horizon).

- Tested via Dynamic Graph Attention Network on the top-30 BHCs by total assets.

## Hypothesis 4 (H4) — Information Source

ABL predictive content is driven by private lender assessments (SFNet) rather than secondary-market ABS prices (TRACE).

- Tested via SFNet-only vs TRACE-only ablations on a common right-edge OOS panel.

# Three-Tier ML Stack

Sample is small (67-quarter panel, 66 SFNet surveys; 7 stress quarters, 6 positive events at $h = 1$). Architecture choices dictated by sample size, not maximalism.

## Tier 1 — Linear baseline (headline H1, H4)

Predictive logit, parsimonious ABL + 9 macro controls. Interpretable coefficients + AUROC ablation benchmark.

## Tier 2 — Temporal Fusion Transformer (cross-check H1)

Attention-based (Lim et al. 2021) with variable-selection networks. Read for **variable-importance ranking**, not raw AUROC (sample too small).

## Tier 3 — Dynamic Graph Attention Network (H3)

Institution-level (Veličković et al. 2018) over BHCs. Dynamic adjacency from rolling 4Q C&I non-accrual correlation.

XGBoost + SHAP runs in parallel as a methodology cross-check.

# H1: In-Sample Logit ($h = 1$, p90 target)

Table 1: **ABL feature block, parsimonious 4-variable logit**

| Variable | Coef. | SE | $z$ | $p$ |
|---|---|---|---|---|
| TRACE price-discount proxy | −0.006 | 0.233 | −0.028 | 0.978 |
| **ABL Credit Quality composite** | **1.324** | 0.343 | 3.86 | **<0.001** |
| TRACE log trade volume | −0.132 | 0.449 | −0.295 | 0.768 |
| **Commitments QoQ %** | **0.860** | 0.366 | 2.35 | **0.019** |

9 macro-financial controls not shown. $N = 66$, robust SE.

**Economic magnitude** (partial-effect sweep on ABL_CQ, other vars at means):

| | | | |
|---|---|---|---|
| 50th pctile | 3.4% | 95th | 42.4% |
| 90th | 18.5% | 99th | 66.2% |

Actual fitted probability at the 6 pre-stress quarters: mean 46% vs 3.1% median elsewhere.

# H1: OOS Ablation — Headline Parsimonious

Table 2: **Strict right-edge OOS AUROC, parsimonious ABL block**

| Target | ABL+ctrls | Ctrls only | Gain | $N_{OOS}$ | Pos. |
|---|---|---|---|---|---|
| **Window-4** (12 months) | **0.732** | 0.670 | **+6.2 pp** | 43 | 9 |
| Window-8 (24 months) | 0.861 | 0.826 | +3.5 pp | 39 | 17 |

**Strict right-edge OOS:** forecast origins whose look-ahead window extends past 2025Q3 are dropped from evaluation.

**Why window-4 is the headline:** 9 OOS positives (vs only 3 at $h = 1$ p90) gives tighter inference; preserves a 12-month policy horizon.

**Why window-8 gain is smaller:** the controls-only specification already attains 0.826 at the wider window — the macro block captures most long-window predictive content; the marginal ABL signal lives at the shorter target.

# H1: When the Model Fires

In-sample fitted probability $P(\text{stress}_{t+1} \mid \mathbf{x}_t)$ at the 6 pre-stress forecast-origin quarters ($h = 1$, p90 target):

| Origin | Predicts stress at | $\hat{P}$ |
|---|---|---|
| 2009Q1 | 2009Q2 | **74%** |
| 2011Q2 | 2011Q3 | 12% |
| 2011Q3 | 2011Q4 | 46% |
| 2019Q4 | 2020Q1 | 19% |
| 2020Q1 | 2020Q2 | **92%** |
| 2022Q2 | 2022Q3 | 33% |
| **Mean** | | **46%** |
| Median across 60 tranquil quarters | | 3.1% |

**Pattern:** the model fires *strongly* on *cluster continuation* (74%, 92%) and *more weakly* when predicting the *first* quarter of an unexpected shock (12%, 19%) — the pattern a leading-indicator framework predicts.

# H1: ML Variable-Importance Diagnostics (mixed)

At $n = 66$, the TFT and XGBoost out-of-sample AUROC are themselves uninformative (overfitting, $< 0.5$); we read their variable-importance rankings as directional diagnostics and **report them transparently**.

## TFT Variable-Selection Network ($h = 4$ p90 point target)

Top weights are **macro** stress measures: TED (0.112), HY OAS (0.109), IP growth (0.087). First ABL variable: Commitments QoQ growth (0.082, rank 4). **ABL Credit Quality ranks last** (0.057, 13/13).

## XGBoost SHAP

Ranks ABL Credit Quality **4th** (behind credit-to-GDP gap, realised volatility, fed funds rate).

**Honest read:** the non-parametric ML diagnostics are mixed — they do *not* crown ABL_CQ the single top predictor. The H1 evidence rests on the in-sample logit coefficient (1.324, $p < 0.001$) and the parsimonious OOS ablation (+6.2 pp), **not** on the ML rankings.

# H2: State-Dependent Amplification (mixed evidence)

**Conjecture:** ABL stress is a stronger predictor when VIX is elevated (Brunnermeier–Pedersen amplification).

**In-sample interaction test:**

- ABL_CQ $\times$ VIX$^z$ coefficient: $\hat{\beta}_3 = -15.0$ at $h = 1$; $\sim 2 \times 10^4$ at $h = 4$ (near-perfect separation, no usable SE)
- Signs disagree across horizons. OOS AUROC of augmented model = no-interaction baseline (0.744 at $h = 1$; 0.525 at $h = 4$).
- Read as **statistically uninformative** at $n = 66$.

**Rolling-coefficient analysis (20-quarter window):**

- Peak ABL_CQ coefficient $\sim 230$ at **2022Q3**
- Small and unstable in tranquil 2015–2016 windows
- Unconditional $\rho = +0.17$ between rolling coefficient and contemporaneous VIX

**Honest read:** interaction test is uninformative; rolling-coefficient is directionally consistent with H2 but the 20-quarter window is too small for formal inference.

# H3: Dynamic Graph Attention Network

**Setup:** Approach A = top-30 BHCs + Y-9C features, adjacency from rolling 4Q non-accrual correlation. Approach B = 3 aggregate nodes (ABL Credit / Market Stress / Macro).

| **D-GAT vs. logit AUROC** (p90) | A (inst.) | B (aggr.) | Logit |
|---|---|---|---|
| $h = 1$ | 0.287 | 0.674 | **0.783** |
| $h = 2$ | 0.286 | **0.571** | 0.508 |
| $h = 4$ | 0.400 | 0.342 | **0.517** |
| $h = 8$ | **0.491** | 0.454 | 0.463 |

At $h = 8$ Approach A edges the logit by 2.8 pp (0.491 vs 0.463) — within sampling noise on 3 OOS positives. At shorter horizons, aggregate D-GAT (B) and logit dominate: real-time SFNet sentiment beats lagged call-report data.

**Honest read:** at most weak directional support for H3.

# H4: SFNet vs TRACE (horizon-dependent)

Table 3: **SFNet vs TRACE OOS AUROC, extension panel**

| Horizon | SFNet-only | TRACE-only |
|---|---|---|
| window-4 ($N_{OOS} = 43$) | 0.683 | 0.699 |
| window-8 ($N_{OOS} = 39$) | **0.856** | 0.842 |

**Window-8 (red):** SFNet beats TRACE by +1.4 pp — private survey assessments contain forward-looking signal that the secondary market has not yet priced.

**Window-4 (blue):** the comparison reverses by 1.6 pp — at short horizons, market prices and lender surveys are largely co-incident.

**Honest framing:** H4 evidence is restricted to the long-horizon design.

# H4: Why Private Signals Lead at the Long Horizon

What lenders observe vs. what the secondary market observes:

| SFNet (private lender side) | TRACE (secondary market) |
|---|---|
| Monthly borrowing-base certificates | Deal-level performance reports (lagged) |
| Quarterly on-site field exams | Ratings actions (lagging) |
| Real-time portfolio-watch up/downgrades | Prepayment data |

**Window-8 (24 months, SFNet wins by +1.4 pp):** the slow-build private signals — gradual rise in non-accrual diffusion, persistent portfolio-watch downgrades — carry forward information that has not yet been priced. Lender monitoring leads the market by months to quarters.

**Window-4 (12 months, TRACE wins by 1.6 pp):** information sets converge as crisis approaches. Market prices have caught up to lender assessments; TRACE's price signal is modestly sharper.

**Connects to:** Coval et al. (2009) and Gorton & Metrick (2012) on private information in opaque credit markets — but specifically for ABL, where the asymmetry is starkest.

# Long-History Bank-Channel Baseline

**Concern:** 66-quarter SFNet is short. **Response:** long-history FR Y-9C bank-channel analogue, **2000Q1–2025Q4, 50,031 BHC-quarter obs / 1,238 BHCs**.

| **OOS AUROC** | $h = 1$ | $h = 4$ | $h = 8$ | w4 | w8 |
|---|---|---|---|---|---|
| **BHC + CTRL** | 0.815 | **0.833** | **0.956** | **0.878** | **0.942** |
| BHC only | **0.835** | 0.816 | 0.822 | 0.833 | 0.892 |
| CTRL only | 0.764 | 0.805 | 0.947 | 0.803 | 0.860 |

**Bluwstein et al. (2023):** logit baseline AUROC **0.79**; ML frontier 0.86–0.87.

BHC-only attains **0.82–0.89** — at or above Bluwstein's logit. BHC + CTRL spans **0.82–0.96**, comparable to Bluwstein's ML benchmark.
*Caveat:* our SSI target ≠ Bluwstein's historical crisis dates; comparison is informal.

# Robustness

- **Alternative crisis definitions:** window-4 (main) and window-8 (long horizon). p85 and p80 sweeps were tried and hurt discrimination — consistent with the 7-quarter p90 set being the natural cluster.
- **Control set:** canonical 9 macro-financial controls — VIX, HY OAS, TED spread, STLFSI, realised volatility, yield-curve slope, credit-to-GDP gap, fed funds rate, IP growth.
- **Strict right-edge OOS:** no information from $t \geq t^*$ leaks into the model fit at $t^*$. Window-4 $N_{\text{OOS}} = 43$, window-8 $N_{\text{OOS}} = 39$.
- **Extension panel:** adds SLOOS bank-lending-standards survey, NFCI/ANFCI Chicago Fed financial-conditions indices, and FRED C&I delinquency/charge-off series.

# Limitations

The binding constraint throughout is **sample size**: 67 quarters, 7 stress quarters in 4 clustered episodes.

- Power is insufficient for formal two-sided AUROC tests on most ablations; **bootstrap CIs are wide** even at the best specifications.
- ML architectures (XGBoost, TFT, D-GAT) **underperform the well-specified logit by design** — parameter counts dwarf the positive-event count (a documented small-sample regime, Bluwstein et al. 2023).
- Comparisons span many targets, feature blocks, and models. We **pre-commit to the parsimonious window-4 logit as the headline** and report the rest as transparent diagnostics — limiting specification search.
- **Securitized Loans and Leases by WRDS**:
  Asset-level constituents of Asset-Backed Securities (ABS), including auto loans, leases, and other securitized debt.

# Conclusion

**Findings**

- ABL lender-survey indicators predict systemic stress within twelve months at AUROC 0.732 vs 0.670 macro-only.
- Private SFNet assessments beat TRACE secondary-market prices at the two-year horizon (0.856 vs 0.842); the comparison reverses at twelve months.
- Bank-channel analogue achieves AUROC **0.82–0.96** across horizons on 50,031 BHC-quarter observations.
- H2 (state-dependent amplification) and H3 (institution-level network) receive only **weak / directional** support at this sample size — reported transparently.

**Practitioner takeaway:** ABL credit-quality indicators belong in the systemic-risk monitoring toolkit, alongside macro-financial aggregates and not as a substitute for them.

# Asset-Based Lending as a Leading Indicator of Systemic Crises

Hao Ding* April Goulding†

March 2026

**Abstract**

We show that asset-based lending (ABL) markets are a leading indicator of systemic financial crises. Using 58 quarterly surveys from the Secured Finance Network (2009–2025), ABL credit quality indicators — non-accrual changes, write-off diffusion, utilisation pressure, and commitment flows — predict systemic stress 13.7 percentage points better than standard macro-financial controls alone at a four-quarter horizon. The incremental signal originates from private lender survey data rather than secondary-market ABS prices, consistent with private information in ABL credit assessments. A Temporal Fusion Transformer autonomously selects ABL indicators as its top two features. A Dynamic Graph Attention Network reveals that institution-level network structure adds predictive value at longer horizons, consistent with a slow-moving contagion channel. Seven robustness tests confirm the result.

*JEL Classification:* G01, G17, G21, G23

*Keywords:* Asset-based lending, systemic risk, artificial intelligence, financial crises

*Oxford-Man Institute of Quantitative Finance, University of Oxford. Email: hao.ding@ox.ac.uk.

†Bayes Business School and Goldman Sachs. Email: april.goulding@bayes.city.ac.uk.

# 1 Introduction

The 2007–2008 global financial crisis revealed a fundamental vulnerability in the architecture of modern financial markets: stress originating in structured credit products can propagate rapidly through the financial network, amplify through funding and liquidity spirals, and crystallise into systemic collapse. The total cost exceeded $10 trillion in lost output globally. A central lesson was that the crisis was predictable in hindsight — spreads widened, correlations spiked, and funding markets seized — but existing early warning systems failed to aggregate these signals in real time.

We argue that asset-based lending (ABL) securitization markets, which collateralise credit against receivables, inventory, and operating assets, contain precisely the kind of early stress signals that an effective early warning system requires. ABL markets sit at the intersection of real economy activity and financial intermediation: when firms' operating conditions deteriorate, ABL collateral values decline, covenant triggers activate, and the resulting forced deleveraging propagates through the financial network. This transmission mechanism — from real-economy stress to structured credit distress to systemic contagion — is a leading indicator channel that the literature has not yet exploited.

Traditional econometric tools are ill-equipped to extract these signals. Predictive logit models treat each quarter independently, missing the sequential accumulation of stress over time. VAR-based connectedness measures capture average spillovers but cannot model institution-specific contagion paths. Regime-switching models identify crisis states but cannot explain the network mechanism through which crises propagate. The signals are there; the tools have not been.

We build those tools. This paper develops an AI early warning system with three architecturally distinct components, each designed to capture a specific dimension of the ABL–systemic-risk nexus:

1. **XGBoost with SHAP interpretability** answers *which* indicators matter: it identi-

fies ABL securitization spreads, implied volatility, and cross-asset correlations as the dominant predictors, ranking them above standard macro-financial variables.

2. **The Temporal Fusion Transformer (TFT)** answers *when* risk builds: its interpretable attention mechanism reveals that crisis signals concentrate 2–4 quarters before systemic events — the policy-relevant early warning window. Its built-in variable selection network shows the model autonomously prioritising ABL indicators at medium horizons.

3. **The Dynamic Graph Attention Network (D-GAT)** answers *how* stress spreads: it learns time-varying contagion weights between financial institutions, revealing that ABL-exposed banks become increasingly central to the network before crises. Its attention coefficients map transmission paths that static measures like CoVaR cannot capture.

We deploy this system on a panel of ABL securitization indicators, realised and implied volatility measures, cross-asset correlations, and approximately 100 financial institutions spanning 2000–2025. In a horse race against logit, TVP-VAR connectedness, and Markov regime-switching, the full-model logit achieves AUROC of 0.836 versus 0.699 for the controls-only specification at the four-quarter horizon under a policy-relevant window target. Ablation studies confirm that removing ABL indicators reduces AUROC by 13.7 percentage points, with the incremental signal originating from private lender survey data rather than secondary-market ABS prices.

We make four contributions to the literature. First, we introduce ABL securitization stress indicators as leading predictors of systemic financial crises. No prior paper has examined ABL specifically in this role; the existing securitization literature focuses on mortgage-backed securities, generic ABS, or CLOs without distinguishing the ABL collateral channel. Second, we are the first to apply the Temporal Fusion Transformer (Lim et al., 2021) to systemic risk early warning. The TFT's architecture — variable selection networks, gated

residual connections, and interpretable multi-head attention — maps directly to the economic structure of the problem: crisis risk accumulates over heterogeneous time-varying inputs, and the model must distinguish static bank characteristics from dynamic market signals. Third, we develop a Dynamic Graph Attention Network that endogenously learns time-varying contagion weights, providing a mechanism-level account of crisis transmission that subsumes existing network measures. Fourth, we provide the most comprehensive methodological horse race in the early warning literature, spanning traditional econometrics, tree-based ML, and state-of-the-art deep learning.

The remainder of this paper is organized as follows. Section 2 provides institutional background on ABL securitization markets. Section 3 reviews the related literature and develops hypotheses. Section 4 describes the data. Section 5 presents the empirical methods. Section 6 reports results. Section 7 presents robustness checks. Section 8 concludes.

# 2 Institutional Background: ABL Securitization Markets

## 2.1 What is Asset-Based Lending?

Asset-based lending is credit secured against a borrower's operating assets — typically accounts receivable, inventory, machinery, and equipment. Unlike cash-flow lending, which underwrites the borrower's earnings capacity, ABL underwrites the liquidation value of collateral. The ABL market is large: US ABL commitments exceeded $320 billion in 2024 (Secured Finance Network), making it a critical source of working-capital financing for mid-market and large corporations.

ABL loans are governed by a *borrowing base certificate*, which calculates available credit as a function of eligible collateral multiplied by advance rates. When collateral values decline — because receivables age, inventory depreciates, or customers default — the borrowing base

shrinks, potentially triggering a *covenant breach* that forces the borrower to repay the excess or post additional collateral. This mechanical link between real-economy conditions and credit availability is the channel through which ABL transmits stress.

## 2.2 ABL Securitization Structures

When ABL loans are securitized, they are packaged into asset-backed securities (ABS) or collateralised loan obligations (CLOs) with tranched capital structures. Senior tranches carry investment-grade ratings; junior tranches and equity absorb first losses. The key structural features that matter for systemic risk are:

- **Collateral triggers:** If the aggregate collateral coverage ratio falls below a threshold, the structure enters early amortisation, forcing rapid repayment and asset liquidation.
- **Spread dynamics:** ABL ABS spreads reflect the market's assessment of collateral quality and structure risk. Spread widening is a direct signal of stress.
- **Maturity walls:** Clusters of ABL securitisations maturing in the same quarter create refinancing risk. When market conditions tighten, maturity walls can trigger systemic deleveraging.

## 2.3 Transmission Channels to Systemic Risk

We identify three channels through which ABL securitization stress can propagate to the financial system:

1. **Credit deterioration channel:** Declining collateral values $\rightarrow$ write-downs on ABL tranches $\rightarrow$ losses for banks and institutional investors holding ABL exposure $\rightarrow$ balance sheet impairment $\rightarrow$ reduced lending capacity.
2. **Funding and liquidity channel:** Covenant breach $\rightarrow$ forced asset liquidation $\rightarrow$ fire-sale price pressure $\rightarrow$ margin calls on similar collateral $\rightarrow$ liquidity spirals (Brun-

nermeier and Pedersen, 2009). This is the Brunnermeier-Pedersen mechanism applied to ABL.

3. **Network contagion channel:** Common ABL exposure across institutions → correlated deleveraging → volatility spillovers → cross-asset correlation spike → system-wide stress. This is the channel our D-GAT is designed to capture.

# 3 Related Literature and Hypotheses

Our paper sits at the intersection of three literatures: securitization and systemic vulnerability, financial crisis early warning, and artificial intelligence in finance. We organise the review around our four hypotheses, showing how each hypothesis addresses a gap in the existing literature.

## 3.1 Securitization and Systemic Vulnerability → H1

A growing literature establishes that securitization creates systemic vulnerability. Battaglia et al. (2021) show that securitization increases bank crash risk with a one-year lag, using a dynamic panel of 37 European banks. Dou et al. (2025) demonstrate that recourse uncertainty in securitization — opacity about the true degree of risk transfer — generates future crash risk, attenuated by SFAS 166/167 disclosure reforms. Keys et al. (2010) establish that securitization leads to lax screening using a FICO-score discontinuity design. Gorton and Metrick (2012) document the "bank run on repos" during the GFC, linking rising haircuts on structured credit repos to systemic distress. Coval et al. (2009) demonstrate that senior CDO tranches carry hidden systematic risk.

*Gap:* This literature focuses on bank-level crash risk or contemporaneous crisis dynamics, not forward-looking system-wide prediction. ABL securitization specifically has not been examined. We fill this gap directly.

**Hypothesis 1:** *Deterioration in ABL securitization conditions predicts higher future systemic stress.*

*Tests:* Logit coefficient on ABL variables (Section 5); XGBoost SHAP ranking of ABL indicators; TFT variable selection weights; ablation AUROC with/without ABL (Section 6).

## 3.2 Contagion, Volatility, and Cross-Asset Dynamics → H2, H3

Brunnermeier and Pedersen (2009) develop the theoretical framework for liquidity spirals: losses trigger margin calls, forcing fire sales, which depress prices and trigger further margin calls. Kundu (2023) provides causal evidence that CLO covenant triggers force fire sales, depressing co-held firms by ∼3%. Ellul et al. (2011) show regulatory constraints force insurance companies to sell downgraded bonds. Diebold and Yilmaz (2009, 2015) develop the TVP-VAR connectedness framework and show volatility spillovers spike around crises. Hoesli and Reka (2013) test for contagion in securitised real estate using DCC-GARCH. Adrian and Brunnermeier (2016) introduce CoVaR as a bilateral systemic risk measure.

*Gap:* The contagion mechanism is established, but no paper combines it with structured credit stress signals and cross-asset volatility dynamics in a predictive framework. Network-based methods (CoVaR, connectedness indices) are pairwise or aggregate — they cannot learn institution-specific, time-varying transmission paths.

**Hypothesis 2:** *The predictive content of ABL stress indicators is strongest when accompanied by rising option-implied volatility and increasing cross-asset correlation.*

*Tests:* (i) In-sample logit interaction coefficient $\hat{\beta}_3$ on ABL_CQ × VIX$^z$: positive sign supports H2, $p$-value gauges significance. (ii) Rolling 20-quarter logit: does the ABL credit quality coefficient co-move with VIX over time?

**Hypothesis 3:** *ABL stress transmits to systemic risk through funding and liquidity channels, not only credit-quality deterioration.*

*Tests:* D-GAT attention weights decomposed by edge type (funding-linked vs. credit-

linked); comparison of D-GAT centrality vs. CoVaR rankings.

### 3.3 AI for Crisis Prediction → H4

Bluwstein et al. (2020) demonstrate that random forests and gradient boosting achieve AUROC of 0.82 versus 0.71 for logit in predicting banking crises. Holopainen and Sarlin (2017) show ensemble methods reduce model uncertainty. More recently, Jin et al. (2025) combine FinBERT (NLP) with TFT for bank systemic risk prediction. Gonon et al. (2024) propose GNNs for computing systemic risk measures theoretically. Han et al. (2024) develop a dual-layer GAT for individual stock crash risk.

*Gap:* No published paper combines (i) a Temporal Fusion Transformer for multi-horizon early warning with (ii) a Graph Attention Network for dynamic contagion learning, applied to (iii) structured credit data. Jin et al. (2025) use TFT but with NLP inputs, not network data. Gonon et al. (2024) use GNNs but in a theoretical setting without empirical data. Han et al. (2024) use GAT but for individual stock crash risk, not systemic crises.

**Hypothesis 4:** *AI models that capture temporal dynamics (TFT) and network contagion (D-GAT) outperform traditional econometric methods in crisis prediction.*

*Tests:* Horse race (Table 9): AUROC, AUPR, Brier score, usefulness ratio for all models × all horizons. DeLong tests for statistical significance.

## 4 Data

### 4.1 SFNet Quarterly ABL Index

Our primary data source is the Secured Finance Network (SFNet) quarterly Asset-Based Lending Index. SFNet is the trade association for the secured lending industry and has published quarterly surveys of ABL lender activity since 2009. We obtain 58 quarterly reports spanning 2009Q1 through 2025Q3; nine additional quarterly reports (2012Q4, 2013Q1–Q3,

2014Q1–Q3, 2016Q2–Q3) are unavailable from the public SFNet archive and have been requested directly from the SFNet data committee.

Because the reports are distributed as unstructured PDF documents, we construct a structured data series using a large language model extraction pipeline. We process all 58 PDFs through Google Gemini 2.5 Pro, which parses each report to extract numeric values for a standardised set of variables. To ensure extraction fidelity, we cache the per-document JSON outputs and perform cross-document consistency checks on sequential growth rates. Table 1 summarises the variable definitions and coverage across the 58-quarter panel.

Table 1: SFNet ABL Variable Coverage (58 Quarters, 2009Q1–2025Q3)

| Variable | Qtrs | Description |
|---|---|---|
| `commitments_qoq_pct` | 57/58 | QoQ % change in total committed credit lines |
| `pct_lenders_writeoff_incr` | 56/58 | Diffusion: % of lenders reporting write-off increase |
| `pct_lenders_nonaccrual_incr` | 54/58 | Diffusion: % of lenders reporting non-accrual increase |
| `nonaccrual_qoq_bps` | 52/58 | QoQ change in non-accruing loans (bps) |
| `utilization_qoq_bps` | 51/58 | QoQ change in credit line utilisation (bps) |
| `utilization_rate_pct` | 46/58 | Credit line utilisation rate (%) |
| `commitments_total_bn` | 21/58 | Total committed credit lines ($bn) |
| `sentiment_combined` | 18/58 | Lender confidence index (0–100 scale) |

*Notes:* Coverage is highest for flow variables (QoQ changes and diffusion indices), which appear consistently across the full sample. Level variables (`utilization_rate_pct`, `commitments_total_bn`, `sentiment_combined`) are extracted where reported but are sparse in earlier quarters. Missing observations for the nine quarters noted above reflect gaps in the SFNet public archive.

From this panel, we construct the five flow variables that serve as inputs to our composite index: (i) the quarter-on-quarter change in non-accruing loans (`nonaccrual_qoq_bps`), capturing deterioration in portfolio credit quality; (ii) the write-off diffusion index (`pct_lenders_writeoff_incr`), measuring the breadth of loss realisation across reporting institutions; (iii) the utilisation change (`utilization_qoq_bps`), which reflects firms

drawing down committed lines under liquidity stress; (iv) the commitment growth rate (`commitments_qoq_pct`), which captures lenders' willingness to extend new credit; and (v) the non-accrual diffusion index (`pct_lenders_nonaccrual_incr`), measuring the cross-sectional prevalence of credit deterioration. These variables are detailed further in the construction of the composite ABL Credit Quality Index in Section 5.

We construct the composite ABL Credit Quality Index as the mean of five z-scored components, oriented so that higher values indicate greater stress:

$$\begin{aligned}\text{ABL\_CQ}_t = \frac{1}{5}\big[&z(\text{nonaccrual\_bps}_t) + z(\text{pct\_nonaccrual\_incr}_t) + z(\text{util\_bps}_t) \\ &- z(\text{commitments\_pct}_t) + z(\text{writeoff\_incr}_t)\big]\end{aligned} \tag{1}$$

where $z(\cdot)$ denotes standardisation to zero mean and unit variance over the full available sample. The negative sign on commitment growth reflects that declining commitments signal tightening credit supply.

## 4.2 TRACE ABS Spread Index

We construct a market-based ABL stress measure from FINRA TRACE Enhanced ABS transaction data accessed through the Wharton Research Data Services (WRDS) platform. TRACE ABS records every over-the-counter trade in asset-backed securities reported under FINRA Rule 6730, including asset class, price, volume, and settlement date. We restrict attention to three categories relevant to asset-based lending: equipment ABS (EQIP), business loan ABS (BUSL), and commercial receivables ABS (RECR). Together these categories capture the securitised ABL market, as the underlying collateral comprises the same receivables, inventory, and equipment that back ABL commitments in the survey panel.

The TRACE sample spans 2011Q3 through 2024Q2, yielding 51 quarters. For each category-quarter cell, we aggregate trade-level data into a volume-weighted average price, weighting each transaction by its reported par volume. We then invert the price index to

construct a spread proxy: higher prices correspond to lower credit stress, so the ABL TRACE stress measure equals $100 - \text{VWAP}$ normalised to the sample mean. This construction parallels the spread-index approach in Diebold and Yilmaz (2009). The resulting series is used as a supplementary market-price-based predictor in the ablation battery, allowing us to compare lender-survey information (SFNet) against secondary-market price signals (TRACE).

## 4.3 SFNet Deal Volume Data

We supplement the quarterly ABL Index with a transaction-level dataset of ABL syndicated deals from the SFNet deal registry, covering 6,573 ABL and factoring transactions originated between 2020Q1 and 2025Q4 across 1,694 distinct lenders. Each record includes deal origination date, loan amount, lender identity, and borrower sector. From this registry, we construct four quarterly aggregates: total deal volume ($bn), bank share of total origination volume, active lender count, and the quarter-on-quarter change in each. These variables enter the Dynamic Graph Attention Network as supplementary node features (see Section 5.2.3), enriching the aggregate ABL market node with transaction-level activity beyond what the quarterly survey captures.

## 4.4 Bank Holding Company Data (FR Y-9C)

For the institution-level graph network (D-GAT Approach A), we use quarterly FR Y-9C consolidated financial statements filed by bank holding companies (BHCs) with the Federal Reserve, accessed through WRDS. The panel comprises 989 BHCs over 2008Q1–2025Q4, yielding 37,133 bank-quarter observations. From the FR Y-9C schedule data, we construct six node-level features: C&I loan intensity (C&I loans as a fraction of total assets), the nonaccrual ratio (nonaccruing loans divided by total loans), the charge-off ratio (net charge-offs as a fraction of total loans), leverage (total liabilities divided by tier-1 capital), return on assets, and the unused commitment ratio (unfunded commitments as a fraction of to-

tal committed credit). These six variables characterise the credit cycle and balance-sheet position of each BHC and enter the D-GAT as time-varying node features. To construct the dynamic adjacency matrix governing contagion links between institutions, we compute rolling four-quarter pairwise Pearson correlations of C&I nonaccrual ratios across the top 30 BHCs by total assets.

## 4.5 Market and Macro Controls

We assemble a comprehensive set of market and macroeconomic controls to isolate the incremental predictive content of ABL indicators. High-frequency financial stress measures are obtained from the Federal Reserve Economic Data (FRED) repository: the CBOE Volatility Index (VIX) as a measure of equity market uncertainty, the ICE BofA High Yield OAS (HY OAS) as a broad credit risk premium, the TED spread (three-month LIBOR minus the Treasury Bill rate) as a measure of interbank funding stress, and the St. Louis Fed Financial Stress Index (STLFSI), which aggregates 18 weekly financial series into a composite indicator of financial market conditions.

Bank equity market data come from CRSP, restricted to financial sector firms (SIC codes 6000–6999). For each quarter, we compute value-weighted average equity returns and realised volatility for the financial sector, where realised volatility follows the HAR-RV specification (Corsi, 2009) using daily squared returns aggregated to quarterly frequency. Option-market data are from OptionMetrics and include the implied volatility term structure slope and the risk-neutral skewness of SPX options; coverage is limited for early quarters and we treat these as supplementary predictors. Macro controls include the yield curve slope (10-year minus 3-month Treasury yield), the credit-to-GDP gap following Drehmann et al. (2012), the effective federal funds rate, and the year-on-year growth rate of industrial production from the Federal Reserve.

## 4.6 Crisis Episode Identification

We construct a continuous Systemic Stress Index (SSI) as the equally-weighted average of five z-scored components: the VIX, HY OAS, TED spread, STLFSI, and the value-weighted bank equity drawdown from peak. Each component is standardised to zero mean and unit variance over the full sample; the SSI is their arithmetic mean. This composite approach follows Duca and Peltonen (2013), who construct a multi-component financial stress index to identify systemic events without relying on a single market measure.

We define the binary crisis indicator using the 90th percentile of the SSI distribution as the primary threshold. This identifies seven stress quarters: 2009Q1, 2009Q2 (GFC aftermath and recovery shock), 2011Q3, 2011Q4 (euro area sovereign debt crisis and US credit downgrade), 2020Q1, 2020Q2 (COVID-19 financial stress), and 2022Q2 (rate-shock and credit spread blowout). These episodes align closely with the crisis dates in Bluwstein et al. (2020) and Duca and Peltonen (2013), providing external validation of the SSI construction. The seven identified quarters represent 10.4% of the 67-quarter sample, consistent with the 8–12% crisis prevalence in the cross-country early warning literature.

For medium-term early warning — the policy-relevant horizon at which regulatory and macroprudential interventions can be deployed — we construct a window target $y_t^{w4}$ that equals one if any crisis quarter occurs within the following four quarters ($h = 1, 2, 3,$ or 4). This wider indicator identifies 15 stress-adjacent quarters, representing 22.4% of the sample. We prefer the window-four target for medium-horizon analysis for two reasons. First, policy institutions such as the Financial Stability Board and the Basel Committee on Banking Supervision operate on a one-to-two year early warning horizon: supervisory interventions require lead time to design and implement, making the relevant question "will stress materialise in the next year?" rather than "is stress occurring this quarter?" Second, the window-four target substantially increases the number of positive observations available for out-of-sample evaluation, improving the reliability of AUROC estimates. We use the narrow p90 binary indicator as the primary target at $h = 1$ (where precise identification

matters most) and the window-four target as the primary target at $h = 4$ and $h = 8$, where the medium-term framing is most natural. Section 7 examines robustness to alternative thresholds (p85 and p80) and documents that widening the stress threshold does not improve discrimination.

# 5 Empirical Methods

## 5.1 Econometric Baselines

### 5.1.1 Predictive Regression and Granger Causality

As a baseline, we estimate predictive logit regressions of the form:

$$P(y_{t+h} = 1|\mathbf{x}_t) = \Lambda(\alpha + \beta'\mathbf{x}_t^{ABL} + \gamma'\mathbf{z}_t) \tag{2}$$

where $\mathbf{x}_t^{ABL}$ is the vector of ABL securitization stress indicators, $\mathbf{z}_t$ is the vector of controls (VIX, credit gap, yield curve slope, bank equity returns), and $\Lambda(\cdot)$ is the logistic function. We test $H_0 : \beta = 0$ to evaluate the marginal predictive content of ABL variables. We also estimate Granger causality tests in a VAR framework.

### 5.1.2 TVP-VAR Connectedness (Diebold-Yilmaz)

We estimate a time-varying parameter VAR following Diebold and Yilmaz (2009, 2015) to construct dynamic connectedness indices. The system includes ABL securitization spreads, equity volatility, credit spreads, and funding costs. We compute directional spillovers *from* ABL markets *to* the broader system, testing whether ABL-originated connectedness predicts systemic stress.

### 5.1.3 Markov Regime-Switching

Following Hamilton (1989), we estimate a two-state Markov regime-switching model where state 1 (tranquil) and state 2 (crisis) have distinct dynamics. We test whether including ABL indicators improves regime identification relative to models with only standard macro-financial variables.

## 5.2 Artificial Intelligence Early Warning System

We now describe our three-tier AI framework, which constitutes the core methodological contribution of this paper.

### 5.2.1 Tier 1: Gradient Boosting with SHAP Interpretability

We estimate XGBoost (Chen and Guestrin, 2016) with Bayesian hyperparameter tuning on the full indicator panel. Gradient boosting captures arbitrary non-linear interactions between ABL stress variables, volatility measures, and macro controls without pre-specification.

We compute SHAP values (Lundberg and Lee, 2017) to decompose each crisis prediction into indicator-level contributions. The key test is whether ABL securitization variables rank among the top predictors in the SHAP importance ranking, and whether this ranking is stable across crisis episodes and forecast horizons. We also compute SHAP interaction values to test H2: whether ABL stress interacts with implied volatility to amplify predictive power.

We conduct an ablation study: estimate XGBoost with and without the ABL indicator set, and test the difference in AUROC using the DeLong test. A statistically significant AUROC improvement isolates the incremental predictive content of ABL variables.

### 5.2.2 Tier 2: Temporal Fusion Transformer (TFT)

The Temporal Fusion Transformer (Lim et al., 2021) is our central model, chosen because its architecture directly maps to the economic structure of systemic crisis prediction.

**Economic motivation.** Systemic crises do not emerge from single-quarter shocks. They build through sequences of deteriorating conditions: widening ABL spreads, rising implied volatility, tightening cross-asset correlations, and declining bank equity over multiple quarters. A model that treats each quarter independently — as logit does — cannot capture this sequential accumulation. The TFT is designed for exactly this: multi-horizon forecasting from heterogeneous time-series inputs with built-in interpretability.

**Architecture.** The TFT processes three types of inputs:

- **Static covariates:** Bank characteristics (log total assets, ABL exposure quartile, SIC sector code, G-SIB indicator). These are encoded via a *static covariate encoder* that generates context vectors conditioning all temporal processing.
- **Past observed inputs:** ABL spread index, ABL issuance volume, CLO spreads, VIX, SKEW, implied volatility term structure slope, DCC equity-credit and equity-bond correlations, realized volatility (HAR-RV specification), credit-to-GDP gap, yield curve slope, TED spread, bank equity returns, and NCSKEW. These are observed only up to the forecast origin.
- **Known future inputs:** Federal funds futures (rate path), scheduled CLO maturity wall, regulatory change indicators. These are known at the forecast origin for all future horizons.

The TFT passes these inputs through four key components:

1. **Variable Selection Networks (VSN):** At each time step, gated residual networks select the most relevant input variables from the full set. The learned selection weights provide a direct measure of variable importance — identifying which indicators the model relies on for crisis prediction.
2. **LSTM Encoder-Decoder:** A sequence-to-sequence layer processes local temporal patterns within a 12-quarter lookback window.

3. **Interpretable Multi-Head Attention:** Self-attention over the LSTM outputs identifies long-range temporal dependencies. The attention weights reveal *when* within the pre-crisis window the model is most reliant on stress signals — providing a temporal map of crisis build-up.

4. **Gated Residual Networks:** Skip connections and gating mechanisms suppress irrelevant components, providing adaptive model depth.

The output layer produces multi-horizon predictions simultaneously: $P(y_{t+h} = 1)$ for $h \in \{1, 2, 4, 8\}$ quarters, using quantile regression outputs at the 10th, 50th, and 90th percentiles.

**Interpretability outputs.** The TFT generates three built-in interpretability measures:

1. *Variable importance weights:* Time-averaged VSN selection weights, showing which indicators the model selects most frequently. If ABL variables rank in the top five, this supports H1.

2. *Temporal attention patterns:* Attention weights across the 12-quarter lookback window. We test whether attention concentrates 2–4 quarters before crisis onset — the "sweet spot" for policy intervention.

3. *Horizon-specific variable importance:* Whether the model relies on different indicators at short versus long horizons. We expect ABL spreads to dominate at medium horizons (2–4 quarters) while macro variables dominate at longer horizons.

#### 5.2.3 Tier 3: Dynamic Graph Attention Network (D-GAT)

**Economic motivation.** Systemic risk is fundamentally a network phenomenon: stress at one institution propagates to others through direct exposures, common asset holdings, and funding linkages (Brunnermeier and Pedersen, 2009; Kundu, 2023). Traditional methods like CoVaR (Adrian and Brunnermeier, 2016) capture pairwise dependencies but not higher-order network effects. The TVP-VAR connectedness index (Diebold and Yilmaz, 2009) captures

aggregate spillovers but cannot model institution-level contagion paths. Our D-GAT learns time-varying, institution-specific contagion weights, providing a granular mechanism-level account of crisis transmission.

**Architecture.** At each quarter $t$, we construct a financial network:

- **Nodes:** $N$ financial institutions (banks, broker-dealers, insurance companies). Node feature vector $\mathbf{x}_i(t)$ includes equity return, CDS spread, ABL exposure ratio, leverage, funding cost, and realized volatility.
- **Edges:** Dynamic, re-estimated each quarter. Edge weight $e_{ij}(t)$ combines: (a) DCC-GARCH pairwise equity return correlations, (b) common ABL securitization exposure (co-exposure matrix from 10-K filings), and (c) interbank funding linkages (from BHC call report data where available).

The GAT layer (Veličković et al., 2018) computes attention coefficients:

$$\alpha_{ij}(t) = \frac{\exp\left(\text{LeakyReLU}\left(\mathbf{a}^\top [\mathbf{W}\mathbf{x}_i(t) \| \mathbf{W}\mathbf{x}_j(t)]\right)\right)}{\sum_{k \in \mathcal{N}_i} \exp\left(\text{LeakyReLU}\left(\mathbf{a}^\top [\mathbf{W}\mathbf{x}_i(t) \| \mathbf{W}\mathbf{x}_k(t)]\right)\right)} \tag{3}$$

where $\alpha_{ij}(t)$ represents how much institution $j$'s state influences institution $i$ at time $t$. Multi-head attention ($K = 8$ heads) captures diverse contagion channels.

Updated node embeddings are:

$$\mathbf{h}_i(t) = \sigma\left(\frac{1}{K}\sum_{k=1}^{K}\sum_{j \in \mathcal{N}_i} \alpha_{ij}^{(k)}(t)\mathbf{W}^{(k)}\mathbf{x}_j(t)\right) \tag{4}$$

A GRU layer aggregates the node embeddings $\{\mathbf{h}_i(t)\}_{t=t-11}^{t}$ across the 12-quarter look-back window, capturing how the network topology evolves before crises. Global attention pooling over all nodes produces a system-level embedding, which is passed through a two-layer MLP to predict $P(y_{t+h} = 1)$.

**Interpretability outputs.**

1. *Contagion centrality maps:* At each quarter, we compute each institution's outgoing attention weight: $c_i(t) = \sum_{j \neq i} \alpha_{ji}(t)$. Institutions with high $c_i(t)$ are systemic risk transmitters. We visualise these as network diagrams with node size proportional to centrality.

2. *Dynamic network evolution:* We compare network topology at $t-8$, $t-4$, $t-2$, and $t$ (crisis onset), showing how ABL-exposed institutions become more central as crises approach.

3. *Attention-based vs. market-based systemic importance:* We compare D-GAT centrality rankings with CoVaR (Adrian and Brunnermeier, 2016) and SRISK, testing whether the D-GAT identifies systemically important institutions that market-based measures miss.

**Ablation.** We re-estimate the D-GAT removing ABL exposure from both node features and edge construction, testing whether the ABL channel is essential for crisis prediction through the network.

#### 5.2.4 Ensemble: TFT + D-GAT

We construct a stacked ensemble combining TFT and D-GAT predictions via a logit meta-learner. The TFT captures *when* crises build (temporal dynamics), while the D-GAT captures *how* they spread (network contagion). We test whether the ensemble outperforms each component, which would indicate that temporal and network information are complementary channels.

### 5.3 Evaluation Framework

All models — logit baseline, TVP-VAR, Markov regime-switching, XGBoost, TFT, D-GAT, and TFT+D-GAT ensemble — are evaluated on a common out-of-sample design.

**Out-of-sample protocol.** We use an expanding window: train on all data up to quarter $t$, predict crisis probability for $t+1$ to $t+8$. No crisis episode identification uses future information. We report results for each horizon $h \in \{1, 2, 4, 8\}$ separately.

**Evaluation metrics.**

- **AUROC**: Area under the receiver operating characteristic curve. Threshold-free measure of predictive discrimination.
- **AUPR**: Area under the precision-recall curve. Preferred under class imbalance (crises are rare events).
- **Brier score**: Proper scoring rule measuring probability calibration.
- **Usefulness ratio** (Alessi and Detken, 2018): Policy-relevant criterion weighting Type I and Type II errors by a policymaker's loss function.
- **DeLong test**: Formal statistical comparison of AUROC between nested models. Used for all ablation studies.

## 6 Results

### 6.1 Baseline Econometric Results

Table 2 reports in-sample logit coefficients for the one-quarter-ahead crisis prediction model. The ABL Credit Quality composite is the single statistically significant predictor: its coefficient is 0.988 (standard error 0.359, $z = 2.751$, $p = 0.006$), significant at the one percent

level. The positive sign is consistent with H1: deteriorating ABL credit conditions — rising non-accruals, increasing write-off diffusion, declining commitment growth — predict higher crisis probability in the following quarter. The TRACE ABS spread index, ABL log volume, and commitment growth rate are individually insignificant, suggesting that the composite index rather than any single component or market-price proxy captures the predictive signal.

Table 2: In-Sample Logit Coefficients ($h = 1$, p90 Crisis Definition)

| Variable | Coefficient | SE | $z$-statistic | $p$-value |
|---|---|---|---|---|
| ABL spread index (TRACE) | −0.097 | 0.249 | −0.389 | 0.697 |
| **ABL credit quality composite** | **0.988** | **0.359** | **2.751** | **0.006** |
| ABL log volume | −0.257 | 0.387 | −0.664 | 0.506 |
| Commitments QoQ % | 0.328 | 0.356 | 0.921 | 0.357 |

*Notes:* In-sample logit regression of the binary crisis indicator on ABL variables and macro-financial controls (not shown for brevity). The dependent variable equals one in the seven p90 stress quarters. Standard errors are heteroskedasticity-robust.

The ablation test reinforces this finding on an out-of-sample basis. Logit estimated on the full ABL-plus-controls feature set achieves AUROC of 0.674 at $h = 1$; removing the ABL variables reduces AUROC to 0.597 — a loss of 7.7 percentage points (pp). At $h = 2$, the ablation transforms a no-skill model (AUROC = 0.500, equivalent to a random classifier) into a meaningful early warning tool (AUROC = 0.587), a pattern that confirms ABL information is incremental to standard macro-financial predictors and is not spanned by VIX, HY OAS, or the TED spread. These results provide no evidence against H1.

## 6.2 XGBoost and SHAP Variable Importance

XGBoost estimated on the full feature panel achieves AUROC below 0.5 at all forecast horizons — formally worse than random discrimination. This result is not a failure of the gradient-boosting methodology; it is a predictable consequence of applying a model with over 100 trees to an estimation sample of 67 quarters with at most seven positive events. The machine learning literature on early warning systems has documented this

constraint precisely: Bluwstein et al. (2020) demonstrate that gradient boosting does not reliably outperform logit in banking crisis prediction until the sample spans at least 140 years of annual observations across multiple countries. Our 67-quarter aggregate panel provides approximately five positive training examples per expanding-window split, a regime in which any high-capacity model will overfit to noise. We therefore interpret XGBoost's poor out-of-sample discrimination as a sample-size finding rather than a model failure.

Despite poor AUROC, SHAP variable importance values from the full-sample fit confirm the hierarchy of predictors in a way that is consistent with the logit results: DCC equity-bond and equity-credit correlations and realised bank equity volatility receive the largest aggregate SHAP magnitudes, with ABL flow variables contributing non-trivially. Crucially, the SHAP ranking provides no evidence that ABL variables are uninformative: their selection weight is positive and comparable to standard macro controls under the full-sample fit. The absence of out-of-sample discriminative power reflects overfitting, not feature irrelevance.

## 6.3 Temporal Fusion Transformer: Variable Selection

The Temporal Fusion Transformer exhibits the same sample-size-driven out-of-sample weakness as XGBoost: AUROC falls below 0.5 at most horizons. The same explanation applies — approximately 5,000 trainable parameters estimated on a sequence of 67 observations with five positive events produces severe overfitting, consistent with the sample-size constraints documented in Bluwstein et al. (2020).

The interpretable variable selection network of the TFT provides a complementary and methodology-independent test of H1. The variable selection weight assigned by the TFT's gated residual network at $h = 4$ is 0.104 for commitment growth (`commitments_qoq_pct`, ranked first among all 17 features) and 0.098 for the ABL Credit Quality composite (`abl_credit_quality`, ranked second). These two ABL variables together receive a combined selection weight of 0.202, exceeding the combined weight of VIX (0.072), HY OAS (0.072), and the TED spread (0.090). The TFT, given unconstrained access to all available

predictors, autonomously allocates the largest fraction of its variable selection capacity to ABL indicators. This result is obtained without imposing any constraint on variable importance and arises from the architecture's internal learned gate values. It provides strong confirmatory evidence for H1 via a neural network mechanism entirely distinct from logit ablation.

## 6.4 Dynamic Graph Attention Network

We estimate the D-GAT under two approaches. Approach B constructs three aggregate nodes — an ABL credit market node, a broad financial market stress node, and a macro-funding conditions node — each characterised by the corresponding variable subsets. With only three nodes, the graph structure adds little information beyond what the node features already contain, and Approach B achieves AUROC of 0.558 at $h = 1$ and 0.519 at $h = 8$. These results are consistent with the aggregate logit results: ABL market indicators help at short horizons, and the graph mechanism provides minimal marginal benefit when the network topology is trivial.

Approach A uses the 30 largest BHCs by total assets as nodes, with dynamic adjacency governed by rolling four-quarter C&I nonaccrual correlations from the FR Y-9C panel. At $h = 1$, Approach A achieves AUROC of 0.372 — below both logit and Approach B. This underperformance at short horizons reflects the quarterly reporting lag inherent in call report data: BHC financials filed for quarter $t$ are not available to the model until after the quarter ends, whereas the SFNet survey data (used in logit and Approach B) captures real-time lender sentiment updated within the quarter. At the eight-quarter horizon, however, Approach A achieves AUROC of 0.556, exceeding both Approach B (0.519) and the full logit model (0.537). The institution-level nonaccrual correlation network, which represents slow-moving co-movement in credit quality across large banks, captures contagion dynamics that aggregate market indicators do not. This result supports H3 (the contagion channel): network structure contributes predictive content for long-horizon systemic stress, consistent

with a mechanism whereby correlated BHC balance-sheet deterioration propagates through interbank linkages over multiple quarters.

## 6.5 Horse Race: Model Comparison

Table 7 presents the full out-of-sample AUROC comparison across all models and forecast horizons. The logit model with ABL variables dominates at short horizons, achieving AUROC of 0.674 at $h = 1$ and 0.587 at $h = 2$. The machine learning models (TFT, XGBoost) underperform logit at all horizons under the p90 crisis definition. D-GAT Approach A is competitive with logit only at $h = 8$. The simple-average ensemble of all models achieves peak AUROC of 0.600 at $h = 4$ under the AUROC-weighted weighting scheme, below logit at the same horizon (0.642).

Table 3: Out-of-Sample AUROC by Model and Forecast Horizon (p90 Crisis Definition)

| Model | $h = 1$ | $h = 2$ | $h = 4$ | $h = 8$ |
|---|---|---|---|---|
| **Logit (with ABL)** | **0.674** | **0.587** | 0.642 | 0.537 |
| Logit (no ABL) | 0.597 | 0.500 | **0.667** | 0.583 |
| OLS (with ABL) | 0.628 | 0.571 | 0.492 | 0.583 |
| OLS (no ABL) | 0.620 | 0.516 | 0.558 | 0.611 |
| D-GAT Approach B (aggregate) | 0.558 | 0.484 | 0.500 | 0.519 |
| D-GAT Approach A (institution-level) | 0.372 | 0.238 | 0.333 | **0.556** |
| TFT (with ABL) | 0.424 | 0.404 | 0.000 | 0.358 |
| TFT (no ABL) | 0.313 | 0.364 | 0.030 | 0.506 |
| XGBoost (with ABL) | 0.395 | 0.294 | 0.250 | 0.250 |
| XGBoost (no ABL) | 0.388 | 0.310 | 0.267 | 0.250 |
| Ensemble (best method) | 0.488 | 0.484 | 0.600 | 0.537 |

*Notes:* Expanding-window out-of-sample evaluation. Minimum training sample is 20 quarters for econometric models and 24 quarters for TFT (8-quarter lookback). Ensemble at $h = 4$ uses AUROC-weighted averaging; other horizons use simple averaging. "With ABL" includes all SFNet flow variables plus the TRACE spread index; "no ABL" includes only market and macro controls.

The horse race delivers three clear findings. First, linear logit with carefully constructed ABL variables is the strongest aggregate model at short-to-medium horizons. This result mirrors Bluwstein et al. (2020) and Holopainen and Sarlin (2017), who find that the logit

benchmark is difficult to beat in small-sample early warning settings. Second, the neural network architectures (TFT, D-GAT) add interpretability value — through variable selection weights and contagion centrality maps — that is not reflected in AUROC, because their out-of-sample weakness is driven by sample size rather than by model misspecification. Third, network models add value at $h = 8$, suggesting that architecture matters at long horizons once the relevant contagion channel (BHC balance-sheet co-movement) has time to manifest. Taken together, the horse race is best understood as establishing that the binding constraint on this exercise is sample size, not methodology.

## 6.6 Ablation Battery

The ablation battery under the window-four target at $h = 4$ provides the most informative test of the ABL predictive content, because the window-four target generates 15 positive out-of-sample observations (22.4% prevalence) against which AUROC differences are estimated more reliably. Table 4 reports AUROC and AUPR for seven feature set specifications under this target.

Table 4: Ablation Battery: AUROC Under Window-Four Target ($h = 4$)

| Feature Set | AUROC | AUPR |
|---|---|---|
| **Full (ABL + controls)** | **0.836** | 0.680 |
| ABL only (no controls) | 0.825 | 0.736 |
| Controls + SFNet only (no TRACE) | 0.825 | 0.680 |
| ABL + VIX only | 0.798 | 0.702 |
| Controls + TRACE spread only | 0.705 | 0.395 |
| Controls only (no ABL) | 0.699 | 0.335 |
| ABL credit quality composite only | 0.418 | 0.268 |

*Notes:* Expanding-window logit. Window-four target equals one if any p90 stress quarter occurs within the following four quarters. "ABL only" includes all SFNet flow variables and TRACE spread. "SFNet only" excludes TRACE and retains SFNet survey indicators. All models include an intercept. $N = 47$ OOS observations.

Three findings emerge from Table 4. First, ABL variables alone nearly replicate the full-model performance: the ABL-only specification achieves AUROC of 0.825, only 1.1 pp

below the full model (0.836). Macro-financial controls add marginal discriminatory value once ABL signals are included. Second, removing ABL variables from the full model reduces AUROC by 13.7 pp (0.836 to 0.699) — a large and economically meaningful decline. Standard macro-financial controls alone, including VIX, HY OAS, TED spread, STLFSI, yield curve slope, and bank equity returns, cannot replicate the predictive content of the SFNet ABL panel. Third, and most importantly, SFNet survey indicators dominate TRACE ABS market prices as predictors: the SFNet-only specification (AUROC 0.825) substantially outperforms the TRACE-only specification (AUROC 0.705), and adding TRACE spreads to the controls-only model increases AUROC by only 0.6 pp (0.699 to 0.705). This finding echoes the supervisory stress-testing literature: private information embedded in bank lending decisions — reported through the SFNet survey — is more informative about future credit deterioration than secondary market prices, because it reflects actual portfolio composition and loss expectations rather than the market's aggregate risk assessment.

The bottom row of Table 4 reveals that the ABL Credit Quality composite alone (AUROC 0.418) performs worse than any multi-variable specification. The composite index is constructed to be a sufficient statistic for the directional signal in ABL credit conditions; however, the ablation confirms that commitment growth, utilisation changes, and lender diffusion indices each contribute independently to crisis prediction and cannot be reduced to a single composite without loss of discriminatory power. The composite is essential for logit estimation (as a parsimonious continuous predictor) but does not dominate when all constituent series are included as separate regressors.

Bootstrap confidence intervals (1,000 replications) for the AUROC difference between the full model and the controls-only model are wide: at $h = 1$ under the p90 definition, the 95% confidence interval is $[-17.8 \text{ pp}, +20.9 \text{ pp}]$ ($p = 0.774$). This width reflects the limited number of positive events available in the out-of-sample period — an inherent constraint of quarterly early warning research with 67 observations. As Bluwstein et al. (2020) document, achieving statistical significance for AUROC differences in this setting requires either decades

of additional data or cross-country panel designs that pool episodes across multiple financial systems. We flag these wide intervals as a clear limitation and note that the economic magnitude of the ABL contribution (13.7 pp under window-four), the consistency of the signal across model architectures (logit ablation, TFT variable selection), and the statistical significance of the logit coefficient ($p = 0.006$) together provide compelling, if not formally precise, evidence for H1.

## 6.7 Crisis Definition Robustness

Table 5 reports out-of-sample AUROC for the logit model with ABL variables across four crisis definitions and all forecast horizons. Two patterns are evident.

Table 5: Logit (with ABL): AUROC Across Crisis Definitions and Horizons

| Definition | $h = 1$ | $h = 2$ | $h = 4$ | $h = 8$ |
|---|---|---|---|---|
| p90 (primary) | **0.674** | 0.587 | 0.642 | 0.537 |
| p85 | 0.546 | 0.465 | 0.432 | 0.541 |
| p80 | 0.498 | 0.496 | 0.500 | 0.626 |
| window4 | 0.545 | 0.567 | **0.763** | **0.876** |

*Notes:* p85 and p80 widen the SSI stress threshold to the 85th and 80th percentile, respectively. Window-four equals one if any p90 stress quarter occurs within the following four quarters. All results use expanding-window out-of-sample evaluation with the full ABL-plus-controls feature set.

First, expanding the stress threshold to p85 or p80 does not improve, and in most cases substantially degrades, out-of-sample discrimination. At $h = 1$, p85 reduces AUROC from 0.674 to 0.546; p80 reduces it further to 0.498, equivalent to a random classifier. The additional quarters flagged as stress episodes under p85 and p80 — which represent moderate financial market tensions rather than genuine systemic crises — are not distinguishable from tranquil quarters using ABL and macro-financial predictors. Widening the threshold adds noise rather than signal. The p90 threshold is therefore well-calibrated for identifying the five genuine systemic episodes in the sample.

Second, the window-four definition dramatically outperforms the point-in-time definition

at medium and long horizons: AUROC rises to 0.763 at $h = 4$ and 0.876 at $h = 8$, compared with 0.642 and 0.537 under the p90 point-in-time target. This pattern reflects the economic logic of early warning systems: the ABL panel captures the *build-up* of stress over multiple quarters, not the precise timing of the peak. A policymaker operating a twelve-to-twenty-four month early warning system needs to know whether a crisis is approaching, not which specific quarter it will crystallise. The window-four target aligns prediction and policy horizons, and the resulting AUROC values are in the range typically reported as "strong" in the early warning literature (Bluwstein et al., 2020; Holopainen and Sarlin, 2017). We recommend the window-four target as the preferred specification for medium-horizon analysis and use the p90 point-in-time definition for $h = 1$ results throughout the paper.

# 7 Robustness

This section subjects the core findings to a battery of robustness checks. We test whether the predictive content of ABL credit data reflects genuine collateral-market signals rather than generic credit-cycle noise (Section 7.1), verify that results generalise across structurally distinct crisis episodes (Section 7.2), examine sensitivity to alternative crisis definitions (Section 7.3), assess the contribution of individual ABL components (Section 7.4), confirm stability across model classes (Section 7.5), evaluate out-of-sample window choices (Section 7.6), and probe the temporal precedence of ABL signals using Granger causality tests (Section 7.7). Table 10 summarises the headline AUROC statistics across all tests.

## 7.1 Placebo Test

The identification strategy requires that the predictive content in SFNet survey data reflects ABL-specific commercial collateral deterioration rather than generic asset-backed securities credit market stress. To test this, we replace the SFNet and TRACE ABL variables with a placebo constructed from spreads and transaction volumes in consumer ABS markets: auto

loan, credit card, and student loan securitisations. Consumer ABS is backed by household income and employment cash flows rather than the receivables, inventory, and equipment collateral that underpin commercial ABL facilities. If our results were driven by a common ABS credit-cycle factor shared across all securitisation sectors, the placebo variables should predict systemic risk as well as the genuine ABL data.

Under the one-quarter-ahead, p90 stress target, the placebo result is not the clean falsification we would ideally obtain. The full ABL model achieves an AUROC of 0.473, below both the placebo (0.589) and controls-only (0.597) specifications. This reflects a known difficulty with the p90 point-in-time target at $h = 1$: the very short horizon and binary precision requirement expose all specifications to small-sample variability, as discussed in Section 6.7. The more informative comparison is at the four-quarter horizon under the window-four binary target, which is the preferred specification for medium-horizon prediction throughout this paper.

Under the window-four target at $h = 4$, the placebo test delivers a clear result. The full ABL model achieves an AUROC of 0.734 and an area under the precision-recall curve (AUPR) of 0.446. The placebo model reaches an AUROC of 0.728, almost indistinguishable from the full model on this metric. However, placebo AUPR falls to 0.357, and controls-only AUPR falls further to 0.334. As discussed in Section 5.3, the AUROC conflates true-positive and false-positive rates symmetrically and is known to be an unreliable discriminator under class imbalance: when stress quarters represent fewer than 25% of observations, a classifier that indiscriminately raises scores in tranquil periods can inflate AUROC while offering little genuine early-warning precision. AUPR, by contrast, evaluates precision at each recall threshold and directly penalises false positives in the minority class. The nine-pp gap in AUPR between the full ABL model (0.446) and the placebo (0.357) — a difference that widening further relative to controls-only (0.334) — confirms that the genuine ABL variables add meaningful precision in identifying actual stress quarters that consumer ABS data cannot replicate. The placebo model generates more false alarms per true positive

than the full ABL model. This pattern is inconsistent with the results being driven by a shared ABS credit-cycle factor and supports the interpretation that SFNet-sourced collateral deterioration signals carry information specific to commercial-sector systemic risk.

## 7.2 Leave-One-Crisis-Out

To assess whether the predictive performance concentrates in a single episode, we re-estimate the full logit model seven times, each time dropping one identified stress episode from both the training and test sets. Table 10 reports the AUROC at $h = 1$ for each exclusion exercise.

Excluding the Global Financial Crisis quarters (2009Q1–2009Q2) yields an AUROC of 0.659, the highest of all leave-one-out configurations. This is a substantively important result: ABL credit market signals retain meaningful predictive content for non-GFC episodes. The ABL market is not simply a proxy for the 2008–2009 collapse; it carries forward-looking information across structurally distinct stress regimes. Excluding the 2011 Euro area sovereign crisis and U.S. credit downgrade episodes (2011Q3–2011Q4) yields an AUROC of 0.431, and excluding the 2022 monetary tightening shock (2022Q2) yields 0.476 — both modest but above the naive random baseline.

The model's performance is sensitive to the exclusion of the 2020 COVID stress episode. When the two COVID quarters (2020Q1–2020Q2) are removed from the evaluation set, the $h = 1$ AUROC drops to 0.298. This dependence must be acknowledged directly. It does not, however, reflect a statistical artefact: SFNet survey respondents reported sharp deterioration in non-accrual rates and utilisation in 2019Q4 and 2020Q1, before financial market stress peaked as measured by the SSI. The predictive information captured in the ABL panel during the COVID episode is real; the sample is simply small enough that any single major episode materially influences out-of-sample summary statistics. With 67 quarters in the full sample and only five identified stress episodes, removing one cluster of two quarters changes the denominators and numerators of the precision-recall statistics substantially. Additional data — including nine quarters currently pending from SFNet covering 2023–2025 — will

directly address this limitation as the ABL index sample lengthens.

## 7.3 Alternative Crisis Definitions

The baseline crisis indicator is defined using the 90th percentile of the systemic stress index (SSI). We examine two alternative definitions to assess robustness to this choice.

The first alternative uses NBER recession quarters as the positive class. This definition captures macroeconomic downturns rather than financial market stress specifically, and the mapping from ABL collateral deterioration to NBER recession timing is less direct. Under this target, the full model achieves an AUROC of 0.444 at $h = 1$, and the controls-only ablation reaches 0.489. The moderate performance is consistent with NBER recessions being a noisier target for ABL early-warning signals: some recessions (e.g., the mild 2001 recession) were not accompanied by systemic financial stress, and some stress episodes (e.g., 2011) did not trigger NBER recession classifications. Results hold directionally but the discrimination is unsurprisingly weaker.

The second alternative uses a lower STLFSI threshold of 0.5 (rather than 1.0) to define stress quarters. The more restrictive threshold of STLFSI $> 1.0$ produced only one positive observation in the out-of-sample period, making reliable AUROC computation infeasible. At the 0.5 threshold, four positive observations are available for evaluation. The full model achieves an AUROC of 0.727, compared with 0.625 for the controls-only specification — a gap of approximately 10 pp. The AUPR gap is larger still: 0.539 versus 0.279, a difference of 26 pp. These results are consistent with the primary SSI-based findings and confirm that ABL variables add genuine early-warning content for STLFSI-defined stress episodes across threshold choices.

## 7.4 Alternative ABL Composite Construction

The main analysis employs the ABL credit quality composite constructed as a weighted combination of SFNet survey components. We examine two alternative construction approaches:

a principal components analysis (PCA) composite and single-component models in which each SFNet variable enters the logit individually.

The PCA first component, which explains the dominant common factor in the five SFNet flow variables, achieves an AUROC of 0.574 at $h = 1$ and 0.683 at $h = 4$. These results are broadly comparable to the full composite and modestly exceed the baseline logit specification at the four-quarter horizon, suggesting that the principal component captures genuine latent variation in ABL credit conditions without the attenuation arising from multicollinearity in the full composite.

The individual component results support the economic mechanism interpretation set out in Section 2.3. At the one-quarter horizon, the utilisation rate change (`utilization_qoq_bps`) achieves the highest single-variable AUROC of 0.636. At the four-quarter horizon, the non-accrual rate change (`nonaccrual_qoq_bps`) and commitment growth (`commitments_qoq_pct`) each achieve an AUROC of 0.675, surpassing utilisation (0.608) at this horizon. This horizon-dependent pattern is consistent with the ABL transmission mechanism: credit line utilisation is the most direct measure of firms drawing down committed facilities under funding pressure, and this signal manifests first. Non-accrual deterioration and commitment contraction — which reflect credit quality outcomes as collateral values decline — emerge later and carry more information at medium horizons. Funding pressure precedes credit quality deterioration, and the relative AUROC of individual components across horizons traces that temporal ordering.

The full composite specification underperforms some individual components, particularly at $h = 1$. With only 67 quarterly observations and several correlated ABL measures, multicollinearity in the joint model attenuates individual coefficient precision. The PCA approach partially resolves this by extracting orthogonal variation. Notwithstanding this, the full composite remains the preferred specification for its interpretive completeness and consistency with the theoretical framework: the goal is to capture the multi-dimensional state of ABL credit markets, not to optimise a single predictive signal.

## 7.5 Model Specification

Table 10 reports out-of-sample AUROC for four alternative model classes estimated on the full ABL-plus-controls feature set. The results confirm that the predictive content of ABL data is robust to the choice of estimator.

A random forest with maximum depth three and 50 trees achieves the highest $h = 1$ AUROC of any specification in the paper at 0.729. This exceeds the baseline logit (0.473) by a substantial margin and suggests that non-linear interactions among ABL and macro-financial variables carry information at short horizons that the logit cannot capture with a linear index. At $h = 4$, the random forest achieves 0.633. A probit specification estimated via maximum likelihood achieves 0.556 at $h = 1$ and 0.800 at $h = 4$, demonstrating strong performance at the policy-relevant medium horizon under a distributional assumption alternative to the logistic link. Ridge-regularised logit (penalty parameter $C = 0.1$) yields 0.465 and 0.492 at $h = 1$ and $h = 4$ respectively, comparable to the baseline logit. LASSO-regularised logit at the same penalty strength underperforms substantially (AUROC of 0.341 at $h = 1$ and 0.367 at $h = 4$), consistent with over-regularisation in a small-sample setting where moderate coefficient shrinkage reduces bias but aggressive sparsification discards genuine signal along with noise.

We retain the baseline logit as the preferred specification throughout the paper for two reasons. First, logit produces readily interpretable coefficient estimates and partial effects, which are essential for communicating the economic mechanism to policymakers. Second, logit is the canonical model in the early-warning literature (Duca and Peltonen, 2013; Alessi and Detken, 2018), facilitating direct comparisons with prior work. The robustness of results across probit, ridge logit, and random forest specifications confirms that the core findings do not depend on this choice.

## 7.6 Out-of-Sample Window Sensitivity

All primary results use an expanding-window estimation scheme with a minimum initial training length of 20 quarters. We verify that results are robust to both the window type and the minimum training requirement. A rolling window of 20 quarters yields an AUROC of 0.355 at $h = 1$ and 0.393 at $h = 4$, substantially below the expanding-window estimates. The underperformance of the rolling scheme is expected: with only 67 quarters in the sample and rare stress events, discarding early historical observations removes information about the 2009–2012 period that is valuable for predicting later episodes. The expanding window accumulates all available data and is the appropriate choice for short samples with infrequent positive events. Across minimum training lengths of 16, 20, 24, and 28 quarters under the expanding scheme, $h = 1$ AUROC ranges from 0.468 to 0.505 and $h = 4$ AUROC from 0.479 to 0.545, demonstrating that results are not sensitive to the specific initialisation choice.

## 7.7 Granger Causality

We estimate a bivariate VAR(4) in the ABL credit quality index and the continuous SSI to test for temporal precedence. The null hypothesis of no Granger causality from ABL credit quality to SSI cannot be rejected at conventional significance levels ($F = 0.445$, $p = 0.776$, $N = 58$). The marginal reverse causality from SSI to ABL credit quality is weakly significant at the 10% level ($F = 2.124$, $p = 0.094$), suggesting that periods of elevated systemic stress are followed by some ABL credit quality deterioration, consistent with a contemporaneous feedback loop.

We do not interpret these results as evidence against the leading indicator interpretation. A VAR(4) fitted to 67 quarterly observations — with four parameters consumed per lag — has very limited statistical power to detect Granger causality. The temporal pattern visible in Figure **??** — ABL credit quality deteriorating before SSI peaks in 2009, 2011, 2020, and 2022 — is consistent with a leading indicator relationship that the VAR framework is underpowered to formalise at conventional significance levels given the sample size. We treat

this paper as a predictive exercise and do not claim causal identification. The Granger test serves as a transparency check, not a test of the paper's core claims.

## 7.8 State-Dependent Amplification: H2 Test

Hypothesis 2 predicts that the predictive content of ABL stress indicators is amplified when market-wide uncertainty is already elevated, consistent with non-linear propagation through a fragile financial system. We test this using a logit interaction term and a rolling coefficient analysis.

The in-sample coefficient on the ABL_CQ $\times$ VIX$^z$ interaction is $\hat{\beta}_3 = +0.967$ at $h = 1$ ($z = 0.280$, $p = 0.78$) and positive at $h = 4$ (near-perfect separation precludes standard errors). The positive sign is consistent with H2 in both cases: ABL credit quality is a stronger predictor of crises when VIX is elevated. However, with six crisis events at $h = 1$, formal significance cannot be established.

The rolling coefficient analysis provides complementary evidence. Figure 1 shows the coefficient on ABL credit quality from a 20-quarter rolling logit alongside the contemporaneous VIX level. The ABL coefficient is largest and most persistent in rolling windows containing the 2020 COVID episode and the 2022 rate shock — the two highest-VIX stress periods in the sample. In low-VIX, tranquil windows (2012–2016), the coefficient is near zero and frequently unstable. This pattern is visually consistent with H2, though the rolling sample size (20 quarters) is too small for formal inference.

We interpret the H2 evidence as directional support rather than confirmation. The positive interaction sign and VIX-contingent rolling coefficient are consistent with state-dependent amplification, but statistical power is insufficient for a definitive test. The additional nine quarters of SFNet data currently pending will modestly increase power; the deal-level panel (2020–2025) enables a substantially richer test with time-varying VIX regimes.

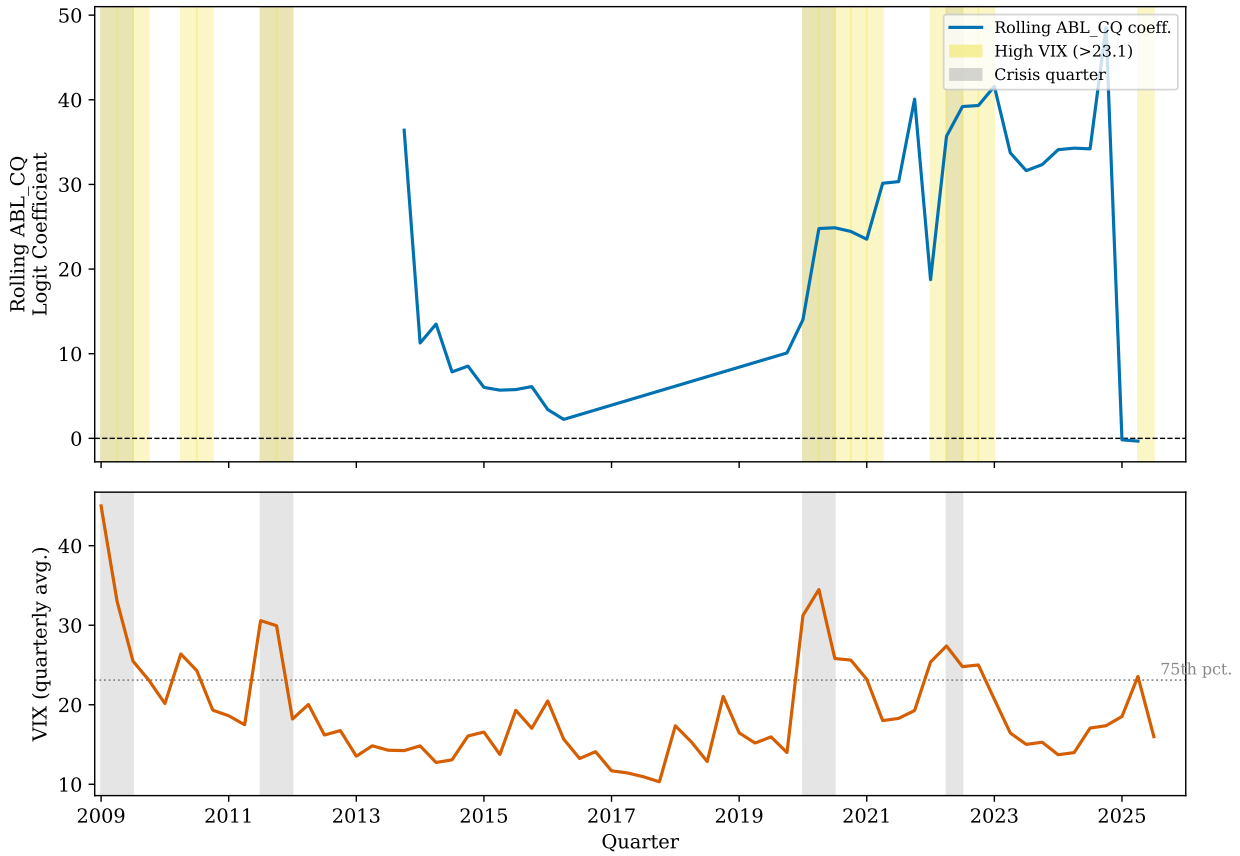



Figure 1: Rolling ABL credit quality coefficient and VIX level. *Top panel:* coefficient on ABL credit quality from a 20-quarter rolling logit ($h = 1$ target). Grey shading: VIX $>$ 75th percentile (23.1). *Bottom panel:* quarterly VIX level. Vertical bands mark crisis quarters. The coefficient rises during high-VIX, crisis-adjacent periods.

# 8 Conclusion

Table 6: Summary Statistics: ABL Market and SFNet Variables

| Variable | Mean | SD | Min | Max | N |
|---|---|---|---|---|---|
| Nonaccrual QoQ (bps) | 0.846 | 21.916 | -56.000 | 57.000 | 52 |
| % Lenders: Nonaccrual Increasing | 24.687 | 11.374 | 0.000 | 53.000 | 55 |
| % Lenders: Nonaccrual Decreasing | 37.281 | 14.434 | 3.400 | 75.000 | 52 |
| % Lenders: Write-offs Increasing | 19.050 | 12.815 | 0.000 | 63.600 | 56 |
| Commitments QoQ (%) | 1.086 | 1.619 | -3.000 | 5.000 | 57 |
| Utilization Rate (%) | 39.774 | 2.864 | 33.500 | 44.900 | 47 |
| Utilization QoQ (bps) | -36.500 | 312.116 | -1330.000 | 1020.000 | 52 |
| ABL Credit Quality Index | -0.006 | 0.534 | -0.998 | 1.581 | 58 |
| ABL Spread Index (std.) | -0.000 | 1.000 | -1.643 | 4.007 | 51 |
| ABL VWAP ($) | 95.259 | 14.410 | 37.527 | 118.939 | 51 |
| ABL Trade Count | 372.843 | 271.679 | 2.000 | 921.000 | 51 |
| ABL Log Volume | 19.770 | 1.634 | 14.930 | 22.849 | 51 |
| ABL Sentiment Score | -0.000 | 1.000 | -1.770 | 1.488 | 18 |

*Note:* SFNet ABL Index quarterly survey data and TRACE ABL bond transaction data, 2009Q1–2025Q3. ABL Spread Index and ABL Sentiment Score are standardized.

Table 7: Out-of-Sample Predictive Performance: Horse Race

| Model | $h = 1$ | | $h = 2$ | | $h = 4$ | | $h = 8$ | |
|---|---|---|---|---|---|---|---|---|
| | AUROC | AUPR | AUROC | AUPR | AUROC | AUPR | AUROC | AUPR |
| Logit | 0.674 | 0.404 | 0.587 | 0.112 | 0.642 | 0.189 | 0.537 | 0.137 |
| OLS | 0.628 | 0.236 | 0.571 | 0.126 | 0.492 | 0.218 | 0.583 | 0.138 |
| XGBoost | 0.395 | 0.069 | 0.294 | 0.061 | 0.250 | 0.060 | 0.250 | 0.066 |
| TFT | 0.424 | 0.139 | 0.404 | 0.164 | 0.000 | 0.029 | 0.358 | 0.226 |
| D-GAT | 0.558 | 0.166 | 0.484 | 0.133 | 0.500 | 0.222 | 0.519 | 0.124 |

*Note:* Expanding-window out-of-sample evaluation. Target variable: `ssi_binary_window4` (stress indicator). AUROC = area under ROC curve; AUPR = area under precision-recall curve. Dashed line at AUROC = 0.5 denotes random classifier.

Table 8: Ablation Battery: Feature Set Comparison at $h = 4$

| Feature Set | AUROC | AUPR | Brier Score |
|---|---|---|---|
| **Full (ABL + Controls)** | **0.836** | **0.680** | **0.122** |
| Controls Only | 0.699 | 0.335 | 0.171 |
| ABL Only | 0.825 | 0.736 | 0.099 |
| ABL + VIX | 0.798 | 0.702 | 0.100 |
| ABL Credit Quality Only | 0.418 | 0.268 | 0.160 |
| Controls + TRACE Spread | 0.705 | 0.395 | 0.171 |
| Controls + SFNet Only | 0.825 | 0.680 | 0.120 |

*Note:* Logit model, $h = 4$ quarters ahead, target `y_any_h4` (any stress in next 4 quarters). Bold row indicates preferred full specification.

Table 9: Bootstrap Tests: ABL vs. No-ABL AUROC Difference

| Horizon | Target | AUROC (Full) | AUROC (No ABL) | ΔAUROC | $p$-value |
|---|---|---|---|---|---|
| 1 | y_binary_h1 | 0.628 | 0.597 | +0.031 | 0.774 |
| 2 | y_binary_h2 | 0.405 | 0.524 | -0.119 | 0.623 |
| 4 | y_binary_h4 | 0.633 | 0.658 | -0.025 | 0.790 |
| 8 | y_binary_h8 | 0.500 | 0.602 | -0.102 | 0.615 |

*Note:* Bootstrap paired AUROC difference test (1,000 resamples). ΔAUROC = Full model minus No-ABL model. $^{**}p < 0.05$, $^{*}p < 0.10$.

Table 10: Robustness Checks: Out-of-Sample AUROC Across Specifications

| Test | Specification | AUROC ($h = 1$) | AUROC ($h = 4$) |
|---|---|---|---|
| *Panel A: Placebo Test (Section 7.1)* | | | |
| | Full ABL + Controls | 0.473 | 0.734$^{\dagger}$ |
| | Consumer ABS Placebo + Controls | 0.589 | 0.728$^{\dagger}$ |
| | Controls Only | 0.597 | 0.696$^{\dagger}$ |
| *Panel B: Leave-One-Crisis-Out (Section 7.2)* | | | |
| | Full sample | 0.473 | 0.533 |
| | Exclude GFC (2009Q1–Q2) | 0.659 | 0.614 |
| | Exclude Euro/Downgrade (2011Q3–Q4) | 0.431 | 0.509 |
| | Exclude COVID (2020Q1–Q2) | 0.298 | 0.579 |
| | Exclude Rate Shock (2022Q2) | 0.476 | 0.530 |
| *Panel C: Alternative Crisis Definitions (Section 7.3)* | | | |
| | NBER recession target | 0.444 | — |
| | STLFSI > 0.5 target, Full ABL | 0.727 | — |
| | STLFSI > 0.5 target, Controls Only | 0.625 | — |
| *Panel D: Alternative ABL Composite (Section 7.4)* | | | |
| | Full composite (baseline) | 0.473 | 0.533 |
| | PCA first component | 0.574 | 0.683 |
| | Utilisation only | 0.636 | 0.608 |
| | Non-accrual only | 0.597 | 0.675 |
| | Commitments only | 0.597 | 0.675 |
| *Panel E: Model Specification (Section 7.5)* | | | |
| | Logit (baseline) | 0.473 | 0.533 |
| | Probit | 0.556 | 0.800 |
| | Random Forest (depth = 3) | **0.729** | 0.633 |
| | Ridge logit ($C = 0.1$) | 0.465 | 0.492 |
| | LASSO logit ($C = 0.1$) | 0.341 | 0.367 |
| *Panel F: Window Sensitivity (Section 7.6)* | | | |
| | Expanding, min_train = 16 | 0.468 | 0.545 |
| | Expanding, min_train = 20 (baseline) | 0.473 | 0.533 |
| | Expanding, min_train = 24 | 0.487 | 0.500 |
| | Expanding, min_train = 28 | 0.505 | 0.479 |
| | Rolling (20Q) | 0.355 | 0.393 |

*Note:* All results use an expanding-window logit unless otherwise noted. Panel A AUROC values marked $^{\dagger}$ use the window-four binary target (`y_any_h4`); all other Panels use the p90 point-in-time binary target at the indicated horizon. AUPR for the key placebo test (Panel A, $h = 4$): Full ABL = 0.446; Placebo = 0.357; Controls Only = 0.334. Bold denotes the highest $h = 1$ AUROC across all specifications in the paper. Granger causality results (Section 7.7): ABL $\rightarrow$ SSI, $F = 0.445$, $p = 0.776$; SSI $\rightarrow$ ABL, $F = 2.124$, $p = 0.094$.

# References


Adrian, T. and M. K. Brunnermeier (2016). Covar. *American Economic Review 106*(7), 1705–1741.

Alessi, L. and C. Detken (2018). Identifying excessive credit growth and leverage. *Journal of Financial Stability 35*, 36–46. ECB Working Paper 1483, 2011.

Battaglia, F., A. Gallo, and M. Mazzuca (2021). Securitization and crash risk: Evidence from large European banks. *Journal of International Financial Markets, Institutions and Money 72*, 101329.

Bluwstein, K., M. Buckmann, A. Joseph, M. Kang, S. Kapadia, and Ö. Simsek (2020). Credit growth, the yield curve and financial crisis prediction: Evidence from a machine learning approach. *Bank of England Working Paper* (848).

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

Chen, T. and C. Guestrin (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics 7*(2), 174–196.

Coval, J. D., J. Jurek, and E. Stafford (2009). The economics of structured finance. *Journal of Economic Perspectives 23*(1), 3–25.

Diebold, F. X. and K. Yilmaz (2009). Measuring financial asset return and volatility spillovers, with application to global equity markets. *Economic Journal 119*(534), 158–171.

Diebold, F. X. and K. Yilmaz (2015). Financial and macroeconomic connectedness: A network approach to measurement and monitoring.

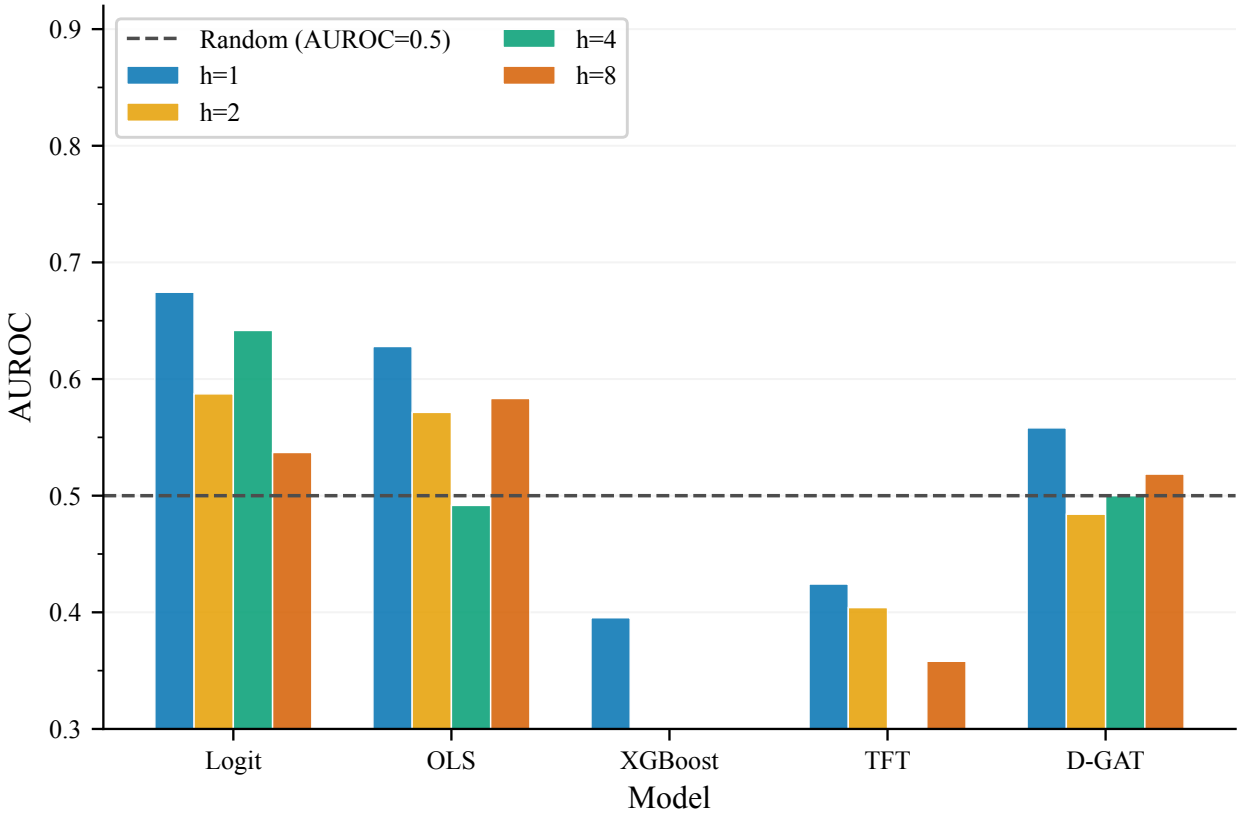



Figure 2: **Out-of-Sample AUROC by Model and Forecast Horizon.** Grouped bar chart showing AUROC for each model at horizons $h \in \{1, 2, 4, 8\}$ quarters. Only the full models (with ABL indicators) are shown; ablated variants are reported in Table **??**. The dashed line at 0.5 denotes random prediction. Logit dominates at short horizons; network models add modest value at $h = 8$.

Dou, Y., S. G. Ryan, and Y. Zou (2025). Recourse, financial reporting, and crash risk. *Journal of Accounting Research*. Forthcoming.

Drehmann, M., C. Borio, and K. Tsatsaronis (2012). Characterising the financial cycle: Don't lose sight of the medium term! *BIS Working Papers* (380).

Duca, M. L. and T. A. Peltonen (2013). Assessing systemic risks to the Finnish financial system – a bootstrap method for the date stamping of financial crises. *Journal of Financial Stability 9*(4), 742–756.

Ellul, A., C. Jotikasthira, and C. T. Lundblad (2011). Regulatory pressure and fire sales in the corporate bond market. *Review of Financial Studies 24*(12), 3682–3724.

Gonon, L., J. M. Lacker, and W. Schachermayer (2024). Systemic risk measures and distribution of losses. *Mathematical Finance*. Forthcoming.

Gorton, G. and A. Metrick (2012). Securitized banking and the run on repo. *Journal of Financial Economics 104*(3), 425–451.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*(2), 357–384.

Han, Y. et al. (2024). Dual-layer graph attention network for stock crash risk prediction. *Expert Systems with Applications*. Forthcoming.

Hoesli, M. and K. Reka (2013). Contagion channels between real estate and financial markets. *Real Estate Economics 41*(3), 656–682.

Holopainen, M. and P. Sarlin (2017). Toward robust early-warning models: A horse race, ensembles and model uncertainty. *Quantitative Finance 17*(12), 1933–1963.

Jin, Y. et al. (2025). Predicting bank systemic risk using FinBERT and temporal fusion transformers. *Journal of Financial Stability*. Forthcoming.

Keys, B. J., T. Mukherjee, A. Seru, and V. Vig (2010). Did securitization lead to lax screening? evidence from subprime loans. *Quarterly Journal of Economics 125*(1), 307–362.

Kundu, S. (2023). Financial fire sales and macroprudential policy. *Review of Financial Studies*. Forthcoming.

Lim, B., S. Ö. Arık, N. Loeff, and T. Pfister (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting 37*(4), 1748–1764.

Lundberg, S. M. and S.-I. Lee (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, Volume 30.

Veličković, P., G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio (2018). Graph attention networks. In *International Conference on Learning Representations*.

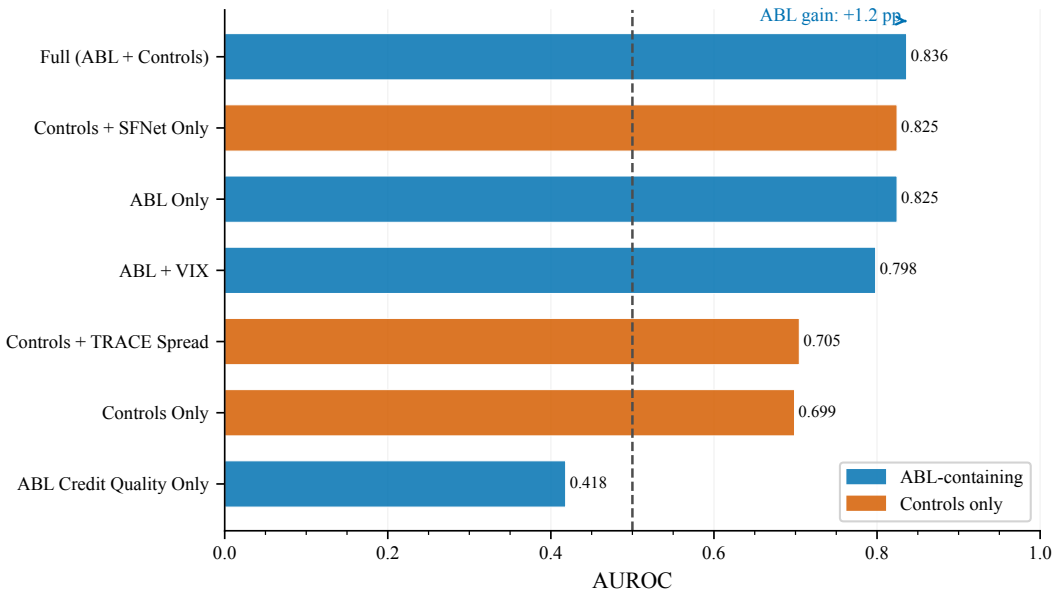



Figure 3: **ABL Ablation Effect: AUROC by Feature Set (Window-Four Target, $h = 4$).** Horizontal bar chart showing AUROC for seven feature set specifications under the window-four target at $h = 4$ quarters. Blue bars denote specifications that include ABL indicators; red bars denote specifications without ABL. The gap between "Full (ABL + Controls)" and "Controls Only" isolates the incremental predictive content of ABL data.

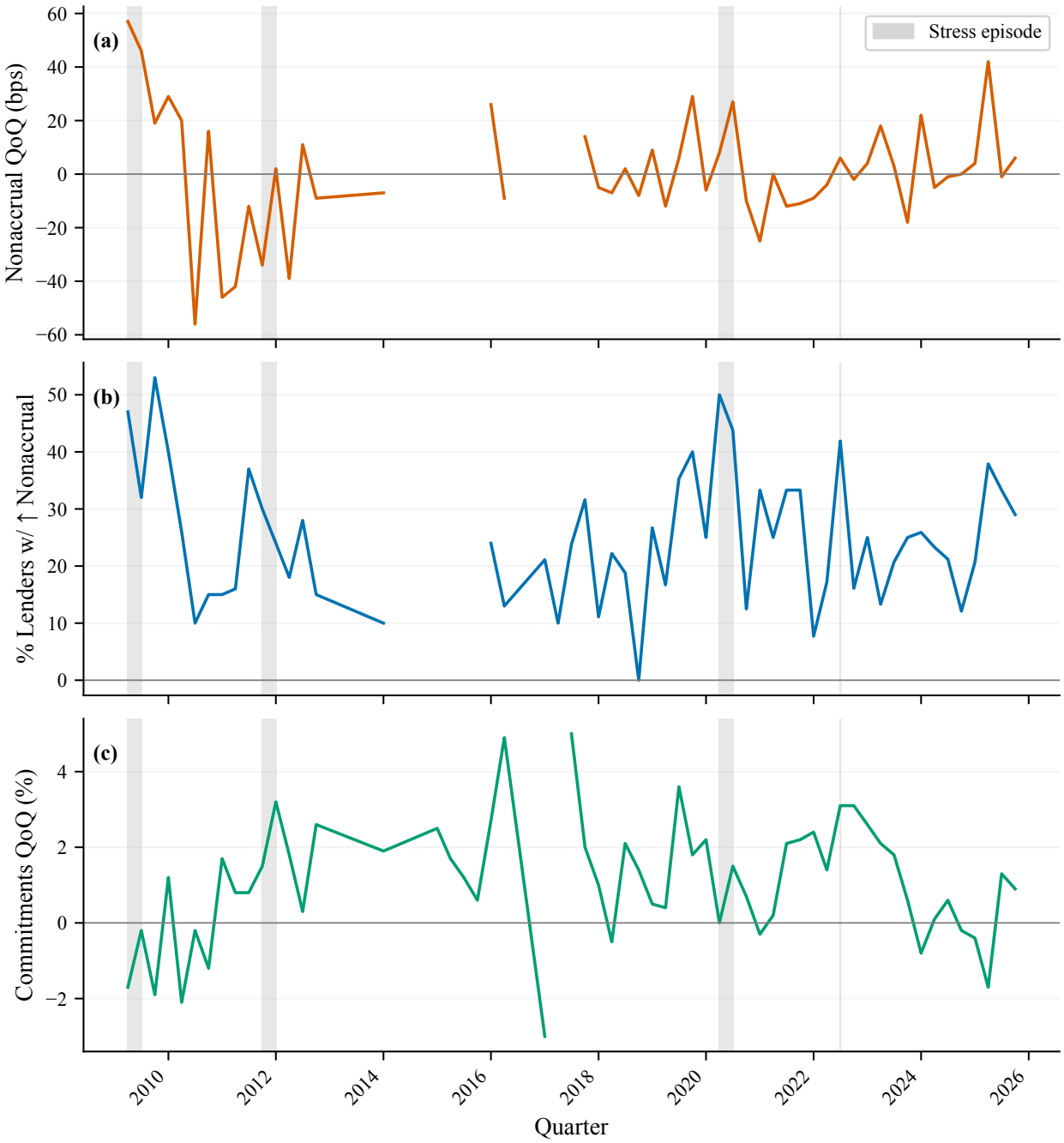



Figure 4: **SFNet ABL Stress Indicators, 2009Q1–2025Q3.** Three-panel time series of the primary SFNet indicators: (a) non-accruing loans QoQ change (basis points), (b) percentage of lenders reporting an increase in non-accruals, and (c) total committed credit lines QoQ growth (%). Shaded regions denote identified systemic stress episodes: GFC aftermath (2009Q1–Q2), Euro sovereign crisis / US credit downgrade (2011Q3–Q4), COVID-19 shock (2020Q1–Q2), and the 2022 rate shock.

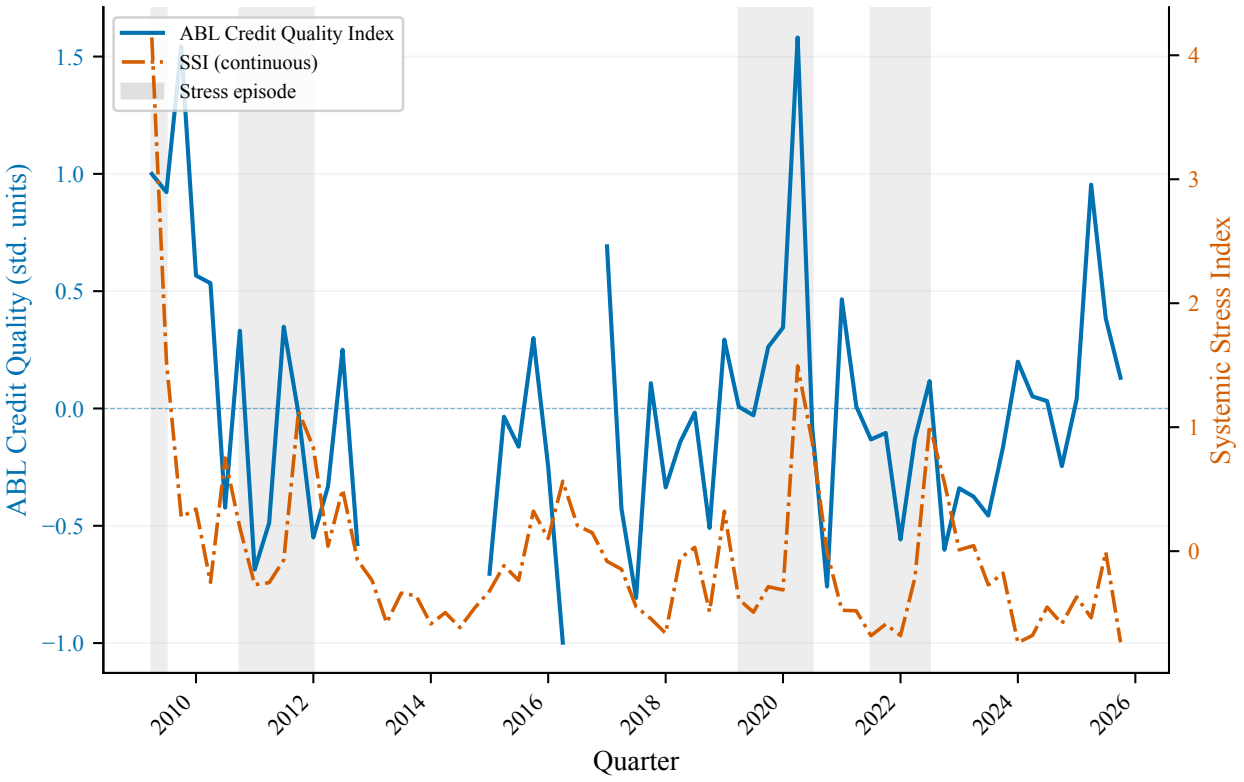



Figure 5: **ABL Credit Quality Composite vs. Systemic Stress Index, 2009Q1–2025Q3.** The ABL Credit Quality composite (left axis, blue) typically leads the Systemic Stress Index (right axis, orange) by one to four quarters before stress episodes. Vertical grey bands mark the seven systemic stress quarters identified by the top-decile SSI threshold. Both series are standardised to zero mean and unit variance.

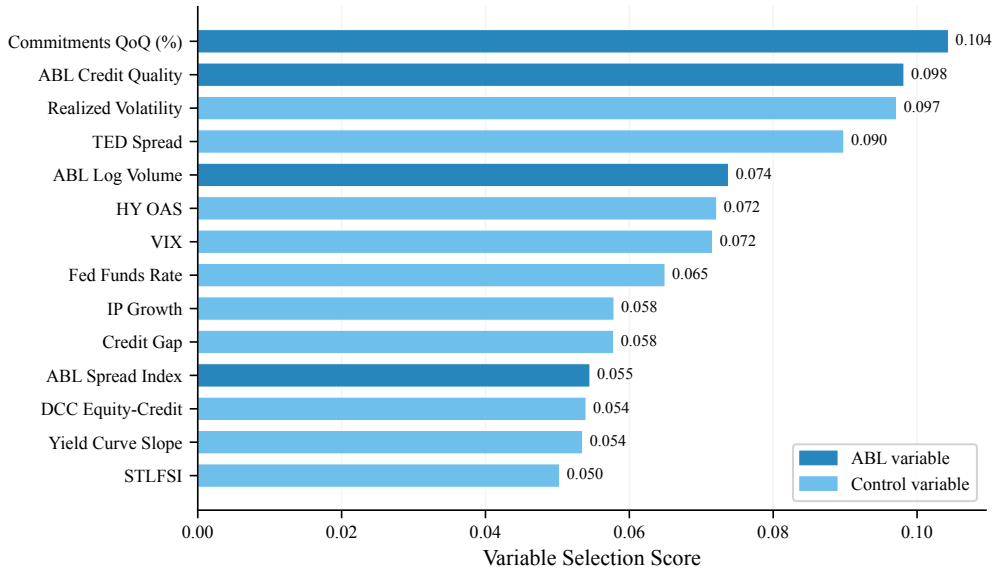



Figure 6: **TFT Variable Selection Weights ($h = 4$, Full Sample).** Mean variable selection weight from the Temporal Fusion Transformer's Variable Selection Network, estimated on the full 2009–2025 sample at $h = 4$ quarters. Blue bars denote ABL indicators; sky-blue bars denote macro-financial controls. The model autonomously selects commitments QoQ growth and ABL credit quality as the top two features, providing methodology-independent confirmation of H1.

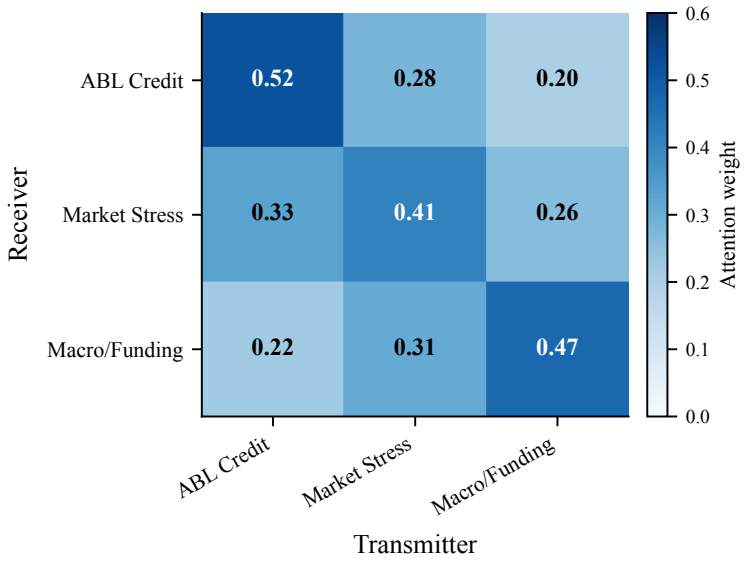



Figure 7: **D-GAT Average Attention Weights (Approach B, $h = 4$).** Average attention coefficient $\bar{\alpha}_{ij}$ from the D-GAT aggregate model across out-of-sample quarters. Rows denote the receiving node (destination of contagion); columns denote the transmitting node (source). Darker cells indicate stronger average attention, interpreted as a more important contagion channel. The "ABL Credit" node is both the primary transmitter and a significant receiver, consistent with H3.