Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Inventory stress as financial risk: Trade shocks, macro-financial conditions and explainable early warning for U.S. retail and manufacturing

Md Mizanur Rahman
The University of Texas at Dallas
Fatema Tuj Johra
The University of Texas at Dallas
Abdullah Al Shohan
The University of Texas at Dallas
Md Mazharul Islam
The University of Texas at Dallas
Md Farhad Hossain
farhad@cou.ac.bd

Comilla University

Research Article

Keywords: inventory stress · working capital · credit spreads · early-warning system · financial stress prediction · international trade · machine learning · forecast evaluation

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11438198/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Inventory stress as financial risk: Trade shocks, macro-financial conditions and explainable early warning for U.S. retail and manufacturing

Md Mizanur Rahman[1], Fatema Tuj Johra[2], Abdullah Al Shohan[3], Md Mazharul Islam[3], and Md Farhad Hossain[4,5,*]

[1]*Department of Finance, The University of Texas at Dallas, Texas 75080-3021, USA*
[2]*Department of Economics, The University of Texas at Dallas, Texas 75080-3021, USA*
[3]*Department of Supply Chain Management, The University of Texas at Dallas, Texas 75080-3021, USA*
[4]*Department of Statistics, Comilla University, Cumilla-3506, Bangladesh*
[5]*Department of Data Science and Statistics, The University of Texas at Dallas, Texas 75080-3021, USA*

[*]Corresponding author: Md Farhad Hossain, Email: farhad390ju@gmail.com

Abstract Inventories are among the largest working-capital investments of retailers and manufacturers, and inventory imbalances destroy value: overhang ties up capital and forces markdowns and production cuts, while shortage costs sales and margins. Lenders, investors and managers would benefit from treating inventory stress like credit or financial-market stress, as a risk event that can be priced and forecast. This study develops an explainable early-warning framework for inventory stress in U.S. retail and manufacturing. Stress is defined from the Hamilton regression-filter gap of the Census inventories-to-sales ratios, separating overhang (excess stocks) from shortage (lean stocks) episodes between 1994 and 2026. Predictors cover trade conditions (imports, exports, the trade balance, import prices, the real dollar and wholesale inventories) and macro-financial conditions (activity, orders, sales, prices, unemployment, sentiment, the policy rate, the term spread, the credit spread and oil), each aligned to its publication lag. A persistence benchmark (AR-logit), a logit with moving-block bootstrap inference, an elastic-net logit, a random forest, XGBoost and an equal-weight ensemble are compared in a pseudo-real-time evaluation (April 2005–July 2026) with information-set ablation and SHAP attributions. In sample, weak industrial production, a flat yield curve, wide credit spreads and wholesale stock accumulation precede manufacturing overhang, the financing channel of inventory investment, but bootstrap standard errors are up to 14.93 times the conventional ones. Out of sample, stress is predictable (mean AUC 0.89 one month and 0.81 three months ahead), yet no model significantly beats the AR-logit, and the best-fitting logit forecasts worst. Trade and macro-financial information helps mainly for shortage stress, especially in retail (three-month AUC 0.83 for the random forest versus 0.80), and its SHAP share rises with the horizon. Persistence benchmarks and dependence-robust evaluation should therefore be standard in financial and operational stress prediction.

Keywords inventory stress · working capital · credit spreads · early-warning system · financial stress prediction · international trade · machine learning · forecast evaluation

JEL Classification G17 · G32 · E22 · E32 · E44 · F14 · C53

1 Introduction

Inventories are a financial investment. For U.S. retailers and manufacturers they are among the largest working-capital items on the balance sheet, financed by trade credit, bank lines and short-term debt, and their carrying cost rises with interest rates and credit spreads (Gertler and Gilchrist 1994; Maccini et

al. 2004). When sales fall faster than orders can be cancelled, unwanted stocks accumulate, capital is tied up, markdowns erode margins and firms cut production to liquidate stocks; when demand outruns replenishment, stock-outs and lost sales follow. The value consequences are measurable: inventory and margin data improve forecasts of retailers' sales (Kesavan et al. 2010), and inventory turnover varies systematically with firm fundamentals (Gaur et al. 2005). At the macroeconomic level, changes in inventory investment account for a disproportionate share of output declines in U.S. recessions (Blinder 1981; Blinder and Maccini 1991; Ramey and West 1999), the accelerator mechanism of Metzler (1941) makes inventory imbalances a source of business-cycle volatility, and the bullwhip effect amplifies them along supply chains (Lee et al. 1997; Cachon et al. 2007; Dooley et al. 2010).

The years since 2020 have made inventory risk a first-order concern for investors and lenders. The retail inventories-to-sales ratio, which averaged 1.53 before 2020, collapsed to a record low of 1.09 in June 2021 as stimulus-fuelled demand met disrupted supply, while the manufacturing ratio spiked to 1.88 in April 2020 when sales froze. Rapid restocking, an import surge ahead of the 2025 tariff increases and renewed swings in the dollar, interest rates and commodity prices followed. These episodes were driven as much by trade and financial conditions, including container shortages, import prices, the exchange rate and front-loading of foreign orders, as by domestic demand (Amiti et al. 2019; Alessandria et al. 2023; Lafrogne-Joussier et al. 2023; Benigno et al. 2022). For credit analysts assessing borrowers' working capital, for investors exposed to retail and industrial firms and for policy-makers monitoring supply–demand imbalances, the relevant question is whether observable trade and macro-financial conditions give early warning of inventory stress, in the same way that financial indicators give early warning of credit and market stress.

Four literatures bear on this question. Macroeconomic and financial research explains inventory behaviour through production smoothing, stock-out avoidance and (S, s) adjustment with non-convex costs (Kahn 1987; Ramey 1991; Bils and Kahn 2000; Khan and Thomas 2007; Iacoviello et al. 2011). Because inventories must be financed, monetary policy and credit conditions transmit to inventory investment, especially for financially constrained firms (Gertler and Gilchrist 1994; Maccini et al. 2004), and better inventory management has been linked to the moderation of output volatility (McConnell and Perez-Quiros 2000; Kahn et al. 2002). Open-economy models show that importers hold larger and lumpier inventories, so exchange rates and import prices are transmitted through stock adjustment (Alessandria et al. 2010), and destocking helps to explain the Great Trade Collapse of 2008–09 (Bems et al. 2013). This literature is structural; it does not ask whether trade and financial data predict sector-level imbalances out of sample.

Operations and forecasting research documents long-run declines in inventory holdings (Chen et al. 2005), cross-industry determinants of inventory levels (Rumyantsev and Netessine 2007) and the links between demand forecasts and replenishment (Syntetos et al. 2016; Fildes et al. 2022). Goltsos et al. (2022) argue that forecasting and inventory control remain fragmented and call for integrative empirical work. Almost all of this research operates at the item or firm level and treats the macro-financial and trade environment as exogenous, even though disruptions propagate across firms and sectors (Carvalho et al. 2021; Ivanov 2020). The trade literature quantifies the aggregate effects of tariffs and supply-chain pressure (Fajgelbaum et al. 2020; Alessandria et al. 2023) and of trade-policy and economic-policy uncertainty (Baker et al. 2016; Caldara et al. 2020), but does not evaluate trade indicators as predictors of sectoral inventory imbalances.

The closest methodological literature is financial early warning. Binary models have a long tradition for currency crises (Kaminsky et al. 1998) and recessions, where the term spread and credit spreads are

robust predictors (Estrella and Mishkin 1998; Gilchrist and Zakrajšek 2012; Liu and Moench 2016), dynamic binary models exploit persistence (Kauppi and Saikkonen 2008) and ROC-based evaluation is standard (Berge and Jordà 2011). Boosted trees and random forests increasingly compete with logit models (Döpke et al. 2017; Vrontos et al. 2021; Goulet Coulombe et al. 2022). Recent studies forecast equity-market stress from macroeconomic predictors with logit models and bootstrap inference (Mofokeng 2026), model the response of stock-market volatility to macroeconomic factors with multivariate GARCH (Jakšić et al. 2025) and predict stock-market movements with machine learning on macroeconomic, technical and sentiment indicators (Patsiarikas et al. 2025). These studies typically rely on in-sample fit, ignore publication lags and rarely benchmark against the persistence of the stress state, so the incremental value of macro-financial information remains unclear.

Research gap. These strands leave a clear gap. Inventory research explains imbalances after the fact, while financial early-warning research has focused on market, credit and real-activity outcomes and has not been applied to the working-capital risk embodied in inventories. To our knowledge, no study (i) frames sector-level inventory imbalance as a probabilistic early-warning problem comparable to financial stress prediction; (ii) separates the predictive content of trade conditions (import and export flows, import prices, the real dollar and the wholesale pipeline) from that of macro-financial conditions (activity, the yield curve, the policy rate and credit spreads); (iii) distinguishes overhang (glut) from shortage (stock-out) stress, which have different drivers and different costs for firm value; or (iv) evaluates such predictions in a pseudo-real-time design that respects publication lags and benchmarks against the persistence of the inventory position itself. The study therefore pursues three research objectives. **RQ1** identifies the trade and macro-financial conditions associated with the onset of overhang and shortage stress in U.S. retail and manufacturing. **RQ2** tests whether these conditions improve out-of-sample predictions one, three and six months ahead relative to climatology and to a model that uses only the sector's own inventory state. **RQ3** measures whether the predictive value of trade information is distinct from that of macro-financial information and how it varies across sectors, stress types and horizons.

The article makes four contributions. First, it proposes a transparent, reproducible definition of sector-level inventory stress based on the Hamilton (2018) regression filter, distinguishing overhang from shortage and documenting a chronology of stress episodes from 1994 to 2026. Second, it assembles a monthly data set of 17 FRED series and the Census advance wholesale inventories series, with every predictor aligned to its publication lag. Third, it brings the evaluation standards of financial early-warning research to inventory risk, providing the first systematic out-of-sample comparison of logit, penalised logit, random forest, gradient boosting and ensemble classifiers, with dependence-robust inference and an information-set ablation that isolates the value of trade data. Fourth, it uses SHAP attributions to show which conditions drive predicted stress in each sector, linking the forecasts to the financing channel of inventory investment and to credit and investment decisions. Section 2 describes the methodology, Section 3 presents the results, Section 4 discusses them, and Section 5 concludes.

2 Methodology

Figure 1 summarises the design. Every number is produced by the Python code supplied with the article from the input files described below.

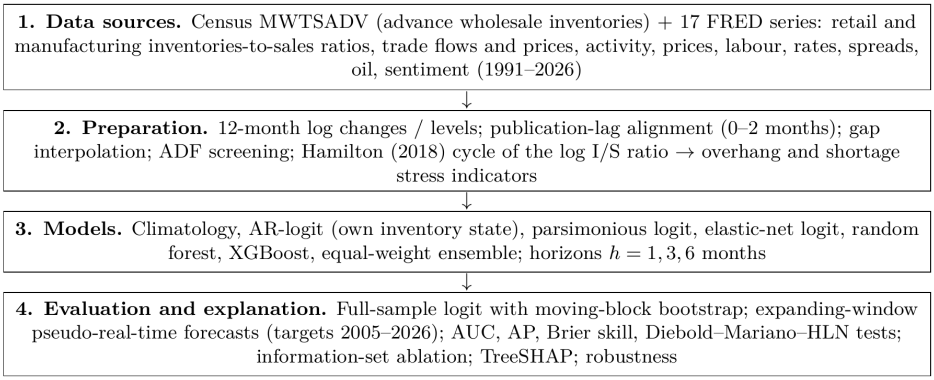


Figure 1 Research workflow

2.1 Data

The targets are the seasonally adjusted inventories-to-sales (I/S) ratios of U.S. retailers (RETAILIRSA) and manufacturers (MNFCTRIRSA) from the Census Bureau's Manufacturing and Trade Inventories and Sales programme, distributed through FRED (McCracken and Ng 2016), monthly from January 1992 to July 2026. The I/S ratio measures the months of sales held in stock and is the standard summary of the inventory position (Blinder and Maccini 1991; Ramey and West 1999; Chen et al. 2005). Predictors fall into two groups (Table 1). *Trade conditions* are imports and exports of goods and services, the import price index, the BIS real broad effective dollar and the Census advance estimate of merchant wholesalers' inventories (MWTSADV), the distribution tier between importers and manufacturers on one side and retailers on the other. *Macroeconomic conditions* are industrial production, manufacturers' new orders, retail sales, consumer prices, unemployment, consumer sentiment, the federal funds rate, the 10-year minus 3-month term spread, the Baa credit spread and the WTI oil price. They represent the demand, cost, financing and expectations channels of inventory theory (Gertler and Gilchrist 1994; Maccini et al. 2004; Kilian 2009) and of the leading-indicator literature (Estrella and Mishkin 1998; Stock and Watson 2002; Gilchrist and Zakrajšek 2012).

2.2 Data preparation and the definition of stress

Trending and not-seasonally-adjusted nominal series enter as 12-month log growth rates, $g_t^X = 100(\ln X_t - \ln X_{t-12})$, which also removes residual seasonality; rates and spreads enter as levels or 12-month changes, and the trade balance enters as the share $100(X_t^{exp} - X_t^{imp})/(X_t^{exp} + X_t^{imp})$. A single missing month (October 2025) in three series is linearly interpolated. All transformed predictors except the trade-balance share and consumer sentiment, both bounded and mean-reverting, reject a unit root in augmented Dickey–Fuller tests at the 5% level. To mimic the information available at the end of month t, each predictor enters at the date of its latest release (Table 1). The data are the latest vintage rather than real-time vintages (Croushore and Stark 2001).

Let $r_{s,t} = 100\ln(I/S_{s,t})$ for sector $s \in \{\text{retail}, \text{manufacturing}\}$. The inventory gap is the cyclical component from the Hamilton (2018) regression filter, which avoids the spurious dynamics and end-point problems of the Hodrick and Prescott (1997) filter. For monthly data it regresses the value two years ahead on the 12 most recent observations:

$$r_{s,t+24} = \beta_{s,0} + \sum_{j=0}^{11} \beta_{s,j+1}\, r_{s,t-j} + v_{s,t+24}, \qquad c_{s,t} \equiv \hat{v}_{s,t}, \tag{1}$$

Table 1 Data sources, transformations and publication lags (monthly series)

Group	Code	Description	Transformation	Lag
Target	RETAILIRSA	Retailers: inventories-to-sales ratio (SA)	Hamilton gap of 100 ln	2
Target	MNFCTRIRSA	Manufacturers: inventories-to-sales ratio (SA)	Hamilton gap of 100 ln	2
Trade	MWTSADV	Merchant wholesalers' inventories, advance estimate (SA)	Δ_{12} ln	1
Trade	BOPTIMP, BOPTEXP	Imports and exports of goods and services (SA)	Δ_{12} ln; balance share	2
Trade	IR	Import price index, all commodities	Δ_{12} ln	1
Trade	RBUSBIS	Real broad effective exchange rate (BIS)	Δ_{12} ln	1
Macro	INDPRO	Industrial production	Δ_{12} ln	1
Macro	AMTMNO	Manufacturers' new orders	Δ_{12} ln	2
Macro	RSAFS	Retail and food-services sales	Δ_{12} ln	1
Macro	CPIAUCSL	Consumer price index	Δ_{12} ln	1
Macro	UNRATE	Unemployment rate	Δ_{12}	1
Macro	UMCSENT	University of Michigan consumer sentiment	level	0
Macro	FEDFUNDS	Effective federal funds rate	Δ_{12}	0
Macro	T10Y3MM	10-year minus 3-month Treasury spread	level	0
Macro	BAA10YM	Baa corporate minus 10-year Treasury yield	level	0
Macro	MCOILWTICO	WTI crude oil price	Δ_{12} ln	0

Notes: Lag is the number of months by which a series is delayed at origin t to reflect its release schedule. Series run from 1991–1994 to July–September 2026. The producer price index (PPIACO) is used descriptively only because it is nearly collinear with import prices and oil.

so that $c_{s,t}$ is the part of the inventory position not predictable from its own history two years earlier. The gap is available from December 1994 to July 2026. Two binary stress states are defined from its tails,

$$S^{O}_{s,t} = \mathbb{1}\left[c_{s,t} > Q_{0.85}(c_s)\right], \qquad S^{U}_{s,t} = \mathbb{1}\left[c_{s,t} < Q_{0.15}(c_s)\right], \tag{2}$$

where Q_q is the q-quantile. Overhang stress (S^O) marks months in which stocks are unusually high relative to sales, the classic glut that precedes production cuts; shortage stress (S^U) marks unusually lean stocks and stock-out risk. Quantile thresholds fix the base rate at 15% and are not inflated by the extreme pandemic observations. Thresholds of 10%/90% and 20%/80% and a one-sided gap relative to the trailing 24-month mean serve as robustness checks.

2.3 Forecasting design and models

For horizon $h \in \{1,3,6\}$ and origin t, the target is $y_{s,t+h} = S_{s,t+h}$ and the predictor vector combines the sector's own inventory state (the gap $c_{s,t-2}$ and the three-month change $\Delta_3 r_{s,t-2}$), six trade predictors and ten macroeconomic predictors, each at its publication lag. Because the label of month τ is known only once its I/S ratio is published, a model used at origin t is trained on $\mathscr{D}_t = \{(\boldsymbol{x}_{s,\tau}, y_{s,\tau+h}) : \tau + h \le t-2\}$, an expanding window. The first origin is January 2005 (at least 112 training observations), models are re-estimated each quarter, and predictors are standardised with training-window moments.

Seven forecasts are compared. *Climatology* is the training stress frequency. The *AR-logit* uses only the two own-state variables and captures the persistence of imbalances, the natural benchmark for any claim that external conditions add information (Kauppi and Saikkonen 2008). The *parsimonious logit* relates stress to twelve predictors chosen to keep variance inflation factors near 5 or below,

$$\Pr(y_{s,t+h} = 1 \mid \boldsymbol{x}_{s,t}) = \Lambda(\alpha + \boldsymbol{x}_{s,t}^{\top}\boldsymbol{\beta}) = \frac{1}{1+\exp\{-(\alpha + \boldsymbol{x}_{s,t}^{\top}\boldsymbol{\beta})\}}, \tag{3}$$

so that e^{β_j} is the odds ratio for a one-standard-deviation increase. The *elastic-net logit* (Zou and Hastie 2005) enters all 18 predictors with penalty and mixing parameters chosen by time-ordered cross-validation on log loss. The *random forest* (Breiman 2001) averages 500 trees with at least five observations per leaf, and *XGBoost* (Chen and Guestrin 2016) uses 300 depth-2 trees with a learning rate of 0.03 and subsampling of 0.8; these values were fixed a priori. The *ensemble* is the equal-weight average of the AR-logit, elastic-net, random forest and XGBoost probabilities (Bates and Granger 1969).

2.4 Inference and evaluation

Conventional standard errors ignore the strong serial dependence of the monthly stress indicator. Following the bootstrap approach of Mofokeng (2026) but respecting time dependence, we use the moving-block bootstrap (Künsch 1989) with 12-month blocks and 999 resamples, reporting percentile intervals and two-sided bootstrap p-values; HAC standard errors (Newey and West 1987) are stored for comparison. Out of sample, discrimination is measured by the AUC, the probability that a randomly chosen stress month receives a higher predicted probability than a randomly chosen calm month (Berge and Jordà 2011), and by average precision. Accuracy and calibration are measured by the Brier score (Brier 1950) and the Brier skill score relative to climatology, $BSS_m = 1 - BS_m/BS_{clim}$. Brier losses are compared with the Diebold and Mariano (1995) test with the small-sample correction of Harvey et al. (1997),

$$DM^* = \frac{\bar{d}}{\sqrt{\hat{\sigma}_d^2/n}} \sqrt{\frac{n+1-2h+h(h-1)/n}{n}} \sim t_{n-1}, \tag{4}$$

where d_t is the loss differential and $\hat{\sigma}_d^2$ its Newey–West long-run variance. AUC intervals and the p-value of $AUC_m - AUC_{AR} \leq 0$ come from 1,000 block-bootstrap resamples. A warning is issued when the predicted probability exceeds the training base rate, and its sensitivity and specificity are reported as in the early-warning literature (Kaminsky et al. 1998). An information-set ablation re-estimates the elastic-net logit and XGBoost on four nested sets: own state; own state plus macro; own state plus trade; and all predictors. Full-sample XGBoost models are explained with exact TreeSHAP values (Lundberg and Lee 2017; Lundberg et al. 2020), and group importance is the share of mean absolute SHAP values attributable to own-state, trade and macroeconomic predictors.

3 Results

3.1 Inventory positions and the stress chronology

The two I/S ratios behave very differently (Fig. 2). The retail ratio drifted down from about 1.7 in the mid-1990s to roughly 1.5 before the pandemic, reflecting lean replenishment and information technology (Chen et al. 2005), collapsed to 1.09 in June 2021 and stood at 1.27 in July 2026, still well below its pre-2020 mean of 1.53. The manufacturing ratio rose after 2012, peaked at 1.88 in April 2020 and ended the sample at 1.47. The two gaps are essentially uncorrelated ($\rho = -0.03$), and joint overhang months account for only 6.3% of the sample, so inventory stress is sector-specific and justifies separate models.

Table 2 summarises the chronology. Manufacturing overhang occurs in six episodes, including the 2008–09 recession, the 2015–16 slowdown associated with the oil collapse and the strong dollar, and a long episode from September 2019 to December 2020 spanning the tariff escalation and the pandemic. Manufacturing shortage is concentrated in the 2003–2006 expansion, 2011 and 2022. Retail shortage is dominated by the pandemic episode of May 2020–July 2022, when stimulus-supported demand outran

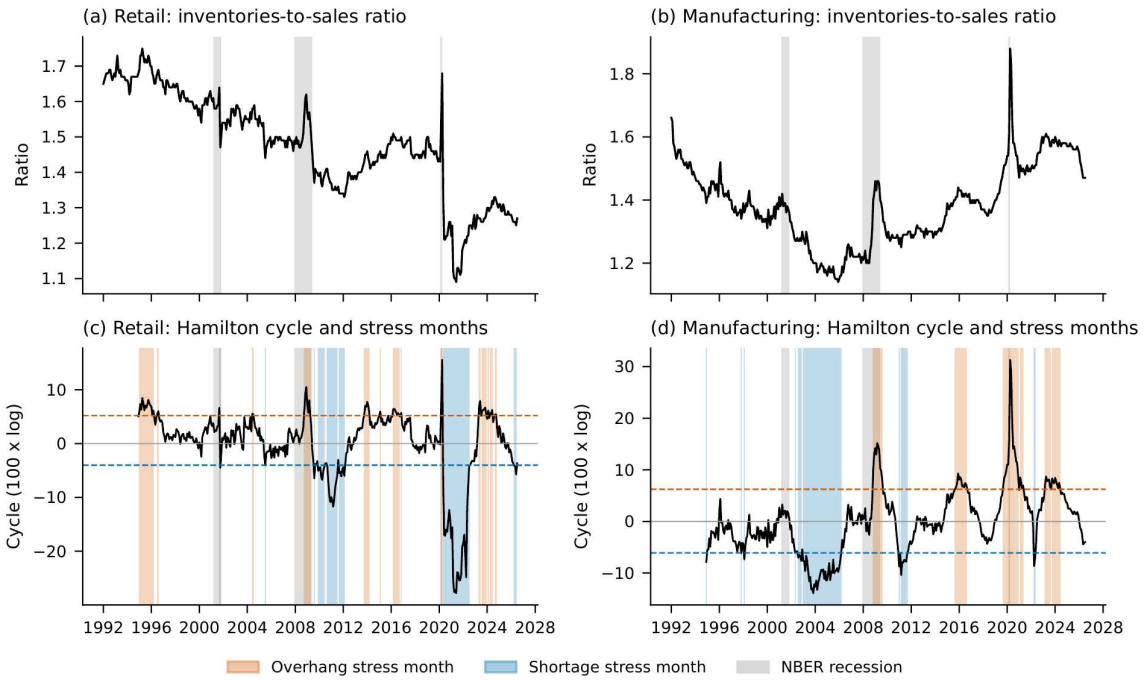


Figure 2 Inventories-to-sales ratios (a, b) and Hamilton-filter gaps with stress months (c, d). Dashed lines are the 85th and 15th percentile thresholds; grey bands are NBER recessions

replenishment, while retail overhang appears in short bursts (15 episodes, mean length 3.8 months). Stress is highly persistent: the probability that a stress month is followed by another ranges from 74% (retail overhang) to 89% (manufacturing overhang). The chronology is not an artefact of the filter, as the trailing-mean definition classifies 85–90% of months identically. Several trade and activity predictors are strongly correlated (import growth with export growth, new orders and import prices), which motivates the parsimonious logit (maximum VIF 6.39 in manufacturing and 6.67 in retail) and the use of regularised and tree-based models.

Table 2 Inventory stress chronology, December 1994–July 2026

Sector	Stress	Threshold	Months	Share (%)	Episodes	Mean length	Longest episode
Retail	Overhang	+5.20	57	15.0	15	3.8	1995-01–1996-03
Retail	Shortage	−4.01	57	15.0	8	7.1	2020-05–2022-07
Manufacturing	Overhang	+6.24	57	15.0	6	9.5	2019-09–2020-12
Manufacturing	Shortage	−6.10	57	15.0	9	6.3	2003-01–2006-03

Notes: Thresholds are the 85th (overhang) and 15th (shortage) percentiles of the Hamilton gap of $100\ln(I/S)$.

3.2 Conditions associated with stress (RQ1)

Table 3 and Fig. 3 report the full-sample logit for stress three months ahead. Three results stand out. First, the own inventory gap dominates: a one-standard-deviation larger gap multiplies the odds of retail overhang by 15.76 (bootstrap interval 3.14 to above 99) and the odds of retail shortage by 0.05, consistent with slow, (S,s)-type adjustment (Ramey 1991; Khan and Thomas 2007).

Second, trade and macroeconomic conditions carry economically meaningful signals that differ by sector and stress type. Manufacturing overhang is more likely when industrial production growth is weak (odds ratio 0.21), when the yield curve is flat or inverted (term spread 0.06) and, at the 10%

level, when the wholesale tier is accumulating stocks (9.07) and credit spreads are wide (13.26). These are the cyclical and financing channels of inventory theory: weakening activity leaves unsold output in the pipeline, and tighter financing raises carrying costs (Gertler and Gilchrist 1994; Maccini et al. 2004; Gilchrist and Zakrajšek 2012). The term spread works in the opposite direction for manufacturing shortage (27.01), in line with its role as a business-cycle predictor (Estrella and Mishkin 1998): a steep curve signals a recovery in which sales outrun production. Manufacturing shortage is also associated, at the 10% level, with a weaker trade balance (0.11) and lower import-price inflation (0.12). In retail, a higher trade-balance share, typically the result of collapsing imports in a demand slump, raises the odds of overhang (2.83), while faster wholesale inventory growth lowers the odds of retail shortage (0.35), consistent with the upstream tier buffering retailers (Lee et al. 1997).

Third, inference must respect time dependence. Block-bootstrap standard errors are on average 2.22 times the conventional ones for retail overhang and 10.55–14.93 times for the other three models, whose in-sample fit is very high (McFadden R^2 0.61–0.77; AUC up to 0.98). Near-separation in a sample with few, long episodes means that conventional standard errors greatly overstate precision, a stronger version of the understatement reported by Mofokeng (2026) for equity-market stress, and that high in-sample fit is a poor guide to predictive value.

Table 3 Full-sample logit for stress three months ahead: odds ratios per one-standard-deviation increase with 95% moving-block bootstrap intervals

		Retail		Manufacturing	
Group	Predictor	Overhang	Shortage	Overhang	Shortage
Own	Own inventory gap	15.76** [3.14, >99]	0.05** [0.00, 0.29]	36.80* [0.95, >99]	0.06* [0.00, 1.01]
Own	Own I/S change (3m)	0.69 [0.26, 2.14]	2.03* [0.85, 32.38]	1.16 [0.00, 30.43]	0.51 [0.00, >99]
Trade	Import growth	1.71 [0.15, 7.04]	1.38 [0.03, 21.28]	3.22 [0.00, >99]	4.43 [0.00, >99]
Trade	Trade balance share	2.83** [1.02, 21.44]	2.79 [0.11, >99]	5.67 [0.23, >99]	0.11* [0.00, 1.16]
Trade	Import price inflation	1.35 [0.25, 8.96]	0.76 [0.11, >99]	0.30 [0.00, >99]	0.12* [0.00, 1.22]
Trade	Real dollar appreciation	0.48 [0.17, 1.71]	0.25 [0.04, 16.30]	0.23 [0.00, 2.67]	0.75 [0.00, 6.17]
Trade	Wholesale inventory growth	1.28 [0.26, 6.04]	0.35* [0.00, 1.27]	9.07* [0.92, >99]	3.34 [0.00, >99]
Macro	Industrial production growth	0.61 [0.09, 2.70]	0.47 [0.02, 21.37]	0.21** [0.00, 0.70]	6.23 [0.02, >99]
Macro	Consumer sentiment	0.64 [0.25, 7.60]	0.81 [0.01, 9.48]	1.65 [0.02, >99]	1.45 [0.01, >99]
Macro	Fed funds change	2.57 [0.51, 11.66]	0.93 [0.01, 16.19]	0.63 [0.00, >99]	0.53 [0.02, >99]
Macro	Term spread	0.77 [0.23, 5.46]	0.57 [0.05, 13.42]	0.06** [0.00, 0.46]	27.01** [5.52, >99]
Macro	Credit spread	2.58 [0.52, 12.28]	2.72 [0.29, 53.10]	13.26* [0.31, >99]	0.39 [0.00, >99]
	McFadden R^2 / AUC	0.35 / 0.88	0.61 / 0.95	0.67 / 0.98	0.77 / 0.98
	Mean SE_{boot}/SE_{ML}	2.22	10.55	14.93	11.70

Notes: Block length 12, 999 resamples. ** / *: bootstrap $p < 0.05$ / $p < 0.10$.

3.3 Out-of-sample prediction (RQ2)

Table 4 and Figs. 4–6 report the expanding-window forecasts for target months April 2005–July 2026 ($n = 256$ at $h = 3$).

Stress is predictable. Every model except the parsimonious logit discriminates well at short horizons: averaged over the four sector–stress combinations, the one-month-ahead AUC is 0.89 for the AR-logit, 0.86 for the elastic-net logit, 0.85 for the random forest, 0.84 for XGBoost and 0.89 for the ensemble. Discrimination decays with the horizon (three months: 0.81 for the AR-logit and 0.79 for the ensemble; six months: 0.68 and 0.66), and probabilistic accuracy decays faster. Brier skill relative to

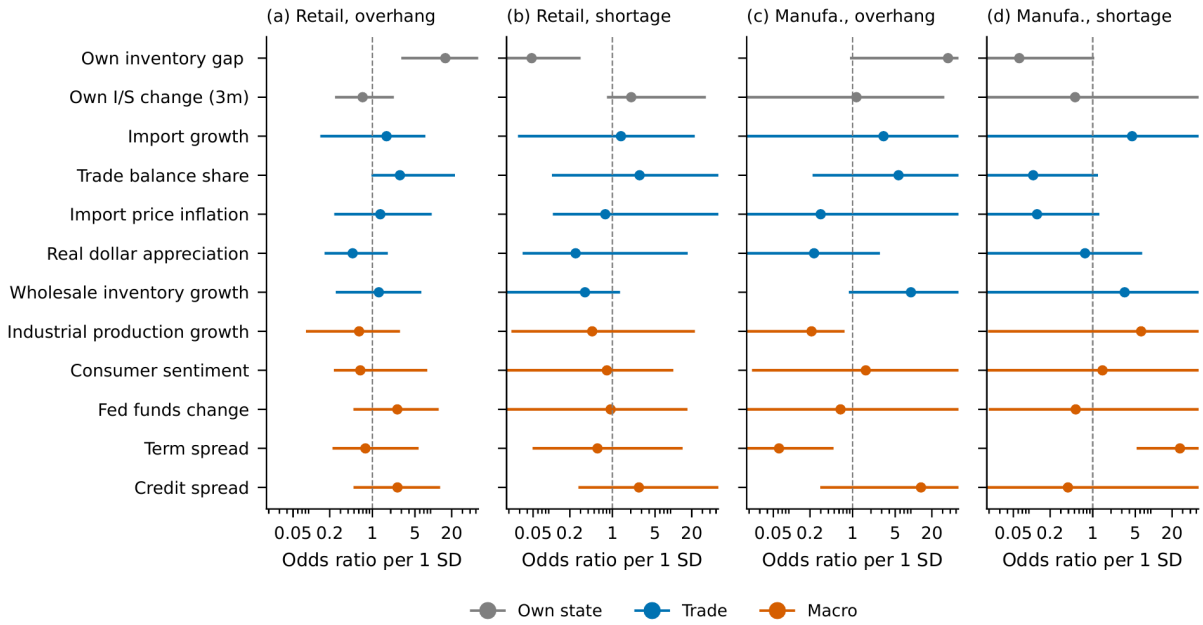


Figure 3 Odds ratios (log scale) per one-standard-deviation increase for stress three months ahead, with 95% moving-block bootstrap intervals (truncated at $e^{\pm 4}$)

climatology is positive in 11 of 12 sector–stress–horizon combinations for the AR-logit and 10 for the ensemble, and the Diebold–Mariano–HLN test rejects equal accuracy with climatology in six combinations for each, almost all at $h \leq 3$. Manufacturing shortage is the most predictable state (three-month AUC 0.92 for the elastic-net logit) and retail overhang the least (0.76 for the AR-logit, with no other model above 0.66).

The persistence benchmark is hard to beat. The AR-logit, which knows only the sector's own gap and its recent change, is the best or statistically indistinguishable from the best in every combination. Trade-and-macro models exceed its AUC in a minority of cases (random forest and XGBoost 3 of 12; elastic-net logit 2), the ensemble in 7 of 12, and none of these gains is significant at 5%; no model has significantly lower Brier loss than the AR-logit. The parsimonious logit, which had the best in-sample fit, has the weakest out-of-sample record (mean AUC 0.71 against 0.79 for the AR-logit) and negative Brier skill in most combinations.

Where external conditions help. Gains are concentrated in two places. The first is retail shortage, where the tree ensembles outperform the AR-logit at every horizon (AUC 0.90 versus 0.87 at $h = 1$, 0.83 versus 0.80 at $h = 3$, 0.72 versus 0.66 at $h = 6$). Once an episode is under way the AR-logit issues sharper probabilities, which is why its Brier skill remains highest, but it assigns near-zero probability to several stress months at the start or in the interior of episodes (January 2012 and September 2020), whereas the tree ensembles, drawing on import growth, monetary conditions and activity, keep moderate probabilities throughout (Fig. 6b). The second is manufacturing shortage, where the elastic-net logit is best at $h = 1$ and $h = 3$ (0.96, 0.92) and the parsimonious logit at $h = 6$ (0.79 versus 0.76). For overhang stress, external conditions add noise rather than signal out of sample. With warnings issued when the predicted probability exceeds the training base rate, the ensemble flags 77% of manufacturing shortage months and 89% of manufacturing overhang months three months ahead, with specificities of 0.93 and 0.56.

Table 4 Out-of-sample performance, expanding window, target months April 2005–July 2026

	$h=1$		$h=3$		$h=6$			$h=1$		$h=3$		$h=6$	
Model	AUC	BSS	AUC	BSS	AUC	BSS	Model	AUC	BSS	AUC	BSS	AUC	BSS
Retail, overhang							*Manufacturing, overhang*						
AR-logit	0.86	0.21*	0.76	0.01	0.64	−0.12	AR-logit	0.89	0.49**	0.79	0.31**	0.67	0.09
Logit	0.76	−0.17	0.66	−0.69	0.48	−0.71	Logit	0.78	0.09	0.61	−0.29	0.59	−0.42
EN	0.81	0.05	0.59	−0.07	0.46	−0.09	EN	0.86	0.28**	0.70	0.06*	0.60	0.01
RF	0.70	0.11*	0.49	−0.05	0.46	−0.09	RF	0.86	0.30**	0.78	0.19**	0.64	0.05
XGB	0.69	0.08	0.51	−0.14	0.45	−0.15	XGB	0.85	0.25**	0.77	0.09	0.67	0.05
Ensemble	0.80	0.17*	0.64	−0.02	0.49	−0.06	Ensemble	0.90	0.40**	0.78	0.20**	0.66	0.10*
Retail, shortage							*Manufacturing, shortage*						
AR-logit	0.87	0.56**	0.80	0.41**	0.66	0.23	AR-logit	0.94	0.64**	0.90	0.47**	0.76	0.21
Logit	0.73	−0.02	0.60	−0.25	0.69	0.05	Logit	0.94	0.06	0.85	0.31**	0.79	−0.28
EN	0.80	0.45**	0.72	0.22**	0.60	0.11	EN	0.96	0.59**	0.92	0.39**	0.73	0.05
RF	0.90	0.39**	0.83	0.28**	0.71	0.15	RF	0.94	0.47**	0.88	0.25**	0.75	−0.06
XGB	0.90	0.39**	0.81	0.22*	0.72	0.06	XGB	0.92	0.59**	0.84	0.27	0.71	−0.07
Ensemble	0.89	0.49**	0.83	0.34**	0.70	0.19*	Ensemble	0.96	0.63**	0.90	0.43**	0.78	0.18

Notes: EN: elastic-net logit; RF: random forest; XGB: XGBoost. BSS: Brier skill score relative to climatology; ** / *: Diebold–Mariano–HLN test of equal Brier loss against climatology, one-sided $p < 0.05$ / $p < 0.10$. No AUC exceeds that of the AR-logit significantly at 5% under the block bootstrap.

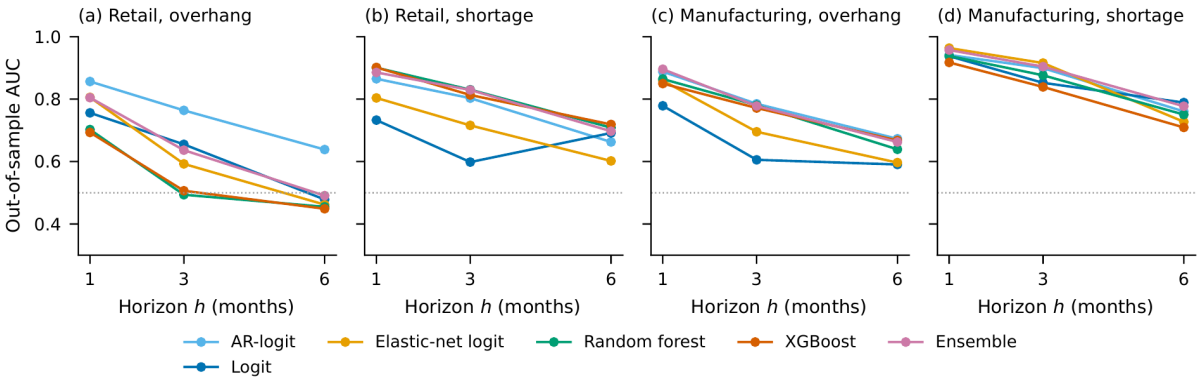


Figure 4 Out-of-sample AUC by forecast horizon

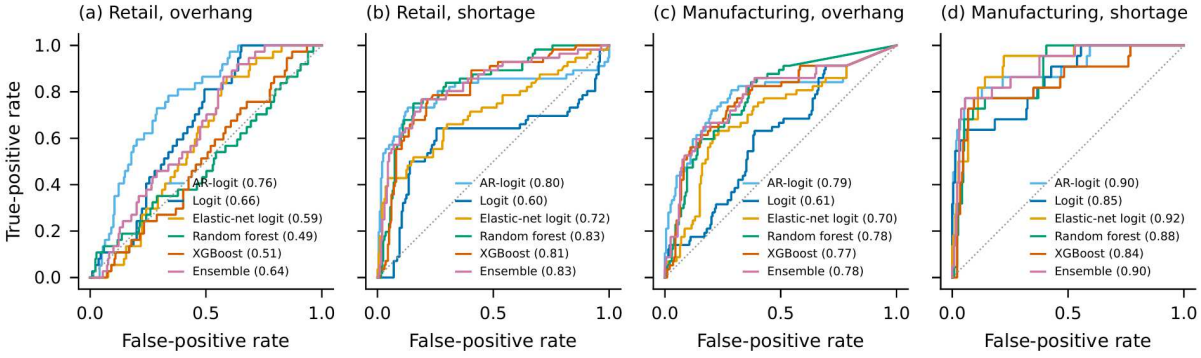


Figure 5 Out-of-sample ROC curves at $h = 3$ (AUC in parentheses)

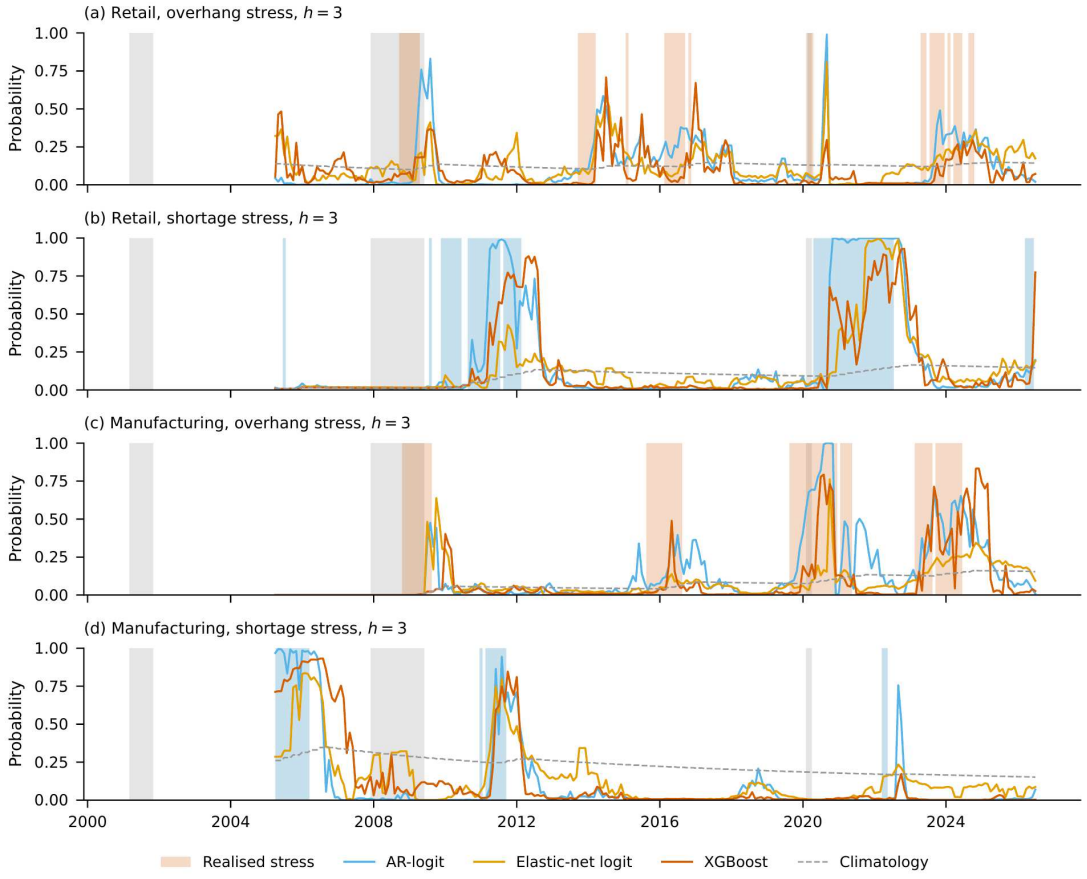


Figure 6 Out-of-sample stress probabilities three months ahead, plotted at the target month, with realised stress months shaded. Grey bands are NBER recessions

3.4 The value of trade versus macro-financial information (RQ3)

Table 5 isolates each information set while holding the model class fixed. Averaged over all 12 combinations, the elastic-net logit attains an AUC of 0.75 with the own state alone, 0.74 with macroeconomic conditions added, 0.72 with trade conditions added and 0.73 with both; for XGBoost the values are 0.75, 0.76, 0.72 and 0.74. On average, neither information set improves on the own inventory state, and trade information alone is the weaker of the two. The averages conceal clear heterogeneity. For retail shortage, adding macroeconomic conditions raises the elastic-net AUC from 0.66 to 0.76 and the XGBoost AUC from 0.76 to 0.82 (means over horizons). For manufacturing shortage, trade and macro together lift the elastic-net AUC from 0.84 to 0.87. For retail overhang, every addition lowers the AUC (elastic net 0.74 to 0.62; XGBoost 0.67 to 0.55), a symptom of overfitting when the target has few, short episodes.

Table 5 Information-set ablation: out-of-sample AUC at $h = 3$ with nested predictor sets

		Elastic-net logit				XGBoost			
Sector	Stress	Own	+Macro	+Trade	+Both	Own	+Macro	+Trade	+Both
Retail	Overhang	0.76	0.65	0.71	0.59	0.68	0.63	0.55	0.51
Retail	Shortage	0.69	0.76	0.72	0.72	0.81	0.84	0.81	0.81
Manufacturing	Overhang	0.74	0.71	0.73	0.70	0.77	0.78	0.77	0.77
Manufacturing	Shortage	0.93	0.91	0.85	0.92	0.87	0.84	0.83	0.84
Mean, all horizons		0.75	0.74	0.72	0.73	0.75	0.76	0.72	0.74

The SHAP decomposition of the full-sample XGBoost models gives a complementary, explanatory view (Table 6, Fig. 7). Across the twelve models, the own state accounts for 27% of mean absolute attributions, trade conditions for 27% and macroeconomic conditions for 46%. Industrial production growth is the most important predictor of manufacturing overhang, ahead of the own gap and the term spread; the own gap, the trade balance and the term spread lead for manufacturing shortage; and the own gap, the fed funds change and import growth lead for retail shortage. The share attributed to trade conditions rises with the horizon in every case except retail shortage: for manufacturing overhang it increases from 10% at $h = 1$ to 34% at $h = 6$. Trade conditions are thus slow-moving, longer-horizon drivers of where inventory positions are heading, whereas at short horizons the inventory position itself carries most of the information.

Table 6 Share (%) of mean absolute SHAP values attributable to the own inventory state, trade and macroeconomic conditions

		$h = 1$			$h = 3$			$h = 6$		
Sector	Stress	Own	Trade	Macro	Own	Trade	Macro	Own	Trade	Macro
Retail	Overhang	31	20	49	25	29	46	16	36	48
Retail	Shortage	31	25	44	33	24	43	24	29	47
Manufacturing	Overhang	44	10	46	20	19	61	11	34	55
Manufacturing	Shortage	38	25	36	34	32	34	17	40	43

3.5 Robustness

Table 7 repeats the three-month-ahead exercise under alternative definitions and samples, and the conclusions are unchanged. With stricter or looser thresholds the AR-logit remains the benchmark to beat

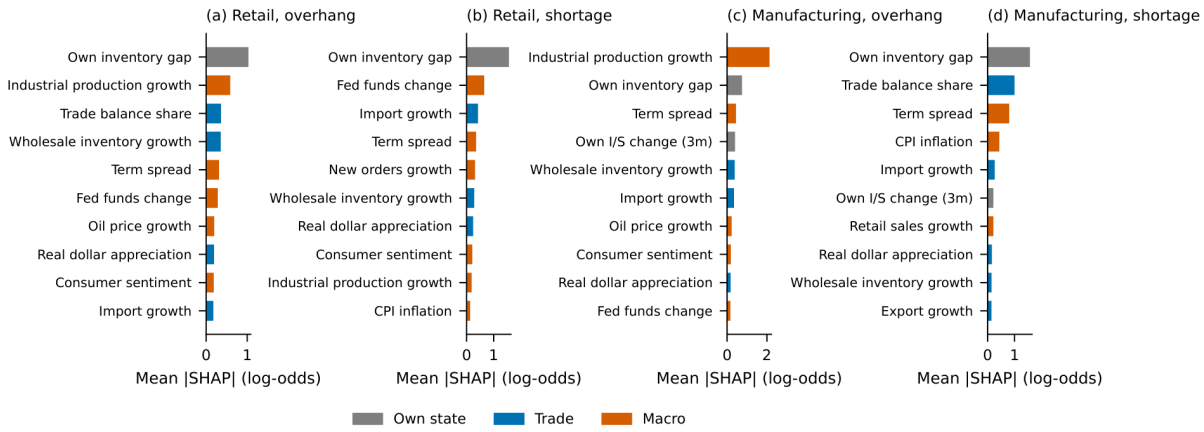


Figure 7 Mean absolute SHAP values of the ten most important predictors in the full-sample XGBoost models at $h = 3$, coloured by predictor group

for retail overhang, and flexible models remain competitive or better for shortage stress. With the one-sided trailing-mean gap, which uses no future information even in the label, the random forest improves markedly on the AR-logit for retail shortage (0.82 versus 0.68). Excluding the pandemic target months or restricting evaluation to targets before 2020 leaves the ranking intact; the random forest's advantage for retail shortage is in fact larger before 2020, so it is not driven by the pandemic.

Table 7 Robustness of out-of-sample discrimination at $h = 3$: AUC of the AR-logit / best other model

	Retail		Manufacturing	
Specification	Overhang	Shortage	Overhang	Shortage
Baseline (Hamilton, 15%/85%)	0.76 / 0.66 (Lg)	0.80 / 0.83 (RF)	0.79 / 0.78 (RF)	0.90 / 0.92 (EN)
Stricter threshold (10%/90%)	0.67 / 0.64 (Lg)	0.81 / 0.82 (RF)	0.72 / 0.71 (Ens)	0.83 / 0.87 (Ens)
Looser threshold (20%/80%)	0.79 / 0.71 (Ens)	0.80 / 0.85 (RF)	0.83 / 0.84 (Ens)	0.88 / 0.87 (Lg)
Trailing 24-month mean gap	0.73 / 0.70 (RF)	0.68 / 0.82 (RF)	0.82 / 0.81 (Ens)	0.86 / 0.79 (Ens)
Pandemic months excluded	0.76 / 0.64 (Lg)	0.78 / 0.81 (RF)	0.76 / 0.74 (RF)	0.89 / 0.91 (EN)
Targets before 2020	0.74 / 0.67 (Lg)	0.74 / 0.87 (RF)	0.72 / 0.74 (RF)	0.93 / 0.91 (EN)

Notes: Lg: logit; EN: elastic-net logit; RF: random forest; Ens: ensemble.

4 Discussion

Economic and financial interpretation. The results support three readings of inventory theory. First, the persistence of the stress state and its dominance in out-of-sample prediction are consistent with costly, lumpy adjustment (Ramey 1991; Khan and Thomas 2007): once a sector is out of balance it tends to stay so for several months, so the inventory position itself is the best short-horizon signal, much as past volatility is the best predictor of near-term market volatility. Second, the asymmetry between overhang and shortage is economically meaningful. Overhang, which arises when sales fall short of plans, is best explained by weak activity, a flat yield curve and wide credit spreads, the demand and financing channels of inventory investment (Blinder and Maccini 1991; Gertler and Gilchrist 1994; Maccini et al. 2004; Gilchrist and Zakrajšek 2012). When financing tightens and the outlook darkens, carrying stocks becomes expensive precisely when they are hardest to sell. These signals are difficult to exploit in real time because demand collapses are abrupt. Shortage builds more gradually as supply and replenishment lag demand, so trade flows, import prices, the trade balance and monetary conditions offer usable lead time,

especially for retailers at the end of import-intensive supply chains (Alessandria et al. 2010; Alessandria et al. 2023). Third, the rising SHAP share of trade conditions at longer horizons accords with delayed propagation of trade shocks through supply chains (Lee et al. 1997; Dooley et al. 2010; Carvalho et al. 2021; Lafrogne-Joussier et al. 2023) and with the role of oil-price and dollar shocks in manufacturing downturns (Hamilton 1983; Kilian 2009).

Methodological implications for financial forecasting. The contrast between in-sample and out-of-sample results is a cautionary tale for stress-prediction studies that report only full-sample fit. The parsimonious logit, whose in-sample AUC was at least 0.88 in every case, was the weakest model out of sample, and its bootstrap standard errors were up to 14.93 times the conventional ones. Studies that evaluate macroeconomic predictors of equity-market stress in sample, even with bootstrap inference (Mofokeng 2026), or that report very high fit from macroeconomic, technical and volatility predictors of stock markets (Jakšić et al. 2025; Patsiarikas et al. 2025), may therefore overstate predictability. Publication-lag alignment, persistence benchmarks and dependence-robust tests are essential. Consistent with the recession-forecasting literature (Döpke et al. 2017; Vrontos et al. 2021; Goulet Coulombe et al. 2022), tree ensembles are most useful where non-linear interactions between demand and supply conditions matter, as in retail shortage, and forecast combination is a robust default (Bates and Granger 1969): the equal-weight ensemble was never far from the best model and beat the AR-logit in 7 of 12 combinations.

Implications for lenders, investors, managers and policy-makers. For lenders, sector-level stress probabilities are an early indicator of working-capital pressure: a rising probability of overhang, especially when credit spreads widen and the yield curve flattens, signals that borrowers' inventories may need to be financed for longer and liquidated at a discount, which matters for asset-based lending and covenant monitoring. For equity investors, inventory stress precedes markdowns and margin compression that appear in reported earnings only with a lag, consistent with the value of inventory data for sales forecasts (Gaur et al. 2005; Kesavan et al. 2010); the warnings can inform sector allocation between retail and industrial firms. For managers, the sector's own inventory gap, computable from public data with a two-month lag, provides most of the short-horizon warning, while trade and macro-financial dashboards add value for shortage risk over the three-to-six-month planning window for purchase orders, capacity and financing (Syntetos et al. 2016; Fildes et al. 2022; Goltsos et al. 2022). For policy-makers, stress probabilities are a timely indicator of supply–demand imbalances that can precede price pressures, as in 2021–22, and production cuts, as in 2008–09 and 2015–16.

5 Conclusions

This article framed inventory imbalance in U.S. retail and manufacturing as a working-capital risk and a probabilistic early-warning problem, comparable to credit and financial-market stress. Using the Census inventories-to-sales ratios, a Hamilton-filter definition of overhang and shortage stress, and 18 trade and macroeconomic series aligned to their publication lags, we compared logit, penalised logit, random forest, gradient boosting and ensemble classifiers in a pseudo-real-time design over April 2005–July 2026. Four conclusions emerge. Inventory stress is sector-specific and persistent. Trade and macro-financial conditions are associated with stress in theoretically coherent ways, with weak activity, a flat yield curve and wide credit spreads preceding manufacturing overhang, but full-sample logit estimates are fragile. Stress is predictable out of sample (one-month AUC around 0.89, three-month around 0.81), yet the sector's own inventory state is a benchmark that external information rarely beats, and the logit

with the best in-sample fit performs worst. Trade and macro-financial information is most valuable for shortage stress, particularly in retail, and at longer horizons, where the SHAP share of trade conditions rises steadily.

Limitations. The study has several limitations. The data are the latest vintage rather than real-time vintages (Croushore and Stark 2001), so revisions to inventories, trade flows and activity data may flatter both benchmarks and external-information models. The stress chronology is defined from full-sample quantiles of the Hamilton gap, so early labels use information not available at the time; the one-sided trailing-mean definition gives similar conclusions but is not a full substitute. With about 30 years of monthly data, each sector has only a handful of distinct episodes, so effective samples are small and out-of-sample comparisons have limited power, as the wide bootstrap intervals show. The analysis is at the level of total retail and total manufacturing, which masks heterogeneity across durable and non-durable goods and across retail segments. The trade predictors are aggregate flows and prices and do not capture shipping costs, delivery times, port congestion or trade-policy announcements directly. Finally, tree-model hyperparameters were fixed a priori, which protects against overfitting the evaluation period but may understate their potential.

Future directions. Real-time vintages from ALFRED and the Philadelphia Fed real-time data set would allow a genuinely real-time evaluation, and disaggregated inventory data by NAICS subsector would support panel models that share information across sectors and increase the number of stress episodes. Direct measures of supply-chain pressure (Benigno et al. 2022), freight rates, supplier delivery times and trade-policy and economic-policy uncertainty (Baker et al. 2016; Caldara et al. 2020) could sharpen the trade channel, with the tariff changes of 2025–26 as a natural test. Dynamic binary-response models that model the persistence of stress explicitly (Kauppi and Saikkonen 2008), mixed-frequency models using weekly indicators, and cost-sensitive evaluation that weighs the different costs of overhang and stock-outs would make the warnings more directly useful for decisions. Linking sector-level stress probabilities to firm-level inventory and margin data (Gaur et al. 2005; Kesavan et al. 2010) would connect the macroeconomic early-warning view to the integration of forecasting and inventory control called for by Goltsos et al. (2022).

Statements and declarations

Author contribution

Md Mizanur Rahman, Fatema Tuj Johra, Abdullah Al Shohan, Md Mazharul Islam, and Md Farhad Hossain contributed equally to the conceptualization, methodology, data curation, software, validation, investigation, formal analysis, visualization, writing, review and editing, supervision, and project administration of the study. All authors have read and approved the final manuscript and agree to be accountable for all aspects of the work.

Data Availability

All data are public: FRED, Federal Reserve Bank of St. Louis (`https://fred.stlouisfed.org`), and the U.S. Census Bureau time-series interface for MWTSADV (`https://www.census.gov/econ/currentdata/`).

Acknowledgments

Authors are grateful to one and all who, directly or indirectly, have lent their helping hand and thoughtful responses of this study.

Funding

This research received no specific grant from any funding agencies in the public, commercial, or not-for-profit sectors.

References

Alessandria G, Kaboski JP, Midrigan V (2010) Inventories, lumpy trade, and large devaluations. American Economic Review 100(5):2304–2339. https://doi.org/10.1257/aer.100.5.2304

Alessandria G, Khan SY, Khederlarian A, Mix C, Ruhl KJ (2023) The aggregate effects of global and local supply chain disruptions: 2020–2022. Journal of International Economics 146:103788. https://doi.org/10.1016/j.jinteco.2023.103788

Amiti M, Redding SJ, Weinstein DE (2019) The impact of the 2018 tariffs on prices and welfare. Journal of Economic Perspectives 33(4):187–210. https://doi.org/10.1257/jep.33.4.187

Baker SR, Bloom N, Davis SJ (2016) Measuring economic policy uncertainty. Quarterly Journal of Economics 131(4):1593–1636. https://doi.org/10.1093/qje/qjw024

Bates JM, Granger CWJ (1969) The combination of forecasts. Operational Research Quarterly 20(4):451–468

Bems R, Johnson RC, Yi KM (2013) The great trade collapse. Annual Review of Economics 5:375–400

Benigno G, di Giovanni J, Groen JJJ, Noble AI (2022) The GSCPI: a new barometer of global supply chain pressures. Staff Report 1017. Federal Reserve Bank of New York

Berge TJ, Jordà Ò (2011) Evaluating the classification of economic activity into recessions and expansions. American Economic Journal: Macroeconomics 3(2):246–277. https://doi.org/10.1257/mac.3.2.246

Bils M, Kahn JA (2000) What inventory behavior tells us about business cycles. American Economic Review 90(3):458–481. https://doi.org/10.1257/aer.90.3.458

Blinder AS (1981) Retail inventory behavior and business fluctuations. Brookings Papers on Economic Activity 1981(2):443–520

Blinder AS, Maccini LJ (1991) Taking stock: a critical assessment of recent research on inventories. Journal of Economic Perspectives 5(1):73–96. https://doi.org/10.1257/jep.5.1.73

Breiman L (2001) Random forests. Machine Learning 45(1):5–32. https://doi.org/10.1023/A:1010933404324

Brier GW (1950) Verification of forecasts expressed in terms of probability. Monthly Weather Review 78(1):1–3

Cachon GP, Randall T, Schmidt GM (2007) In search of the bullwhip effect. Manufacturing & Service Operations Management 9(4):457–479

Caldara D, Iacoviello M, Molligo P, Prestipino A, Raffo A (2020) The economic effects of trade policy uncertainty. Journal of Monetary Economics 109:38–59

Carvalho VM, Nirei M, Saito YU, Tahbaz-Salehi A (2021) Supply chain disruptions: evidence from the Great East Japan Earthquake. Quarterly Journal of Economics 136(2):1255–1321. https://doi.org/10.1093/qje/qjaa044

Chen H, Frank MZ, Wu OQ (2005) What actually happened to the inventories of American companies between 1981 and 2000? Management Science 51(7):1015–1031

Chen T, Guestrin C (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. https://doi.org/10.1145/2939672.2939785

Croushore D, Stark T (2001) A real-time data set for macroeconomists. Journal of Econometrics 105(1):111–130

Diebold FX, Mariano RS (1995) Comparing predictive accuracy. Journal of Business & Economic Statistics 13(3):253–263. https://doi.org/10.1080/07350015.1995.10524599

Dooley KJ, Yan T, Mohan S, Gopalakrishnan M (2010) Inventory management and the bullwhip effect during the 2007–2009 recession: evidence from the manufacturing sector. Journal of Supply Chain Management 46(1):12–18. https://doi.org/10.1111/j.1745-493X.2009.03183.x

Döpke J, Fritsche U, Pierdzioch C (2017) Predicting recessions with boosted regression trees. International Journal of Forecasting 33(4):745–759

Estrella A, Mishkin FS (1998) Predicting U.S. recessions: financial variables as leading indicators. Review of

Economics and Statistics 80(1):45–61. https://doi.org/10.1162/003465398557320

Fajgelbaum PD, Goldberg PK, Kennedy PJ, Khandelwal AK (2020) The return to protectionism. Quarterly Journal of Economics 135(1):1–55. https://doi.org/10.1093/qje/qjz036

Fildes R, Ma S, Kolassa S (2022) Retail forecasting: research and practice. International Journal of Forecasting 38(4):1283–1318. https://doi.org/10.1016/j.ijforecast.2019.06.004

Gaur V, Fisher ML, Raman A (2005) An econometric analysis of inventory turnover performance in retail services. Management Science 51(2):181–194

Gertler M, Gilchrist S (1994) Monetary policy, business cycles, and the behavior of small manufacturing firms. Quarterly Journal of Economics 109(2):309–340

Gilchrist S, Zakrajšek E (2012) Credit spreads and business cycle fluctuations. American Economic Review 102(4):1692–1720. https://doi.org/10.1257/aer.102.4.1692

Goltsos TE, Syntetos AA, Glock CH, Ioannou G (2022) Inventory–forecasting: mind the gap. European Journal of Operational Research 299(2):397–419. https://doi.org/10.1016/j.ejor.2021.07.040

Goulet Coulombe P, Leroux M, Stevanovic D, Surprenant S (2022) How is machine learning useful for macroeconomic forecasting? Journal of Applied Econometrics 37(5):920–964

Hamilton JD (1983) Oil and the macroeconomy since World War II. Journal of Political Economy 91(2):228–248. https://doi.org/10.1086/261140

Hamilton JD (2018) Why you should never use the Hodrick–Prescott filter. Review of Economics and Statistics 100(5):831–843. https://doi.org/10.1162/rest_a_00706

Harvey D, Leybourne S, Newbold P (1997) Testing the equality of prediction mean squared errors. International Journal of Forecasting 13(2):281–291. https://doi.org/10.1016/S0169-2070(96)00719-4

Hodrick RJ, Prescott EC (1997) Postwar U.S. business cycles: an empirical investigation. Journal of Money, Credit and Banking 29(1):1–16

Iacoviello M, Schiantarelli F, Schuh S (2011) Input and output inventories in general equilibrium. International Economic Review 52(4):1179–1213

Ivanov D (2020) Predicting the impacts of epidemic outbreaks on global supply chains: a simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoV-2) case. Transportation Research Part E: Logistics and Transportation Review 136:101922

Jakšić P, Pjanić M, Milošević M, Marjanović D, Dudić B, Stošić L (2025) Impact of macroeconomic factors on stock indices volatility: a GARCH analysis for USA, Germany, Serbia and Croatia. Computational Economics. https://doi.org/10.1007/s10614-025-11170-1

Kahn JA (1987) Inventories and the volatility of production. American Economic Review 77(4):667–679

Kahn JA, McConnell MM, Perez-Quiros G (2002) On the causes of the increased stability of the U.S. economy. Federal Reserve Bank of New York Economic Policy Review 8(1):183–202

Kaminsky G, Lizondo S, Reinhart CM (1998) Leading indicators of currency crises. IMF Staff Papers 45(1):1–48

Kauppi H, Saikkonen P (2008) Predicting U.S. recessions with dynamic binary response models. Review of Economics and Statistics 90(4):777–791

Kesavan S, Gaur V, Raman A (2010) Do inventory and gross margin data improve sales forecasts for U.S. public retailers? Management Science 56(9):1519–1533

Khan A, Thomas JK (2007) Inventories and the business cycle: an equilibrium analysis of (S, s) policies. American Economic Review 97(4):1165–1188. https://doi.org/10.1257/aer.97.4.1165

Kilian L (2009) Not all oil price shocks are alike: disentangling demand and supply shocks in the crude oil market. American Economic Review 99(3):1053–1069. https://doi.org/10.1257/aer.99.3.1053

Künsch HR (1989) The jackknife and the bootstrap for general stationary observations. Annals of Statistics 17(3):1217–1241. https://doi.org/10.1214/aos/1176347265

Lafrogne-Joussier R, Martin J, Mejean I (2023) Supply shocks in supply chains: evidence from the early lockdown in China. IMF Economic Review 71(1):170–215. https://doi.org/10.1057/s41308-022-00166-8

Lee HL, Padmanabhan V, Whang S (1997) Information distortion in a supply chain: the bullwhip effect. Management Science 43(4):546–558. https://doi.org/10.1287/mnsc.43.4.546

Liu W, Moench E (2016) What predicts US recessions? International Journal of Forecasting 32(4):1138–1150

Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems 30, pp 4765–4774

Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, Lee SI (2020) From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence 2(1):56–67. https://doi.org/10.1038/s42256-019-0138-9

Maccini LJ, Moore BJ, Schaller H (2004) The interest rate, learning, and inventory investment. American Economic Review 94(5):1303–1327

McConnell MM, Perez-Quiros G (2000) Output fluctuations in the United States: what has changed since the early 1980's? American Economic Review 90(5):1464–1476. https://doi.org/10.1257/aer.90.5.1464

McCracken MW, Ng S (2016) FRED-MD: a monthly database for macroeconomic research. Journal of Business & Economic Statistics 34(4):574–589. https://doi.org/10.1080/07350015.2015.1086655

Metzler LA (1941) The nature and stability of inventory cycles. Review of Economics and Statistics 23(3):113–129

Mofokeng TSV (2026) Predicting JSE top 40 equity market stress using the logistic regression approach with bootstrap inference for South Africa. Preprints.org (preprint, not peer reviewed)

Newey WK, West KD (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55(3):703–708. https://doi.org/10.2307/1913610

Patsiarikas M, Papageorgiou G, Tjortjis C (2025) Using machine learning on macroeconomic, technical, and sentiment indicators for stock market forecasting. Information 16(7):584. https://doi.org/10.3390/info16070584

Ramey VA (1991) Nonconvex costs and the behavior of inventories. Journal of Political Economy 99(2):306–334. https://doi.org/10.1086/261752

Ramey VA, West KD (1999) Inventories. In: Taylor JB, Woodford M (eds) Handbook of Macroeconomics, vol 1B. Elsevier, Amsterdam, pp 863–923

Rumyantsev S, Netessine S (2007) What can be learned from classical inventory models? A cross-industry exploratory investigation. Manufacturing & Service Operations Management 9(4):409–429

Stock JH, Watson MW (2002) Macroeconomic forecasting using diffusion indexes. Journal of Business & Economic Statistics 20(2):147–162

Syntetos AA, Babai Z, Boylan JE, Kolassa S, Nikolopoulos K (2016) Supply chain forecasting: theory, practice, their gap and the future. European Journal of Operational Research 252(1):1–26

Vrontos SD, Galakis J, Vrontos ID (2021) Modeling and predicting U.S. recessions using machine learning techniques. International Journal of Forecasting 37(2):647–671

Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67(2):301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x