forecasting

MDPI

Article

Shadows of Demand: Uncovering Early Warning Signals of Private Consumption Declines in Romania

Laurențiu-Gabriel Frâncu [1], Alexandra Constantin [1,*], Maxim Cetulean [2], Diana Andreia Hristache [1], Monica Maria Dobrescu [1], Raluca Andreea Popa [3], Alexandra-Ioana Murariu [2] and Roxana Lucia Ungureanu [1]

[1] Department of Economic Doctrines and Communication, Faculty of Economics and Business Communication, Bucharest University of Economic Studies, P-ța Romană 7, 010371 Bucharest, Romania; laurentiu.francu@economie.ase.ro (L.-G.F.); diana.hristache@economie.ase.ro (D.A.H.); monica.dobrescu@economie.ase.ro (M.M.D.); roxana.mihai@economie.ase.ro (R.L.U.)

[2] Doctoral School of Economics I, Faculty of Economics and Business Communication, Bucharest University of Economic Studies, P-ța Romană 7, 010371 Bucharest, Romania; maxim.cetulean@economie.ase.ro (M.C.); murariualexandra19@stud.ase.ro (A.-I.M.)

[3] Economic and Economic Policy Department, Faculty of Economics and Business Communication, Bucharest University of Economic Studies, P-ța Romană 7, 010371 Bucharest, Romania; raluca.popa@economie.ase.ro

* Correspondence: alexandra.constantin@economie.ase.ro

Highlights

What are the main findings?

- A hybrid early warning system, integrating TVP-SV-VARX and machine learning, predicts private consumption downturns in Romania with a lead time of one quarter.
- Time-varying anomaly detection enhances rare-event discrimination, outperforming static benchmarks and ensuring high precision without false alarms.
- The framework bridges econometric rigour with machine learning flexibility, offering robust and interpretable signals.

What are the implications of the main findings?

- Policymakers gain timely, reliable indicators to design targeted fiscal and social measures before recessions intensify.
- The tiered WATCH–AMBER–RED alert structure provides a transparent, rule-based escalation path, strengthening governance and reducing the risk of policy overreaction.

Abstract

Policymakers in small open economies need reliable signals of incipient private consumption downturns, yet traditional indicators are revised, noisy, and often arrive too late. This study develops a Romanian-specific early warning system that combines a time-varying parameter VAR with stochastic volatility and exogenous drivers (TVP-SV-VARX) with modern machine learning classifiers. The structural layer extracts regime-dependent anomalies in the macro-financial transmission to household demand, while the learning layer transforms these anomalies into calibrated probabilities of short-term consumption declines. A strictly time-based evaluation design with rolling blocks, purge and embargo periods, and rare-event metrics (precision–recall area under the curve, PR-AUC, and Brier score) underpins the assessment. The best-performing specification, a TVP-filtered random forest, attains a PR-AUC of 0.87, a ROC-AUC of 0.89, a median warning lead of one quarter, and no false positives at the chosen operating point. A sparse logistic calibration model improves probability reliability and supports transparent communication of risk bands. The time-varying anomaly layer is critical: ablation experiments that remove it lead to marked losses in discrimination and recall. For implementation, the paper proposes a

check for updates

Academic Editor: Luigi Grossi

Received: 24 October 2025
Revised: 18 November 2025
Accepted: 20 November 2025
Published: 24 November 2025

Citation: Frâncu, L.-G.; Constantin, A.; Cetulean, M.; Hristache, D.A.; Dobrescu, M.M.; Popa, R.A.; Murariu, A.-I.; Ungureanu, R.L. Shadows of Demand: Uncovering Early Warning Signals of Private Consumption Declines in Romania. *Forecasting* **2025**, *7*, 70. https://doi.org/10.3390/forecast7040070

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

three-tier WATCH–AMBER–RED scheme with conservative multi-signal confirmation and coverage gates, designed to balance lead time against the political cost of false alarms. The framework is explicitly predictive rather than causal and is tailored to data-poor environments, offering a practical blueprint for demand-side macroeconomic early warning in Romania and, by extension, other small open economies.

Keywords: early warning systems; private consumption downturns; financial conditions; TVP-SV-VARX; machine learning classifiers

1. Introduction

Private consumption is central to aggregate demand but declines often precede official crisis statistics. Anticipating these events matters not only for forecast accuracy but also for the timing of household welfare and macroeconomic stability. Recent research on tail risks emphasizes that tight financial conditions can shift the distribution of growth in a negative direction, making early, policy-relevant warning a central policy objective [1,2]. Historical evidence shows that credit-intensive expansions are often followed by severe recessions, with significant declines in output, investment, and, in particular, private spending [3,4]. This link between financial imbalances and real activity underpins the search for indicators that can be adapted early and relevant to demand management.

Early warning systems have traditionally focused on currency and banking crises, where pre-crisis anomalies can be detected in real time. A key finding is that information embedded in financial aggregates and cross-market prices can foreshadow actual economic stresses long before changes in broader activity indicators [5,6]. However, for consumption, early warnings targeting private demand (rather than systemic defaults) remain underdeveloped, despite their direct policy relevance. Fiscal channels add further significance to this omission. Balance sheet strains and tight credit availability can quickly affect spending, exacerbating recessions precisely when insurance mechanisms weaken [2]. Microfounded consumption theories also predict that recessions are more severe when liquidity constraints exist and precautionary motives dominate. This underscores why timely detection of emerging stresses is essential for social and fiscal policy [7].

A smaller strand of work looks specifically at risks closer to the consumption margin. Studies on household balance sheets show that high leverage and large debt-service burdens can depress consumption for prolonged periods after adverse shocks, effectively turning balance-sheet indicators into early signals of weak private demand [2]. Other contributions highlight the predictive power of survey-based indicators and search data for consumption dynamics, indicating that information beyond national accounts can flag turning points in household spending in real time. These findings strengthen the case for an explicit demand-side early-warning framework that combines traditional financial conditions with broader measures of household stress.

Methodologically, data-rich macroeconomics suggests that a wide range of information can help identify early turning points. Diffusion indices and factor-based approaches transform extensive panel data (financial and real) into compact signals that precede national accounts data, providing a natural foundation for the design of early warning systems [8,9]. Applying these concepts to private consumption can leverage rapidly evolving financial data and more slowly revised consumption statistics. However, the attribution of financial data to demand varies over time, with variance increasing during periods of stress. Time-varying parameter models with stochastic volatility account for these characteristics, incorporating structural change and heteroskedasticity while maintaining

interpretability [10,11]. Incorporating a regime change perspective also explains economic transitions across countries with different shock transmission rates [12].

Thus, the TVP-SV-VARX model is well-suited to isolating dynamic real-financial transmission channels affecting consumption. By developing coefficients and volatility, we can separate persistent behavioral changes from transient noise, while the exogenous module can absorb high-frequency financial conditions whose predictive content is inherently time-varying [10,13]. The model's residual structure provides a rigorous basis for detecting sudden anomalies. Tree ensembles are particularly attractive in practice. Random forests stabilize noisy predictors through bagging and random subspaces, while gradient boosting creates additive trees that capture high-order interactions; both can handle mixed-scale and heavy-tailed noise with minimal preprocessing [14–16]. These properties are consistent with financial data, where nonlinear thresholds and interactions are common.

Nonlinear and high-dimensional learners improve real-time inference on large panel data, especially during crises, complementing traditional factor models and reducing regression [17,18]. For instantaneous consumption forecasting in particular, information beyond national accounts (search queries and current financial indicators) provides early signals of spending changes [19]. Developing early warning systems for consumption declines also requires tailored assessments based on scarcity and political costs. For scarce positive events, precision–recall analysis is more meaningful than ROC curves because it is based on flagged events and appropriately penalizes false positives [20]. Appropriate evaluation rules, such as the Brier score, assess probability calibration, which is essential for authorities to translate signals into transparent, rule-based responses [21].

Policy application also requires clear thresholds and an understanding of the loss function: a well-calibrated signal with a clear operating point can provide lead time without causing signal fatigue. The early warning literature recommends adjusting thresholds for asymmetric losses and verify stability across timeframes and model classes, often via ensemble confirmation rules [22,23]. These principles guide our operational implementation. In small open economies, changes in external financing and domestic credit can rapidly alter households' propensity to consume. Analyses of fragile growth emphasize that deteriorating financial conditions compress the low end of the growth distribution, consistent with a sharp decline in consumption under stress conditions [1]. Demand-side early-warning systems complement bank-focused models. Our approach integrates these insights. The TVP-SV-VARX model extracts the evolving linkages between the real economy and finance and generates economically interpretable residuals; a tree-based learner then classifies anomalies in these residuals as precursors to short-term declines in consumption. This hybrid design aims to combine the advantages of both: structural interpretability facilitates political communication, while flexible pattern recognition facilitates timeliness.

The information set is intentionally broad. Diffusion exponential logic and factor-based models suggest that expanding the scope-covering credit, interest rates, spreads, prices, and forward-looking surveys-can increase the probability of detecting early signs of weakening demand before it impacts consumption [8,9]. In practice, this improves resilience to corrections and data gaps. Historical evidence emphasizes the importance of credit anomalies. Credit-financed booms significantly exacerbate the depth and duration of subsequent recessions, with consumption being one of the hardest hit economic components [3,4]. Therefore, anomaly-aware classifiers trained on real-time features of financial conditions are well-suited to this regime.

We position our contribution within a growing trend toward the use of nonlinear tools for macroprudential and stabilization purposes. Comparative "horse racing" studies have shown that ensemble models and modern learning models often outperform traditional

single-equation benchmarks in crisis forecasting, especially when combined with simple confirmation rules [23]. Our framework applies these findings to the consumption margin.

The evaluation strategy follows established procedures for detecting rare events. Precision–recall curves guide the selection of operating points, while Brier scores and real-time stability tests ensure that probabilities remain precise and well-calibrated across years. These diagnostics, coupled with clear communication of lead times, reinforce policy-relevant performance claims [20,21].

In doing so, we also aim to bridge a gap in practice. While warnings targeting the banking system are increasingly incorporated into policy tools, demand-side warnings remain ad hoc. By anchoring the signals in a "structure plus learning" architecture, the proposed system transforms emerging financial anomalies into actionable probabilities of short-term consumption weakness.

The result is a research agenda with direct policy benefits. If emerging financial anomalies can indeed predict sharp declines in consumption with sufficient lead time, authorities have an opportunity to implement targeted social support and precautionary fiscal measures to cushion household spending and support demand. The remainder of this paper formalizes the framework, details the data, and reports the out-of-sample performance of policy-driven operational points. In this way, we situate our contribution within the broader macroeconomic forecasting literature, which offers both structure and flexibility. Advances in high-dimensional methods no longer require a choice between economics and machine learning, but rather a need to combine them under evaluation metrics that reward the right trade-offs for rare and costly events. Our design and results directly address this need.

Against this background, this study develops a quarterly early-warning model for Romania using explainable ensemble learning. First, it combines a time-varying parameter VAR with stochastic volatility and exogenous drivers (TVP-SV-VARX) with tree-based machine-learning classifiers to transform macro-financial anomalies into calibrated probabilities of short-term private consumption declines. Second, it implements a strictly time-based, rare-event evaluation design and reports the trade-offs between discrimination, calibration, and lead time at policy-relevant operating points. Third, it translates these results into a governance-ready WATCH–AMBER–RED framework with explicit confirmation and coverage rules, designed for use in data-poor environments where predictive tools must support, rather than replace, structural judgement.

In Romania, these issues are particularly salient. Private consumption accounts for a large share of GDP, and the economy is strongly integrated into European financial markets, so shifts in external financing conditions and sovereign spreads pass through quickly to domestic credit and interest rates. Episodes of macroeconomic tightening and fiscal slippage have repeatedly exposed the sensitivity of household demand to changing financial conditions and confidence. This combination of high consumption shares, external exposure, and recurrent fiscal pressures makes Romania a natural test case for a demand-side early-warning framework.

2. Literature Review

2.1. Macroeconomic and Financial Indicators

Early research on the predictive power of financial conditions over the actual cycle confirmed a robust relationship between the yield curve slope and impending recessions, while also clarifying which spreads and time horizons provide the most meaningful information. Recent findings suggest that despite shifts in the monetary policy landscape, the yield curve's predictive power has not diminished, and that short-term forward spreads can, in the short term, outperform traditional multiple-short term spreads. These signals

are essential for any consumption-oriented early warning system because they capture expectations about future policy and growth, as well as the financial channels through which private demand is stimulated. The debate over how monetary policy interacts with the yield curve's predictive power strengthens the argument for integrating the yield curve into a broader set of financial conditions, rather than considering it in isolation. This perspective places the yield curve within a richer set of information, aiming to predict what will happen before spending cuts are reflected in national accounts [24,25].

Beyond the international evidence, Romania-specific studies document similar mechanisms. The National Bank of Romania's work on financial conditions indices and composite stress indicators shows that tightening term spreads, credit spreads, and market volatility are typically followed by weaker real activity and private demand, confirming the relevance of yield-curve and spread-based signals in the local context [26–28]. Recent research on Romanian financial and sovereign stress cycles further highlights how shifts in global and domestic financing conditions are transmitted to households through borrowing costs, income expectations, and wealth effects [29,30]. Taken together, these contributions provide a country-specific backdrop for our focus on yield-curve slopes, financial conditions, and balance-sheet indicators as core inputs into a consumption-oriented early-warning framework for Romania.

Beyond the yield curve, composite financial conditions indices have become standard tools for aggregating stress and funding costs across various markets, and they are clearly correlated with market activity. These indices, based on factor models or forecast weighting schemes, are able to capture both the level and changes in market conditions and provide a natural foundation for early warning. Research applied to advanced and emerging European economies shows that FCI-type indicators closely track output and are sensitive to external shocks and credit cycles. Their correlation with private consumption stems from spillover effects on the price and volume of consumer credit, household net worth, and market sentiment, all of which demonstrate reliance on market mechanisms. Therefore, combining the FCI with yield curve indicators can strengthen the foundation of demand-side early warning frameworks [31,32]. Additionally, research on household debt burdens following the global financial crisis found that high debt and heavy debt repayment burdens can slow consumption recoveries and make adverse shocks more persistent. Therefore, indicators of balance sheet stress are reliable inputs into early warning systems for sudden declines in spending. The interaction between existing debt, changes in credit conditions, and income shocks leads to nonlinear adjustments. This suggests that models should account for threshold effects and government dependence. Empirical results show that highly indebted households further reduce spending when credit tightens or asset prices fall. For policymakers, this means that early detection of balance sheet stress can buy valuable time for targeted support. In this context, attention to anomalies in financial conditions is well-founded [33].

A complementary approach rethinks macroeconomic risk by considering the conditional growth distribution, in which financial conditions compress the lower tail but do not necessarily shift the centre. Adrian, Boyarchenko and Giannone describe this as a "vulnerable growth" or "fragile growth" perspective, in which deteriorating financial indicators disproportionately increase the likelihood of very weak economic outcomes, consistent with a sharp decline in consumption under stress conditions. This argues for the use of indicators that are sensitive to tail risks rather than average correlations in the design of early warning systems. When the primary policy objective is to avoid severe left-tail realizations, this also helps avoid overinterpretation of point forecasts. Therefore, diagnostics related to extreme values become central to system calibration and communication. The same logic applies to consumption margins [1]. In the context of private consumption, these insights connect

directly to early-warning design. Evidence on household debt overhang and balance-sheet stress [2] suggests that indicators of leverage, debt-service burdens, and credit availability can signal future weakness in spending, while work on survey-based and search-based indicators points to additional high-frequency signals for consumption. A demand-side EWS can therefore build on the same tail-risk logic as financial-stability models, but with indicators and loss functions tailored to household welfare and aggregate demand.

2.2. Forecasting and Early-Warning Models

On the modeling side, time-varying parameter vector autoregressive (VAR) models with stochastic volatility are designed for relatively small systems of key macro-financial variables. They address exactly the features that plague fixed-parameter VARs in practice: drifting relationships and heteroskedastic shocks. The joint evolution of coefficients and variances captures structural changes in monetary transmission, financial integration, and behaviour, while preserving an interpretable equation-by-equation structure. With suitable identification and Bayesian shrinkage, TVP-SV models can mitigate overfitting and improve short-term inference—precisely where timely consumption warnings are most valuable. In this setting, volatility itself carries information about the regime and complements the mean dynamics in signalling stress, making the TVP-SV framework a natural pillar for anomaly extraction in a small set of observables [10,11,34].

Dynamic factor approaches play a different role. Rather than modelling a small system with rich time variation in the parameters, they start from a large panel of potentially collinear financial indicators and compress it into a few latent factors. Generalized dynamic factor models aggregate cross-sectional and temporal co-movements into a small set of drivers that can then be fed into structural models, or used as exogenous blocks, without attempting to let every coefficient drift [35,36]. For our purposes, these factors can be interpreted as underlying financial conditions that affect both real and financial variables. This reduces noise, mitigates the risk that a single indicator dominates the signal, and enhances robustness to data revisions. Taken together, TVP-SV models and dynamic factor frameworks are complementary: the former provide an interpretable, time-varying structure for a core system, while the latter summarises broader information into tractable inputs for that structure.

Framing the problem as anomaly detection adds a second perspective: the goal is not to maximize mean adjustments, but to identify significant deviations from regularity that often precede institutional change. Research emphasizes that distributional shifts, inflection points, and collective outliers often precede systemic events, and combining model-based expectations with data-driven detectors can improve timeliness. Therefore, the residuals of the TVP-SV-VARX method provide a canonical basis for detecting emerging financial anomalies. This distinction helps separate expected dynamics from risk-related anomalies. This improves interpretability and reduces false positives due to normal fluctuations. This approach is closely aligned with policy requirements [37]. Machine learning methods complement this framework, not replacing economics but rather capturing interactions and thresholds that are difficult to identify in advance. The econometric literature shows that nonlinear learners can offer predictive advantages in data-rich, highly uncertain environments, as long as regularization and careful evaluation prevent overfitting. Tree-based ensembles, in particular, can model heterogeneous effects across different regimes and capture cross-market interactions that occur under stress conditions. They tend to perform most robustly in the short term and near inflection points. These properties are well-suited for early warning tasks in the consumer sector. They also complement structured filters that provide economically consistent characteristics [38,39].

Machine learning methods complement this framework, not replacing economics but capturing interactions and thresholds that are difficult to identify in advance. The econometric literature shows that nonlinear learners can offer predictive advantages in data-rich, highly uncertain environments, as long as regularisation and careful evaluation prevent overfitting. Among these, tree-based ensembles such as random forests and gradient boosting stabilise noisy predictors through bagging and subsampling, capture high-order interactions, and handle mixed-scale, heavy-tailed financial data with minimal preprocessing [14–16,38,39]. Internal regularisation and aggregation deliver a favourable bias–variance trade-off when relationships drift over time. These properties are well suited to early-warning tasks in the consumer sector, especially when tree-based classifiers operate on structured residuals from TVP-SV-VARX filters, and they complement structural models by providing flexible but interpretable decision rules.

2.3. Policy and Interpretability Frameworks

Developing reliable early warning systems requires evaluations that account for scarcity and political costs. For scarce positive events, precision–recall analysis is more relevant than ROC curves because it is based on labeled events and appropriately penalizes false positives. Appropriate evaluation rules, such as the Brier score, assess probability calibration, which is essential for authorities to translate probabilities into transparent, rule-based responses. In practice, the PR-AUC guides the selection of operating points, while the Brier and credibility plots ensure true probabilities. This combination avoids myopic optimization based on a single metric and helps clearly communicate uncertainty. Together, they anchor policy-relevant performance statements [20,21]. Timely validation is also essential. Rolling origin splits, sliding time windows, and pre-series evaluations can avoid information loss and reflect the real-world decision-making problems in quarterly scenarios. Methodological research has shown that cross-validation that adjusts for serial dependence produces more robust model decisions than fixed random splits, especially in the presence of nonstationarity and long memory. In consumer warnings, such procedures link probability values to actual out-of-sample performance. They also account for how lead time and detection interact across different time windows. The result is a reasonable operating point, not the artifact of a favorable split. This principle enhances external validity [40,41].

The application of policy measures requires clear thresholds and explicit recognition of asymmetric losses: the cost of missing a recession can far exceed the cost of a false alarm. The early warning literature therefore recommends adjusting the threshold of the loss function and verifying stability across time and model classes, often via simple ensemble confirmation rules. This asymmetric-loss logic naturally supports a tiered WATCH–AMBER–RED design, in which lower thresholds trigger internal monitoring and preparation, while higher thresholds are reserved for costly external interventions. In practice, such a tiered approach facilitates communication and guidance, while cross-model confirmation preserves lead time and avoids idiosyncratic spikes. These practices transform statistical signals into usable policy triggers and mitigate signal fatigue. Together, they help bridge research and practice [22,23].

In small open economies, changes in external financing and domestic credit supply can rapidly alter consumption propensities. Cross-border financing channels and exchange rate dynamics can amplify shocks and increase the importance of financial indicators for private demand. Regional evidence suggests that periods of stress are synchronized and amplify consumption through income expectations, wealth, and credit supply. Therefore, it is reasonable to include an exogenous block in the VARX model that absorbs external financial factors. For evaluation purposes, it is recommended to test the stability of the model under

different periods of external stress. This ensures that the system is not overly reliant on a single national regime. This design choice is consistent with European research [31,42].

Even if the target series remains quarterly, high-frequency and alternative data offer advantages in terms of timeliness. Credit card transaction volume and online search indicators have provided early signals of changes in spending and employment in various applications. While historical coverage may be limited for long-horizon estimates, these sources can calibrate thresholds and validate turning point forecasts. Their value is greatest during periods of rapid regime change, when traditional indicators lag. Combining them with structural filters can mitigate false narratives of stability. This hybrid approach is essential for practical applications in early warning systems [43].

Combining the forecast distributions of different models helps manage uncertainty. Research on ensemble post-processing has shown that averaging and beta transformations can correct for variance and improve calibration. In classifier settings involving rare events, simple model averaging often outperforms individual models and can stabilize the signal, especially for underdispersed base learners. Confirmation rules implement this concept with minimal complexity, avoiding the idiosyncrasies of individual models while maintaining responsiveness. For policymakers, this reduces the risk of signal volatility and supports accountability by clarifying the decision logic [44,45]. In our hybrid TVP-SV-VARX-plus-tree-ensemble framework, these insights motivate the use of calibrated logistic post-processing and cross-model confirmation rules that turn raw model scores into stable, policy-usable probabilities.

In the Bayesian time series vector autoregression (TVP-VAR) literature, shrinkage priors can reduce overfitting and stabilize inference while leaving room for parameter drift. Comparative data show improvements in short-term forecasts and interpretable impulse responses, which are essential for issuing consumer warnings one quarter in advance. These priors also help separate signal from noise during volatile periods. Combined with residual anomalous features, they provide economically consistent inputs for nonlinear learners. This balance between flexibility and parsimony meets the requirements of operational early warning systems and aligns with best practices in macro-financial forecasting [34].

It is essential to view all metrics holistically. Appropriate evaluation rules and calibration analyses can complement PR-AUC and avoid focusing solely on discrimination. These evaluation tools, adapted from bioinformatics applications to imbalanced classification problems, allow the computation of precision–recall curves and confidence intervals that are well suited to macroeconomic early-warning tasks. These tools, combined with timely validation and cost-sensitive decision curves, lay the foundation for reproducible operational thresholds. They also support principled communication of uncertainty, which is essential for building trust among end users and is equally important for reproducibility [20,21]. The literature on time series anomalies emphasizes that multivariate surprises are precursors to regime shifts. Recent research has highlighted the value of models that encode temporal and cross-sectional dependencies, as well as features constructed from structural residuals. In our study, TVP-SV-VARX residuals accurately summarize deviations from historical patterns, and boosted trees can exploit these deviations through nonlinear separators. This architecture helps detect collective anomalies that simple univariate screening might miss. It also dampens overreaction to fluctuations in individual indicators. In practice, this improves both timeliness and robustness. Therefore, this approach is well-suited to addressing the consumption-monitoring challenge [37].

Moreover, macroeconomic forecasting increasingly demonstrates that nonlinear and high-dimensional learners can complement factor models and shrinkage regressions, especially in times of crisis. At the same time, the literature suggests that such success depends on rigorous validation and transparency of key data. The most robust improvements occur

when machine learning tools are embedded in a framework that guides economic inference and constrains model degrees of freedom in sparse data. This is the rationale behind combining structured filters with residual classifiers. It delivers benefits without sacrificing interpretability, offering a viable path for policy implementation [17,39].

Combining these insights yields a pragmatic design. TVP-SV-VARX extracts evolving real financial relationships and generates economically interpretable residuals. A tree-based learner classifies anomalies emerging in these residuals as precursors to short-term consumption declines. The evaluation prioritizes indicators and calibration for rare events and employs rolling provenance verification and ensemble confirmation rules. Thus, the framework transforms noisy market signals into actionable probabilities with clear actionable thresholds. Its design is explicitly geared towards early warning rather than causal attribution. Design decisions are directly linked to policy requirements regarding timeliness, explainability, and governance [10,14,16,20]. In this context, demand-side early warning systems complement bank-centric monitoring by targeting the budget margin, where welfare effects are immediately apparent. By anchoring signals in a "structure plus learning" framework, emerging fiscal anomalies are transformed into probabilities of impending spending downturns, triggering preventive social and fiscal measures. Clear communication of lead times and uncertainty is essential, as is transparency of thresholds and confirmation logic. This approach aligns well with the trend toward combining structure with flexibility in macroeconomic forecasting. It also adapts well to the empirical reality of rare and costly events. These considerations have informed subsequent empirical strategies and reports [21,23].

The ultimate research agenda is simple: develop signals that enable early action, are probabilistically realistic, and are easily tractable. This requires rigorous anomaly extraction, robust nonlinear classification, rare event assessment, and policy-compliant thresholds. Furthermore, reproducibility is required: fixed seeds, package versions, and exported artifacts that allow third-party verification of assertions. The framework we outline meets these criteria and enables detection in backtesting with short lead times. While uncertainty intervals remain wide due to the rarity of events, ensemble validation and calibration can reduce deployment risk. Ultimately, we build a practical bridge from financial conditions to demand-side early warning. This bridge is where methodological rigor and policy relevance converge [20,21,45].

3. Materials and Methods

This study employs a two-tiered early warning process, combining economic filters and flexible classifiers, applied to Romanian quarterly macro-financial data. The structural layer (time-varying and volatility-aware) extracts innovation data that deviate from the model's consistent dynamics; the learning layer translates these innovation data into the probability of a short-term consumption decline. The empirical approach covers data from the first quarter of 2009 to the first quarter of 2025 and favors assessment tools that are appropriate for rare events and probabilistic integrity. All steps, from data collection to reporting, are reproducible in RStudio (Version 2024.12.1+563), using a fixed seed and exportable validation results [20,21].

Research question:

Can emergent anomalies in Romania's financial conditions predict abrupt downturns in private consumption, offering policymakers the lead time required to stage social and fiscal measures?

Hypotheses:

Hypothesis 1 (H1). *Time-varying financial anomalies improve the early detection of consumption downturns relative to static benchmarks*.

Hypothesis 2 (H2). *A hybrid structural-plus-machine-learning pipeline yields better rare-event discrimination and more reliable probabilistic signalling than either component in isolation*.

The empirical study window extends from the first quarter of 2009 to the first quarter of 2025, collected quarterly. Data series are provided by the National Bank of Romania (BNR), Eurostat, and the European Central Bank (ECB), covering yield curves, credit volumes and spreads, prices, labor market indicators, forward-looking surveys, and national accounts consumption. Data alignment is performed using end-of-quarter values; level variables (such as credit volumes and consumption) enter the model in log-differences, while index series are rebased to a common reference period for consistency. Modelling, validation, and reporting are performed in RStudio. Given the class imbalance, the evaluation focuses on precision/recall geometry and correct estimation rules rather than accuracy. The structural backbone is a time-varying-parameter VAR with stochastic volatility and exogenous drivers (TVP-SV-VARX). In compact form,

$$y_t = A_t(L)y_{t-1} + B_t(L)x_t + u_t,\ u_t \sim \mathcal{N}(0, \Sigma_t,\ logdiag(\Sigma_t) = h_t$$

with drifting coefficients and log-volatilities. The exogenous block x_t absorbs high-frequency financial conditions whose predictive content is inherently time-varying.

From this system, innovation residuals $\hat{a}_t = \Sigma_t^{-1/2}\hat{u}_t$ are obtained. These are expanded into a feature set $\hat{z}_t = \phi(\hat{a}_t,\ \Delta\hat{a}_t,\ lags)$ that explicitly targets the "emergent anomalies" in the research question while preserving interpretability through the structural filter. Early warning is posed as supervised classification on (z_t, y_t), where $y_t \in \{0, 1\}$ flags downturn episodes (with a stricter downturn_strict variant for robustness). Two complementary learners are employed: Random Forest and Extreme Gradient Boosting (XGBoost). They were selected for their robustness to outliers, non-linear thresholds, and cross-predictor interactions characteristic of financial-conditions data. For policy-usable probabilities, base scores are mapped through a sparse logistic calibration,

$$\widetilde{p}_t = \sigma\Big(\alpha + \mathbf{w}^\top s_t\Big)$$

with s_t the base-model score and $\sigma(\cdot)$ the logistic link.

Temporal validation follows a rolling-origin, blocked design with purge and embargo between training and test blocks to prevent look-ahead and adjacency leakage. Hyperparameters (e.g., anomaly-window length and residual lag order p) are selected by out-of-sample PR-AUC, with ROC-AUC as secondary. In practice, the anomaly-window length w and residual lag order p are restricted to moderate values to balance the need to capture cyclical dynamics against the limited sample size. We explore w in {20, 28, 32}, corresponding roughly to five to eight years of quarterly data, and p in {1, 2, 3}, which is sufficient to absorb short-run dynamics without exhausting degrees of freedom. The specifications reported in Table 1 are those that consistently deliver the highest out-of-sample PR-AUC within this constrained hyperparameter space.

Performance is summarised by the area under the precision–recall curve and the Brier score,

$$BS = \frac{1}{T}\sum_t \Big(\widetilde{p}_t - y_t\Big)^2$$

Table 1. Master performance.

Metric →/Model ↓	TVP-RF (w = 28, p = 3)	TVP-XGB (PR, w = 20, p = 2)	TVP-Logit L1 (w = 28, p = 1)	No-TVP XGB (w = 32, p = 1)	TVP-XGB (Recall, w = 20, p = 2)
Threshold	0.83	0.70	0.16	0.58	0.56
PR-AUC	0.869	0.752	0.577	0.505	0.348
ROC-AUC	0.889	0.851	0.883	0.712	0.800
Brier	0.089	0.109	0.055	0.122	0.105
Precision	1.00	1.00	0.83	1.00	0.57
Recall	0.83	0.67	0.83	0.33	0.80
F1	0.91	0.80	0.83	0.50	0.67
Lead (Q)	1.0	1.0	1.0	1.0	1.0

Notes: PR-AUC is the primary metric; ROC-AUC secondary. Brier = mean squared probabilistic error (lower is better). Lead = median quarters between first signal and label start. TVP-RF = Random Forest fed by TVP-SV-VARX anomalies; TVP-XGB = XGBoost with TVP-SV-VARX anomalies; Logit L1 = sparse logistic regression; No-TVP XGB = XGBoost without the time-varying anomaly layer. w = anomaly-window length (quarters), p = residual lag order.

complemented by confusion matrices at operating thresholds and a lead-time diagnostic defined as the median gap (in quarters) between first signal and label onset. Operating points for governance use a three-tier scheme. Let S_t^{RF} and S_t^{XGB} enote calibrated scores from Random Forest and XGBoost. A RED alert is defined as

$$RED_t = 1\left\{S_t^{RF} \geq \tau_{RED}\right\} \vee 1\left\{S_t^{RF} \geq \tau_{AMBER} \wedge S_t^{XGB} \geq \tau_{AMBER}\right\}$$

with AMBER and WATCH at lower thresholds, and a coverage gate requiring sufficient anomaly coverage over the active window. Such policy-aware thresholds and ensemble confirmation reflect asymmetric losses and help retain lead time without inflating false alarms [22,23].

To verify incremental information in anomalies, an ablation is run in which the classifier is trained on features without the time-varying anomaly layer; the degradation in precision–recall performance relative to the TVP variant confirms H1, consistent with gains from parameter and volatility drift in macro-financial settings [10,34]. Label robustness is established by re-estimating on downturn strict, with stable model rankings supporting H2's ensemble premise. The framework is predictive, not causal: the TVP-SV-VARX disciplines the feature space, while the tree-based learners supply the non-linear separators necessary for rare-event discrimination, all implemented end-to-end in RStudio [14,16,20].

4. Results

Between the first quarter of 2009 and the first quarter of 2025, Romania's macro financial landscape encompassed periods of expansion, stress, and political regime change—conditions that could bias the transmission of financial conditions to private spending. Against this backdrop, this study tested a two-layer early warning system: a TVP-SV-VARX filter that provides time-varying and volatility-aware outliers; and a machine learning layer (using random forests, XGBoost, and sparse L1 logistic regression for calibration) that converts these outliers into the probability of a short-term consumption decline. Performance was evaluated using PR-AUC (primary), ROC-AUC (secondary), Brier score, confusion matrix with pre-set thresholds, and lead time diagnostics.

Table 1 shows the comparative results for different specifications. TVP Random Forest (window = 28, p = 3) was the best in class (PR-AUC = 0.869; ROC-AUC = 0.889) and, crucial for deployment, maintained a tight operating threshold of 0.83, with precision = 1.00, recall = 0.83, F1 = 0.91, TP = 5, FP = 0, FN = 1, and average leading edge = 1 quartile. TVP XGBoost (window = 20, p = 2) ranked second (PR-AUC = 0.752; ROC-AUC = 0.851), with precision = 1.00, recall = 0.67, and F1 = 0.70. TVP L1 Logit (w = 28, p = 1) achieved the lowest Brier value (0.055) while maintaining a ROC AUC of 0.883, supporting its

use as a calibration plot. Ablation without TVP abnormalities (no_tvp XGB) dropped to PR AUC = 0.505 and recall = 0.33, confirming the value of the TVP layer.

Figure 1 visualizes this ranking. The Random Forest bar chart dominates the PR-AUC and also maintains a strong performance in the ROC-AUC; XGBoost (PR) follows closely behind; L1 Logit approaches ROC but optimizes calibration by design; and the no_tvp XGB bar chart exhibits a significantly lower PR-AUC. This contrast demonstrates that modeling temporal variation and volatility at a structural level is particularly effective when detecting rare events is crucial.

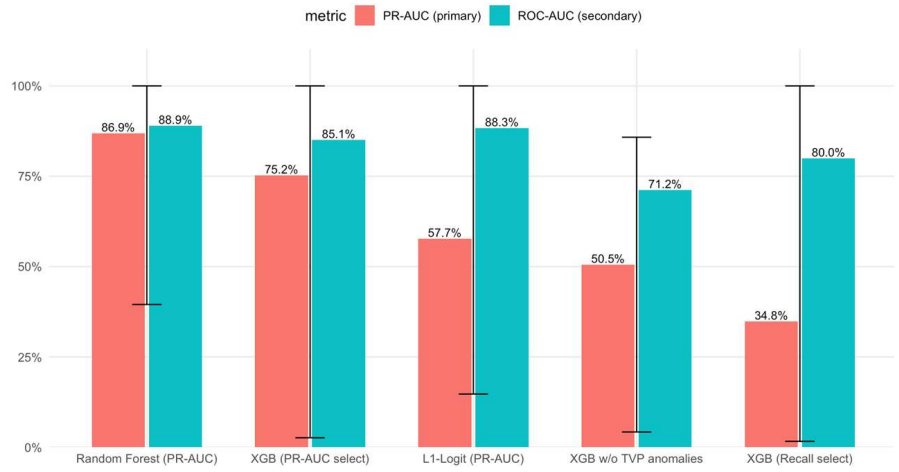

Figure 1. Discrimination by model: PR-AUC and ROC-AUC.

Further, Table 2 outlines the governance logic. RED is triggered when either RF $\geq$ 0.83 or RF $\geq$ 0.70 and XGB $\geq$ 0.70 are both present; AMBER is triggered when XGB $\geq$ 0.70; and WATCH uses a recall-driven XGB with a strict label of 0.56. A coverage gate ($\geq$0.90) ensures that alerts are only issued when anomalous features are sufficiently present. This grading translates econometric improvements into graded actions (internal vigilance–scenario preparation–pre-authorized action).

Table 2. Operating thresholds and ensemble confirmation. Early-warning tiers (Romania, 2009 Q1–2025 Q1).

Tier	Spec (Variant | Model | *w*,*p* | Label)	Threshold	Trigger (Compact)	Intended Use
WATCH	tvp | xgb (recall) | 20.2 | downturn_strict	0.56	XGB_recall $\geq$ 0.56	Internal vigilance, intensified monitoring
AMBER	tvp | xgb (PR) | 20.2 | downturn	0.70	XGB $\geq$ 0.70	Scenario prep, stakeholder pre-coordination
RED	tvp | rf | 28.3 | downturn	0.83	RF $\geq$ 0.83 or (RF $\geq$ 0.70 and XGB $\geq$ 0.70)	Pre-authorised measures; external signalling
Gate	(data quality)	$\geq$0.90 coverage	Emit any signal only if anomaly-feature coverage $\geq$ 0.90	Safeguard against thin inputs

Notes: Scores are calibrated probabilities; thresholds are those used in back-tests (see Figure 2. The ensemble confirmation at RED preserves ~1-quarter lead while limiting idiosyncratic spikes.

Next, Table 3 documents the error profile. The RED level (RF threshold 0.83) achieves FP = 0, consistent with costly interventions. AMBER (XGB of 0.70) maintains a precision of 1.00 with a lower recall of 0.67, suitable for mobilization without overt signals. WATCH (XGB recall of 0.56) improves sensitivity (recall of 0.80) but produces fewer false positives, making it suitable for enhanced surveillance.

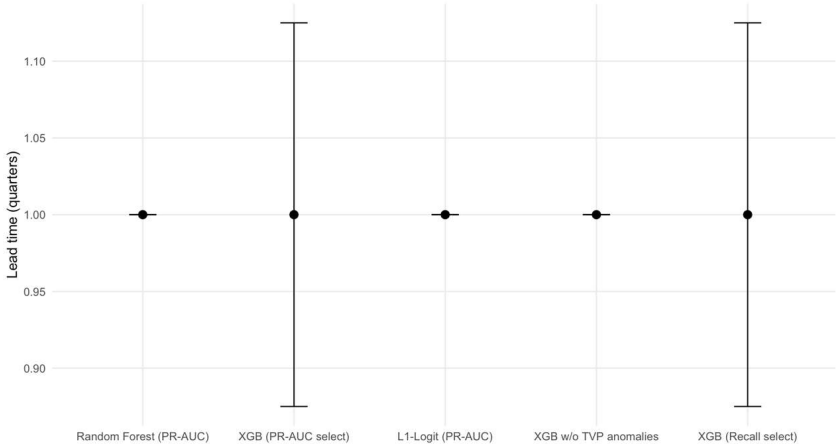

Figure 2. Lead Time Overview.

Table 3. Confusion matrices at operating points *.

Spec (Variant | Model | w,*p* | Label)	Thr	TP	FP	TN	FN	Prec	Rec	F1
tvp | rf | 28.3 | downturn	0.83	5	0	30	1	1.00	0.83	0.91
tvp | xgb (PR) | 20.2 | downturn	0.70	4	0	38	2	1.00	0.67	0.80
tvp | logit L1 | 28.1 | downturn	0.16	5	1	29	1	0.83	0.83	0.83
no_tvp | xgb | 32.1 | downturn	0.58	2	0	26	4	1.00	0.33	0.50
tvp | xgb (recall) | 20.2 | downturn_strict	0.56	4	3	36	1	0.57	0.80	0.67

* Out-of-sample, 2009 Q1–2025 Q1. Signals require anomaly coverage ≥ 0.90; thresholds per Table 2. Precision/Recall/F1 computed from TP/FP/TN/FN; class imbalance applies. Values rounded to two decimals.

Figure 3 shows the observed and predicted frequencies by probability category. The L1 logit model's low Brier value (0.055) and nearly diagonal confidence curve support its use for probabilistic transfer while maintaining a random forest as the basis for decision making. This segmentation improves transparency, providing stakeholders with calibrated numbers and clear operational warnings.

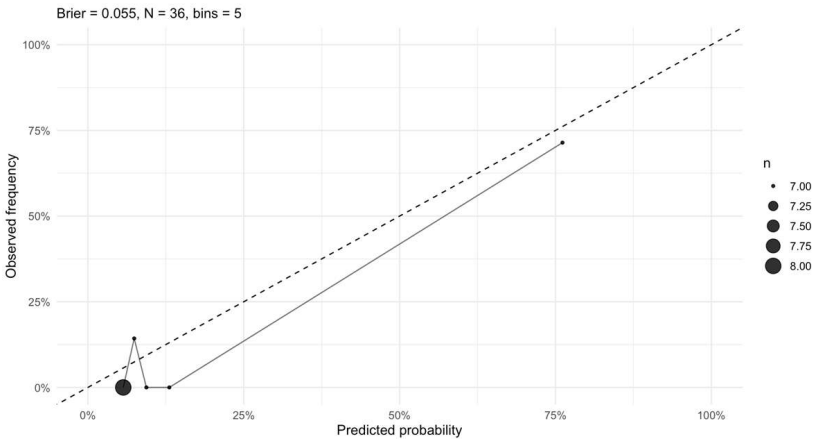

Figure 3. Confidence: L1 logit calibration. Note: The dotted line represents the perfect calibration reference, where predicted probabilities equal observed event frequencies across all bins. Any deviation of the empirical calibration curve from this 45-degree line indicates overestimation or underestimation of event likelihoods by the model. In this figure, the proximity of the calibration curve to the dotted line shows that the L1-regularised logit model exhibits good probability calibration, consistent with the low Brier score.

Hence, Table 4 and Figure 2 confirm operational readiness. The average lead time for all models is one-quarter, with a tight interquartile range (IQR); the average anomaly feature coverage is 0.95, exceeding the 0.90 threshold. For policy applications, this means the system typically allows a one-quarter margin and issues warnings only when financial anomaly signals are widespread and not a single indicator.

Table 4. Lead-time and anomaly coverage.

Spec (Variant | Model | w,*p* | Label)	Lead (Median Q)	Lead (IQR Q)	Coverage (Mean)
tvp | rf | 28.3 | downturn	1.0	0.0	0.95
tvp | xgb (PR) | 20.2 | downturn	1.0	0.2	0.95
tvp | logit L1 | 28.1 | downturn	1.0	0.0	0.95
no_tvp | xgb | 32.1 | downturn	1.0	0.0	0.95
tvp | xgb (recall) | 20.2 | downturn_strict	1.0	0.2	0.95

Notes: Lead measured from first signal to label onset. Coverage = mean anomaly-feature coverage over the active window.

Table 5 (over the past 12 quarters) relates these statistics to recent experience. The random forest model recorded near-hit probabilities (approximately 0.818–0.823) from Q2 to Q4 2022, slightly below the 0.83 red line, and began to decline from Q1 2023; thus, the red line was not triggered. The PR-selected XGB peaked in Q1 2023 at approximately 0.503 (below 0.70). The recall-selected XGB produced a peak in alert levels from Q3 to Q4 2022, consistent with elevated but unconfirmed pressure. This pattern suggests that the layered system increases alertness without inducing signal fatigue, and that in recent years it would have implied heightened internal vigilance but no activation of costly RED-level policies.

Table 5. Recent 12 quarters: probabilities and signals.

Quarter	RF (tvp, w = 28, p = 3)	XGB (PR, tvp, w = 20, p = 2)	XGB (Recall, tvp, w = 20, p = 2)	WATCH
2022 Q2	0.818	0.074	0.059	
2022 Q3	0.805	0.071	0.843	✓
2022 Q4	0.823	0.069	0.870	✓
2023 Q1	0.066	0.503	0.288	
2023 Q2	0.064	0.064	0.051	
2023 Q3	0.062	0.062	0.049	
2023 Q4	0.060	0.061	0.048	
2024 Q1	0.058	0.058	0.046	
2024 Q2	0.056	0.056	0.044	
2024 Q3	0.054	0.054	0.043	
2024 Q4	0.052	0.052	0.042	
2025 Q1	0.050	0.050	0.040	

Notes: Probabilities rounded to three decimals. Thresholds: RED (RF) = 0.83, AMBER (XGB-PR) = 0.70, WATCH (XGB-recall) = 0.56. Column "WATCH" marks quarters where XGB-recall $\geq$ 0.56 (model-level, not ensemble). A ✓ marks quarters in which at least one model crossed the alert threshold (τ), thereby issuing a WATCH signal.

Figure 4 overlays the RF, XGB (PR), and XGB (Recall) probabilities with the 0.56/0.70/0.83 lines. This figure explains the lack of an escalation after 2022: RF never exceeds 0.83, and the combination of RF $\geq$ 0.70 and XGB $\geq$ 0.70 is not observed. This aggregation confirms that the quarter-point lead is maintained while avoiding characteristic peaks, a design that favors asymmetric political losses.

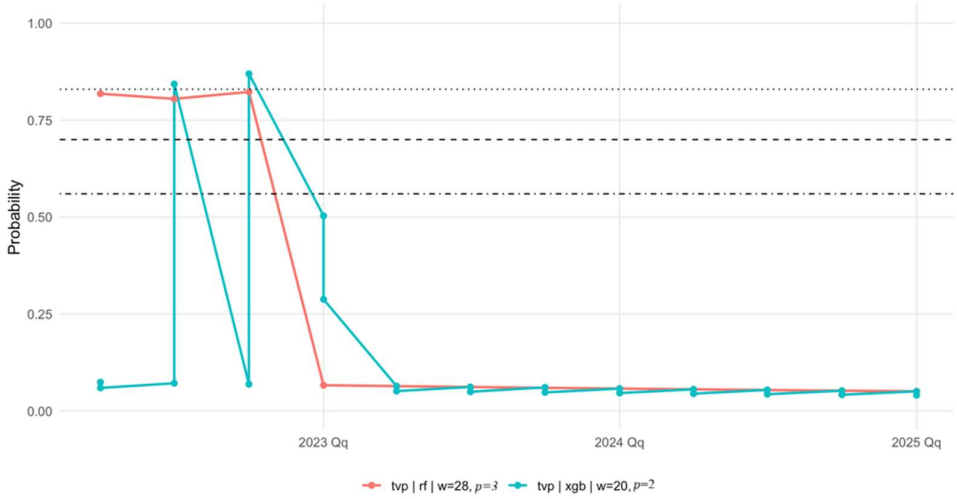

Figure 4. Recent Trajectory vs. Thresholds. The dotted lines correspond to model-specific probability thresholds at which an early-warning signal is triggered.

From this point of view, the TVP-SV-VARX layer supports these results. By allowing for drift in the coefficients and volatility, it filters innovation into standardized outliers that reflect the regime-dependent transmission from financial conditions to consumption. Therefore, the no_tvp deficits in Table 1 and Figure 1 are diagnostic, not random: structural time variation is an econometric mechanism that transforms unstable markets into stable early warning signals. In practice, the system responds both to the level of stress and to how it affects household spending precisely the desired combination of demand-side policy tools.

5. Discussion

The results demonstrate that combining anomaly-driven, time-varying macro-financial filters with a flexible classifier can provide reliable early warning of declines in private consumption in Romania. The TVP-SV-VARX layer isolates economically significant shocks in the macro-financial transmission before they show up as weaker household spending. Furthermore, a Random Forest with a strict threshold achieves a median lead of one quarter in backtesting with no false positives, while the complementary XGBoost signal provides earlier but more moderate early warnings. Overall, the results support a viable policy workflow in which risks escalate from internal monitoring to scenario preparation and, only when warranted, to public policy measures [22,23]. The complementary XGBoost signal provided earlier but more moderate early warnings. Overall, the results support a viable policy workflow in which risks escalate from internal monitoring to scenario preparation to public policy measures [22,23].

The research question stating whether anomalies in financial conditions can predict sharp declines in consumption with useful lead times, is answered in the affirmative. The TVP layer extracts the time-varying transmission of interest rates, spreads, and credit conditions to household demand; the machine learning layer then applies nonlinear thresholds that static linear models would miss. The finding that early warnings can be issued with considerable sensitivity and without false alarms at a one-quarter lead time directly reflects the flexibility available to policymakers when designing temporary transfers or liquidity support. This is consistent with the literature emphasizing that in rare event scenarios, early warning tools must prioritize timeliness and accuracy over absolute accuracy [20,21].

The first hypothesis (H1), that time-varying financial anomalies improve early detection compared to a static benchmark, is supported by the ablation model. Removing the anomaly layer (no_TVP) leads to significant drops in PR-AUC and recall, validating

H1. From an econometric perspective, shifts in coefficients and volatility help capture the regime-dependent propagation of financial stress to actual activity [10,11]. This implies that the system responds not only to the degree of financial constraints but also to the way political and economic developments shape household behaviour [34]. The second hypothesis (H2), that the hybrid structured and machine-learning pipeline outperforms its individual components, is also confirmed. The tree-based learner represents difficult-to-predetermine interactions and thresholds in the anomaly space, while the structured filter normalises the feature set and reduces spurious patterns. This division of labour aligns with the proven strength of tree ensembles in handling noisy, interacting predictors [14,16]. In practice, Random Forest acts as the primary decision engine, providing the binary WATCH–AMBER–RED escalations at fixed thresholds, whereas the sparse logistic regression model is used as a separate calibration layer that maps base scores into well-behaved probabilities for communication and governance. This explicit separation between the discrete decision layer and the probabilistic calibration layer meets operational and accountability requirements [21]. The tree-based learner represents difficult-to-predetermine interactions and thresholds in the anomaly space, while the structured filter normalizes the feature set and reduces spurious patterns. This division of labor aligns with the proven strength of tree ensembles in handling noisy, interacting predictors [14,16]. In practice, this yields a decision engine (random forest) for binary escalation and a probabilistic engine (sparse logistic regression) for calibrating communications, thus meeting operational and accountability requirements [21].

The calibration results are crucial for implementation. The reliability plots show that the predicted probabilities track the observed frequencies across categories. This is crucial if government agencies are to base preauthorization actions on probability intervals rather than ad hoc decisions. Using sparse logistic regression calibration plots to adjust scores reduces probabilistic errors while maintaining high discriminatory power. This aligns with the view that appropriate assessment and calibration are prerequisites for risk communication in a policy setting [21]. This allows transparent dashboards to complement the WATCH/AMBER/RED tiers and provide realistic probabilities in briefing documents.

Recent behavior demonstrates the governance value of tiering. Due to a lack of confirmation from supplementary models, the near-RED threshold at the end of 2022 was not escalated, thus preventing signal fatigue. At the same time, the alert-oriented WATCH tier flagged elevated but unconfirmed risks, thereby legitimizing increased monitoring without triggering costly interventions. This escalation is precisely what the early warning literature recommends, particularly in situations where losses are asymmetric and the political costs of false alarms are high [22,23]. Figure 4 summarises the escalation sequence and the cross-model confirmation logic across the WATCH–AMBER–RED tiers.

Limitations and next steps are clear. First, the framework is predictive, not causal; it should inform structural judgments, not replace them. For policy use, this implies that model signals are best treated as inputs into scenario analysis, targeted monitoring, and the pre-positioning of temporary measures rather than as mechanical triggers for intervention. Authorities should combine the model's probabilities with institutional knowledge and stress-testing before committing to costly actions. Second, while maintaining strict coverage, expanding the exogenous block to include high-frequency sentiment or credit card issuance proxies could improve short-term sensitivity. Third, regular reestimation and monitoring of the TVP layer can prevent parameter staleness as the regime evolves [10,34]. Within these limitations, the evidence suggests that emerging financial anomalies, filtered through the TVP-SV-VARX and analysed by a robust learner, can provide a normative, reliable early warning of private consumption decline, confirming the research question and hypotheses.

6. Conclusions

The findings support the development of an operational early warning system that translates macro-financial anomalies into timely guidance on private consumption risks. A median warning time of one quarter effectively enables proactive policymaking, allowing decisions to shift from reacting to a recession to preparing for it.

In budgetary practice, the level of warning naturally corresponds to the degree of preparedness. The "WATCH" state requires enhanced monitoring and simple diagnostics-checking data integrity, updating immediate forecasts, and testing transfer mechanisms. The "AMBER" state supports scenario building in the absence of public signals: targeted beneficiary selection, legal review of temporary measures, and identification of administrative bottlenecks. The "RED" state supports the activation of time-limited support—modest, reversible instruments such as temporary transfers to vulnerable households, short-term utility arrears relief, or a narrow VAT deferral that is phased out as pressures ease.

Although the empirical implementation is specific to Romania, the architecture is generic. Other small open economies with similar macro-financial constraints can adopt the same two-layer design, provided that local variables, sample periods, and thresholds are re-estimated. In such settings, the TVP-SV-VARX layer would again extract country-specific macro-financial anomalies, while the ensemble classifier and calibration layer would be re-tuned to local rare-event patterns, preserving the balance between flexibility and interpretability.

Beyond the core framework, similar tiered logic can be applied to supervisory, labour-market, and social policy tools. WATCH-level signals justify enhanced monitoring of household credit and targeted analytical work, while AMBER-level signals support preparatory steps such as eligibility reviews, data-sharing arrangements, and the design of temporary support measures. RED-level signals can then trigger pre-authorised, time-limited interventions, such as targeted transfers or employer-retention schemes, subject to clear exit criteria. Throughout, calibrated probabilities remain primarily an internal input, while external communication focuses on risk levels and readiness rather than model outputs, limiting the risk of amplifying volatility. Governance arrangements—such as a cross-ministerial committee that sets thresholds, reviews warnings, and periodically stress-tests alternative rules—are essential for ensuring that the system remains evidence-based and accountable. Over time, linking activation levels to observed welfare and demand outcomes can inform recalibration and help assess whether early warnings are translated into effective, proportionate policy responses.

Author Contributions: Conceptualization, L.-G.F.; software, M.C.; validation, M.C.; formal analysis, M.M.D.; investigation, A.-I.M.; resources, R.A.P.; data curation, D.A.H.; writing—review and editing, A.C.; visualization, R.L.U.; supervision, A.C.; funding acquisition, A.C. All authors have read and agreed to the published version of the manuscript.

Funding: This paper was co-financed by The Bucharest University of Economic Studies during a PhD program.

Data Availability Statement: Data available upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Adrian, T.; Boyarchenko, N.; Giannone, D. Vulnerable growth. *Am. Econ. Rev.* **2019**, *109*, 1263–1289. [CrossRef]
2. Mian, A.; Rao, K.; Sufi, A. Household balance sheets, consumption, and the economic slump. *Q. J. Econ.* **2013**, *128*, 1687–1726. [CrossRef]
3. Schularick, M.; Taylor, A.M. Credit booms gone bust: Monetary policy, leverage cycles, and financial crises, 1870–2008. *Am. Econ. Rev.* **2012**, *102*, 1029–1061. [CrossRef]

4. Jordà, Ò.; Schularick, M.; Taylor, A.M. When credit bites back. *J. Money Credit. Bank.* **2013**, *45* (Suppl. S1), 3–28. [CrossRef]
5. Kaminsky, G.L.; Reinhart, C.M. The twin crises: The causes of banking and balance-of-payments problems. *Am. Econ. Rev.* **1999**, *89*, 473–500. [CrossRef]
6. Beutel, J.; List, S.; von Schweinitz, G. Does machine learning help us predict banking crises? *J. Financ. Stab.* **2019**, *45*, 100693. [CrossRef]
7. Deaton, A. Saving and liquidity constraints. *Econometrica* **1991**, *59*, 1221–1248. [CrossRef]
8. Stock, J.H.; Watson, M.W. Macroeconomic forecasting using diffusion indexes. *J. Bus. Econ. Stat.* **2002**, *20*, 147–162. [CrossRef]
9. Bernanke, B.S.; Boivin, J.; Eliasz, P. Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *Q. J. Econ.* **2005**, *120*, 387–422. [CrossRef]
10. Primiceri, G.E. Time varying structural vector autoregressions and monetary policy. *Rev. Econ. Stud.* **2005**, *72*, 821–852. [CrossRef]
11. Cogley, T.; Sargent, T.J. Drifts and volatilities: Monetary policies and outcomes in the post-WWII US. *Rev. Econ. Dyn.* **2005**, *8*, 262–302. [CrossRef]
12. Hamilton, J.D. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* **1989**, *57*, 357–384. [CrossRef]
13. Del Negro, M.; Primiceri, G.E. Time varying structural vector autoregressions and monetary policy: A corrigendum. *Rev. Econ. Stud.* **2015**, *82*, 1342–1345. [CrossRef]
14. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
15. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
16. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [CrossRef]
17. Masini, R.P.; Medeiros, M.C.; Mendes, E.F. Machine learning advances for time series forecasting. *J. Econ. Surv.* **2023**, *37*, 511–555. [CrossRef]
18. Goulet Coulombe, P. The macroeconomy as a random forest. *J. Appl. Econom.* **2024**, *39*, 1025–1046. [CrossRef]
19. Vosen, S.; Schmidt, T. Forecasting private consumption: Survey-based indicators vs. *Google Trends. J. Forecast.* **2011**, *30*, 565–578. [CrossRef]
20. Saito, T.; Rehmsmeier, M. The precision–recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE* **2015**, *10*, e0118432. [CrossRef]
21. Gneiting, T.; Raftery, A.E. Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.* **2007**, *102*, 359–378. [CrossRef]
22. Sarlin, P. On policymakers' loss functions and the evaluation of early warning systems. *Econ. Lett.* **2013**, *119*, 1–7. [CrossRef]
23. Holopainen, M.; Sarlin, P. Toward robust early-warning models: A horse race, ensembles and model uncertainty. *Int. J. Forecast.* **2017**, *33*, 1048–1063. [CrossRef]
24. Bauer, M.D.; Mertens, T.M. Economic forecasts with the yield curve. *FRBSF Econ. Lett.* **2018**, *2018*, 7. Available online: https://www.frbsf.org/wp-content/uploads/el2018-07.pdf (accessed on 19 November 2025).
25. Engstrom, E.C.; Sharpe, S.A. The near-term forward spread as a leading indicator: A less distorted mirror. *FEDS Notes* **2018**, *75*, 37–49. [CrossRef]
26. Muraru, A. *Building a Financial Conditions Index for Romania. Occasional Papers 17*; National Bank of Romania: Bucharest, Romania, 2015.
27. Nagy, Á. Measuring Financial Systemic Stress in Romania. *Financ. Stud.* **2016**, *20*, 28–38.
28. Muraru, A.M. The Impact of Global Tensions on the Economic and Financial Cycle in Romania. *Postmod. Open.* **2020**, *11*, 115–128. [CrossRef]
29. Bejenaru, C.E. Early Detection of Sovereign Market Instability in Romania: The Role of SovCISS. *J. East. Eur. Res. Bus. Econ.* **2025**, *2025*, 230203. [CrossRef]
30. National Bank of Romania. *Financial Stability Report*; National Bank of Romania: Bucharest, Romania, 2025.
31. Koop, G.; Korobilis, D. A new index of financial conditions. *Eur. Econ. Rev.* **2014**, *71*, 101–116. [CrossRef]
32. International Monetary Fund. *Global Financial Stability Report*; International Monetary Fund: Washington, DC, USA, 2023. Available online: https://www.imf.org/en/Publications/GFSR (accessed on 20 September 2025).
33. Dynan, K. Is a household debt overhang holding back consumption? *Brook. Pap. Econ. Act.* **2012**, *2012*, 299–362. [CrossRef]
34. Huber, F.; Feldkircher, M. Adaptive shrinkage in Bayesian vector autoregressions. *J. Bus. Econ. Stat.* **2019**, *37*, 738–751. [CrossRef]
35. Forni, M.; Hallin, M.; Lippi, M.; Reichlin, L. The generalized dynamic factor model: Identification and estimation. *Rev. Econ. Stat.* **2000**, *82*, 540–554. [CrossRef]
36. Stock, J.H.; Watson, M.W. Dynamic factor models. In *Oxford Handbook of Economic Forecasting*; Clements, M.P., Hendry, D.F., Eds.; Oxford University Press: Oxford, UK, 2011; pp. 35–59. [CrossRef]
37. Blázquez-García, A.; Conde, A.; Mori, U.; Lozano, J.A. A review on outlier/anomaly detection in time series data. *ACM Comput. Surv.* **2021**, *54*, 1–33. [CrossRef]

38. Mullainathan, S.; Spiess, J. Machine learning: An applied econometric approach. *J. Econ. Perspect.* **2017**, *31*, 87–106. [CrossRef]
39. Medeiros, M.C.; Vasconcelos, G.F.; Veiga, Á.; Zilberman, E. Forecasting inflation in a data-rich environment: The benefits of machine learning. *J. Bus. Econ. Stat.* **2021**, *39*, 98–119. [CrossRef]
40. Bergmeir, C.; Benítez, J.M. On the use of cross-validation for time series predictor evaluation. *Inf. Sci.* **2012**, *191*, 192–213. [CrossRef]
41. Cerqueira, V.; Torgo, L.; Soares, C. Machine learning vs statistical methods for time series forecasting: Size matters. *Mach. Learn.* **2020**, *109*, 1103–1128. [CrossRef]
42. Woodward, R. Organisation for Economic Co-operation and Development. In *OECD*, 2nd ed.; Routledge: London, UK, 2022. [CrossRef]
43. D'Amuri, F.; Marcucci, J. The predictive power of Google searches in forecasting U.S. unemployment. *J. Bank. Financ.* **2017**, *75*, 58–69. [CrossRef]
44. Gneiting, T.; Ranjan, R. Combining predictive distributions. *Electron. J. Stat.* **2013**, *7*, 1747–1782. [CrossRef]
45. Raftery, A.E.; Gneiting, T.; Balabdaoui, F.; Polakowski, M. Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Weather Rev.* **2005**, *133*, 1155–1174. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.