Macroeconomic Fundamentals and the Shape of Sovereign Credit Risk*

Daniele Bianchi
Queen Mary, University of London
d.bianchi@qmul.ac.uk

Teng Jiao
Queen Mary, University of London
t.jiao@qmul.ac.uk

Abstract

Sovereign CDS spreads price the credit risk embedded in government borrowing. We ask how macroeconomic fundamentals predict that risk, and whether a linear mapping from fundamentals to spreads is adequate. Across a large panel of OECD economies, linear methods perform well across countries but lose predictive accuracy after 2022 and fail to capture how spreads vary across maturities. Non-linear methods recover both. While linear forecasts mainly rely on fiscal ratios, non-linear predictions draw on the term-structure slope throughout, inflation during the 2022–2024 monetary tightening, and consumer confidence during the Eurozone crisis. These predictive relationships hold panel-wide rather than country-by-country.

Keywords: Sovereign Credit Risk, Credit Default Swap, Macroeconomic Indicators, Extreme Gradient Boosting, Decision Trees, Interpretable Machine Learning.

JEL codes: F30, F37, G13, G17, C45

*The authors would like to thank Hendrik Bessembinder (the editor), an anonymous referee, Laura Coroneo, Hao Ma, Moqi Groen-Xu, Lucrezia Reichlin, and seminar participants at Queen Mary, University of London and Now-Casting Ltd for their valuable comments and suggestions.

I. Introduction

The price that financial markets place on sovereign default risk is a core component of the cost of borrowing for governments and shapes how capital moves across countries. The post-pandemic rise in public debt and the 2022–2024 monetary tightening have renewed a long-standing question: how do macroeconomic fundamentals enter the price of sovereign risk (e.g., Rogoff, 2022)? That they enter at all is well established; country-specific fundamentals retain substantial explanatory power even after controlling for global factors and credit ratings (Hilscher and Nosbusch, 2010), and a sovereign country's fiscal capacity governs how growth shocks pass through to its default risk (Augustin, Sokolovski, Subrahmanyam, and Tomio, 2022).

What remains less clear is where in the data those fundamentals appear and in what form. Longstaff, Pan, Pedersen, and Singleton (2011) find that sovereign Credit Default Swaps (CDS) spreads are driven mainly by global factors. Ang and Longstaff (2013) documents a systemic component common to U.S. and European sovereigns, whereas Augustin (2018) shows the balance between global and domestic risk shifts with the slope of the CDS term structure. A separate thread reads the time variation in estimated macro-sensitivities as episodes in which spreads detach from fundamentals (Balduzzi, Savona, and Alessi, 2023). These accounts share a modeling choice: each estimates the mapping from fundamentals to spreads within the class of linear models.

We ask whether this linearity assumption, rather than the fundamentals, drives the literature's conclusion. Theory gives reason to doubt it. Doshi, Ericsson, Jacobs, and Turnbull (2013) derive a no-arbitrage framework in which the CDS term structure responds non-linearly to the macroeconomic state, generating threshold effects and regime-contingent loadings that no linear interaction of state variables and maturity can reproduce. We therefore study the *shape* of the mapping from macroeconomic fundamentals to sovereign CDS premiums: the functional form through which fundamentals price sovereign risk, and in particular whether it is non-monotone, varies across maturities, and shifts with the economic regime. If the mapping has this shape, a linear model misstates both how much fundamentals matter and which fundamentals matter.

Our premise is that sovereign CDS premiums measure the price of default risk, isolating it from the interest-rate and liquidity effects that also move bond yields. The analysis uses a panel of 25 macroeconomic indicators for 25 OECD economies from January 2008 to December 2024, with daily CDS spreads at four maturities as the forecasting target. This is a broader, higher-frequency sample than the panels used in prior CDS studies (see Section III). The panel spans five macroeconomic regimes, from the global financial crisis through the 2022–2024 monetary tightening cycle. We use macro variables at their first-release ("real-time") value to avoid look-ahead contamination rather than as an identifying treatment; our findings are quantitatively similar when we use the corresponding revised vintages, which we report in Section C of the Internet Appendix.

We compare two broad classes of forecasting models: linear specifications spanning model-averaging, penalized regression, and principal component methods, and nonlinear specifications including non-parametric methods (tree ensembles and support vector regression), and a feed-forward neural network. Because out-of-sample R^2 in a panel setting is highly sensitive to the benchmark used to evaluate forecasts, we report results against four nested benchmarks that isolate the cross-country, cross-maturity, and within-country-and-maturity components of CDS variation.

The paper makes three main contributions. First, we document that the linear class fails to forecast sovereign CDS premiums across both the cross-maturity and cross-country dimensions during the 2022–2024 monetary tightening, whereas tree-based ensembles forecast accurately across both dimensions. Every linear specification delivers a negative yearly R^2 in every year when forecasts are evaluated against a benchmark that isolates maturity-level variation. Maturity interactions and principal-component reduction do not change this. Extreme gradient boosting (XGBoost), in contrast, is positive in 15 of 16 years at the 1-month horizon. Against a benchmark that isolates country-level variation, linear performance is initially strong but deteriorates post-2022. Elastic Net's yearly 1-month cross-country R^2 on the Global panel falls from an average of 0.53 over 2011–2021 to 0.31 over 2022–2024, while XGBoost shows no such deterioration. Support vector regression forecasts accurately along both dimensions but by smaller margins, while a neural network does so only weakly.

Second, the functional form of XGBoost forecasts is state-dependent in economically interpretable ways. Out-of-sample Shapley values (Shapley et al., 1953; Lundberg, Erion, Chen, DeGrave, Prutkin, Nair, Katz, Himmelfarb, Bansal, and Lee, 2020) show that the predictor attribution is not time-invariant: the term-structure slope is the persistently dominant predictor, year-over-year inflation is activated primarily during the 2022–2024 tightening cycle, and consumer confidence is mostly relevant during the Eurozone sovereign debt crisis. Elastic Net, by contrast, reads sovereign credit risk through the debt-to-GDP ratio with an attribution pattern that is nearly time-invariant. Partial dependence plots refine the reading. The response of expected CDS premiums to the term structure is monotone increasing with a regime-dependent slope, and the regime in which the slope flattens most is the monetary tightening cycle—where the inflation channel activates through a threshold-like step. A linear specification that fits a single coefficient per predictor across all regimes cannot capture these state-dependent shapes.

A key concern is whether these shapes are panel-wide regularities or artifacts of country-specific fits. Our third contribution is to show that they hold over the whole panel of countries. A leave-one-country-out exercise, in which each sovereign is held out of the training sample in turn, shows that the state-dependent mappings tree ensemble fits transport across countries. Because cross-country variation accounts for roughly 73% of the total log-CDS variance, the leave-one-country-out MSE primarily measures how well each method extrapolates the cross-country ordering of CDS levels to a country absent from the training set. XGBoost beats Elastic Net in 22 of 25 countries, with a 59% pooled MSE reduction, and the magnitude of the advantage orders by how idiosyncratic the held-out country's macro fundamentals are: largest for emerging markets, intermediate for the European periphery, and smallest for core OECD economies.

Our findings speak directly to the empirical literature on macroeconomic fundamentals and sovereign CDS spreads. Longstaff et al. (2011) attribute variation in sovereign CDSs mainly to global factors; our design also considers country-specific indicators, so our evidence on predictor relevance complements their decomposition rather than testing it. Once nonlinearity is allowed, predictor relevance reorders relative to the linear class: the term-structure slope—the signal Augustin (2018) identifies as central—is the dominant predictor across regimes even though the linear specifications leave it unused, while the debt and deficit ratios on which those specifications and the broader literature load fall to secondary. Inflation, the short-rate response, and consumer confidence enter only in specific regimes. Finally, what the LASSO of Balduzzi et al. (2023) reads as price variation unrelated to fundamentals is, in our setting, reconciled through the lens of non-linearity.

Our methodological approach connects to two further strands of literature. Machine-learning methods have been applied extensively to sovereign default classification (e.g., Manasse, Schimmelpfennig, and Roubini, 2003; Fioramanti, 2008; Basu and Perrelli, 2019; Silva, Rego, and Frascaroli, 2019) but remain underused for sovereign risk pricing, where Belly, Boeckelmann, Caicedo Graciano, Di Iorio, Istrefi, Siakoulis, and Stalla-Bourdillon (2023) is a notable exception, applying ML to euro-area sovereign bond spreads over the pre-COVID period. We adapt the out-of-sample R^2 evaluation framework that Gu, Kelly, and Xiu (2020) and Bianchi, Büchner, and Tamoni (2021) apply to equity and bond return prediction to the sovereign CDS setting, and characterize predictor-level functional forms with Shapley-value feature attribution and partial dependence plots. The Shapley decomposition replaces the reduction-in-R^2 importance measure used in that literature with a fully allocated attribution of the drivers of expected CDS premiums.

II. A Simple Motivating Framework

A single-name CDS is an over-the-counter credit derivative contract that provides protection against the default of a specific reference entity. The buyer pays a premium, known as the CDS spread, to the seller in exchange for compensation in the event of a default. The premium reflects the perceived risk of default.

To motivate the use of machine learning methods, we build on a standard no-arbitrage argument that equates the discounted cash flows of the CDS buyer and seller. The resulting pricing expression nests the log-linear specifications used in the empirical sovereign-CDS literature as a special case, and makes explicit the functional-form assumptions they impose.

Assume that the buyer seeks compensation equal to $L = 1 - R$, where $R \in [0, 1)$ is the recovery rate and L represents the expected loss as a fraction of the notional. The seller charges the buyer a continuously paid premium $S^*_{j,t}$ at time t. At time t, the two parties agree on the maturity $t + j$ and the compensation L payable if the reference entity defaults. For simplicity, we assume no counterparty risk, meaning the premium depends solely on the reference entity's default risk.

Under these conditions, the *premium leg* of the CDS represents the seller's cash inflows, discounted to present value:

$$PL(S^*_{j,t}; t, t+j) = \mathbb{I}_{\{\tau > t\}} \int_t^{t+j} \mathbb{E}\left[S^*_{j,t} e^{-\int_t^u (r_s + g(\mathbf{x}_s))ds} \middle| \mathcal{F}_t \right] du, \tag{1}$$

where $\mathbb{I}_{\{\tau > t\}}$ is an indicator of no default before time t, τ is the random default time, $\mathcal{F}_s = \{\mathbf{x}_{s-i}\}_{i \geq 0}$ represents the filtration generated by the state variables, and r_s and $g(\mathbf{x}_s)$ are the risk-free discount rate and default intensity rate, respectively, at time s. The expression implies that the seller's cash flows are discounted at a higher rate of $r_s + g(\mathbf{x}_s)$, reflecting both the time value of money and default risk.

The *default leg* of the CDS represents the discounted compensation paid in the event of default before maturity:

$$DL(S^*_{j,t}; t, t+j) = L \times \mathbb{I}_{\{\tau > t\}} \int_t^{t+j} \mathbb{E}\left[e^{-\int_t^u (r_s + g(\mathbf{x}_s))ds} g(\mathbf{x}_u) \middle| \mathcal{F}_t \right] du. \tag{2}$$

Under no-arbitrage conditions (e.g., Augustin and Tédongap, 2016), the CDS premium is considered "fair" when the present values of the premium and default legs are equal, $PL(S^*_{j,t}; t, t+j) = DL(S^*_{j,t}; t, t+j)$. Assuming a constant recovery rate R and a first-order small-spread approximation, this yields an approximate pricing expression for the CDS spread of maturity j:

$$S^*_{j,t} \approx (1-R)\,\mathbb{E}_t\left[\tfrac{1}{j}\int_t^{t+j} g(\mathbf{x}_u)\,du\right], \quad t \geq 0.$$

The dependence of the spread on the state is therefore mediated through the default intensity process $g(\mathbf{x}_u)$. Appendix A provides the explicit derivation and the conditions under which the approximation holds.

We collapse the forward-looking expectation in the pricing expression into a reduced-form function of the state by defining $\gamma(\mathbf{x}_{t-1}, j) \equiv \mathbb{E}_t[\frac{1}{j}\int_t^{t+j} g(\mathbf{x}_u)\,du \mid \mathbf{x}_{t-1}]$, which gives the log spread

$$s_{j,t} \approx \log(1-R) + \log\left(\gamma(\mathbf{x}_{t-1}, j)\right), \tag{3}$$

where $\gamma(\cdot, j)$ is a (possibly unknown) function of the state vector $\mathbf{x}_{t-1}$ for each maturity j. Equation (3) is the starting point for our empirical exercise: it is deliberately agnostic about the functional form of γ, and the predictability of $s_{j,t}$ from one-lag conditioning macroeconomic variables $\mathbf{x}_{t-1}$ depends on how well a chosen specification for γ approximates the true mapping.

The benchmarking approach in the empirical CDS literature assumes a log-linear approximation of equation (3),

$$\log\big(\gamma(\mathbf{x}_{t-1}, j)\big) \approx \boldsymbol{\beta}(j)'\mathbf{x}_{t-1} + c(j), \tag{4}$$

which yields the familiar linear predictive regression $s_{j,t} \propto \boldsymbol{\beta}(j)'\mathbf{x}_{t-1}$ (e.g., Gourieroux, Monfort, and Polimenis, 2006). Under equation (4), term-structure slopes $\boldsymbol{\beta}(j_1) - \boldsymbol{\beta}(j_2)$ are deterministic functions of j alone, and maturity interactions with the state take a strictly additive form (Balduzzi et al., 2023). If equation (4) is an adequate approximation, then linear regressions and their penalized variants with maturity dummies exhaust the predictable variation in sovereign spreads.

In contrast, if the default intensity depends on the state non-monotonically and the shape of that dependence varies with maturity, the CDS term structure is itself a non-linear function of the state. Doshi et al. (2013) derive this result in a no-arbitrage, discrete-time setting and show that it produces threshold effects and regime-contingent loadings that cannot be spanned by linear interactions between the state variables and maturity dummies. Determining whether cross-maturity non-linearity is present in the data is one of the tasks in Section IV.

A distinct implication concerns the short-horizon dynamics of the spread at fixed country-maturity (i, j). The persistence of log CDS spreads, combined with infrequent macroeconomic releases at a daily frequency, implies that $s_{j,t}$ exhibits little predictable variation relative to its own country-maturity trailing mean, $\bar{s}_{ij}$. Any macro-driven specification—linear or non-linear—is therefore expected to underperform $\bar{s}_{ij}$ at short horizons. Section IV documents this and evaluates forecasts against benchmarks that capture cross-country or cross-maturity variation rather than within-country-maturity variation. We interpret the within-country-maturity underperformance as a mechanical consequence of state dynamics and release frequency rather than as evidence against the macro-to-CDS link.

Related evidence for non-linearity in sovereign spread dynamics is provided by Augustin and Tédongap (2016), Doshi, Jacobs, and Zurita (2017), and Chernov, Schmid, and Schneider (2020), with complementary work on endogenous structural breaks in credit risk (e.g., Galil, Shapir, Amiram, and Ben-Zion, 2014; Jeanneret, 2015; Blommestein, Eijffinger, and Qian, 2016; Lahiani, Hammoudeh, and Gupta, 2016) pointing to further sources of non-linearity in the mapping from macro fundamentals to sovereign credit risk. Building on this foundation, our empirical exercise takes equation (3) as the object to estimate and tests whether the log-linear approximation equation (4) is an adequate representation of $\gamma(\cdot, j)$. Evidence of improved out-of-sample prediction accuracy from a non-linear specification over the linear benchmark implies that the former is a closer approximation to $\gamma(\cdot, j)$ than the log-linear null.

III. Data and Research Design

We collect daily CDS premiums for 3-, 5-, 7-, and 10-year maturity contracts on a panel of 25 OECD economies spanning developed and emerging markets over January 2008 to December 2024. Sovereign CDS spreads are sourced from Bloomberg. All contracts use the Complete Restructuring (CR) clause and are denominated in US dollars, with the exception of Canada (No Restructuring, XR) and the United States (CR, denominated in euros).[1] Section B of the Internet Appendix reports the restructuring clause, currency, and earliest available date for each

[1]CR is the standard market convention for European and Asian sovereign CDS. The US sovereign CDS trades in euros because the reference entity is the US dollar issuer. Not all country-maturity pairs are available from the start of

country-maturity pair. The panel includes Austria, Belgium, Canada, Chile, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Ireland, Italy, Japan, Korea, Mexico, the Netherlands, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, the United Kingdom, and the United States.

The seventeen-year window encompasses the 2008–2009 Global Financial Crisis, the 2010–2012 European sovereign debt crisis, the post-crisis low-rate decade, the 2020 COVID-19 pandemic, and the 2022–2024 monetary tightening cycle in response to post-COVID inflationary pressures, providing substantial variation in the economic environments under which CDS premiums are priced.

Figure 1 presents a snapshot of the sample variation in the 5-year CDS spread over time and across countries. Countries with weaker economic fundamentals, such as Ireland, Italy, Spain, and Portugal, exhibit consistently higher and more volatile premiums, and cross-country dispersion peaked during the European sovereign debt crisis. Despite the severity of the 2020 shock, CDS spreads during the COVID-19 outbreak remained relatively subdued compared to historical crisis periods.

[Insert Figure 1 here]

The combination of daily frequency and a broad cross-section places our sample at the recent end of the sovereign CDS literature. Studies that extend further back do so at a lower frequency. Longstaff et al. (2011) use monthly spreads for 26 countries over 2000–2010, and Augustin (2018) uses monthly spreads for 44 countries from 2001 to 2012. Broad daily panels

the sample, and several countries lack certain maturities entirely; the panel is therefore unbalanced along both the country and maturity dimensions.

appear only after the global financial crisis, once single-name sovereign contracts became liquid across a wide set of reference entities, as in the 38-sovereign daily network of Bostanci and Yilmaz (2020) from 2007 to 2019. The closest design to ours is Balduzzi et al. (2023), who pair daily CDS with real-time macroeconomic data for a small cross-section of European sovereigns over 2009–2020. Relative to these later studies, our panel is wider and longer, covering the COVID pandemic and the 2022–2024 tightening cycle.

The panel's variance structure shapes how forecasting performance should be interpreted. On the pooled 2009–2024 sample of log CDS spreads, cross-country differences account for approximately 73% of total variance, cross-maturity differences for approximately 16%, and within-country-maturity time variation for the remaining 11%.[2] The dominance of the cross-country component and the small within-country-maturity share motivate the nested-benchmark evaluation design in Section IV, which isolates where, in the variance decomposition, each forecasting method succeeds or fails.

A. Macroeconomic Indicators and Preprocessing

Our predictor set combines macroeconomic indicators from the "Revisions Analysis Dataset – Infra-annual Indicators" of the OECD Main Economic Indicators (MEI) database with sentiment indicators from the OECD Key Short-Term Economic Indicators dataset, country-specific economic policy uncertainty indexes developed by Baker, Bloom, and Davis

[2]Decomposition computed on the pooled 2009–2024 panel via the identity $\mathrm{Var}(s_{ij,\tau}) = \mathrm{Var}(\bar{s}_{i\cdot}) + \mathrm{Var}(\bar{s}_{\cdot j}) + \mathrm{Var}(s_{ij,\tau} - \bar{s}_{i\cdot} - \bar{s}_{\cdot j} + \bar{s}_{\cdot\cdot})$, normalized to sum to unity.

(2016),[3] and financial variables from Bloomberg. The predictor block covers real-sector indicators, fiscal and inflation measures, financial variables (interest-rate spreads, stock indices, USD exchange rates), and global factors such as the VIX and the global economic activity index of Kilian (2009), incorporating the correction discussed in Kilian (2019).

Two small variable choices are worth noting. The OECD *Composite Leading Indicator* (CLI) was discontinued in the OECD Revisions Database after 2022; we replace it with two sentiment-based indicators from the OECD Key Short-Term Economic Indicators dataset, the Business Confidence Indicator (BCI) and the Consumer Confidence Indicator (CCI), both available over the full sample. The EPU index of Baker et al. (2016) is not available for all countries in our panel; we approximate missing values using the EPU series of a neighboring country with close economic and geographical ties (Sweden for Finland, Spain for Portugal, and so on). Table A1 in Appendix B lists each variable together with its source and transformation code, and the average correlation structure is reported in Section B of the Internet Appendix.

We align our predictor transformations with the FRED-MD convention of McCracken and Ng (2016) to ensure that the predictor block is stationary by construction. Each indicator is mapped to one of four transformation codes (`tcode`). Let x_t denote the original series and $\tilde{x}_t$ the

[3]The EPU data can be accessed at `https://www.policyuncertainty.com`. We are grateful to the authors for making these indicators publicly available.

transformed predictor; we use

$$\tilde{x}_t = \begin{cases} x_t & (\texttt{tcode} = 1, \text{level}) \\ x_t - x_{t-p} & (\texttt{tcode} = 2, \text{first difference}) \\ \log(x_t) - \log(x_{t-p}) & (\texttt{tcode} = 5, \text{log first difference}) \\ (x_t/x_{t-p} - 1) \times 100 & (\texttt{tcode} = 7, \text{percentage change}) \end{cases} \tag{5}$$

with $p = 12$ for the monthly series and $p = 4$ for the quarterly series, so that differences and growth rates are measured year-over-year. The variable-by-variable mapping is reported in Table A1 in Appendix A. GDP and its real-activity components enter as year-over-year percentage changes; price and activity indices as year-over-year log differences; unemployment and short-term interest rates as year-over-year first differences; and scale-free ratios, sentiment indicators, and global financial factors in levels.

The stationarity of the target variable — the log CDS spread — is a separate question. Panel unit-root tests (Pesaran, 2007) and stationarity tests (Hadri, 2000; Kwiatkowski, Phillips, Schmidt, and Shin, 1992) reject their respective nulls for all four CDS maturities at the 5% level or better; the combined pattern points to high persistence of log CDS spreads with slowly drifting levels, consistent with the characterization of sovereign credit spreads in Longstaff et al. (2011) and Augustin (2018). The transformed predictor block is stationary by construction under the FRED-MD transformations. The forecasting results reported in Section IV are unlikely to be affected by spurious correlations. Detailed test results are reported in Section B of the Internet Appendix.

Timing and frequency alignment. The target variable is observed daily, while macroeconomic predictors are released monthly or quarterly. We use each macroeconomic variable at its first-release ("real-time") value rather than at ex-post revised values, which aligns the predictor block with the information set available to market participants at each forecast date (following the broader real-time tradition of Croushore and Stark, 2001; Orphanides, 2001; Ghysels, Horan, and Moench, 2018). The daily frequency of the target, aligned to these first-release indicators, is what makes a real-time macroeconomic predictor block of this breadth usable: monthly sovereign-spread studies, with far fewer target observations per estimation window, are typically restricted to more compact, lower-frequency predictor sets. Appendix A provides an example of the original data structure from the OECD MEI database.

Aligning daily CDS observations with lower-frequency predictor releases requires a choice of interpolation. We follow the forward-filling convention used in the yield-curve forecasting literature (Coroneo, Giannone, and Modugno, 2016; Caruso and Coroneo, 2023). This implies that each macroeconomic variable is held constant at its most recent release value until a new release occurs, so that between two consecutive releases, the last observed value is the only predictor realization in any investor's information set. Figure A2 in Appendix B illustrates the alignment procedure and the release-date assumptions used when the exact release date is not recorded in the source data.

We note that the frequency mismatch in our setting is the reverse of that motivating the MIDAS family (Ghysels, Sinko, and Valkanov, 2007; Ghysels et al., 2018), where a lower-frequency target is forecast from higher-frequency predictors. In our setting, the target is higher-frequency than the predictors, so MIDAS lag polynomials do not apply directly. The natural MIDAS analogs—reverse-MIDAS specifications or mixed-frequency state-space models

(Foroni, Guérin, and Marcellino, 2018)—impose a parametric stochastic model on the high-frequency component of the predictor between macro releases, effectively treating the intra-release period as containing news inferred from interpolation rather than observed.

In contrast, forward-filling treats the intra-release period as containing no new available information, consistent with the literal structure of the data-release process by policymakers. Each forward-filled predictor block is then the common input to all forecasting methods in Section B, linear and nonlinear alike, so that differences in out-of-sample performance reflect the mapping from predictors to CDS rather than the handling of the frequency mismatch.

B. Forecasting Methods and Evaluation Protocol

We compare eight linear specifications against five nonlinear specifications in a unified out-of-sample protocol. The linear models implement the log-linear benchmark of equation (4); the three nonlinear families relax it through distinct representational mechanisms — recursive partitioning of the feature space, kernel-based feature expansions, and hierarchical nonlinear transformations.

Every method receives the same set of macroeconomic indicators $\mathbf{x}_{i,t-1}$ and four CDS maturity dummies, with the maturity dummies entering as interactions in the linear specifications and as ordinary predictors in the nonlinear specifications. For every method that either shrinks coefficients toward zero or trains on cross-sectionally demeaned targets, we subtract the cross-country mean at each time and maturity before estimation, and re-center the maturity-specific means at the forecast stage, thereby preventing shrinkage or demeaning from contaminating the maturity-interaction structure.

Linear specifications. Following Balduzzi et al. (2023), we adopt a "level+slope" form that interacts the real-time macroeconomic indicators $\mathbf{x}_{i,t-1}$ (plus an intercept) with the CDS maturity j,

$$(6) \qquad s_{ij,t} = \mathbf{x}'_{i,t-1}\beta + \sum_{j \in \mathcal{J} \setminus \{3\}} (\mathbf{x}_{i,t-1} \times m_j)' \, \delta_j + \varepsilon_{ij,t},$$

where m_j is a dummy variable that takes the value 1 if the contract has a j-year maturity and 0 otherwise, $\mathcal{J} = \{3, 5, 7, 10\}$, and the 3-year CDS is the reference group. On top of this common form, we consider three strategies for managing the high dimensionality of the predictor block in short rolling windows: coefficient regularization (Ridge, Lasso, Elastic Net; Hsiang, 1975; Tibshirani, 1996; Zou and Hastie, 2005), forecast combination via Complete Subset Regression (CSR; Elliott, Gargano, and Timmermann, 2013) and its equal-weighted special case (EWC; Rapach, Strauss, and Zhou, 2010), and latent-factor dimension reduction via Principal Component Regression (PCR; Ludvigson and Ng, 2009; Stock and Watson, 2011; Kelly and Pruitt, 2013) and Partial Least Squares (PLS). Implementation details are in Section A of the Internet Appendix. For robustness, we also estimate "level-only" variants of each linear specification that omit the maturity-interaction terms in equation (6), and report the corresponding results in Section C of the Internet Appendix.

Tree-based methods. The first nonlinear family is based on recursive partitioning of the feature space (e.g., Breiman, Friedman, Stone, and Olshen, 1984; Döpke, Fritsche, and Pierdzioch, 2017; Medeiros, Vasconcelos, Veiga, and Zilberman, 2021). At each node, the feature space is split on a predictor $x_{i,t-1,p}$ (the p-th component of the state vector $\mathbf{x}_{i,t-1}$) and a cutoff value c, chosen jointly to minimize the within-region sum of squared errors of the target $s_{ij,t}$,

$$\ell = \sum_{(i,j,t)\in I_<} (s_{ij,t} - \bar{s}_{I_<})^2 + \sum_{(i,j,t)\in I_>} (s_{ij,t} - \bar{s}_{I_>})^2, \tag{7}$$

with $I_< = \{(i,j,t) \mid x_{i,t-1,p} \leq c\}$, $I_> = \{(i,j,t) \mid x_{i,t-1,p} > c\}$, and leaf predictions equal to the mean target value in each region. We consider three variants that control a fully grown tree's tendency to overfit. A *pruned regression tree* penalizes leaf count directly. A *random forest* (Breiman, 2001) averages over an ensemble of trees trained on bootstrap samples, with a random subset of predictors evaluated at each split. *XGBoost* (Chen and Guestrin, 2016) trains a sequence of trees in which each new tree is fit to the gradient of a regularized loss function, yielding an additive ensemble

$$\widehat{s}_{ij,t} = \sum_{k=1}^{K} f_k(\mathbf{x}_{i,t-1}), \quad f_k \in \mathcal{F}, \tag{8}$$

with $\mathcal{F}$ the space of regression trees. Implementation details for all these tree-based methods are in Section A of the Internet Appendix.

Kernel-based methods. We estimate an ε-support vector regression (SVR) model with a radial basis function (RBF) kernel (e.g., Cortes and Vapnik, 1995; Drucker, Burges, Kaufman, Smola, and Vapnik, 1996). SVR minimizes a regularized ε-insensitive loss, trading residuals smaller than ε against the Euclidean norm of the feature-space coefficients; the RBF kernel $K_\kappa(u,v) = \exp(-\kappa\|u-v\|^2)$ — where κ is the kernel bandwidth — implicitly maps predictors into an infinite-dimensional feature space, yielding a globally smooth predictor. While recursive partitioning approximates nonlinearities via threshold splits, SVR approximates them via radial basis function expansions centered at support vectors. Implementation details are in Section A of the Internet Appendix.

Neural networks. We consider a feedforward neural network (NN) following the architecture of Gu et al. (2020), adapted to our demeaned panel-data setting. The network has four fully connected hidden layers with 32, 16, 8, and 4 units, `ReLU` activations, batch normalization after each hidden layer, and dropout regularization (0.50 in the first three hidden layers, 0.25 in the last). The model is trained with the Adam optimizer, early stopping on the validation loss, and a `ReduceLROnPlateau` schedule. As for SVR, training uses cross-sectionally demeaned inputs and targets, and forecasts are re-centered at the maturity mean. Neural networks approximate nonlinearities through compositions of affine transformations and nonlinear activations, a third distinct representational mechanism relative to the tree-based and kernel-based families. Full implementation details are documented in Section A of the Internet Appendix.

Training, validation, and testing protocol. We evaluate the predictive performance of each model on an expanding panel of non-overlapping testing months. At each testing month t and for a given forecasting horizon $h \in \{1\text{M}, 3\text{M}, 6\text{M}, 12\text{M}\}$, we apply the following step-by-step procedure.

(i) *Sample partition.* Define the training window as the 12 months preceding t, $[t-12, t-1]$, pooling all 25 countries and four maturities, and the testing window as $[t, t+h]$. No observation dated t or later is used at any step before step (v).

(ii) *Hyperparameter search by time-series cross-validation.* Within the training window, we implement a five-fold forward-chaining cross-validation, illustrated for the one-month testing horizon in Section A of the Internet Appendix. Each fold uses the earliest portion of the training window to fit model parameters and the subsequent block to compute validation loss; fold boundaries do not overlap, preserving chronological order. For each candidate hyperparameter configuration on a pre-specified grid, we record the average validation loss across the five folds.

(iii) *Hyperparameter freezing.* We select the configuration $\widehat{\boldsymbol{\vartheta}}_t^\star$ that minimizes the average validation loss. From this point on, $\widehat{\boldsymbol{\vartheta}}_t^\star$ is fixed for the entire testing window $[t, t+h]$: it is not retuned, updated, or perturbed based on test-period realizations.

(iv) *Final fit.* Holding $\widehat{\boldsymbol{\vartheta}}_t^\star$ fixed, we refit the model on the full 12-month training window $[t-12, t-1]$ to obtain the final parameter estimates $\widehat{\boldsymbol{\psi}}_t$. The validation observations' role is limited to hyperparameter selection in step (iii); they do not contribute to $\widehat{\boldsymbol{\psi}}_t$ through any channel other than their presence in the common training window.

(v) *Out-of-sample forecasting.* Using $(\widehat{\boldsymbol{\vartheta}}_t^\star, \widehat{\boldsymbol{\psi}}_t)$, we generate daily out-of-sample forecasts for the testing window $[t, t+h]$. No model object, hyperparameter, or fit weight is updated during the testing window.

(vi) *Rolling advance.* We advance t by one month and repeat steps (i)–(v). Each model is therefore retrained monthly on a rolling basis.

This protocol has three features that address concerns about the statistical reliability of machine-learning out-of-sample results. First, the data partition at each t satisfies a strict temporal ordering: training precedes validation, which precedes test, with no observations from the testing window entering the training or validation steps. Second, test-period observations do not feed back into hyperparameter choices or parameter estimates, since $\widehat{\boldsymbol{\vartheta}}_t^\star$ and $\widehat{\boldsymbol{\psi}}_t$ are frozen at the end of step (iv). Third, the rolling structure of step (vi) ensures that any parameter drift between testing months reflects only new training data, not repeated peeks at the test block. A time-series split preserves the temporal dependencies inherent in macroeconomic variables and sovereign CDS premiums, avoiding the look-ahead bias that would arise from random shuffling. The rolling-window estimation also accommodates the structural shifts and regime changes highlighted in Section II and documented empirically in Section IV (see also Pesaran and Timmermann, 2007; Clark and McCracken, 2009; Pesaran and Pick, 2011).

IV. On the Predictability of CDS Premiums

We compare the forecasting performance of each model based on the out-of-sample $R^2_{t,oos}$, as defined in Campbell and Thompson (2008):

$$R^2_{t,oos} = 1 - \frac{\sum_i^n \sum_{j \in \mathcal{J}} \sum_\tau^{\mathcal{T}} (s_{ij,\tau} - \widehat{s}_{ij,\tau})^2}{\sum_i^n \sum_{j \in \mathcal{J}} \sum_\tau^{\mathcal{T}} (s_{ij,\tau} - \bar{s})^2}, \tag{9}$$

where $s_{ij,\tau}$ is the observed log CDS premium for country i, maturity j, at day $\tau = 1, \ldots, \mathcal{T}$ within the monthly testing period $t = 1, \ldots, T$, and $\widehat{s}_{ij,\tau}$ refers to the log premium predicted by a given model. Equation (9) compares a given forecast with a "no-predictability" benchmark from a prevailing sample mean $\bar{s}$, estimated over the training period.

We consider four distinct settings for the naive forecast $\bar{s}$, each absorbing a different source of variation in the panel:

$$\bar{s} = \begin{cases} \bar{s}_{\cdot\cdot} & \text{the log CDS averaged across countries and maturities (grand mean).} \\ \bar{s}_{\cdot j} & \text{the log CDS averaged across countries } i = 1, \ldots, n\text{, for each maturity } j. \\ \bar{s}_{i\cdot} & \text{the log CDS averaged across maturities } j \in \mathcal{J}\text{, for each country } i. \\ \bar{s}_{ij} & \text{the country-maturity mean computed on the 12-month training window.} \end{cases}$$

The four benchmarks are ordered along a natural hierarchy of predictive severity, and they admit a direct interpretation in terms of the dimension of CDS variation each benchmark leaves in the denominator of the R^2. Against $\bar{s}_{\cdot j}$, which does not vary across countries, a positive R^2 reflects the model's ability to generate country-specific predictions that improve on a maturity-level constant; the R^2 against this benchmark measures *cross-country* predictability. Against $\bar{s}_{i\cdot}$, which does not vary across maturities, a positive R^2 reflects the ability to generate maturity-specific predictions that improve on a country-level constant; the R^2 measures *cross-maturity* predictability. The grand mean $\bar{s}_{\cdot\cdot}$ is the least demanding benchmark, since it absorbs neither dimension; $\bar{s}_{ij}$ is the most demanding, since it absorbs both and leaves only within-country-maturity time variation in the denominator.

The severity of each benchmark reflects the panel's variance structure.[4] As discussed in

[4]The set $\{\bar{s}_{\cdot\cdot},\, \bar{s}_{\cdot j},\, \bar{s}_{i\cdot},\, \bar{s}_{ij}\}$ is a nested sequence of benchmarks, ordered by the share of total panel variance each absorbs. Their role in the out-of-sample R^2 parallels that of fixed effects in an in-sample panel regression. An R^2 against $\bar{s}_{\cdot\cdot}$ is inflated by the same between-country and between-maturity variance that country and maturity fixed effects would absorb in-sample, and moving to more demanding benchmarks removes that component from the denominator.

Section III, the panel's variance is dominated by the cross-country component, with the cross-maturity and within-country-maturity components accounting for progressively smaller shares. This asymmetry is reflected in the R^2 levels each benchmark delivers: the more of the total variance the benchmark absorbs, the smaller the denominator of the R^2, and the harder it is for any model to achieve a positive score.

All forecasting methods in Section B are fitted once per rolling month on the full 25-country panel, and subgroup results are obtained by restricting the summations in equation (9) to the countries in each subgroup rather than retraining on a country subset. This design choice keeps the training sample size — 25 countries and four maturities over a 12-month window — constant across subgroups, so that subgroup comparisons are apples-to-apples in their training input. The three sub-groups are the South-West Eurozone Periphery (SWEAP: Spain, Portugal, Italy, Ireland), OECD non-EU, and EU ex-SWEAP; the results for SWEAP and OECD non-EU are reported in the main text alongside the Global panel, whereas the results for EU ex-SWEAP are reported in Section C of the Internet Appendix.

A. Time-Varying Predictability

We begin with the yearly out-of-sample R^2, which we compute as the average of $R^2_{t,oos}$ within each calendar year. Figure 2 reports the yearly R^2 on the Global panel, 2010–2024, across two benchmarks (rows) and two forecasting horizons (columns).[5] The heatmap colors each yearly

[5]Sample coverage begins in 2010 because the 12-month rolling training window uses the 2008–2009 data, a period dominated by the Lehman shock, to produce the 2009 forecasts. We retain 2009 observations in the underlying computation but report from 2010 onward. The Global panel at 6-month and 12-month horizons is in Section C of the Internet Appendix.

cell by its R^2 value, with blue shading for positive values, red shading for negative values, and white at zero; cells with $|R^2| \geq 0.7$ are annotated.

[Insert Figure 2 here]

The results pin down where linear methods succeed, where they fail, and where nonlinear methods extend the region of positive out-of-sample performance. Against the cross-country benchmark $\bar{s}_{\cdot j}$ (panels A and B), Elastic Net and CSR are positive in nearly every year of the 2010–2024 sample, averaging around 0.43 at the 1-month horizon and 0.34 at the 3-month horizon; PCR, restricted to five principal components, is weaker by roughly 0.20 but preserves the sign pattern. The equal-weighted combination (EWC), which strips the linear class of its coefficient structure, collapses to essentially zero at both horizons, confirming that the cross-country signal is not an aggregation artifact. This suggests that macro fundamentals—debt-to-GDP, growth, unemployment, financing conditions—carry sufficient cross-country signal for linear specifications with regularization to recover positive predictability. Tree-based ensembles improve on this benchmark (XGBoost averages 0.81 at 1-month, Random Forest 0.76), so nonlinearity contributes in the cross-country dimension as well.

[Insert Figure 3 here]

The picture changes for the cross-maturity benchmark $\bar{s}_{i\cdot}$. Linear methods fail in every cell of panels (C) and (D), with negative R^2 in every year at both horizons. The antecedent is Doshi et al. (2013): in a no-arbitrage CDS pricing model, slope and curvature of the term structure depend on state variables through threshold effects and regime-contingent slope changes that linear maturity interactions cannot span. Tree-based ensembles close the cross-maturity gap most completely. XGBoost is positive against $\bar{s}_{i\cdot}$ at 1-month in every year of the 2010–2024 sample, averaging $+0.49$. Random Forest is positive in 12 of 15 years (mean $+0.31$), with 2010, 2011, and 2022 being the three negative cells. SVR is positive in 12 of 15 years from 2013 onward with systematically smaller magnitudes (mean $+0.23$). The feedforward network is negative in 9 of 15 years. The cleavage lies between tree-based ensembles, which capture interactions between maturity and macro state without researcher-specified structure, and the kernel and neural alternatives.

Linear cross-country predictability also weakens under regime change. Elastic Net's 1-month R^2_{oos} against $\bar{s}_{\cdot j}$ drops from an average of 0.54 over 2011–2021 to 0.32 over 2022–2024. At a 3-month horizon, Elastic Net's 2022 cell is 0.04, compared to 0.68 for XGBoost. Section C of the Internet Appendix shows that the pattern sharpens at 6-month and 12-month horizons. Nonlinear methods retain the signal (XGBoost's 1M mean moves from 0.85 in 2011–2021 to 0.82 in 2022–2024), so the weakening is not about fundamentals becoming uninformative but about linear coefficients failing to transport when the mapping from fundamentals to CDS is state-dependent. This suggests that the "disconnect" suggested by prior literature (e.g., Balduzzi et al., 2023) is more a statement about the rigidity of linear specifications than about the information content of macro fundamentals.

[Insert Figure 4 here]

Figure 3 reports the SWEAP sub-panel, representing countries sharing a monetary regime and a common crisis history. Elastic Net averages 0.57 at 1-month and 0.53 at 3-month against $\bar{s}_{\cdot j}$, against 0.47 and 0.40 in Global and 0.41 and 0.30 in OECD non-EU. The same specification nonetheless shows negative cross-maturity in most years (Elastic Net mean: -0.52 at 1-month).

The Doshi et al. (2013) term-structure nonlinearity is a within-country property of the pricing mapping and is therefore not attenuated by restricting to a more homogeneous sub-panel. The linear cross-maturity failure holds in the group where the linear cross-country fit is strongest.

In the OECD non-EU sub-panel, the institutional setting is the opposite: no shared monetary regime, no common crisis history, and country-level fundamentals ranging from AAA-rated borrowers to investment-grade emerging markets. Figure 4 shows that nonlinear methods nevertheless deliver positive cross-maturity R^2. XGBoost averages $+0.39$ over a 1-month horizon (14 of 15 years positive), and Random Forest $+0.21$ (10 of 15), while every linear specification has a negative mean. The nonlinear cross-maturity advantage, therefore, reflects the functional-form gap at the sovereign level rather than the institutional commonality visible in SWEAP.

B. Four-Benchmark Aggregate Results

We now turn to full-sample averages of the yearly R^2, organized by benchmark and horizon. Table 1 reports the average $\overline{R}^2_{oos}$ on the Global panel for the eight level+slope linear specifications and five nonlinear specifications, against each of the four benchmarks, at each horizon.[6] The table is organized so that scanning a single model column from top to bottom traces the R^2 decomposition as the benchmark becomes more demanding; scanning a single row compares models at a fixed benchmark and horizon.

[Insert Table 1 here]

[6]Level-only linear specifications and the analogous tables for the three country sub-groups are reported in Section C of the Internet Appendix.

The decomposition reading follows the panel's variance structure. Moving from $\bar{s}_{..}$ to $\bar{s}_{.j}$ removes the approximately 16% of total variance attributable to between-maturity differences from the denominator, which modestly lowers every R^2. Elastic Net at 1-month drops from 0.52 to 0.45, XGBoost from 0.83 to 0.81. Moving from $\bar{s}_{..}$ to $\bar{s}_{i.}$ removes approximately 73% of the variance attributable to between-country differences, which is a much larger shift: Elastic Net at a 1-month horizon falls to -0.76, XGBoost to $+0.44$. Linear specifications lose nearly all of their R^2 in this step; tree-based nonlinear specifications retain a substantial portion of it. The magnitude of the R^2 drop between the two middle benchmarks, therefore, suggests that linear specifications' out-of-sample performance is concentrated in variance that country fixed effects would absorb in-sample, whereas tree-based ensembles generate predictions in the dimension that fixed effects cannot.

The bottom row of each horizon block reports the R^2 against $\bar{s}_{ij}$, the 12-month rolling country-maturity training mean. Every model is negative against this benchmark at every horizon; XGBoost is the least negative (-0.64 at the 1-month horizon, -2.77 at the 12-month horizon), and the model ordering is preserved (XGBoost $>$ RF $>$ SVR $>$ linear). This is not evidence against the macro-to-CDS link. Daily log CDS spreads are near-persistent (see Section B of the Internet Appendix), so a trailing country-maturity mean is close to a random-walk forecast, consistent with the characterization of sovereign credit spreads in Longstaff et al. (2011) and Augustin (2018). The difficulty of beating the $\bar{s}_{ij}$ benchmark suggests that the productive macroeconomic signal lies more in the cross-country and cross-maturity dimensions than in the time-series dimension (Duffee, 2002).

Reading within horizon blocks, the results are consistent with the yearly evidence in Section A whereby linear methods capture cross-country but not cross-maturity predictability, and nonlinear methods extend the region of positive R^2 into the cross-maturity dimension and cross-country during crisis regimes. Reading across horizon blocks, the R^2 degrades as the horizon extends, but the relative ordering of model classes—linear versus nonlinear, and within nonlinear, XGBoost≈RF ahead of SVR ahead of NN—is preserved. At the 12-month horizon, Elastic Net's cross-country R^2 is indistinguishable from zero.

Across country groups (Section C of the Internet Appendix), the pattern is preserved without qualitative change. Linear methods capture cross-country variation but fail to capture cross-maturity variation in the OECD non-EU, EU ex-SWEAP, and SWEAP panels; nonlinear methods—XGBoost in particular—deliver positive cross-maturity R^2 in every subgroup at short horizons. The magnitudes differ: SWEAP countries show higher raw R^2 against the grand-mean benchmark at short horizons, reflecting the larger cross-country spread introduced by peripheral Eurozone sovereigns, but the within-group decomposition (the fall from $\bar{s}_{\cdot j}$ to $\bar{s}_{i\cdot}$ and the sign pattern it produces) is invariant.

C. Testing Predictive Accuracy

This subsection reports three pieces of evidence on whether the differences in out-of-sample performance documented in Table 1 and Figure 2 are statistically distinguishable. First, we report pairwise tests of predictive accuracy following Qu, Timmermann, and Zhu (2023), a modified Diebold–Mariano procedure robust to panel dependence that delivers the raw significance of each model-versus-model comparison. Second, we report the Hansen, Lunde, and Nason (2011) Model Confidence Set at 75% and 90% confidence levels, which identifies the set of models that cannot be rejected as the best performer given the data. Third, we report Romano and Wolf (2005, 2016) step-down-adjusted p-values for the pairwise comparisons, controlling the family-wise error rate across the grid of tested hypotheses; this addresses the data-mining concern that the raw Diebold–Mariano p-values ignore the multiplicity of comparisons.

Panels A and B of Table 2 report the first and third tests for XGBoost against each of the twelve rivals, at the one-month testing horizon. Technical details and the Model Confidence Set output are in Appendix A.

[Insert Table 2 here]

Under the unadjusted test, every XGBoost-vs-rival z-statistic is large and negative, indicating lower mean squared error for XGBoost; importantly, every comparison is significant at the 1% level in every group ($|z| \geq 3.86$). The Romano–Wolf adjustment is more demanding: in the Global, OECD non-EU, and EU ex-SWEAP panels, XGBoost's advantage survives familywise error control against all twelve rivals at $\alpha = 0.10$ (maximum adjusted p-value = 0.005 in Global and 0.011 in OECD non-EU, where the marginal rival is SVR in each case). In the SWEAP sub-group, only two of twelve comparisons survive at $\alpha = 0.10$ (PCR at $p = 0.050$, CSR at $p = 0.098$). The remaining models, both linear and non-linear, are indistinguishable from XGBoost under FWER control. The SWEAP pattern reflects the narrower cross-section (five sovereigns versus twenty-five in the Global panel) and the correspondingly wider posterior on the loss-differential distribution, while XGBoost retains the lowest mean squared error in every cell, as shown in Section C of the Internet Appendix.

The Model Confidence Set delivers a stronger verdict (see Table A2): at both confidence levels and across all country groups over the one-month horizon, the MCS contains only XGBoost. That is, no rival model can be retained at the 90% confidence level as statistically indistinguishable from the best forecaster. Put differently, every non-XGBoost specification is rejected from the set of equivalent-best models.

V. Dissecting the Expected CDS Premiums

Section IV documents a sharp split between the linear and nonlinear forecaster classes. Linear methods can capture a great deal of cross-country variation in sovereign CDS premiums but fail to predict cross-maturity variation at any horizon, while tree-based ensembles—XGBoost in particular—deliver positive out-of-sample R^2 against both benchmarks at short horizons. This section opens the XGBoost forecaster to identify the predictors that carry this performance, trace the functional shapes they induce, and test whether those mappings are panel-wide regularities or country-specific fits. The comparison is drawn throughout against the best-performing linear specification, the "level + slope" Elastic Net.

We treat the exercise as descriptive. Shapley values are properties of a given prediction, not of the data-generating process; the XGBoost–Elastic Net contrast therefore identifies which dimensions of the predictor space each functional form reaches, not the structural determinants of sovereign default risk. In that sense, the attribution evidence is the mechanism behind the R^2 ordering in Section IV, not a claim about the equilibrium CDS premiums.

A. Forecasting Performance Attribution

We measure predictor importance with the Shapley value (Shapley et al., 1953), which represents the unique additive and monotone allocation of a model's forecast to its inputs. Additivity means that the predictor contributions sum to the forecast minus a model-specific baseline, so attribution is fully allocated with no residual hidden in an intercept. Monotonicity means that if predictor i raises the forecast by at least as much under one model as under another on every subset, then its Shapley value is at least as large under the first model. The two properties serve complementary roles in our setting. Additivity yields fully allocated attribution within each forecaster, and monotonicity ensures that the attributions are comparable across models fitted to the same inputs.

Formally, let $D = \{1, \ldots, d\}$ index the predictors and $v : \mathcal{P}(D) \to \mathbb{R}$ be a value function that returns a model's forecast as a function of which predictors are "present." The Shapley value of predictor i is

$$\overbrace{\phi_i(v)}^{i\text{'s Shapley value}} = \sum_{S \subseteq D \setminus \{i\}} \underbrace{\frac{|S|!(|D| - |S| - 1)!}{|D|!}}_{S\text{'s weight}} \overbrace{(v(S \cup \{i\}) - v(S))}^{i\text{'s marginal contribution}}, \tag{10}$$

where the weight $|S|!(|D| - |S| - 1)!/|D|!$ is the probability of forming coalition S in a random permutation of the d predictors.[7] equation (10) satisfies $\sum_{i \in D} \phi_i(v) = v(D) - v(\emptyset)$, so Shapley values decompose the forecast-minus-baseline into predictor-specific contributions.

A two-predictor example makes additivity concrete. For a forecaster of the term-structure slope (Term) and the debt-to-GDP ratio (Debt/GDP) with $v(\emptyset) = 0.01$, $v(\{\text{Term}\}) = 0.04$, $v(\{\text{Debt/GDP}\}) = 0.03$, and $v(\{\text{Term}, \text{Debt/GDP}\}) = 0.05$, equation (10) gives $\phi_{\text{Term}} = 0.025$

[7]The two remaining axiomatic properties, symmetry and missingness, ensure respectively that predictors with identical marginal contributions across all subsets receive identical Shapley values and that predictors that never alter the forecast receive a Shapley value of zero. Neither property is binding in our setting.

and $\phi_{\text{Debt/GDP}} = 0.015$.[8] The two sum to $0.04 = v(\{\text{Term}, \text{Debt/GDP}\}) - v(\emptyset)$, and Term receives the larger attribution because its marginal forecast increase is larger whether it is added first or second.

In the empirical implementation, the value function is the model's out-of-sample forecast with the predictors in S set at their explicand values and the predictors in $\bar{S} = D \setminus S$ integrated over baseline observations drawn from the testing distribution:

$$v(S) = \frac{1}{|E|} \sum_{x^b \in E} f(x^e_S, x^b_{\bar{S}}), \tag{11}$$

where x^e is the explicand (the observation being explained), E is the set of baseline observations,[9] and $f(x^e_S, x^b_{\bar{S}})$ denotes the forecast when the predictors in S take the explicand values and those in $\bar{S}$ take the baseline values.[10]

For XGBoost, equation (11) is computed with the `TreeSHAP` algorithm of Lundberg et al. (2020), which exploits the tree structure to return all ϕ_i simultaneously in time linear in the

[8]For $|D| = 2$, this reduces to $\phi_{\text{Term}} = \frac{1}{2}\left[v(\{\text{Term}\}) - v(\emptyset)\right] + \frac{1}{2}\left[v(\{\text{Term}, \text{Debt/GDP}\}) - v(\{\text{Debt/GDP}\})\right]$ and $\phi_{\text{Debt/GDP}} = \frac{1}{2}\left[v(\{\text{Debt/GDP}\}) - v(\emptyset)\right] + \frac{1}{2}\left[v(\{\text{Term}, \text{Debt/GDP}\}) - v(\{\text{Term}\})\right]$.

[9]Setting $S = \emptyset$ in equation (11) gives $v(\emptyset) = |E|^{-1} \sum_{x^b \in E} f(x^b)$, the model's average out-of-sample forecast across the testing distribution. This is the baseline against which every predictor's Shapley contribution in equation (10) is measured.

[10]Chen, Janizek, Lundberg, and Lee (2020) refer to this construction as "model-true" attribution because it preserves the forecaster's functional form rather than conditioning on the empirical distribution of $x_{\bar{S}}$. We prefer it here because our interest is precisely in how XGBoost and Elastic Net use predictors within their own functional form, not in how they would behave under a counterfactual data-generating process. See Chen, Covert, Lundberg, and Lee (2023) for a taxonomy of alternative estimators.

number of leaves.[11] For Elastic Net, equation (11) admits the closed form $\phi_i(f, x^e) = \beta_i(x_i^e - \mu_i)$, where β_i is the penalized coefficient on predictor i and $\mu_i = |E|^{-1} \sum_{x^b \in E} x_i^b$ is the baseline mean.[12] Linear attribution is therefore the deviation of the explicand from the baseline mean, scaled by a single penalized coefficient—fixed across countries, across regimes, and across maturities.

In both cases, the forecaster f at each month is the recursive one-month-ahead estimate used to produce the out-of-sample R^2 reported in Section IV. The explicand x^e is the corresponding testing observation, and the baseline set E is the set of testing observations, so both the evaluation point and the distribution over which $\bar{S}$-predictors are integrated lie outside the training window.

B. Time-Varying and Cross-Sectional Attribution

Figure 5 reports the time-varying mean absolute Shapley values for XGBoost (Panel A) and the "level + slope" Elastic Net (Panel B), averaged across countries in the Global panel.[13] The out-of-sample window spans February 2009 through November 2024, covering the GFC aftermath, the Eurozone sovereign debt crisis, the decade of accommodative policy, the COVID-19 shock, and the 2022–2024 monetary tightening cycle.

---

[11] The `TreeSHAP` package is available on GitHub at https://github.com/ModelOriented/treeshap.

[12] Substituting the linear forecaster $f(x) = \alpha + \sum_i \beta_i x_i$ into equation (11) gives $v(S) = \alpha + \sum_{j \in S} \beta_j x_j^e + \sum_{j \notin S} \beta_j \mu_j$, and the summation in equation (10) telescopes to the stated expression.

[13] Two predictors — the VIX and the index of global real economic activity (IGREA) — are omitted from the heatmap: both are globally uniform on each date, so after cross-sectional demeaning, they are identically zero in the predictor block and carry no Shapley attribution by construction.

[Insert Figure 5 here]

We organize the discussion around three features that together characterize the attribution-level footprint of the cross-maturity R^2 failure in Section IV, namely which predictors XGBoost uses and in which regimes, which predictors Elastic Net uses and how its use differs, and how the contrast holds across the cross-section of countries.[14]

XGBoost uses the term-structure slope (Term Spread) persistently and inflation-related predictors conditionally. Panel A of Figure 5 shows a dominant Term Spread band across most of the 2011–2024 window; this is the only predictor that carries non-trivial attribution in every regime of the sample. The Consumer Confidence Indicator (CCI) is activated during the Eurozone sovereign debt crisis (2011–2014) and flares again during the COVID period (2020–2021). The year-over-year change in consumer prices (CPI) and the first difference of the short-term interest rate (INT) are essentially dormant through 2020 and then activate sharply in 2021–2024, together giving the monetary tightening cycle its distinct signature.

Fiscal ratios (Debt/GDP, EGS/GDP, IGS/GDP), the business sentiment indicator (BCI), and country-specific risk proxies (EPU, FX, equity prices) carry secondary or near-zero attribution throughout. The same three-predictor core — Term Spread, CCI, and the tightening-cycle CPI/INT pair — appears in each of the three country sub-groups (Figures A3–A5 in Appendix B);

[14] The mean absolute Shapley value is roughly an order of magnitude smaller for XGBoost than for Elastic Net throughout the sample. The asymmetry is mechanical. In the Elastic Net closed form $\phi_i = \beta_i(x_i^e - \mu_i)$, attribution scales with the cross-country dispersion of predictor levels, which is large for variables such as Debt/GDP. The tree ensemble bounds each predictor's contribution by the local prediction range. Comparisons throughout this subsection are therefore in terms of attribution structure — which predictors are active, when, in which ranking — not numerical magnitudes.

the Term Spread band is sharpest in SWEAP and the tightening-cycle CPI/INT activation sharpest in OECD non-EU, but the set of active predictors and their timing are stable across groups.

Elastic Net uses debt and sentiment, not the term-structure slope or inflation. Panel B of Figure 5 is dominated by a persistent Debt/GDP band from 2011 through the mid-2020s, with CCI and BCI as secondary bands of similar structure. The three predictors that carry XGBoost's attribution most distinctively — Term Spread, CPI, INT — are essentially white in the heatmap throughout Panel B. The tightening-cycle activation visible in Panel A is not present in Panel B; the Eurozone-crisis CCI flare is present but at a much smaller amplitude; Term Spread is unused at any point in the sample.

Expected CDS premiums from Elastic Net are therefore not a muted version of XGBoost's but depend on a different attribution pattern entirely. The attribution asymmetry is a consequence of functional form.[15] The linear specification does not adapt its predictor use across regimes, and the predictors it does use are disjoint from those XGBoost relies on. The same ranking holds in the three country sub-groups, as shown in Appendix B.

[Insert Figure 6 here]

The asymmetry holds country by country. Figure 6 reports the time-averaged attribution separately for each country. For XGBoost (Panel A), Term Spread is the leading or near-leading predictor for essentially every country in the panel, with secondary concentrations on CCI (sharpest in SWEAP and the EU ex-SWEAP periphery) and on Debt/GDP (concentrated on Italy,

[15] The Elastic Net closed form requires a single penalized coefficient per predictor, fitted on the 12-month training window and held fixed during the testing month. This form can load strongly only on predictors whose relationship to CDS spreads is approximately linear and stable. The XGBoost ensemble faces no such restriction.

Hungary, and Mexico). For Elastic Net (Panel B), Debt/GDP dominates, with the largest loadings falling on countries whose debt-to-GDP ratios deviate most from the panel mean — Japan, Korea, Mexico, Hungary, Italy, and the Slovak Republic. The Japan case is the cleanest illustration of the linear mechanism: sovereign CDS premiums on Japan are among the lowest in the sample, yet Elastic Net attributes the largest Debt/GDP-driven component to Japan because the closed form $\phi_i = \beta_i(x_i^e - \mu_i)$ rewards deviation from the panel mean rather than spread level. The three predictors XGBoost uses most actively — Term Spread, CPI, INT — are white across every country in Panel B. The attribution asymmetry documented in Figure 5 is therefore not an artifact of cross-country averaging; it is a predictor-level property that manifests consistently across all 25 countries in the panel.

C. Pricing of Macroeconomic Risks

The Shapley evidence in Section B shows which predictors XGBoost uses and in which regimes, but not the functional shape of the mapping. Partial dependence plots (Friedman, 2001) recover that shape. We report PDPs for the six predictors with the largest XGBoost attribution in Figure 5: the term-structure slope (Spread), year-over-year log change in consumer prices (CPI), year-over-year first difference of the short-term interest rate (INT), Consumer Confidence Indicator (CCI), debt-to-GDP ratio (Debt/GDP), and Business Confidence Indicator (BCI).

The partial dependence function of a forecaster $\widehat{f}$ with respect to a predictor $x_S \in \mathbb{R}$ is

$$f_{x_S}(x_S) = \int \widehat{f}(x_S, x_C)\, dP(x_C), \tag{12}$$

where x_C are the remaining predictors and the expectation is over the testing-period distribution of x_C. Shapley values measure predictor i's contribution to a single forecast; PDPs measure the conditional mean forecast as a function of i alone.[16]

[Insert Figure 7 here]

The linear specification in Section B constrains each predictor's slope to be constant across the training window. If CDS pricing is stable across regimes, the regime curves in Figure 7 should be parallel shifts; if not, they should differ in slope or shape. The results show that they differ in slope or shape, with the largest divergences occurring during the 2022–2024 tightening cycle.

XGBoost's response to the term spread (Panel A) is monotone increasing in the Eurozone crisis, low-rate, and COVID regimes, with ranges of approximately 1.3, 0.85, and 0.75 log CDS points, respectively. The monetary tightening-cycle curve lies near $+0.29$ over most of the spread range and is no longer monotonic. Similarly, CCI (Panel D) shows a monotone-decreasing Eurozone-crisis curve: the model assigns lower predicted spreads to countries with higher-than-average consumer confidence.

The economic reading of the term spread and consumer confidence is coupled with CPI and INT (Panels B and C) and suggests that during 2022–2024, a flatter curve reflected the short-rate response to inflation rather than to deterioration in growth. Both inflation and short-term interest rates primarily affect expected CDS premiums in the tightening cycle. Thus,

[16]For Spread, CCI, Debt/GDP, and BCI, the horizontal axis is the variable in levels; for CPI (tcode 5), it is the year-over-year percentage change; for INT (tcode 2), it is the year-over-year change in percentage points. All axes are demeaned relative to the cross-sectional mean in the corresponding month. Monthly PDPs are interpolated onto a common within-regime grid (10th–90th percentile of pooled feature values) and averaged across the 24 months of each regime window.

XGBoost treats inflation above the cross-sectional mean and the short-rate response to it as separate channels.

XGBoost's response to Debt/GDP (Panel E) is approximately monotone increasing across three regimes, with mild non-monotonicity around the cross-country mean, and flatter in tightening regimes. The linear component is what Elastic Net captures (see Panel B in Figure 5); XGBoost's secondary attribution picks up the non-monotone corrections. Panel F shows BCI as nearly flat in the pre-tightening regimes and a threshold drop from $+0.04$ to -0.11 log points during tightening. Elastic Net captures the pre-tightening linearity; XGBoost adds the tightening threshold.

Overall, Figure 7 describes one mechanism. XGBoost fits a different mapping in each regime, and each predictor embodies regime dependence that a linear model cannot capture. Term spread breaks in slope and level during tightening; CPI and INT switch from inert to threshold-active; CCI is monotone during the Eurozone crisis and flat elsewhere; and Debt/GDP and BCI each decompose into a linear component, which Elastic Net captures, plus regime-specific non-linear corrections that only the tree ensemble reaches.

D. Cross-Country Extrapolation

Sections B and C establish that XGBoost learns sovereign CDS spreads via state-dependent mappings that the linear class cannot capture. They do not establish *where* those mappings come from. Two readings are consistent with the attribution and partial-dependence evidence. The mappings might be country-specific—that is, what XGBoost fits when it sees Italy's macro history and Italy's CDS history jointly—in which case the nonlinear advantage documented in Section IV reflects country-by-country overfitting that is opaque to a coefficient-based estimator. Or the mappings might be panel-wide—regularities in how macro fundamentals price sovereign risk that hold in common across the sample—in which case the nonlinear advantage reflects a structural feature of sovereign pricing that the tree ensemble learns from cross-country variation.

To distinguish the two readings, we conduct a leave-one-country-out (LOO) exercise. For each country i, we train every model on the remaining 24 countries and forecast country i's daily log CDS spreads from its macroeconomic fundamentals. The MSE is computed over country i's four maturities and the full out-of-sample window. Because cross-country variation accounts for roughly 73% of total panel variance (Section III), the LOO MSE measures how well each forecaster's predictions for a held-out country preserve the cross-country ordering of CDS levels implied by the training set. Linear methods extrapolate using coefficient combinations fit to the other 24 countries; tree ensembles extrapolate using a partition of the feature space that depends on where predictor values lie in the joint distribution. The LOO exercise is thus an indirect test of which form of extrapolation transports.

[Insert Table 3 here]

Table 3 reports the results. Panel A shows that XGBoost has the lowest mean MSE across all models, at 0.45 compared to 1.11 for Elastic Net, with a median country-level ratio (calculated as $\mathrm{MSE}_{\text{model}}/\mathrm{MSE}_{\text{ElasticNet}} - 1$) equal to -0.53, and wins in 22 of 25 countries. The advantage is not specific to the Elastic Net benchmark. XGBoost has a lower mean MSE than every linear specification, including PLS, which has the best median MSE and the most country wins in the linear class.

The gap between mean and median MSE for the linear methods indicates that their poor aggregate performance is driven by a handful of countries whose macro fundamentals lie in the tails of the panel distribution, where coefficient-based extrapolation fails most severely.

Among nonlinear methods, XGBoost is the strongest. Its median country-level ratio against Elastic Net (-0.53) exceeds those of SVR (-0.39), the regression tree (-0.31), the random forest (-0.25), and the neural network (-0.16). The cleavage between gradient-boosted trees and other nonlinear methods, identified in Section IV at the within-panel cross-maturity benchmark, reappears here at the cross-country extrapolation benchmark.

Panel B reports the XGBoost-to-Elastic Net mean MSE ratio for the three country sub-groups of Section IV. The XGBoost advantage is largest for OECD non-EU (-0.63), intermediate for SWEAP (-0.60), and smallest for EU ex-SWEAP (-0.48). The OECD non-EU group contains most of the emerging-market sovereigns in the sample, which is consistent with the extrapolation reading; the relatively smaller advantage for EU ex-SWEAP reflects these countries' macro fundamentals being closer to the panel-typical case.

Overall, the cross-country extrapolation evidence in Table 3 supports a panel-wide reading of the mechanism. Tree-based partitioning fits state-dependent shapes in Section C and transports those shapes to held-out countries at a rate that orders by how close each country's macro fundamentals are to the panel average. This is the cross-country complement to the cross-maturity result in Section IV: XGBoost's advantage persists after removing country-specific training data because the shapes it fits are panel regularities rather than country-specific fits.

VI. Conclusion

This paper studies the mechanism by which macroeconomic fundamentals map onto expected sovereign CDS premiums. Evaluating thirteen specifications against four nested benchmarks on a panel of 25 OECD economies over 2008–2024, we find that linear methods recover the cross-country ordering of spreads but fail to predict their cross-maturity variation in every year and at every horizon. Tree-based ensembles predict the maturity structure that the linear class misses and retain their cross-country accuracy throughout the 2022–2024 monetary policy tightening cycle, while linear performance degrades.

The two classes of models, i.e., linear vs non-linear, rely on different predictors. Ensemble learning loads on the term-structure slope, consumer confidence, and inflation through regime-dependent mappings. Elastic Net loads on the debt-to-GDP ratio with near-constant attribution. A leave-one-country-out exercise confirms that these mappings are panel-wide regularities rather than country-specific fits.

The evidence suggests that macroeconomic fundamentals contain more predictive information about sovereign credit risk than the linear class can extract. The "disconnect" between fundamentals and sovereign spreads documented in earlier work is, in this reading, a statement about the rigidity of the linear specification rather than about the information content of macro fundamentals.

References

Ang, A., and F. A. Longstaff. "Systemic Sovereign Credit Risk: Lessons from the U.S. and Europe." *Journal of Monetary Economics*, 60 (2013), 493–510.

Augustin, P. "The term structure of CDS spreads and sovereign credit risk." *Journal of Monetary Economics*, 96 (2018), 53–76.

Augustin, P.; V. Sokolovski; M. G. Subrahmanyam; and D. Tomio. "In sickness and in debt: The COVID-19 impact on sovereign credit risk." *Journal of Financial Economics*, 143 (2022), 1251–1274.

Augustin, P., and R. Tédongap. "Real economic shocks and sovereign credit risk." *Journal of Financial and Quantitative Analysis*, 541–587.

Baker, S. R.; N. Bloom; and S. J. Davis. "Measuring Economic Policy Uncertainty*." *The Quarterly Journal of Economics*, 131 (2016), 1593–1636. ISSN 0033-5533. doi:10.1093/qje/qjw024.

Balduzzi, P.; R. Savona; and L. Alessi. "Anatomy of a sovereign debt crisis: machine learning, real-time macro fundamentals, and CDS spreads." *Journal of Financial Econometrics*, 21 (2023), 1728–1758.

Basu, S. S., and R. Perrelli. "External Crisis Prediction Using Machine Learning: Evidence from Three Decades of Crises Around the World." (2019).

Belly, G.; L. Boeckelmann; C. M. Caicedo Graciano; A. Di Iorio; K. Istrefi; V. Siakoulis; and A. Stalla-Bourdillon. "Forecasting sovereign risk in the Euro area via machine learning." *Journal of Forecasting*, 42 (2023), 657–684.

Bianchi, D.; M. Büchner; and A. Tamoni. "Bond risk premiums with machine learning." *The Review of Financial Studies*, 34 (2021), 1046–1089.

Blommestein, H.; S. Eijffinger; and Z. Qian. "Regime-dependent determinants of Euro area sovereign CDS spreads." *Journal of Financial Stability*, 22 (2016), 10–21.

Bostanci, G., and K. Yilmaz. "How connected is the global sovereign credit risk network?" *Journal of Banking & Finance*, 113 (2020), 105761.

Breiman, L. "Random Forests." *Machine Learning*, 45 (2001), 5–32. ISSN 1573-0565. doi:10.1023/A:1010933404324.

Breiman, L.; J. Friedman; C. J. Stone; and R. A. Olshen. *Classification and Regression Trees*. Taylor & Francis (1984). ISBN 978-0-412-04841-8.

Campbell, J. Y., and S. B. Thompson. "Predicting Excess Stock Returns out of Sample: Can Anything Beat the Historical Average?" *The Review of Financial Studies*, 21 (2008), 1509–1531. ISSN 0893-9454.

Caruso, A., and L. Coroneo. "Does Real-Time Macroeconomic Information Help to Predict Interest Rates?" *Journal of Money, Credit and Banking*, 55 (2023), 2027–2059.

Chen, H.; I. C. Covert; S. M. Lundberg; and S.-I. Lee. "Algorithms to estimate Shapley value feature attributions." *Nature Machine Intelligence*, 5 (2023), 590–601.

Chen, H.; J. D. Janizek; S. Lundberg; and S.-I. Lee. "True to the model or true to the data?" *arXiv preprint arXiv:2006.16234*.

Chen, T., and C. Guestrin. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. "Xgboost: A scalable tree boosting system.", 785–794 (2016).

Chernov, M.; L. Schmid; and A. Schneider. "A macrofinance view of US Sovereign CDS premiums." *The Journal of Finance*, 75 (2020), 2809–2844.

Clark, T. E., and M. W. McCracken. "Improving forecast accuracy by combining recursive and rolling forecasts." *International Economic Review*, 50 (2009), 363–395.

Coroneo, L.; D. Giannone; and M. Modugno. "Unspanned macroeconomic factors in the yield curve." *Journal of Business & Economic Statistics*, 34 (2016), 472–485.

Cortes, C., and V. Vapnik. "Support-vector networks." *Machine learning*, 20 (1995), 273–297.

Croushore, D., and T. Stark. "A real-time data set for macroeconomists." *Journal of econometrics*, 105 (2001), 111–130.

Diebold, F., and R. Mariano. "Comparing Predictive Accuracy." *Journal of Business & Economic Statistics*, 13 (1995), 253–63. Publisher: American Statistical Association.

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

Doshi, H.; J. Ericsson; K. Jacobs; and S. M. Turnbull. "Pricing credit default swaps with observable covariates." *The Review of Financial Studies*, 26 (2013), 2049–2094.

Doshi, H.; K. Jacobs; and V. Zurita. "Economic and financial determinants of credit risk premiums in the sovereign CDS market." *The Review of Asset Pricing Studies*, 7 (2017), 43–80.

Drucker, H.; C. J. Burges; L. Kaufman; A. Smola; and V. Vapnik. "Support vector regression machines." *Advances in neural information processing systems*, 9.

Duffee, G. R. "Term premia and interest rate forecasts in affine models." *The Journal of Finance*, 57 (2002), 405–443.

Duffie, D. "Credit swap valuation." *Financial Analysts Journal*, 55 (1999), 73–87.

Elliott, G.; A. Gargano; and A. Timmermann. "Complete subset regressions." *Journal of Econometrics*, 177 (2013), 357–373.

Fioramanti, M. "Predicting sovereign debt crises using artificial neural networks: a comparative approach." *Journal of Financial Stability*, 4 (2008), 149–164.

Foroni, C.; P. Guérin; and M. Marcellino. "Using low frequency information for predicting high frequency variables." *International Journal of Forecasting*, 34 (2018), 774–787.

Friedman, J. H. "Greedy function approximation: a gradient boosting machine." *Annals of statistics*, 1189–1232.

Galil, K.; O. M. Shapir; D. Amiram; and U. Ben-Zion. "The determinants of CDS spreads." *Journal of Banking & Finance*, 41 (2014), 271–282.

Ghysels, E.; C. Horan; and E. Moench. "Forecasting through the Rearview Mirror: Data Revisions and Bond Return Predictability." *The Review of Financial Studies*, 31 (2018), 678–714. ISSN 0893-9454. doi:10.1093/rfs/hhx098.

Ghysels, E.; A. Sinko; and R. Valkanov. "MIDAS regressions: Further results and new directions." *Econometric reviews*, 26 (2007), 53–90.

Gourieroux, C.; A. Monfort; and V. Polimenis. "Affine models for credit risk analysis." *Journal of Financial Econometrics*, 4 (2006), 494–530.

Gu, S.; B. Kelly; and D. Xiu. "Empirical asset pricing via machine learning." *The Review of Financial Studies*, 33 (2020), 2223–2273.

Hadri, K. "Testing for stationarity in heterogeneous panel data." *The econometrics journal*, 3 (2000), 148–161.

Hansen, P. R. "Asymptotic tests of composite hypotheses."

Hansen, P. R.; A. Lunde; and J. M. Nason. "The model confidence set." *Econometrica*, 79 (2011), 453–497.

Hilscher, J., and Y. Nosbusch. "Determinants of Sovereign Risk: Macroeconomic Fundamentals and the Pricing of Sovereign Debt." *Review of Finance*, 14 (2010), 235–262. doi:10.1093/rof/rfq005.

Hsiang, T. "A Bayesian view on ridge regression." *Journal of the Royal Statistical Society: Series D*, 24 (1975), 267–268.

Jeanneret, A. "The dynamics of sovereign credit risk." *Journal of Financial and Quantitative Analysis*, 963–985.

Kelly, B., and S. Pruitt. "Market Expectations in the Cross-Section of Present Values." *The Journal of Finance*, 68 (2013), 1721–1756. doi:10.1111/jofi.12060.

Kilian, L. "Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market." *American economic review*, 99 (2009), 1053–1069.

Kilian, L. "Measuring global real economic activity: Do recent critiques hold up to scrutiny?" *Economics Letters*, 178 (2019), 106–110.

Kwiatkowski, D.; P. C. Phillips; P. Schmidt; and Y. Shin. "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?" *Journal of econometrics*, 54 (1992), 159–178.

Lahiani, A.; S. Hammoudeh; and R. Gupta. "Linkages between financial sector CDS spreads and macroeconomic influence in a nonlinear setting." *International Review of Economics & Finance*, 43 (2016), 443–456.

Longstaff, F. A.; J. Pan; L. H. Pedersen; and K. J. Singleton. "How Sovereign Is Sovereign Credit Risk?" *American Economic Journal: Macroeconomics*, 3 (2011), 75–103. ISSN 1945-7707.

Ludvigson, S. C., and S. Ng. "Macro Factors in Bond Risk Premia." *Review of Financial Studies*, 22 (2009), 5027–5067.

Lundberg, S. M.; G. Erion; H. Chen; A. DeGrave; J. M. Prutkin; B. Nair; R. Katz; J. Himmelfarb; N. Bansal; and S.-I. Lee. "From local explanations to global understanding with explainable AI for trees." *Nature machine intelligence*, 2 (2020), 56–67.

Manasse, P.; A. Schimmelpfennig; and N. Roubini. "Predicting Sovereign Debt Crises." *IMF Working Papers*, 03. doi:10.5089/9781451875256.001.

McCracken, M. W., and S. Ng. "FRED-MD: A monthly database for macroeconomic research." *Journal of Business & Economic Statistics*, 34 (2016), 574–589.

Medeiros, M. C.; G. F. Vasconcelos; Á. Veiga; and E. Zilberman. “Forecasting inflation in a data-rich environment: the benefits of machine learning methods.” *Journal of Business & Economic Statistics*, 39 (2021), 98–119.

Newey, W. K., and K. D. West. “A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix.” (1987).

Orphanides, A. “Monetary policy rules based on real-time data.” *American Economic Review*, 91 (2001), 964–985.

Pesaran, M. H. “A simple panel unit root test in the presence of cross-section dependence.” *Journal of applied econometrics*, 22 (2007), 265–312.

Pesaran, M. H., and A. Pick. “Forecast combination across estimation windows.” *Journal of Business & Economic Statistics*, 29 (2011), 307–318.

Pesaran, M. H., and A. Timmermann. “Selection of estimation window in the presence of breaks.” *Journal of Econometrics*, 137 (2007), 134–161.

Qu, R.; A. Timmermann; and Y. Zhu. “Comparing forecasting performance with panel data.” *International journal of forecasting*.

Rapach, D. E.; J. K. Strauss; and G. Zhou. “Out-of-sample equity premium prediction: Combination forecasts and links to the real economy.” *The Review of Financial Studies*, 23 (2010), 821–862.

Rogoff, K. “Emerging market sovereign debt in the aftermath of the pandemic.” *Journal of Economic Perspectives*, 36 (2022), 147–166.

Romano, J. P., and M. Wolf. “Exact and approximate stepdown methods for multiple hypothesis testing.” *Journal of the American Statistical Association*, 100 (2005), 94–108.

Romano, J. P., and M. Wolf. “Efficient computation of adjusted p-values for resampling-based stepdown multiple testing.” *Statistics & Probability Letters*, 113 (2016), 38–40.

Schönbucher, P. J. *Credit derivatives pricing models: models, pricing and implementation*. John Wiley & Sons (2003).

Shapley, L. S., et al. “A value for n-person games.”

Silva, D. R. B. d.; T. G. d. Rego; and B. Frascaroli. “Sovereign risk ratings’ country classification using machine learning.” (2019).

Stock, J. H., and M. Watson. “Dynamic factor models.” *Oxford Handbooks Online*.

Tibshirani, R. “Regression shrinkage and selection via the lasso.” *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.

Zou, H., and T. Hastie. “Regularization and variable selection via the elastic net.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67 (2005), 301–320.

Table 1

Four-benchmark decomposition of out-of-sample $\overline{R}^2_{oos}$, Global panel.

The table reports the full-sample average of yearly $R^2_{t,oos}$ (2009–2024) for each model, forecasting horizon, and naive benchmark. Benchmarks are ordered top-to-bottom within each horizon block by increasing severity: $\bar{s}_{..}$ (grand mean), $\bar{s}_{.j}$ (cross-country predictability), $\bar{s}_{i.}$ (cross-maturity predictability), $\bar{s}^{\text{train}}_{ij}$ (within-cell training mean). Within each model class (linear vs non-linear) and at each row, the best-performing specification is in bold. Level-only linear specifications and country sub-group tables are reported in the companion Internet Appendix.

		Level + Slope linear								*Nonlinear models*				
Horizon	Benchmark	OLS	EWC	CSR	Lasso	Ridge	EN	PCR	PLS	SVR	RT	RF	XGB	NN
1-month	$\bar{s}_{..}$	0.44	0.19	0.51	0.51	0.49	**0.52**	0.36	0.48	0.77	0.69	0.79	**0.83**	0.65
	$\bar{s}_{.j}$	0.36	0.07	0.44	0.44	0.41	**0.45**	0.25	0.41	0.74	0.65	0.76	**0.81**	0.60
	$\bar{s}_{i.}$	−1.06	−2.11	−0.79	−0.81	−0.90	**−0.76**	−1.42	−0.93	0.19	−0.10	0.26	**0.44**	−0.31
	$\bar{s}_{ij}$	−5.77	−10.25	−5.27	−5.14	−5.50	**−5.05**	−8.41	−5.92	−1.50	−2.45	−1.25	**−0.64**	−3.17
3-month	$\bar{s}_{..}$	0.10	0.17	**0.44**	0.44	0.37	0.44	0.30	0.31	0.69	0.61	0.73	**0.77**	0.59
	$\bar{s}_{.j}$	0.01	0.04	0.36	0.36	0.28	**0.36**	0.19	0.21	0.65	0.56	0.70	**0.74**	0.54
	$\bar{s}_{i.}$	−2.46	−2.12	−1.03	−1.05	−1.35	**−1.02**	−1.56	−1.55	−0.12	−0.41	0.05	**0.20**	−0.49
	$\bar{s}_{ij}$	−7.86	−9.83	−5.67	−5.66	−6.49	**−5.62**	−8.42	−7.57	−2.43	−3.42	−1.83	**−1.30**	−3.55
6-month	$\bar{s}_{..}$	−0.07	0.15	**0.36**	0.35	0.23	0.35	0.23	0.03	0.59	0.52	0.67	**0.70**	0.52
	$\bar{s}_{.j}$	−0.18	0.02	**0.28**	0.26	0.13	0.25	0.10	−0.11	0.54	0.45	0.62	**0.66**	0.45
	$\bar{s}_{i.}$	−2.68	−2.13	**−1.26**	−1.34	−1.74	−1.35	−1.77	−2.48	−0.44	−0.70	−0.16	**−0.04**	−0.73
	$\bar{s}_{ij}$	−8.78	−9.26	**−6.06**	−6.33	−7.42	−6.35	−8.49	−10.38	−3.27	−4.24	−2.37	**−1.94**	−4.10
12-month	$\bar{s}_{..}$	−0.30	0.11	**0.23**	0.14	−0.01	0.12	0.09	−0.51	0.45	0.40	0.58	**0.61**	0.39
	$\bar{s}_{.j}$	−0.46	−0.03	**0.12**	0.02	−0.15	0.00	−0.06	−0.73	0.37	0.31	0.53	**0.56**	0.30
	$\bar{s}_{i.}$	−3.56	−2.22	**−1.76**	−2.14	−2.63	−2.18	−2.22	−4.41	−0.95	−1.17	−0.49	**−0.41**	−1.20
	$\bar{s}^{\text{train}}_{ij}$	−11.14	−8.64	**−6.92**	−8.04	−9.45	−8.14	−8.82	−16.16	−4.45	−5.12	−3.04	**−2.77**	−5.11

Table 2

Pairwise statistical significance vs XGBoost.

Panel A reports the panel-robust Diebold–Mariano z-statistics of Qu et al. (2023) for XGBoost against each rival. Negative values indicate lower loss for XGBoost. Panel B reports Romano–Wolf step-down adjusted p-values for the twelve XGBoost-vs-rival comparisons. These are computed on monthly-aggregated squared-error losses via stationary block bootstrap with $B = 5{,}000$ replications and block length $\lceil \sqrt{T} \rceil = 14$, controlling the familywise error rate over this family. Sample: 2009–2024.

	Panel A: DM z-statistic				**Panel B:** RW adj. p-value			
Rival	Global	OECD/EU	EU/SWEAP	SWEAP	Global	OECD/EU	EU/SWEAP	SWEAP
OLS	−10.73	−8.14	−9.81	−5.32	0.004	0.011	0.001	0.165
EWC	−40.14	−23.99	−14.45	−10.28	0.000	0.000	0.000	0.124
CSR	−22.95	−12.28	−10.54	−10.20	0.000	0.003	0.001	0.098
Lasso	−20.15	−12.52	−9.48	−6.85	0.000	0.003	0.004	0.124
Ridge	−19.09	−13.13	−9.84	−6.40	0.000	0.002	0.003	0.124
EN	−21.19	−13.05	−9.32	−7.95	0.000	0.003	0.004	0.124
PCR	−12.76	−11.95	−8.69	−11.63	0.004	0.003	0.008	0.050
PLS	−9.53	−9.26	−7.35	−4.86	0.005	0.011	0.008	0.167
SVR	−7.54	−7.26	−3.86	−3.88	0.005	0.011	0.063	0.167
RT	−7.81	−6.31	−6.20	−4.45	0.005	0.011	0.008	0.167
RF	−11.52	−7.50	−5.14	−4.83	0.003	0.011	0.014	0.167
NN	−9.81	−7.71	−7.90	−7.24	0.005	0.011	0.008	0.124
# rej. @ 10%	12/12	12/12	12/12	12/12	12/12	12/12	12/12	2/12

Table 3

Leave-one-country-out forecasting performance.

Panel A reports the mean and median mean-squared error (MSE) of the one-month-ahead forecasts of the log CDS spread for each of the 13 models considered, when each country is held out of the training sample in turn. Linear models (level + slope specification) are reported in the top block; nonlinear models in the bottom block. Columns 3–4 report the mean and median of the country-level ratio $\mathrm{MSE}_{\text{model}}/\mathrm{MSE}_{\text{ElasticNet}} - 1$ across the 25 held-out countries; column 5 counts the countries in which the model achieves lower MSE than Elastic Net. Panel B reports the ratio of the cross-country mean MSE of XGBoost to that of Elastic Net, separately for the three country groups: OECD non-EU, SWEAP, and EU ex-SWEAP. The out-of-sample window is 15th February 2009 to 15th November 2024.

Panel A: Cross-country extrapolation by method

	MSE		Ratio -1 vs. ENet		
Model	Mean	Median	Mean	Median	Wins
Linear (level+slope)					
OLS	0.87	0.77	-0.02	$+0.00$	13
EWC	0.80	0.80	$+0.12$	$+0.19$	11
CSR	0.80	0.76	$+0.08$	$+0.15$	11
Lasso	1.10	0.79	-0.02	$+0.01$	12
Ridge	1.11	0.77	$+0.02$	-0.00	13
Elastic Net	1.11	0.76	0.00	0.00	—
PCR	0.84	0.61	-0.03	-0.06	14
PLS	0.98	0.52	-0.19	-0.26	21
Nonlinear					
SVR	0.57	0.37	$+0.26$	-0.39	20
RT	0.59	0.50	-0.12	-0.31	20
RF	0.57	0.52	-0.25	-0.25	20
XGBoost	0.45	0.39	-0.34	-0.53	22
NN	0.64	0.56	-0.20	-0.16	20

Panel B: XGBoost vs. Elastic Net by country group

Group	n	XGBoost mean MSE	Elastic Net mean MSE	Ratio -1
OECD non-EU	12	0.58	1.55	-0.63
EU ex-SWEAP	9	0.34	0.65	-0.48
SWEAP	4	0.33	0.83	-0.60
All	25	0.45	1.11	-0.59

Figure 1

Sovereign CDS sample variation.

The figure reports the month-year cross-sectional distribution of the daily 5-year CDS premium (left panel) and the country time series distribution of the 5-year CDS premium (right panel). The sample period is daily from January 2008 to December 2024.

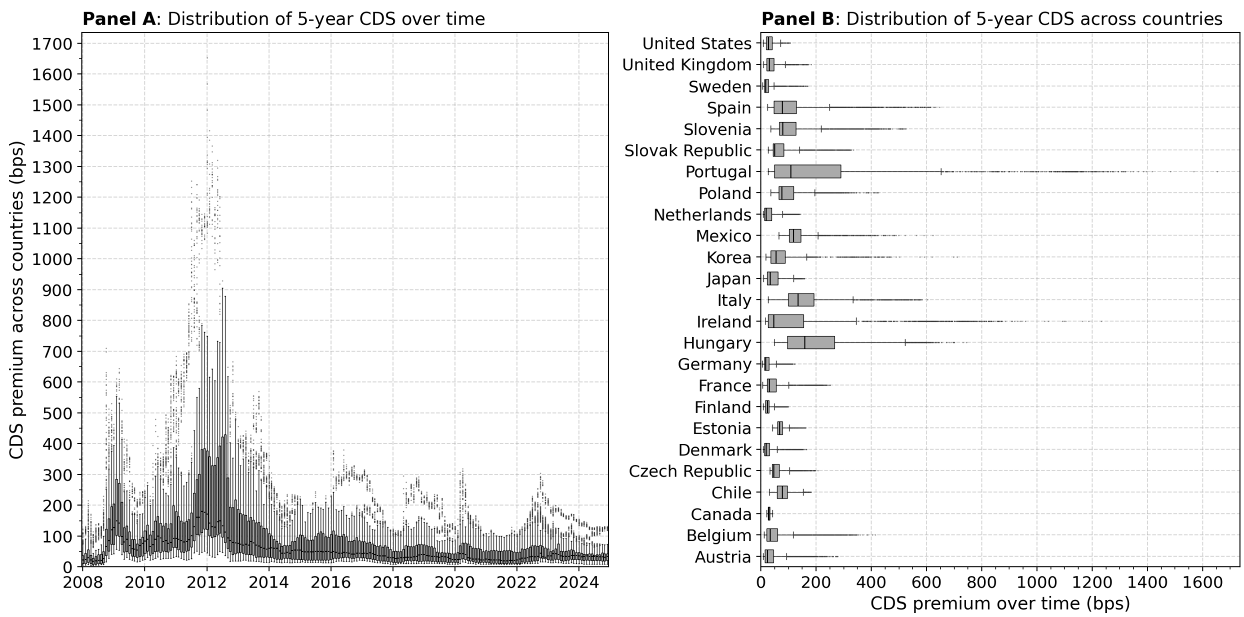


Figure 2

Yearly R^2_{OOS} across benchmarks and horizons, Global panel.

The figure reports the annual average of $R^2_{t,oos}$ on the Global panel of 25 countries. Rows: the top row uses the benchmark $\bar{s}_{\cdot j}$ (log CDS averaged across countries); the bottom row uses $\bar{s}_{i\cdot}$ (averaged across maturities). Columns: 1-month (left) and 3-month (right) testing horizons. Cells are colored by R^2 using a diverging scale centered at zero; values with $|R^2| \geq 0.7$ are annotated.

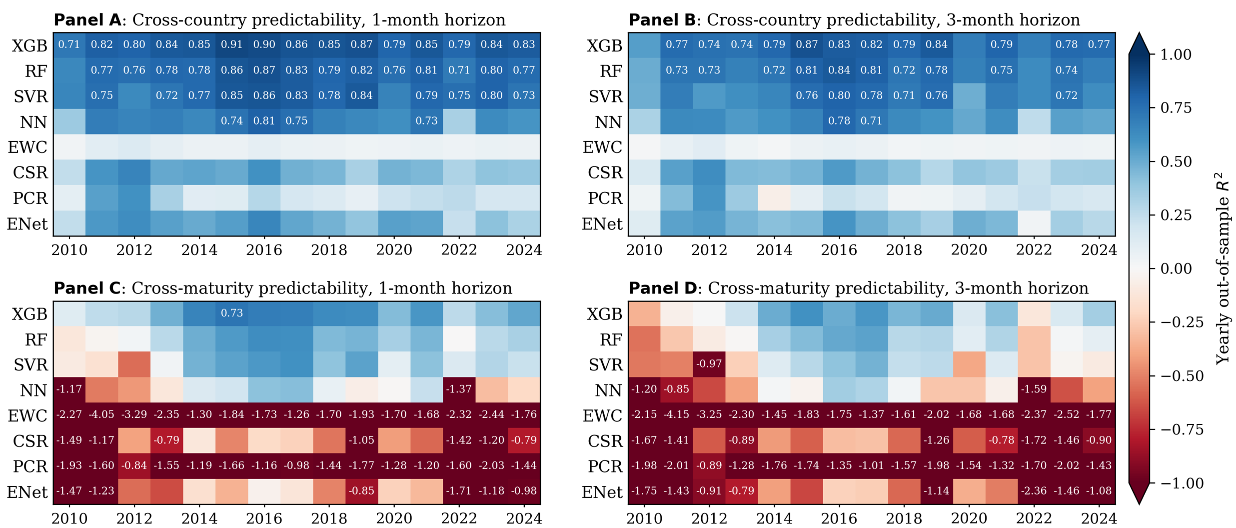


Figure 3

Yearly R^2_{OOS} across benchmarks and horizons, SWEAP panel.

The figure reports the annual average of $R^2_{t,oos}$ for the SWEAP country sub-group (Spain, Portugal, Italy, Ireland). Rows: cross-country predictability against $\bar{s}_{\cdot j}$ (top) and cross-maturity predictability against $\bar{s}_{i\cdot}$ (bottom). Columns: 1-month (left) and 3-month (right) testing horizons. Models and formatting follow Figure 2.

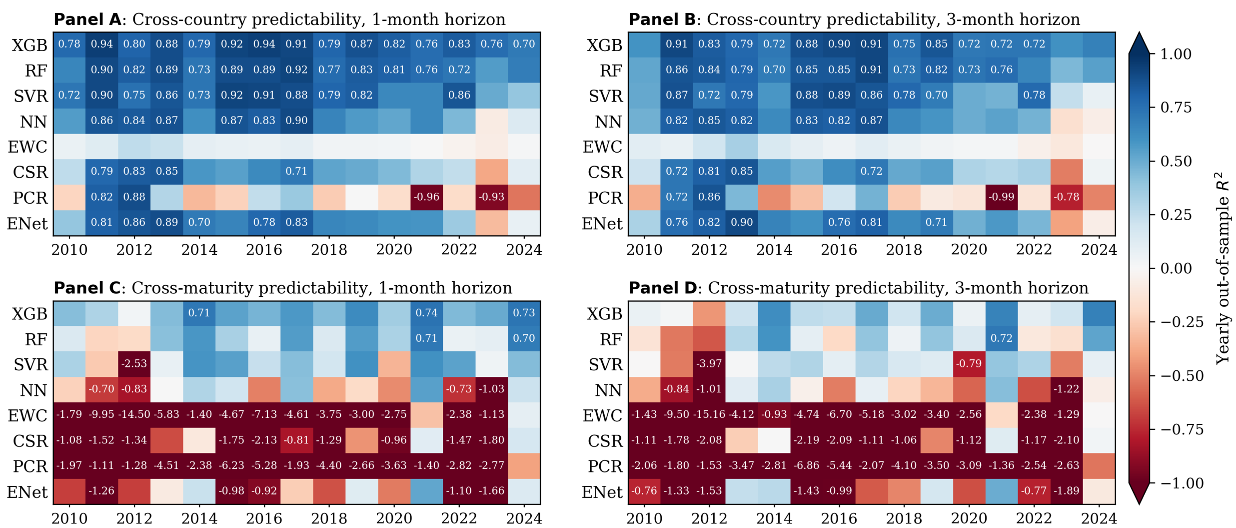


Figure 4

Yearly R^2_{OOS} across benchmarks and horizons, OECD non-EU panel.

The figure reports the annual average of $R^2_{t,oos}$ for the OECD non-EU country sub-group. Rows: cross-country predictability against $\bar{s}_{\cdot j}$ (top) and cross-maturity predictability against $\bar{s}_{i\cdot}$ (bottom). Columns: 1-month (left) and 3-month (right) testing horizons. Models and formatting follow Figure 2.

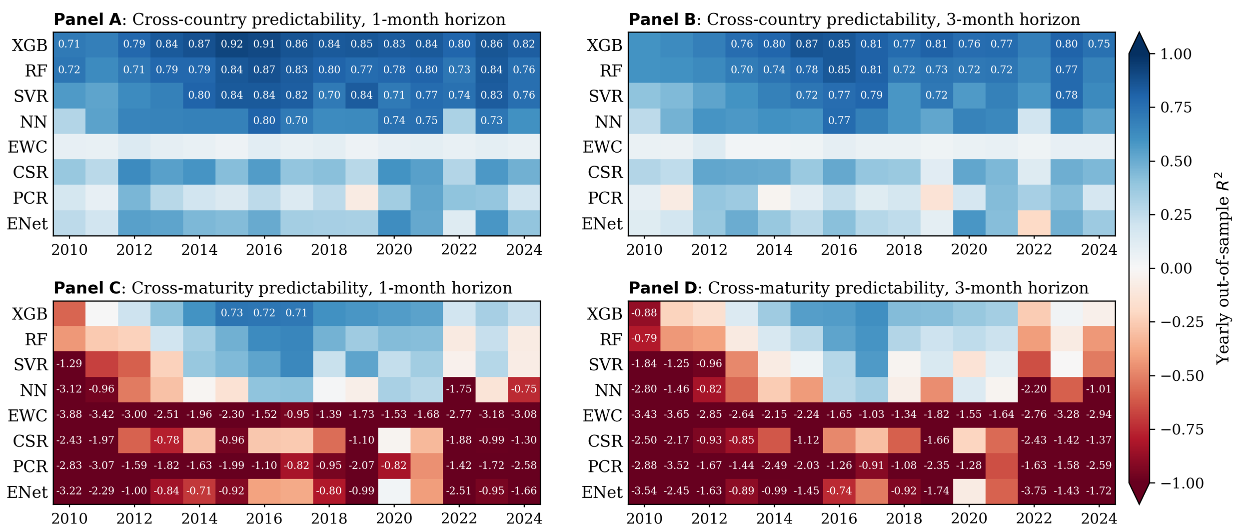


Figure 5

Time-varying importance of macroeconomic variables.

The figure displays the out-of-sample mean absolute Shapley values from XGBoost (Panel A) and the "level+slope" Elastic Net (Panel B), averaged across countries in the Global panel at each one-month testing period. The color scale is sequential (darker indicates higher importance). Rows correspond to the predictors described in Section A. The out-of-sample period is February 2009 through November 2024.

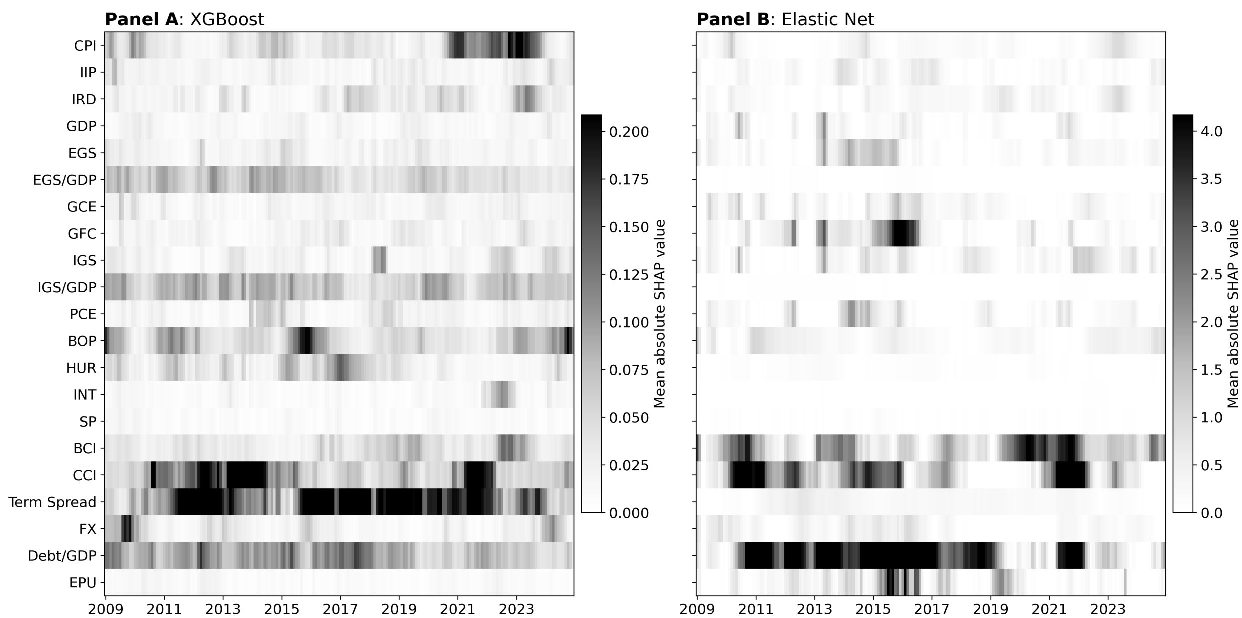


Figure 6

Country-specific feature attribution.

The figure displays the mean absolute Shapley values from XGBoost (Panel A) and the "level+slope" Elastic Net (Panel B), with each cell reporting a predictor's monthly out-of-sample Shapley values averaged across the testing period for the indicated country. Columns are grouped by country group (SWEAP; EU ex-SWEAP; OECD non-EU) and ordered alphabetically within group; rows correspond to the predictors described in Section A. The out-of-sample period is February 2009 through November 2024.

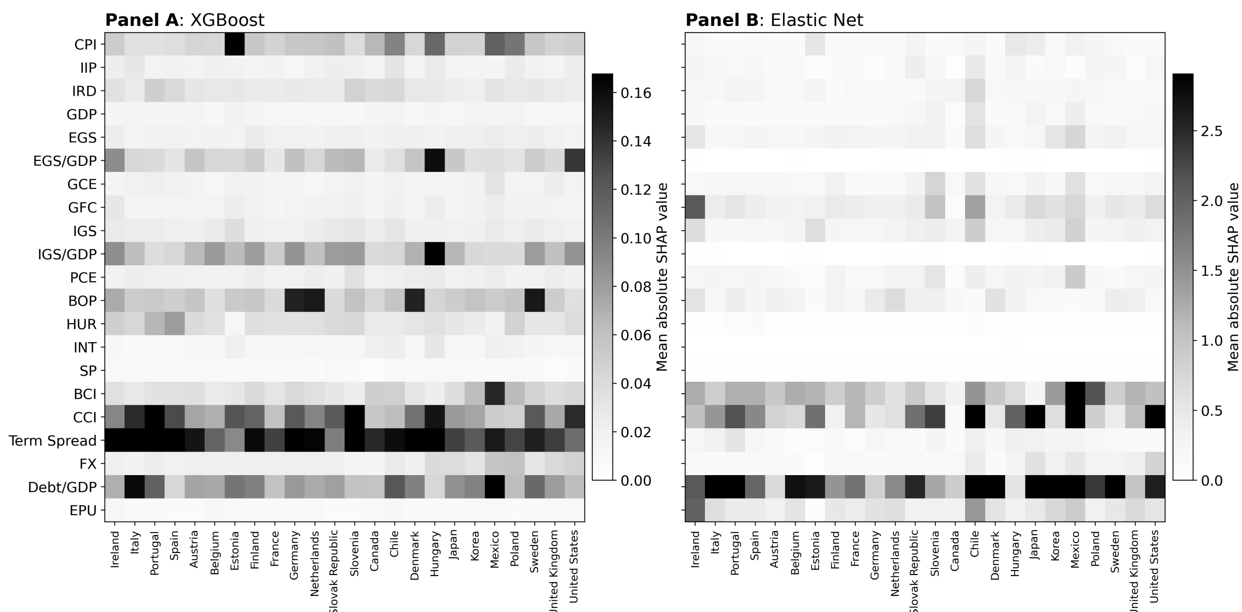
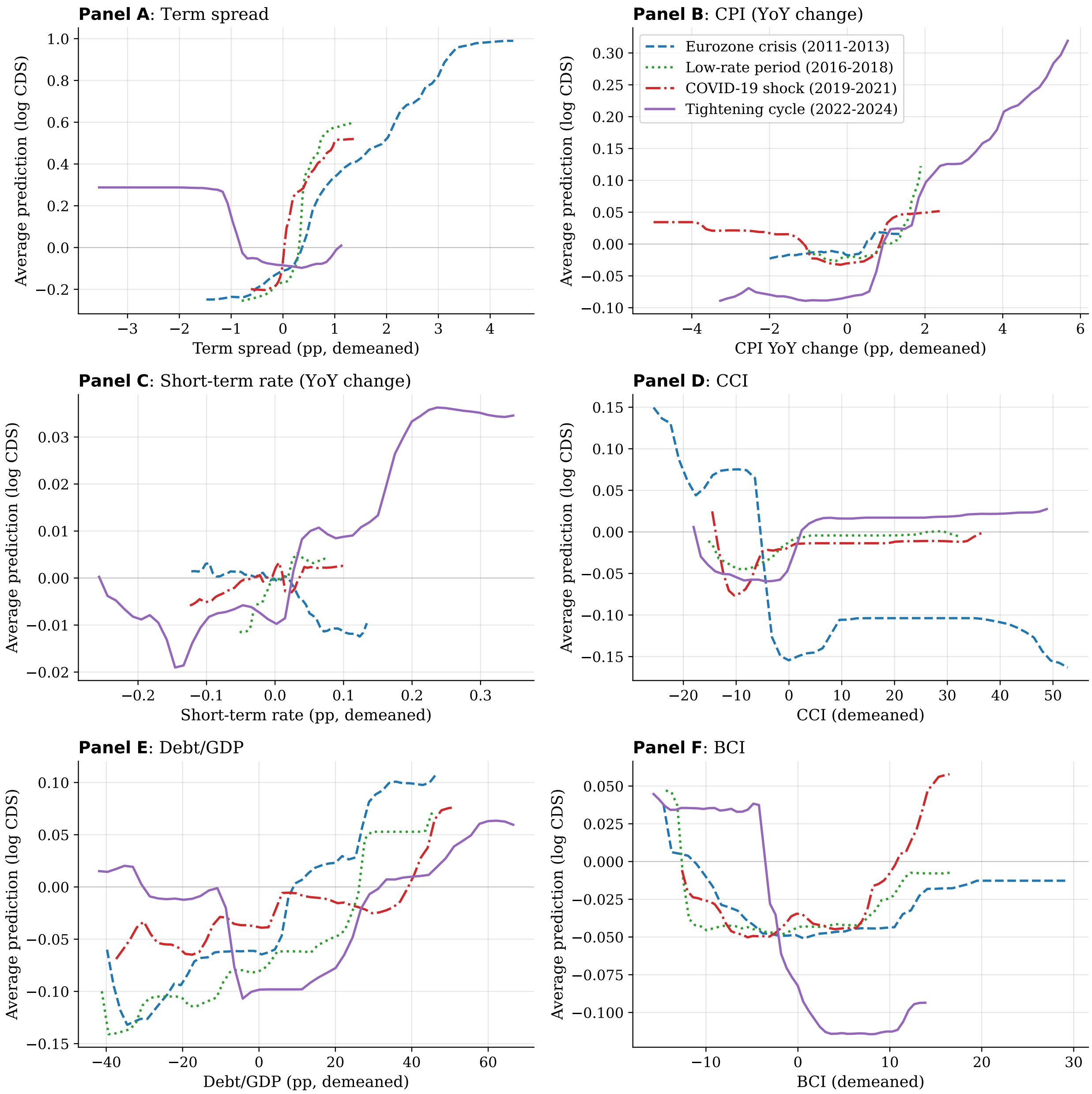


Figure 7

Partial dependence plots.

Regime-averaged partial dependence functions from XGBoost for Spread (Panel A), CPI (Panel B), INT (Panel C), CCI (Panel D), Debt/GDP (Panel E), and BCI (Panel F). Regime windows: Eurozone crisis (2011-04 to 2013-03), low-rate period (2016-04 to 2018-03), COVID-19 (2019-05 to 2021-04), tightening cycle (2022-07 to 2024-06). Horizontal axes are demeaned relative to the cross-sectional mean.



Supplementary appendix for:

Macroeconomic Fundamentals and the Shape of Sovereign Credit Risk

This appendix provides auxiliary theoretical results (Section A), data documentation and descriptive statistics (Section B), and additional empirical results (Section C). Implementation details for all forecasting methods are collected in the Internet Appendix.

A. Auxiliary Theoretical Results

This section provides the explicit steps connecting the no-arbitrage condition $PL(S^*_{j,t}; t, t+j) = DL(S^*_{j,t}; t, t+j)$ in Section II to the pricing expression that precedes Equation (3). Our derivation follows the standard reduced-form setting of Duffie (1999), Schönbucher (2003), and the exposition in Augustin and Tédongap (2016).

We define the premium and default legs of a CDS contract with maturity j as

$$PL(S^*_{j,t}; t, t+j) = \mathbb{I}_{\{\tau>t\}} \int_t^{t+j} \mathbb{E}\left[S^*_{j,t}\, e^{-\int_t^u (r_s + g(\mathbf{x}_s))ds} \,\middle|\, \mathcal{F}_t \right] du, \tag{A.1}$$

$$DL(S^*_{j,t}; t, t+j) = L \cdot \mathbb{I}_{\{\tau>t\}} \int_t^{t+j} \mathbb{E}\left[e^{-\int_t^u (r_s + g(\mathbf{x}_s))ds}\, g(\mathbf{x}_u) \,\middle|\, \mathcal{F}_t \right] du, \tag{A.2}$$

where $L = 1 - R$ is the loss-given-default and all other notation follows Section II. Since the CDS premium $S^*_{j,t}$ is fixed at contract inception and does not depend on the integration variable u, we can factor it out of (A.1):

$$PL(S^*_{j,t}; t, t+j) = S^*_{j,t} \cdot \mathbb{I}_{\{\tau>t\}} \int_t^{t+j} \mathbb{E}\left[e^{-\int_t^u (r_s + g(\mathbf{x}_s))ds} \,\middle|\, \mathcal{F}_t \right] du. \tag{A.3}$$

Equating (A.3) and (A.2) and dividing through by the no-default indicator $\mathbb{I}_{\{\tau>t\}}$ (which is unity conditional on $\mathcal{F}_t$ prior to default) yields

$$S^*_{j,t} = (1 - R) \cdot \frac{\int_t^{t+j} \mathbb{E}_t\left[e^{-\int_t^u (r_s + g(\mathbf{x}_s))ds}\, g(\mathbf{x}_u)\right] du}{\int_t^{t+j} \mathbb{E}_t\left[e^{-\int_t^u (r_s + g(\mathbf{x}_s))ds}\right] du}, \tag{A.4}$$

where we use the shorthand $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot \mid \mathcal{F}_t]$ and make the dependence of the fair premium on the contract maturity j explicit through the subscript. Equation (A.4) is a valuation formula under reduced-form pricing with deterministic recovery and no counterparty risk. The CDS spread for maturity j equals the loss given default multiplied by the ratio of expected discounted intensities to expected discount factors, integrated over the contract life.

The ratio in equation (A.4) does not admit a closed-form expression without additional structure on the joint dynamics of $(r_s, g(\mathbf{x}_s))$. A standard simplification in the empirical CDS literature exploits the fact that sovereign CDS spreads in our sample are typically small: average 5-year spreads for the 25 countries over 2008–2024 lie below 150 basis points outside crisis episodes, and even during the European sovereign debt crisis remain below 5% for most of the panel. For $g(\mathbf{x}_u)j$ small, a first-order Taylor expansion of the exponential $e^{-\int_t^u g(\mathbf{x}_s)ds}$ around unity gives

$$e^{-\int_t^u g(\mathbf{x}_s)ds} \approx 1 - \int_t^u g(\mathbf{x}_s)ds,$$

and the joint discount factor factorizes to first order as $e^{-\int_t^u (r_s + g(\mathbf{x}_s))ds} \approx e^{-\int_t^u r_s ds} \cdot (1 - \int_t^u g(\mathbf{x}_s)ds)$. Under this approximation, both integrands in equation (A.4) are dominated by the pure risk-free discount factor, and the ratio simplifies. Taking the risk-free rate as approximately constant over the maturity window gives

$$S^*_{j,t} \approx (1 - R) \cdot \frac{1}{j} \int_t^{t+j} \mathbb{E}_t[g(\mathbf{x}_u)]\, du = (1 - R) \cdot \mathbb{E}_t\left[\frac{1}{j} \int_t^{t+j} g(\mathbf{x}_u)\, du\right], \tag{A.5}$$

which is the expression invoked before Equation (3) in Section II. The Fubini step in the second equality uses the interchangeability of integration and expectation, which is valid under standard regularity conditions on $g(\mathbf{x}_u)$ (see, e.g., Duffie, 1999).

Taking logs of equation (A.5),

$$s_{j,t} \equiv \log S^*_{j,t} \approx \log(1 - R) + \log \mathbb{E}_t \left[\frac{1}{j} \int_t^{t+j} g(\mathbf{x}_u)\, du \right]. \tag{A.6}$$

Equation (A.6) still depends on the entire intensity path over $[t, t + j]$. Because the instantaneous intensity $g(\mathbf{x}_u)$ is a function of the contemporaneous state, and because the state vector is adapted to $\mathcal{F}_t$ through its predictable component, we summarize the joint dependence on the current state and the maturity in a single functional form, writing $\log \mathbb{E}_t[\frac{1}{j} \int g(\mathbf{x}_u)du] = \log \gamma(\mathbf{x}_{t-1}, j)$, which yields equation (3) in Section II. This is where our empirical framework departs from the classical empirical CDS literature. Rather than imposing a log-linear specification on $\log \gamma(\mathbf{x}_{t-1}, j)$, we treat it as a flexible function of the state and maturity and estimate it non-parametrically.

Two observations about the small-spread approximation are worth stating explicitly. First, the approximation is tightest when spreads are small and worst when they are large. The approximation, therefore, is likely to attenuate the degree of non-linearity in our pricing framework relative to a fully non-linear reduced-form implementation; that is, if anything, the log-linear benchmark in Equation (4) understates the non-linearity the data would reveal under exact pricing. Second, the approximation does not affect the subsequent non-parametric treatment. Equation (3) and the agnostic specification $s_{j,t} \propto f(\mathbf{x}_{t-1}, j)$ are stated in terms of the unknown function $\gamma(\cdot, j)$, which Equation (A.6) derives from the underlying intensity process. Any further discrepancy between the first-order approximation and the exact expression is absorbed into the error between f and the true data-generating process that our empirical exercise estimates.

B. Data Description and Descriptives

This appendix describes the real-time macroeconomic dataset used in the main empirical analysis. Section A illustrates the real-time release structure of the OECD Main Economic Indicators (MEI) data. Section B sets out the release date and forward-filling conventions used to align the daily CDS target with the lower-frequency macroeconomic predictors. The correlation structure of the predictor block and panel stationarity tests on the log CDS target are reported in the Internet Appendix.

The empirical analysis covers daily sovereign CDS premiums for 3-, 5-, 7-, and 10-year contracts across 25 OECD economies from 15 January 2008 to 15 December 2024, paired with 23 country-specific and two global macroeconomic predictors. The panel of countries is composed of Austria, Belgium, Canada, Chile, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Ireland, Italy, Japan, Korea, Mexico, the Netherlands, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, the United Kingdom, and the United States.

A. Macroeconomic Indicators

The predictor set combines macroeconomic indicators from the "Revisions Analysis Dataset – Infra-annual Indicators" of the OECD Main Economic Indicators (MEI) dataset and sentiment indicators from the OECD Key Short-Term Economic Indicators dataset with financial variables from Bloomberg (such as the VIX), country-specific economic policy uncertainty indexes from Baker et al. (2016), and global factors including the VIX and the index of global economic activity (IGREA) of Kilian (2009), incorporating the correction discussed in Kilian (2019). Table A1 summarizes each variable, its description, the FRED-MD transformation code, and its category.

[Insert Table A1 here]

The OECD MEI dataset records every historical release of each macroeconomic indicator, allowing us to reconstruct the information set available to market participants on each date. Figure A1 illustrates the structure using the harmonized unemployment rate of Austria: each row indexes the reference period of the observation, while each column indexes the date on which the OECD published the series. The multiple columns per reference period reflect the OECD's standard practice of data revision. For our analysis, we retain only the first-release value of each observation — the entry in the first column of each row — yielding a simulated real-time dataset in which no observation contains information that was not publicly available on the date assigned to it. For instance, the first available value for the Austrian unemployment rate in February 2011 is 4.8%, published in April 2011; for March 2011, the first available value is 4.3%, published in May 2011.

[Insert Figure A1 here]

B. Handling Data Timeliness

The OECD MEI dataset records the publication month of each release but not the exact publication date. To assign a daily timestamp to each real-time value, we assume that a monthly indicator is released on the 15th of the following month and a quarterly indicator on the 15th of the first month following the reference quarter; if the 15th falls on a weekend or public holiday, we use the closest prior business day. A daily predictor series is then constructed by forward-filling; that is, each macroeconomic variable is held constant at its most recent release value until the next release.

Figure A2 illustrates the resulting alignment. The dashed red vertical line represents a specific observation date at which CDS premiums are available; for that date, the latest available release of each quarterly or monthly predictor is carried forward from its own release date, and the forward-filled predictor block is combined with daily financial variables (FX, VIX, stock indices) that require no alignment. Returning to the Austrian unemployment example in Figure A1, the April 2011 release value of 4.8% is assigned to April 15, 2011 and held constant through May 14, 2011, after which the May 2011 release value of 4.3% replaces it. The same convention is applied uniformly to all OECD MEI macroeconomic indicators.

[Insert Figure A2 here]

C. Additional Empirical Results

This section collects supplementary empirical results referenced in the main text. Section A documents the statistical inference procedures underlying Section C. Section B reports sub-group Shapley value decompositions. The yearly R^2 heatmaps for the sub-groups, the four-benchmark decomposition for level-only linear specifications, and the first-release versus vintage comparison are reported in the Internet Appendix.

A. Testing Forecasting Accuracy

All three tests reported in Section C use the 2009–2024 out-of-sample forecast errors of the thirteen candidate specifications. The Qu et al. (2023) test operates directly on the country-maturity-day loss differentials, as described below. The MCS and Romano–Wolf procedures operate on the $T \times K$ matrix $\mathcal{L}$ of monthly-aggregated squared losses, where $T = 192$ is the number of testing months, $K = 13$ is the number of models, and

$$\mathcal{L}_{t,k} = \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} \sum_{\tau \in \mathcal{T}_t} \left(s_{ij,\tau} - \widehat{s}^{(k)}_{ij,\tau}\right)^2,$$

with $\mathcal{T}_t$ the daily observations within testing month t, $\mathcal{I}$ the countries in the relevant group, and $\mathcal{J}$ the maturities. Summing within-month across the panel units absorbs cross-sectional dependence into the monthly total loss, so the bootstrap-based inference on $\mathcal{L}$ does not require explicit panel-dependence corrections.

Pairwise Diebold–Mariano test. We adopt a modified version of the pairwise test of Diebold and Mariano (1995) as proposed by Qu et al. (2023). Specifically, we define the squared error loss associated with a given forecast $\widehat{s}_{ij,t}$ for country i, maturity $j \in \mathcal{J}$, and testing period $t = 1, \ldots, T$ as $L_{ij,t} = (s_{ij,t} - \widehat{s}_{ij,t})^2$. We test the null hypothesis $\mathcal{H}_0 : E\left[\overline{\Delta L}\right] = 0$ where $\overline{\Delta L} \equiv \frac{1}{nT|\mathcal{J}|} \sum_{t=1}^{T} \sum_{j \in \mathcal{J}} \sum_{i=1}^{n} \Delta L_{ij,t}$ is the pooled average loss differential between two models.

The corresponding t-statistic is defined as,

$$t_s = \frac{1}{nT|\mathcal{J}|}\frac{\sum_{t=1}^{T}\sum_{j\in\mathcal{J}}\sum_{i=1}^{n}\Delta L_{ij,t}}{\widehat{\sigma}\left(\Delta L_{ij,t}\right)}, \tag{C.1}$$

where $\widehat{\sigma}\left(\Delta L_{ij,t}\right)$ is a consistent estimator for the aggregate pooled loss variance. The latter can be computed by a Newey and West (1987) estimator as,

$$\widehat{\sigma}\left(\Delta L_{ij,t}\right) = \sqrt{\sum_{l=-L}^{L}\left(1-\frac{|l|}{L}\right)\widehat{\rho}\left(l\right)}, \tag{C.2}$$

where $L > 0$ is the maximum lag length and $\widehat{\rho}\left(l\right) = \frac{1}{T}\sum_{t=|l|+1}^{T}\widetilde{R}_{t-|l|}\widetilde{R}_t$ with $\widetilde{R}_t = R_t - T^{-1}\sum_{s=1}^{T}R_s$ and $R_s = \left(n|\mathcal{J}|\right)^{1/2}\sum_{j\in\mathcal{J}}\sum_{i=1}^{n}\Delta L_{ij,s}$. For $l < 0$, we set $\widehat{\rho}\left(l\right) = \widehat{\rho}\left(-l\right)$. When T is large, $t_s \xrightarrow{d} N(0,1)$ (see Qu et al., 2023), so the usual rejection rules apply.

Model Confidence Set. The Model Confidence Set of Hansen et al. (2011) is computed using the range statistic T_R, which avoids the coherency failure documented in the corrigendum to Hansen (2003). Starting from the full set of K models, the procedure iteratively tests the null of equal predictive accuracy within the current set via a stationary block bootstrap with block length $\lceil\sqrt{T}\rceil = 14$ and $B = 5{,}000$ replications; when the null is rejected, the worst-performing model in the current set is eliminated. The procedure terminates when the null cannot be rejected at level α, and the surviving set is the α-level MCS. We report the 90% MCS ($\alpha = 0.10$) and the 75% MCS ($\alpha = 0.25$). Per-model MCS p-values are reported in Table A2. Implementation uses the `arch` Python package.

Table A2 reports per-model MCS p-values by country group at the one-month horizon. A model belongs to the α-level MCS if its p-value is at least α; under this rule, the 90% MCS ($\alpha = 0.10$) and the 75% MCS ($\alpha = 0.25$) both contain only XGBoost in every group. XGBoost's p-value is uniformly 1.000, indicating that it is the last model to survive the sequential-elimination procedure; every rival is rejected at or near the bootstrap resolution of $1/B = 0.0002$.

The strength of the MCS result contrasts with the Romano–Wolf output in the SWEAP sub-panel (Table 2, Panel B), where FWER control over twelve pairwise comparisons leaves ten of twelve rivals indistinguishable from XGBoost. The two procedures answer different questions: the MCS tests whether each model can be retained as among the best given the current active set (a sequential, elimination-driven hypothesis), while the Romano–Wolf adjustment tests all twelve pairwise comparisons simultaneously against the null of equal predictive accuracy. The SWEAP MCS remains a singleton because the elimination sequence rejects each non-XGB model at the step where it performs worst in the current set, well before the procedure terminates; the Romano–Wolf test, by contrast, has reduced power against all twelve hypotheses simultaneously in the smaller cross-section.

[Insert Table A2 here]

Romano–Wolf step-down. For a family of P pairwise hypotheses $\{H_p : E[d_p] = 0\}_{p=1}^{P}$ on the monthly-aggregated loss differentials $d_{p,t} = \mathcal{L}_{t,a} - \mathcal{L}_{t,b}$, we compute the Romano–Wolf (Romano and Wolf, 2005, 2016) step-down adjusted p-values by stationary block bootstrap with the same settings as the MCS. Bootstrap samples are drawn from the centered series $d_{p,t} - \bar{d}_p$; for each sample, we compute the studentized statistic $|t^*_{p,b}|$ for each hypothesis. Ordering the original hypotheses by $|t_p|$ descending, the adjusted p-value for hypothesis p at rank r is

$$p_p^{\text{RW}} = \max\left(\frac{1}{B} \sum_{b=1}^{B} \mathbf{1}\left\{ \max_{q \in \mathcal{A}_r} |t^*_{q,b}| \geq |t_p| \right\},\ p^{\text{RW}}_{(r-1)} \right),$$

where $\mathcal{A}_r$ is the set of hypotheses still active at rank r and the outer maximum enforces monotonicity of adjusted p-values along the elimination sequence. We report the twelve XGBoost-vs-rival pairwise comparisons (Panel B of Table 2). The twelve-hypothesis family is the natural test of the paper's headline claim; the 78-hypothesis family is the conservative multiplicity cut that the referee would expect under the strictest FWER control.

B. Sub-group Shapley values

This appendix reports the time-varying Shapley values of Figure 5 disaggregated into the three country sub-groups (SWEAP, EU ex-SWEAP, OECD non-EU) for both XGBoost and Elastic Net, shown in Figures A3–A5. Shapley values are computed from the full 25-country fits used in the main text and averaged within each country group; no model is retrained on a country subset. The three main-text patterns of Section B are preserved across all three sub-groups: (i) XGBoost loads persistently on Term Spread and secondarily on CCI, with CPI and INT activating in the 2021–2024 tightening cycle; (ii) Elastic Net loads persistently on Debt/GDP with CCI and BCI as secondary bands; (iii) the set of active predictors in each model is stable across sub-groups and disjoint between models. What varies across subgroups is the intensity and timing of regime activations, which, in turn, reflect each group's exposure to the dominant sovereign-risk episodes in the sample.

[Insert Figure A3 here]

SWEAP. Figure A3 shows the sharpest Term Spread concentration in the sample for XGBoost, with the band most intense during the Eurozone sovereign debt crisis (2011–2014) and through the subsequent low-rate period (2015–2020). This coincides with the period during which Spanish, Portuguese, Italian, and Irish CDS premiums were at their most volatile. The tightening-cycle CPI/INT activation is present but comparatively muted, which may reflect the fact that 2022–2024 inflation dynamics were less novel for countries already operating under elevated risk premiums. On the Elastic Net side, the Debt/GDP band is present but less intense than in the OECD non-EU, reflecting that SWEAP countries' Debt/GDP deviations from the panel mean, while elevated during the Eurozone crisis, are not the largest in the sample once Japan and Mexico are included.

[Insert Figure A4 here]

EU ex-SWEAP. Figure A4 is the closest sub-group match to the Global pattern of Figure 5, reflecting the fact that the EU ex-SWEAP countries (Austria, Belgium, Estonia, Finland, France, Germany, Netherlands, Slovak Republic, Slovenia) span the largest share of the panel and their CDS dynamics track the Global aggregate most closely. The XGBoost Term Spread band and the 2021–2024 CPI/INT activation are both visible; the Eurozone-crisis CCI flare is sharper for this group than for OECD non-EU, consistent with their proximity to the crisis epicenter through the EU institutional mechanism.

[Insert Figure A5 here]

OECD non-EU. Figure A5 shows the sharpest tightening-cycle CPI/INT activation in the sample, with both rows visibly bright through 2022–2024. This coincides with the group containing the economies most exposed to the 2022–2024 inflation shock (the United States, Canada, the United Kingdom, and the emerging-market sovereigns Chile, Mexico, and Hungary, whose inflation dynamics diverged sharply from the euro-area core during this period). The Elastic Net Debt/GDP band is also the most intense for this group. Both Panel A and Panel B of Figure A5 therefore load most sharply during the tightening cycle, though for different predictors — Term Spread, CPI, and INT in XGBoost, and Debt/GDP in Elastic Net — a within-regime echo of the contrast documented in the main text.

Table A1

Real-time macroeconomic indicators and financial variables.

This table summarizes the real-time macroeconomic indicators and global risk factors used as explanatory variables in the main empirical analysis. The data are sampled from the "Revisions Analysis Dataset – Infra-annual Indicators" of the OECD Main Economic Indicators, the OECD Key Short-Term Economic Indicators, the Bloomberg terminal, and `https://www.policyuncertainty.com`. The transformation code (`tcode`) refers to the FRED-MD convention of McCracken and Ng (2016), defined in Eq. (5): 1 for levels, 2 for first differences, 5 for log first differences, and 7 for percentage changes. Unless otherwise noted, differences and growth rates are computed year-over-year, with $p = 12$ for monthly series and $p = 4$ for quarterly series. † SP and FX enter as monthly ($p = 1$) percentage changes following the standard financial-literature convention for these series, rather than the year-over-year ($p = 12$) transformation applied to other monthly indicators.

Variable	**Description**	**`tcode`**	**Category**
BOP	Current Account Balance (% of GDP)	1	Fiscal
Debt/GDP	Debt to GDP Ratio (%)	1	Fiscal
INT	Short-Term Interest Rate (%)	2	Monetary
Spread	Spread Between Long- and Short-Term Interest Rates (%)	1	Monetary
CPI	Consumer Price Index	5	Prices
GDP	Total GDP	7	Output
PCE	Private Consumption Expenditure	7	Output
GCE	Government Consumption Expenditure	7	Output
GFC	Gross Fixed Capital Formation	7	Output
EGS	Exports of Goods and Services	7	Output
IGS	Imports of Goods and Services	7	Output
EGS/GDP	Exports of Goods and Services (% of GDP)	1	Output
IGS/GDP	Imports of Goods and Services (% of GDP)	1	Output
BCI	Business Confidence Indicator	1	Expectation
CCI	Consumer Confidence Indicator	1	Expectation
IIP	Index of Industrial Production	5	Activity
IRD	Index of Retail Trade Volume	5	Activity
HUR	Harmonised Unemployment Rate (%)	2	Labour Market
SP	Share Price Index	$7^{\dagger}$	Financial
FX	National Currency per USD	$7^{\dagger}$	Financial
VIX	CBOE Volatility Index	1	Global Risk
EPU	Economic Policy Uncertainty Index	5	Policy Uncertainty
IGREA	Index of Global Real Economic Activity	1	Business Cycle

Table A2

Model Confidence Set p-values, one-month testing period.

The table reports per-model MCS p-values computed on the $T \times K$ matrix of monthly-aggregated squared-error losses, by country group. Values are from the step-down range-statistic procedure of Hansen et al. (2011) with stationary block bootstrap, block length $\lceil \sqrt{T} \rceil = 14$ and $B = 5{,}000$ replications. A model is in the α-level MCS if $p \geq \alpha$; values above 0.10 are in the 90% MCS, values above 0.25 in the 75% MCS. Entries of 0.000 are at or below the bootstrap floor $1/B = 0.0002$.

Model	Global	OECD non-EU	EU ex-SWEAP	SWEAP
OLS	0.000	0.000	0.000	0.001
EWC	0.000	0.000	0.000	0.000
CSR	0.000	0.000	0.000	0.000
Lasso	0.000	0.000	0.000	0.001
Ridge	0.000	0.000	0.000	0.000
EN	0.000	0.000	0.000	0.000
PCR	0.000	0.000	0.000	0.000
PLS	0.000	0.000	0.000	0.001
SVR	0.000	0.000	0.003	0.002
RT	0.000	0.000	0.000	0.002
RF	0.000	0.000	0.000	0.002
XGB	**1.000**	**1.000**	**1.000**	**1.000**
NN	0.000	0.000	0.000	0.000

Figure A1

Example of data structure for Austria.

The figure illustrates the organization of the monthly HUR data in the OECD database. Each row corresponds to the reference period of the observation, and each column to the publication date. Multiple columns per reference period reflect the OECD's practice of data revision.

Country	Austria																	
Variable	Harmonised unemployment rate																	
Frequency	Monthly																	
Unit	Percentage																	
Edition	April 2011	May 2011	June 2011	July 2011	August 2011	September 2011	October 2011	November 2011	December 2011	January 2012	February 2012	March 2012	April 2012	May 2012	June 2012	July 2012	August 2012	September 2012
Time																		
Feb-2011	4.8	4.7	4.6	4.6	4.6	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.4	4.4	4.4	4.4	4.4	4.5
Mar-2011	..	4.3	4.4	4.4	4.3	4.3	4.3	4.3	4.3	4.3	4.3	4.3	4.2	4.2	4.2	4.2	4.3	4.3
Apr-2011	..	..	4.2	4.2	4.2	4.1	4.1	4.1	4.1	4.1	4.1	4.2	4.1	4.2	4.2	4.2	4.2	4.2
May-2011	..	..	..	4.3	4.2	4.2	4.2	4.2	4.2	4.2	4.2	4.2	4.2	4.2	4.2	4.2	4.2	4.2
Jun-2011	..	..	..	..	4.0	3.9	3.9	3.9	3.9	3.9	3.9	4.0	4.0	4.0	4.0	4.0	3.9	3.9
Jul-2011	..	..	..	..	..	3.7	3.7	3.7	3.7	3.7	3.7	3.8	3.8	3.8	3.8	3.8	3.7	3.8
Aug-2011	..	..	..	..	..	..	3.7	3.7	3.7	3.7	3.7	3.8	3.8	3.8	3.8	3.8	3.7	3.8

Figure A2

An illustration of the data alignment.

The figure illustrates the alignment of daily CDS premiums with the most recent available releases of quarterly and monthly macroeconomic indicators. Release dates follow the 15th-of-the-following-month convention, with forward-filling between consecutive releases.

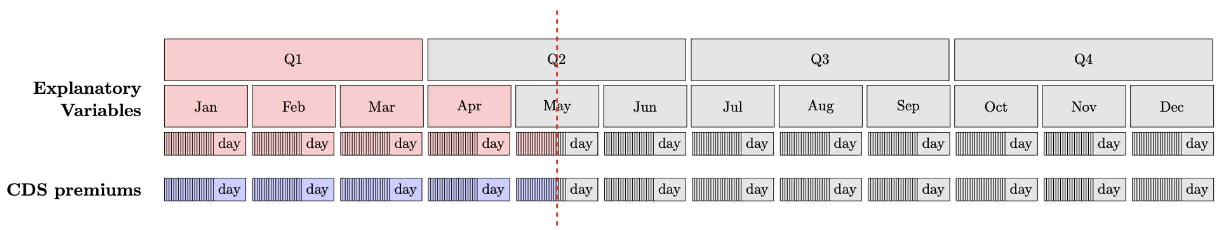


Figure A3

Time-varying importance of macroeconomic variables: SWEAP.

Out-of-sample mean absolute Shapley values from XGBoost (Panel A) and the "level+slope" Elastic Net (Panel B), averaged across SWEAP countries (Spain, Portugal, Italy, Ireland) at each one-month testing period. Shapley values are computed from the full 25-country XGBoost and Elastic Net fits of Section IV and averaged over countries in the sub-group; no model is retrained. VIX and IGREA are omitted as in Figure 5. The out-of-sample period is January 2009 through November 2024.

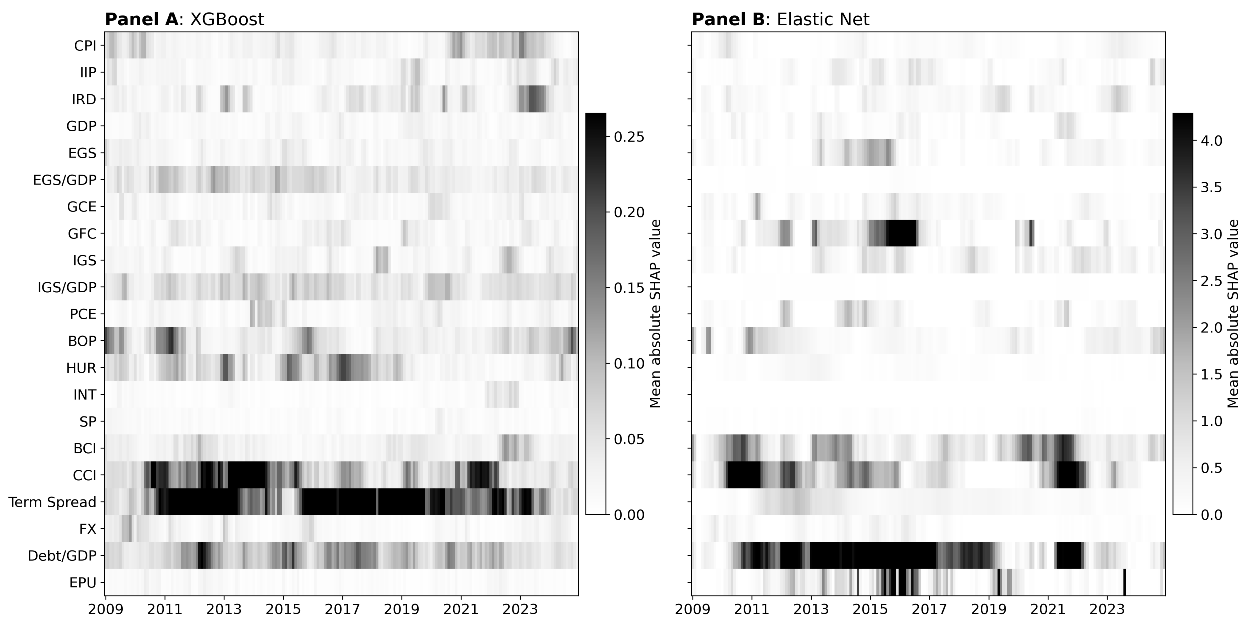


Figure A4

Time-varying importance of macroeconomic variables: EU ex-SWEAP.

Out-of-sample mean absolute Shapley values from XGBoost (Panel A) and the "level+slope" Elastic Net (Panel B), averaged across EU ex-SWEAP countries (Austria, Belgium, Estonia, Finland, France, Germany, Netherlands, Slovak Republic, Slovenia) at each one-month testing period. Shapley values are computed from the full 25-country XGBoost and Elastic Net fits of Section IV and averaged over countries in the sub-group; no model is retrained. VIX and IGREA are omitted as in Figure 5. The out-of-sample period is February 2009 through November 2024.

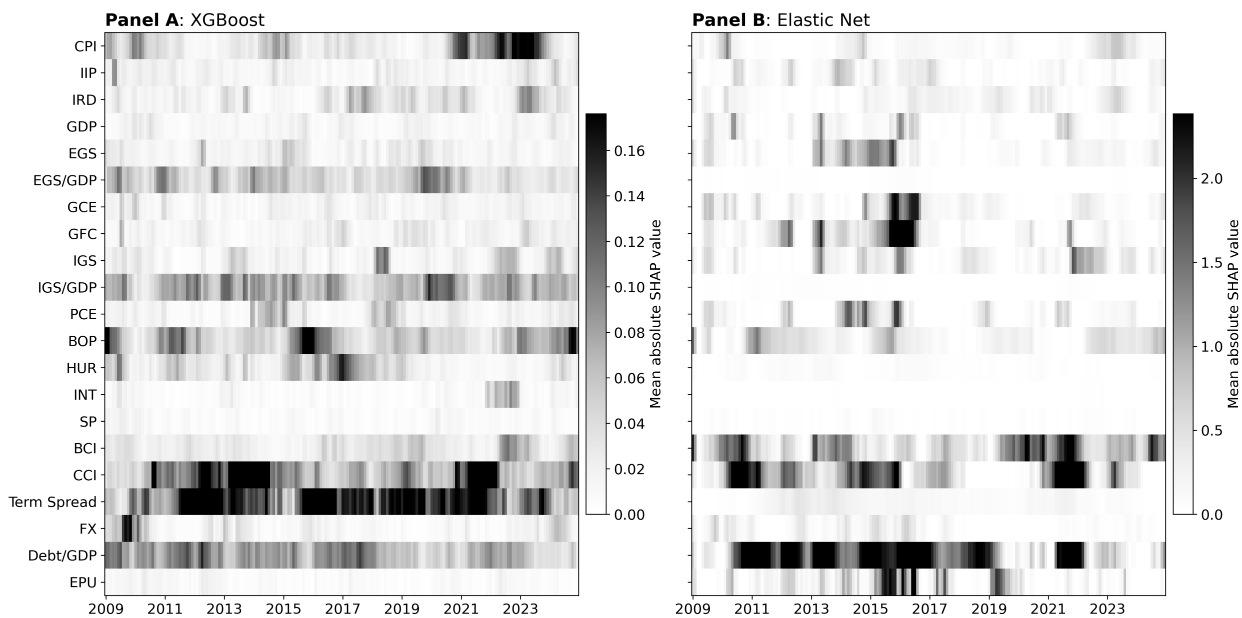
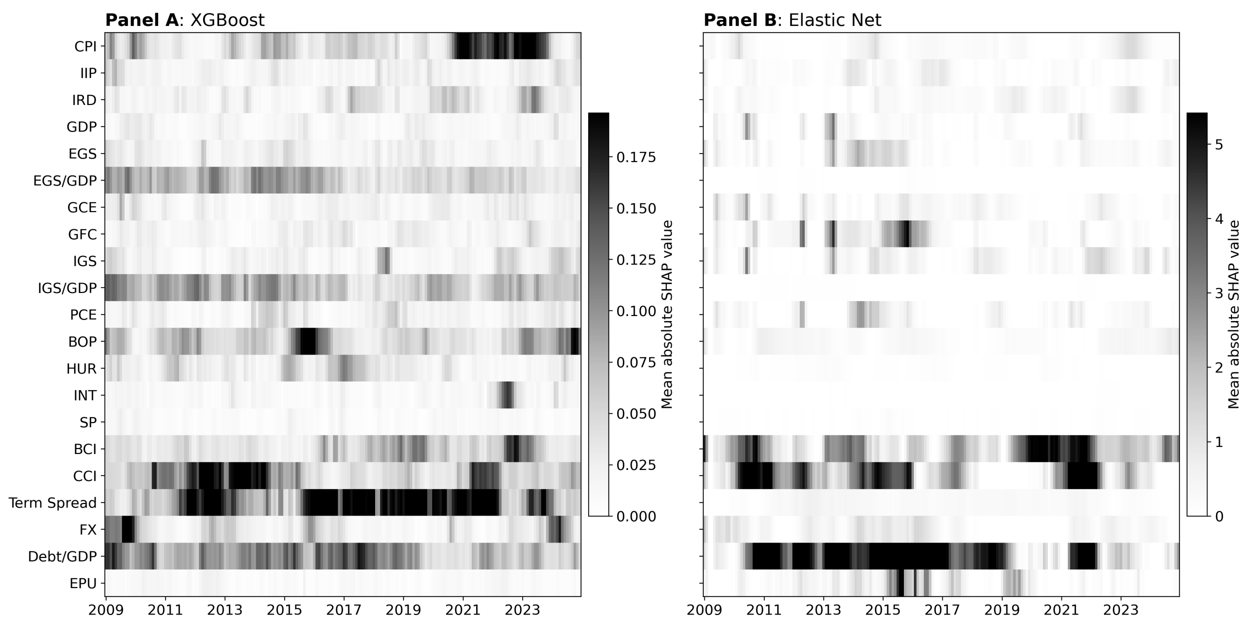


Figure A5

Time-varying importance of macroeconomic variables: OECD non-EU.

Out-of-sample mean absolute Shapley values from XGBoost (Panel A) and the "level+slope" Elastic Net (Panel B), averaged across OECD non-EU countries (Canada, Chile, Czech Republic, Denmark, Hungary, Japan, Korea, Mexico, Poland, Sweden, United Kingdom, United States) at each one-month testing period. Shapley values are computed from the full 25-country XGBoost and Elastic Net fits of Section IV and averaged over countries in the sub-group; no model is retrained. VIX and IGREA are omitted as in Figure 5. The out-of-sample period is February 2009 through November 2024.



Internet Appendix for:

Macroeconomic Fundamentals and the Shape of Sovereign Credit Risk

This Internet Appendix collects supplementary material not included in the published manuscript: implementation details for all forecasting methods (Section A), additional data description (Section B), and additional empirical results (Section C). Cross-references to the main text resolve against the published manuscript.

A Estimation Details

This appendix provides implementation details for the forecasting methods described in Section III.B of the main paper. Sections A.1–A.3 cover the eight linear specifications, Sections A.4–A.6 the five nonlinear specifications, and Section A.7 the time-series cross-validation procedure used to select hyperparameters for every method. Throughout, we use the panel notation established in the main paper: $s_{ij,t}$ for the log CDS spread of country i at maturity j and date t, and $\mathbf{x}_{i,t-1} \in \mathbb{R}^P$ for the real-time predictor vector. Hyperparameters reported in each subsection are tuned over the 12-month training window using the time-series cross-validation procedure in Section III.B of the main paper, with the final estimates frozen before testing.

A.1 Penalized Regressions

We follow the approach outlined by Balduzzi et al. (2023) and estimate a "level+slope" linear regression specification in which the real-time macroeconomic indicators $\mathbf{x}_{i,t-1}$ (plus an intercept) interact with the CDS maturity j of the respective CDS contract for country i,

$$s_{ij,t} = \mathbf{x}_{i,t-1}'\beta + \sum_{j \in \mathcal{J} \setminus \{3\}} (\mathbf{x}_{i,t-1} \cdot m_j)' \delta_j + \varepsilon_{ij,t}, \tag{1}$$

where $s_{ij,t}$ denotes the logarithm of the CDS spread for country i, maturity j, at time t, m_j is a dummy variable that takes value one if the contract has j-year maturity and zero otherwise, and $\mathcal{J} = \{3, 5, 7, 10\}$. The reference group is the 3-year CDS spread. In the main empirical analysis, we start from Eq. (1) and estimate penalized variants, such as Ridge regression (Hsiang, 1975), Lasso (Tibshirani, 1996), and Elastic Net (Zou and Hastie, 2005).

To prevent the maturity-dummy coefficients from being penalized, we subtract the cross-country mean at each time and maturity before estimation, as discussed in Section III.B:

$$s_{ij,t} - \bar{s}_{\cdot j,t} = (\mathbf{x}_{i,t-1} - \bar{\mathbf{x}}_{\cdot,t-1})'\beta + \sum_{j \in \mathcal{J} \setminus \{3\}} \left((\mathbf{x}_{i,t-1} - \bar{\mathbf{x}}_{\cdot,t-1}) \cdot m_j \right)' \delta_j + \varepsilon'_{ij,t}, \tag{2}$$

where $\bar{s}_{\cdot j,t} = \frac{1}{N_t(j)} \sum_{i=1}^{N_t(j)} s_{ij,t}$ is the cross-country average of the log CDS spread at maturity j and time t, with $N_t(j)$ the number of countries for which maturity-j data are available at t, and $\bar{\mathbf{x}}_{\cdot,t-1} = \frac{1}{N_t(j)} \sum_{i=1}^{N_t(j)} \mathbf{x}_{i,t-1}$ the corresponding predictor average.[1] The demeaned predictions $\widehat{s}^{d}_{ij,t}$ are re-centered at the forecast stage by adding back the maturity-specific training-window mean $\bar{s}^{(\text{train})}_{\cdot j,\cdot}$ computed across i and t within the training window, yielding $\widehat{s}_{ij,t} = \widehat{s}^{d}_{ij,t} + \bar{s}^{(\text{train})}_{\cdot j,\cdot}$.

For Lasso, we search for the optimal L1 penalty α over $[10^{-2}, 10^{3}]$; for Ridge, we search for the optimal L2 penalty λ over $[10^{-2}, 10^{4}]$. In both cases, we select 20 values from the respective ranges using logarithmic spacing. For Elastic Net, we search jointly over (α, λ) using the same ranges. All penalized regressions are estimated on the demeaned inputs without further standardization, since the FRED-MD transformations in Section III.A render the predictors approximately scale-comparable.

For robustness, we also estimate a "level-only" variant of Eq. (2) in which the maturity-interaction terms are omitted. The level-only specification fits a single coefficient vector β across all maturities. Results are reported in Table C.1 in Appendix C.2.

A.2 Latent Factor Models

For the latent-factor specifications, we apply PCA or PLS to the level-only specification, then interact the resulting components with the CDS maturity dummies. The components serve as a condensed representation of the original predictor set, simultaneously achieving dimension reduction and preserving the level+slope structure of Eq. (1). The number of retained components is selected by time-series cross-validation; we fix $K = 5$ throughout the empirical analysis, as this value is selected in the majority of rolling windows.

[1] After cross-country demeaning, the maturity-dummy coefficients are orthogonal to the demeaned intercept component, so penalizing the slope and interaction coefficients does not shrink them.

A.3 Complete Subset Regressions

Complete Subset Regression (CSR) uses equal-weighted combinations of forecasts from all p-variable subsets, where $p \leq P$ and P is the total number of predictors. For a given subset size p, we run the regression of $s_{ij,t}$ on each p-dimensional subset of $\mathbf{x}_{i,t-1}$ and average the resulting coefficient vectors across subsets to obtain the CSR estimator $\widehat{\beta}^{\text{CSR}}$.

As an illustration, the univariate case $p = 1$ has P subset regressions, one per predictor. The CSR forecast is the equal-weighted average $\widehat{s}_{ij,\tau}^{\text{CSR}} = \frac{1}{P}\sum_{k=1}^{P} \mathbf{x}'_{i,\tau-1}\widehat{\beta}_k$, where $\widehat{\beta}_k$ is the P-vector with the least-squares estimate on the k-th predictor in the k-th position and zeros elsewhere. The equal-weighted forecast combination (EWC) is the special case $p = 1$ of CSR.

In our empirical analysis, we treat the subset size p as a hyperparameter. When the number of p-variable subsets exceeds 5,000, we estimate 5,000 subsets drawn without replacement from the full set; otherwise, we estimate every possible subset.

A.4 Tree Ensemble and Regression Trees

At iteration k, XGBoost updates the prediction of the log CDS spread as $\widehat{s}_{ij,t}^{(k)} = \widehat{s}_{ij,t}^{(k-1)} + f_k(\mathbf{x}_{i,t-1})$, with each new tree f_k trained to minimize the regularized objective

$$\mathcal{L}^{(k)} = \sum_{(i,j,t)} \ell\Big(s_{ij,t},\, \widehat{s}_{ij,t}^{(k-1)} + f_k(\mathbf{x}_{i,t-1})\Big) + \Omega(f_k), \qquad \Omega(f_k) = \eta\, T_k + \tfrac{1}{2}\lambda \sum_{l=1}^{T_k} w_l^2, \tag{3}$$

where T_k is the number of leaves in tree k, w_l the weight of leaf l, and (η, λ) control the leaf-count penalty and the L2 shrinkage of leaf weights.[2] Approximating the loss around $\widehat{s}_{ij,t}^{(k-1)}$ by a second-order Taylor expansion and writing I_l for the set of observations assigned to leaf l, the optimal leaf weight and the tree-quality score are

$$w_l^\star = -\frac{\sum_{(i,j,t)\in I_l} g_{ij,t}}{\sum_{(i,j,t)\in I_l} h_{ij,t} + \lambda}, \qquad \widetilde{\mathcal{L}}^{(k)}(q) = -\tfrac{1}{2}\sum_{l=1}^{T_k} \frac{\big(\sum_{(i,j,t)\in I_l} g_{ij,t}\big)^2}{\sum_{(i,j,t)\in I_l} h_{ij,t} + \lambda} + \eta\, T_k, \tag{4}$$

with $g_{ij,t}$ and $h_{ij,t}$ the gradient and Hessian of the loss ℓ at $\widehat{s}_{ij,t}^{(k-1)}$. The sparsity-aware split-finding algorithm of Chen and Guestrin (2016) makes training scalable even when predictors are missing or sparse.

For the XGBoost model, we tune the `number of estimators` (100, 150, 200), the `learning rate` (from 0.05 to 0.30 in increments of 0.05), the `maximum depth of trees` (3, 4, 5), the

[2] The symbol η in Eq. (3) is the XGBoost leaf-count penalty.

leaf-count penalty `gamma` (from 0.0 to 0.4 in increments of 0.1), the `subsample ratio of training instances` (from 0.5 to 0.9 in increments of 0.1), and the `subsample ratio of columns per tree` (0.3, 0.4, 0.5, 0.7).

For the regression tree, we cross-validate `max_depth` (tree depth limited to 3, 5, 7, or 9 levels), `min_samples_split` (minimum of 100, 200, 500, or 1500 observations required to allow a node split), and `min_samples_leaf` (each leaf node must contain at least 500, 1500, or 2000 observations).

For the random forest, we cross-validate the `number of trees` (10, 50, 100, 500), the `maximum depth of trees` (5, 6, 7), the `minimum number of observations` required to split an internal node (100, 200, 500), and the `maximum number of features` evaluated at each split (3, 4, 5).

A.5 Support Vector Regression

We estimate nonlinear forecasts with an ϵ-support vector regression (SVR) model (e.g., Cortes and Vapnik, 1995; Drucker et al., 1996) using a radial basis function (RBF) kernel. To maintain consistency with the linear specifications, we train SVR on cross-sectionally demeaned inputs and targets and recover final forecasts by adding back the maturity-specific training-window mean, following the same procedure as for the penalized regressions.

For country i, maturity $j \in \{3, 5, 7, 10\}$, and time t, let $s_{ij,t}$ be the log CDS spread and $\mathbf{x}_{i,t-1} \in \mathbb{R}^P$ the vector of real-time macroeconomic predictors. We form demeaned pairs by subtracting the cross-country average at the same time and maturity:

$$s^d_{ij,t} = s_{ij,t} - \bar{s}_{\cdot j,t}, \qquad \mathbf{x}^d_{i,t-1} = \mathbf{x}_{i,t-1} - \bar{\mathbf{x}}_{\cdot,t-1}, \tag{5}$$

where $\bar{s}_{\cdot j,t}$ and $\bar{\mathbf{x}}_{\cdot,t-1}$ are defined as in Eq. (2). We recover the final forecasts by adding back the maturity mean computed within the training window:

$$\widehat{s}_{ij,t} = \widehat{s}^d_{ij,t} + \bar{s}^{(\text{train})}_{\cdot j,\cdot}, \qquad \bar{s}^{(\text{train})}_{\cdot j,\cdot} := \frac{1}{|\mathcal{T}_{\text{train}}|} \sum_{t \in \mathcal{T}_{\text{train}}} \bar{s}_{\cdot j,t}. \tag{6}$$

Before fitting SVR, we standardize each predictor to zero mean and unit variance using the training window:

$$\tilde{\mathbf{x}}^d_{i,t-1} = \operatorname{diag}(\widehat{\boldsymbol{\sigma}})^{-1}\big(\mathbf{x}^d_{i,t-1} - \widehat{\boldsymbol{\mu}}\big), \tag{7}$$

with $\widehat{\boldsymbol{\mu}} \in \mathbb{R}^P$ and $\widehat{\boldsymbol{\sigma}} \in \mathbb{R}^P_+$ the per-predictor mean and standard deviation. The standardized design $\tilde{\mathbf{x}}^d_{i,t-1}$ enters SVR as the feature vector.

SVR with RBF Kernel. Let $\{(\tilde{\mathbf{x}}_n, y_n)\}_{n=1}^N$ denote the standardized, demeaned training pairs, with $y_n = s_{ij,t}^d$ and $\tilde{\mathbf{x}}_n = \tilde{\mathbf{x}}_{i,t-1}^d$. We use ϵ-SVR with RBF kernel

$$K_\kappa(u, v) = \exp\big(-\kappa \, \|u - v\|_2^2\big), \tag{8}$$

where κ is the kernel bandwidth. The SVR hyperparameter triple is (C, ϵ, κ). The primal problem is

$$\begin{aligned} \min_{w,b,\{\xi_n,\xi_n^\star\}} \quad & \tfrac{1}{2}\|w\|_2^2 + C \sum_{n=1}^N (\xi_n + \xi_n^\star) \\ \text{s.t.} \quad & y_n - (w^\top \phi(\tilde{\mathbf{x}}_n) + b) \leq \epsilon + \xi_n^\star, \\ & (w^\top \phi(\tilde{\mathbf{x}}_n) + b) - y_n \leq \epsilon + \xi_n, \\ & \xi_n, \xi_n^\star \geq 0 \qquad (n = 1, \ldots, N), \end{aligned} \tag{9}$$

where ϕ is the implicit feature map of K_κ. Its dual maximizes

$$\max_{\{\alpha_n,\alpha_n^\star\}} -\epsilon \sum_{n=1}^N (\alpha_n + \alpha_n^\star) + \sum_{n=1}^N y_n(\alpha_n^\star - \alpha_n) - \tfrac{1}{2} \sum_{n,m=1}^N (\alpha_n^\star - \alpha_n)(\alpha_m^\star - \alpha_m) \, K_\kappa(\tilde{\mathbf{x}}_n, \tilde{\mathbf{x}}_m), \tag{10}$$

subject to $0 \leq \alpha_n, \alpha_n^\star \leq C$ and $\sum_{n=1}^N (\alpha_n^\star - \alpha_n) = 0$. The resulting predictor is the standard support-vector expansion

$$\widehat{s}^d(\tilde{\mathbf{x}}) = \sum_{n=1}^N (\alpha_n^\star - \alpha_n) \, K_\kappa(\tilde{\mathbf{x}}_n, \tilde{\mathbf{x}}) + b, \tag{11}$$

where only the support vectors (observations with $\alpha_n^\star - \alpha_n \neq 0$) contribute.

For the SVR model, we tune the `regularization parameter` C (0.01, 0.1, 1, 10), the `epsilon-insensitive tube width` ϵ (0.01, 0.1, 0.2, 0.5), and the `RBF kernel bandwidth` κ (`scale`, 10^{-3}, 10^{-2}). The `scale` option corresponds to $\kappa = 1/(P \cdot \mathrm{Var}(\tilde{\mathbf{x}}))$, which simplifies to $\kappa = 1/P$ after standardization. The selected hyperparameter combination $(\widehat{C}, \widehat{\epsilon}, \widehat{\kappa})$ is applied within each expanding training window to generate demeaned forecasts $\widehat{s}_{ij,t}^d$ via Eq. (11), which are re-centered using Eq. (6).

A.6 Feedforward Neural Network

We implement a feedforward neural network (NN) following the architecture of Gu et al. (2020), adapted to our demeaned panel setting. As with SVR, the NN is trained on demeaned inputs and targets, and final forecasts are obtained by re-adding the maturity-specific training-window means to maintain consistency with the linear specifications.

The NN is a fully connected feedforward network with four hidden layers of sizes 32, 16, 8, and 4, each with `ReLU` activations and `he_normal` weight initialization. To improve numerical stability when training on demeaned inputs, each hidden layer is followed by `batch normalization`. To mitigate overfitting, we apply `dropout` regularization with rate 0.50 in the first three hidden layers and 0.25 in the final hidden layer. The output layer produces a single demeaned prediction through a linear activation.

All inputs are standardized to zero mean and unit variance using a `StandardScaler`. The network is trained with the `Adam` optimizer at `learning rate` 10^{-3} and `batch size` 32, for up to 100 `epochs` with `early stopping` on the validation loss and a patience of 20 epochs. A `ReduceLROnPlateau` scheduler halves the learning rate whenever the validation loss stagnates, with a floor of 10^{-5}. Demeaned out-of-sample predictions $\widehat{s}^{d}_{ij,t}$ are re-centered using the same maturity-mean formula as for SVR (Eq. (6)).

A.7 Cross-Validation Strategy

To cross-validate the hyperparameters of each model, we use a time-series-split approach that preserves the temporal ordering of the data and prevents look-ahead bias (e.g., Gu et al., 2020; Bianchi et al., 2021). Figure A.1 illustrates the procedure.

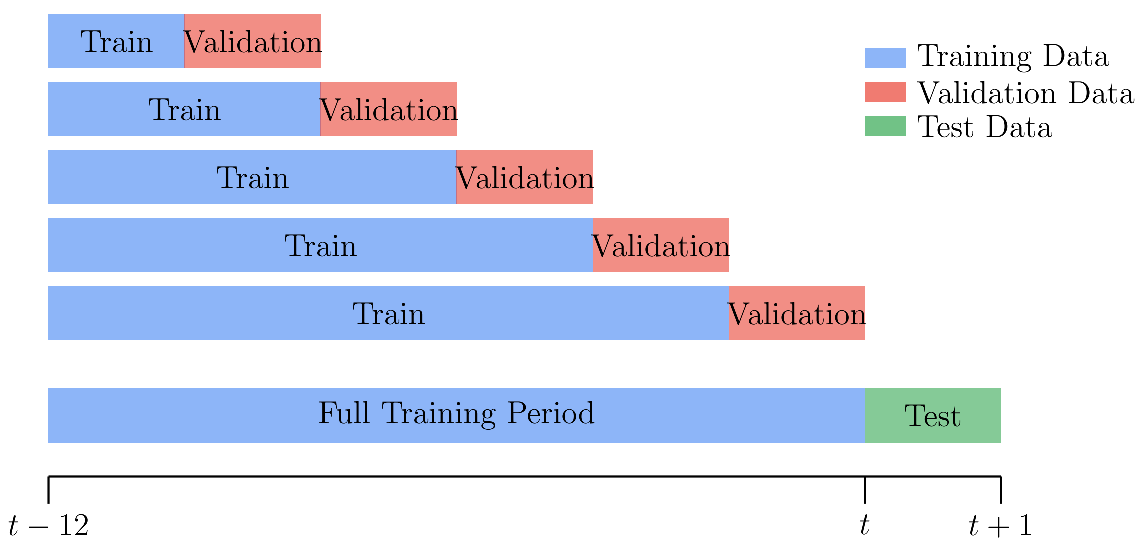


Figure A.1: **Recursive time-series cross-validation.** The figure illustrates the time-series cross-validation strategy with five sequential, non-overlapping validation folds and a final test set. The procedure is used to train each model's hyperparameters, such as penalty terms in penalized regressions or learning rates in XGBoost.

For each fold, the model is trained on the earlier block and validated on the subsequent block. This preserves the temporal dependence structure of the data and ensures that validation errors reflect each model's ability to generalize to subsequent, unseen observations. Hyperparameters are selected to minimize the average validation mean squared error across the five folds. Once selected, they remain fixed throughout the testing window, as described in Section III.B.

B Supplementary Data Description

B.1 CDS Contract Specifications

Table B.1 reports the restructuring clause, currency denomination, and earliest available date for each country-maturity pair. All 25 sovereigns trade under the Complete Restructuring (CR) clause, except Canada, which uses the No Restructuring (XR) clause. All contracts are denominated in US dollars, except the United States, whose sovereign CDS trades in euros. The panel is unbalanced: five countries (Belgium, Denmark, Germany, Italy, and the Netherlands) are missing one or two maturities, and several countries enter the sample after the panel start date of 15 January 2008. The most notable late entry is Canada, which becomes available only in November 2017.

B.2 Correlation Structure

Panel A of Figure B.1 reports the per-country correlation between each macroeconomic variable and the 5-year CDS spread. Three patterns stand out. First, business and consumer confidence indicators (BCI, CCI) and consumer price inflation (CPI) are negatively correlated with CDS spreads across most countries in the panel, consistent with sentiment and activity-slowdown signals co-moving with rising credit risk. Second, the interest-rate term-structure slope (Spread) and the debt-to-GDP ratio (Debt/GDP) are positively correlated with CDS spreads, with the strongest positive values concentrated in SWEAP countries and advanced emerging markets — consistent with these variables carrying elevated sovereign-risk content when debt sustainability is a concern. Third, the magnitude and even the sign of the correlations vary substantially across countries for most variables, including the real-activity indicators (GDP, PCE, GCE, GFC) and the global financial factors (VIX, IGREA). This advocates for quite heterogeneous country-level relationships.

Panel B reports the average pairwise correlation among the 23 country-specific macroeconomic variables (excluding VIX and IGREA, which are globally uniform). Real-activity variables (GDP, EGS, GCE, GFC, IGS, PCE) form a dense block of positive correlations, reflecting the common business-cycle component. Confidence indicators (BCI, CCI) and CPI

Country	Clause	CCY	3Y	5Y	7Y	10Y
SWEAP						
Ireland	CR	USD	2008-01	2008-01	2008-01	2008-01
Italy	CR	USD	2004-10	2004-10	—	2004-10
Portugal	CR	USD	2006-03	2004-10	2004-10	2004-10
Spain	CR	USD	2005-07	2004-10	2004-10	2004-10
EU ex-SWEAP						
Austria	CR	USD	2007-05	2004-10	2004-10	2004-10
Belgium	CR	USD	—	2004-10	2004-10	2004-10
Estonia	CR	USD	2006-10	2006-02	2006-02	2006-02
Finland	CR	USD	2007-08	2006-02	2006-02	2006-02
France	CR	USD	2007-05	2005-08	2005-08	2005-08
Germany	CR	USD	2007-05	2005-01	—	2004-10
Netherlands	CR	USD	—	2005-09	—	2005-09
Slovak Republic	CR	USD	2004-10	2004-10	2004-10	2004-10
Slovenia	CR	USD	2004-11	2004-11	2010-09†	2006-01
OECD non-EU						
Canada	XR	USD	2017-11†	2017-11†	2017-11†	2017-11†
Chile	CR	USD	2004-10	2004-10	2004-10	2004-10
Czech Republic	CR	USD	2009-07†	2009-04†	2009-04†	2009-04†
Denmark	CR	USD	—	2009-06†	—	2009-06†
Hungary	CR	USD	2004-10	2004-10	2008-07†	2004-10
Japan	CR	USD	2007-07	2004-10	2004-10	2004-10
Korea	CR	USD	2004-10	2004-10	2004-10	2004-10
Mexico	CR	USD	2004-10	2004-10	2004-10	2004-10
Poland	CR	USD	2004-10	2004-10	2004-10	2004-10
Sweden	CR	USD	2007-11	2006-02	2006-02	2006-02
United Kingdom	CR	USD	2007-11	2006-04	2007-11	2007-11
United States	CR	EUR	2007-12	2007-12	2007-12	2007-12

Table B.1: **CDS contract details.** The table reports the restructuring clause, currency denomination, and earliest available date (year-month) for each country-maturity pair. CR = Complete Restructuring; XR = No Restructuring. "—" indicates that the maturity is not available in Bloomberg for the respective sovereign. Entries marked † indicate country-maturity pairs that enter the sample after the panel start date of 15 February 2008; the panel is therefore unbalanced along both the country and maturity dimensions. Source: Bloomberg.

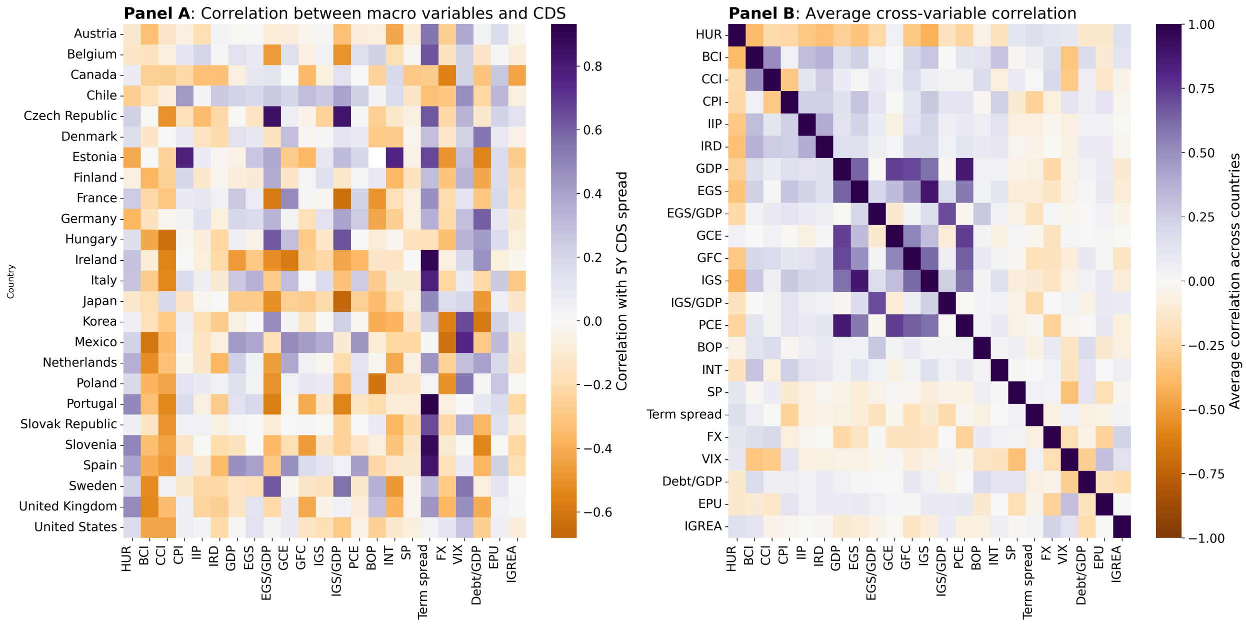


Figure B.1: **Sample correlation.** The figure reports the sample correlation between macroeconomic variables and CDS premiums per country (left panel) and the correlation among macroeconomic variables averaged across countries (right panel). The sample period is daily from 15 January 2008 to 15 December 2024.

form a second, weaker cluster. Global financial and policy-uncertainty variables (Spread, FX, VIX, Debt/GDP, EPU, IGREA) correlate weakly with each other and the real-activity block, indicating that these are distinct sources of variation rather than proxies for the same underlying factor.

B.3 Stationarity Tests

Table B.2 reports panel stationarity tests on daily log CDS spreads for each of the four contract maturities over the 2008–2024 sample. We implement two complementary tests that together bracket the order of integration. The Pesaran (Pesaran, 2007) CIPS test maintains the null of a panel unit root; it is robust to cross-sectional dependence, a first-order feature of sovereign CDS spreads, and invalidates panel tests that assume independence across countries. The Hadri (Hadri, 2000) LM test maintains the null of panel stationarity; it is the panel analog of the KPSS statistic (Kwiatkowski et al., 1992) and provides complementary evidence against the alternative of non-stationarity. We report both panel statistics and the number of countries for which the per-country test statistics reject their respective nulls at the 5% level as a disaggregated diagnostic.

The two tests reject in opposite directions at every maturity. CIPS rejects the panel unit-

Maturity	N	CIPS	# CADF rej.	Hadri Z	# KPSS rej.
3Y	22	-3.559**	19	129.117***	21
5Y	25	-3.374**	20	145.888***	24
7Y	21	-3.437**	17	125.530***	19
10Y	25	-3.711**	20	133.869***	25

Table B.2: **Panel stationarity tests for log CDS spreads.** The table reports panel unit-root and stationarity tests on daily log CDS spreads over 15 February 2008 to 15 December 2024. Column (2) gives the number of countries retained (at least 500 daily observations per country). Column (3) reports the CIPS statistic (Pesaran, 2007) for the null of a panel unit root, computed as the cross-country average of per-country CADF statistics with four augmentation lags. Column (4) reports the number of countries for which the per-country CADF statistic rejects the unit-root null at the 5% level. Column (5) reports the Hadri (Hadri, 2000) LM statistic for the null of panel stationarity, standardized to $N(0,1)$ under the null. Column (6) reports the number of countries for which the per-country KPSS statistic (Kwiatkowski et al., 1992) rejects the stationarity null at the 5% level. Significance: * 10%, ** 5%, *** 1%.

root null at the 5% level for all four maturities, with statistics ranging from -3.37 to -3.71 against a 5% critical value of approximately -2.20; per-country rejections cover 17 to 20 of the 21–25 countries in each maturity panel. Hadri rejects the panel stationarity null at the 1% level for all four maturities, with LM statistics exceeding 125 against an $N(0,1)$ reference distribution; per-country KPSS rejections cover 19 to 25 of the included countries. The combination — decisive rejection of both the strict-unit-root null and the strict-stationary null — is the signature of high persistence with slowly drifting levels rather than either pure random-walk or pure mean-reverting dynamics.

The implication for the forecasting exercise follows the high-persistence characterization in Section II: at the daily horizon, predictable variation in $s_{ij,t}$ is small relative to the country-maturity trailing mean $\bar{s}_{ij}$. The within-country-maturity underperformance documented in Section IV.B is a direct consequence of this property and does not reflect an absence of predictive content in the predictor block more broadly.

C Supplementary Empirical Results

C.1 Yearly R^2_{oos} for sub-groups and long horizons

We present three sets of figures: the 1-month and 3-month heatmaps for each of the three country sub-groups (Figures C.1–C.3); the corresponding 6-month and 12-month heatmaps for the Global panel (Figure C.4); and the 6-month and 12-month heatmaps for each country

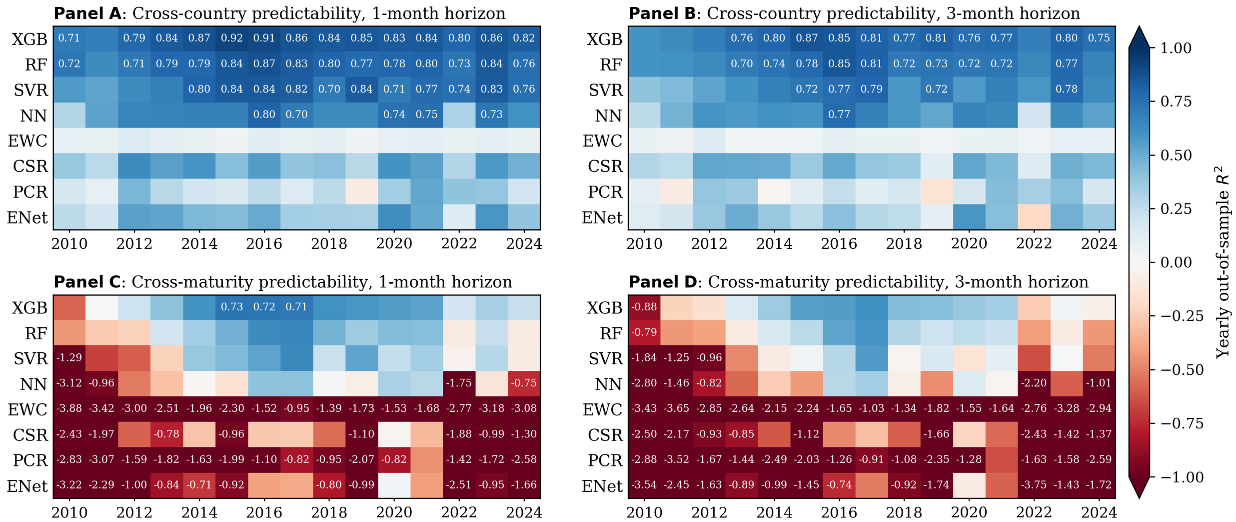


Figure C.1: **Yearly R^2_{oos} across benchmarks and horizons, OECD non-EU panel.** The figure reports the annual average of $R^2_{t,oos}$ for the OECD non-EU country sub-group. Rows: cross-country predictability against $\bar{s}_{\cdot j}$ (top) and cross-maturity predictability against $\bar{s}_{i\cdot}$ (bottom). Columns: 1-month (left) and 3-month (right) testing horizons. Models and formatting follow Figure 2 in the main paper.

sub-group (Figures C.5–C.7). Formatting follows the main-text figure: cells are colored by R^2 using a diverging scale centered at zero, and values with $|R^2| \geq 0.7$ are annotated.

The three sub-group figures confirm the main-text pattern qualitatively: linear methods are broadly positive against $\bar{s}_{\cdot j}$ and systematically negative against $\bar{s}_{i\cdot}$, while XGBoost and Random Forest maintain positive cross-maturity R^2 across most years at the 1-month horizon.

The EWC in the OECD non-EU panel (Figure C.1) is essentially flat at zero across all years, reinforcing the observation that cross-country predictability cannot be obtained from trivial aggregation. Second, the EU ex-SWEAP panel (Figure C.2) shows the cleanest separation between tree-based ensembles and all other methods at the cross-maturity benchmark. Third, the SWEAP panel (Figure C.3) exhibits catastrophic cross-maturity cells for several linear methods in the early years (e.g., EWC below -9 in 2011–2012), illustrating that small-panel linear methods are sensitive to the extreme volatility of the Eurozone crisis years. The within-row ordering of nonlinear methods (XGBoost $\approx$ RF $>$ SVR $>$ NN against $\bar{s}_{i\cdot}$) is preserved across all three sub-groups.

Figure C.4 reports the yearly R^2 pattern at the 6-month and 12-month horizons for the Global panel. The main message is the extension of framing B: linear cross-country predictability degrades as the horizon lengthens, and the 2022–2024 tightening-cycle cells turn visibly red at the 12-month horizon (Elastic Net panel (b) shows -1.08 for 2022 and -0.86 for 2023).

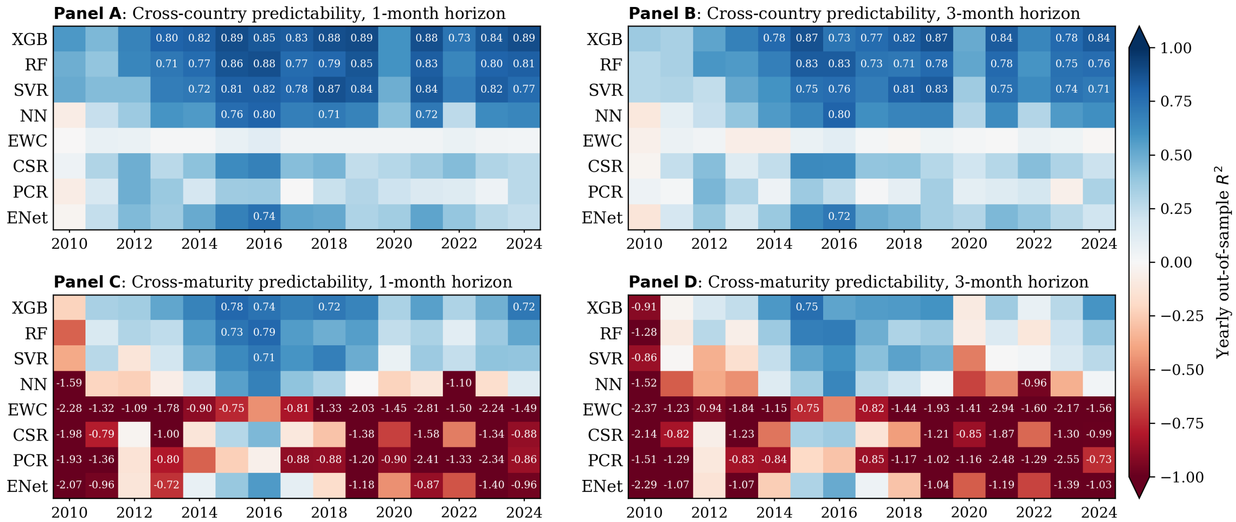


Figure C.2: **Yearly R^2_{oos} across benchmarks and horizons, EU ex-SWEAP panel.** The figure reports the annual average of $R^2_{t,oos}$ for the EU ex-SWEAP country sub-group. Rows, columns, and formatting follow Figure C.1.

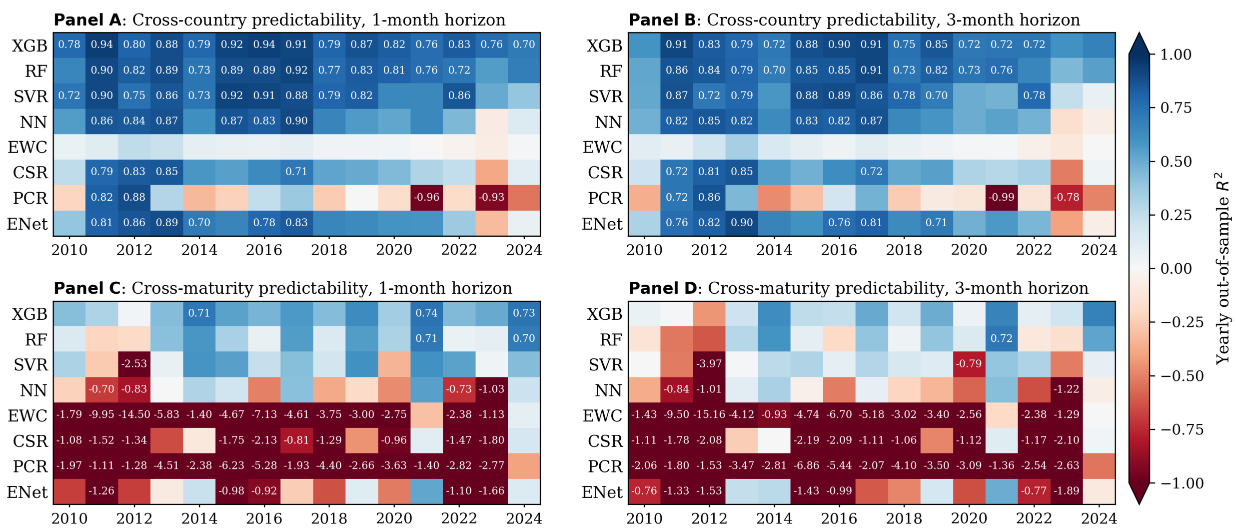


Figure C.3: **Yearly R^2_{oos} across benchmarks and horizons, SWEAP panel.** The figure reports the annual average of $R^2_{t,oos}$ for the SWEAP country sub-group. Rows, columns, and formatting follow Figure C.1.

XGBoost retains positive R^2 against $\bar{s}_{\cdot j}$ across the full 2010–2024 window at both horizons, and the relative ordering of nonlinear methods remains stable. Against $\bar{s}_{i\cdot}$ (bottom row), every method delivers negative R^2 at the 12-month horizon—daily cross-maturity prediction at this horizon lies beyond the reach of any specification in our set, regardless of functional form.

Figures C.5–C.7 confirm that the long-horizon pattern observed on the Global panel holds

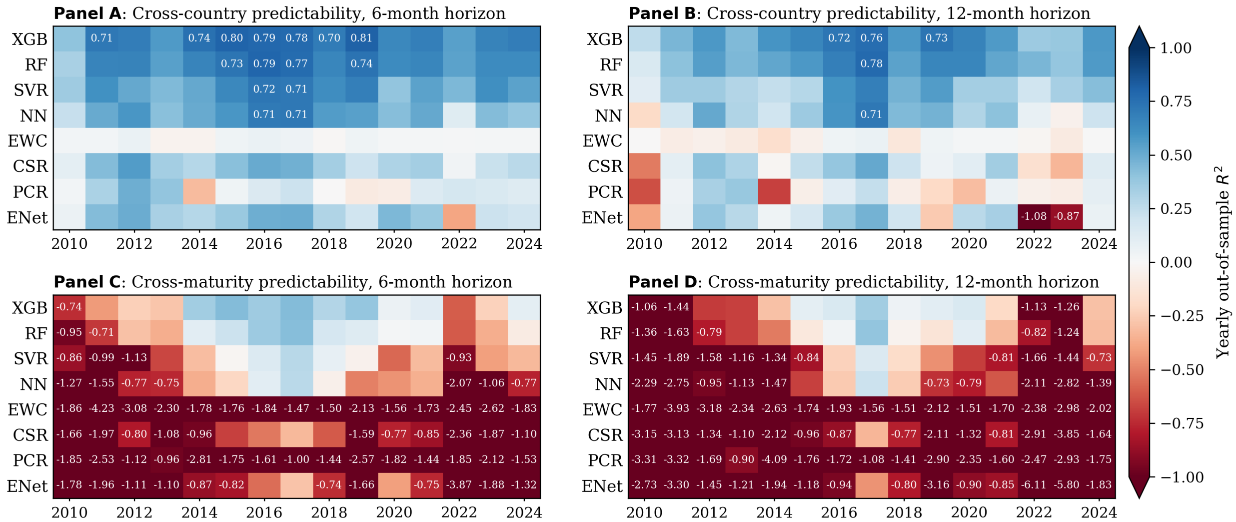


Figure C.4: **Yearly R^2_{oos} across benchmarks at long horizons, Global panel.** The figure reports the annual average of $R^2_{t,oos}$ on the Global panel for the 6-month (left columns) and 12-month (right columns) testing horizons. Rows: cross-country predictability against $\bar{s}_{\cdot j}$ (top) and cross-maturity predictability against $\bar{s}_{i\cdot}$ (bottom). Models and formatting follow Figure C.1.

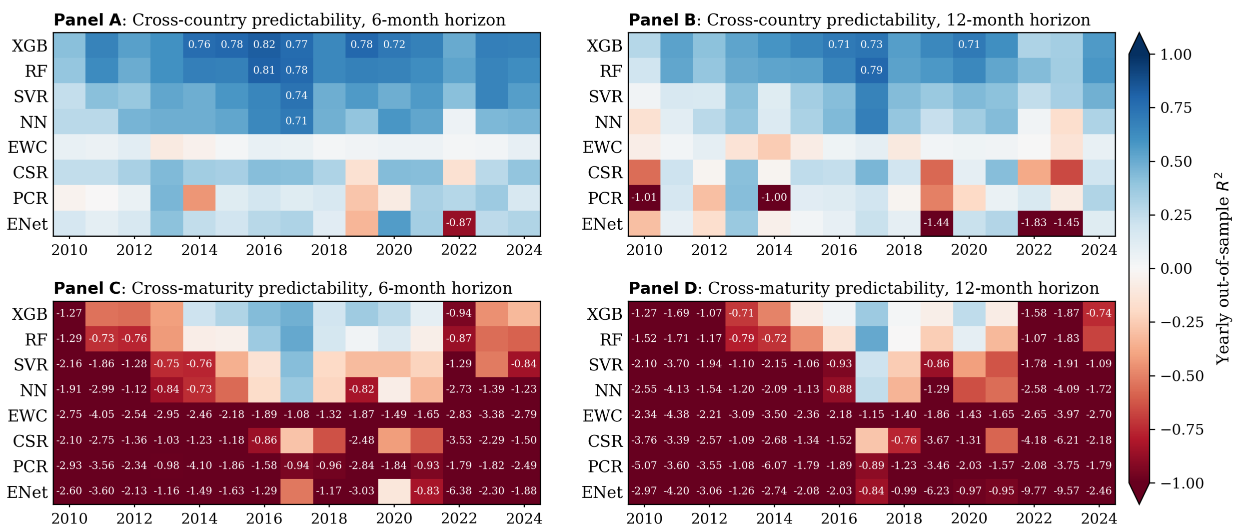


Figure C.5: **Yearly R^2_{oos} across benchmarks at long horizons, OECD non-EU panel.** The figure reports the annual average of $R^2_{t,oos}$ for the OECD non-EU country sub-group at the 6-month (left columns) and 12-month (right columns) testing horizons. Rows, columns, and formatting follow Figure C.4.

across country sub-groups. In each sub-group, the cross-country R^2 against $\bar{s}_{\cdot j}$ remains positive for most years and all nonlinear methods at the 6-month horizon, but linear methods weaken in the 2022–2024 tightening-cycle cells, with SWEAP showing the sharpest linear fragility at the 12-month horizon (Elastic Net in panel (b) of Figure C.7) and OECD non-EU the most resilient. Against $\bar{s}_{i\cdot}$ at 12-month horizon, all methods deliver negative R^2 across

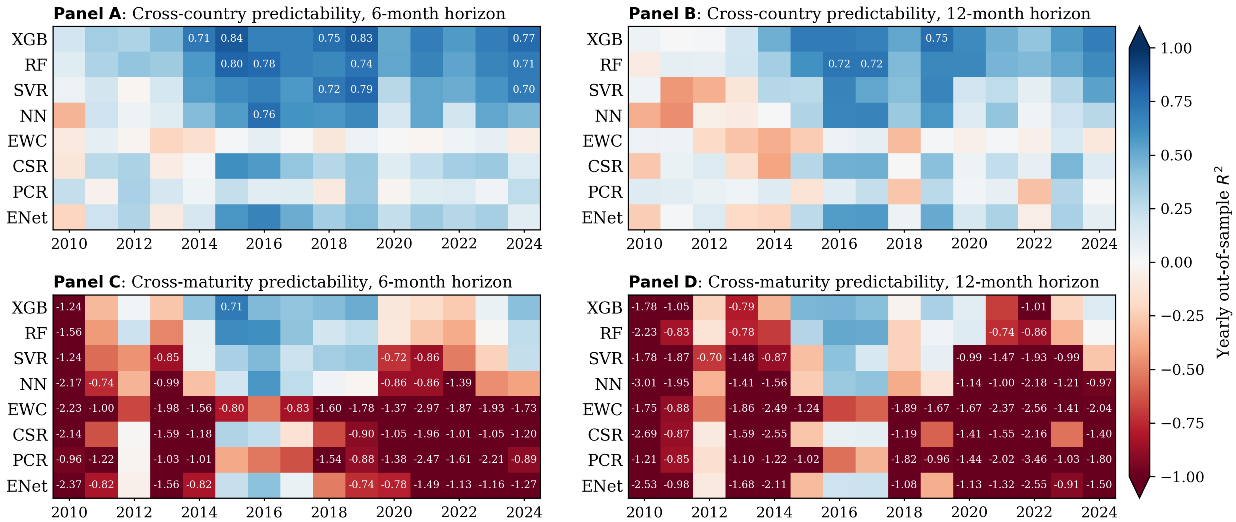


Figure C.6: **Yearly R^2_{oos} at long horizons, EU ex-SWEAP panel.** The figure reports the annual average of $R^2_{t,oos}$ for the EU ex-SWEAP country sub-group at the 6-month and 12-month testing horizons. Rows, columns, and formatting follow Figure C.4.

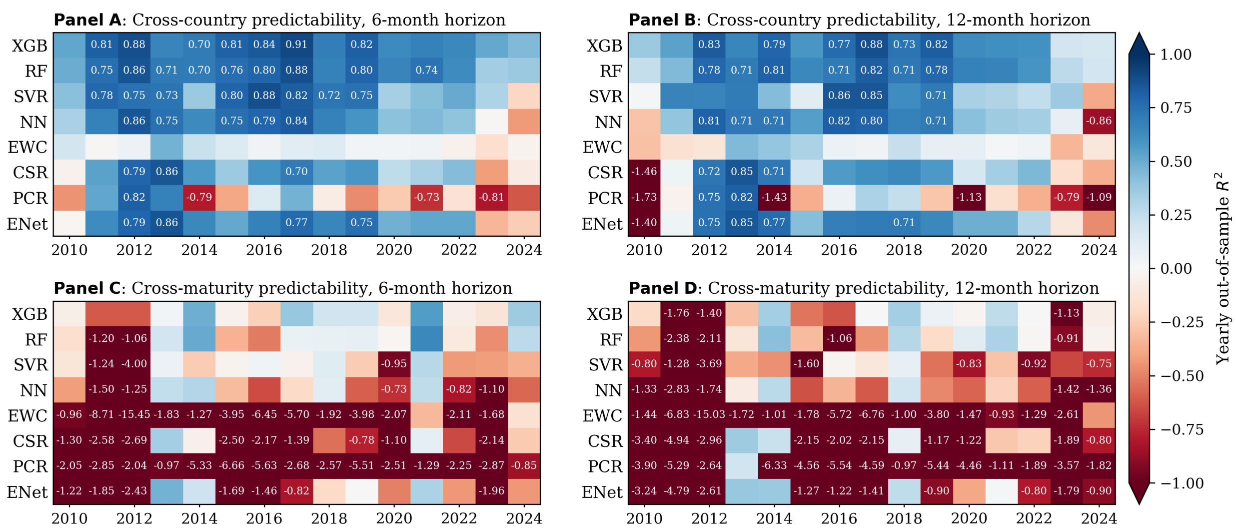


Figure C.7: **Yearly R^2_{oos} at long horizons, SWEAP panel, 2010–2024.** The figure reports the annual average of $R^2_{t,oos}$ for the SWEAP country sub-group at the 6-month and 12-month testing horizons. Rows, columns, and formatting follow Figure C.4.

all country sub-groups, reinforcing the main-text reading that cross-maturity prediction at the 12-month horizon is not within reach of either linear or nonlinear methods on a daily panel of this size.

Horizon	Benchmark	*Level-only linear* OLS	EWC	CSR	Lasso	Ridge	EN	PCR	PLS
1-month	$\bar{s}_{..}$	0.43	0.19	0.49	0.51	0.47	**0.52**	0.35	0.48
	$\bar{s}_{.j}$	0.35	0.06	0.42	0.43	0.40	**0.45**	0.24	0.40
	$\bar{s}_{i.}$	−1.09	−2.12	−0.85	−0.83	−0.90	**−0.77**	−1.45	−0.97
	$\bar{s}_{ij}$	−5.92	−10.29	−5.49	−5.20	−5.49	**−5.11**	−8.50	−6.04
3-month	$\bar{s}_{..}$	0.15	0.17	0.43	0.44	0.35	**0.44**	0.30	0.30
	$\bar{s}_{.j}$	0.06	0.04	0.35	0.36	0.26	**0.36**	0.18	0.20
	$\bar{s}_{i.}$	−2.20	−2.13	−1.09	−1.05	−1.37	**−1.02**	−1.58	−1.59
	$\bar{s}_{ij}$	−7.60	−9.86	−5.85	−5.68	−6.39	**−5.65**	−8.48	−7.69
6-month	$\bar{s}_{..}$	−0.02	0.14	**0.35**	0.35	0.24	0.35	0.22	0.01
	$\bar{s}_{.j}$	−0.13	0.01	**0.26**	0.26	0.13	0.25	0.10	−0.12
	$\bar{s}_{i.}$	−2.48	−2.14	**−1.31**	−1.35	−1.70	−1.34	−1.79	−2.52
	$\bar{s}_{ij}$	−8.55	−9.28	**−6.20**	−6.35	−7.25	−6.37	−8.54	−10.51
12-month	$\bar{s}_{..}$	−0.27	0.10	**0.22**	0.14	−0.01	0.12	0.08	−0.52
	$\bar{s}_{.j}$	−0.43	−0.03	**0.11**	0.02	−0.15	0.00	−0.06	−0.74
	$\bar{s}_{i.}$	−3.44	−2.23	**−1.79**	−2.14	−2.59	−2.17	−2.23	−4.46
	$\bar{s}_{ij}$	−10.90	−8.66	**−6.99**	−8.05	−9.18	−8.15	−8.84	−16.29

Table C.1: **Four-benchmark aggregate $\overline{R}^2_{oos}$, Global panel, linear "level-only".** The table reports the full-sample average of yearly $R^2_{t,oos}$ (2009–2024) for each level-only linear specification (without maturity interactions), forecasting horizon, and naive benchmark. Benchmarks and horizons are as in Table 1 of the paper; within each row, the best-performing specification is in bold.

C.2 Aggregate Results for Sub-Groups and Linear "Level-Only"

Table C.1 repeats the four-benchmark decomposition for the eight level-only linear specifications on the Global panel. Tables C.2–C.4 report the decomposition separately for the three country sub-groups—OECD non-EU, EU ex-SWEAP, and SWEAP—using the same level+slope linear and nonlinear specifications as the main-text table. The layout of each appendix table mirrors Table 1 in the paper, so the decomposition reading is preserved: scanning a single model column from top to bottom within a horizon block traces the R^2 decomposition as the benchmark becomes more demanding.

The level-only results in Table C.1 closely track the level+slope pattern in Table 1 of the paper. Elastic Net without maturity interactions achieves an average R^2 of 0.52 against $\bar{s}_{..}$ at 1M, 0.45 against $\bar{s}_{.j}$, and −0.77 against $\bar{s}_{i.}$; the corresponding level+slope values in the main-text table are 0.52, 0.45, and −0.76. The gain from maturity interactions in linear specifications is quantitatively negligible across all four benchmarks and all four horizons, confirming the observation in prior work (e.g., Balduzzi et al., 2023) that simple maturity-variable interactions do not recover the term-structure nonlinearity captured by nonlinear methods. We focus on the level+slope specifications in the main text because they are the more favorable comparison for the linear class.

The three sub-group tables establish that the four-benchmark pattern of the Global panel is preserved across country groups. In every sub-group and at every horizon, linear specifications deliver positive R^2 against $\bar{s}_{\cdot j}$ and negative R^2 against $\bar{s}_{i\cdot}$; nonlinear specifications—XGBoost in particular—retain positive R^2 on the cross-maturity benchmark at short horizons. In addition to the sign pattern being preserved across sub-groups, two further patterns are worth noting.

First, the OECD non-EU panel (Table C.2) yields the lowest linear cross-country R^2 among the three subgroups (Elastic Net at 1M is 0.40, relative to 0.45 for Global and 0.50 for SWEAP), with modest degradation at longer horizons. The smaller magnitudes relative to Global reflect the lower cross-country variance in this subset—OECD non-EU countries have relatively homogeneous macroeconomic fundamentals—which reduces the denominator of the R^2 statistic with respect to $\bar{s}_{\cdot j}$.

Second, the SWEAP panel (Table C.4) exhibits the largest cross-country R^2 values among the sub-groups at short horizons (XGBoost at 1M against $\bar{s}_{\cdot\cdot}$ is 0.82, comparable to the Global panel). This reflects the pronounced cross-country dispersion introduced by peripheral Eurozone sovereigns, which enlarges the R^2 denominator and makes cross-country prediction visibly rewarding. However, the SWEAP panel also shows increasing linear fragility at long horizons: at the 12-month horizon, Elastic Net's full-sample R^2 against $\bar{s}_{\cdot j}$ remains positive (0.32) but its yearly values collapse sharply in the 2022–2024 tightening window (see Figure C.7). Against $\bar{s}_{i\cdot}$ at 12M, Elastic Net averages -1.22, versus -0.34 for XGBoost, preserving the benchmark-based decomposition pattern observed in the Global panel.

C.3 First-Release vs Vintage Data

A concern in real-time forecasting is whether first-release macroeconomic data contain sufficient signal relative to revised vintages. In this subsection, we compare the out-of-sample R^2 of XGBoost trained on first-release information with that of XGBoost trained on the latest available data vintage for each forecast date. Table C.5 reports the results for one-month, three-month, six-month, and 12-month testing horizons, for the Global panel and for each country sub-group (OECD non-EU, EU ex-SWEAP, SWEAP). The naive benchmark is the cross-country or cross-maturity mean, as in the main text.

		Level + Slope linear								*Nonlinear models*				
Horizon	Benchmark	OLS	EWC	CSR	Lasso	Ridge	EN	PCR	PLS	SVR	RT	RF	XGB	NN
1M	$\bar{s}_{..}$	0.37	0.21	**0.52**	0.46	0.43	0.47	0.37	0.44	0.76	0.68	0.79	**0.83**	0.65
	$\bar{s}_{.j}$	0.28	0.08	**0.45**	0.39	0.35	0.40	0.26	0.35	0.72	0.64	0.76	**0.81**	0.60
	$\bar{s}_{i.}$	−1.71	−2.37	**−1.00**	−1.26	−1.39	−1.20	−1.65	−1.27	−0.00	−0.30	0.16	**0.34**	−0.59
	$\bar{s}_{ij}$	−8.35	−11.99	**−6.33**	−7.20	−7.55	−7.02	−9.92	−8.06	−2.27	−3.36	−1.78	**−0.99**	−4.23
3M	$\bar{s}_{..}$	−0.10	0.19	**0.44**	0.37	0.27	0.38	0.30	0.24	0.66	0.59	0.73	**0.77**	0.58
	$\bar{s}_{.j}$	−0.21	0.06	**0.36**	0.28	0.17	0.28	0.19	0.12	0.62	0.53	0.70	**0.74**	0.53
	$\bar{s}_{i.}$	−4.79	−2.36	**−1.34**	−1.59	−2.14	−1.59	−1.83	−2.02	−0.37	−0.66	−0.06	**0.10**	−0.79
	$\bar{s}_{ij}$	−12.63	−11.30	**−6.93**	−7.99	−9.26	−7.96	−9.90	−10.27	−3.44	−4.56	−2.46	**−1.77**	−4.76
6M	$\bar{s}_{..}$	−0.40	0.16	**0.34**	0.24	0.08	0.23	0.20	−0.10	0.56	0.50	0.67	**0.70**	0.49
	$\bar{s}_{.j}$	−0.54	0.03	**0.24**	0.13	−0.05	0.12	0.08	−0.28	0.51	0.43	0.62	**0.66**	0.43
	$\bar{s}_{i.}$	−4.67	−2.36	**−1.65**	−2.05	−2.74	−2.09	−2.11	−3.21	−0.74	−0.99	−0.30	**−0.17**	−1.06
	$\bar{s}_{ij}$	−14.05	−10.72	**−7.76**	−9.29	−11.02	−9.36	−10.17	−14.70	−4.56	−5.46	−3.15	**−2.55**	−5.51
12M	$\bar{s}_{..}$	−0.77	0.12	**0.14**	−0.07	−0.28	−0.09	0.04	−0.84	0.42	0.39	0.58	**0.60**	0.34
	$\bar{s}_{.j}$	−0.97	−0.02	**0.02**	−0.23	−0.46	−0.25	−0.11	−1.14	0.34	0.30	0.52	**0.55**	0.25
	$\bar{s}_{i.}$	−5.84	−2.46	**−2.37**	−3.27	−4.12	−3.34	−2.66	−6.03	−1.31	−1.45	−0.67	**−0.60**	−1.63
	$\bar{s}_{ij}$	−17.91	−10.11	**−9.45**	−12.39	−14.44	−12.55	−10.81	−25.66	−6.03	−6.32	−3.95	**−3.60**	−6.88

Table C.2: **Four-benchmark aggregate $\overline{R}^2_{oos}$, OECD non-EU panel.** The table reports the full-sample average of yearly $R^2_{t,oos}$ (2009–2024) for the OECD non-EU country sub-group. Benchmarks, horizons, models, and formatting follow Table 1 in the paper. Analogous tables for the Global panel and other country sub-groups are reported in Tables C.3 and C.4.

		Level + Slope linear								*Nonlinear models*				
Horizon	Benchmark	OLS	EWC	CSR	Lasso	Ridge	EN	PCR	PLS	SVR	RT	RF	XGB	NN
1-month	$\bar{s}_{..}$	0.39	0.22	0.45	0.47	0.44	**0.48**	0.37	0.43	0.72	0.59	0.73	**0.78**	0.55
	$\bar{s}_{.j}$	0.28	0.04	0.34	0.37	0.32	**0.37**	0.22	0.31	0.67	0.52	0.69	**0.74**	0.47
	$\bar{s}_{i.}$	−0.81	−1.51	−0.73	−0.66	−0.79	**−0.64**	−1.06	−0.77	0.30	−0.10	0.28	**0.44**	−0.23
	$\bar{s}_{ij}$	−7.76	−13.39	−8.20	−7.22	−8.06	**−7.10**	−10.61	−8.22	−2.05	−3.32	−1.87	**−1.16**	−4.11
3-month	$\bar{s}_{..}$	0.27	0.20	0.40	**0.43**	0.37	0.42	0.34	0.29	0.63	0.52	0.67	**0.70**	0.49
	$\bar{s}_{.j}$	0.14	0.01	0.28	**0.32**	0.24	0.31	0.18	0.13	0.57	0.43	0.61	**0.65**	0.40
	$\bar{s}_{i.}$	−1.12	−1.51	−0.84	−0.75	−0.95	**−0.74**	−1.08	−1.13	0.04	−0.38	0.07	**0.20**	−0.38
	$\bar{s}_{ij}$	−8.08	−12.36	−7.97	−7.08	−8.15	**−7.02**	−9.99	−9.21	−2.92	−4.18	−2.55	**−1.94**	−4.45
6-month	$\bar{s}_{..}$	0.25	0.18	0.36	**0.39**	0.31	0.38	0.29	0.10	0.53	0.43	0.61	**0.64**	0.43
	$\bar{s}_{.j}$	0.11	−0.01	0.23	**0.27**	0.17	0.26	0.13	−0.11	0.44	0.32	0.54	**0.57**	0.33
	$\bar{s}_{i.}$	−1.18	−1.50	−0.91	**−0.82**	−1.06	−0.83	−1.10	−1.61	−0.25	−0.61	−0.09	**0.01**	−0.58
	$\bar{s}_{ij}$	−8.15	−11.34	−7.65	−6.95	−8.21	**−6.94**	−9.36	−10.98	−3.96	−5.06	−3.04	**−2.63**	−5.04
12-month	$\bar{s}_{..}$	0.13	0.13	0.29	**0.31**	0.20	0.30	0.24	−0.14	0.35	0.29	0.52	**0.54**	0.31
	$\bar{s}_{.j}$	−0.05	−0.08	0.13	**0.16**	0.03	0.15	0.05	−0.40	0.22	0.15	0.43	**0.46**	0.18
	$\bar{s}_{i.}$	−1.61	−1.58	−1.12	**−1.05**	−1.38	−1.08	−1.27	−2.47	−0.82	−1.00	−0.36	**−0.28**	−0.98
	$\bar{s}_{ij}$	−8.74	−10.31	−7.46	**−7.09**	−8.58	−7.13	−9.03	−14.71	−5.53	−6.19	−3.73	**−3.42**	−6.17

Table C.3: **Four-benchmark aggregate $\overline{R}^2_{oos}$, EU ex-SWEAP panel.** The table reports the full-sample average of yearly $R^2_{t,oos}$ (2009–2024) for the EU ex-SWEAP country sub-group. Benchmarks, horizons, models, and formatting follow Table 1 in the paper. Analogous tables for the Global panel and other country sub-groups are reported in Tables C.2 and C.4.

Horizon	Benchmark	*Level + Slope linear*								*Nonlinear models*				
		OLS	EWC	CSR	Lasso	Ridge	EN	PCR	PLS	SVR	RT	RF	XGB	NN
1-month	$\bar{s}_{..}$	0.44	0.15	0.48	0.52	0.53	**0.57**	0.08	0.30	0.76	0.67	0.79	**0.82**	0.63
	$\bar{s}_{.j}$	0.37	0.06	0.41	0.46	0.46	**0.50**	−0.07	0.20	0.72	0.61	0.76	**0.80**	0.57
	$\bar{s}_{i.}$	−0.71	−4.02	−1.04	−0.65	**−0.57**	−0.60	−2.73	−1.12	0.06	−0.05	0.26	**0.45**	−0.28
	$\bar{s}_{ij}$	**−5.25**	−15.98	−7.44	−5.37	−5.61	−5.48	−16.28	−7.68	−1.75	−2.89	−1.75	**−0.69**	−3.75
3-month	$\bar{s}_{..}$	0.23	0.14	0.41	0.44	0.41	**0.50**	0.05	−0.03	0.67	0.58	0.74	**0.77**	0.60
	$\bar{s}_{.j}$	0.16	0.06	0.34	0.38	0.33	**0.44**	−0.09	−0.14	0.61	0.51	0.71	**0.75**	0.53
	$\bar{s}_{i.}$	−1.27	−3.91	−1.23	−0.85	−0.92	**−0.78**	−2.83	−1.99	−0.27	−0.30	0.07	**0.23**	−0.38
	$\bar{s}_{ij}$	−5.61	−12.80	−6.31	**−4.71**	−5.33	−4.79	−12.59	−8.09	−2.46	−3.37	−1.85	**−1.12**	−3.33
6-month	$\bar{s}_{..}$	0.02	0.14	0.34	0.40	0.28	**0.46**	−0.04	−0.45	0.60	0.49	0.69	**0.71**	0.54
	$\bar{s}_{.j}$	−0.06	0.06	0.27	0.33	0.21	**0.40**	−0.17	−0.57	0.53	0.41	0.65	**0.68**	0.47
	$\bar{s}_{i.}$	−1.75	−3.71	−1.38	−1.00	−1.24	**−0.92**	−3.05	−3.12	−0.53	−0.60	−0.16	**−0.00**	−0.54
	$\bar{s}_{ij}$	−6.05	−10.44	−5.55	**−4.36**	−5.34	−4.41	−10.88	−9.89	−2.74	−4.15	−2.10	**−1.51**	−3.23
12-month	$\bar{s}_{..}$	−0.01	0.12	0.33	**0.41**	0.26	0.39	−0.22	−0.69	0.50	0.40	0.64	**0.65**	0.49
	$\bar{s}_{.j}$	−0.09	0.04	0.26	**0.34**	0.18	0.32	−0.34	−0.84	0.43	0.32	0.60	**0.61**	0.41
	$\bar{s}_{i.}$	−2.06	−3.45	−1.50	**−1.17**	−1.51	−1.22	−3.46	−4.06	−0.82	−1.04	−0.46	**−0.34**	−0.78
	$\bar{s}_{ij}$	−6.47	−8.45	−4.99	**−4.10**	−5.46	−4.26	−10.15	−11.72	−3.04	−4.41	−2.39	**−2.18**	−3.38

Table C.4: **Four-benchmark aggregate $\overline{R}^2_{oos}$, SWEAP panel.** The table reports the full-sample average of yearly $R^2_{t,oos}$ (2009–2024) for the SWEAP country sub-group. Benchmarks, horizons, models, and formatting follow Table 1 in the paper. Analogous tables for the Global panel and other country sub-groups are reported in Tables C.2 and C.3.

	Cross-country ($\bar{s}_{\cdot j}$)		Cross-maturity ($\bar{s}_{i\cdot}$)	
	First release	Vintages	First release	Vintages
1-month testing				
Global	0.890	0.893	0.631	0.651
OECD non-EU	0.905	0.906	0.675	0.681
EU ex-SWEAP	0.846	0.859	0.621	0.660
SWEAP	0.881	0.883	0.362	0.345
3-month testing				
Global	0.808	0.810	0.376	0.389
OECD non-EU	0.825	0.826	0.409	0.419
EU ex-SWEAP	0.747	0.758	0.353	0.400
SWEAP	0.831	0.820	0.096	0.070
6-month testing				
Global	0.723	0.720	0.101	0.116
OECD non-EU	0.746	0.746	0.150	0.161
EU ex-SWEAP	0.644	0.653	0.052	0.145
SWEAP	0.769	0.729	−0.194	−0.341
12-month testing				
Global	0.588	0.590	−0.316	−0.284
OECD non-EU	0.611	0.624	−0.294	−0.225
EU ex-SWEAP	0.487	0.477	−0.367	−0.320
SWEAP	0.663	0.665	−0.596	−0.651

Table C.5: **XGBoost with first-release vs. vintage data.** The table compares the R^2_{oos} of XGBoost trained on first-release macroeconomic data against XGBoost trained on the latest available data vintage. Results are reported for 1-month, 3-month, 6-month, and 12-month testing horizons, for the Global panel and for each country sub-group. The R^2_{oos} values are obtained by averaging the period-by-period $R^2_{t,oos}$ over the full out-of-sample period. The naive benchmark is the average of the log CDS spread across countries (left panel) or across maturities (right panel).

References

Balduzzi, P., Savona, R., and Alessi, L. (2023). Anatomy of a sovereign debt crisis: machine learning, real-time macro fundamentals, and cds spreads. *Journal of Financial Econometrics*, 21(5):1728–1758.

Bianchi, D., Büchner, M., and Tamoni, A. (2021). Bond risk premiums with machine learning. *The Review of Financial Studies*, 34(2):1046–1089.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.

Drucker, H., Burges, C. J., Kaufman, L., Smola, A., and Vapnik, V. (1996). Support vector regression machines. *Advances in neural information processing systems*, 9.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273.

Hadri, K. (2000). Testing for stationarity in heterogeneous panel data. *The econometrics journal*, 3(2):148–161.

Hsiang, T. (1975). A bayesian view on ridge regression. *Journal of the Royal Statistical Society: Series D*, 24(4):267–268.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of econometrics*, 54(1-3):159–178.

Pesaran, M. H. (2007). A simple panel unit root test in the presence of cross-section dependence. *Journal of applied econometrics*, 22(2):265–312.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288.

Zou, H. and 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.