Economic Political Uncertainty Index for Peru Using X and DeepSeek-V3

Marcelo Gallardo* Manuel Loaiza† Gabriel Rodríguez‡

June 13, 2026

Abstract

We build a new daily economic political uncertainty (EPU) index for Peru from the public discourse on X (formerly Twitter) of influential figures in politics, economics, journalism, and business, classifying each message along economic, political, and uncertainty dimensions with a large language model (LLM). Recombining these dimensions yields a family of complementary daily indices of economic and/or political uncertainty—an instrument that does not yet exist for Peru and is especially valuable in an emerging economy subject to acute and recurrent political turmoil. We study the dynamic relationship between the index and domestic and international financial and macroeconomic variables using correlations, rolling correlations, and linear and nonlinear Granger causality tests. Over the full sample, linear predictability runs from financial conditions to the index, indicating that influential commentators react to market movements. This unconditional result masks pronounced state dependence: during crisis episodes—dated endogenously from market stress (equity-return volatility and exchange-rate pressure), with documented events used only as ex-post validation—the index gains predictive power over sovereign-yield changes and, less robustly, over stock returns, while remaining uninformative in calm periods. A broader political-uncertainty (PU) index captures intensely political episodes that the stricter EPU measure misses, making the two complementary. A social-media-based uncertainty index is thus a conditionally useful indicator whose signal activates precisely during the crises when it matters most.

Keywords: Economic Political Uncertainty, X (formerly Twitter), Large Language Models, social media, Granger causality, DeepSeek-V3, GARCH, Peru.

JEL Classification: C32, C43, C55, C82, D80, E44, G15.

Declaration on the Use of AI Tools: The authors used Claude Fable 5 (Anthropic) for drafting and editing the LaTeX source, and Refine Ink for manuscript feedback. The authors reviewed all AI-assisted content and take full responsibility for it; all errors are their own.

Declaration on Political Neutrality: This paper is a purely academic and methodological contribution. Its analysis, findings, and conclusions do not reflect, endorse, or promote any political position, party, candidate, or agenda, and are not influenced by any political interest or affiliation of the authors.

*Department of Mathematics, Pontificia Universidad Católica del Perú.
†Autodesk Inc.
‡Department of Economics, Pontificia Universidad Católica del Perú.

1 Introduction

Uncertainty about economic and political conditions shapes real and financial outcomes through well-documented channels (Bloom, 2009, 2014). When uncertainty rises, firms postpone irreversible investment and hiring while investors reprice risk, so uncertainty shocks act much like adverse demand shocks for output and employment. These effects are amplified in emerging economies, where external financing is procyclical and sovereign-risk premia react sharply to political news (Frankel, 2010; Carrière-Swallow and Céspedes, 2013). Measuring uncertainty at high frequency is therefore valuable to policymakers and investors alike.

Much of the relevant information arrives as narrative, and how it propagates through public discourse is itself an economic force: narratives spread contagiously and drive aggregate fluctuations (Shiller, 2017), and media coverage can generate frenzies and comovement in asset prices (Veldkamp, 2006). Empirically, the tone of the financial press predicts short-run stock returns, especially in downturns (Tetlock, 2007; García, 2013), and high-frequency proxies built from internet search (Da et al., 2015) and social-media text (Baker et al., 2021; Tumasjan et al., 2010; Bovet and Makse, 2019) track sentiment and forecast returns and volatility.

The dominant measure of policy uncertainty is the newspaper-based index, as in Baker et al. (2016), which counts articles jointly mentioning the economy, policy, and uncertainty; it relies on traditional media and is typically available only monthly. To capture real-time dynamics, recent literature leverages social media platforms like X (formerly Twitter); for instance, Becerra and Sagner (2023) use the official X handles of mainstream Chilean news outlets, newspapers, and radio stations. While this provides a higher-frequency alternative, it remains bound to centralized, institutional media narratives. We depart from this institutional focus by tracking the accounts of key individual opinion leaders including politicians, journalists, analysts and policy influencers capturing a more direct and responsive measure of uncertainty as it unfolds.

Turning such discourse into a quantitative index is a measurement problem that the text-as-data literature has progressively addressed (Gentzkow et al., 2019). Early measures relied on fixed dictionaries and word counts (Loughran and McDonald, 2011; Baker et al., 2016; Becerra and Sagner, 2023), which are transparent but miss context, negation, and irony; supervised and contextual models then improved accuracy (Hassan et al., 2019). The transformer architecture behind modern large language models—from Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019) to the generative models that followed (Brown et al., 2020)—now classifies short, informal text with near-human reliability, opening new possibilities for economic measurement (Korinek, 2023). We exploit this advance: instead of keyword matching, a large language model reads each tweet and judges whether it concerns the economy, politics, and uncertainty. A human-coded validation of the classifier on Peruvian-Spanish posts is left for future work.

We construct a daily index for Peru over January 2018–January 2023. The setting is informative: within five years the country had six presidents, a dissolved Congress, a contested election, an attempted self-coup, and sustained protests, all overlapping with the COVID-19 pandemic and the 2021–2022 inflation surge. This concentration of political–economic crises in a single small open economy makes Peru a natural laboratory, and a near-real-time measure of uncertainty especially valuable.

This paper makes three contributions. First, we build a new daily index of economic political uncertainty for Peru from a novel data source: the X discourse of influential accounts in poli-

tics, economics, journalism, and business. Unlike the newspaper-based *economic policy uncertainty* index of Baker et al. (2016), which is monthly and relies on press archives, our index exploits high-frequency social-media text, classifying each tweet along economic, political, and uncertainty dimensions with a large language model (DeepSeek-V3); by economic political uncertainty (EPU) we mean the joint intersection $E \cap P \cap U$ of economic, political, and uncertainty-laden discourse—a social-media analogue of, but conceptually distinct from, Baker et al. (2016). Recombining these dimensions yields a family of daily indices spanning economic and political uncertainty that does not yet exist for Peru and can be replicated. Second, we relate the index to domestic and international financial and macroeconomic variables using Pearson and rolling correlations, linear Granger causality tests (Granger, 1969, 1980), the nonlinear test of Diks and Panchenko (2006), and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) family models with the index as an exogenous regressor in the variance equation. Over the full sample, linear predictability runs in one direction only: financial variables help forecast the index, while the index does not help forecast them. The nonlinear test does reject independence in some pairs, but this reflects shared higher-order dynamics—such as common volatility clustering—rather than exploitable predictive content, so we draw no causal conclusions from it. Third, this unconditional non-result masks state dependence. Dating crisis episodes *endogenously* from the financial data themselves—the top quintile of a market-stress index built from equity-return volatility and exchange-rate pressure, rather than from a hand-picked news calendar in the tradition of Romer and Romer (2010) and event-based risk indices (Caldara and Iacoviello, 2022), which we retain only as ex-post validation—we find that within crises the index predicts sovereign-yield changes—robustly across alignment conventions—and, more fragilely, stock returns, while remaining inert in calm periods. A complementary Political Uncertainty index, built from the political–uncertainty intersection $P \cap U$, captures intensely political episodes that the stricter measure misses.

The paper is organized as follows: Section 2 describes the data; Section 3 constructs the index family; Section 4 presents the statistical analysis—co-movement, linear and nonlinear causality, and the state-dependent crisis split; Section 5 models the conditional volatility of the index; and Section 6 concludes.

2 Data

2.1 Tweet Collection

We download tweets from 97 accounts representing the main agents in Peruvian political and economic discourse, spanning 23 January 2018 to 31 January 2023. The accounts cover five categories: journalists (24), politicians (42), economists (16), political analysts (13), and private sector representatives (2). Accounts are selected based on three criteria: (i) public visibility in Peruvian economic or political discourse, measured by follower count and media citations; (ii) sustained activity throughout the sample window (at least 100 tweets); and (iii) coverage of the main institutional actors: Congress, the executive, the central bank, major media outlets, and business chambers (see Table 1). While this elite-focused design limits representativeness relative to a full-population sample, it mirrors the approach of Baker et al. (2016) of targeting "informed agents" whose discourse is more likely to reflect genuine policy uncertainty rather than noise.

The raw data is obtained via the X (formerly Twitter) application programming interface (API). For each account, we retrieve all available tweets within the sample window, producing one CSV

file per account. These individual files are merged and cleaned in an Extract–Transform–Load step, yielding a consolidated dataset of 213 249 tweets. Four institutional accounts (corporate and academic broadcast feeds) present in the raw download are excluded at this stage, reducing the initial 101 accounts of Figure 1 to the 97 individual accounts of Table 1. All items are retained, including original tweets, retweets, replies, and quoted tweets; each retained item is therefore one tweet-level observation, so the index measures the intensity of discourse, rather than only unique original content. Replies and quoted posts are classified on their own text. Figure 1 illustrates the full pipeline.

2.2 Semantic Classification via a Large Language Model (LLM)

The original Baker et al. (2016) approach classifies a newspaper article as policy-uncertainty-related if it contains at least one keyword from each of three pre-specified dictionaries: an "economic" list D_E (e.g., "economy", "inflation"), a "policy" list D_P (e.g., "Congress", "legislation"), and an "uncertainty" list D_U (e.g., "uncertain", "risk").

Formally, let x_a denote the raw text of article a (a string of tokens) and let $\text{tokens}(x_a)$ be the multi-set of words it contains. Let $w_E \in D_E$ denote any word belonging to the economic dictionary, and define $w_P \in D_P$, $w_U \in D_U$ analogously. The Baker–Bloom–Davis (BBD) classifier assigns a binary label $z_a^{\text{BBD}} \in \{0, 1\}$ to article a:

$$z_a^{\text{BBD}} = \mathbf{1}\{\exists\, w_E \in D_E,\, w_P \in D_P,\, w_U \in D_U : w_E, w_P, w_U \in \text{tokens}(x_a)\}, \tag{1}$$

so $z_a^{\text{BBD}} = 1$ when the article contains at least one representative term of each of the three dictionaries simultaneously. This deterministic rule misses synonyms, idiomatic expressions, irony, and context-dependent meaning.

We replace this rule with a semantic classifier based on DeepSeek-V3. Let x_{ijt} denote the text of the j-th tweet posted by user i on day t, where $i = 1, \ldots, 97$ indexes the X accounts, t indexes calendar days in the sample window, and $j = 1, \ldots, n_{it}$ indexes that user's tweets on day t. The classifier returns a triple

$$z_{ijt} = (E_{ijt},\, P_{ijt},\, U_{ijt}) \in \{0, 1\}^3, \tag{2}$$

where $E_{ijt} = 1$ indicates that tweet (i, j, t) has economic content, $P_{ijt} = 1$ indicates political content, and $U_{ijt} = 1$ indicates the expression of uncertainty. The three dimensions are classified independently: for each tweet, three separate expert-level prompts are evaluated, each producing a binary output. Figures 2–3 illustrate the prompt structure and the three-dimensional labelling. We adopt DeepSeek-V3 for three practical reasons: strong multilingual instruction-following performance, including Spanish; an open-weights release that makes the classification step replicable; and an inference cost low enough to label the full corpus along the three dimensions—approximately 160 United States dollars (USD)—a small fraction of the cost of comparable proprietary models at the time of classification.

Table 2 reports the eight possible (E, P, U) label combinations with their frequencies, and Table 3 summarizes the cleaned sample. Of the 213 249 tweets posted by the 97 accounts between January 2018 and January 2023, about 38% are unrelated to economics, politics, or uncertainty, while close to half carry political content. Uncertainty in Peruvian elite discourse is overwhelmingly political rather than economic: the political-uncertainty cell $(0, 1, 1)$ alone contains 59 978 tweets, against only 1 318 tagged as purely economic uncertainty $(1, 0, 1)$. The strict EPU label $(1, 1, 1)$, which requires the simultaneous presence of all three dimensions, is comparatively rare,

16 451 tweets, 7.7% of the corpus, and this sparsity is precisely what motivates the complementary political-uncertainty index built from the $P \cap U$ cells.

The strict EPU count $Y_t = \sum_{i=1}^{97} \sum_{j=1}^{n_{it}} \mathbf{1}\{E_{ijt} = 1, P_{ijt} = 1, U_{ijt} = 1\}$ is positive on 1 462 of the 1 835 days, with a conditional mean of 11.25 tweets per active day. Daily tweet volume N_t is strongly right-skewed: its mean (116.21) far exceeds its median (56), and the maximum is 1 016 tweets in a single day, reflecting sharp bursts of activity around political–economic events. The index is thus well populated yet concentrated on event-driven spikes, consistent with the state-dependent behavior documented below.

3 Index Construction

3.1 Tweet-level events and daily counts

We observe $I = 97$ elite accounts over $T = 1\,835$ days (23 January 2018–31 January 2023). Let n_{it} be account i's tweets on day t, $N_t = \sum_{i=1}^{I} n_{it}$ the daily volume, and $\sum_t N_t = 213\,249$ the corpus total. Each tweet carries a binary triple $(E_{ijt}, P_{ijt}, U_{ijt})$ for economic, political, and uncertainty content; for any event A on the triple write $Y_t^A = \sum_{i=1}^{I} \sum_{j=1}^{n_{it}} \mathbf{1}\{A\}$. Three events organize the family: the strict intersection $E \cap P \cap U$ (count Y_t), economic uncertainty $E \cap U$ (Y_t^{EU}), and political uncertainty $P \cap U$ (Y_t^{PU}), containing 16 451, 17 769, and 76 429 tweets respectively (Tables 2 and 4).

3.2 The index family

Strict EPU rate (1). The starting point is the daily share of tweets that are simultaneously economic, political, and uncertainty-laden,

$$\mathrm{EPU}_t^{\mathrm{raw}} = \frac{Y_t}{N_t}, \tag{3}$$

the social-media analogue of Baker et al. (2016); it is the raw ingredient from which the headline index EPU_t is built (Section 3); the analysis of Section 4 uses that headline index, not this raw rate. It is specific but sparse: only 7.7% of tweets satisfy it and the count is positive on 1 462 of 1 835 days (Figure 4), hugging a low baseline with isolated bursts rather than a drifting trend.

Economic and political rates (2&3). The strict event $E \cap P \cap U$ is the overlap of two broader, overlapping events—economic uncertainty $E \cap U$ and political uncertainty $P \cap U$—with rates $\mathrm{EU}_t = Y_t^{EU}/N_t$ and $\mathrm{PU}_t = Y_t^{PU}/N_t$; the two are supersets of the strict rate, not disjoint parts of it. Figure 5 overlays them: the political rate dominates throughout (76 429 vs. 17 769 tweets) and carries most of the variation. This is the corpus's central descriptive fact—Peruvian elite uncertainty is voiced in political far more than economic terms—and it motivates carrying PU_t forward in Section 4.6. Figure 7 previews this: across six dated crises the strict EPU rate rises clearly and persistently above its full-sample average only at the COVID-19 shock—the political episodes produce at most brief, marginal crossings of that average—whereas every political episode registers strongly in PU_t.

Union. The union rate $(Y_t^{EU} + Y_t^{PU} - Y_t)/N_t$ counts any economically *or* politically uncertain tweet (Figure 6); inclusion–exclusion is exact because $(E \cap U) \cap (P \cap U) = E \cap P \cap U$. With PU_t at 98.3% of the union, the curve nearly coincides with it and is reported only as context.

3.3 Normalization and the fragility of the raw rate

Normalization by N_t is necessary but insufficient. The raw count Y_t (Figure 8) is near zero before 2020 and peaks at 82 tweets (19 September 2022), tracking platform growth—median 14-day volume rises from $\approx$ 210 tweets in 2018 to $\approx$ 3 900 in 2022—rather than uncertainty; dividing by N_t removes this confound. But the daily rate then carries its own defect. With $Y_t \sim \mathrm{Binomial}(N_t, \pi_t)$, the estimator $\widehat{\pi}_t^{\mathrm{raw}} = Y_t/N_t$ has variance $\pi_t(1-\pi_t)/N_t$, which increases as $N_t \to 1$: the sample maximum of (3) equals 1.0 on a single-tweet day (3 February 2019, $N_t = 1$), and all four days with $\widehat{\pi}_t^{\mathrm{raw}} \geq 0.5$ have $N_t \leq 4$—pure denominator artifacts.

We therefore prefer the volume-weighted 14-day rolling rate[1],

$$\widehat{\pi}_t \;=\; \frac{\sum_{s=t-W+1}^{t} Y_s}{\sum_{s=t-W+1}^{t} N_s}, \qquad W = 14, \tag{4}$$

reported with the binomial band $\widehat{\pi}_t \pm 1.96\,\sqrt{\widehat{\pi}_t(1-\widehat{\pi}_t)/\sum_s N_s}$.[2] Pooling volume across the window keeps the denominator from collapsing while preserving invariance to platform growth. Figure 9 contrasts the raw rate with (4) and an empirical-Bayes shrinkage alternative: the 2019 artifact vanishes and the maximum relocates to the genuine April 2020 COVID-19 episode.

3.4 Standardization

Following Baker et al. (2016) we rescale (4) to base 100,

$$\mathrm{EPU}_t^{\mathrm{std}} \;=\; 100 + \frac{f_t - \mu_{\mathrm{base}}}{\sigma_{\mathrm{base}}}, \qquad f_t = 100\,\widehat{\pi}_t, \tag{5}$$

an affine, information-preserving rescaling of the rate: a shifted z-score that subtracts the base mean μ_{base} and divides by the base standard deviation σ_{base}. This differs from the Baker et al. (2016) normalization, which divides the rate by its mean so that readings are percentages of the average; ours instead *de-means* and scales by the standard deviation, fixing the variance at one, so a reading is interpreted in standard-deviation units (a value of 102 is two standard deviations above average uncertainty). Two cautions follow: the daily standardization (Figure 10) inherits thin-day fragility and should not be used, whereas the rolling version (Figure 11) is the headline series, on which the only sustained elevation is the 2020 quarantine plateau; and because 2018 is the thinnest and most crisis-laden year, we standardize on the full sample rather than a 2018 base.

[1] $W = 14$ balances precision against resolution: a stable rate needs the expected count $\widehat{\pi}\sum_s N_s \gtrsim 10$, i.e. $\sum_s N_s \gtrsim 130$ tweets. In the sparse 2018–2019 sample only 32% of 7-day windows clear this, against 97% at $W = 14$, at negligible cost to the April 2020 peak ($0.214 \to 0.199$). For the first $t < W$ days the window is truncated to the available days $s = 1, \ldots, t$ (an expanding window), so $\widehat{\pi}_t$ is defined from the first observation.

[2] This band treats the pooled 14-day window as a single binomial draw and is therefore indicative: it ignores both the overlap of successive windows and the clustering of classifications within an account on a given day, so it understates the true sampling uncertainty. A clustered or heteroskedasticity- and autocorrelation-consistent (HAC) standard error widens it but leaves the qualitative picture unchanged.

3.5 Choosing the working index

Table 5 weighs the candidates. The working index is the volume-weighted rolling rate (4), standardized via (5) on the full sample, for three reasons. (i) It is a *rate*, not a count: dividing the daily uncertainty tweets Y_t by the volume N_t strips out the secular growth in posting and measures intensity rather than platform size. (ii) It is *rolling*: the raw daily ratio Y_t/N_t is a binomial proportion with variance $\pi_t(1-\pi_t)/N_t$ that explodes on thin days (a single tweet yields a spurious rate of one, the 3 February 2019 artifact), whereas pooling numerator and denominator over fourteen days, $\widehat{\pi}_t = \sum_s Y_s / \sum_s N_s$, stabilizes the denominator while preserving real episodes such as the 2020 plateau. (iii) It is *standardized* to mean 100 and unit variance: de-meaning removes the level of the share, which depends on the labeling threshold and carries no economic meaning, and scaling by the standard deviation makes the index scale-free and directly comparable to the macro–financial series, so a value of 102 reads as "two standard deviations above normal uncertainty." We standardize on the full sample rather than a thin, crisis-laden 2018 base, which would give an unstable variance normalizer.

4 Statistical Analysis of the Index

This section studies the headline index of Section 3 and measures it against the macro–financial block. We work throughout with the volume-weighted 14-day rolling strict rate (4), standardized on the full sample to base 100 (5); we do *not* use the raw daily rate Y_t/N_t, whose maximum is the single-tweet artifact of 3 February 2019. Writing n_{it} for the tweets of account $i \in \{1, \dots, 97\}$ on day t and $N_t = \sum_i n_{it}$ for the daily volume,

$$Y_t = \sum_{i=1}^{97} \sum_{j=1}^{n_{it}} \mathbf{1}\{E_{ijt} = 1,\, P_{ijt} = 1,\, U_{ijt} = 1\}, \qquad \widehat{\pi}_t = \frac{\sum_{s=t-13}^{t} Y_s}{\sum_{s=t-13}^{t} N_s}, \qquad \mathrm{EPU}_t = 100 + \frac{100\widehat{\pi}_t - \mu}{\sigma}, \tag{6}$$

with (μ, σ) the full-sample mean and standard deviation of $100\widehat{\pi}_t$, so EPU_t has sample mean 100 and unit variance and peaks at the April 2020 COVID-19 plateau (Figure 12). We measure $\{\mathrm{EPU}_t\}_{t=1}^{T}$, $T = 1\,835$, against six macro–financial comparators, asking how strongly it co-moves with each, at what lead or lag, and in which direction predictability runs—over the full sample and, in Section 4.6, conditional on a financially defined crisis regime.

4.1 Macro–financial series

All six comparators are observed at daily frequency over $t = 1, \dots, T$. Asset prices enter as log returns, $r_t = \ln(P_t/P_{t-1})$, where P_t denotes the asset's closing price (or index level) on day t; the log ratio measures the continuously compounded one-day percentage change in the asset's value, so that, e.g., $r_t = 0.01$ corresponds to a gain of approximately 1% between days $t-1$ and t. Table 6 collects their definitions and sources; the comparators fall into two blocks. Because the index is defined on every calendar day while several comparators (the Chicago Board Options Exchange (CBOE) Volatility Index (VIX), the equity indices, the exchange rate, and the sovereign yield) trade only on business days, we align all series to the common calendar-day grid by linearly interpolating each comparator's *level* across non-trading days[3] and computing returns and realized volatilities

[3] Linear interpolation fills each non-trading day with the value lying on the straight line between the last observed close before the gap and the first observed close after it: if the market closes at P_{t_0} and reopens k days later at P_{t_0+k},

on the interpolated levels; this fixes the sample at $T = 1\,835$ for every series. Interpolation injects near-zero increments on non-trading days, which mildly attenuates measured return volatility and inflates short-horizon autocorrelation. It also carries a timing cost: an interpolated non-trading-day level is a function of the *next* trading-day close, so lagged values measured on such days embed information that was not yet available in real time, injecting a mild look-ahead into the dynamic tests; the Granger results below should therefore be read as statements on the calendar-day alignment. Sections 4.5 and 4.6 therefore re-estimate the key tests on the business-day (Monday–Friday) subsample, which removes the interpolated weekend observations that account for the bulk of non-trading days (holidays, a second-order share, remain interpolated).

(i) International risk and return.

- VIX_t — the Chicago Board Options Exchange (CBOE) Volatility Index, an observed implied-volatility level (% annualized) entering in levels; the benchmark gauge of global risk aversion (Figure 13).
- $r_t^{\mathrm{msci}} = \ln(P_t^{\mathrm{msci}}/P_{t-1}^{\mathrm{msci}})$ — daily log return of the Morgan Stanley Capital International (MSCI) Latin America index P_t^{msci}; a regional risk factor (Figure 14).

(ii) Domestic market.

- $r_t^{\mathrm{bvl}} = \ln(P_t^{\mathrm{bvl}}/P_{t-1}^{\mathrm{bvl}})$ — daily log return of the Standard & Poor's / Lima Stock Exchange (Bolsa de Valores de Lima, BVL) Perú General Index P_t^{bvl} (Figure 15).
- $\sigma_t^{\mathrm{bvl}} = \left(\frac{1}{6}\sum_{k=1}^{7}(r_{t-k}^{\mathrm{bvl}} - \bar{r}_t^{\mathrm{bvl}})^2\right)^{1/2}$, with $\bar{r}_t^{\mathrm{bvl}} = \frac{1}{7}\sum_{k=1}^{7} r_{t-k}^{\mathrm{bvl}}$ — its seven-day realized volatility, the rolling sample standard deviation of the previous seven daily returns (Figure 16).
- $r_t^{\mathrm{fx}} = \ln(e_t/e_{t-1})$ — daily log return of the exchange rate e_t (Peruvian soles, PEN, per USD); $r_t^{\mathrm{fx}} > 0$ is a depreciation of the sol (Figure 17; the level e_t, which is $I(1)$, is shown in Figure 18).
- b_t — ten-year sovereign (*soberano*) yield (% per annum), the market price of sovereign risk; being $I(1)$, it enters the dynamic models in first differences Δb_t (Figure 19).

4.2 Descriptive statistics and stationarity

Moments. Table 7 reports the moments. The index is smooth and persistent rather than spiky: centered at 100 with unit variance, it ranges only over [97.7, 104.0], is mildly right-skewed (skewness 0.52) and modestly leptokurtic (excess kurtosis 0.81), with first-order autocorrelation $\widehat{\rho}_1 = 0.979$—a slow-moving *level* of uncertainty whose single sustained elevation is the 2020 plateau (Figures 12 and 20); its broadly symmetric, near-Gaussian shape is confirmed by the histogram and kernel density in Figure 21. The three return series are by contrast near-white ($\widehat{\rho}_1 \lesssim 0.10$) and sharply leptokurtic (excess kurtosis 13–19); we note this only for contrast. The conditional-variance analysis of Section 5 concerns the index's *own* innovations—the AR(1) residual of the EPU index, which is itself heavy-tailed once its conditional variance is modeled—not the unconditional moments of these return series.

the imputed level on day $t_0 + j$ $(0 < j < k)$ is $P_{t_0} + \frac{j}{k}(P_{t_0+k} - P_{t_0})$, i.e., the gap is bridged in equal daily steps.

Why this matters? Before relating the index to the comparators in a vector autoregression we must classify each series as either stationary, $I(0)$—a stable mean it reverts to—or integrated, $I(1)$—a level that wanders without returning, as a random walk does. Regressing across an $I(1)$ level and stationary variables yields spurious results (Granger y Newbold, 1974), so every $I(1)$ series must enter in first differences (Engle and Granger, 1987). The return series need no such treatment: being (log) first differences of prices, they are stationary whenever log prices are integrated of order one, the standard case—a presumption the unit-root tests below confirm rather than replace. The realized-volatility series is a rolling seven-day standard deviation of those stationary returns, hence stationary as well, as the unit-root tests confirm; none of these must be differenced again. Only a genuine level can carry a unit root.

Tests and verdict. We use three unit-root tests (Table 8). The augmented Dickey–Fuller (ADF) regression (Dickey and Fuller, 1979)

$$\Delta x_t = \alpha + \rho\, x_{t-1} + \sum_{j=1}^{p} \phi_j\, \Delta x_{t-j} + \varepsilon_t, \qquad H_0 : \rho = 0, \tag{7}$$

tests the unit-root null $H_0 : \rho = 0$, with lag length p by the Akaike information criterion (AIC). The ADF-GLS test (Elliott et al., 1996) runs the same regression after generalized-least-squares detrending; this makes it the most powerful of the three against highly persistent alternatives, so its verdict carries the most weight. The Phillips–Perron Z_t statistic instead runs the *unaugmented* Dickey–Fuller regression—that is, (7) without the lagged differences Δx_{t-j}—and corrects the resulting statistic nonparametrically for serial correlation through a Newey–West long-run-variance estimator, rather than absorbing it parametrically with added lags as ADF does. A statistic more negative than the critical value rejects the unit root in favor of stationarity.

For the index, ADF and Phillips–Perron reject decisively (-3.76 and -5.54, against the 5% value -2.86), while the most powerful test, ADF-GLS, rejects only at the 10% level (-1.69 vs -1.62), not at 5% (-1.94). Because ADF-GLS is the most powerful of the three, its non-rejection at 5% counsels caution: on purely statistical grounds the evidence is mixed and the verdict may be sensitive to the deterministic specification and to structural shifts. Because the index is a bounded standardized share and cannot follow an unbounded random walk, we nonetheless treat it as a highly persistent $I(0)$ process, reading its persistence as close to the unit-root boundary. Among the comparators, the VIX, realized volatility, and the three returns are $I(0)$ on all three tests; the ten-year sovereign yield is the only $I(1)$ series and enters the dynamic models in first differences, Δb_t. For the yield, all three tests fail to reject the unit root (ADF -0.52, ADF-GLS -0.01, Phillips–Perron -0.63), and its level visibly wanders without reverting (Figure 19); we therefore classify it $I(1)$.

4.3 Co-movement against the macro–financial block

What we measure. We ask how strongly, in what shape, and with what timing the index moves with each comparator. For each x_t we compute the Pearson correlation

$$\rho_x^P = \frac{\sum_t (\mathrm{EPU}_t - \overline{\mathrm{EPU}})(x_t - \overline{x})}{[\sum_t (\mathrm{EPU}_t - \overline{\mathrm{EPU}})^2]^{1/2} [\sum_t (x_t - \overline{x})^2]^{1/2}}, \tag{8}$$

which captures *linear* co-movement; its Spearman rank counterpart ρ_x^S (equation (8) on ranks), which captures any *monotone* co-movement, so a gap between ρ_x^S and ρ_x^P would signal nonlinearity;

and the cross-correlation function (CCF) $\rho_x(k) = \text{corr}(\text{EPU}_t, x_{t-k})$, with $k > 0$ meaning x leads the index, which locates the relationship in time. These statistics are purely descriptive—correlation is symmetric and says nothing about *direction*—which is why this pass precedes the causality tests of Section 4.5.

Three facts emerge (Table 9, Figure 22). First, contemporaneous co-movement is essentially a single relationship (Figure 23): the index co-moves with risk—most strongly with global risk through the VIX ($\rho^P = +0.44$, $\rho^S = +0.41$) and, far more weakly, with domestic equity-market risk through realized BVL volatility (+0.20)—while the equity, currency, and regional returns and the sovereign-yield change are contemporaneously near zero. Second, that relationship is monotone and close to linear: $\rho^P \approx \rho^S$ for the leading comparators, so rank and linear correlation agree and no transformation is needed. Third, the lead–lag profile is dominated by persistence rather than by a sharp timing: the CCF (Figure 24) is positive at every displayed lag and rises monotonically across the whole ±15-day window, from $\approx +0.32$ at $k = -15$ through $\approx +0.44$ at $k = 0$ to $\approx +0.54$ at $k = +15$, with no isolated peak. This broad, slowly rising shape is the signature of two highly persistent series ($\widehat{\rho}_1(\text{EPU}) = 0.979$) and is therefore only weakly informative about direction; the modest tilt toward the positive (VIX-leading) side hints that global risk moves first, but the direction of predictability is settled by the Granger tests of Section 4.5, not by the cross-correlation.

Why a single number is not enough? These full-sample coefficients are averages over very different states. The 90-day rolling correlation (Figure 25) swings from −0.64 to +0.78 against the VIX and from −0.42 to +0.39 against the BVL return, changing sign around every crisis. A single coefficient therefore conceals as much as it reveals: the index–risk link is not a stable constant but a regime-dependent one. This is precisely why the rest of the section moves beyond contemporaneous correlation to the direction of predictability (Section 4.5) and its dependence on the financial regime (Section 4.6).

4.4 Multicollinearity among the comparators

Before testing predictability we settle a practical question: are the six comparators mutually collinear, which would make their separate effects impossible to identify? We answer with the variance inflation factor $\text{VIF}_j = 1/(1 - R_j^2)$, where R_j^2 is from regressing x_j on the other five comparators, and the condition index

$$\kappa = \sqrt{\lambda_{\max}/\lambda_{\min}}$$

of the scaled regressor cross-product matrix.

The regressors are well-conditioned (Table 10): every VIF lies between 1.0 and 1.55, no eigenvalue is near zero, and $\kappa \approx 6.9$, well below the conventional threshold of 30. Restricting the comparison to the daily series removes the only potential source of collinearity, the lower-frequency block—domestic consumer price index (CPI) inflation and the policy rate of the Central Reserve Bank of Peru (Banco Central de Reserva del Perú, BCRP), together with the external US federal funds rate—which is shown for context in Figure 22 but excluded from the daily comparator set of Section 4.1; each retained comparator carries distinct information and none is a near-combination of the others. This is a statement about the contemporaneous series only: it establishes that the comparator set is not redundant, not that a joint lagged system would be well-conditioned—adjacent

lags of these highly persistent series would be strongly collinear regardless of the contemporaneous VIFs. The diagnostic does not bear on the bivariate regressions below, in which the other comparators are simply absent; it establishes only that the comparator set is not redundant.

We proceed one comparator at a time. For each x_t we fit a two-equation (bivariate) vector autoregression (VAR) in the index and x_t, within which Granger causality is the F-test that the cross-lags jointly vanish (Section 4.5). We prefer this to a single six-regressor VAR, which at five lags would be heavily over-parameterized and subject to the lag collinearity just noted. The VAR here is a reduced-form predictive device, not a structural macroeconomic system: at daily frequency no measure of real activity exists, which is precisely why the comparator set is financial, and we make no structural identification claims—Granger causality is read strictly as predictive precedence.

4.5 Predictive direction: linear and nonlinear causality

Correlation is symmetric: it cannot tell whether the index leads the macro–financial variables or merely follows them—the property that would make it useful in real time. We therefore test Granger causality in both directions, one comparator at a time (bivariate, for the parsimony argued in Section 4.4). For each comparator x_t we fit the bivariate vector autoregression (VAR) (Sims, 1980)

$$\mathrm{EPU}_t = c + \sum_{j=1}^{L} a_j \,\mathrm{EPU}_{t-j} + \sum_{j=1}^{L} b_j \, x_{t-j} + u_t, \tag{9}$$

which predicts the index from its own L lags—the coefficients a_j, capturing its persistence—and from L lags of the comparator, the coefficients b_j. The hypothesis "x does not Granger-cause the index" is $H_0 : b_1 = \cdots = b_L = 0$: once the index's own past is controlled for, the comparator's past adds nothing. The classical homoskedastic test compares the residual sum of squares of the restricted regression (own lags only, RSS_r) with that of the unrestricted one (RSS_u),

$$F = \frac{(\mathrm{RSS}_r - \mathrm{RSS}_u)/L}{\mathrm{RSS}_u/(n - 2L - 1)},$$

where $n = T - L$ is the number of usable observations after conditioning on the first L lags. Because the 14-day rolling construction induces serial dependence beyond the modeled autoregressive dynamics, we instead report the heteroskedasticity- and autocorrelation-consistent (HAC, Newey–West) Wald version of the same joint restriction $H_0 : b_1 = \cdots = b_L = 0$, $W = (R\widehat{\theta})'[R\widehat{V}_{\mathrm{HAC}}R']^{-1}(R\widehat{\theta})$, where $\widehat{\theta}$ stacks all coefficients of (9) and R selects the cross-lag block $(b_1, \ldots, b_L)$, rescaled by L to an F for comparability; the F and p-values in Table 11 are these HAC-robust values, and the same convention applies to the crisis-split tests of Section 4.6. The HAC correction yields valid inference on the coefficients of the fitted linear projection under weak dependence; the residual caveat is one of dynamic specification, not orthogonality: a finite-lag VAR only approximates the overlap-induced dynamics, so the cross-lag coefficients capture the projection at the chosen lag length and the precise horizon of predictability is blurred—we therefore read the lead/lag timing in all Granger tests below as approximate. A large statistic means the x-lags matter; the reverse direction swaps EPU and x. We use $L = 5$ lags, selected by the Akaike information criterion (AIC), and difference the one integrated comparator, the bond yield.

Over the full sample (Table 11) the result is sharply asymmetric and, on its own, modest: what little linear predictability exists runs from risk to the index—global risk through the VIX ($F = 4.11$, $p = 0.001$) and domestic equity-market volatility through realized BVL volatility ($F = 3.82$, $p =$

0.002)—while the index Granger-causes none of the six comparators ($p > 0.36$). Unconditionally the index is a follower, not a leader: it absorbs risk but moves no market. This is the right baseline, not a disappointment—it shows that any predictive content is not a permanent feature, which is what makes the regime split of Section 4.6 informative. Two robustness checks confirm the full-sample direction. First, re-running the full-sample test for $W \in \{7, 14, 21, 30\}$ confirms the smoothing choice: the risk→index direction is significant at every window—the VIX Granger-causes the index with $F = 5.43, 4.11, 2.75, 2.59$ at $W = 7, 14, 21, 30$ (the window-dependence of the crisis-subsample channels is examined in Section 4.6). Second, re-estimating on the business-day (Monday–Friday) subsample (1 311 days)—the non-anticipative alignment check announced in Section 4.1, which removes the interpolated weekend observations—leaves the conclusion intact: the VIX and realized volatility still Granger-cause the index ($F = 5.02$, $p < 0.001$ and $F = 3.39$, $p = 0.005$) and the index still Granger-causes none of the comparators (all $p > 0.21$).

As a robustness check against dependence the linear VAR cannot capture, we apply the non-parametric test of Diks and Panchenko (2006) (see also Bekiros and Diks, 2008) to the VAR-filtered residuals. Let x_t and y_t denote the (lagged) residual vectors of the candidate cause and effect, with lag orders $\ell_x = \ell_y = 1$, and let $z_t = y_{t+1}$. The null hypothesis of no nonlinear Granger causality states that x_t contains no information about z_t beyond that in y_t, i.e. $z_t \mid (x_t, y_t) \sim z_t \mid y_t$. The test targets the functional

$$q = \mathbb{E}[f_{X,Y,Z}(x_t, y_t, z_t)\, f_Y(y_t) - f_{X,Y}(x_t, y_t)\, f_{Y,Z}(y_t, z_t)]\,, \tag{10}$$

where $f_W(\cdot)$ denotes the joint density of the random vector W evaluated at the sample point; $q = 0$ under conditional independence and $q > 0$ otherwise. Each density is replaced by the local indicator estimator

$$\hat{f}_W(W_i) = \frac{(2\varepsilon_n)^{-d_W}}{n-1} \sum_{j \neq i} \mathbf{1}(\|W_i - W_j\| < \varepsilon_n)\,, \tag{11}$$

with d_W the dimension of W, $\|\cdot\|$ the supremum norm, and ε_n a bandwidth shrinking at rate $n^{-\beta}$, $\beta \in (1/4, 1/3)$. The test statistic is

$$T_n(\varepsilon_n) = \frac{n-1}{n(n-2)} \sum_i \left[\hat{f}_{X,Y,Z}(X_i, Y_i, Z_i)\, \hat{f}_Y(Y_i) - \hat{f}_{X,Y}(X_i, Y_i)\, \hat{f}_{Y,Z}(Y_i, Z_i)\right], \tag{12}$$

which, suitably studentized, is asymptotically standard normal under the null; the test is one-sided. Following Diks and Panchenko (2006), the residuals are standardized to unit variance and we set $\varepsilon_n = 1.5$. The test flags a nonlinear lead–lag relation between the index and both volatility series that the linear test misses (Table 12). We read this only as shared nonlinear dynamics, not as economic causation of global risk by a domestic text index, and do not lean on it below.

4.6 Crisis periods defined from the macro–financial variables

The tests so far treat the five years as one regime, yet Figure 25 showed the index–risk link reversing sign around every crisis. The right question is therefore not whether the index predicts markets on average but whether it does so when markets are under stress. We split the sample on the financial regime, dating it endogenously from the market itself rather than from a hand-picked news calendar. The dating rule combines three measures of market turbulence, each a seven-day transform of a daily return series: σ_t^{bvl}, the seven-day realized volatility of the BVL log return; $|r^{\text{bvl}}|_t$, the seven-day moving average of the absolute BVL log return; and $|r^{\text{fx}}|_t$, the seven-day moving average of the

absolute PEN/USD log return. Because the three series are on different scales, each is standardized over the full sample: $\tilde{z}(u_t) = (u_t - \bar{u})/\hat{\sigma}_u$ denotes the z-score of a generic series u_t, with $\bar{u}$ and $\hat{\sigma}_u$ its full-sample mean and standard deviation, so that each component has mean zero and unit variance. The financial-stress index s_t is the sum of the three z-scores, and the crisis indicator D_t flags the days on which stress lies in its top quintile:

$$s_t = \tilde{z}(\sigma_t^{\text{bvl}}) + \tilde{z}(|r^{\text{bvl}}|_t) + \tilde{z}(|r^{\text{fx}}|_t), \qquad D_t = \mathbf{1}\{s_t \geq Q_{0.80}(s)\}, \tag{13}$$

where $Q_{0.80}(s)$ is the 80th percentile of the empirical distribution of s_t over the sample and $\mathbf{1}\{\cdot\}$ is the indicator function, equal to 1 when the condition inside the braces holds and 0 otherwise. By construction D_t marks the top quintile of stress days: $T^{-1}\sum_t D_t = 20.0\%$, that is, 366 of the 1 835 days. Two of the three components, σ_t^{bvl} and $|r^{\text{bvl}}|_t$, are distinct seven-day transforms of the BVL-return window (a root-mean-square and a mean-absolute level), so equity dispersion receives more weight than currency pressure; this reflects the centrality of the BVL to domestic financial conditions. Replacing s_t with its first principal component, or down-weighting the equity block, is a natural robustness check on the dating. A second, more important caveat concerns endogeneity: because s_t loads on BVL volatility and $|r^{\text{bvl}}|$, conditioning on D_t when the target is the BVL return links regressor and target mechanically through the second moment. The clean check—an important robustness exercise for the state-dependent result—re-dates crises from currency pressure and the VIX alone, excluding every BVL component, when predicting r_t^{bvl}. Carrying it out, the result survives: the two dating rules agree on 89% of days, and under the BVL-free rule EPU_t still Granger-causes r_t^{bvl} in crisis ($F = 2.01$, $p = 0.077$) but not in calm ($F = 1.07$, $p = 0.38$), essentially identical to the baseline ($F = 2.02$, $p = 0.075$). The state-dependent predictability therefore does not arise from the BVL components entering the crisis-dating rule.

The baseline rule (13) already recovers every major episode (Table 13): the March-2020 COVID-19 crash, the November-2020 vacancy crisis (Vizcarra–Merino–Sagasti), the April–June-2021 election (Castillo vs. Fujimori) and its transition, and the 2022 inflation surge and instability before the December-2022 self-coup. The omissions are equally telling: the late-2018 referendum and the September-2019 dissolution of Congress leave almost no footprint in equity volatility or the currency and do not register. The dating thus isolates genuinely financial turmoil, not political noise. The raw indicator D_t marks its 366 stress days as 40 disjoint runs, many of them lasting only one to nine days; for the event annotation of Table 13 we merge runs separated by ≤ 7 days, which collapses them into 22 clusters, and keep the clusters of at least 10 days, yielding the 13 windows shown, which span 395 calendar days (more than the raw 366 because merging absorbs the short calm gaps between adjacent runs). The estimation subsamples of Table 14 use the raw indicator D_t itself; the crisis-subsample conditional-variance estimates of Section 5.5 use the stress runs of at least ten consecutive days, as detailed there.

Re-estimating the bivariate VAR (9) on the crisis ($D_t = 1$) and calm ($D_t = 0$) subsamples of the raw indicator (Table 14)—not on the merged annotation windows of Table 13—gives the section's main result: the index–market relationship is dormant in calm periods and comes alive in crises, in both directions that the full sample obscured. Global risk drives the index almost only under stress—the VIX Granger-causes it in crisis ($F = 6.68$, $p < 0.001$) but not in calm ($p = 0.15$)—and, more importantly, the index acquires predictive content over the local market exactly when that market is stressed: EPU_t Granger-causes the BVL return in crisis ($F = 2.02$, $p = 0.075$) but not in calm ($p = 0.43$), and the sovereign-yield channel shows the same activation ($F = 2.26$, $p = 0.048$ in crisis; $p = 0.90$ in calm). Crucially this is a dynamic effect, not a co-movement one: contemporaneous correlation does not account for it in any of the three pairs. The index–VIX correlation is +0.41 in crisis vs +0.45 in calm, the index–BVL-return correlation is +0.04 vs +0.02,

and only the index–yield-change correlation moves at all (-0.16 in crisis vs 0.00 in calm), a modest contemporaneous shift that cannot generate the lagged cross-coefficient Fs above. The index is therefore a conditionally useful indicator—uninformative about the market on average, predictive precisely in the states where prediction matters.

Two further checks establish which crisis channel is robust. Across rolling windows, the crisis EPU $\to r^{\text{bvl}}$ channel is sharpest at $W = 14$ ($F = 2.02$, $p = 0.075$) and attenuates at both shorter, noisier windows ($W = 7$: $p > 0.5$) and longer, over-smoothed ones ($W = 30$: $p > 0.5$). And under the business-day re-estimation of Section 4.5—the non-anticipative alignment that removes the interpolated weekend observations—the sovereign-yield channel strengthens (EPU_t Granger-causes Δb_t in crisis with $F = 3.02$, $p = 0.011$, against $p = 0.76$ in calm) and the crisis risk→index link remains (VIX $\Rightarrow$ EPU: $F = 4.37$, $p = 0.001$; in calm the link turns marginal, $F = 2.42$, $p = 0.034$, so "almost only under stress" is best read as "far more strongly under stress"), but the equity-return channel does not survive ($F = 0.73$, $p = 0.60$). We therefore read the sovereign-yield channel as the robust state-dependent result—it survives the BVL-free dating and the trading-day alignment—and the equity-return channel as suggestive only: marginal to begin with, specific to $W = 14$, and sensitive to the calendar-day alignment.

4.7 Summary of stylized facts

Five facts orient the modeling that follows. (i) The index is stationary but highly persistent ($\widehat{\rho}_1 = 0.98$), near-Gaussian and slow-moving, with the 2020 plateau as its only sustained elevation; in persistence it sits with the macro–financial levels, not the near-white returns. (ii) Its only material contemporaneous co-movement is with global risk (VIX, +0.44); the cross-correlation is dominated by persistence and shows no isolated peak, so the lead–lag direction is established by the Granger tests rather than read off the timing. (iii) Restricted to the daily comparators the regressor set is well-conditioned ($\kappa \approx 6.9$, all VIF < 1.55), so a joint specification would be reliable; inference nonetheless proceeds one comparator at a time, a choice of parsimony rather than a remedy for collinearity. (iv) Unconditionally, linear causality runs only from risk to the index, never the reverse, and the index moves no market. (v) Once crises are dated endogenously from equity volatility and currency pressure, the picture is state-dependent: global risk drives the index far more strongly in crisis, and the index in turn predicts sovereign-yield changes (robustly) and equity returns (more fragilely) almost only in crisis, even though contemporaneous correlation does not rise. Together these facts justify bivariate dynamics in mixed levels and differences, the conditional-variance analysis of Section 5, and the regime split that organizes the results.

5 Conditional Volatility of the Index

5.1 Why a conditional-variance model?

The descriptive analysis showed that the dispersion of the index is not constant: its 60-day rolling standard deviation rises around the COVID-19 and late-2022 episodes (Figure 20). We make this precise with the Engle autoregressive conditional heteroskedasticity (ARCH) Lagrange-multiplier (LM) test on the residuals of the first-order autoregressive, AR(1), mean equation $\text{EPU}_t = \phi_0 + \phi_1 \text{EPU}_{t-1} + \varepsilon_t$: under conditional homoskedasticity the squared residual today carries no information about its size tomorrow, so regressing $\widehat{\varepsilon}_t^2$ on five of its own lags should yield

no fit. It does: LM $= 92.9$ ($p < 10^{-8}$), decisively rejecting conditional homoskedasticity. Volatility clustering—large movements followed by large movements—is therefore a genuine feature of the uncertainty process and must be modeled, both because it is economically informative (it measures "uncertainty about uncertainty") and because, in dynamic models with conventional (non-robust) standard errors, ignoring it would bias inference. The HAC-robust tests of Section 4.5 do not require this correction; the object of this section is the conditional variance itself, as a measure, not a repair of those tests.

One caveat applies throughout this section: the headline index is a 14-day rolling rate, whose overlapping construction induces a moving-average dependence that an AR(1) mean equation removes only imperfectly, so part of the ARCH-LM rejection and of the near-integrated persistence $\widehat{\alpha} + \widehat{\beta}$ could in principle reflect the smoothing rather than genuine conditional heteroskedasticity. Two checks gauge how much. First, pre-whitening the index with an AR(14) mean equation—enough own lags to absorb the moving-average dependence of the 14-day window—attenuates but does not remove the rejection: the ARCH-LM statistic falls from 92.9 to 27.0 ($p < 10^{-4}$), and the conditional-variance estimates on the pre-whitened residuals are essentially those reported below ($\widehat{\alpha} + \widehat{\beta} = 0.997$, $\widehat{\nu} \approx 5.0$). Volatility clustering is therefore not purely a smoothing artifact. Second, on the unsmoothed daily rate ($W = 1$) the same test does not reject (LM $= 8.5$, $p = 0.13$); this is uninformative rather than contradictory, because at the daily frequency binomial sampling noise dominates the signal—the median day contains 56 classified tweets, so the sampling standard deviation of the daily rate is roughly half its mean—and such noise masks any clustering in the underlying process, which is one more reason the index is read at $W = 14$. The alternative windows $W \in \{7, 21, 30\}$ of Sections 4.5–4.6 are examined there for the Granger-causality tests only; the conditional-variance models of this section are estimated on the headline $W = 14$ index alone.

5.2 The model and what it represents

We model the AR(1) residual ε_t as $\varepsilon_t = \sigma_t \eta_t$, where η_t is an independent and identically distributed (i.i.d.) innovation with mean zero and unit variance, so that σ_t^2 is the variance of ε_t conditional on the information available at $t-1$. The conditional variance follows the generalized ARCH model, GARCH(1,1), of Bollerslev (1986),

$$\sigma_t^2 = \omega + \alpha\, \varepsilon_{t-1}^2 + \beta\, \sigma_{t-1}^2, \qquad \omega > 0,\ \alpha, \beta \geq 0, \tag{14}$$

which decomposes today's variance into a constant floor ω, a reaction to yesterday's squared shock (α, the ARCH or "news" effect), and the carry-over of yesterday's variance (β, the GARCH or "memory" effect). The sum $\alpha + \beta$ measures the persistence of volatility: values near unity imply that an episode of high variance decays only slowly. To allow the variance to react asymmetrically to the sign of the shock we add the Glosten–Jagannathan–Runkle (GJR) term of Glosten et al. (1993),

$$\sigma_t^2 = \omega + \alpha\, \varepsilon_{t-1}^2 + \gamma\, \varepsilon_{t-1}^2 \mathbf{1}\{\varepsilon_{t-1} < 0\} + \beta\, \sigma_{t-1}^2, \tag{15}$$

where the indicator activates only after negative shocks, so $\gamma > 0$ would mean that negative shocks raise volatility more than positive ones—the leverage effect familiar from equity returns. Finally, because the index's innovations are conditionally heavy-tailed, we estimate each specification under both a Gaussian and a Student-t innovation η_t; the degrees-of-freedom parameter ν measures tail thickness—small ν makes extreme draws frequent, and $\nu \to \infty$ recovers normality. Estimation is by maximum likelihood.

Before estimating, we state what to expect. For a slow-moving uncertainty level we expect high volatility persistence ($\alpha + \beta$ close to one) and fat-tailed innovations (low ν), since uncertainty arrives in clustered bursts. We do not expect a leverage effect. A rise in the index is, of course, economically adverse; but the leverage mechanism is specific to asset returns, where a price fall mechanically raises financial and operating leverage and hence future volatility. An uncertainty index has no such channel linking the sign of an innovation to next-period variance, so $\gamma \approx 0$ would be the economically coherent outcome.

5.3 Estimation results

Table 15 reports the estimates; model comparisons use the Akaike information criterion, where lower values indicate a better fit after penalizing added parameters. Three findings stand out. First, volatility is highly persistent: $\widehat{\alpha}+\widehat{\beta} = 0.992$ in the Gaussian GARCH(1,1) and 0.995 under Student-t innovations, so shocks to the variance of the index die out only gradually, mirroring the persistence of its level. Second, the innovations are strongly fat-tailed: the Student-t model estimates $\widehat{\nu} \approx 4.8$ and improves the fit massively over the Gaussian (AIC -1382 vs -1180), confirming that extreme daily moves in the index are far more frequent than a normal law allows. Third, there is no leverage of the asset-return type: the GJR term is $\widehat{\gamma} \approx 0$ and the GJR model does not improve on the symmetric GARCH-t (AIC -1380 vs -1382), exactly as anticipated for a non-return series. Because the GJR indicator isolates negative shocks, $\widehat{\gamma} \approx 0$ rules out the asset-style asymmetry directly; the mirrored form—positive uncertainty innovations raising volatility more—would surface as $\widehat{\gamma} < 0$, and the unconstrained estimate is likewise indistinguishable from zero. The fitted conditional standard deviation is shown with $\pm 1.96\,\widehat{\sigma}_t$ bands in Figure 26—the interval that would contain 95% of the residuals under normality—which here contain 94.6%; the quantile plot (Figure 27) shows the Student-t law matching the central quantiles, with mildly asymmetric extreme tails, and the news-impact curves (Figure 28)—the variance response σ_t^2 traced as a function of the previous shock ε_{t-1}—are coincident for positive and negative shocks, confirming the absence of sign asymmetry in the variance response. The preferred specification is the symmetric GARCH(1,1) with Student-t innovations.

5.4 Volatility in crisis periods

We next ask whether the conditional volatility of the index is higher during the market-stress crises of Section 4.6. Using the crisis indicator D_t of (13), the average fitted conditional standard deviation $\widehat{\sigma}_t$ from the GARCH-t model is in fact lower in crisis than in calm (0.15 vs 0.20). This is not a contradiction but a clarification: the index's elevation during the COVID-19 episode is a sustained shift in its level, not a burst of day-to-day variance. A measurement mechanism explains the direction of the gap: the index is a rate, whose sampling precision rises with the volume of classified tweets, and volume surges under stress—its median 14-day total is 2 303 in crisis against 488 in calm. The link is tight: regressing the log conditional variance of the index on log tweet volume yields $R^2 = 0.60$ with a negative elasticity, so most of the variation in the index's volatility measures how precisely the rate is estimated, not how turbulent the underlying uncertainty is—a caveat that applies to any text-based rate index. The volatility of uncertainty and the stress of the market are thus distinct objects: the index becomes predictively linked to the market in crises (Section 4.6) without itself becoming more volatile then.

Formally, the crisis–calm contrast of this and the preceding section is a two-regime model with

an observable regime—the indicator D_t of (13). The natural generalization lets the regime be latent and evolve as a Markov chain, as in the regime-switching framework of Hamilton (1989) and its conditional-variance extensions (Hamilton and Susmel, 1994; Haas et al., 2004), estimating regimes, transition probabilities, and regime-specific dynamics jointly. We leave that to future work; the observable-regime design ties the regimes directly to measurable financial stress, at the cost of taking the dating rule as given.

5.5 Do the daily risk variables drive the index's volatility?

The model so far is univariate: the index's conditional variance is explained by its own past. We now ask whether the daily risk variables of Section 4.1 add to it, augmenting the variance equation with one-period-lagged exogenous regressors (a GARCH-X, the "X" denoting the exogenous block),

$$\sigma_t^2 = \omega + \alpha\,\varepsilon_{t-1}^2 + \beta\,\sigma_{t-1}^2 + \sum_k \theta_k\, z_{k,t-1}, \qquad \theta_k \geq 0, \tag{16}$$

where the z_k are non-negative volatility proxies (the VIX, the BVL realized volatility, and the absolute exchange-rate, MSCI, and bond-yield-change returns), each scaled to unit mean so that the θ_k are comparable. We estimate (16) with Student-t innovations on the full sample and, separately, on the crisis and calm subsamples (segment by segment, as the crisis windows are not contiguous; the segment convention is detailed in the caption of Table 16), each time with a parsimonious specification (the VIX alone) and the full set (Table 16).

Three results follow. First, over the full sample the external variables add nothing: every $\widehat{\theta}$ is essentially zero, the AIC rises (+1.0 with the VIX alone, +8.9 with all five regressors, against the ceiling of +10 reached when five added parameters leave the likelihood exactly unchanged), and the persistence stays at $\widehat{\alpha}+\widehat{\beta}=0.995$. Second, the calm subsample behaves identically ($\widehat{\theta}_{\mathrm{VIX}}=0.000$, $\Delta\mathrm{AIC}=+0.8$ and +8.8; $\widehat{\alpha}+\widehat{\beta}=0.990$). Third, the crisis subsample—too small and fragmented for precise variance estimation, with 188 usable days in 13 segments—points the same way: $\widehat{\theta}_{\mathrm{VIX}}\approx 0$ and no fit improvement ($\Delta\mathrm{AIC}=+1.4$ and +8.5), with weakly identified tail and persistence parameters (the baseline sits at the $\alpha+\beta$ boundary with a near-Gaussian $\widehat{\nu}$). An earlier draft reported a large crisis VIX loading with a collapse in persistence; that estimate came from a fit whose implied variance scale is inconsistent with the residuals—a unit-mean VIX loading of 0.14 would by itself contribute an average conditional variance an order of magnitude above the residual variance—and it does not survive a convergence-checked re-estimation, so we discard it.

The reverse direction is equally empty: adding the lagged index—its headline form or the unsmoothed daily rate—to the conditional-variance equations of the BVL return, the PEN/USD return, and the VIX log-change leaves every likelihood essentially unchanged, on the full sample and on the crisis segments alike (all likelihood-ratio statistics below one). Variance therefore flows nowhere through the index. Its own conditional variance is internally—indeed, largely measurement—governed (Section 5.4), and it transmits none to the market: the crisis activation documented in Section 4.6 is a property of the conditional mean, not of the variance. The index anticipates the direction of prices under stress; it does not anticipate their dispersion, which the market's own persistence and global risk already forecast. Estimating segment by segment discards the per-segment initialization observations and the runs shorter than ten days, which is why the crisis (188) and calm (1 284) subsamples do not sum to the full 1 834 filtered observations.

6 Conclusion

Daily measurement of policy uncertainty has so far rested on newspaper archives, a design that presupposes a dense, digitized press and that, in most emerging economies, delivers monthly indices at best. Social networks offer different raw material: the literature documents that platforms such as X mirror, in near real time, the state of the public debate—elite users react to policy news within hours, and Twitter-based uncertainty measures co-move closely with their newspaper counterparts in the economies where both exist. The motivation for this paper follows directly: where the press archive is thin, the timestamped discourse of the political and economic elite is itself a high-frequency record of how uncertainty is perceived, and a large language model can read it at scale.

We construct a daily index of Economic Policy Uncertainty (EPU) for Peru over 2018–2023 from the X discourse of 97 elite accounts, replacing the dictionary rule of the newspaper-based literature with a large language model that classifies each of 213 249 tweets along economic, political, and uncertainty dimensions. The design yields a replicable family of daily indices for an economy that had none, at a classification cost of about 160 USD, and the corpus delivers a descriptive finding of independent interest: Peruvian elite uncertainty is voiced overwhelmingly in political rather than economic terms, so the strict EPU measure and the broader political-uncertainty index are complements, the latter capturing the intensely political episodes that the former misses.

The econometric results form a coherent asymmetry. Over the full sample, predictability runs from financial conditions to the index: global risk and domestic equity volatility help forecast it, while the index forecasts no market—it absorbs risk rather than anticipating it. Conditioning on the financial regime overturns this in one specific direction. Within crises, dated endogenously from equity volatility and currency pressure, the index acquires predictive content over sovereign-yield changes—a channel that survives a dating rule excluding every component of the Bolsa de Valores de Lima (BVL) and a trading-day-only re-estimation—and, more fragilely, over equity returns. The conditional-variance analysis sharpens the result's location: the index's volatility is highly persistent, fat-tailed, free of leverage effects, and internally governed in every regime, with no variance flowing between the index and the market in either direction. The crisis activation is therefore a property of the conditional mean alone: under stress the index anticipates the direction in which sovereign risk is repriced, not the amount of turbulence, which the market's own dynamics already forecast.

The limitations delimit the claims. The elite account design trades representativeness for signal; the calendar-day alignment interpolates non-trading days, which is why the trading-day re-estimation accompanies every key test; the classifier, though replicable, awaits a human-coded validation on a stratified subsample; and the observable-regime crisis split takes the dating rule as given. These point to the natural extensions: a latent-regime Markov-switching specification that estimates the crisis states rather than conditioning on them, and a monthly vector autoregression with a real-activity or investment variable to test whether the uncertainty that this index measures at daily frequency carries the real effects that the literature documents at lower ones.

References

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *Quarterly Journal of Economics*, **131**(4):1593–1636.

Baker, S. R., Bloom, N., Davis, S. J., and Renault, T. (2021). Twitter-derived measures of economic uncertainty. Working paper, available at `policyuncertainty.com`.

Becerra, J. S. and Sagner, A. (2023). Twitter-based economic policy uncertainty index for Chile. *Revista de Análisis Económico*, **38**(1):41–69.

Bekiros, S. D. and Diks, C. G. H. (2008). The nonlinear dynamic relationship of exchange rates: Parametric and nonparametric causality testing. *Journal of Macroeconomics*, **30**(4):1641–1650.

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, **77**(3):623–685.

Bloom, N. (2014). Fluctuations in uncertainty. *Journal of Economic Perspectives*, **28**(2):153–176.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**(3):307–327.

Bovet, A. and Makse, H. A. (2019). Influence of fake news in Twitter during the 2016 US presidential election. *Nature Communications*, **10**(1):7.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., et al. (2020). Language models are few-shot learners. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33.

Caldara, D. and Iacoviello, M. (2022). Measuring geopolitical risk. *American Economic Review*, **112**(4):1194–1225.

Carrière-Swallow, Y. and Céspedes, L. F. (2013). The impact of uncertainty shocks in emerging economies. *Journal of International Economics*, **90**(2):316–325.

Da, Z., Engelberg, J., and Gao, P. (2015). The sum of all FEARS: Investor sentiment and asset prices. *Review of Financial Studies*, **28**(1):1–32.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, 4171–4186.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, **74**(366):427–431.

Diks, C. and Panchenko, V. (2006). A new statistic and practical guidelines for nonparametric Granger causality testing. *Journal of Economic Dynamics and Control*, **30**(9–10):1647–1669.

Elliott, G., Rothenberg, T. J., and Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica*, **64**(4):813–836.

Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, **55**(2):251–276.

Frankel, J. A. (2010). Monetary policy in emerging markets: A survey. In Friedman, B. M. and Woodford, M. (Eds.), *Handbook of Monetary Economics*, volume 3B, 1439–1520. Elsevier.

García, D. (2013). Sentiment during recessions. *Journal of Finance*, **68**(3):1267–1300.

Gentzkow, M., Kelly, B., and Taddy, M. (2019). Text as data. *Journal of Economic Literature*, **57**(3):535–574.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, **48**(5):1779–1801.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, **37**(3):424–438.

Granger, C. W. J. and P. Newbold (1974). Spurious regressions in econometrics. *Journal of Econometrics*, **2**(2), 111–120.

Granger, C. W. J. (1980). Testing for causality: A personal viewpoint. *Journal of Economic Dynamics and Control*, 2:329–352.

Haas, M., S. Mittnik, and M. S. Paolella (2004). A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics*, **2**(4), 493–530.

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

Hamilton, J. D. and R. Susmel (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, **64**(1–2), 307–333.

Hassan, T. A., Hollander, S., van Lent, L., and Tahoun, A. (2019). Firm-level political risk: Measurement and effects. *Quarterly Journal of Economics*, **134**(4):2135–2202.

Korinek, A. (2023). Generative AI for economic research: Use cases and implications for economists. *Journal of Economic Literature*, **61**(4):1281–1317.

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance*, **66**(1):35–65.

Romer, C. D. and Romer, D. H. (2010). The macroeconomic effects of tax changes: Estimates based on a new measure of fiscal shocks. *American Economic Review*, **100**(3):763–801.

Shiller, R. J. (2017). Narrative economics. *American Economic Review*, **107**(4):967–1004.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, **48**(1):1–48.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance*, **62**(3):1139–1168.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., and Welpe, I. M. (2010). Predicting elections with Twitter: What 140 characters reveal about political sentiment. In *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM)*, 178–185.

Veldkamp, L. L. (2006). Media frenzies in markets for financial information. *American Economic Review*, **96**(3):577–601.

Table 1: Complete list of X accounts (97)

Username	Name	Category
RosanaCuevaM	Rosana Cueva	Journalist
NicolasLucar	Nicolás Lucar	Journalist
PolloFarsantePe	Beto Ortiz	Journalist
MilagrosLeivaG	Milagros Leiva	Journalist
DeltaMdelta	Mónica Delta	Journalist
MavilaHuertasC	Mavila Huertas	Journalist
SolCn	Sol Carreño	Journalist
VertizPamela	Pamela Vértiz	Journalist
claudia_cisneros	Claudia Cisneros	Journalist
recisneros	Renato Cisneros	Journalist
larryportera	Paola Ugaz	Journalist
tuesta	Fernando Tuesta Soldevilla	Journalist
Federicoagust	Federico Salazar B.	Journalist
julianaoxenford	Juliana Oxenford	Journalist
mariateguiperu	Aldo Mariátegui	Journalist
JaimeChincha	Jaime Chincha	Journalist
VeroLinaresC	Verónica Linares	Journalist
tenoriopedro	Pedro Tenorio	Journalist
JBCPERU	José Barba Caballero	Journalist
jdealthaus	Jaime de Althaus	Journalist
ensustrece	Hildebrandt en sus trece	Journalist
BaylyOficial	Jaime Bayly	Journalist
jctafur	Juan Carlos Tafur	Journalist
cparedesr	Carlos Paredes	Journalist
PCaterianoB	Pedro Cateriano B.	Politician
Carlos_Bruce	Carlos Bruce	Politician
JuanSheput	Juan Sheput	Politician
VLADIMIR_CERRON	Vladimír Cerrón	Politician
anibaltorresv	Aníbal Torres V.	Politician
pedrofrancke	Pedro Francke	Politician
Ollanta_HumalaT	Ollanta Humala Tasso	Politician
sigridbazan	Sigrid Bazán	Politician
MirtyVas	Mirtha Vásquez	Politician
FlorPabloMedina	Flor Pablo Medina	Politician
rlopezaliaga1	Rafael López Aliaga	Politician
patarevalo	Patricia Arévalo	Politician
George_Forsyth	George Forsyth	Politician
JorgeMunozPe	Jorge Muñoz	Politician
DanielUrresti1	Daniel Urresti	Politician
VictorAndresGB	Víctor Andrés García Belaunde	Politician
KeikoFujimori	Keiko Fujimori	Politician
MartinVizcarraC	Martín Vizcarra	Politician
JorgeDCG	Jorge del Castillo	Politician
Mauriciomulder	Mauricio Mulder	Politician

Username	Name	Category
nidiavilchez	Nidia Vílchez	Politician
FSagasti	Francisco Sagasti	Politician
AlbertoBelaunde	Alberto de Belaunde	Politician
MaricarmenAlvaP	Maricarmen Alva	Politician
PattyChirinos1	Patricia Chirinos	Politician
adrianatudelag	Adriana Tudela Gutiérrez	Politician
CesarAcunaP	César Acuña Peralta	Politician
MesiasGuevara	Mesías Guevara	Politician
vozdelatierra	Marco Arana	Politician
MarisaGlave	Marisa Glave	Politician
GuilleBermejoR	Guillermo Bermejo Rojas	Politician
robertochiabra	Roberto Chiabra	Politician
RichardArcePeru	Richard Arce	Politician
Alm_Montoya	Jorge Montoya	Politician
AlejandroCavero	Alejandro Cavero	Politician
HectorValer_PER	Héctor Valer Pinto	Politician
RosselliAmuruz	Rosselli Amuruz	Politician
HDeSotoPeru	Hernando de Soto	Politician
UrsulaLetonaP	Ursula Letona	Politician
AlvaroVargasLl	Álvaro Vargas Llosa	Politician
AlbertoOtarolaP	Alberto Otárola	Politician
HPerezDeCuellar	Hania Pérez de Cuéllar	Politician
alosegura	Alonso Segura Vasi	Economist
BustamantePao18	Paola Bustamante S.	Economist
aethorne	Alfredo E. Thorne	Economist
WaldoMendozaB	Waldo Mendoza Bellido	Economist
JuanManuelGarc3	Juan Manuel García C.	Economist
CarlosAnderso1	Carlos A. Anderson	Economist
hugonopo	Hugo Ñopo	Economist
tuestadavid	David Tuesta	Economist
pablo_lavado	Pablo Lavado	Economist
OswaldoMolinaC	Oswaldo Molina	Economist
CParodiT	Carlos Parodi Trece	Economist
psecadae	Pablo Secada	Economist
DWinkelried	Diego Winkelried	Economist
romerocaroperu	Manuel Romero Caro	Economist
manupulgarvidal	Manuel Pulgar Vidal	Economist
BancalariA	Antonella Bancalari	Economist
eljorobado	Carlos Meléndez	Political analyst
DargentEduardo	Eduardo Dargent	Political analyst
DelaPuenteJuan	Juan De la Puente	Political analyst
lauermirko	Mirko Lauer	Political analyst
rmapalacios	Rosa María Palacios	Political analyst
Frospigliosi	Fernando Rospigliosi	Political analyst
BetoAdrianzen	Alberto Adrianzén	Political analyst
ocram	Marco Sifuentes	Political analyst

Username	Name	Category
brunoschaaf	Bruno Schaaf	Political analyst
luisbenaventegi	Luis Benavente	Political analyst
camcesar	César Campos	Political analyst
AlfredoMTorres	Alfredo M. Torres G.	Political analyst
gonza_banda	Gonzalo Banda	Political analyst
Roque_Benavides	Roque Benavides	Private sector
CayetanaAljovin	Cayetana Aljovín	Private sector

Table 2: Interpretation of binary labels (E, P, U) assigned to each tweet

E	P	U	Interpretation	Count
0	0	0	Unrelated to economics, politics, or uncertainty	81 353
1	0	0	Economic content without politics or uncertainty	2 051
0	1	0	Political content without economics or uncertainty	19 280
0	0	1	Uncertainty not linked to economics or politics	23 144
1	1	0	Economic and political content, no uncertainty	9 674
1	0	1	Economic uncertainty	1 318
0	1	1	Political uncertainty	59 978
1	1	1	Economic and political uncertainty (strict EPU)	16 451

Table 3: Descriptive statistics of the cleaned sample

Statistic	Value
Total tweets	213 249
Unique accounts	97
Date range	2018-01-23 to 2023-01-31
Calendar days	1 835
Daily N_t: Mean / Median / Max	116.21 / 56 / 1 016
Total EPU tweets $(\sum_t Y_t)$	16 451
Days with $Y_t > 0$	1 462
Y_t mean (cond. on $Y_t > 0$)	11.25
Y_t maximum	82

Event	Tweets	Share of corpus
Strict EPU $(E \cap P \cap U)$	16 451	7.7%
Economic uncertainty $(E \cap U)$	17 769	8.3%
Political uncertainty $(P \cap U)$	76 429	35.8%
Union $(E \cap U) \cup (P \cap U)$	77 747	36.5%

Table 4: Tweet-level event totals over the full corpus ($\sum_t N_t = 213\,249$). Political uncertainty is 98.3% of the union; only 1 318 tweets express economic uncertainty with no political frame.

Table 5: Candidate uncertainty series: advantage, problems, and role.

Series	Advantage	Problems	Role
Strict rate, raw daily Y_t/N_t (3)	Direct Baker et al. (2016) analogue; economically specific	Sparse (7.7%); variance $\propto 1/N_t$; spurious max = 1.0 on a single-tweet day	Rejected
Strict rate, 14-day rolling (4)	Stable denominator; growth-invariant; artifact removed, COVID signal emerges	Mild smoothing/lag; window choice	**Primary**
Standardized base 100, rolling (5)	Mean-100, unit-variance scale (cf. Baker et al., 2016)	Base-period sensitive (use full sample)	Primary (scale)
Empirical-Bayes shrinkage rate	Principled thin-day correction	Extra prior assumptions	Robustness
Economic rate EU_t	Isolates the economic frame	Almost no unique signal (1 318 tweets); little variation	Not used alone
Political rate PU_t	Most variation; lights up every political crisis (45–54%)	Not an economic-policy measure	Backup (political risk)
Union rate	Broadest coverage	98.3% redundant with PU_t	Context only

Table 6: The index and the six *daily* macro–financial comparators, observed over 23 January 2018–31 January 2023.

Symbol	Series	Definition	Source
Index			
EPU_t	EPU index	14-day volume-weighted rolling strict rate, standardized to base 100 (6)	X corpus
(i) International risk and return			
VIX_t	VIX	Implied volatility of the S&P 500 (% annualized)	CBOE
r_t^{msci}	MSCI LatAm	Daily log return of the MSCI Latin America index	MSCI
(ii) Domestic market			
r_t^{bvl}	BVL return	Daily log return, S&P/BVL Perú General Index	BVL
σ_t^{bvl}	BVL volatility	Seven-day realized volatility of r_t^{bvl}	BVL
r_t^{fx}	PEN/USD return	Daily log return, sol per US dollar	BCRP
b_t	Bond yield	Ten-year sovereign (*soberano*) yield (% per annum)	MEF/BCRP

Acronyms. EPU: economic political uncertainty; VIX: Chicago Board Options Exchange (CBOE) Volatility Index; S&P: Standard & Poor's; MSCI: Morgan Stanley Capital International; BVL: Lima Stock Exchange (Bolsa de Valores de Lima); PEN/USD: Peruvian soles per United States dollar; BCRP: Central Reserve Bank of Peru (Banco Central de Reserva del Perú); MEF: Ministry of Economy and Finance (Ministerio de Economía y Finanzas).

Table 7: Descriptive moments, full sample ($T = 1\,835$ calendar days). The index is the headline 14-day volume-weighted rolling rate standardized to base 100 (6); returns are daily log returns; the bond yield enters in first differences (percentage points). Skew. is skewness, Exc. kurt. is excess kurtosis, and $\widehat{\rho}_1$ is the first-order autocorrelation.

Variable	Mean	S.D.	Min	Max	Skew.	Exc. kurt.	$\widehat{\rho}_1$
EPU_t	100.00	1.000	97.66	104.02	0.52	0.81	0.979
VIX_t	21.36	8.284	3.81	82.69	2.32	9.91	0.976
r_t^{bvl}	−0.0001	0.0092	−0.1101	0.0541	−1.12	19.01	0.074
σ_t^{bvl}	0.0071	0.0054	0.0000	0.0489	2.79	13.34	0.923
r_t^{fx}	0.0001	0.0035	−0.0240	0.0358	−0.38	13.53	0.083
r_t^{msci}	−0.0002	0.0137	−0.1618	0.1141	−0.92	17.85	0.101
Δb_t	0.0018	0.0571	−0.3060	0.7760	3.47	46.00	0.294

Table 8: Unit-root tests (intercept, no trend). Each test classifies a series as stationary, $I(0)$, or integrated, $I(1)$: a statistic more negative than the critical value rejects the unit root in favor of stationarity, which determines whether the series enters the dynamic models in levels ($I(0)$) or in first differences ($I(1)$). Augmented Dickey–Fuller (ADF) and GLS-detrended ADF (ADF-GLS) lag length by the Akaike information criterion; Phillips–Perron is the Z_t statistic with a Newey–West long-run variance. ADF-GLS is the most powerful test against persistent alternatives. Critical values: ADF and Phillips–Perron = −2.86 (5%); ADF-GLS = −1.94 (5%), −1.62 (10%).

Variable	ADF	ADF-GLS	Phillips–Perron	Verdict
EPU_t	−3.76	−1.69	−5.54	$I(0)$
VIX_t	−4.30	−2.73	−4.52	$I(0)$
r_t^{bvl}	−26.72	−25.17	−40.36	$I(0)$
σ_t^{bvl}	−6.35	−4.19	−9.01	$I(0)$
r_t^{fx}	−39.39	−12.04	−39.44	$I(0)$
r_t^{msci}	−38.68	−4.14	−38.95	$I(0)$
b_t	−0.52	−0.01	−0.63	$I(1)$

Table 9: Contemporaneous correlation of each daily comparator with the EPU index: Pearson (linear) and Spearman (rank). The bond yield enters in first differences (Δb_t), consistent with its $I(1)$ classification in Table 8 and with the dynamic models of Section 4.5.

Variable	Pearson	Spearman
VIX_t	+0.437	+0.408
r_t^{bvl}	+0.024	+0.022
σ_t^{bvl}	+0.198	+0.214
r_t^{fx}	−0.018	+0.009
r_t^{msci}	−0.012	−0.033
Δb_t	−0.048	−0.009

Table 10: Multicollinearity diagnostics for the six daily comparators. VIF is the variance inflation factor (each comparator regressed on the other five); κ is the Belsley scaled condition index.

Variable	VIF	Variable	VIF
VIX_t	1.41	r_t^{msci}	1.50
r_t^{bvl}	1.43	Δb_t	1.12
σ_t^{bvl}	1.39		
r_t^{fx}	1.18		
All VIF < 1.55; no eigenvalue near zero; $\kappa \approx 6.9$.			

Table 11: Linear Granger causality, $L = 5$ lags, full sample. HAC-robust (Newey–West) Wald statistics rescaled to F, with p-values in parentheses. $I(1)$ variables differenced.

	$x \Rightarrow$ EPU		EPU $\Rightarrow x$	
x	F	(p)	F	(p)
VIX_t	4.11	(0.001)	0.69	(0.630)
r_t^{bvl}	1.12	(0.349)	0.63	(0.678)
σ_t^{bvl}	3.82	(0.002)	0.36	(0.874)
r_t^{fx}	0.97	(0.438)	1.09	(0.364)
r_t^{msci}	1.10	(0.361)	0.42	(0.833)
Δb_t	0.88	(0.493)	0.88	(0.495)

Table 12: Nonlinear Diks–Panchenko (10) causality on VAR(5)-filtered residuals, full sample. One-sided T-statistics ($N(0,1)$) with p-values; bandwidth $\varepsilon = 1.5$. Large positive T rejects the no-nonlinear-causality null. The exchange-rate and MSCI returns are omitted: like the BVL return and the yield change they show no linear relationship in either direction (Table 11), but unlike them they play no role in the crisis analysis of Section 4.6.

	$x \Rightarrow$ EPU		EPU $\Rightarrow x$	
x	T	(p)	T	(p)
VIX_t	0.50	(0.31)	9.15	(< 0.001)
r_t^{bvl}	-8.92	(ns)	-10.01	(ns)
σ_t^{bvl}	-8.32	(ns)	3.57	(< 0.001)
Δb_t	-9.02	(ns)	-13.23	(ns)

Table 13: Crisis windows defined endogenously from the financial-stress rule (13). The raw indicator $D_t = 1$ (top 20% of s_t, 366 days in 40 disjoint runs) is collapsed, for event annotation only, by merging runs separated by ≤ 7 days and keeping the resulting clusters of ≥ 10 days, yielding the 13 windows below (spanning 395 calendar days, more than the raw 366 because merging absorbs short calm gaps). The dates are determined entirely by the market data (σ_t^{bvl}, $|r^{\text{bvl}}|_t$, $|r^{\text{fx}}|_t$); the rule flags 20.0% of the 1 835 days. The *event* column is our annotation, matching each data-defined window to the documented episode of that period; it plays no role in the dating.

Window	Days	Corresponding event
2020-02-28 – 2020-04-13	46	COVID-19 global crash; first national lockdown
2020-04-30 – 2020-05-20	21	COVID-19 market volatility
2020-06-04 – 2020-06-18	15	COVID-19 market volatility
2020-10-30 – 2020-11-23	25	Vacancy crisis: Vizcarra–Merino–Sagasti
2021-04-08 – 2021-06-22	76	2021 presidential election (Castillo vs. Fujimori); sol depreciation
2021-07-21 – 2021-08-05	16	Castillo inauguration; cabinet uncertainty
2021-08-19 – 2021-08-31	13	Cabinet uncertainty
2021-10-08 – 2021-10-20	13	Cabinet crisis
2022-02-09 – 2022-03-20	40	Cabinet instability; onset of the inflation shock
2022-04-22 – 2022-07-22	92	Inflation peak; protests; global tightening; currency pressure
2022-08-05 – 2022-08-16	12	Political instability
2022-09-24 – 2022-10-09	16	Instability preceding the December-2022 self-coup
2022-11-05 – 2022-11-14	10	Escalating crisis before the December-2022 self-coup

Table 14: Crisis vs calm: the index–market link is *dynamic*, not contemporaneous. Upper panel: linear Granger causality ($L = 5$ lags), HAC-robust F with p-values in parentheses. Lower panel: the contemporaneous index–VIX correlation. Crises are dated endogenously from equity volatility and currency pressure (13) (20% of days).

	Crisis ($D_t = 1$)		Calm ($D_t = 0$)	
	F	(p)	F	(p)
Granger causality				
$\text{VIX}_t \Rightarrow \text{EPU}$	6.68	(0.000)	1.62	(0.153)
$\text{EPU} \Rightarrow r_t^{\text{bvl}}$	2.02	(0.075)	0.97	(0.432)
$\text{EPU} \Rightarrow \Delta b_t$	2.26	(0.048)	0.32	(0.902)
Contemporaneous correlation				
$\text{corr}(\text{EPU}_t, \text{VIX}_t)$	+0.41		+0.45	

Table 15: Conditional-variance models for the AR(1)-filtered EPU index. GARCH(1,1) (14) and GJR-GARCH (15) under Gaussian (N) and Student-t innovations. ARCH-LM(5) = 92.9 ($p < 10^{-8}$).

Model	ω	α	β	γ	ν	$\alpha+\beta$	AIC
GARCH(1,1)-N	0.0003	0.063	0.930	—	—	0.992	−1179.8
GARCH(1,1)-t	0.0003	0.080	0.915	—	4.8	0.995	−1382.3
GJR-GARCH-t	0.0003	0.080	0.915	0.000	4.8	0.995	−1380.3

Table 16: GARCH-X(1,1) with Student-t innovations (16): the index's conditional variance augmented with one-lag daily risk regressors, estimated on the full sample and on the crisis and calm subsamples. "+VIX" adds the VIX alone; "+all" adds the VIX, BVL realized volatility, and absolute FX, MSCI, and bond-yield-change returns. θ_{VIX} is the VIX loading (variables scaled to unit mean; $\alpha+\beta<1$ imposed). We report ΔAIC relative to the baseline specification *within each sample block* (baseline $\equiv 0$); lower is better, and with five added parameters an unchanged likelihood yields exactly +10, so values near +10 indicate uninformative regressors. Subsamples are estimated segment by segment on maximal runs of the raw crisis indicator (crisis) or of its complement (calm) lasting at least ten consecutive days; within each segment the variance recursion is initialized at the segment residual variance and the first five observations enter only the initialization. This leaves 188 usable crisis days (13 segments) and 1 284 usable calm days (22 segments), which do not sum to the full 1 834 filtered observations. Crisis-block tail and persistence parameters are weakly identified (boundary estimates) given the short, fragmented subsample (Section 5.5).

Sample	Spec.	α	β	$\alpha+\beta$	ν	θ_{VIX}	ΔAIC
Full (1 834)	baseline	0.080	0.915	0.995	4.80	—	0.0
	+VIX	0.079	0.916	0.995	4.77	0.000	+1.0
	+all	0.079	0.918	0.996	4.78	0.000	+8.9
Crisis (188)	baseline	0.000	0.999	0.999	24.5	—	0.0
	+VIX	0.000	0.999	0.999	24.4	0.000	+1.4
	+all	0.000	0.999	0.999	50.0	0.000	+8.5
Calm (1 284)	baseline	0.071	0.919	0.990	5.27	—	0.0
	+VIX	0.071	0.919	0.990	5.21	0.000	+0.8
	+all	0.071	0.920	0.990	5.21	0.000	+8.8

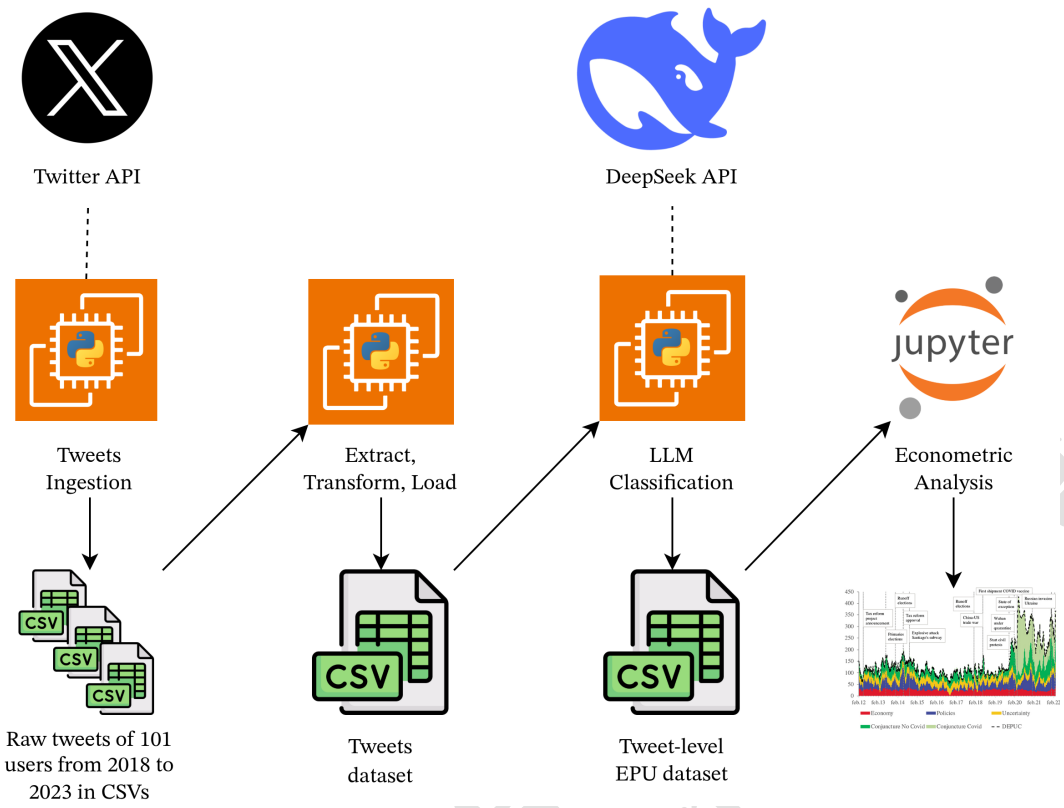

Figure 1: Pipeline architecture, from tweet ingestion to econometric analysis. The final sample comprises the 97 accounts selected by the criteria of Section 2; any larger initial count shown in the schematic is illustrative of the pre-selection stage.

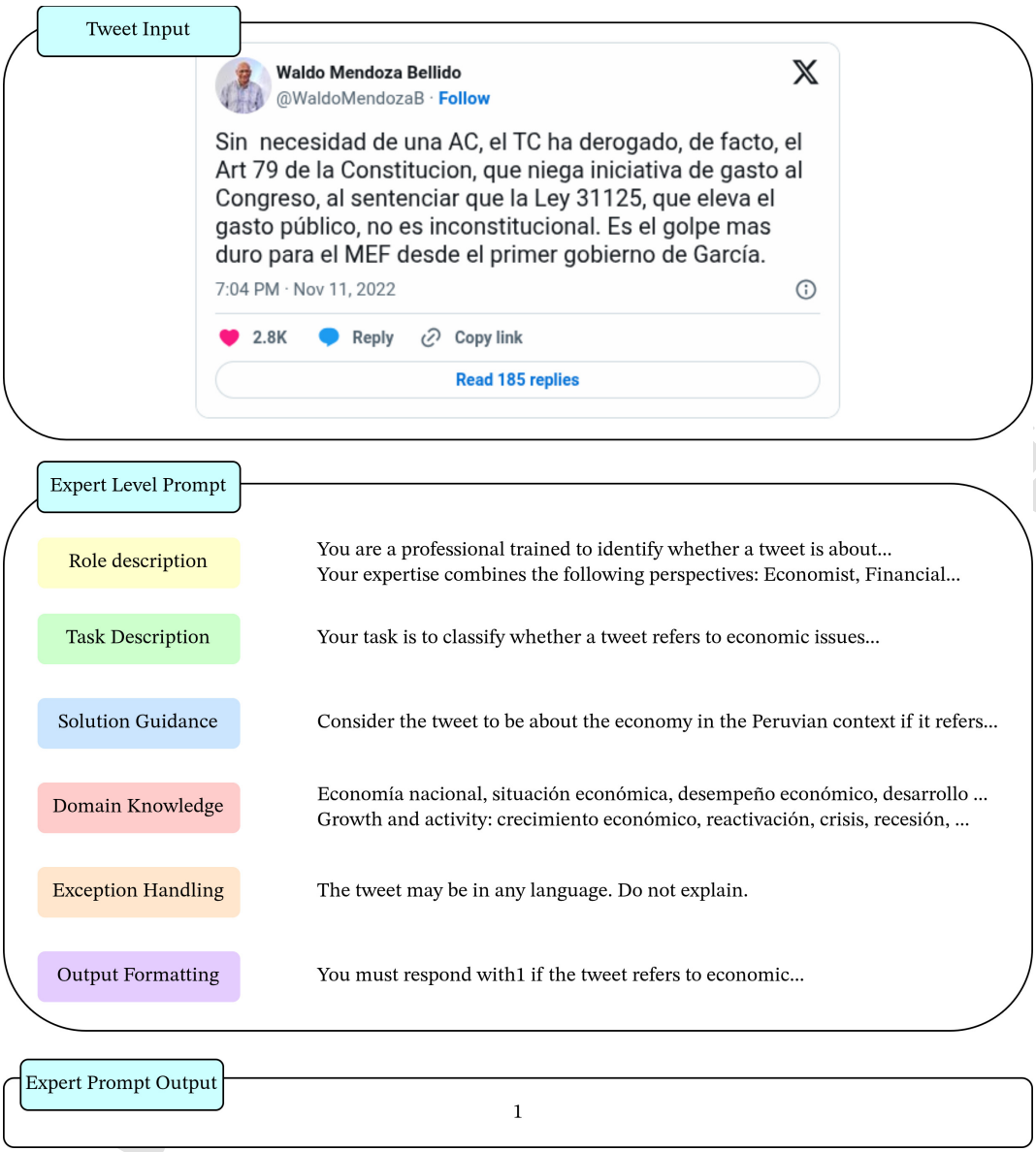

Figure 2: Expert-level prompt structure for LLM classification.

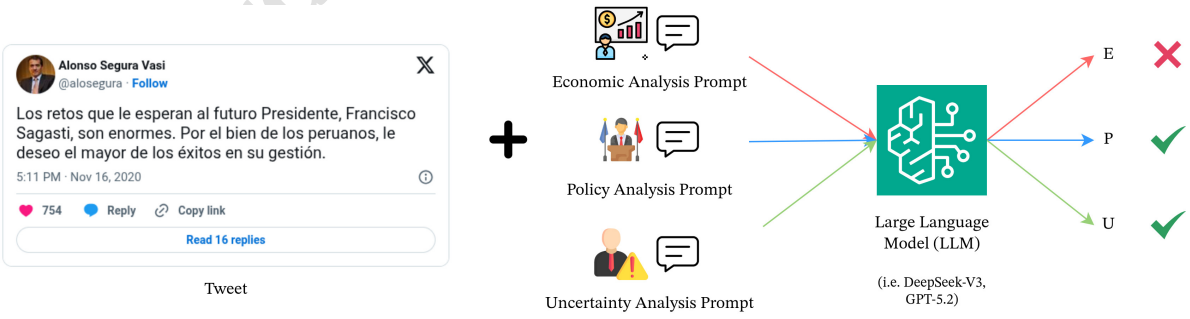

Figure 3: Three-dimensional labelling: each tweet is classified independently for E_{ijt}, P_{ijt}, U_{ijt}.

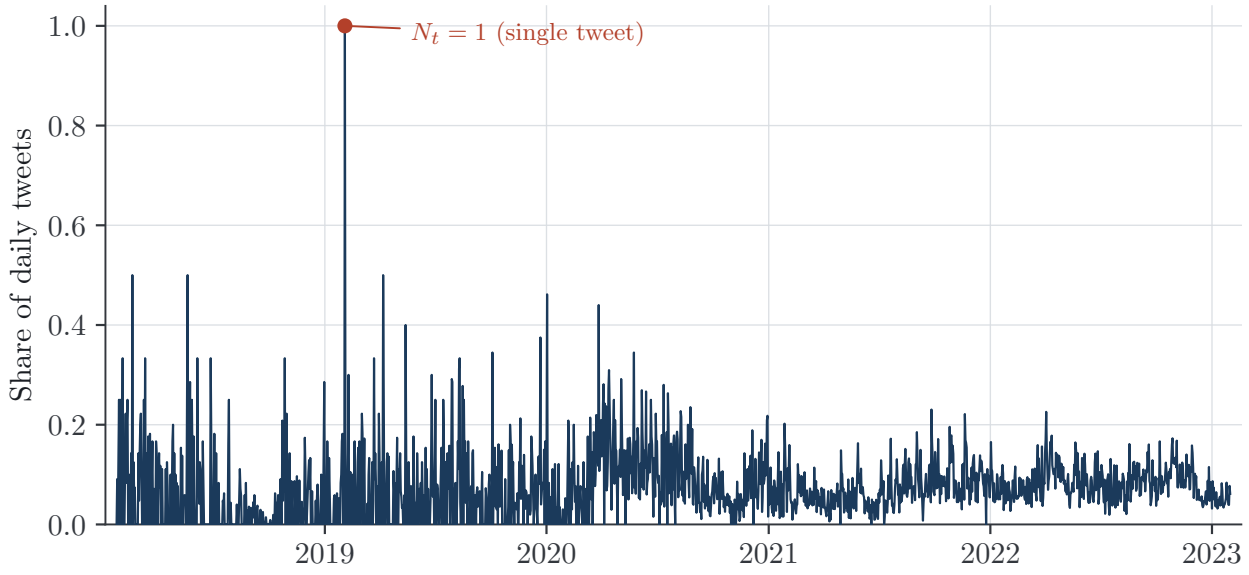

Figure 4: **EPU rate,** $\text{EPU}_t^{\text{raw}} = Y_t/N_t$, the daily share of tweets that are simultaneously economic, political, and uncertainty-laden ($E \cap P \cap U$). The maximum on 3 February 2019 is a single-tweet day ($N_t = 1$) and is a denominator artifact, not a surge; see Eq. (4) and Figure 9 for the robust construction.

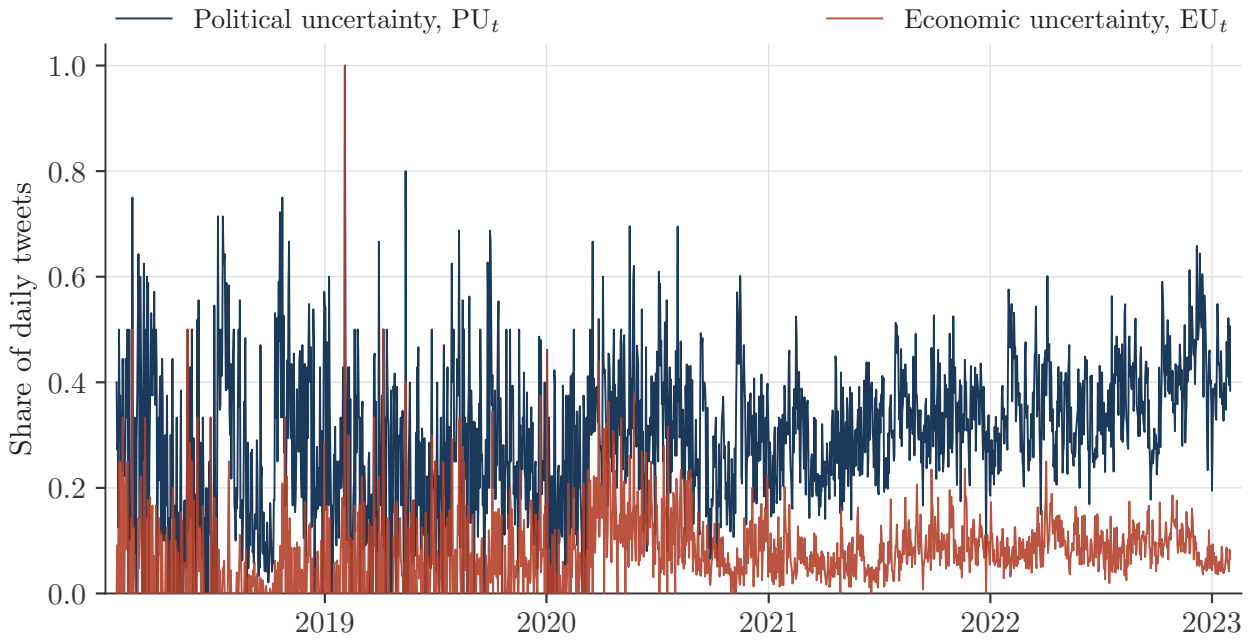

Figure 5: **Economic- and political-uncertainty rates,** $\text{EU}_t = Y_t^{EU}/N_t$ and $\text{PU}_t = Y_t^{PU}/N_t$. The political rate dominates the economic rate throughout: in this corpus uncertainty is voiced overwhelmingly in political terms.

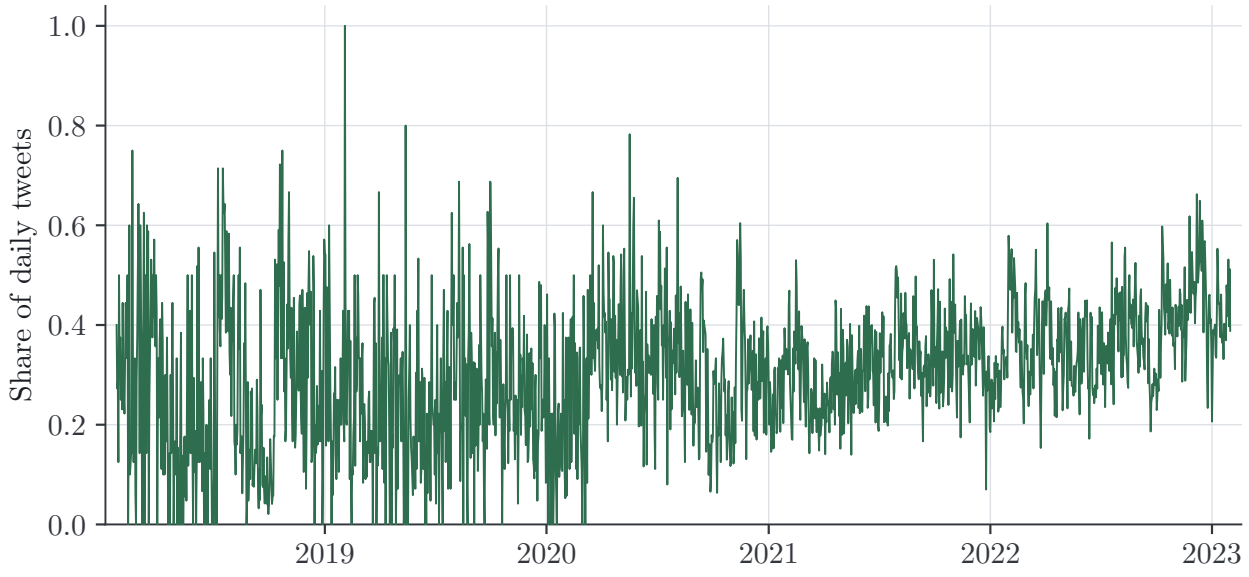

Figure 6: **Union rate,** $(Y_t^{EU} + Y_t^{PU} - Y_t)/N_t$, the share of tweets that are economically *or* politically uncertain. Because political uncertainty is 98.3% of the union, the curve nearly coincides with PU_t.

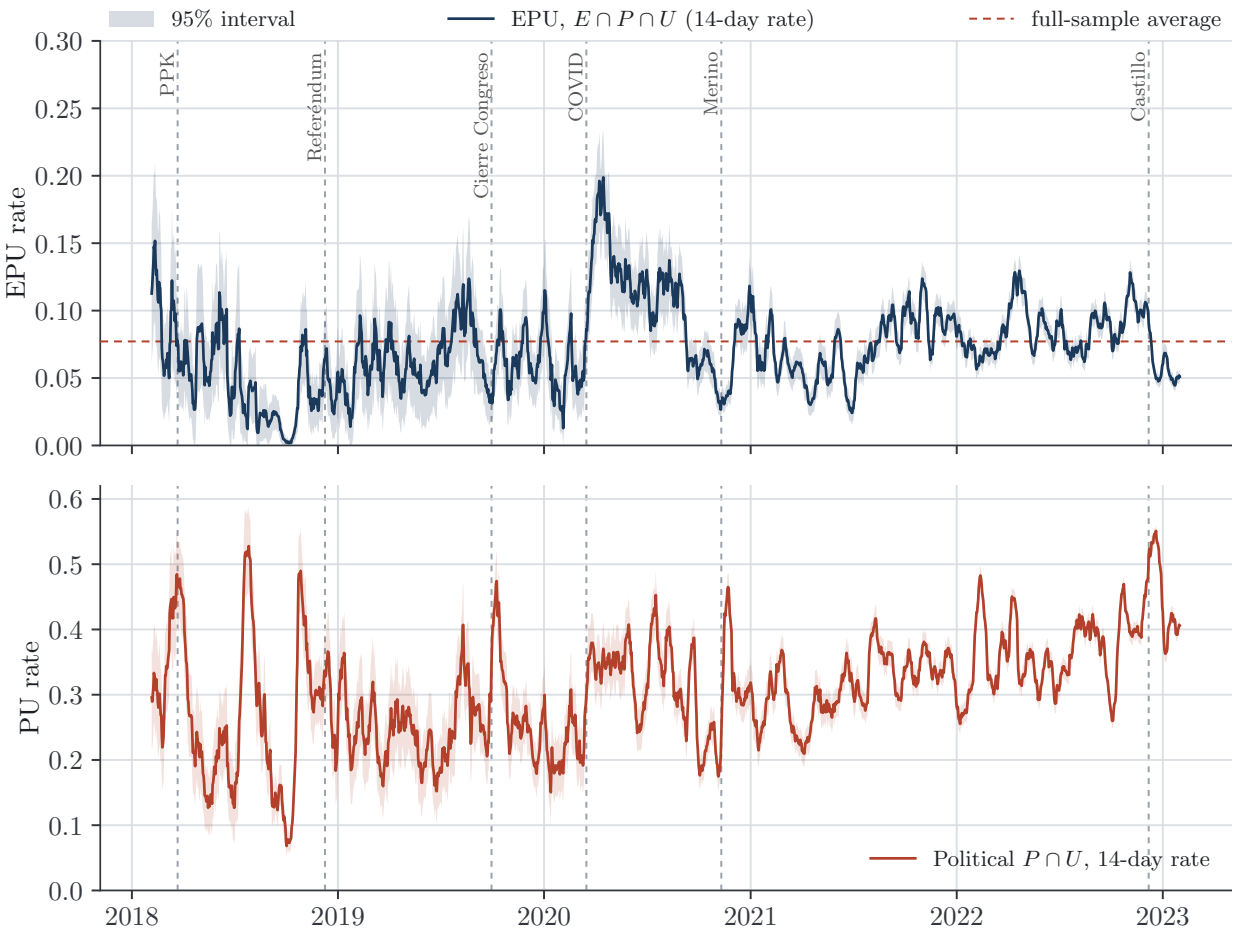

Figure 7: **Validation against Peruvian crises.** EPU rate ($E \cap P \cap U$), top, and political rate PU_t ($P \cap U$), bottom, both on a 14-day volume-weighted basis with 95% binomial bands; vertical lines mark six events. The EPU rate rises clearly and persistently above its full-sample average only at the COVID-19 shock; the five political episodes produce at most brief, marginal crossings of that average in EPU but register strongly in PU_t. Wide bands in 2018–2019 reflect the low tweet volume of the early sample.

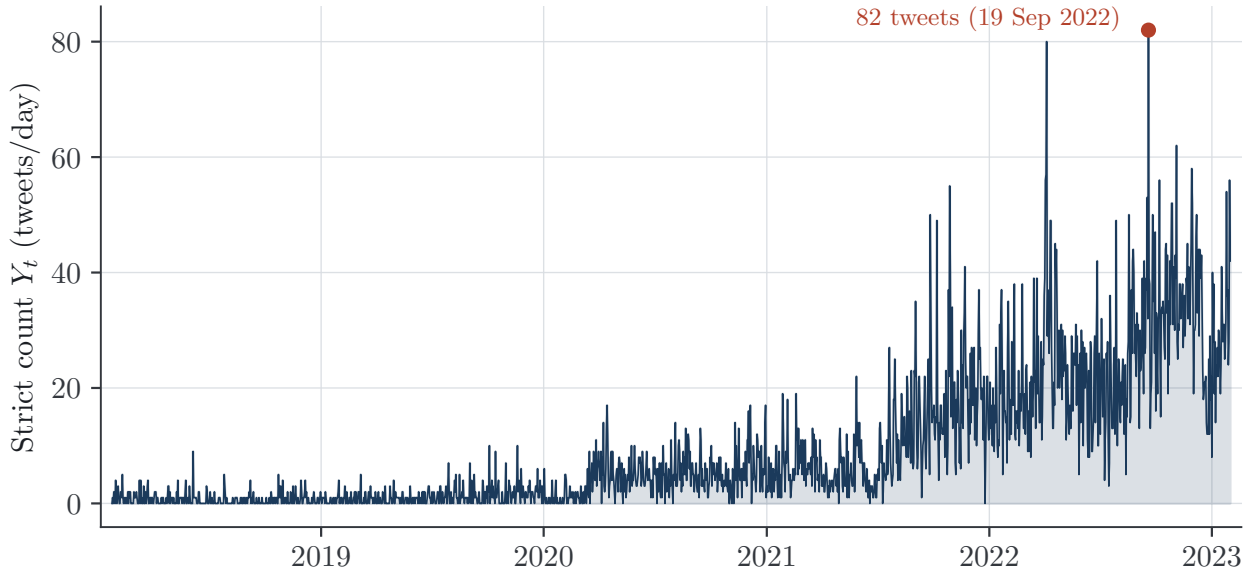

Figure 8: **Raw strict count,** Y_t, without normalization. The count is near zero before 2020 and rises sharply afterward, peaking at 82 tweets in a single day (19 September 2022), tracking the secular growth in platform activity rather than uncertainty—the confound that dividing by N_t removes.

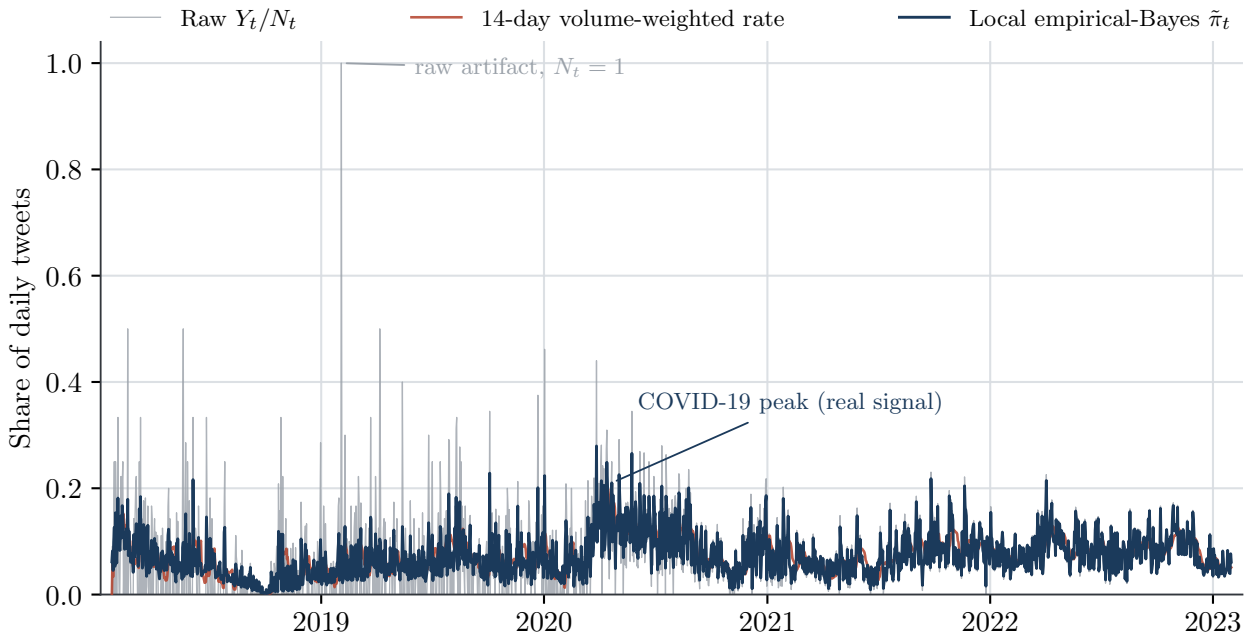

Figure 9: **Raw vs. robust EPU rate.** The raw daily rate Y_t/N_t (grey) spikes to 1.0 on the single-tweet day of 3 February 2019. The 14-day volume-weighted rate of Eq. (4) (red) and a local empirical-Bayes shrinkage estimator (navy) remove the artifact and relocate the maximum to the April 2020 COVID-19 episode, the genuine signal, while remaining invariant to platform growth.

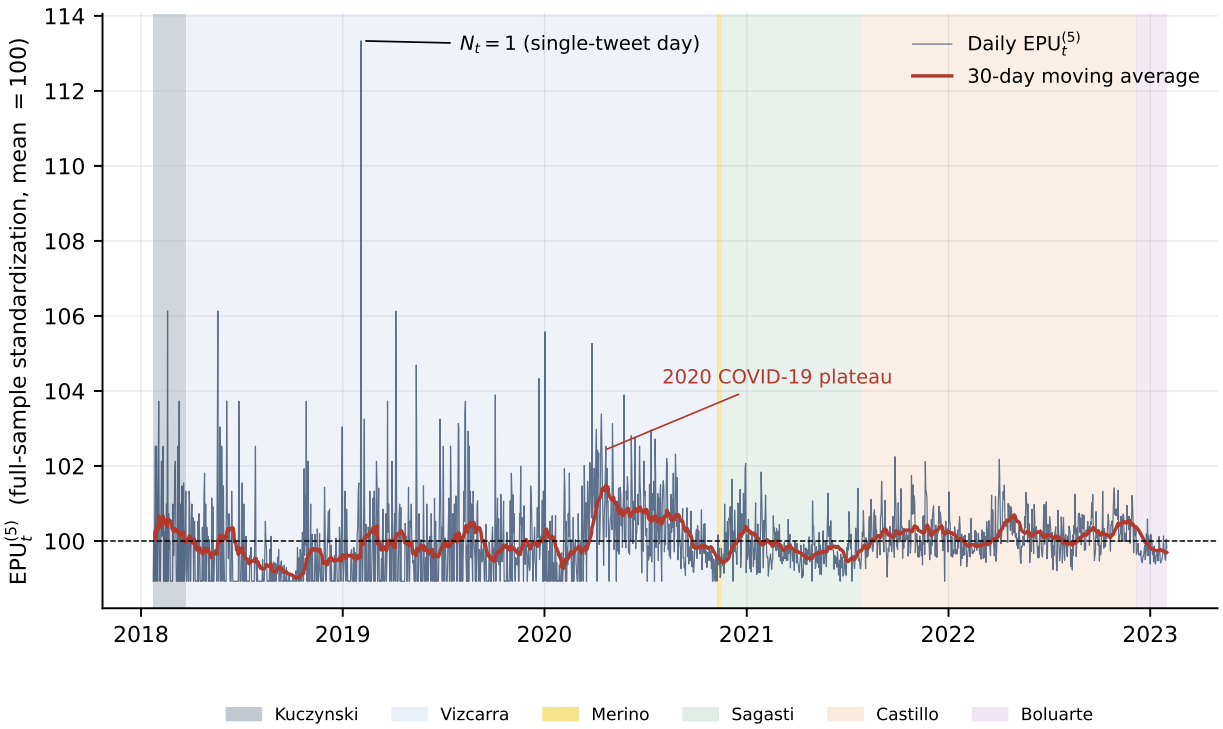

Figure 10: **Standardized EPU, raw daily basis.** Eq. (5) applied to the daily rate Y_t/N_t, standardized on the full sample. The spurious spike on 3 February 2019 is the single-tweet day ($N_t = 1$). This construction inherits the thin-day fragility and is shown only to motivate the rolling version below.

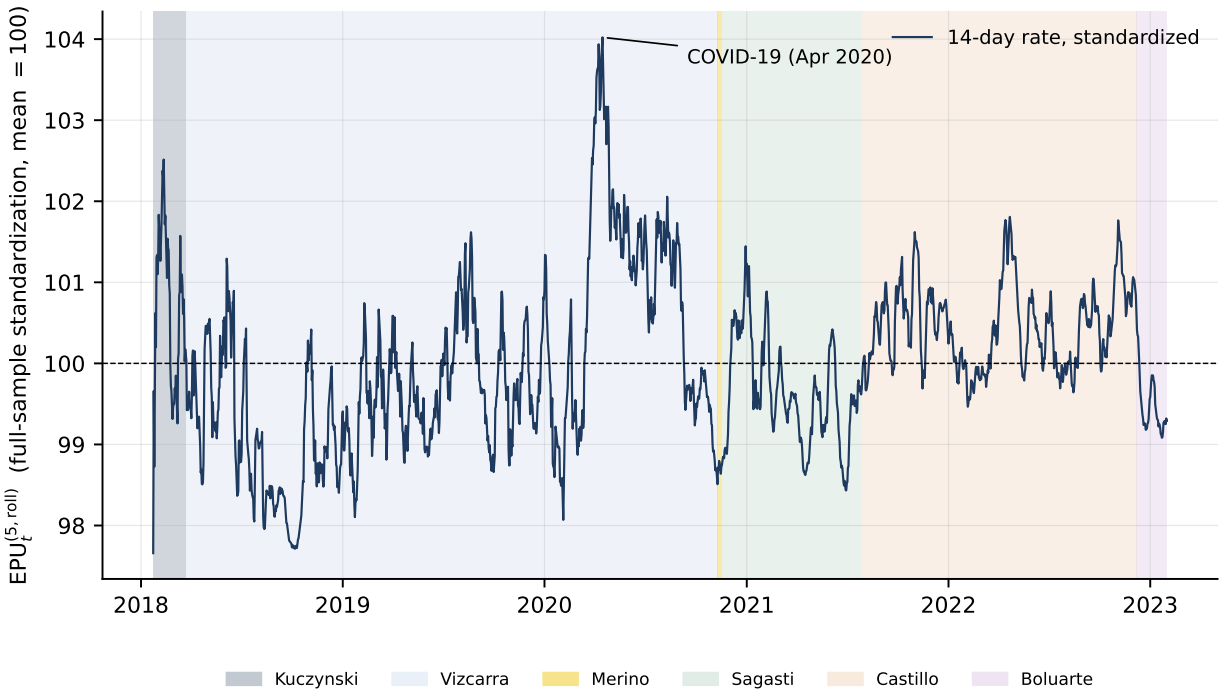

Figure 11: **Standardized EPU, 14-day basis (headline series).** Eq. (5) applied to the volume-weighted rolling rate (4), with presidential sub-periods shaded. The only sustained elevation is the 2020 COVID-19 plateau; the 2019 artifact is gone. Cross-period averages are highest under Kuczynski and Castillo and lowest in the brief November-2020 Merino interim.

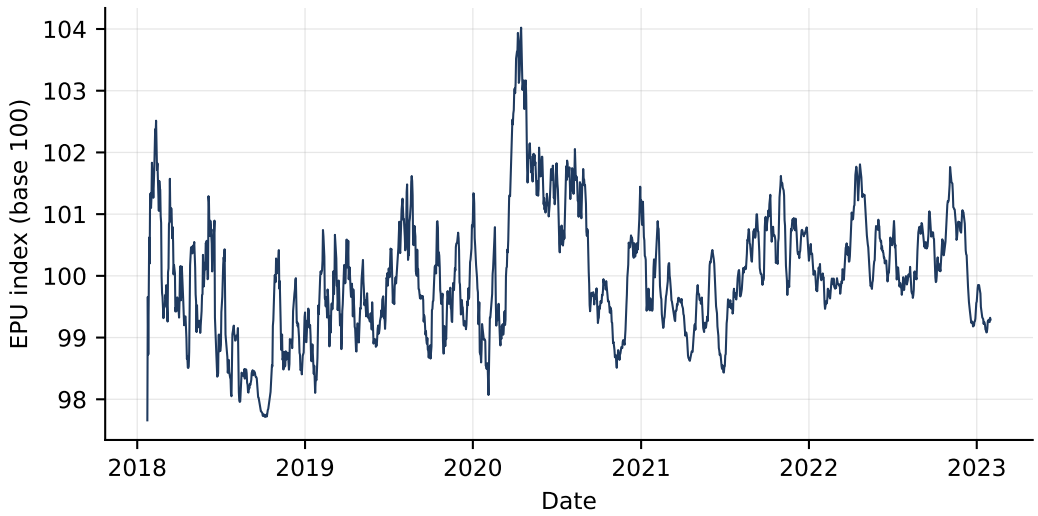

Figure 12: EPU index (standardized, base 100): the working series, exactly as defined in (6), that enters every test of Sections 4 and 5.

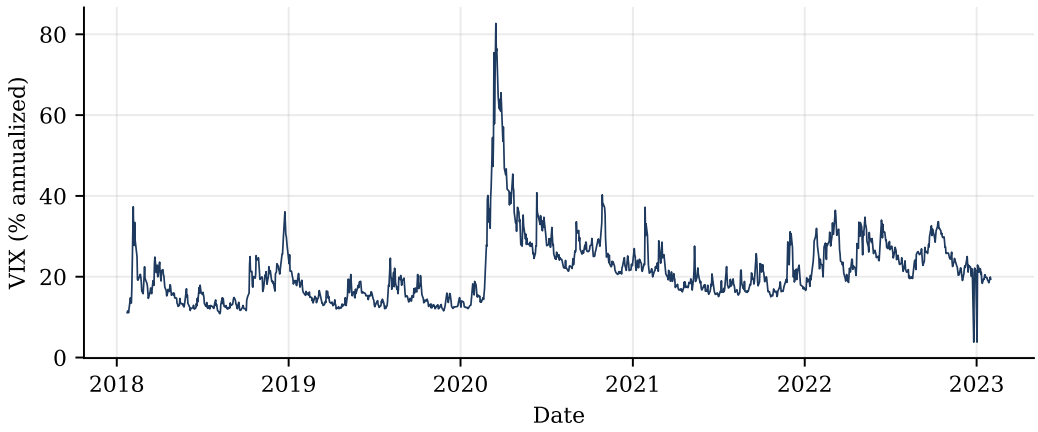

Figure 13: CBOE Volatility Index (VIX), % annualized.

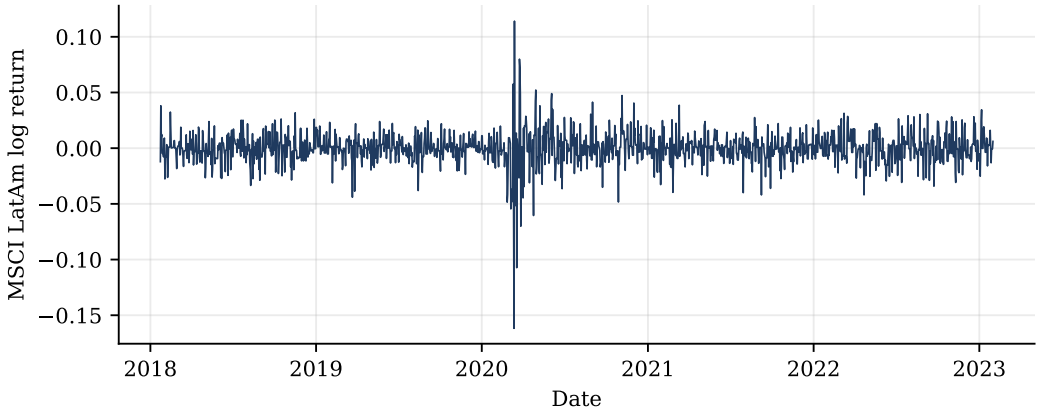

Figure 14: Daily log return of the MSCI Latin America index.

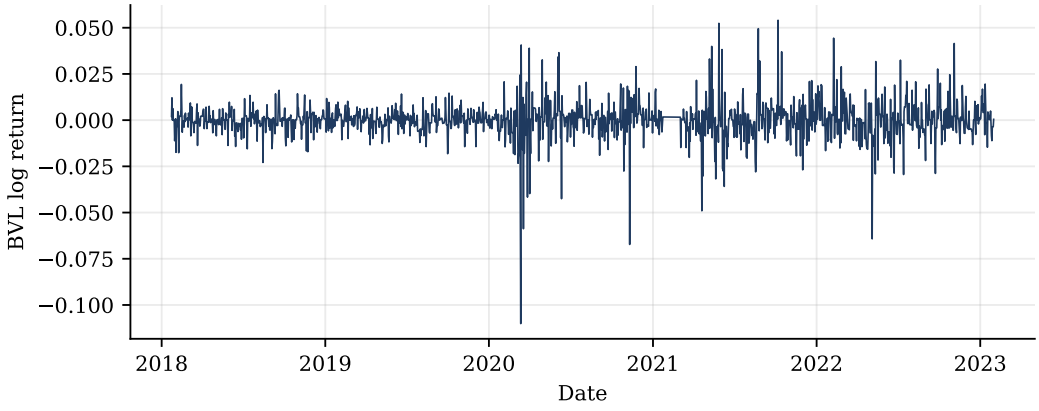

Figure 15: Daily log return of the S&P/BVL Perú General Index.

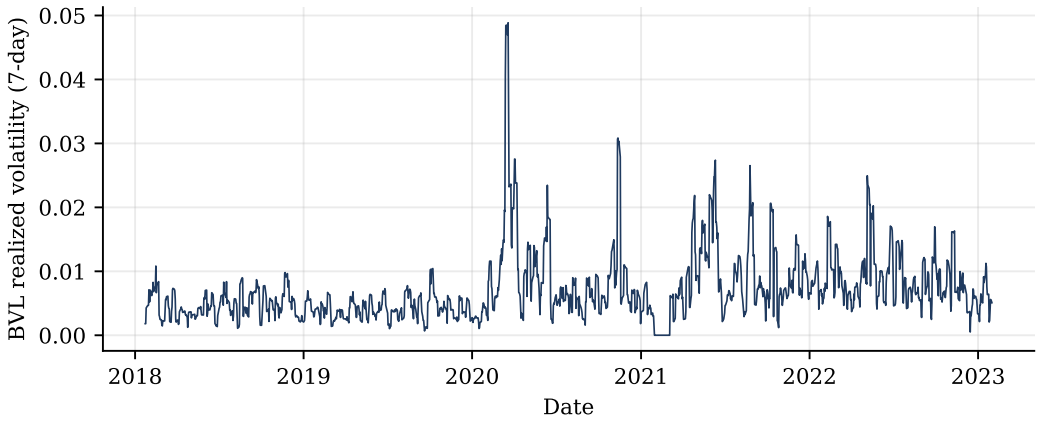

Figure 16: Seven-day realized volatility of the BVL return.

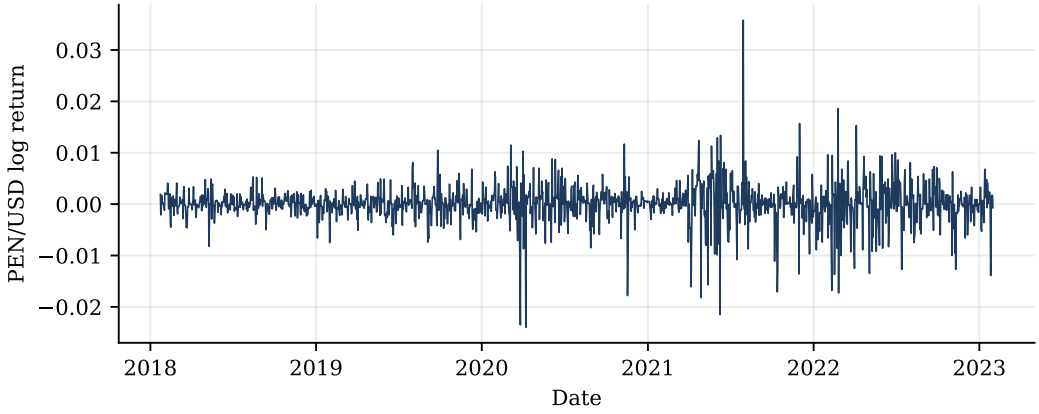

Figure 17: Daily log return of the PEN/USD exchange rate (+ a depreciation of the sol).

Figure 18: PEN/USD exchange rate (soles per US dollar), reconstructed from the daily log returns anchored at its January-2018 value. The level is non-stationary (I(1)); the analysis uses the stationary log return.

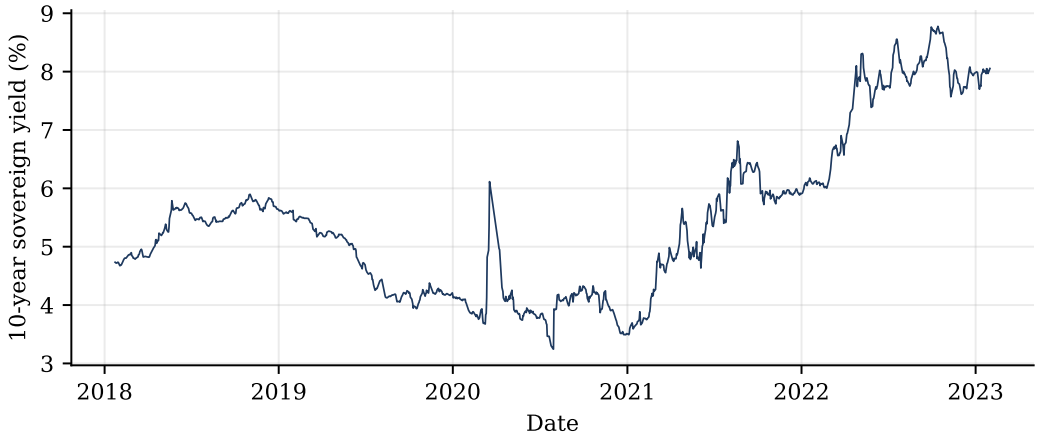

Figure 19: Ten-year sovereign (*soberano*) yield, % per annum.

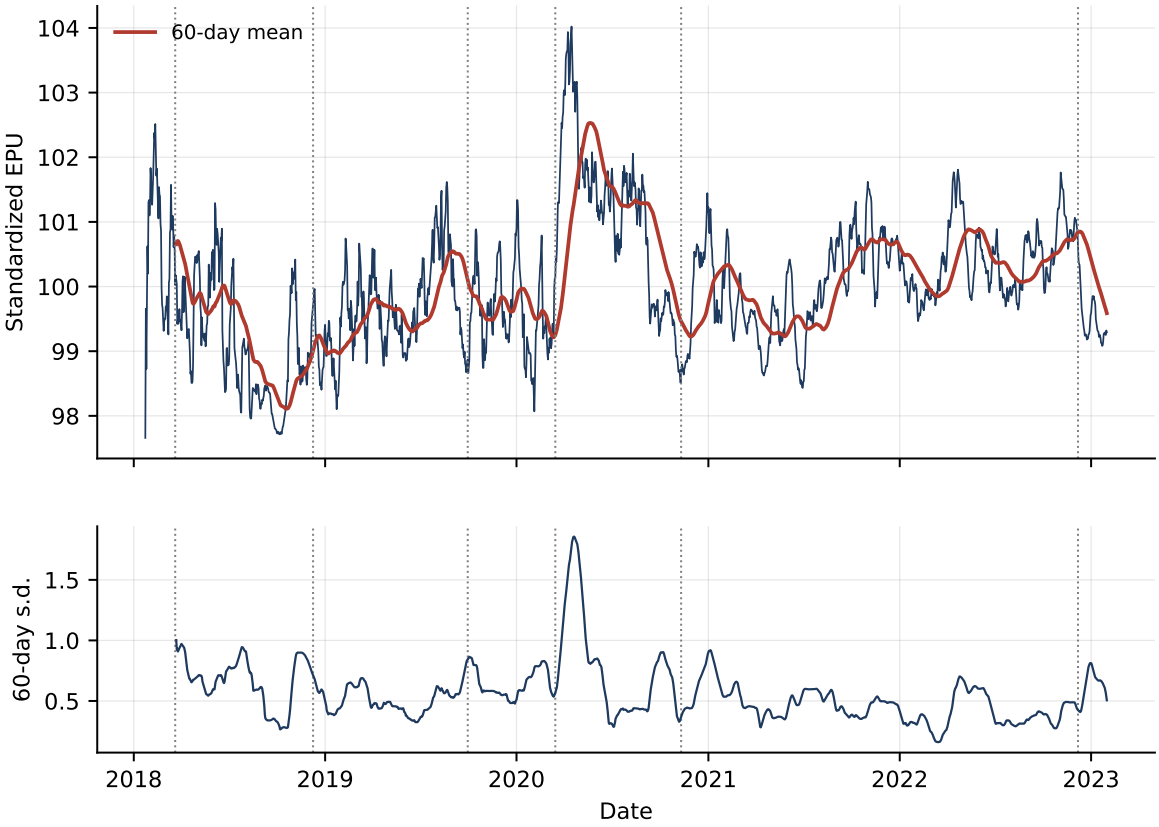

Figure 20: The EPU index (upper panel, with 60-day rolling mean) and its 60-day rolling standard deviation (lower panel). Dotted lines mark six dated political–economic episodes. The flat rolling mean confirms the absence of a trend; the time-varying rolling dispersion shows pronounced volatility clustering around the COVID-19 and late-2022 episodes.

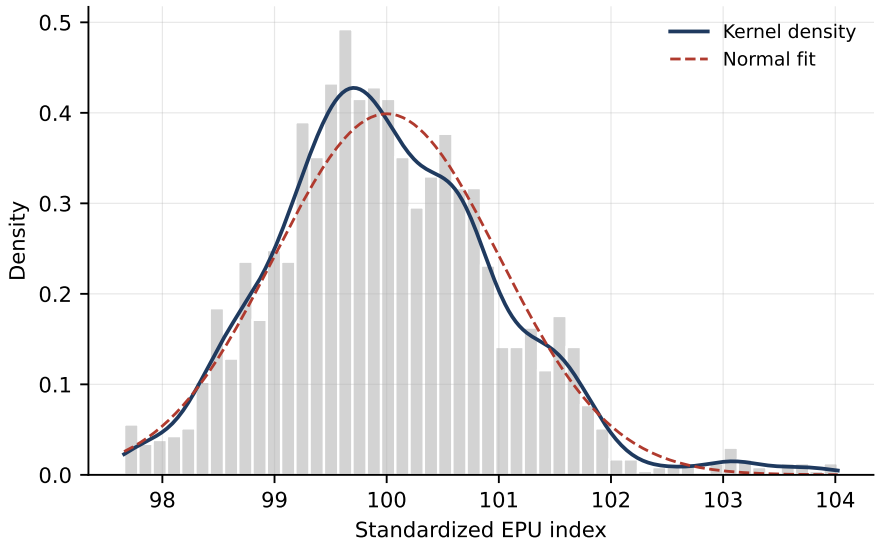

Figure 21: Histogram and kernel density of the standardized EPU index with a fitted normal (dashed). The series is near-symmetric and close to Gaussian, consistent with its smooth, persistent level dynamics.

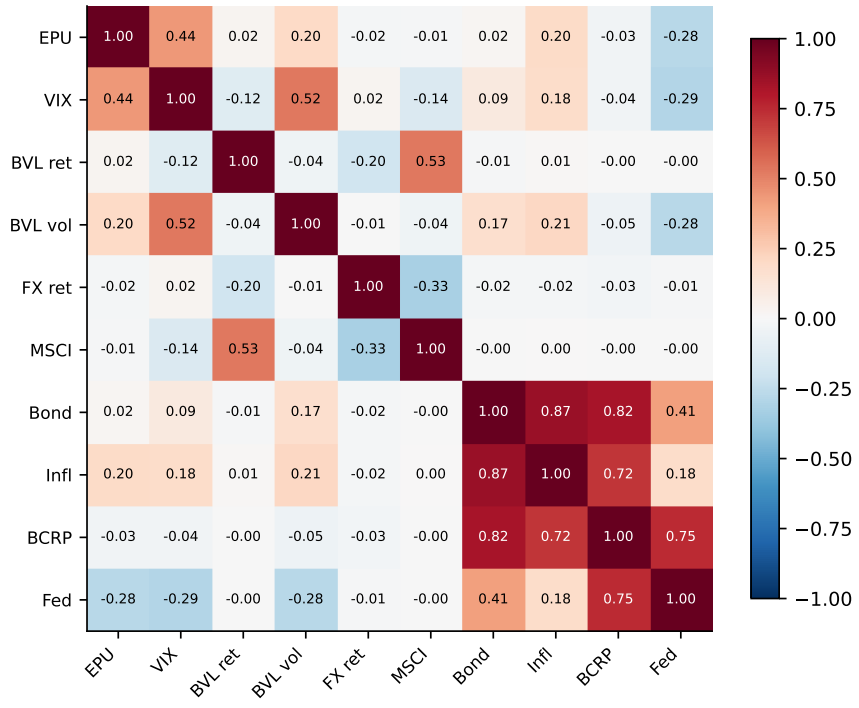

Figure 22: Pearson correlation matrix of the EPU index and the daily macro–financial comparators. Contemporaneous dependence among the daily series is modest, the two material cells being the index–VIX correlation (+0.44) and the index–realized-volatility correlation (+0.20). The lower-frequency block (domestic CPI inflation and the BCRP policy rate, together with the external US federal funds rate) is included for context only and is excluded from the daily analysis of Section 4.

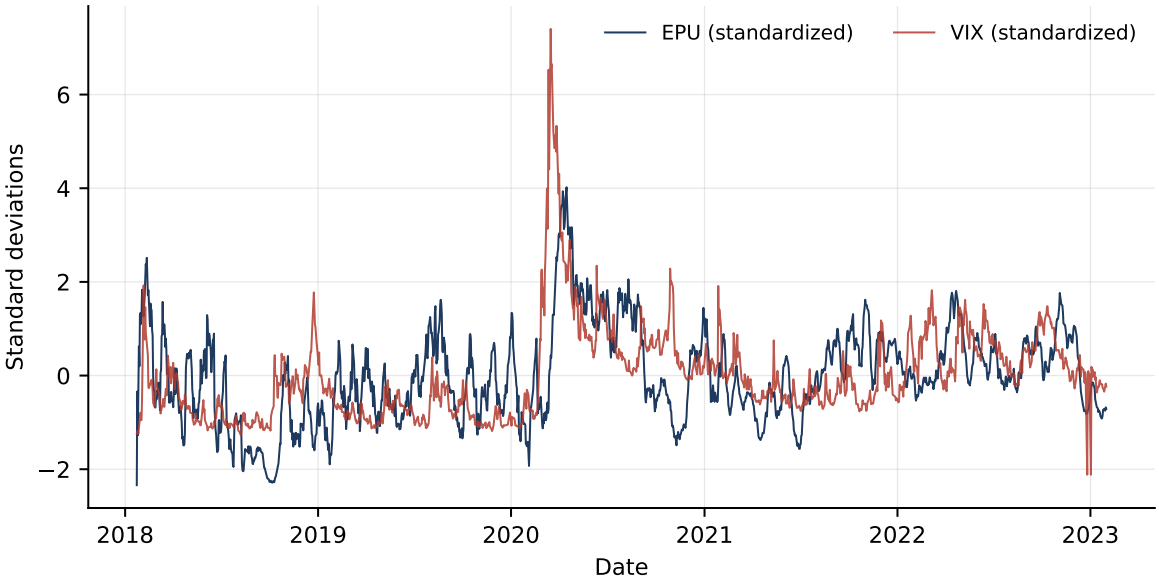

Figure 23: The EPU index and the VIX, each standardized to zero mean and unit variance. Comovement is loose in calm windows and tight during the March 2020 shock, an early indication of state-dependent comovement.

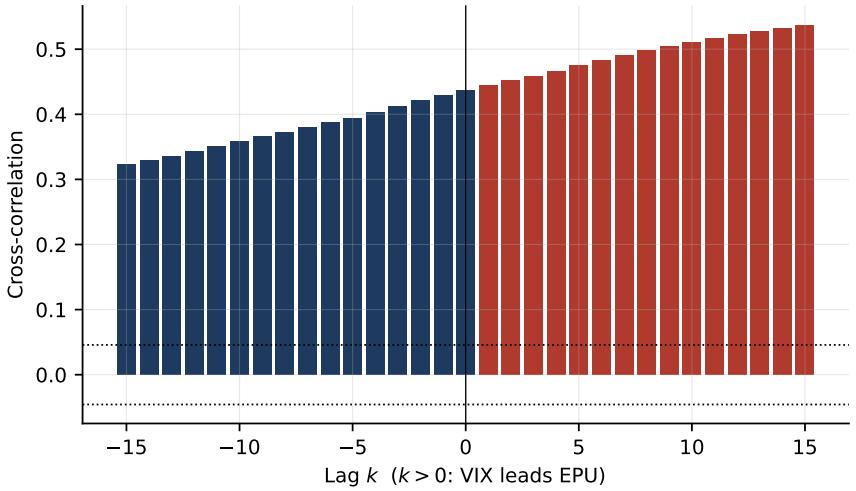

Figure 24: Cross-correlation function $\mathrm{corr}(\mathrm{EPU}_t, \mathrm{VIX}_{t-k})$ with 95% bands, where $k > 0$ means the VIX leads. The function is positive at every displayed lag and rises monotonically from $\approx +0.32$ at $k = -15$ through $\approx +0.44$ at $k = 0$ to $\approx +0.54$ at $k = +15$, with no isolated peak. The broad shape reflects the joint persistence of the two series, so the cross-correlation is only weakly informative about direction; the direction of predictability is settled by the Granger tests of Section 4.5.

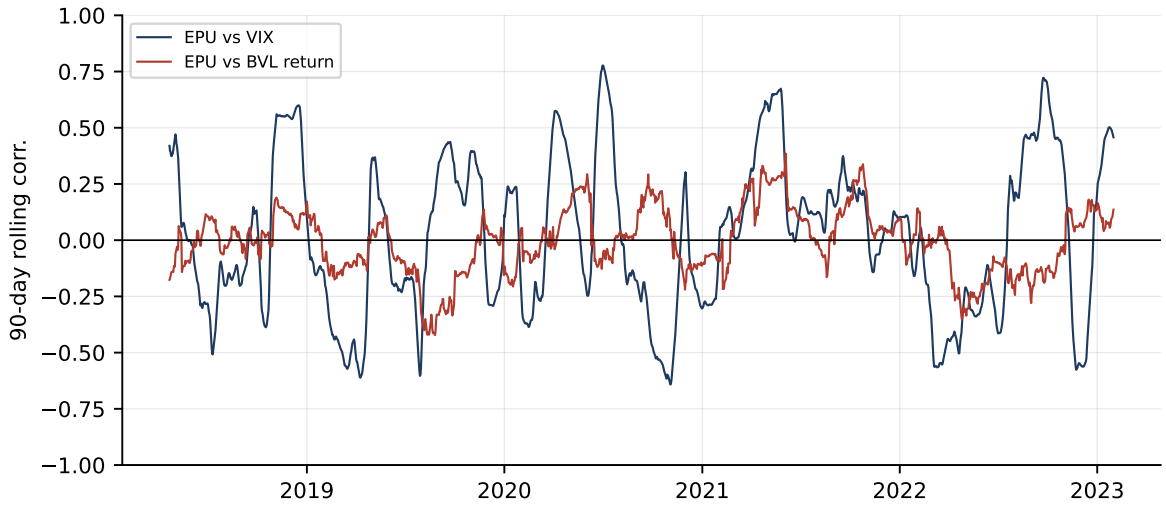

Figure 25: Ninety-day rolling correlations of the EPU index with the VIX and with the BVL return. Both wander across the full sign range and switch sign around each dated crisis; a single full-sample coefficient averages over economically distinct regimes.

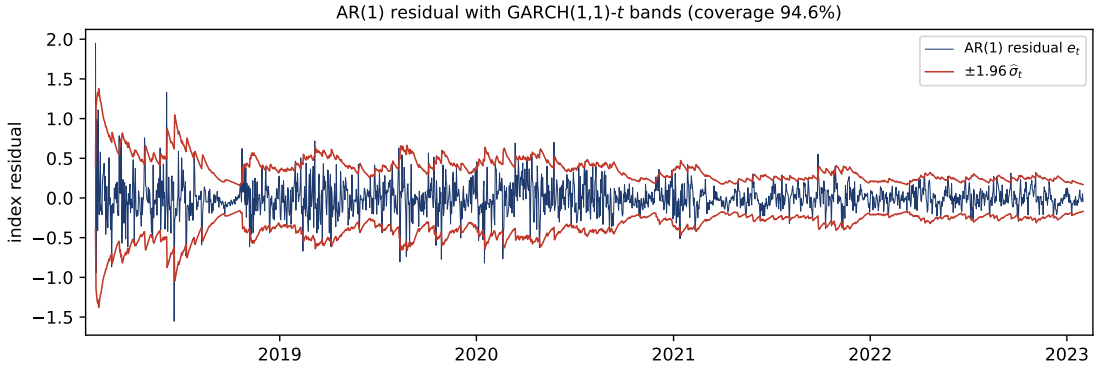

Figure 26: AR(1) residual of the EPU index with $\pm 1.96\,\hat{\sigma}_t$ bands from the GARCH(1,1)-t model; the bands contain 94.6% of the residuals.

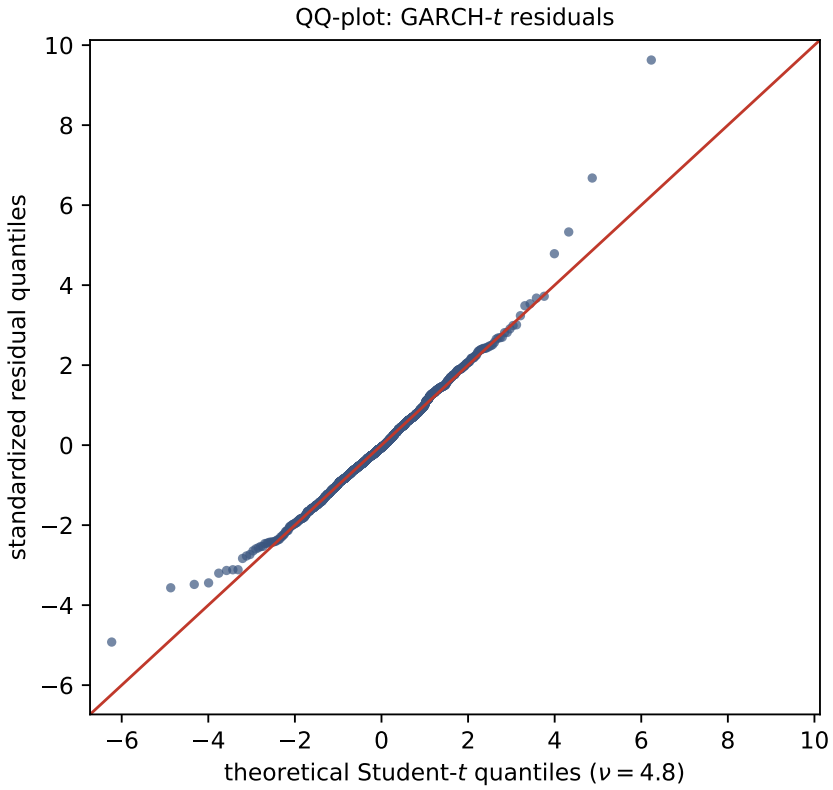

Figure 27: Quantile–quantile plot of the GARCH(1,1)-t standardized residuals against the fitted Student-t law ($\widehat{\nu} \approx 4.8$). The central quantiles track the 45-degree line; the largest positive residuals lie above it and the left tail is milder than the fitted law, so the symmetric Student-t captures the tail thickness only approximately.

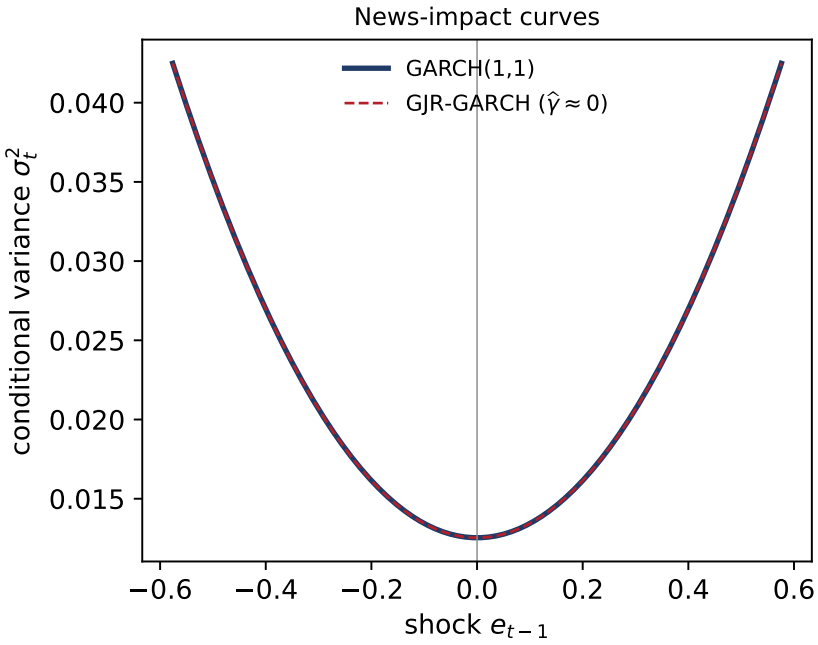

Figure 28: News-impact curves for the symmetric GARCH(1,1) and the GJR-GARCH specifications; the two coincide, consistent with the absence of a leverage effect ($\widehat{\gamma} \approx 0$).