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

Marcelo Gallardo* Manuel Loaiza† Gabriel Rodríguez‡

July 14, 2026

Abstract

We propose the first daily index of economic and political uncertainty (EPU) for Peru, and its defining novelty is the source: rather than newspapers, we read the real-time discourse on X (formerly Twitter). The dominant policy-uncertainty indices count newspaper articles and are typically monthly and lagged; X instead records, at daily frequency, how the country's most influential voices in politics, economics, journalism, and business react to events as they happen, before traditional media process them. We classify each of their messages along economic, political, and uncertainty dimensions with a large language model (LLM), DeepSeek-V3, and recombine these dimensions into a family of complementary daily indices—an instrument that does not yet exist for Peru and is especially valuable in an emerging economy whose political turmoil is acute and fast-moving, the kind a monthly newspaper index would smooth away. Our aim is to assess how closely this index, as a time series, behaves like two established market-based measures of uncertainty: the Chicago Board Options Exchange Volatility Index (VIX) and the volatility of the Bolsa de Valores de Lima (BVL). We compare their levels through static (Pearson) and rolling correlation. Because the VIX is already a volatility while the BVL return is not, we fit a generalized autoregressive conditional heteroskedasticity model of order one, GARCH$(1, 1)$, to the BVL return and take its conditional volatility as a third level. We also build a broader political-uncertainty (PU) index from the intersection of the political and uncertainty categories. The index co-moves with both benchmarks at a moderate level over the full sample and more strongly within documented crises. We read this as the expected signature of measures that share a common uncertainty component while each captures a distinct dimension: for the index a domestic, political and economic one; for the VIX a global financial one and for BVL volatility a domestic market one. The moderate co-movement is therefore not a shortcoming but evidence that the index is not redundant with the benchmarks.

Keywords: Economic Political Uncertainty, X (formerly Twitter), Large Language Models, social media, 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.

*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 work leverages social-media platforms such as 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—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 high reliability in benchmark studies, 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.

Our sample runs from January 2018 to January 2023. The setting is informative: within five years Peru 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—a concentration of political and economic crises in a single small open economy that makes it a natural laboratory for a near-real-time measure of uncertainty. We do not extend the window in either direction, for reasons that are practical rather than conceptual: classifying the full corpus with a large language model carries a monetary cost that scales with the number of tweets; the platform's data-access terms tightened sharply during 2023, raising the cost of retrieving tweets beyond our window; and before 2018 the platform was used only sparsely by the Peruvian political and economic elite we track, so earlier coverage would be thin and unrepresentative. Within these bounds the index still delivers what no existing measure for Peru does: a daily-frequency reading of

economic and political uncertainty, built with a new large-language-model methodology that reads each message in context rather than matching keywords.

This paper makes three contributions. First, we build a new daily index of economic and political uncertainty for Peru from a novel data source: the X discourse of influential accounts in politics, 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 and 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. The construction is portable: it requires only a list of influential accounts and access to the platform's text, so the same pipeline can be replicated for any country in which X hosts an active political and economic conversation.

Second, rather than asking whether the index leads or lags markets, we ask how similar it is to existing measures of uncertainty: the Chicago Board Options Exchange Volatility Index (VIX) and the Bolsa de Valores de Lima (BVL). We compare the co-movement of their levels through static (Pearson) and rolling correlation. We also extract the conditional volatility of the BVL return from a generalized autoregressive conditional heteroskedasticity (GARCH) model of order $(1, 1)$ (Bollerslev, 1986), and compare it with the index and the VIX.

Third, we ask whether this similarity is constant or state-dependent. Splitting the sample into calm periods and documented political and economic crisis windows (dated from event chronologies in the tradition of Romer and Romer (2010) and event-based risk indices (Caldara and Iacoviello, 2022)) lets us assess whether the index tracks market-based uncertainty more closely when uncertainty is acute.

The paper is organized as follows. Section 2 describes the data; Section 3 constructs the index family; and Section 4 sets out the methodology. Section 5 defines a set of time periods that correspond to crises in Peru, whether political or economic in nature, including, for example, the pandemic. Section 6 reports the empirical results: the static and rolling correlation of the index with the VIX and BVL volatility, together with the crisis–calm split (Section 6.1), and the conditional volatility of the Lima market obtained from a GARCH(1,1) model, compared as a third level with the index and the VIX (Section 6.2). Section 7 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, across five categories: journalists (24), politicians (42), economists (16), political analysts (13), and private-sector representatives (2). Accounts are selected for public visibility (follower count and media citations), sustained activity across the window, and coverage of the main institutional actors (Congress, the executive, the central bank, major media outlets, among others). The full set is listed in Table 1. This elite focus limits representativeness relative to a full-population sample but follows Baker et al. (2016) in targeting "informed agents," whose discourse is more likely to reflect genuine policy uncertainty than noise.

The data are obtained via the X (formerly Twitter) application programming interface (API), which yields one comma-separated (CSV) file per account. Figure 1 summarizes the full processing

pipeline in four stages. First, in the ingestion stage, all available tweets of each account within the sample window are pulled from the API into per-account raw files—the 101 accounts of the initial download. Second, an Extract–Transform–Load step merges, deduplicates, and cleans these files; it is here that the four institutional accounts (corporate and academic broadcast feeds) are dropped, reducing the 101 ingested accounts to the 97 individual accounts retained for analysis and leaving a consolidated corpus of 213 249 tweets. Third, every surviving tweet is sent to the DeepSeek-V3 API and labelled along the economic, political, and uncertainty dimensions, yielding a tweet-level uncertainty dataset. Fourth, this labelled dataset feeds the econometric analysis. We keep all items—original tweets, retweets, replies, and quoted posts—so that each retained item is a single tweet-level observation; the index therefore measures the intensity of discourse.

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 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. Baker et al. (2016) classifier assigns a binary label $z_a^{\text{BBD}} \in \{0, 1\}$ to article a:

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

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

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

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 produced independently, by three separate expert-level prompts evaluated on the same tweet, each returning a single binary token. Figure 2 shows the anatomy of one such prompt. It is assembled from six blocks: a role description that casts the model as a panel of domain experts (economist, financial analyst, and so on); a task description stating the binary question for that dimension; solution guidance and a domain-knowledge block that delimit, with Peruvian-context vocabulary, what is to count as economic, political, or uncertainty-laden content; an exception-handling block instructing the model to read tweets in any language and to answer without explanation; and an output-formatting block that constrains the reply to a single 0 or 1. Figure 3 shows how the three prompts combine: the same tweet is passed in parallel to the economic, policy, and uncertainty prompts, and the large language model returns the triple (E, P, U). For the tweet illustrated there, the model returns $(0, 1, 1)$—no explicit economic content, but political content and an expression of uncertainty—so the tweet is counted in the political-uncertainty cell rather than the strict EPU cell.

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 The Index

We build a single daily index. Because every step of its construction is a choice—what to count, how to normalize it, how much to smooth it, and how to scale it—we make those choices, and the reasons for them, explicit. Each is defended below on statistical and substantive grounds, and its effect is shown on the data.

For each calendar day t, let $N_t = \sum_{i=1}^{97} n_{it}$ be the number of classified tweets posted that day and let Y_t be the number that are economic, political, and uncertain at the same time,

$$Y_t \;=\; \sum_{i=1}^{97} \sum_{j=1}^{n_{it}} \mathbf{1}\left\{E_{ijt} = 1,\; P_{ijt} = 1,\; U_{ijt} = 1\right\}, \tag{1}$$

where $E_{ijt}, P_{ijt}, U_{ijt} \in \{0, 1\}$ are the economic, political, and uncertainty labels of the j-th tweet posted by account i on day t, and n_{it} is the number of tweets that account posted that day. The two broader cells that drop one requirement are defined over the same index set,

$$Y_t^{E \cap U} = \sum_{i=1}^{97} \sum_{j=1}^{n_{it}} \mathbf{1}\left\{E_{ijt} = 1,\, U_{ijt} = 1\right\}, \qquad Y_t^{P \cap U} = \sum_{i=1}^{97} \sum_{j=1}^{n_{it}} \mathbf{1}\left\{P_{ijt} = 1,\, U_{ijt} = 1\right\}.$$

Figure 4 plots these three daily counts. Two things stand out. First, the strict count Y_t (bottom panel) and the economic-uncertainty count $Y_t^{E \cap U}$ (top panel) are almost indistinguishable in level and shape: over the sample the strict cell holds 16 451 tweets and $E \cap U$ holds 17 769, and 93 percent of the latter are also political, so the two are nearly the same object. The political-uncertainty count $Y_t^{P \cap U}$ (middle panel) is an order of magnitude larger—76 429 tweets, hundreds on a busy day against tens for the other two—and moves on its own schedule, spiking on episodes of purely political turbulence that carry no economic content. This is why the headline counts only the strict intersection: relaxing the economic requirement would swap in the much larger, differently-timed political signal. Second, all three counts are spiky and drift upward through the sample as these accounts post more, so a raw count confounds uncertainty with sheer activity.

We therefore measure the share of the day's discourse that is uncertain, normalizing Y_t by N_t,

$$s_t \;=\; \frac{Y_t}{N_t} \qquad (0 \le s_t \le 1), \tag{2}$$

and likewise for the two broader cells. Figure 5 plots these shares. The mechanical volume component is gone—the upward drift of the counts flattens—but the daily series is now noisy and ill-behaved on thin days: when only a handful of tweets are posted the denominator N_t is tiny, so a single tweet moves s_t by a large amount and the share can jump close to 1. Daily volume averages 116 tweets but its median is only 56, so thin days are common, and the early, low-volume part of the sample (2018) is correspondingly jagged.

To control that small-sample noise we pool counts over a trailing window of W days and take the ratio of the summed counts—a volume-weighted rate—rather than the average of the daily shares,

$$\rho_t \;=\; \frac{\sum_{s=t-W+1}^{t} Y_s}{\sum_{s=t-W+1}^{t} N_s}. \tag{3}$$

Pooling counts is not the same as averaging the daily shares of Eq. (2). Writing the window total $\mathcal{N}_t = \sum_{s=t-W+1}^{t} N_s$, a short rearrangement of Eq. (3) gives

$$\rho_t \;=\; \sum_{s=t-W+1}^{t} \frac{N_s}{\mathcal{N}_t}\, s_s \;=\; \sum_{s=t-W+1}^{t} w_s\, s_s, \qquad w_s = \frac{N_s}{\mathcal{N}_t}, \qquad \sum_{s=t-W+1}^{t} w_s = 1, \tag{4}$$

so ρ_t is the mean of the daily shares s_s weighted by daily volume N_s. A high-volume day contributes in proportion to how much it actually says, and a low-volume day cannot dominate the window. For $t < W$ the sums in Eq. (3) run over the available days $s = 1, \ldots, t$, so the first $W-1$ dates use an expanding window; with $W = 14$ this affects only the opening fortnight of a five-year sample.

We set $W = 14$ days, and Figure 6 reports what that costs. Each point is one value of W. The horizontal axis is the phase lag a trailing window imposes: a window over the last W days dates its estimate at $(W-1)/2$ days in the past. The left axis is the sampling noise of ρ_t: treating the window's tweets as draws with success rate ρ_t, its standard error is $\left(\rho_t\,(1-\rho_t)\,/\mathcal{N}_t\right)^{1/2}$, expressed relative to the standard deviation of ρ_t itself, a scale-free noise ratio (sampling error against the day-to-day variation of the rate), not a formal variance share. The right axis is the March 2020 maximum of the index, in multiples of its sample mean.

Both curves fall monotonically. Pooling more days buys precision: the noise ratio drops from 0.48 at $W = 3$ to 0.33 at $W = 14$, 0.25 at $W = 30$, and 0.19 at $W = 60$. It pays for that precision in lag and in amplitude, the March 2020 peak falling from 3.54 times the sample mean at $W = 3$ to 2.14 at $W = 60$. No W minimizes both, and the data supply no interior optimum: minimizing noise alone drives W to the sample length and returns a constant, while minimizing lag returns the raw daily share of Eq. (2), whose defects Figure 5 displays. We therefore fix W by convention, at the fortnight, and report where that lands: a 6.5-day phase lag and a noise ratio of 0.33. Nothing of substance turns on it. Recomputing the index with $W \in \{7, 21, 30\}$ leaves the peaks, troughs, and their timing in place, the three series correlating with the $W = 14$ baseline at 0.88, 0.94, and 0.89.

Following Baker et al. (2016), we rescale the rate by its own sample mean so that the series reads in interpretable units,

$$\mathrm{EPU}_t \;=\; 100 \cdot \frac{\rho_t}{\bar{\rho}}, \qquad \bar{\rho} \;=\; \frac{1}{T}\sum_{t=1}^{T} \rho_t, \tag{5}$$

so that a value of 100 is an average day and a reading of, say, 200 is twice as uncertain as usual. On an average day $\bar{\rho} \approx 0.073$: about seven percent of these accounts' discourse is economic-policy uncertainty. Rescaling is a pure change of scale, leaving the dynamics untouched.

Figure 7 plots the finished index against its two cruder forms, the raw count $100\,Y_t/\bar{Y}$ and the daily share $100\,s_t/\bar{s}$, each rescaled by its own mean so that all three average 100 by construction. What separates them is the distance between that mean and the body of the distribution. The raw count sits at a median of 44.6 and reaches 915, because a handful of high-volume days carry the average; the daily share sits at a median of 86.3 and reaches 1 343, because a thin denominator lets one tweet dominate. The refined index has a median of 96.4 and a maximum of 272: its typical day is close to the base, and its extremes are episodes rather than artifacts of volume or of a small denominator. It is the volume-weighted rate of Eq. (3) on the scale of Eq. (5).

Figure 8 shows the index at work. The top panel is the headline EPU; the two below are the VIX and the log return of the Lima exchange, with the eight domestic crisis windows shaded. The index rises sharply inside the market-driven windows, most clearly the March 2020 COVID-19 emergency, when it reaches several times its average; the purely domestic political ruptures (the November 2020 vacancy of President Vizcarra, the December 2022 removal of President Castillo) register instead in the companion PU index of Section 3 where the strict headline series can stay near or below its base. Where the index rises, the VIX and the market returns move with it, a first, informal sign that the index moves with established gauges of uncertainty—the question the sections that follow take up formally, asking how closely the index resembles such measures rather than whether it leads or causes them.

4 Methodology

This section assesses how closely the headline economic and political uncertainty (EPU) index of Section 3 resembles established measures of uncertainty. We ask a question of similarity: whether a text-based attention index built from Peruvian elite discourse behaves like market-based gauges of uncertainty, and where it departs from them. A close co-movement is evidence that the index tracks the same underlying phenomenon (Baker et al., 2016; Bloom, 2009); a divergence is itself informative about what each measure captures. We use a set of association measures—static and rolling correlations against the VIX and BVL realized variance, and the conditional volatility of the BVL return from a GARCH$(1,1)$ model, compared as a third level with the index and the VIX.

4.1 Benchmark indicators

We compare the EPU index against two external uncertainty measures, and no others.

The first is the Chicago Board Options Exchange (CBOE) Volatility Index (VIX), the implied volatility of options on the Standard and Poor's 500 (S&P 500) index. It is the most widely used forward-looking gauge of global financial uncertainty.

The second is the realized volatility of the Bolsa de Valores de Lima (BVL). Let P_t denote the closing level of the BVL index on day t and $r_t = \log(P_t/P_{t-1})$ its daily log-return. We measure realized volatility by the squared log-return,

$$\mathrm{RV}_t = r_t^2. \tag{6}$$

This is the standard one-day proxy: if the conditional mean of r_t is zero, then r_t^2 is an unbiased, though noisy, estimate of that day's return variance.

The two benchmarks differ in every dimension except one. The VIX is global, forward-looking, and read off option prices; RV_t is domestic, backward-looking, and read off observed returns. What

they share is that both are anchored to traded financial instruments, unlike our text-based index. Figure 8 plots the three series together. Both are quoted on trading days only, while the index is defined on every calendar day. We align them on the calendar panel of $T = 1\,835$ days, carrying the last quoted value forward through weekends and holidays, so that every statistic below is computed on the same dates.

4.2 Static co-movement

For each benchmark $b_t \in \{\mathrm{VIX}_t,\ \mathrm{RV}_t\}$ we report two association statistics between the EPU index E_t and b_t over a sample of T days. The Pearson correlation measures linear co-movement,

$$\rho \;=\; \frac{\sum_{t=1}^{T}(E_t-\bar{E})(b_t-\bar{b})}{\sqrt{\sum_{t=1}^{T}(E_t-\bar{E})^2}\,\sqrt{\sum_{t=1}^{T}(b_t-\bar{b})^2}}, \tag{7}$$

with $\bar{E}$ and $\bar{b}$ the sample means. The Spearman correlation is the Pearson correlation of the ranks: with $R_t = \mathrm{rank}(E_t)$ and $Q_t = \mathrm{rank}(b_t)$,

$$\rho_S \;=\; \frac{\sum_{t=1}^{T}(R_t-\bar{R})(Q_t-\bar{Q})}{\sqrt{\sum_{t=1}^{T}(R_t-\bar{R})^2}\,\sqrt{\sum_{t=1}^{T}(Q_t-\bar{Q})^2}}, \tag{8}$$

which captures any monotone relation and is robust to the heavy tails of uncertainty series. Together they discriminate the shape of the association: $\rho_S \approx \rho$ signals a linear link, while $\rho_S > \rho$ signals a monotone but nonlinear one. We report ρ_S as a diagnostic of that shape and conduct inference on ρ, for which the sampling variance derived below is available in closed form.

Before measuring co-movement we classify each of the three series under study (the EPU index, the VIX, and the BVL volatility), generically denoted x_t, because its persistence governs how a correlation is read; Table 4 reports the diagnostics over the $T = 1\,835$ days of the sample. Two tests with opposite nulls are needed, because neither settles the question alone. The augmented Dickey–Fuller (ADF) test of Dickey and Fuller (1979), in the augmented form of Said and Dickey (1984), regresses Δx_t on x_{t-1}, a constant, and enough lagged differences to whiten the residual; its null is a unit root—a random-walk-like process with no fixed mean—rejected when the t-statistic on x_{t-1} falls below the 5% critical value of -2.86. The Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test of Kwiatkowski et al. (1992) reverses the null. It writes $x_t = \kappa_t + \upsilon_t$ with $\kappa_t = \kappa_{t-1} + u_t$ and tests $\mathrm{Var}\,(u_t) = 0$, the absence of the random-walk component κ_t, through the scaled partial sums of the demeaned series,

$$\hat{\eta} \;=\; \frac{1}{T^2}\,\frac{\sum_{t=1}^{T} S_t^2}{\hat{s}^2(\ell)}, \qquad S_t \;=\; \sum_{i=1}^{t}(x_i-\bar{x}), \tag{9}$$

where $\hat{s}^2(\ell)$ is a Bartlett-weighted long-run variance of the same form as Eq. (11) below, computed on the demeaned series $x_t - \bar{x}$ rather than on the score and with its own bandwidth ℓ. Under stationarity the deviations $x_i - \bar{x}$ cancel and S_t stays of order $T^{1/2}$; under a unit root they accumulate and S_t grows like T, so $\hat{\eta}$ diverges. The test rejects for large values, above 0.463 at 5%.

The two verdicts are jointly informative, and Table 4 delivers them. The EPU index and the VIX reject the unit root (-4.18, -4.30) and also reject stationarity (2.247, 2.683). The two failures are of different kinds: the augmented Dickey–Fuller test has low power against processes near the unit-root boundary, whereas the KPSS statistic over-rejects in finite samples when the series is stationary but strongly persistent. With $T = 1\,835$ the first test is powerful enough to detect that

the autoregressive root lies below unity, while the second is distorted enough to reject stationarity, and that combination is the signature of the near-unit-root region: first-order autocorrelations of 0.979 and 0.973, and a shock to the level with a half-life of a few weeks, 32.7 days for the index and 25.3 for the VIX. Both series are highly persistent before any benchmark is brought in, the index somewhat more so than the VIX, decaying over a horizon of weeks rather than days. This pattern is also consistent with structural breaks or regime shifts, so we read it as descriptive rather than as uniquely justifying the levels specification. BVL realized variance behaves differently, reverting within a day with an autocorrelation of 0.165, so its shocks are gone before the index has begun to respond to them. Its KPSS statistic (0.826) also exceeds the critical value, for an unrelated reason: the mean of RV_t rises from 2.5×10^{-5} before 2020 to 1.2×10^{-4} afterwards, and that shift in level, not serial dependence, is what inflates the partial sums. All three series therefore enter the correlations in levels, without differencing.

The persistence of E_t and of the VIX bears directly on inference. Writing the correlation of Eq. (7) as the slope of the ordinary least squares regression of the standardized benchmark on the standardized index, $\tilde{b}_t = \rho\, \tilde{E}_t + u_t$ with $\tilde{E}_t = \left(E_t - \bar{E}\right)/\hat{\sigma}_E$ and $\tilde{b}_t = \left(b_t - \bar{b}\right)/\hat{\sigma}_b$, the slope equals the Pearson correlation numerically. The estimator solves $T^{-1}\sum_{t=1}^{T} g_t\left(\hat{\rho}\right) = 0$ with score

$$g_t(\rho) = \tilde{E}_t\left(\tilde{b}_t - \rho\, \tilde{E}_t\right), \tag{10}$$

the contribution of day t to the least squares gradient. Generalized method of moments gives the sandwich variance $A^{-1}SA^{-1}/T$ with $A = \mathbb{E}\left[\tilde{E}_t^2\right]$ and $S = \sum_{j=-\infty}^{\infty} \mathrm{Cov}\left(g_t, g_{t-j}\right)$ the long-run variance of the score. Standardizing the regressor sets $A = 1$, the sandwich collapses to S/T; this is the HAC variance of the standardized-slope estimand, which coincides numerically with $\hat{\rho}$ and conditions on the sample means and variances used to standardize. The full influence function of $\hat{\rho}$ adds the terms from estimating those moments; under the near-unit-root persistence documented below the serial-dependence correction in S dominates them. We report the standardized-slope HAC standard error and, for reference, the classical Fisher standard error $\left(1-\hat{\rho}^2\right)/\sqrt{T}$.

We estimate S with the heteroskedasticity- and autocorrelation-consistent (HAC) estimator of Newey and West (1987). Let $\hat{\gamma}_j = T^{-1}\sum_{t=j+1}^{T} \hat{g}_t\, \hat{g}_{t-j}$ denote the sample autocovariances of the fitted score. Truncating at lag L and weighting by the triangular window of Bartlett (1950) gives the long-run variance

$$\hat{s}^2\left(L\right) = \hat{\gamma}_0 + 2\sum_{j=1}^{L} w_j\, \hat{\gamma}_j, \qquad w_j = 1 - \frac{j}{L+1}, \tag{11}$$

and the variance of the estimated correlation is

$$\widehat{\mathrm{Var}}\left(\hat{\rho}\right) = \frac{\hat{s}^2\left(L\right)}{T}. \tag{12}$$

The weights are not cosmetic. An unweighted truncation ($w_j = 1$) can return a negative variance, whereas the Bartlett window is the autoconvolution of a rectangular window, its Fourier transform is the nonnegative Fejér kernel, and $\hat{s}^2\left(L\right) \geq 0$ holds by construction.

The bandwidth follows the autoregressive plug-in of Andrews (1991) applied to $\hat{g}_t$, which selects a long window for the VIX and a short one for RV_t. We report $L \in \{7, 30, 60, 100\}$, the lower end being the sample-size rule $L = \left\lfloor 4\left(T/100\right)^{2/9} \right\rfloor = 7$ of Newey and West (1994).

Under independence $\hat{\gamma}_j = 0$ for $j \geq 1$, Eq. (12) collapses to $\hat{\gamma}_0/T$, and for jointly Gaussian data this is the classical variance $\left(1-\rho^2\right)^2/T$ of a Pearson correlation. That collapse fails here. The index is a fourteen-day rolling proportion, so consecutive observations share thirteen of their

fourteen days of tweets by construction, and the VIX inherits the persistence of the option market. The fitted scores are therefore autocorrelated: at first order 0.95 for the VIX against 0.12 for RV_t. The neglected $\hat{\gamma}_j$ are large and positive, so the classical formula understates the standard error of $\hat{\rho}$. Section 6.1 reports both. The correction is large for the VIX and small for RV_t, matching their first-order autocorrelations.

Each correlation is computed on the full sample and, separately, on crisis and calm days using the eight windows of Section 5. Dated from the historical record rather than from market movements, the split is exogenous to the VIX and to RV_t and cannot be an artefact of the benchmarks under evaluation. A series that tracks uncertainty should resemble them more closely when uncertainty is high, $\rho^{\text{crisis}} > \rho^{\text{calm}}$. The full-sample value is not a midpoint: pooling the regimes absorbs the difference in their means, so it need not lie between the two. Because the boundary dates embody judgement, we recompute each correlation with every window edge displaced by ± 5 days and report the resulting range.

4.3 Time-varying co-movement

The crisis/calm split treats similarity as constant within each regime. To let it vary over time we compute the Pearson correlation of Eq. (7) on a trailing window of $m = 90$ days. The window length trades sampling noise against time resolution: the standard error of a correlation falls roughly like $m^{-1/2}$, so a shorter window tracks change faster but at each date estimates ρ from fewer observations and is noisier, while a longer one is smoother but blends distinct episodes and lags turning points by about $m/2$ days. We set $m = 90$ days, one trading quarter. Because the index is a 14-day rate and both series are highly persistent, the effective number of independent observations is well below 90, so we read the rolling curve as descriptive rather than as precise local inference. Hence, we define

$$\rho_t^{(m)} = \frac{\sum_{s=t-m+1}^{t}(E_s - \bar{E}_t)(b_s - \bar{b}_t)}{\sqrt{\sum_{s=t-m+1}^{t}(E_s - \bar{E}_t)^2}\,\sqrt{\sum_{s=t-m+1}^{t}(b_s - \bar{b}_t)^2}}, \tag{13}$$

where $\bar{E}_t$ and $\bar{b}_t$ are the means of E_s and b_s over the window $s = t-m+1, \ldots, t$. A quarter-long window resolves only the two crises longer than 100 days, the pandemic (C3) and the contested election (C5); the six shorter windows are averaged into the surrounding calm and are left to the discrete split of Section 4.2.

Figure 9 plots $\rho_t^{(90)}$ for the VIX and for RV_t with the crisis windows shaded, and three features of each curve are informative. The level of a curve away from the shaded bands measures the similarity in ordinary times; its height inside the bands, relative to that level, measures how much the similarity changes in crises. The sign locates disagreement: an interval where $\rho_t^{(90)} < 0$ is one in which the index and the benchmark moved in opposite directions. And the alignment of any rise with a shaded band distinguishes co-movement tied to a documented crisis from a rise outside every band, which flags an episode the narrative dating did not mark. The curve is read as the time profile of the similarity, not as evidence that the index leads or drives the benchmark.

4.4 Similarity in conditional volatility

The VIX is already a volatility, the option-implied volatility of the S&P 500, so its level is directly comparable to the level of the index. The BVL return is not a volatility; it is the daily price change, and what plays the role of a volatility is its dispersion over time. To place the Lima market on the same footing as the VIX we extract its conditional volatility with a GARCH$(1,1)$ filter, which

turns the return into a volatility series comparable, as a level, with the VIX and with the index. We fit the same filter to the index and to the VIX as well, but for a different purpose: to report the diagnostics of Table 5 and to establish, in Section 6.2, why the conditional variance of the index measures how precisely its rate is estimated rather than how much uncertainty there is, so that it is the level of the index, and not a volatility of its own, that carries the comparison.

Let x_t denote the series being modelled and let $\mathcal{F}_{t-1} = \sigma(x_{t-1}, x_{t-2}, \ldots)$ denote the information set available at the end of day $t-1$; all conditional moments below are taken with respect to it. The conditional mean and conditional variance of the series are

$$m_t \;=\; \mathbb{E}[x_t \mid \mathcal{F}_{t-1}]\,, \qquad \sigma_t^2 \;=\; \mathrm{Var}(x_t \mid \mathcal{F}_{t-1}) \;=\; \mathbb{E}\Big[(x_t - m_t)^2 \mid \mathcal{F}_{t-1}\Big]\,, \tag{14}$$

the mean and dispersion of today's observation given everything observed through yesterday. The conditional variance is the object of interest: it is the day-by-day counterpart of the unconditional variance $\mathrm{Var}(x_t)$, and its square root σ_t, the conditional volatility, is the daily measure of turbulence the model delivers.

The model is the generalized autoregressive conditional heteroskedasticity model of order $(1,1)$, GARCH$(1,1)$, of Bollerslev (1986). We impose two restrictions on it: the innovation is Gaussian, and the variance responds to the magnitude of yesterday's innovation and not to its sign. It generalizes the autoregressive conditional heteroskedasticity (ARCH) model of Engle (1982), in which today's variance responds only to recent squared shocks; adding yesterday's variance itself as a regressor lets a single lag of each capture the long, smooth decay of volatility that would otherwise require many ARCH lags.

The mean equation is a constant for the return series r_t and adds a first-order autoregressive, AR(1), term for the level series (the VIX and the EPU index), so that in every case the GARCH recursion describes the conditional volatility of a mean-zero innovation:

$$x_t = \mu + \phi\, x_{t-1} + \varepsilon_t, \qquad \varepsilon_t = \sigma_t\, z_t, \qquad z_t \sim \mathrm{N}(0,1) \text{ i.i.d.}, \tag{15}$$

$$\sigma_t^2 = \omega + \alpha\, \varepsilon_{t-1}^2 + \beta\, \sigma_{t-1}^2, \qquad \omega > 0, \quad \alpha \geq 0, \quad \beta \geq 0, \quad \alpha + \beta < 1, \tag{16}$$

where μ is the intercept and ϕ the AR(1) coefficient, set to zero for the return series. The innovation $\varepsilon_t \equiv x_t - \mathbb{E}[x_t \mid \mathcal{F}_{t-1}]$ is a martingale-difference sequence, $\mathbb{E}[\varepsilon_t \mid \mathcal{F}_{t-1}] = 0$ and $\mathbb{E}[\varepsilon_t^2 \mid \mathcal{F}_{t-1}] = \sigma_t^2$: the sign of today's surprise is unpredictable, its expected magnitude is not, and that predictability of magnitudes is what the variance equation models. In Eq. (16), ω is the variance intercept, α governs the response of today's variance to yesterday's squared innovation (the news term), and β the persistence inherited from yesterday's variance (the memory term).

The recursion of Eq. (16) is initialized at σ_0^2 and, by induction, σ_t^2 is a measurable function of $x_{t-1}, \ldots, x_0$: it is known at $t-1$, as a conditional variance must be. Three further properties make the model interpretable.

First, taking unconditional expectations in Eq. (16) requires $\mathbb{E}[\varepsilon_{t-1}^2]$. Because σ_{t-1}^2 is $\mathcal{F}_{t-2}$-measurable and z_{t-1} is independent of $\mathcal{F}_{t-2}$ with $\mathbb{E}[z_{t-1}^2] = 1$, the law of iterated expectations gives

$$\mathbb{E}\Big[\varepsilon_{t-1}^2\Big] \;=\; \mathbb{E}\Big[\mathbb{E}\Big[\sigma_{t-1}^2\, z_{t-1}^2 \mid \mathcal{F}_{t-2}\Big]\Big] \;=\; \mathbb{E}\Big[\sigma_{t-1}^2\, \mathbb{E}\Big[z_{t-1}^2 \mid \mathcal{F}_{t-2}\Big]\Big] \;=\; \mathbb{E}\Big[\sigma_{t-1}^2\Big]\,.$$

In stationarity, $\mathbb{E}[\sigma_t^2] = \mathbb{E}[\sigma_{t-1}^2] = \bar{\sigma}^2$, so $\bar{\sigma}^2 = \omega + (\alpha+\beta)\,\bar{\sigma}^2$, that is,

$$\bar{\sigma}^2 \;=\; \frac{\omega}{1-\alpha-\beta}, \tag{17}$$

which is why $\omega > 0$ and $\alpha + \beta < 1$ are imposed. Second, iterating the conditional expectation of Eq. (16) forward gives the h-step variance forecast

$$\mathbb{E}\Big[\sigma_{t+h}^2 \mid \mathcal{F}_t\Big] - \bar{\sigma}^2 \;=\; (\alpha+\beta)^{\,h-1}\Big(\sigma_{t+1}^2 - \bar{\sigma}^2\Big)\,, \qquad h \geq 1,$$

so deviations of volatility from its long-run level die out geometrically at rate $\alpha+\beta$, and a volatility shock has an implied half-life of $\ln(1/2)/\ln(\alpha+\beta)$ days.

Third, repeated back-substitution of $\beta\sigma_{t-1}^2$ yields

$$\sigma_t^2 \;=\; \frac{\omega}{1-\beta} + \alpha \sum_{j=0}^{\infty} \beta^j \, \varepsilon_{t-1-j}^2,$$

an ARCH model of infinite order with geometrically decaying weights: the single pair (α, β) encodes a complete memory profile, which is what makes the $(1,1)$ order sufficient in practice.

Estimation is by Gaussian maximum likelihood. The distributional assumption implies $\varepsilon_t \mid \mathcal{F}_{t-1} \sim \mathrm{N}(0, \sigma_t^2)$, so the joint density factorizes by the prediction-error decomposition into the product of these one-step conditional densities, $f(x_1, \ldots, x_T; \theta) = \prod_t f(x_t \mid \mathcal{F}_{t-1}; \theta)$ with $\theta = (\mu, \phi, \omega, \alpha, \beta)$, and the log-likelihood is

$$\ell(\theta) \;=\; -\frac{1}{2} \sum_{t=1}^{T} \left[\ln(2\pi) + \ln \sigma_t^2(\theta) + \frac{\varepsilon_t(\theta)^2}{\sigma_t^2(\theta)} \right], \tag{18}$$

where $\varepsilon_t(\theta)$ and $\sigma_t^2(\theta)$ are computed recursively from Eqs. (15)–(16), the recursion being initialized at the sample variance of the fitted innovations. The estimator $\hat{\theta}$ maximizes $\ell(\theta)$; no closed form exists, and the maximization is carried out numerically from several starting points.

Estimation delivers a daily volatility series, which is the object we use. Given $\hat{\theta} = (\hat{\mu}, \hat{\phi}, \hat{\omega}, \hat{\alpha}, \hat{\beta})$, the fitted innovations are $\hat{\varepsilon}_t = x_t - \hat{\mu} - \hat{\phi}\, x_{t-1}$, and the variance recursion of Eq. (16), evaluated at $\hat{\theta}$, is iterated forward through the sample,

$$\hat{\sigma}_t^2 \;=\; \hat{\omega} + \hat{\alpha}\, \hat{\varepsilon}_{t-1}^2 + \hat{\beta}\, \hat{\sigma}_{t-1}^2, \qquad t = 2, \ldots, T, \tag{19}$$

and its square root $\hat{\sigma}_t$ is the filtered conditional volatility: one number per day, in the units of the underlying series, summarizing how turbulent that day was expected to be given the information available the day before. The standardized residuals $\hat{z}_t = \hat{\varepsilon}_t / \hat{\sigma}_t$ provide the model's own specification check: if the filter has absorbed the volatility clustering, the squared standardized residuals $\hat{z}_t^2$ should display no remaining autocorrelation, and we report their first-order autocorrelation for every fitted model. We fit the same GARCH$(1,1)$ variance equation to the three series, with an AR(1) conditional mean for the two level series (the index and the VIX) and a constant mean for the BVL return; the three filtered volatilities $\hat{\sigma}_t$ are placed on a common footing by standardizing each to zero mean and unit variance.

The EPU index and the VIX are persistent levels. Their fitted persistence $\hat{\alpha} + \hat{\beta}$ measures how slowly a shock to the conditional variance dies out: the unconditional variance $\omega / (1 - \alpha - \beta)$ exists only when this sum is below one, and at the boundary $\alpha + \beta = 1$ the process is the integrated GARCH (IGARCH) of Engle and Bollerslev (1986), in which shocks never decay and no long-run variance is defined. Our estimates lie below that boundary, though close to it, so the fitted variances mean-revert slowly.

We estimate the model in levels, with the AR(1) mean of Eq. (15). The correlations reported throughout, the parameters (ϕ, α, β), and the standardized residuals $\hat{z}_t$ are invariant to positive affine re-standardizations $x_t \mapsto a + b\, x_t$ with $b > 0$, so the results do not depend on the base-100 convention of Eq. (5). Only $\hat{\mu}$, $\hat{\omega}$, and the units of $\hat{\sigma}_t$ change with the scale.

Section 6.2 reports, for each series, the estimates $(\hat{\mu}, \hat{\phi}, \hat{\omega}, \hat{\alpha}, \hat{\beta})$, the persistence $\hat{\alpha} + \hat{\beta}$, the maximized log-likelihood, and the first-order autocorrelation of $\hat{z}_t^2$, and then places three levels on one plane: the index, the VIX, and the conditional volatility $\hat{\sigma}_t$ of the BVL return, each standardized so that series measured in different units share a panel, against the crisis windows of

Section 5. Pearson correlations among the three levels, on the full sample and within the crisis and calm subsamples, complete the comparison. We ask whether the three conditional-volatility series cluster in the same episodes; Section 6.2 shows that the two market series do, while that of the index does not, for a reason that is itself informative about the instrument.

4.5 What the index may add, and its limitations

The three tools speak to convergent validity: agreement with the VIX and with BVL realized volatility would support the claim that the index measures uncertainty. Perfect agreement is neither expected nor desirable. The index is domestic and political, built from the discourse of Peruvian politicians, journalists, economists, and analysts, so it registers events (presidential vacancy votes, cabinet collapses, contested elections) that a United States equity-option index reflects weakly and that domestic volatility may reflect only with delay. We document where the index moves while the benchmarks do not, and the reverse.

Table 6 sets these differences out. Beyond the domestic and political focus, the measures differ in orientation and anchoring: the index reads uncertainty as expressed in discourse, the VIX as the market's expectation of future volatility, and BVL realized volatility as a summary of past returns, with only the benchmarks tied to a traded price. This is why we expect moderate rather than perfect co-movement, and why the index is not redundant with either benchmark.

Both comparisons separate crisis days from calm ones, on the expectation that a measure of uncertainty tracks the benchmarks most closely when uncertainty is high. The split rests on a calendar fixed in advance, dated from the historical record rather than from the benchmarks it evaluates. Section 5 sets out these windows for Peru over 2018–2023, including the pandemic.

5 Crisis Periods in Peru, 2018–2023

The crisis/calm split of Section 4.2 requires an explicit list of crisis windows. We date them from the documented political, institutional, and macroeconomic record—following Romer and Romer (2010) and narrative event-based risk indices (Caldara and Iacoviello, 2022)—and never from realized or implied market volatility, so the split stays exogenous to the VIX and to BVL realized volatility. An episode qualifies when it combines (i) a datable triggering event, (ii) a rupture in an institution—the presidency, Congress, the electoral authority—or a nationwide economic shock, and (iii) sustained attention rather than a one-day headline. In five years Peru had six presidents, a dissolved Congress, a contested election, an attempted self-coup, and a pandemic. Table 7 lists the eight windows.

C1. Fall of Kuczynski (March 2018). Pedro Pablo Kuczynski resigned the presidency on 21 March 2018 after the release of vote-buying recordings, and Martín Vizcarra was sworn in on 23 March. We date it 12 March–6 April 2018.

C2. Dissolution of Congress (September–October 2019). On 30 September 2019 Vizcarra dissolved Congress, which contested the act and swore in Vice President Mercedes Aráoz, who resigned within a day; the clash briefly left two claimants to the executive. We date it 27 September–15 October 2019.

C3. COVID-19 emergency (March–June 2020). The state of emergency and lockdown were declared on 15 March 2020, the sharpest output contraction in the sample and coincident with the global risk-off shock. It is the one window macroeconomic and global rather than domestic-political. We date the acute phase 15 March–30 June 2020, ending at the first major easing.

C4. Vacancy of Vizcarra and Merino interregnum (November 2020). Congress vacated Vizcarra on 9 November 2020; Manuel Merino took office on 10 November and resigned on

15 November after protests in which two demonstrators were killed; Francisco Sagasti was sworn in on 17 November. Three presidents in one week. We date it 9–18 November 2020.

C5. Contested general election (April–July 2021). The first round (11 April 2021) and the runoff (6 June) between Pedro Castillo and Keiko Fujimori were followed by weeks of annulment disputes before the electoral authorities; Castillo was proclaimed on 19 July and inaugurated on 28 July. We date it 8 April–28 July 2021, from the first round to the inauguration.

C6. Fuel-price protests and Lima stay-at-home order (March–April 2022). Rising fuel and fertilizer prices after the invasion of Ukraine triggered a nationwide transport strike. Near midnight on 4 April 2022 the government issued Supreme Decree 034-2022-PCM, ordering a 22-hour mandatory stay-at-home period for Lima and Callao on 5 April; it was revoked that same afternoon amid open defiance. Economic in origin, institutional in expression. We date it 28 March–12 April 2022.

C7. Castillo self-coup and removal (December 2022). On 7 December 2022 Castillo announced the dissolution of Congress; within hours he was impeached, removed, and detained, and Dina Boluarte was sworn in. With C4, this is the sharpest institutional rupture of the sample. We date it 7–20 December 2022.

C8. Post-Castillo protests (December 2022–January 2023). Castillo's removal set off sustained, at times fatal, nationwide protests demanding new elections and Boluarte's resignation, past the sample end. We date it 20 December 2022–31 January 2023, right-censored by the sample end.

The eight windows cover 347 of the 1 835 sample days (18.9%), leaving 1 488 calm days. Coverage is uneven by design: the pandemic (C3, 108 days) and the contested election (C5, 112 days) supply 63% of the window days, while the other six are 10-to-43-day shocks. Every year holds at least one window, so the split is not identified off a single episode. C7 and C8 are adjacent, sharing 20 December 2022; we keep them separate, as an elite institutional act differs from the popular mobilization that followed, and count the shared day once.

Figure 10 plots the windows against the headline EPU index (top) and the political-uncertainty (PU) index of Section 3 (bottom). The windows capture turbulence: the headline index averages 113 inside them against 97 outside, and its maximum (272 on 15 April 2020, 2.7 times the sample average) falls in C3. The response is heterogeneous. The strict index rises where the crisis is economic—the strict tweet share runs at 1.76 times its full-sample rate within C3 and 1.49 within the fuel-price window C6—and falls in the purely political ruptures, to half its rate in C4 and about two-thirds in C7 and C8, where the PU index spikes instead (mean 174 in C7, a $P \cap U$ share 1.54 times its rate). Two mechanisms drive the gap. Composition: in an institutional collapse the elite conversation turns political and uncertain without an economic frame, satisfying $P \cap U$ but not the triple condition $E \cap P \cap U$. Dilution: daily volume explodes, 691 tweets per day in C7 against a sample mean of 116, inflating the denominator of Eq. (3). The sparsity of the strict condition in political crises is what motivates carrying the PU index alongside it.

Two caveats govern how the figure is read. Because the index is a trailing 14-day rate, in short windows its local maximum sits near the window's end—15 October in C2, 18 November in C4—rather than on the trigger date. And the largest readings outside any window, on 8–12 February 2018 (up to 208) and 10 August 2020 (188), fall in the thin-volume opening months, C1 itself averaging 16 tweets per day, so early-sample readings carry sampling noise. The eight-window list is therefore conservative: tension outside the windows works against the crisis/calm contrasts, which can be read as lower bounds. Every window edge is displaced by ±5 days and each correlation recomputed, so that the split is assessed against its own dating rather than assumed exact.

6 Empirical Results

We apply the three tools of Section 4 to the headline index of Eq. (5) over the $T = 1\,835$ days of the sample. Every statistic is computed on the frozen daily series and is invariant to the base-100 convention of Eq. (5).

6.1 Similarity Analysis

Table 8 reports the static co-movement of the index with the two benchmarks. Over the full sample the index and the VIX correlate at 0.437 (Pearson) and 0.408 (Spearman); the two agree, so the association is linear rather than merely ordinal. The regime split is in the direction a measure of uncertainty should show, $\rho^{\text{crisis}} = 0.665$ against $\rho^{\text{calm}} = 0.259$: the similarity more than doubles when Peru is in a documented rupture, and the Spearman values move the same way (0.680 against 0.324).

Inference on these correlations must account for persistence. The fitted score of Eq. (10) has a first-order autocorrelation of 0.949 for the VIX, so consecutive days carry almost the same information and the classical standard error, 0.021 on the full sample, treats each as fresh. The Newey–West estimator of Eq. (12) returns 0.117 (Andrews plug-in $L = 99$), a factor of 5.6 larger. Even so, the full-sample correlation is 3.7 HAC standard errors from zero, the crisis correlation 6.7, and the calm correlation 3.3: the similarity survives the most conservative inference available.

The index shows no such relation with BVL realized variance $\text{RV}_t = r_t^2$. The Pearson correlation is 0.033 against a HAC standard error of 0.025, and the crisis and calm values, 0.012 and 0.025, are equally indistinguishable from zero; the ± 5-day dating check keeps them inside $[-0.031, 0.052]$. Here the HAC correction barely moves the standard error (0.025 against a classical 0.023), because the score autocorrelation is 0.119 rather than 0.949: the correction is large exactly where the persistence is and negligible where it is not, which is the check promised in Section 4.2. The absence is a statement about frequency, not about the index. RV_t is a one-day object that reverts within a day, while the index is a trailing fortnightly rate; a day on which Lima moves sharply is invisible to a two-week average. What the two might share is not the level of a squared return but the smooth, persistent component of variance, which the GARCH filter of Section 6.2 recovers.

The nature of the EPU–VIX similarity is between-episode, not day-to-day, and three readings agree on this. Computed window by window, the correlation ranges from -0.799 in C1 to 0.774 in C6 and is only 0.059 inside the pandemic window C3; what lifts the pooled crisis figure to 0.665 is that C3 carries an index mean of 185.4 and a VIX mean of 39.32, far above the sample averages of 100 and 21.36, so the two are jointly elevated in the same episodes rather than tracking each other within them. The rolling correlation of Figure 9 tells the same story: it averages 0.018 over the sample and is negative on 47.8% of windows, rising to 0.212 inside the crisis bands (peaking at 0.778 at the close of C3, the window ending 1 July 2020) and sitting at -0.026 outside. And in first differences the association vanishes, -0.012 over the sample, 0.023 in crises, and -0.021 in calm, all indistinguishable from zero (Figure 11). The similarity is therefore a low-frequency, between-episode agreement: the index and the VIX are elevated together in the same crises, but the index does not echo the daily movements of the VIX. This is not a weakness. It is the signature of a measure that is not redundant with the benchmark, and its sharpest expression is the trough of -0.642 on 3 November 2020, when the conversation before the Vizcarra vacancy turned political without an economic frame and the index fell while the VIX rose into the United States election: the two diverged because the index was measuring what was happening in Peru. The EPU–RV_t rolling curve, by contrast, is flat throughout (mean -0.006, negative on 56.9% of windows).

The correlations of Table 8 measure co-movement over the whole support of each series. A

different question, and the one a user of the index would ask first, is whether the index separates the days a historian would call a crisis from the days around them. Table 9 and Figure 12 answer it with the area under the receiver operating characteristic curve, the Mann–Whitney estimate of $\Pr\left(X^{\text{crisis}} > X^{\text{calm}}\right)$. The statistic suits the similarity framing of this paper because it fits no model, imposes no lag, and reads in one direction only: 0.5 is a coin flip, values above it mean the series runs higher on crisis days, values below it mean the reverse. The labels come from Table 7, dated from event chronologies rather than from market data, so the exercise is not circular.

Pooled over all 347 crisis days the headline index scores 0.541, barely above the coin flip, and read alone that number would condemn the index. It is a mixture artifact. Splitting the windows by the character of the episode, which Table 7 already records and which uses no information from any series, resolves it: the index scores 0.944 on the 124 days of the two economic windows and 0.317 on the 223 days of the six political ones. Panel A of Table 9 shows the pattern holds window by window. The COVID-19 emergency C3 gives 0.962 and the fuel-price episode C6 gives 0.823, while the Merino interregnum C4 gives 0.094 and the post-Castillo protests C8 give 0.264. The index is close to a perfect detector of the economic windows and an inverted one of the political windows.

This is a property of the construction rather than a defect of it. The headline index counts a tweet only in the $E \cap P \cap U$ cell of Section 3: a constitutional crisis with no immediate economic content does not qualify, and the fortnightly denominator falls as political traffic that fails the economic test crowds the timelines. The political-uncertainty index, which requires only $P \cap U$, is the mirror image. Its pooled score of 0.694 on the political windows understates it for the same reason the headline figure of 0.541 understated the index, and the culprit is identifiable: the contested general election C5 spans 112 of the 223 political days and returns 0.488, so half that pooled sample is a single window no series in the table separates. Dropping C5 leaves 111 political days on which the political index scores 0.902, against 0.359 for the headline index. The difference of 0.543 carries a DeLong z of 23.13 and a block-bootstrap interval of [0.408, 0.653]. On the economic windows the same difference is -0.247 with interval $[-0.386, -0.111]$. The sign reverses with the character of the episode, and the two scores are close to mirrors: 0.944 for the headline index on economic days, 0.902 for the political index on political days. The indices are not competing measures of one quantity. They partition the crisis calendar between them.

The benchmarks place this reading. The VIX scores 0.901 on the economic windows and 0.477 on the political ones, which is the coin flip; restricting to the political windows other than C5 lifts it only to 0.544, against 0.902 for the political index. It registers the pandemic because the pandemic was global, and it is close to blind to every domestic political episode in the sample. Realized volatility scores near 0.58 on all windows and 0.520 once C5 is dropped, discriminating weakly and indiscriminately. No series available off the shelf separates Peruvian political crises from calm days, and this is the gap the index family of this paper fills. C5 is a window of 112 days that averages an acute fortnight together with three quiet months, and its resistance to every series in Table 9 is a property of the dating in Table 7, which is coarse by construction, rather than of the indices.

6.2 Conditional Volatility

Table 5 reports the GARCH$(1,1)$ estimates for the three series. The VIX has a large news coefficient and a moderate memory coefficient, $\hat{\alpha} = 0.335$ and $\hat{\beta} = 0.620$: implied volatility reprices sharply and forgets within weeks, a half-life of about two weeks. The BVL return sits at the opposite corner, $\hat{\alpha} = 0.080$ and $\hat{\beta} = 0.918$, the familiar equity configuration in which the variance carries almost all the memory; its persistence, 0.9979, is so close to the integrated boundary that the implied half-life describes how slowly the filter forgets rather than a horizon anyone would forecast

over. The index is nearer the equity corner, $\hat{\alpha} = 0.072$ and $\hat{\beta} = 0.921$, with persistence 0.9926, and its mean coefficient $\hat{\phi} = 0.988$ reproduces the near-unit-root persistence Table 4 found in the level; its half-life is to be read with the same caution as the BVL one.

The filter absorbs the volatility clustering. The first-order autocorrelation of the squared standardized residuals is 0.009, −0.005, and 0.019 for the index, the VIX, and the BVL return, and the Ljung–Box statistic of Ljung and Box (1978) at ten lags gives p-values of 0.99, 1.00, and 0.06. Two caveats attach. The mean equation is deliberately minimal, so the level residuals retain autocorrelation (a fourteen-day rate is a moving average of order thirteen, which an AR(1) cannot span); and the standardized residuals are far from Gaussian, so Eq. (18) is a quasi-likelihood, consistent for (ϕ, α, β) under Eqs. (15)–(16) even when the innovation is not normal (Bollerslev and Wooldridge, 1992).

Figure 13 places the three levels on one standardized scale: the index, the VIX, and the conditional volatility $\hat{\sigma}_{\text{BVL}}$. All three rise together in the crisis windows, most sharply in the pandemic window C3, and settle together between them; the two market measures track each other closely and the index moves with both, its swings a little wider and its timing a little looser, as a domestic text-based measure and a global financial one should. Table 10 confirms this. The two market measures agree most, correlating at 0.625 over the sample and 0.702 within crises, which shows the GARCH filter recovers a genuine volatility. The index sits alongside both, at 0.437 with the VIX and 0.334 with $\hat{\sigma}_{\text{BVL}}$, the latter stable across regimes (0.349 in crises, 0.274 in calm). The three correlations are moderate, between 0.33 and 0.44 for the index: the co-movement is real and the series share a common uncertainty component, while each carries a dimension the others do not.

Figure 14 explains why the index enters through its level and not a volatility of its own. The index is a rate estimated on a finite number of tweets, with sampling error $(\rho_t (1 - \rho_t) / \mathcal{N}_t)^{1/2}$, where $\mathcal{N}_t$ is the trailing fourteen-day tweet volume, so a day-to-day movement is part signal and part measurement noise, and the noise shrinks as the platform gets busier. Regressing $\ln \hat{\sigma}_t^2$ on $\ln \mathcal{N}_t$ returns a slope of −0.506 with

$R^2 = 0.601$; a pure binomial sampling term would deliver a slope of −1, so a large part of the variation in the index's conditional variance tracks the precision of the measurement rather than turbulence in the underlying rate. The slope is a log-log elasticity and R^2 measures the variation in $\ln \hat{\sigma}_t^2$ explained by volume. $\hat{\sigma}_t$ correlates at 0.760 with $\mathcal{N}_t^{-1/2}$. Because volume is procyclical to crises (156.9 tweets a day inside the windows against 106.8 outside), the index is measured most precisely exactly when the market measures are most turbulent: the two second moments are pushed in opposite directions by construction. The variance of EPU_t is therefore a measurement object, telling us how well the day's rate is estimated more than how turbulent discourse is. The similarity with the benchmarks lives in the first moment, and a variance equation augmented with volume would model the instrument rather than the uncertainty it measures. We report the symmetric Gaussian GARCH(1, 1) and stop there.

6.3 What the index contributes

The evidence supports a bounded claim. The index is a daily, interpretable, and replicable series of economic and political uncertainty for Peru, built from public platform text and a documented classification, and it exists where no comparable instrument did. Its full-sample correlation is 0.437 with the VIX and 0.334 with Lima market volatility, rising to 0.665 with the VIX in crises; this places it among established uncertainty measures without making it a copy of any. What it adds is domestic and narrative. The VIX is a global financial gauge and BVL volatility a market reaction; neither registers a presidential vacancy vote, a cabinet collapse, or a contested election as an uncertainty event in its own right. The index does, and the episodes on which it and the

VIX diverge, C4 and C7 above all, are exactly the domestic institutional ruptures a United States equity-option index has no reason to price. Paired with the political-uncertainty index built from $P \cap U$, it separates economic-policy uncertainty from political rupture. We do not claim the index is a better measure of uncertainty in general; we claim it is the more appropriate instrument for domestic economic and political uncertainty in Peru, while the VIX and BVL volatility capture the global-financial dimension the index is not built to see.

7 Conclusions

We have built the first daily index of economic and political uncertainty for Peru from the X discourse of 97 influential accounts, classifying 213 249 tweets along economic, political, and uncertainty dimensions with a large language model. The index is the volume-weighted fortnightly rate of Eq. (3), rescaled to base 100 by Eq. (5).

In levels the index tracks the VIX, and the association is state-dependent: a full-sample Pearson correlation of 0.437 rises to 0.665 in the 347 crisis days and falls to 0.259 in the 1 488 calm days, an ordering that survives Newey–West inference at every bandwidth and a ± 5-day shift of every crisis boundary. It does not track the one-day realized variance of the BVL (correlation 0.033, within one HAC standard error of zero), because a squared daily return reverts within a day while the index averages a fortnight. Filtered through a symmetric Gaussian GARCH$(1, 1)$, the conditional volatility of the BVL return enters as a third level, correlating 0.625 with the VIX and 0.334 with the index: the two market volatilities agree most, confirming the filter recovers a genuine volatility. We do not add the index's own conditional volatility as a fourth series, because it is a measurement object: regressing $\ln \hat{\sigma}_t^2$ on $\ln \mathcal{N}_t$ gives a slope of -0.506 ($R^2 = 0.601$), so tweet volume, which rises in crises, makes the index most precise exactly when the benchmarks are most turbulent. The index carries its signal in the first moment (Figure 15).

The correlations lie below unity, so the index is not a rescaling of either benchmark, and above zero, so the three share a common component: the index is not redundant. What it adds is domestic and low-frequency. The VIX similarity is between-episode, not daily (in first differences the correlation is -0.012), and the episodes where the two diverge, C4 and C7, are the domestic institutional ruptures a United States equity-option index has no reason to price; there the headline index falls below its base while the political-uncertainty index built from $P \cap U$ rises to a mean of 174. These are claims of similarity, not causation: the trailing construction and near-unit-root persistence would make any leading or causal reading fragile.

References

Andrews, D. W. K., 1991, "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica* **59(3)**, 817–858.

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.

Bartlett, M. S., 1950, "Periodogram analysis and continuous spectra," *Biometrika* **37(1–2)**, 1–16.

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

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.

Bollerslev, T., and Wooldridge, J. M., 1992, "Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances," *Econometric Reviews* **11(2)**, 143–172.

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., et al., 2020, "Language models are few-shot learners," in *Advances in Neural Information Processing Systems (NeurIPS)* **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.

DeLong, E. R., DeLong, D. M., Clarke-Pearson, D. L., 1988, "Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach," *Biometrics* **44(3)**, 837–845.

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.

Engle, R. F., 1982, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica* **50(4)**, 987–1007.

Engle, R. F. and T. Bollerslev, 1986, "Modelling the persistence of conditional variances," *Econometric Reviews* **5(1)**, 1–50.

Frankel, J. A., 2010, "Monetary policy in emerging markets: A survey," in B. M. Friedman and M. Woodford (Eds.), *Handbook of Monetary Economics* **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.

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.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y., 1992, "Testing the null hypothesis of stationarity against the alternative of a unit root," *Journal of Econometrics* **54(1–3)**, 159–178.

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.

Ljung, G. M., and Box, G. E. P., 1978, "On a measure of lack of fit in time series models," *Biometrika* **65(2)**, 297–303.

Newey, W. K., and West, K. D., 1987, "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica* **55(3)**, 703–708.

Newey, W. K., and West, K. D., 1994, "Automatic lag selection in covariance matrix estimation," *Review of Economic Studies* **61(4)**, 631–653.

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.

Said, S. E., and Dickey, D. A., 1984, "Testing for unit roots in autoregressive-moving average models of unknown order," *Biometrika* **71(3)**, 599–607.

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

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.

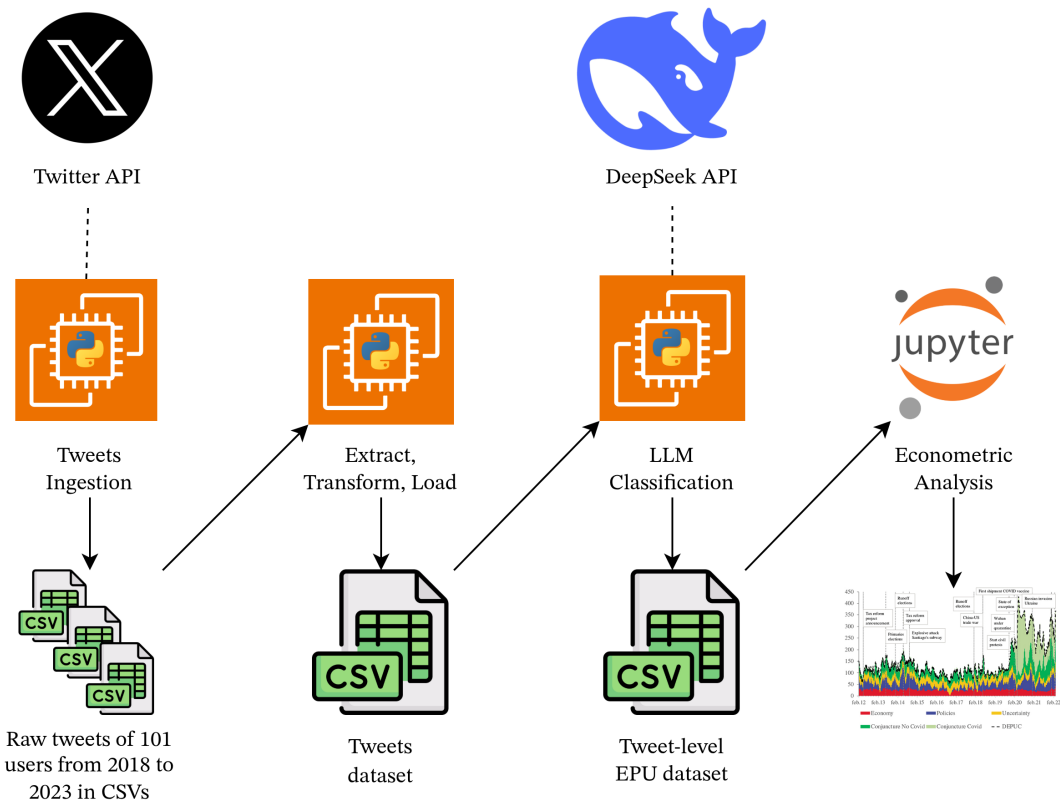

Figure 1. Pipeline architecture, from tweet ingestion to econometric analysis. The final sample comprises the 97 accounts selected by the criteria of Section 2; any larger initial count shown in the schematic is illustrative of the pre-selection stage.

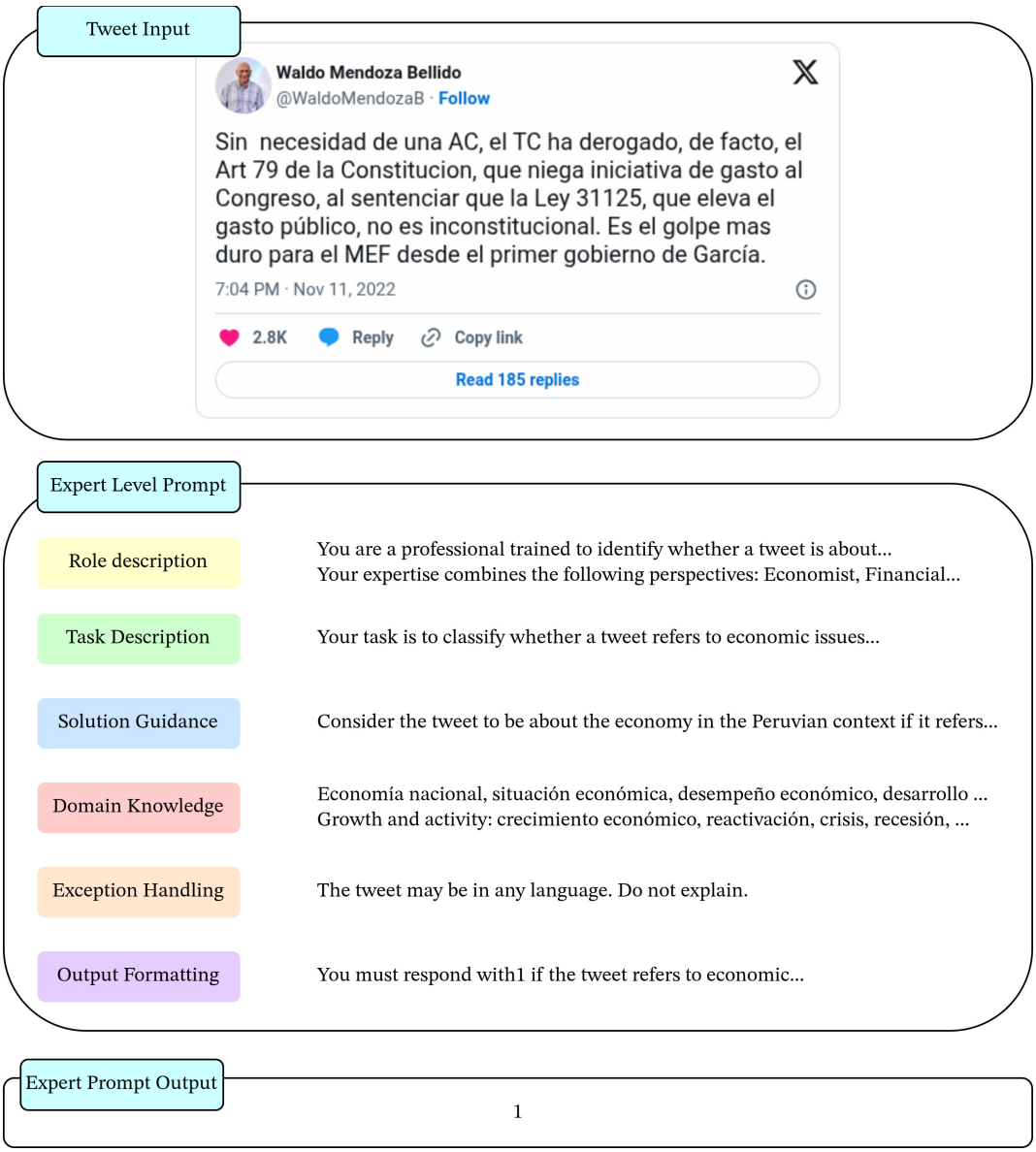

Figure 2. Expert-level prompt structure for LLM classification.

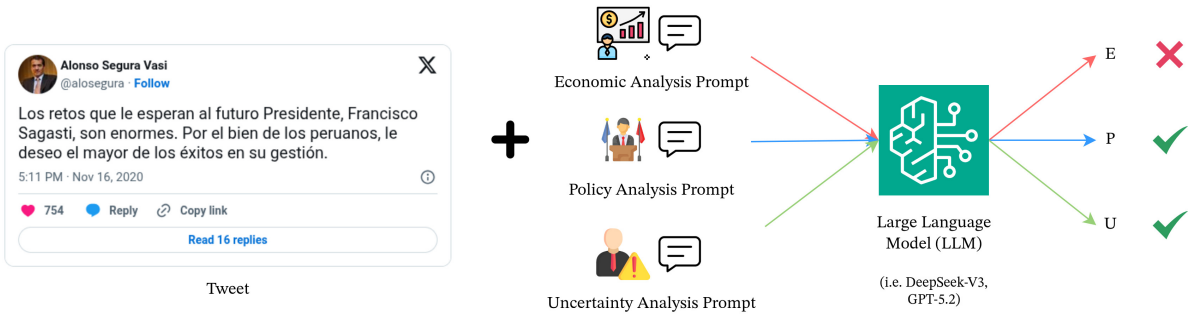

Figure 3. Three-dimensional labelling: each tweet is classified independently for E_{ijt}, P_{ijt}, U_{ijt}.

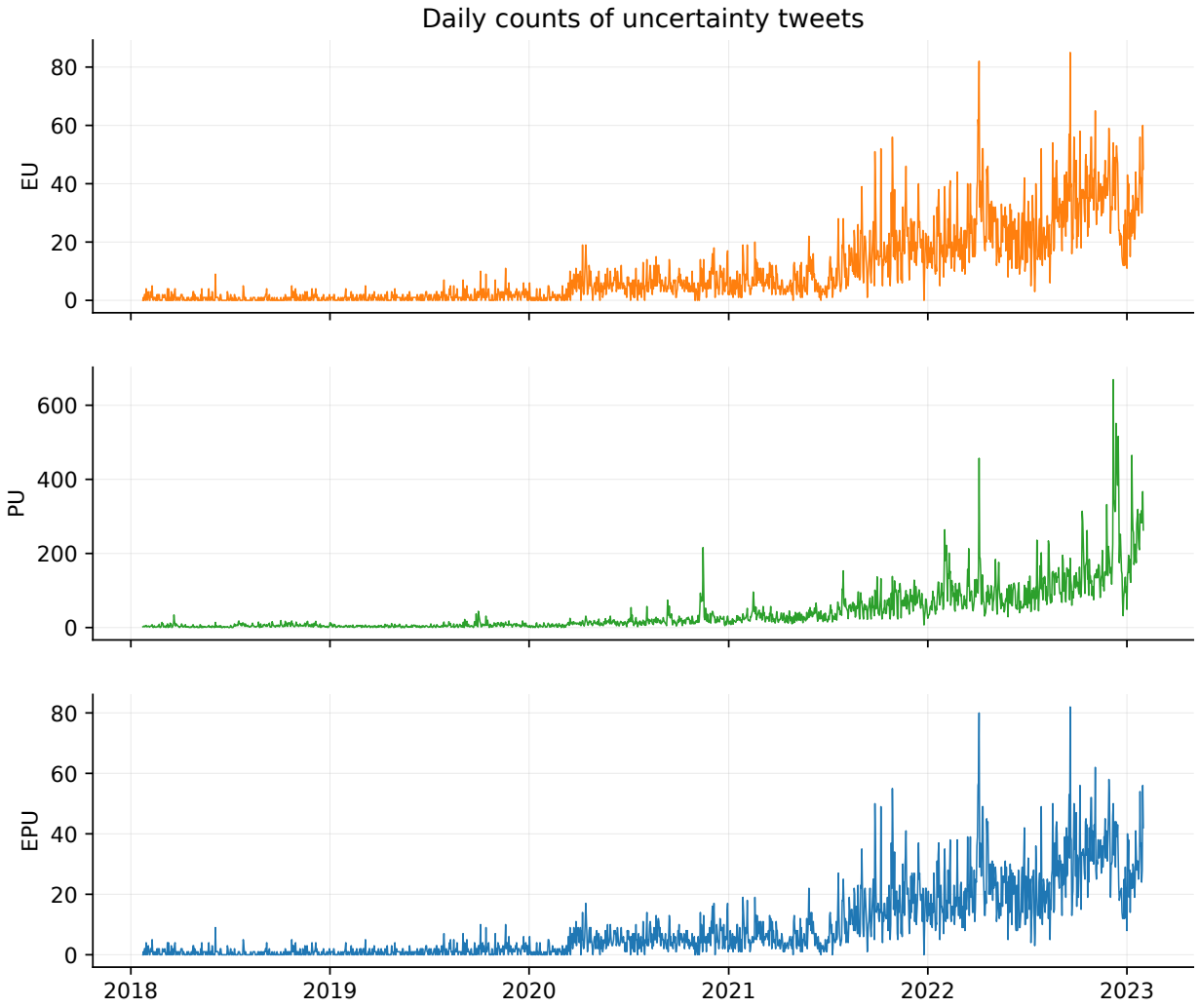

Figure 4. Daily counts by cell.

Daily counts of tweets that are, top to bottom, economic and uncertain ($E \cap U$), political and uncertain ($P \cap U$), and economic, political, and uncertain at once ($E \cap P \cap U$, the strict cell of Eq. (1)). The strict count and $E \cap U$ are nearly identical; $P \cap U$ is an order of magnitude larger and differently timed. All three rise with posting volume. Sample January 2018–January 2023; 97 accounts.

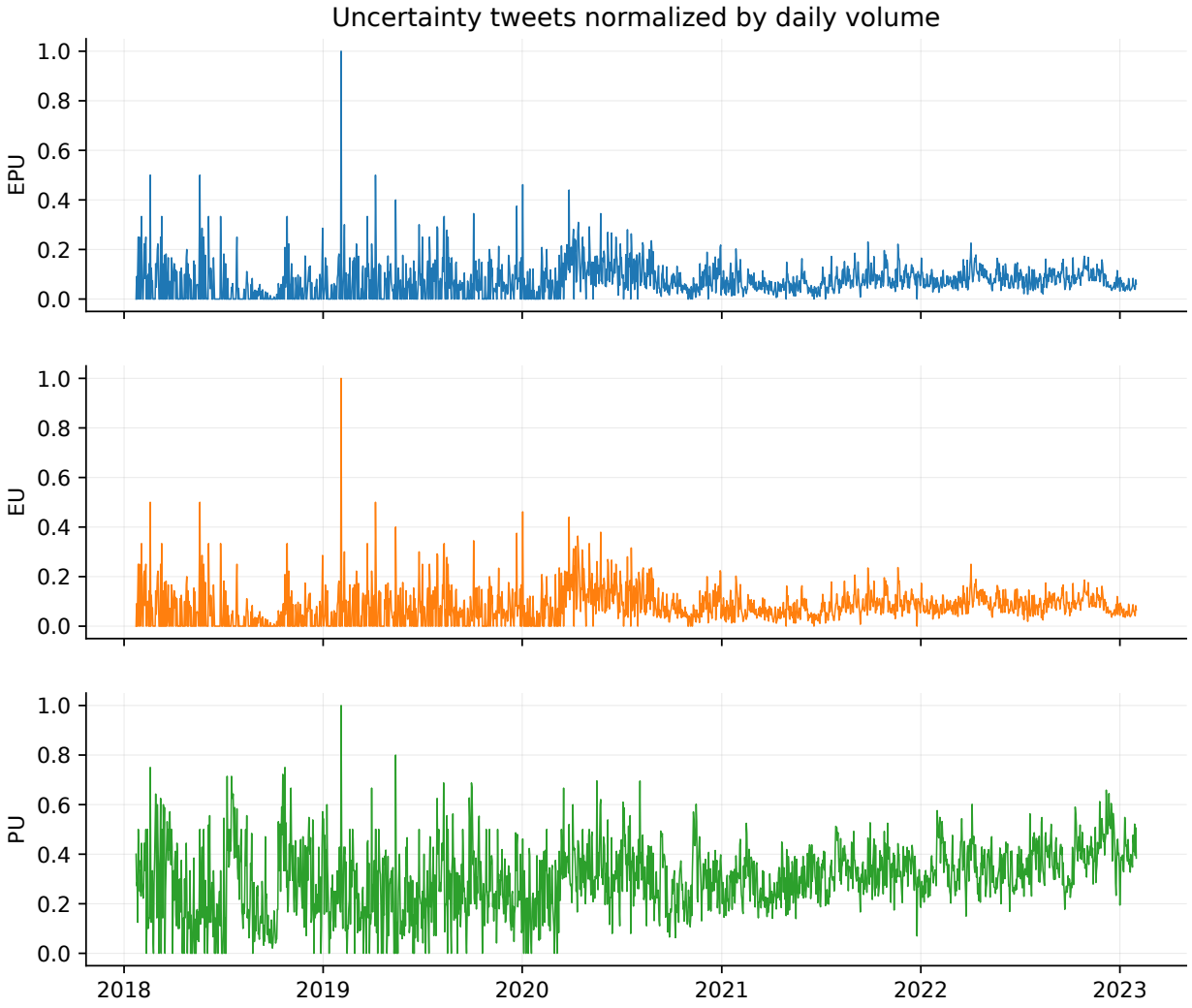

Figure 5. Daily shares.

The daily share $s_t = Y_t/N_t$ (Eq. (2)) for the strict cell and the two broader cells. Normalizing by daily volume removes the upward drift of Figure 4 but leaves heavy small-sample noise on low-volume days, when a tiny denominator lets a single tweet move the share sharply.

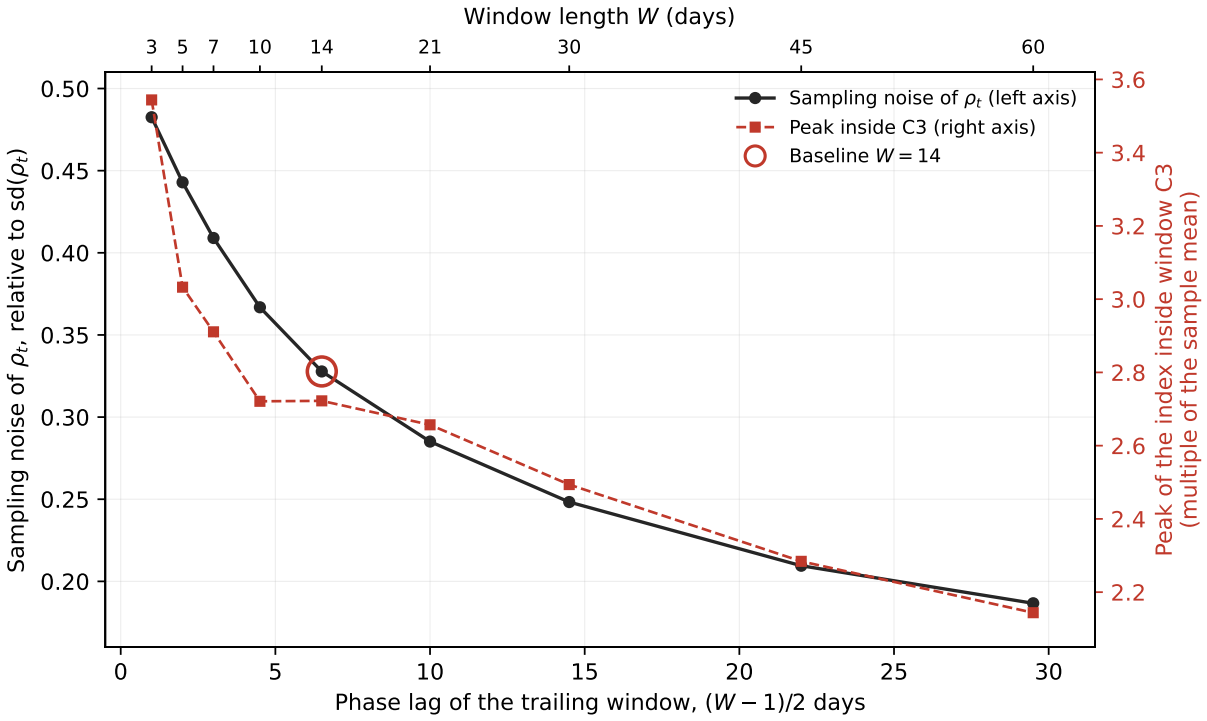

Figure 6. The window-length trade-off.

Each point is one trailing window W of Eq. (3), read on the top axis and placed on the bottom axis at the phase lag $(W-1)/2$ it imposes. Black, left axis: the mean sampling error of ρ_t, $\left(\rho_t\left(1-\rho_t\right)/\mathcal{N}_t\right)^{1/2}$, divided by the standard deviation of ρ_t, a scale-free noise ratio rather than a formal variance share. Red, right axis: the peak the index reaches inside the pandemic window C3, on 15 April 2020, in multiples of the sample mean. Pooling more days buys precision and pays in lag and amplitude; both curves fall monotonically, so no W minimizes them jointly. The circled baseline $W = 14$ costs a 6.5-day lag, a noise ratio of 0.33, and a retained peak of 2.72.

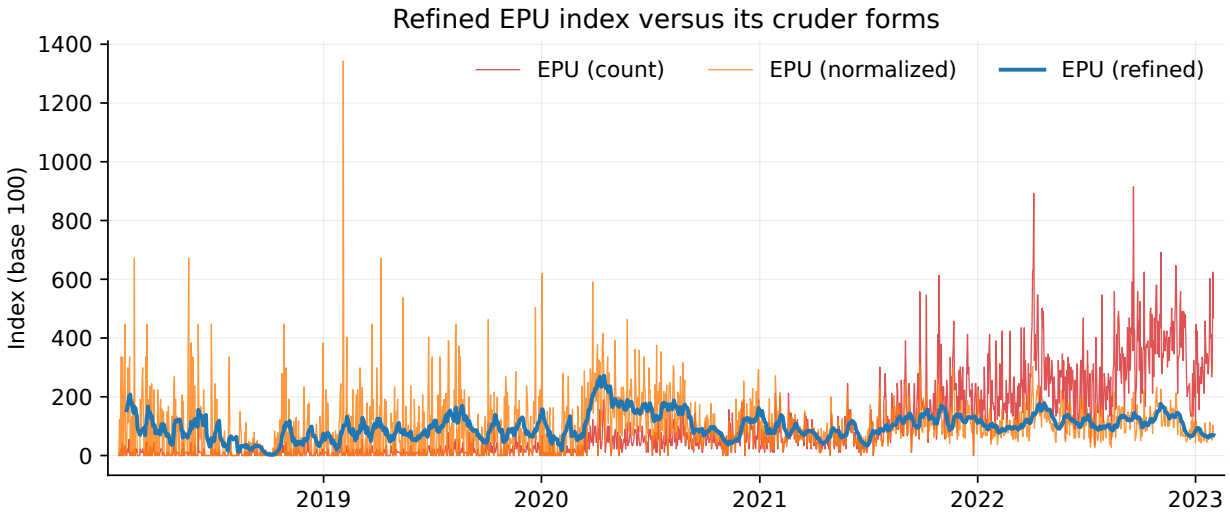

Figure 7. Refined index versus cruder forms.

The headline index (refined: the 14-day volume-weighted rate of Eq. (3), scaled to base 100 by Eq. (5)) against its raw count and normalized-share counterparts on the same scale. The cruder forms swing to many times their average; the refined index stays near its base and retains the episodes. Base 100.

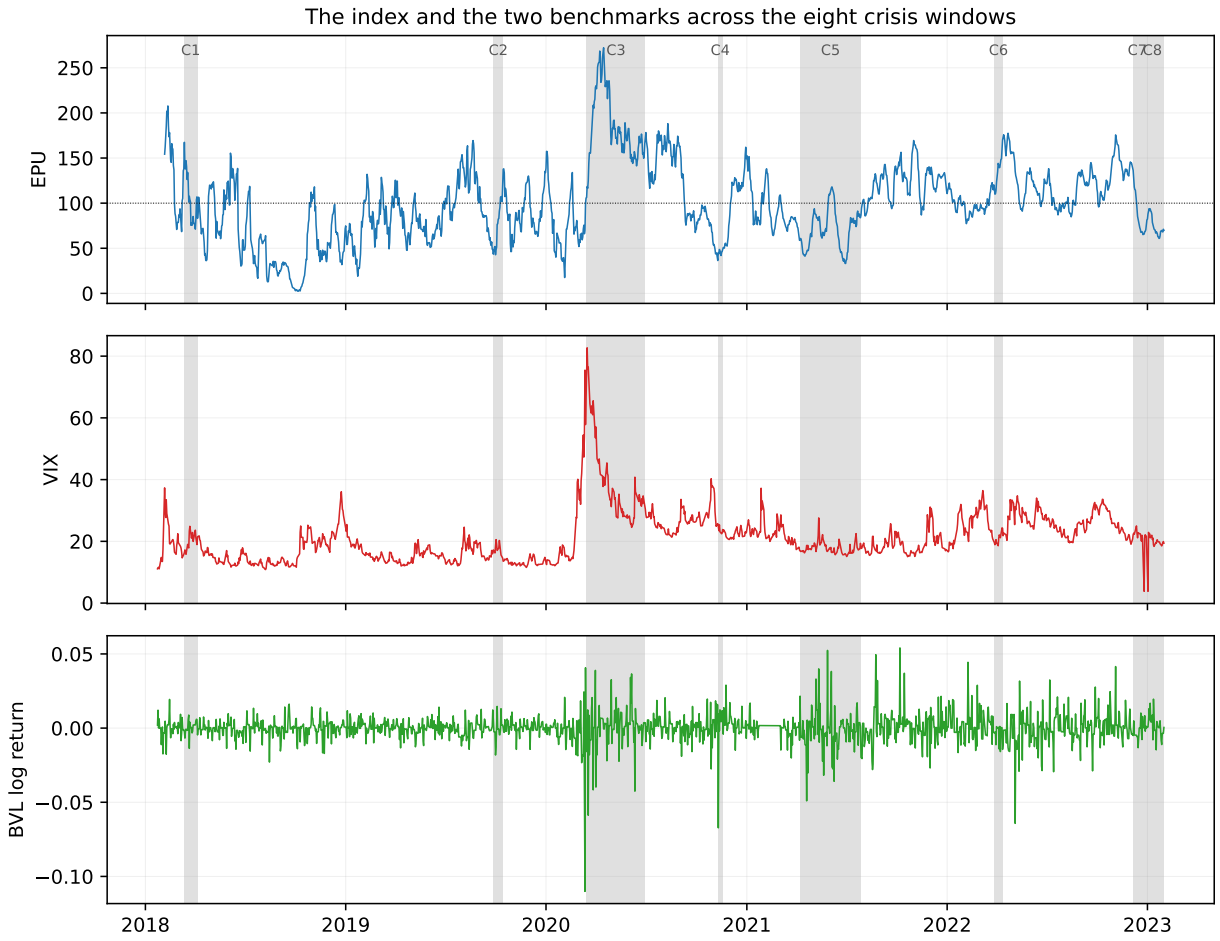

Figure 8. The index and the two benchmarks across the eight crisis windows.

Top to bottom: the headline EPU, the VIX, and the Lima-exchange log return, with the eight domestic crisis windows C1–C8 (2018–2023) shaded. The index rises inside the shaded windows, alongside spikes in the VIX and in market-return volatility.

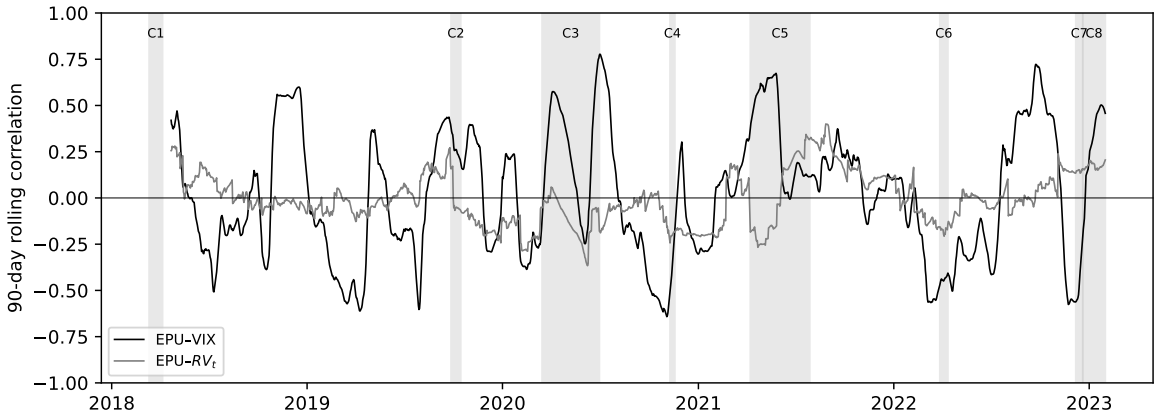

Figure 9. Rolling 90-day correlation of the EPU index with the VIX and with BVL realized variance.

Note: 90-day rolling Pearson correlation of Eq. (13). Shaded bands are the crisis windows C1–C8; $\mathrm{RV}_t = r_t^2$ is BVL realized variance.

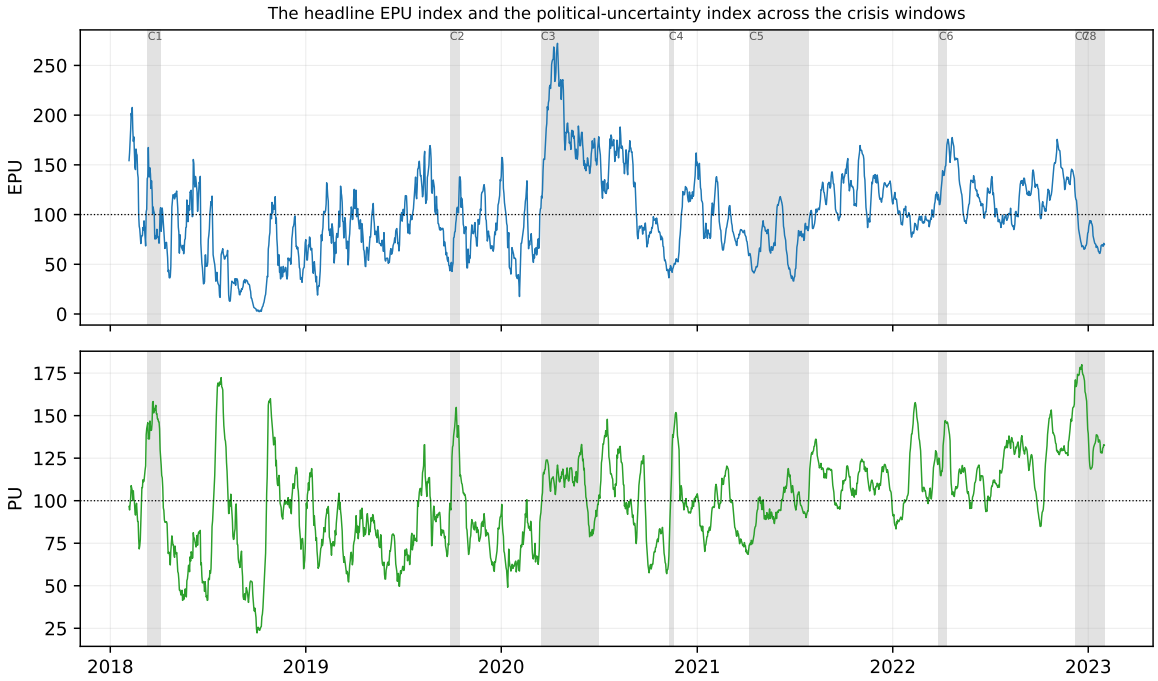

Figure 10. The headline index and the political-uncertainty index across the crisis windows.

Top: the headline EPU index of Eq. (5). Bottom: the political-uncertainty (PU) index, the same volume-weighted 14-day rate applied to the $P \cap U$ cell and rescaled to base 100. Shaded bands are the crisis windows C1–C8 of Table 7; the dotted line marks the base of 100. The headline index peaks at 272 on 15 April 2020, inside C3, and falls below its base in the purely political ruptures C4 and C7, where the PU index instead averages 174 in C7. The two indices resolve different crises: the strict cell requires an economic frame that an institutional collapse does not supply.

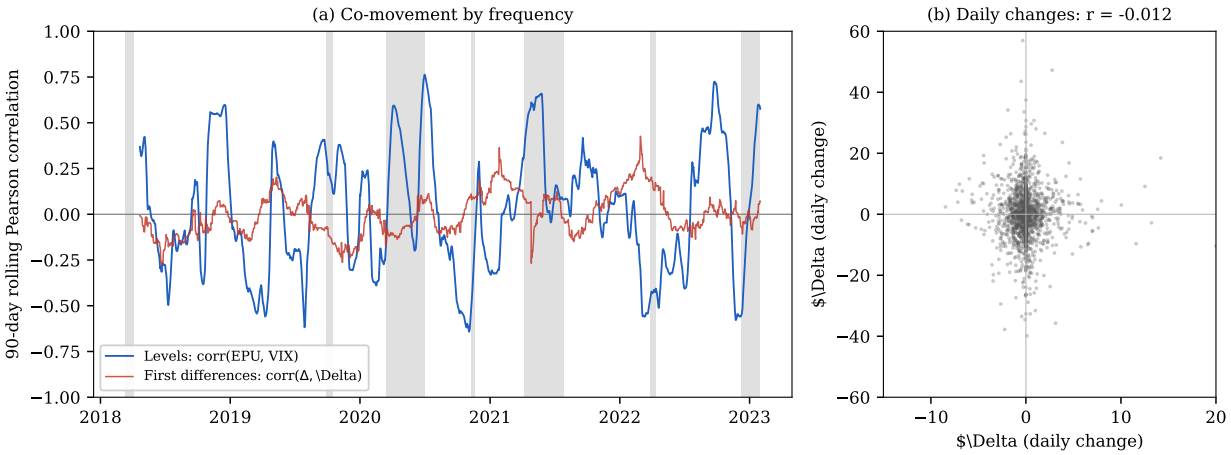

Figure 11. Level co-movement and first-difference co-movement of the index and the VIX.

Note: (a) 90-day rolling Pearson correlation of Eq. (13) between the index and the VIX in levels (solid) and in first differences (dashed), with the crisis windows C1–C8 shaded; (b) scatter of the daily change in the index against the daily change in the VIX. The level correlation swings with the crisis regimes while the difference correlation stays near zero throughout (−0.012 over the sample), so the similarity is a low-frequency, between-episode agreement rather than a daily one.

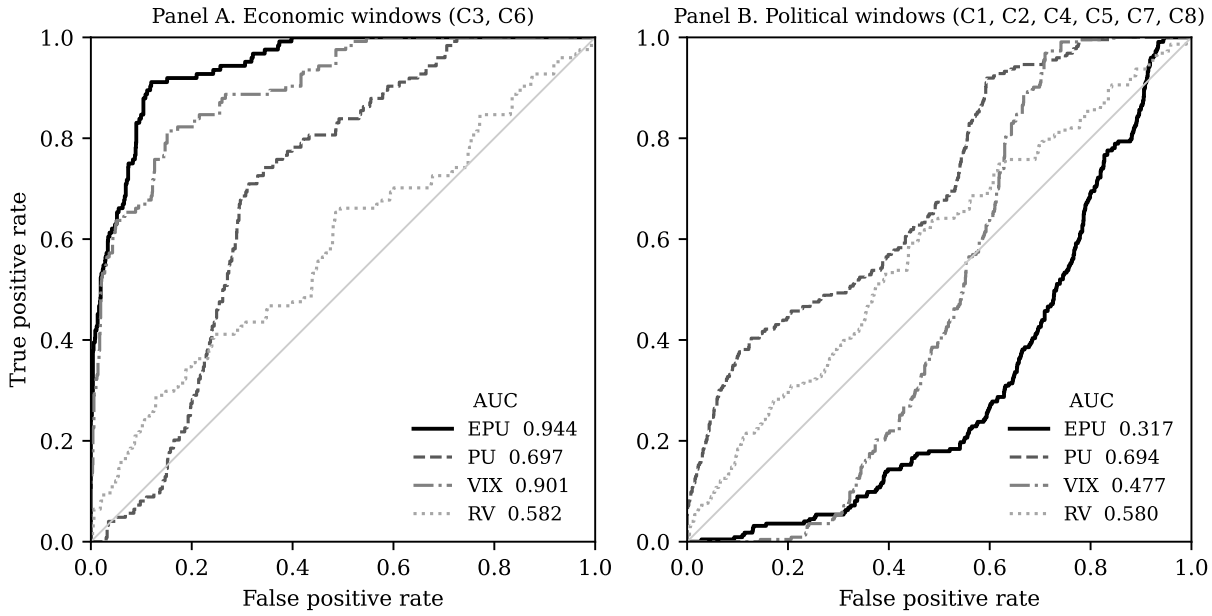

Figure 12. Receiver operating characteristic curves for crisis days.

Note: Panel (a) pools the economic windows C3 and C6 of Table 7 (124 days); Panel (b) pools the six political windows (223 days). In both panels the negative class is the same 1 488 calm days. Each curve traces the true positive rate against the false positive rate as the classification threshold on the series moves across its support; the diagonal is the coin flip. Legend entries report $\widehat{\text{AUC}}$. The headline index runs above the diagonal in Panel (a) and below it in Panel (b); the political-uncertainty index runs above the diagonal in both.

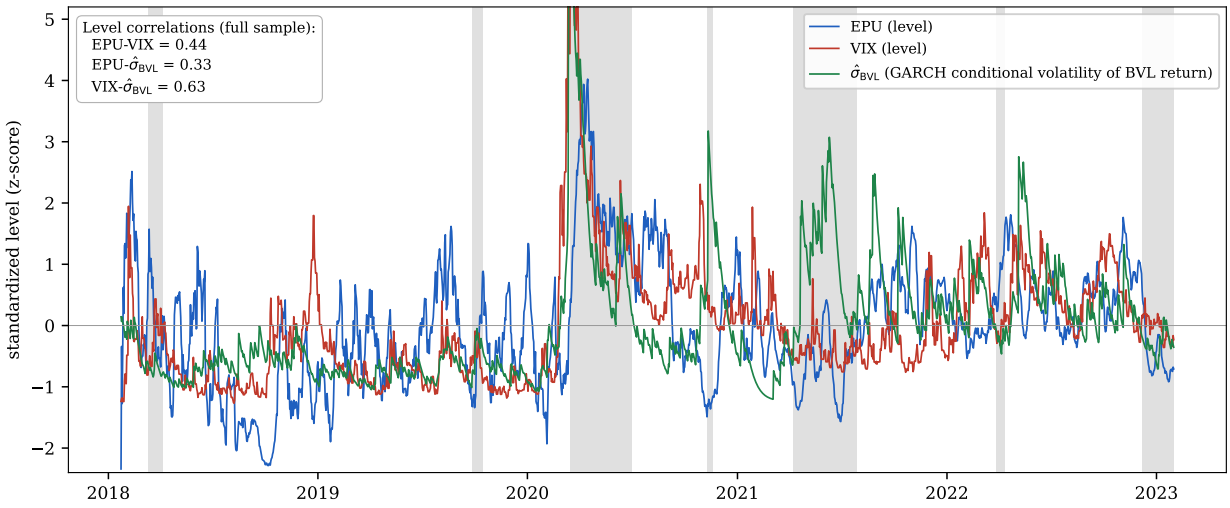

Figure 13. Three levels of uncertainty on the same plane.

Standardized levels of the headline index, the VIX, and the conditional volatility of the BVL return $\hat{\sigma}_{\text{BVL},t}$ from the GARCH$(1,1)$ model of Table 5, over the $T = 1\,835$ daily panel, with the crisis windows C1–C8 of Table 7 shaded. The GARCH filter is used only to turn the BVL return into a volatility level; the VIX enters as its own level and the index as its own level. The three series rise together in the crisis windows, most sharply in the pandemic window C3, and settle together outside them. The pairwise correlations are in Table 10.

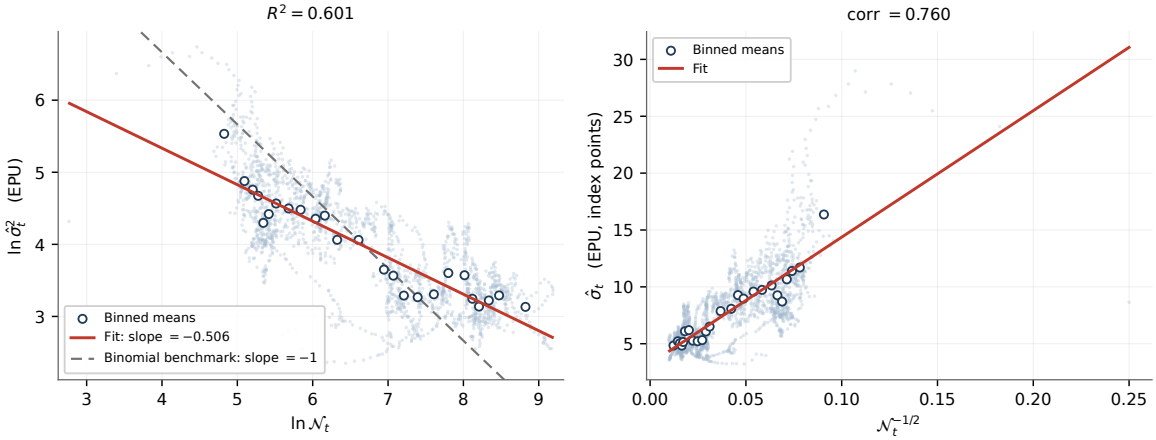

Figure 14. The conditional volatility of the index and the precision of its measurement.

Left: $\ln \hat{\sigma}_t^2$ of the index against $\ln \mathcal{N}_t$, the logarithm of the trailing fourteen-day tweet volume, with the fitted line; the slope is -0.506 and $R^2 = 0.601$, against the slope of -1 that a pure binomial sampling term would deliver. Right: $\hat{\sigma}_t$ against $\mathcal{N}_t^{-1/2}$, the shape of the sampling standard error of the rate ρ_t of Eq. (3); the correlation is 0.760. About half of the movement in the conditional variance of the index is the precision with which the day's rate is estimated.

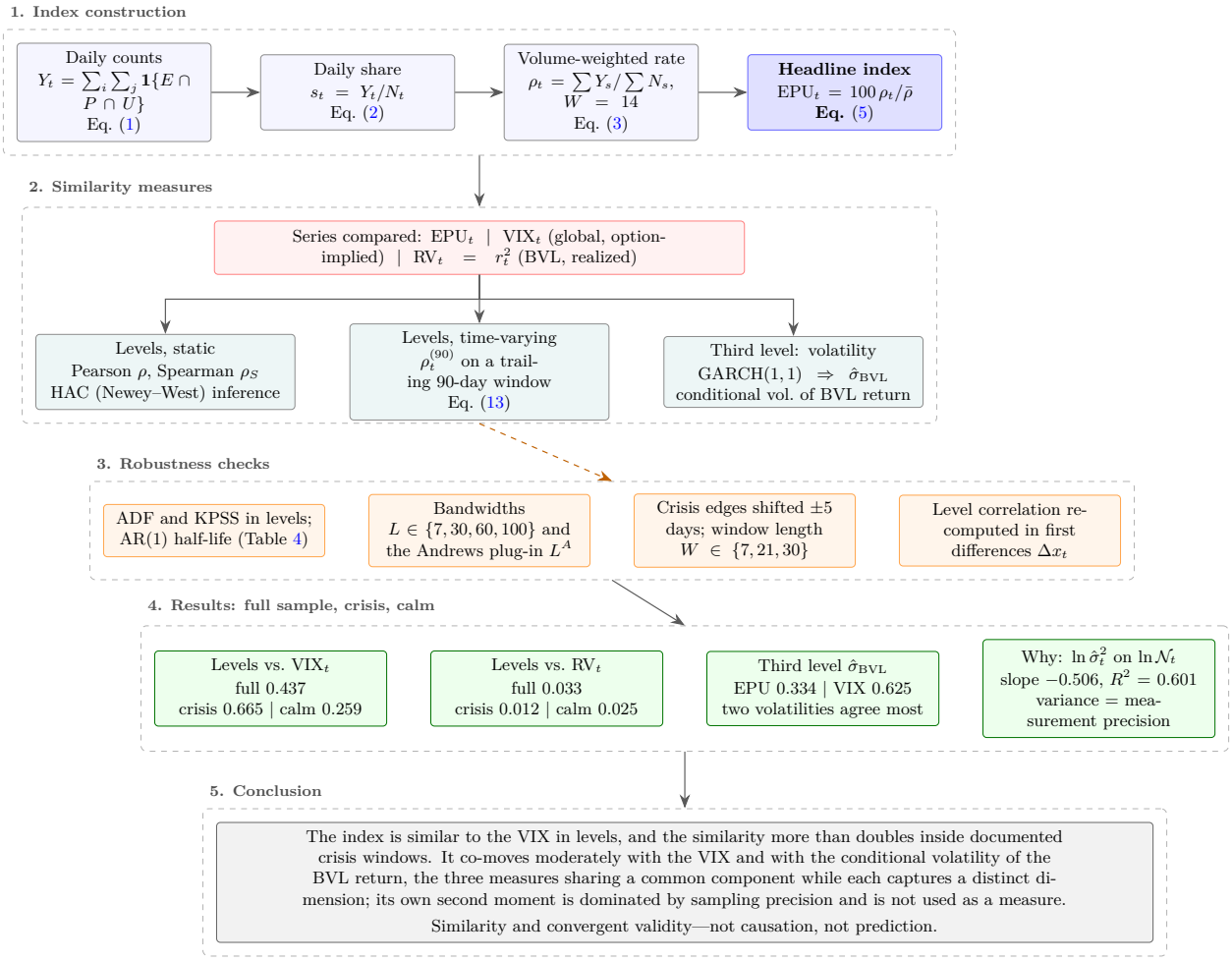

Figure 15. Synthesis of the paper.

The construction of the headline index (top), the three measures of similarity applied to the two benchmarks (middle left), the diagnostics that discipline them (middle right), the results on the full sample and on the crisis and calm subsamples (bottom), and what the evidence licenses (foot). Solid arrows carry data; dashed arrows carry robustness checks. Except where noted (the 14-day window, the HAC bandwidths, and the volume elasticity −0.506), the reported numbers are Pearson correlations computed on the frozen daily series of Section 2.

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. Label combinations (E, P, U).

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

Note: the eight cells partition the 213 249 classified tweets and sum to that total.

Table 3. Descriptive statistics.

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 strict 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

Note: N_t is the daily tweet volume and Y_t the daily strict-EPU count ($E \cap P \cap U$); the conditional mean is taken over the 1 462 days with $Y_t > 0$.

Table 4. Persistence and stationarity diagnostics.

Series	ADF t	Lags	KPSS $\hat{\eta}$	$\hat{\rho}_1$	Half-life (days)
EPU_t	-4.18	23	2.247	0.979	32.7
VIX_t	-4.30	17	2.683	0.973	25.3
$\mathrm{RV}_t = r_t^2$	-36.23	0	0.826	0.165	0.4

Note: $T = 1\,835$ daily observations, 23 January 2018 to 31 January 2023. Augmented Dickey–Fuller regression with a constant and no trend; the lag order is chosen by the Bayesian information criterion over $p \le \lfloor 12\,(T/100)^{1/4} \rfloor = 24$, and the 5% critical value is -2.86. The KPSS statistic of Eq. (9) tests level stationarity with a Bartlett bandwidth $\ell = \lfloor 4\,(T/100)^{1/4} \rfloor = 8$; the 5% critical value is 0.463. $\hat{\rho}_1$ is the first-order sample autocorrelation and the half-life is $\ln(1/2)/\ln \hat{\rho}_1$. The index and the VIX reject both nulls, the signature of the near-unit-root region; RV_t reverts within a day and its KPSS rejection reflects the rise in its mean from 2.5×10^{-5} before 2020 to 1.2×10^{-4} afterwards.

Table 5. GARCH$(1,1)$ estimates.

	EPU_t	VIX_t	r_t (BVL)
Mean equation	AR(1)	AR(1)	Constant
$\hat{\mu}$	1.001	0.409	0.018
$\hat{\phi}$	0.988	0.974	—
$\hat{\omega}$	0.5286	0.2915	0.00754
$\hat{\alpha}$	0.0715	0.3351	0.0798
$\hat{\beta}$	0.9210	0.6202	0.9182
Persistence $\hat{\alpha}+\hat{\beta}$	0.9926	0.9553	0.9979
Volatility half-life (days)	92.9	15.2	335.4
Long-run $\bar{\sigma}$	8.43	2.55	1.91
Log-likelihood	−6 289.62	−3 178.68	−2 156.15
Observations	1 834	1 834	1 835
$\hat{\rho}_1(\hat{z}_t^2)$	0.009	−0.005	0.019
Ljung–Box(10) on $\hat{z}_t^2$	2.68	1.44	17.81
p-value	0.99	1.00	0.06
Kurtosis of $\hat{z}_t$	9.8	35.8	9.9

Note: Gaussian maximum likelihood of Eq. (18), computed from scratch and maximized numerically from several starting points. The mean equation of Eq. (15) carries an AR(1) term for the two level series and a constant for the BVL log-return, which is expressed in percent for numerical conditioning; $\hat{\omega}$, $\hat{\mu}$, and $\bar{\sigma}$ are in the units of each series and are the only quantities that depend on scale. The volatility half-life is $\ln(1/2)/\ln\left(\hat{\alpha}+\hat{\beta}\right)$ and $\bar{\sigma}=\left(\hat{\omega}/\left(1-\hat{\alpha}-\hat{\beta}\right)\right)^{1/2}$ is the long-run conditional standard deviation of Eq. (17). The Ljung–Box statistic on $\hat{z}_t^2$ is the specification check of Section 4.4; the kurtosis of $\hat{z}_t$ is reported because it makes Eq. (18) a quasi-likelihood.

Table 6. Comparison of the index and the two benchmarks.

	EPU index	VIX	BVL volatility
Underlying data	Elite tweets	Index options	Squared returns
Geographic origin	Peru	United States	Peru
Uncertainty type	Political, economic	Financial (implied)	Financial (realized)
Temporal orientation	Contemporaneous	Forward-looking	Backward-looking
Price-anchored	No	Yes	Yes
Availability	Calendar days	Trading days	Trading days

Note: the VIX is option-implied from Standard & Poor's 500 (S&P 500) options and is forward-looking over about one month; BVL volatility is realized from squared log-returns of the Bolsa de Valores de Lima index and, in the GARCH$(1, 1)$ analysis, filtered as a conditional volatility; the index reflects uncertainty as it is expressed in discourse. Price-anchored indicates whether the measure derives from a traded price.

Table 7. Crisis windows in Peru, 2018–2023.

ID	Start	End	Days	Episode
C1	2018-03-12	2018-04-06	26	Fall of Kuczynski; transition to Vizcarra
C2	2019-09-27	2019-10-15	19	Dissolution of Congress
C3	2020-03-15	2020-06-30	108	COVID-19 emergency and lockdown
C4	2020-11-09	2020-11-18	10	Vizcarra vacancy; Merino interregnum
C5	2021-04-08	2021-07-28	112	Contested general election
C6	2022-03-28	2022-04-12	16	Fuel-price protests; Lima curfew
C7	2022-12-07	2022-12-20	14	Castillo self-coup and removal
C8	2022-12-20	2023-01-31	43	Post-Castillo protests
Union of windows			347	18.9% of the 1 835 sample days

Note: windows are dated from the historical record, not market volatility, keeping the crisis/calm split exogenous to the benchmarks. Inclusive calendar days. C7 and C8 share 2022-12-20, counted once in the union ($26 + 19 + 108 + 10 + 112 + 16 + 14 + 43 = 348$ window days; union = 347). C8 is right-censored by the sample end. Edges flagged as judgment calls are checked by shifting all windows ± 5 days.

Table 8. Static co-movement of the index with the two benchmarks.

				Standard error of $\hat{\rho}$					
Sample	T	$\hat{\rho}$	$\hat{\rho}_S$	Classical	$L=7$	$L=30$	$L=60$	$L=100$	L^A
Panel A. Benchmark $b_t = \mathrm{VIX}_t$									
Full	1 835	0.437	0.408	0.021	0.063	0.098	0.109	0.117	0.117
Crisis	347	0.665	0.680	0.040	0.082	0.101	0.093	0.079	0.099
Calm	1 488	0.259	0.324	0.025	0.052	0.072	0.080	0.084	0.080
Panel B. Benchmark $b_t = \mathrm{RV}_t$									
Full	1 835	0.033	0.110	0.023	0.026	0.022	0.020	0.020	0.025
Crisis	347	0.012	0.002	0.054	0.063	0.066	0.068	0.070	0.058
Calm	1 488	0.025	0.138	0.026	0.024	0.027	0.029	0.032	0.023

Panel C. Crisis boundaries displaced by ±5 days		
Benchmark	$\hat{\rho}^{\mathrm{crisis}}$ range	$\hat{\rho}^{\mathrm{calm}}$ range
VIX_t	$[0.605,\ 0.756]$	$[0.246,\ 0.293]$
RV_t	$[-0.031,\ 0.052]$	$[0.017,\ 0.085]$

Note: $\hat{\rho}$ is the Pearson correlation of Eq. (7) between the headline index EPU_t and the benchmark b_t; $\hat{\rho}_S$ is the Spearman correlation of Eq. (8). Classical is $\left(\left(1-\hat{\rho}^2\right)/T\right)^{1/2}$. The remaining columns are the Newey–West standard error of Eq. (12) at the stated Bartlett bandwidth; L^A is the Andrews autoregressive plug-in bandwidth, equal to 99, 39, and 55 for the VIX rows and to 5, 2, and 3 for the RV_t rows. The first-order autocorrelation of the fitted score $\hat{g}_t$ is 0.949 for the VIX and 0.119 for RV_t, which is why the correction inflates the VIX standard error by a factor of 5.6 and leaves the RV_t standard error essentially unchanged. Crisis days are the union of the eight windows of Table 7. Panel C shifts every window edge by -5, 0, and $+5$ days and reports the range of the resulting correlations.

Table 9. Crisis discrimination: area under the receiver operating characteristic curve.

ID	Type	Days	EPU_t	PU_t	VIX_t	RV_t
Panel A. Window by window, against the 1 488 calm days						
C1	Political	26	0.558	0.956	0.523	0.461
C2	Political	19	0.363	0.856	0.380	0.485
C3	Economic	108	0.962	0.671	0.950	0.612
C4	Political	10	0.094	0.652	0.714	0.757
C5	Political	112	0.276	0.488	0.412	0.640
C6	Economic	16	0.823	0.874	0.568	0.380
C7	Political	14	0.456	0.999	0.664	0.559
C8	Political	43	0.264	0.919	0.552	0.511
Panel B. Pooled by episode type						
All windows		347	0.541	0.695	0.629	0.581
Economic (C3, C6)		124	0.944	0.697	0.901	0.582
Political (C1–C2, C4–C5, C7–C8)		223	0.317	0.694	0.477	0.580
Political, excluding C5		111	0.359	0.902	0.544	0.520

Panel C. Paired difference $\widehat{\text{AUC}}\,(\text{PU}_t) - \widehat{\text{AUC}}\,(\text{EPU}_t)$

Subsample	Days	Difference	DeLong z	Bootstrap 95% interval
Economic	124	−0.247	−13.00	[−0.386, −0.111]
Political	223	+0.377	18.42	[+0.234, +0.516]
Political, excluding C5	111	+0.543	23.13	[+0.408, +0.653]

Note: entries in Panels A and B are $\widehat{\text{AUC}}$, the Mann–Whitney estimate of $\Pr\left(X^{\text{crisis}} > X^{\text{calm}}\right)$ for each series X, computed on the calendar panel of $T = 1\,835$ days. Crisis days are the windows of Table 7, dated from the historical record and therefore exogenous to every series compared here; calm days are the 1 488 days outside the union of the windows and form the same negative class in every row. A value of 0.5 is the coin-flip benchmark, 1 is perfect separation, and a value below 0.5 means the series runs lower inside the window than outside it. Hanley–McNeil standard errors range from 0.006 to 0.095 in Panel A and from 0.008 to 0.029 in Panel B. Panel C compares two areas estimated on the same pair of samples; the interval comes from a circular block bootstrap with mean block length 20 days and 2 000 replications, which allows for the serial dependence that the DeLong et al. (1988) variance ignores. Windows are typed as economic or political from the episode descriptions of Table 7, not from any series. PU_t is the political-uncertainty index of Section 3, built from the $P \cap U$ cell; $\text{RV}_t = r_t^2$ is Lima-exchange realized volatility. The area under the curve involves no fitted model and no lag structure: it compares two unconditional distributions and carries no directional reading.

Table 10. Similarity as three levels of volatility.

Pair of levels	Full	Crisis	Calm
$\rho\left(\mathrm{EPU}_t, \mathrm{VIX}_t\right)$	0.437	0.665	0.259
$\rho\left(\mathrm{EPU}_t, \hat{\sigma}_{\mathrm{BVL},t}\right)$	0.334	0.349	0.274
$\rho\left(\mathrm{VIX}_t, \hat{\sigma}_{\mathrm{BVL},t}\right)$	0.625	0.702	0.505

Note: Pearson correlation on the $T = 1\,835$ daily panel. $\hat{\sigma}_{\mathrm{BVL},t}$ is the conditional volatility of the BVL log-return from the GARCH$(1,1)$ model of Table 5; the VIX and the index enter as their own levels. The two volatilities, the VIX and $\hat{\sigma}_{\mathrm{BVL}}$, agree the most, which confirms that the filter recovers a genuine volatility; the index co-moves with both. Crisis days are the union of the eight windows of Table 7. The correlations are moderate rather than high, between 0.33 and 0.44 for the index over the full sample and 0.259 to 0.665 across regimes, consistent with three measures that share a common uncertainty component while each also captures a distinct dimension.