NLP-Measured Political Risk and Volatility*

Chris Li, Derrick Wang, Shitong Xiao, Colbert Yu

December 2025

Abstract

We study how to measure firm-level political risk using quarterly earnings call transcripts and its predictability to stock market volatility. Using an S&P 500 sample from 2008 to 2023, we replicate the dictionary approach proposed by Hassan et al. (2019) and confirm their finding of a significantly positive correlation between political risk measure and both realized and implied volatility. We then propose a new embedding-based measure that embeds policy-relevant sentences and constructs the risk score from each sentence's cosine similarity to predefined uncertainty prototypes net of its similarity to certainty/irrelevance prototypes. We validate our proposed measure by showing its alignment with major political episodes and plausible sector exposure patterns. We further show that our measure exhibits a stronger contemporaneous association with both realized and implied volatility than the dictionary-based measure. Finally, we find that adding our political risk measure to an existing volatility forecasting model reduces out-of-sample forecast loss, while the inclusion of the dictionary-based measure worsens forecast accuracy.

*We owe special thanks to Professor David Thesmar for his continuing guidance and support throughout the project, and to the seminar participants of 15.S06 AI and Machine Learning Research in Finance for their helpful feedback.

1 Introduction

A large literature on investment under uncertainty and political economy emphasizes that policy and political risk are first-order determinants of firms' real decisions including corporate investment and employment (Julio and Yook, 2012; Gulen and Ion, 2016; Jens, 2017; Hassan et al., 2019). However, measuring *firm-level* political risk remains challenging due to the lack of relevant and high-quality data. Recent advances in computational linguistics and textual analysis enable a new measurement of political risk by analyzing firms' discussions of political concerns in narrative disclosures and managerial communications. Yet such discussions often contain contextual and forward-looking language where the same words can have different implications for political uncertainty depending on context. Therefore, text-based measurement requires interpreting *how* political issues are discussed in addition to *whether* political terms appear.

Hassan et al. (2019) pioneer a text-based measure of firm-level political risk by identifying political content in earnings call transcripts via a political bigram dictionary and then counts nearby "risk/uncertainty" synonyms. Building on this work, we revisit the text-based measurement of firm-level political risk with two goals: (i) improve the dictionary-based political risk measures, and (ii) evaluate whether a cleaner and more semantic measure helps forecast stock return volatility.

In this paper, we start by replicating the construction of firm-level political risk in Hassan et al. (2019). Using an S&P 500 sample over 2008 to 2023, we confirm the documented positive association between firm-level political risk and both realized and implied volatility. We also find that this construction of political risk measure is exposed to variation tied to call format rather than political-related issues. We document that a key source of such distortion is the inclusion of generic Q&A language (e.g., "question" and "questions") in the risk synonym set, which can inflate measured "risk" during calls with longer Q&A session even when the surrounding discussion contains little genuine political uncertainty. To address this issue, we construct a revised dictionary measure that removes Q&A filler terms while keeping the remaining architecture unchanged. We show that this revised measure exhibits sector exposures and macro-level correlations that align more closely with economic intuition.

Next, we propose a new embedding-based measure of firm-level political risk that shifts the unit of analysis from token-level to sentence-level. The pipeline first filters transcripts to policy-related sentences using political bigrams. It then embeds each retained sentence and scores it by comparing its semantic similarity to manually curated prototype sets that represent (i) policy uncertainty, (ii) policy certainty/resolution, and (iii) irrelevance. A sentence contributes positively to political risk only when it is closer (in embedding space) to uncertainty prototypes than to certainty or irrelevant prototypes, which helps separate genuinely uncertain policy discussion from reassurance, resolution, or irrelevance. This design directly targets two limitations of dictionary-proximity counting: its difficulty distinguishing "uncertainty increased" from "uncertainty resolved," and its inability to capture paraphrases and richer linguistic constructions that express policy risk without exact dictionary tokens.

To provide validation for our embedding-based measure, we first show that our measure behaves intuitively at both macro and cross-sectional levels, including alignment with major political episodes and plausible sector exposure patterns. We then examine its relation to risk in financial markets by showing that our new measure of firm-level political risk are significantly correlated with both realized and implied volatility. We also document that, in comparable specifications, the embedding-based measure's estimated coefficients are about three times as large as those for the reference dictionary measure.

Finally, we evaluate whether including firm-level political risk as an additional predictor improves stock market volatility forecasting. We augment a benchmark heterogeneous autoregressive realized volatility (HAR-RV) model with firm-level political risk measure. We find that incorporating our embedding-based measure reduces out-of-sample forecast loss relative to the benchmark, while adding the measure from Hassan et al. (2019) can worsen accuracy. The improvements of our measure are not statistically significant in our implementation, but the results are consistent with the view that our approach improves on the dictionary-based measure and captures volatility-related information in corporate policy discussion not fully captured by lagged realized volatility alone.

The remainder of the paper is organized as follows. Section 2 reviews related literature. Section 3 describes the data. Section 4 replicates the dictionary measure in Hassan et al. (2019) and introduces our revised dictionary version. Section 5 details the embedding-based political risk measure. Section 6 presents validation for our proposed measure. Section 7 reports volatility forecasting results. Section 8 concludes.

2 Literature Review

This paper draws on three related strands of research that motivate the construction and evaluation of a firm-level political risk measure. The first strand focuses on aggregate policy uncertainty. The Economic Policy Uncertainty (EPU) index developed by Baker, Bloom, and Davis (2016) highlights the macroeconomic relevance of political shocks and provides a benchmark for measuring policy-related risk. The second strand shifts attention to the firm level. Using earnings conference calls, Hassan et al. (2019) show that text-based indicators capture substantial cross-sectional and time-series variation in firms' exposure to political developments. The third strand examines the link between political uncertainty and market volatility. Liu and Zhang (2015) find that EPU helps explain movements in both realized and implied volatility and improves forecasts of future fluctuations. Taken together, these studies motivate our effort to construct an improved firm-level political risk measure and to study its relationship with realized and implied volatility.

2.1 Aggregate Policy Uncertainty and the Economic Policy Uncertainty Index

Political and policy uncertainty is widely viewed as an important source of macroeconomic fluctuations, but systematic measurement over long horizons was limited until recently. Baker, Bloom, and Davis (2016) address this problem by introducing the EPU index, a high-frequency and historically consistent measure of uncertainty related to government actions, regulation, and macroeconomic policy. In their empirical analysis, Baker, Bloom, and Davis (2016) show that higher EPU is associated with weaker economic performance. Elevated uncertainty predicts declines in industrial production, investment, and GDP growth. The authors interpret these effects through a real-options lens: when policy is uncertain, firms delay irreversible investment. Responses also vary across sectors, with stronger effects in industries more exposed to regulation or government spending. In financial markets, increases in EPU coincide with higher stock-market volatility, larger risk premia, and elevated implied volatility. The index has then become a standard empirical tool in economics and finance (e.g., Gulen and Ion, 2016; Nguyen and Phan, 2017; Bonaime et al., 2018; Leippold and Matthys, 2022; Lin et al., 2025).

Despite its broad influence, the EPU index has important limitations. As an aggregate measure, it cannot capture cross-sectional differences in firms' exposure to political risk. Its reliance on dictionary-based media counts limits its ability to distinguish changes in sentiment or meaning from shifts in coverage intensity. Moreover, the focus on national policy obscures firm-level shocks arising from industry-specific regulation, geographic exposure, or international political developments.

These limitations motivate the development of firm-level, text-based measures of political risk. A leading contribution is Hassan et al. (2019), who construct interpretable political risk indicators from earnings conference calls. Their approach provides granular variation across firms, industries, topics, and time, addressing key shortcomings of aggregate uncertainty indices. We discuss this methodology in the next subsection.

2.2 Firm-Level Political Risk from Earnings Calls

Aggregate indices such as the EPU capture economy-wide movements in policy uncertainty but are not designed to measure political risk at the firm level. Hassan et al. (2019) fill this gap by constructing a firm-level political risk measure from the textual content of quarterly earnings conference calls. Their approach rests on the observation that managers and analysts routinely discuss political and regulatory developments relevant to firm performance, making these calls a natural source of information on perceived political risk.

The measure is defined as the fraction of an earnings call devoted to political risk. To identify relevant passages, the authors apply simple tools from computational linguistics to classify sentences with political content and to isolate statements that explicitly describe

political risks. While the paper does not provide full procedural detail, it is clear that the method systematically separates political from non-political discussion and quantifies the share of each call devoted to risk-related political issues.

A series of validation exercises shows that the measure captures intuitive and economically meaningful variation. The index reliably identifies calls with extensive political discussion, and its time-series and cross-sectional patterns align with major political events. Moreover, most of the variation arises at the firm level rather than at the industry or aggregate level, indicating substantial heterogeneity in firms' political exposure.

The measure is also linked to observable firm behavior. Firms that devote more attention to political risk in their earnings calls subsequently reduce hiring and investment, while increasing lobbying activity and political donations. These responses suggest that the text-based measure captures uncertainty that is relevant for both operational and political decision-making.

Financial-market responses further support the measure's relevance. Firms with higher political risk exposure experience greater stock return volatility. This result parallels evidence from aggregate uncertainty measures but shows that political risk also operates at the firm level, with pronounced cross-sectional variation.

Overall, Hassan et al. (2019) demonstrate that earnings calls contain rich information about firm-level political risk. Systematic analysis of this textual content yields a high-resolution measure that complements aggregate indices such as the EPU. Their framework establishes political risk as an important driver of corporate behavior and market outcomes and provides a foundation for later text-based approaches, including embedding and large-language-model methods. This firm-level political risk measure is widely used in recent studies to examine how firms respond to political uncertainty and how markets price it (e.g., Gad et al., 2024; Grotteria, 2024; Ho et al., 2024).

2.3 Political Uncertainty and Stock Market Volatility

While firm-level text measures quantify political uncertainty, a related question is how such uncertainty translates into financial market outcomes, particularly stock market volatility. Using the EPU index developed by Baker, Bloom, and Davis (2016), Liu and Zhang (2015) examine whether policy uncertainty explains and predicts equity-market volatility. Their analysis considers both the contemporaneous relationship between EPU and realized volatility and the out-of-sample forecasting value of EPU, a dimension largely absent from earlier studies.

The authors construct daily realized volatility for the S&P 500 using 5-minute intraday returns over the 1996–2013 period. To assess the role of policy uncertainty, they augment eight standard HAR-type volatility models with lagged EPU, introducing uncertainty as a parsimonious predictor in volatility dynamics.

Across all specifications, lagged EPU enters with a positive and statistically significant coefficient, indicating that higher policy uncertainty predicts higher subsequent volatility. Estimated marginal effects suggest that a one–standard-deviation increase in EPU raises realized volatility by roughly 3–4 percent, depending on the model.

Forecasting results reinforce this finding. Including EPU improves one-day-ahead volatility forecasts in every model considered, with statistically significant gains in predictive accuracy. At longer horizons, forecast improvements are smaller but remain positive.

Liu and Zhang (2015) interpret these results through the general equilibrium framework of Pastor and Veronesi (2012), which predicts higher stochastic discount factor volatility following policy changes with uncertain economic effects. Increased discount-rate volatility then translates into greater asset-price volatility, providing an economic mechanism for the observed predictability.

Overall, the evidence shows that aggregate-level political uncertainty, as captured by EPU, has meaningful explanatory and predictive power for stock market volatility. More recent studies further show that firm-level political risk predict downside equity risk (e.g., Aye et al., 2018; Makrychoriti and Pyrgiotakis, 2024). These results complement firm-level measures of political exposure by highlighting a channel through which political uncertainty shapes financial risk.

3 Data

We collect 25,709 quarterly earnings call transcripts for S&P 500 firms from 2008 to 2023 from Standard and Poors' Capital IQ. To replicate the political risk measure in Hassan et al. (2019), we obtain the political bigrams, non-political bigrams, and risk-synonyms lists from the paper's replication repository[1]. We also download their political risk scores from the Firm-level Risk website[2].

In addition, we obtain the daily EPU index from the Federal Reserve Economic Data (FRED) database. We obtain firm-quarter implied and realized volatility from OptionMetrics, and intraday volatility from the NYSE Trade and Quote (TAQ) database. We obtain firm-quarter total assets and industry classification codes from Standard and Poors' Compustat. Table 1 provides summary statistics.

[1] See https://github.com/mschwedeler/firmlevelrisk.
[2] See https://www.firmlevelrisk.com/download.

4 Replication

4.1 Construction of the Hassan et al. (2019) Political Risk Measure

We begin our empirical analysis by replicating the firm-level political risk index of Hassan et al. (2019), which we denote by $PRisk_{\text{ref}}$. This subsection provides a detailed description of the original construction and documents how we implement each component in our replication.

Bigrams and Preprocessing

For each earnings conference call transcript, we follow Hassan et al. (2019) and preprocess the raw text by lowercasing, removing non-informative punctuation, and splitting the text into tokens. Overlapping bigrams are then constructed from the token sequence. If a transcript contains B_{it} bigrams, we index them by $b = 1, \ldots, B_{it}$.

Political and Non-Political Bigrams

The political-content classifier in Hassan et al. (2019) is trained using two text libraries: (i) a political library consisting of an undergraduate American politics textbook and articles from the politics sections of major newspapers, and (ii) a non-political library consisting of an accounting textbook and articles from the non-political sections of the same newspapers.

For each bigram b, the authors compute how frequently it appears in the political library relative to the non-political library. Bigrams that occur far more often in political texts are labeled *political*, while those that occur more frequently in non-political texts are labeled *non-political*. Let P denote the set of political bigrams and N denote the set of non-political bigrams; by construction these sets are disjoint.

We replicate this step using the list of exclusively political and non-political bigrams provided by the authors.

Risk Synonyms and the Proximity Criterion

Political content does not necessarily correspond to political risk. Following the original specification, a political bigram contributes to the political risk index only if it appears close to a reference to "risk" or "uncertainty." Let $\mathcal{R}$ denote the dictionary of risk-related synonyms. For each bigram b located at position p_b in the transcript, and each risk word at position r, we compute the word distance $|p_b - r|$. A bigram satisfies the proximity criterion

if it lies within ten words of the nearest risk word:

$$|p_b - r| < 10.$$

Political Frequency Weight

A distinctive feature of the $PRisk_{\text{ref}}$ index is that bigrams belonging to P are weighted using their empirical frequency within the political training library. Let $f_{b,P}$ denote the number of times political bigram b appears in the political library, and let B_P denote the total number of bigrams in that library. The weight assigned to bigram b is $f_{b,P}/B_P$. More frequent political bigrams therefore contribute more to the political risk score.

Call-Level Political Risk Score

Putting these elements together, the call-level political risk measure is:

$$PRisk_{it} = \frac{1}{B_{it}} \sum_{b=1}^{B_{it}} \Big(1[b \in P \setminus N] \times 1[|p_b - r| < 10] \times \frac{f_{b,P}}{B_P} \Big). \tag{1}$$

A bigram contributes to $PRisk_{it}$ if and only if (i) it is classified as political, (ii) it is not classified as non-political, (iii) it occurs within ten words of a risk synonym, and (iv) it is weighted by its political frequency. The denominator B_{it} normalizes by transcript length so that the measure is comparable across firms and across time.

Firm–Quarter Aggregation

Most firms conduct only one earnings call per quarter; when multiple calls occur, we follow the original paper and average the call-level scores to construct a firm–quarter index:

$$PRisk_{\text{ref},i,q} = \frac{1}{|\mathcal{T}_{i,q}|} \sum_{t \in \mathcal{T}_{i,q}} PRisk_{it}. \tag{2}$$

This firm–quarter political risk measure serves as the benchmark for our volatility correlation and forecasting exercises. In Section 6, we contrast this dictionary-based index with our newly proposed embedding-based measure.

4.2 Volatility Correlation

Hassan et al. (2019) reports a significantly positive correlation between their firm-level political risk measure and volatility to support the validity of their risk measure. They estimate the following regression:

$$y_{i,t} = \beta\, PRisk_{i,t} + \gamma X_{i,t} + \delta_t + \delta_i + \epsilon_{i,t}, \tag{3}$$

where y_{it} is either the 90-day realized volatility of firm i in quarter t or 90-day implied volatility derived from at-the-money options. X_{it} is the log of total assets of firm i in quarter t to control for firm size. δ_t and δ_i are time and firm fixed effects, respectively. We replicate their results in Columns (1) and (5) of Table 4 using 91-day realized and implied volatility, restricting the sample to S&P 500 firms and extending the sample period to 2023.

The results are reported in Table 2. Consistent with Hassan et al. (2019), we find that the correlation between realized or implied volatility and political risk is positive and statistically significant. For example, as shown in Columns 1 where we only control for firm size, the correlation coefficient on PRisk is 0.046 when the dependent variable is realized volatility and 0.045 when it is implied volatility. The corresponding coefficients in Hassan et al. (2019) are 0.048 and 0.056, respectively. When we add additional controls for time and firm fixed effects in Column 2, our estimated coefficients on PRisk are 0.017 for realized volatility and 0.015 for implied volatility, compared to 0.014 and 0.013 in Hassan et al. (2019). Therefore, both the sign and the magnitude of our replicated correlation between PRisk and volatility closely align with Hassan et al. (2019). This suggests that the positive association between firm-level political risk and volatility is robust to restricting the sample to S&P 500 firms and to an extended time period.

4.3 Revised Hassan Political Risk Score

While the original political risk measure of Hassan et al. (2019) provides the first firm-level text-based index of political risk, its construction relies critically on a dictionary of "risk" and "uncertainty" synonyms. A careful examination of the dictionary reveals a robustness issue that systematically biases the resulting score. In particular, the Hassan risk dictionary includes the words *"question"* and *"questions"*. These terms occur frequently in the analyst Q&A portion of earnings calls, where they typically appear in purely mechanical phrases such as:

- "Thank you for the question,"
- "My question is about ...",
- "We have time for one more question."

These Q&A filler expressions are unrelated to political developments or policy-driven uncertainty. Yet, under the original scoring rule, any political bigram that appears near these Q&A statements meets the proximity criterion for political risk. This produces three systematic distortions:

1. **Mechanical correlation with call structure.** Firms that host longer Q&A sessions accumulate more instances of the word "question", mechanically inflating their $PRisk_{\text{ref}}$ scores.
2. **Inflated scores for high-coverage firms.** Firms with active analyst coverage appear riskier, even when the surrounding dialogue contains no true political uncertainty.
3. **Reduced cross-sectional interpretability.** Variation in $PRisk_{\text{ref}}$ reflects differences in call format rather than differences in political exposure.

To address this issue, we construct a revised version of the Hassan score, denoted $PRisk_{\text{revised}}$, by removing "question", "questions", and other Q&A filler terms from the risk dictionary before computing the proximity indicator $1[|p_b - r| < 10]$. All other components of the original measure, including political bigram filtering, exclusion of non-political bigrams, and political frequency weights $f_{b,P}/B_P$, are kept unchanged.

Formally, if $\mathcal{R}$ is the original Hassan risk dictionary and $\mathcal{Q}$ is the set of Q&A filler terms such as {question, questions}, our revised dictionary is

$$\mathcal{R}^{\text{rev}} = \mathcal{R} \setminus \mathcal{Q}.$$

The revised call-level score is then

$$PRisk_{revised,it} = \frac{1}{B_{it}} \sum_{b=1}^{B_{it}} \Big(1[b \in P \setminus N] \times 1[|p_b - r^{\text{rev}}| < 10] \times \frac{f_{b,P}}{B_P} \Big), \tag{4}$$

where r^{rev} denotes the nearest occurrence of a risk synonym in the revised dictionary $\mathcal{R}^{\text{rev}}$. The firm–quarter aggregation remains identical to the replication of the baseline Hassan measure.

This revised score eliminates the structural bias caused by the Q&A portion of earnings calls and produces a cleaner measure of political uncertainty that is not mechanically driven by call length or analyst engagement. In Section 6, we show that the revised index yields sector exposures and macro-level correlations that better align with economic intuition than the original Hassan score.

5 Embedding-based Political Risk Scoring Method

This section describes in detail how we construct a firm–quarter political risk score from earnings call transcripts. The method combines a dictionary-based filter for policy-related content with sentence-level semantic embeddings and a prototype-based similarity rule. The aim is to obtain a measure that (i) focuses explicitly on political and policy discussion, (ii) distinguishes sentences that raise uncertainty from those that reduce or neutralize it, and (iii) is comparable across firms and quarters despite large differences in call length and communication style.

5.1 Overview of the Pipeline

Let $\mathcal{T}$ denote the set of earnings call transcripts in our sample. A transcript $t \in \mathcal{T}$ is represented as an ordered sequence of sentences

$$t = (s_1, \ldots, s_{n_t}),$$

where n_t is the number of sentences. For each transcript we construct a scalar political risk score $S(t)$ through five steps.

First, we identify policy-related sentences using a political bigram dictionary. Only sentences that contain at least one dictionary bigram are retained for subsequent analysis. Second, we define three sets of *prototype sentences* that represent (a) policy uncertainty, (b) policy certainty, and (c) irrelevant sentences. These prototypes are manually curated examples of language that clearly belongs to each semantic category.

Third, we map all retained sentences and all prototypes into a common d_θ–dimensional vector space via a representation function ϕ_θ. We consider both a sparse TF–IDF representation and several dense neural sentence embedding models; all downstream operations are defined abstractly in terms of ϕ_θ and are therefore model-agnostic.

Fourth, for each policy-related sentence we compute cosine similarity between its embedding and the average embedding (centroid) of each prototype set. A sentence contributes positively to the political risk score only if it is more similar to uncertainty prototypes than to certainty and irrelevance prototypes.

Finally, we aggregate the sentence-level contributions within the transcript and normalize by the total number of word tokens. The resulting score $S(t)$ can be interpreted as the expected amount of political uncertainty language per thousand words of transcript. If a firm has multiple calls in the same fiscal quarter, we average $S(t)$ across those calls to obtain the firm–quarter political risk score used in the empirical analysis.

5.2 Policy-Sentence Filter

We start from raw transcripts and perform light preprocessing. Each transcript is segmented into sentences by a sentence splitter. Within each sentence we apply lowercasing, remove non-informative punctuation and vendor markup, and standardize whitespace. We intentionally avoid aggressive text normalization: in particular, we do not remove negations or lemmatize words, since these operations can alter the meaning of policy statements.

To focus on political content we rely on a dictionary of political bigrams $B = \{b_1, \ldots, b_{|B|}\}$. A bigram $b \in B$ is an ordered pair of tokens (w_1, w_2) associated with political institutions, policy instruments, regulatory authorities, or geopolitical entities. The dictionary is adapted from the political phrase list of Hassan et al. (2019), with minor cleaning to remove entries that are either extremely rare or clearly unrelated to policy.

For a sentence s let $b \subset s$ denote that bigram b appears in s after preprocessing. We define the indicator that s is policy-related as

$$I_{\text{pol}}(s) = \begin{cases} 1, & \text{if there exists } b \in B \text{ such that } b \subset s, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

Given transcript $t = (s_1, \ldots, s_{n_t})$, the corresponding set of policy-related sentences is

$$P(t) = \big\{ s_i \in t : I_{\text{pol}}(s_i) = 1 \big\}. \tag{6}$$

All subsequent steps of the scoring procedure operate only on sentences in $P(t)$. Non-policy sentences are ignored, except that they enter the total word count used for normalization.

If $P(t)$ is empty, meaning that transcript t does not contain any political bigram, we set its political risk score to zero:

$$S(t) = 0 \quad \text{whenever } P(t) = \varnothing. \tag{7}$$

In this sense our measure preserves the core design of dictionary-based measures: only transcripts that explicitly touch on political topics can receive a non-zero political risk score.

5.3 Prototype Sentences

A distinctive feature of our approach is the use of prototype sentences to anchor the semantic interpretation of political uncertainty. Instead of treating all policy-related sentences as equally informative, we compare each sentence to curated examples of different types of policy talk.

We define three finite sets of prototype sentences:

$$U^+ = \{u_1^+, \ldots, u_{K_+}^+\}, \quad U^- = \{u_1^-, \ldots, u_{K_-}^-\}, \quad I = \{i_1, \ldots, i_{K_I}\}. \tag{8}$$

The set U^+ contains sentences that clearly describe uncertainty about current or future government actions, regulatory outcomes, legislation, or geopolitical developments, in a way that is interpretable as risk for the firm. Typical elements mention difficulties in forecasting policy, delays in rule-making, or ambiguity in implementation.

The set U^- contains sentences that provide reassurance about the policy environment. Examples state that regulation has been finalized and is well understood, that the firm does not expect material regulatory changes, or that known policy changes have already been incorporated into business plans. Such sentences often appear in response to analyst questions about risk and therefore play an important role in distinguishing uncertainty from certainty.

The set I consists of sentences that are superficially similar to earnings call language but are essentially irrelevant. These include greetings, instructions for the Q&A session, standard legal disclaimers, and generic expressions of gratitude. Some of these sentences may contain tokens that are weakly related to policy, such as references to "forward-looking statements", but they do not describe any specific political or regulatory situation.

The prototypes are selected by reading transcripts around major political events, identifying sentences that clearly fall into one of the three categories, and then cleaning and deduplicating the candidate sets. The number of prototypes in each set is modest (on the order of tens), which keeps them interpretable and makes it easy to inspect the embedded centroids defined below.

5.4 Sentence Representations

We now describe how sentences are mapped into vectors. Let

$$\mathcal{C} = \left(\bigcup_{t \in \mathcal{T}} P(t) \right) \cup U^+ \cup U^- \cup I$$

denote the collection of all policy-related sentences and all prototypes. A representation model indexed by θ is a function

$$\phi_\theta : \mathcal{C} \to \mathbb{R}^{d_\theta}, \tag{9}$$

which maps each sentence in $\mathcal{C}$ to a real-valued vector of dimension d_θ. The dimension d_θ and the geometry of the representation depend on the chosen model class.

Sparse TF–IDF representation

The first representation is a standard TF–IDF bag-of-ngrams. We fix a vocabulary $\mathcal{V}$ of unigrams and bigrams constructed from $\mathcal{C}$. For each sentence s we define $\phi_{\mathrm{TFIDF}}(s) \in \mathbb{R}^{|\mathcal{V}|}$ with components

$$\big[\phi_{\mathrm{TFIDF}}(s)\big]_j = \mathrm{tf}_j(s)\,\mathrm{idf}_j, \quad j = 1, \ldots, |\mathcal{V}|,$$

where $\mathrm{tf}_j(s)$ is the term frequency of token j in sentence s, and idf_j is the inverse document frequency of token j over all sentences in $\mathcal{C}$. The resulting vector is high-dimensional and sparse: only a small number of coordinates are non-zero for any given sentence. This representation captures lexical overlap but does not directly account for word order or deeper semantics.

Dense neural sentence embeddings

To incorporate semantic information that is not captured by shared tokens, we also use off-the-shelf neural sentence encoders. These models take a sentence as input and output a dense embedding of moderate dimension. Sentences that are semantically similar are placed close to each other in this embedding space even when they do not share many words.

In practice we experiment with several encoders, such as a compact MiniLM-based model, a medium-sized MPNet-based model, and a T5-based encoder, with embedding dimension d_θ ranging from a few hundred to around one thousand. All encoders are used without fine-tuning. For a given choice of encoder θ we write $\phi_\theta(s) \in \mathbb{R}^{d_\theta}$ for the embedding of sentence s. These embeddings are dense (most coordinates are non-zero) and have no straightforward coordinate-wise interpretation; their meaning is encoded in the relative positions and angles between vectors.

All subsequent steps of the method are expressed in terms of ϕ_θ and do not depend on whether the representation is sparse or dense. This allows us to compare different models on equal footing.

5.5 Prototype Centroids and Similarities

Once all sentences in $\mathcal{C}$ have been embedded, we summarize each prototype set by its centroid. For a fixed representation ϕ_θ we define

$$\mu_+ = \frac{1}{K_+} \sum_{u \in U^+} \phi_\theta(u), \qquad \mu_- = \frac{1}{K_-} \sum_{u \in U^-} \phi_\theta(u), \qquad \mu_I = \frac{1}{K_I} \sum_{u \in I} \phi_\theta(u). \tag{10}$$

The vector μ_+ is the average embedding of uncertainty prototypes, μ_- is the average embedding of certainty prototypes, and μ_I is the average embedding of irrelevant prototypes. Each centroid can be viewed as the "typical" vector for its semantic class.

To compare two vectors $x, y \in \mathbb{R}^{d_\theta}$ we use cosine similarity,

$$\text{sim}(x, y) = \frac{\langle x, y \rangle}{\|x\|_2 \, \|y\|_2}, \tag{11}$$

where $\langle x, y \rangle$ is the Euclidean inner product and $\| \cdot \|_2$ is the ℓ_2 norm. Cosine similarity takes values in $[-1, 1]$ and is invariant to rescaling of either argument. This is useful because the overall magnitude of embedding vectors is not meaningful; only their directions matter.

For a policy-related sentence $s_i \in P(t)$ we write $v_i = \phi_\theta(s_i)$. Its similarity to each prototype centroid is then given by

$$u_i^+ = \text{sim}(v_i, \mu_+), \quad u_i^- = \text{sim}(v_i, \mu_-), \quad u_i^I = \text{sim}(v_i, \mu_I), \tag{12}$$

which produces a three-dimensional similarity profile for each sentence. A sentence that closely resembles uncertainty prototypes will have u_i^+ relatively large; a sentence resembling certainty or irrelevance will have u_i^- or u_i^I larger instead.

5.6 Sentence-Level Political Risk Scores

We now transform the similarity profile in (9) into a scalar contribution to political risk. The key idea is that a sentence should only increase the risk score if it is closer to the uncertainty centroid than to both alternatives.

For a policy-related sentence $s_i \in P(t)$ we define its sentence-level political risk contribution σ_i as

$$\sigma_i = \max\Big\{0, \; u_i^+ - \max\{u_i^-, u_i^I\}\Big\}. \tag{13}$$

The term in parentheses compares the sentence's similarity to the uncertainty centroid with its best match among certainty and irrelevance. If s_i is at least as close to certainty or irrelevance as it is to uncertainty, then the inner difference is non-positive and the outer maximum sets σ_i to zero. In this case the sentence does not contribute to political risk.

If instead s_i is closer to uncertainty prototypes than to both alternatives, the difference $u_i^+ - \max\{u_i^-, u_i^I\}$ is positive and σ_i equals that difference. The magnitude of σ_i is then proportional to the degree by which the sentence leans towards uncertainty. Because cosine similarity is bounded, σ_i is also bounded and typically takes moderate values.

Several features of the definition in 13 are worth emphasizing. First, the use of the maximum with zero ensures that political risk is non-negative at the sentence level: we do not interpret reassuring sentences as generating *negative* risk, but rather as sentences that do not add to the stock of uncertainty. Second, we compare uncertainty to the strongest competing semantic class, which gives conservative scores. Third, the measure is continuous in the embedding vectors and stable under small changes in the representation.

5.7 Transcript-Level Aggregation and Normalization

To obtain a transcript-level political risk score we aggregate the sentence-level contributions across all policy-related sentences in $P(t)$. We define the unnormalized political uncertainty mass as

$$\Sigma(t) = \sum_{s_i \in P(t)} \sigma_i. \tag{14}$$

By construction, $\Sigma(t) \geq 0$ and equals zero when $P(t)$ is empty. A transcript with many sentences that strongly resemble uncertainty prototypes will have a large value of $\Sigma(t)$.

However, transcripts differ substantially in length. Some calls last for more than an hour and contain many pages of text, while others are shorter. Using $\Sigma(t)$ directly would mechanically assign higher scores to longer calls even if the density of uncertainty language is the same. To address this we normalize by the total number of word tokens in transcript t, denoted by $W(t)$.

The final transcript-level score is defined as

$$S(t) = \frac{1000}{W(t)} \sum_{s_i \in P(t)} \sigma_i = \frac{1000}{W(t)} \Sigma(t). \tag{15}$$

The factor $1/W(t)$ adjusts for transcript length, and the factor 1000 is a scaling constant that makes typical values of $S(t)$ numerically convenient (on the order of one). Thus $S(t)$ can be interpreted as the expected sentence-level contribution to political uncertainty per thousand words of transcript.

If a firm f has several calls in quarter q, indexed by $t \in \mathcal{T}_{f,q}$, we define the firm–quarter score as the simple average

$$S_{f,q} = \frac{1}{|\mathcal{T}_{f,q}|} \sum_{t \in \mathcal{T}_{f,q}} S(t). \tag{16}$$

These $S_{f,q}$ constitute the main political risk series used in the empirical analysis.

5.8 Implementation Choices and Robustness

The method above is defined at a high level and admits multiple implementation choices. In this subsection we summarize the main choices and discuss how they affect the resulting scores.

Choice of representation model. The function ϕ_θ can be instantiated using different representations. In our baseline specification we use an MPNet-based sentence encoder with $d_\theta = 768$. This choice provides a reasonable balance between computational cost and semantic richness. We also implement the same pipeline with a smaller MiniLM encoder, a larger T5-

based encoder, and with the sparse TF–IDF representation. In each case the only change is the mapping from sentences to vectors; the definition of σ_i and $S(t)$ is identical. This permits a clean comparison of the effect of representation choice on the behavior of the score.

Prototype specification. The prototype sets U^+, U^-, I play a central role in anchoring the semantics of uncertainty and certainty. To construct them we begin from a broad pool of candidate sentences, drawn from transcripts with high dictionary-based political risk scores and from periods of known political stress. We then hand-label each candidate as uncertainty, certainty, or irrelevance, discard ambiguous cases, and remove near duplicates. The resulting sets are small enough to be manually reviewed and stable across embedding models. In robustness checks we consider minor variations of the prototype sets, such as adding a small number of sector-specific prototypes, and obtain similar aggregate patterns.

Alternative sentence scoring rules. Expression 13 compares u_i^+ with the maximum of u_i^- and u_i^I. One could instead compare u_i^+ to a convex combination $\alpha u_i^- + (1-\alpha)u_i^I$ for some $\alpha \in [0,1]$, or introduce an explicit threshold τ and define σ_i as $\max\{0, u_i^+ - \max\{u_i^-, u_i^I\} - \tau\}$. We explored these variants and found that they mainly rescale or slightly truncate the distribution of $S(t)$ but do not change the relative ranking of transcripts or the time-series behavior at the aggregate level. This suggests that the main driver of the score is the frequency of sentences that clearly resemble uncertainty prototypes rather than the exact functional form of the sentence-level transformation.

5.9 Relation to Dictionary-Based Measures

Empirically, we find that the transcript-level series $S(t)$ is positively correlated with the dictionary-based score but exhibits several desirable features. At the aggregate level the time-series pattern of our measure continues to align with major political events, such as elections or episodes of policy gridlock, but the firm-level series are somewhat smoother and less sensitive to generic question phrases. This supports the interpretation of $S(t)$ as a refined, semantically informed measure of firm-specific political risk.

6 Validation

We assess the validity of our political risk measures using three complementary approaches: (i) macro-level validation by comparing quarterly aggregates of our PRisk measures to the EPU index, (ii) industry-level validation using SIC2 sectors, and (iii) firm-level validation by examining how our PRisk measure correlates with volatility. These exercises evaluate whether the measures we construct capture intuitive patterns in political and regulatory

exposure at both the aggregate and cross-sectional levels and are related to financial market risk.

6.1 Macro-Level Comparison with EPU

To examine whether firm-level political risk moves with macroeconomic policy uncertainty, we compare quarterly aggregates of our PRisk measures: the original Hassan score ($PRisk_{\text{ref}}$), the revised dictionary score ($PRisk_{\text{revised}}$), and our embedding-based score ($PRisk_{\text{emb}}$) with the national EPU index of Baker, Bloom, and Davis (2016).

Step 1: Convert Daily EPU to Quarterly Frequency

Quarterly EPU is constructed as the simple average of daily EPU values within quarter q:

$$EPU_q = \frac{1}{N_q} \sum_{d \in q} EPU_d, \tag{17}$$

where N_q is the number of trading days in quarter q.

Step 2: Aggregate Firm-Level PRisk to Quarterly Level

For each PRisk measure $k \in \{\text{ref}, \text{revised}, \text{emb}\}$, quarterly PRisk is computed as the cross-sectional average:

$$\overline{PRisk}_q^k = \frac{1}{N_q^{\text{firm}}} \sum_{i \in q} PRisk_{iq}^k, \tag{18}$$

where N_q^{firm} is the number of firms with calls in quarter q.

The macro-level time-series plots in Figure 1 and Figure 2 reveal two key findings. First, the revised dictionary score produces smoother and more interpretable macro dynamics. Removing Q&A filler terms (e.g. "question/questions") eliminates a source of mechanical variation and results in a series that more clearly tracks macroeconomic uncertainty. Second, the embedding-based PRisk measure displays the strongest macro-level alignment. It captures major political shocks while suppressing noise from irrelevant language, producing time-series patterns that closely aligned with future EPU.

We also examine whether firm-level PRisk contains information about future macroeconomic uncertainty. We show the correlation between quarterly PRisk and next-quarter EPU in Table 3. The embedding-based measure achieves the highest predictive correlation, suggesting that it captures forward-looking components of political risk at macro level.

6.2 Industry-Level Validation (SIC2)

To evaluate whether our PRisk measures reflect industry differences in political sensitivity, we examine variation across SIC2 industries.

Step 1: SIC2 Aggregation

Each firm has a four-digit SIC code. We collapse it to the two-digit industry level:

$$SIC2_i = \left\lfloor \frac{SIC_i}{100} \right\rfloor . \tag{19}$$

SIC2 captures broad sector groupings (e.g., 62 = brokers, 63 = insurance carriers) and is appropriate for sector-level political risk comparisons.

Step 2: Sector-Level PRisk Aggregation

For each sector j and PRisk measure k, we compute the sector exposure as the cross-sectional median:

$$\widetilde{PRisk}_j^k = \text{median}_{i:SIC2_i=j}(PRisk_i^k). \tag{20}$$

For Hassan PRisk (original), highest exposures occur in insurance, securities brokerage, and financial services, industries heavily regulated and especially sensitive to policy developments. For revised Hassan PRisk, strongest exposures appear in insurance, banking, securities dealers, and utilities. These industries have structural sensitivity to legislation, regulatory changes. For embedding-based PRisk, top-ranked sectors include healthcare, manufacturing, transportation, and energy. This reflects operational uncertainty discussed in earnings calls, rather than policy-specific uncertainty.

While the dictionary-based measures emphasize explicitly regulated industries, the embedding score captures a wider notion of uncertainty expressed in managerial language, consistent with its semantic design.

6.3 Industry-Level Analysis During the Financial Crisis (2008–2009)

To further validate the measures, we examine sector exposure specifically during the 2008–2009 crisis window. This period provides a natural stress test due to credit market disruptions.

For the Hassan score, highest exposure appears in banking, nonbank credit institutions, securities dealers, insurance carriers, and insurance brokers, precisely the sectors at the center of the crisis. For the revised PRisk, the revised measure identifies insurance agents/brokers, insurance carriers, securities dealers, health services, and auto dealers. For the embedding score, top sectors include transportation, business/engineering services, financial holding companies, nonbank credit institutions, and insurance. This is more closely aligned with our expectation.

Across all three measures, sectors with meaningful exposure to the financial system and regulatory interventions display elevated political risk, validating the economic content of the measures.

6.4 Firm-level Volatility Correlation

Finally, we show that our new PRisk measure, $PRisk_{emb}$, is significantly correlated with both realized and implied volatility from the same quarter. We estimate the following specification:

$$RV_{i,t} = \beta\, PRisk_{emb,i,t} + \gamma X_{i,t} + \delta_t + \delta_i + \epsilon_{i,t}, \tag{21}$$

where RV_{it} is the 91-day realized volatility of firm i in quarter t, and we use IV_{it} for 91-day implied volatility. X_{it} is the log of total assets of firm i in quarter t to control for firm size. δ_t and δ_i are time and firm fixed effects, respectively. The results are reported in Table 4.

Panel A shows the results for realized volatility. Column (1) indicates that the correlation between $PRisk_{emb}$ and realized volatility is significantly positive when controlling for firm size only, and a one-standard-deviation increase in $PRisk_{emb}$ is correlated with 0.135-standard-deviation increase in realized volatility. Column (2) shows that the significantly positive correlation still holds after adding both time and firm fixed effects, in addition to the size control. However, the magnitude of such correlation drops by about one-half. In particular, a one-standard-deviation increase in $PRisk_{emb}$ now correlates with 0.072-standard-deviation increase in realized volatility. Panel B shows parallel results for implied volatility.

In Columns (3) and (4) of both panels, we also report the correlation between realized and implied volatility and the original Hassan score ($PRisk_{ref}$) for comparison. The estimated coefficients on $PRisk_{ref}$ are roughly one-third the magnitude of those on $PRisk_{emb}$, suggesting that our embedding-based measure of firm-level political risk explains more variation in realized and implied volatility. This implies that $PRisk_{emb}$ may capture aspects of firm-level political risk that are more relevant for market pricing or firm outcomes. Although it does not by itself prove that $PRisk_{emb}$ is a strictly superior measure of political risk, this comparison suggests that our measure captures additional politically relevant variation that is not fully reflected in $PRisk_{ref}$.

7 Volatility Analysis

We next analyze whether adding firm-level political risk improves volatility predictability.

Volatility forecasts provide crucial implication for market participants in terms of portfolio allocation and risk management. Recent research shows that political risk is an important source of downside equity risk. At the aggregate level, Liu and Zhang (2015) show that adding EPU to the existing volatility prediction models significantly improve forecasting performance for realized market volatility. More recent studies find that cross-sectional measures of firm-level political risk exposure strongly predict "bad" realized volatility of aggregate stock returns (Aye et al., 2018), and firms with higher political risk face a greater likelihood of subsequent stock price crashes (Makrychoriti and Pyrgiotakis, 2024). Against this backdrop, we examine whether a firm-level measure of political risk improves the forecasting of future return volatility beyond the information contained in past realized volatility.

Findings of Andersen and Bollerslev (1998) and Andersen et al. (2003) suggest that realized volatility constructed from high-frequency returns provides a near-unbiased and highly efficient ex-post estimator of daily return volatility under suitable regularity conditions. For stock i on day t, we construct daily realized volatility as the sample variance of intraday log returns:

$$RV_{i,t} = \frac{1}{M_{i,t} - 1} \sum_{k=1}^{M_{i,t}} \left(Ret_{i,t,k} - \overline{Ret}_{i,t} \right)^2, \tag{22}$$

where $Ret_{i,t,k} = \ln(P_{i,t,k}/P_{i,t,k-1})$ is the k-th intraday log return on day t, $P_{i,t,k}$ denotes the transaction price of trade k, $M_{i,t}$ is the number of intraday trades for stock i on day t, and $\overline{Ret}_{i,t}$ is the corresponding intraday average return. This measure provides a nonparametric estimate of the conditional variance of daily returns implied by intraday price variation.

We adopt the HAR-RV model of Corsi (2009) for volatility forecasting. The HAR-RV model captures the distinct persistence of volatility at different horizons by including daily, weekly, and monthly realized volatility components as predictors.

For each stock i, day t, and forecast horizon h, we define the horizon-specific realized volatility target as the simple average of realized volatility from day $t+1$ to $t+h$:

$$RV_{i,h,t} = \frac{1}{h} \sum_{\ell=1}^{h} RV_{i,t+\ell}, \tag{23}$$

We also define the lagged weekly and monthly realized volatility terms as:

$$RV_{i,w,t} = \frac{1}{5}\sum_{\ell=1}^{5} RV_{i,t-\ell}, \tag{24}$$

$$RV_{i,m,t} = \frac{1}{22}\sum_{\ell=1}^{22} RV_{i,t-\ell}, \tag{25}$$

The baseline HAR-RV panel specification is then:

$$RV_{i,h,t} = \beta_0 + \beta_d RV_{i,t} + \beta_w RV_{i,w,t} + \beta_m RV_{i,m,t} + \varepsilon_{i,h,t}, \tag{26}$$

where $\varepsilon_{i,h,t}$ is an error term. We consider two economically relevant horizons: $h = 1$ (one-day-ahead volatility) and $h = 63$ (approximately one trading quarter ahead).

To evaluate the predictive power of firm-level political risk for volatility, we augment the benchmark HAR-RV model with our firm-specific political risk measure. For each firm i and day t, let $PRisk_{i,\leq t}$ denote the value of the political risk index constructed from the most recent quarterly earnings conference call held on or before day t. This timing convention ensures that the political risk variable is determined before the volatility outcomes we seek to forecast. The augmented panel HAR-RV specification is:

$$RV_{i,h,t} = \beta_0 + \beta_d RV_{i,t} + \beta_w RV_{i,w,t} + \beta_m RV_{i,m,t} + \theta\, PRisk_{i,\leq t} + \varepsilon_{i,h,t}, \tag{27}$$

We evaluate the forecasting performance of the baseline and augmented models using recursive rolling windows. The full sample is partitioned into an initial in-sample estimation period and a subsequent out-of-sample evaluation period. Specifically, we use the first 2,000 trading days as the initial in-sample window to estimate the model parameters. We then generate one-step-ahead and 63-step-ahead volatility forecasts for the next day, expand the estimation window by adding the new day and dropping the earliest day, re-estimate the model, and iterate this procedure through the remaining (roughly 2,000) trading days. This rolling-window approach yields a sequence of out-of-sample forecasts for each stock and each horizon, while allowing for gradual evolution in the underlying volatility dynamics.

Following Bollerslev and Ghysels (1996), we assess forecast accuracy using a heteroskedasticity-adjusted mean squared error. Let $h_{i,t}$ denote the volatility forecast for stock i and day t implied by a given model, and let $RV_{i,t}$ be the corresponding realized volatility. The heterogeneous mean squared error (HMSE) is defined as:

$$HMSE = \frac{1}{N}\sum_{i,t}\left(1 - \frac{RV_{i,t}}{h_{i,t}}\right)^2, \tag{28}$$

where the summation runs over all firm–day observations in the out-of-sample period and N is the total number of out-of-sample forecasts. By scaling realized volatility by its forecast,

HMSE penalizes forecast errors in a way that is robust to cross-sectional and time-series variation in the unconditional level of volatility. Lower values of HMSE indicate better forecast performance. We also compare the model differences using Diebold and Mariano (1995) (DM) test as suggested by Liu and Zhang (2015).

Table 5 reports the out-of-sample HMSE for the benchmark HAR-RV model and the HAR-RV models augmented with $PRisk_{\text{ref}}$ and $PRisk_{\text{emb}}$ at the daily ($h = 1$) and quarterly ($h = 63$) horizons. For both horizons, incorporating our new embedding-based political risk measure $PRisk_{\text{emb}}$ leads to lower HMSE relative to the benchmark HAR-RV specification. In comparison, adding the dictionary-based measure $PRisk_{\text{ref}}$ from Hassan et al. (2019) does not improve, and in fact worsens, the forecast accuracy. However, none of the differences in HMSE across models is statistically significant at the 10% level. These results suggest that our embedding-based measure of firm-level political risk may contain information about future volatility beyond that captured by past realized volatility and $PRisk_{\text{ref}}$, but the incremental predictive power is overall limited.

To summarize, our embedding-based political risk measure captures additional politically relevant variation in realized and implied volatility and, in our setting, delivers better realized-volatility forecasts than the existing dictionary-based measure. However, firm-level political risk overall has only limited predictive power for realized volatility.

8 Conclusion

This paper studies how political risk discussed in earnings calls is measured and how such risk relates to equity volatility. We replicate the dictionary-based approach of Hassan et al. (2019) on an S&P 500 sample from 2008 to 2023 and confirm their finding that firms discussing more political risk exhibit higher realized and implied volatility. At the same time, we identify a structural weakness in the original construction. Specifically, the inclusion of Q&A filler terms in the risk dictionary mechanically inflates political risk during analyst-heavy calls. A simple revision that removes these terms yields a cleaner dictionary measure with more interpretable sector patterns and closer alignment with macroeconomic uncertainty.

We then propose a sentence-level, embedding-based measure that focuses on semantic content rather than token proximity. The method filters transcripts to policy-related sentences and scores each sentence by comparing its similarity to curated prototypes representing policy uncertainty, policy certainty, and irrelevance. This design allows the measure to distinguish uncertainty from reassurance and to suppress irrelevant language. Empirically, the embedding-based measure shows a stronger association with both realized and implied volatility than the measure in Hassan et al. (2019), suggesting that semantic information captures aspects of political risk that dictionary-based counts miss.

Finally, we examine whether firm-level political risk improves volatility forecasts beyond the standard HAR-RV model. Adding the embedding-based measure to the forecasting model

reduces out-of-sample forecast loss relative to the benchmark, while the dictionary-based measure does not. The improvements are not statistically significant, but they still provide some evidence that our new measure contain more information about political uncertainty.

More broadly, our results highlight the importance of measurement design in text-based risk indicators. Future work could extend this framework by formalizing prototype construction, allowing language representations to evolve over time, and testing whether semantically grounded political risk measures are informative for other firm outcomes, such as crash risk, investment, or political engagement.

Another important direction for future research is to deepen the use of large language models in measuring political risk in corporate disclosures. While our approach incorporates sentence-level embeddings, it still summarizes meaning through reduced-form similarity scores. More expressive language models could be used to directly assess how firms characterize political developments, including whether policy outcomes remain uncertain, have been resolved, or are contingent on future events. Such classifications would be able to separate the firm-specific political uncertainty from reassurance or compliance language more explicitly than token-based or proximity-based methods.

Future work could also move beyond scalar measures of political risk by extracting structured attributes from earnings-call discussions. Language models may help identify the policy domain, geographic scope, or economic channel referenced in political statements, enabling the construction of multi-dimensional risk measures. Combining these structured outputs with embedding-based representations would preserve semantic richness while improving interpretability.

References

Andersen, Torben G., and Tim Bollerslev, 1998. Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts. *International Economic Review* 39(4), 885–905. 21

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys, 2003. Modeling and Forecasting Realized Volatility. *Econometrica* 71(2), 579–625. 21

Aye, Goodness C., Mehmet Balcilar, Riza Demirer, and Rangan Gupta, 2018. Firm-level Political Risk and Asymmetric Volatility. *The Journal of Economic Asymmetries* 18, e00110. 6, 21

Baker, Scott R., Nicholas Bloom, and Steven J. Davis, 2016. Measuring Economic Policy Uncertainty. *The Quarterly Journal of Economics* 131(4), 1593–1636. 3, 4, 5, 18

Bollerslev, Tim, and Eric Ghysels, 1996. Periodic Autoregressive Conditional Heteroscedasticity. *Journal of Business & Economic Statistics* 14(2), 139–151. 22

Bonaime, Alice A., Huseyin Gulen, and Mihai Ion, 2018. Does policy uncertainty affect mergers and acquisitions? *Journal of Financial Economics* 129(3), 531–558. 4

Corsi, Fulvio, 2009. A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics* 7(2), 174–196. 21

Diebold, Francis X., and Roberto S. Mariano, 1995. Comparing Predictive Accuracy. *Journal of Business & Economic Statistics* 13(3), 253–263. 23

Gad, Mahmoud, Valeri Nikolaev, Ahmed Tahoun, and Laurence van Lent, 2024. Firm-level political risk and credit markets. *Journal of Accounting and Economics* 77(2–3), Article 101642. 5

Grotteria, Marco, 2024. Follow the Money. *The Review of Economic Studies* 91(2), 1122–1161. 5

Gulen, Huseyin, and Mihai Ion, 2016. Policy Uncertainty and Corporate Investment. *The Review of Financial Studies* 29(3), 523–564. 2, 4

Hassan, Tarek A., Stephan Hollander, Laurence van Lent, and Ahmed Tahoun, 2019. Firm-Level Political Risk: Measurement and Effects. *The Quarterly Journal of Economics* 134(4), 2135–2202. 1, 2, 3, 4, 5, 6, 7, 9, 12, 23, 30, 32, 33

Ho, Thang, Anastasios Kagkadis, and George Wang, 2024. Is Firm-Level Political Risk Priced in the Equity Option Market? *The Review of Asset Pricing Studies* 14(1), 153–195. 5

Jens, Candace E., 2017. Political Uncertainty and Investment: Causal Evidence from U.S. Gubernatorial Elections. *Journal of Financial Economics* 124(3), 563–579. 2

Julio, Brandon, and Youngsuk Yook, 2012. Political Uncertainty and Corporate Investment Cycles. *The Journal of Finance* 67(1), 45–83. 2

Leippold, Markus, and Felix Matthys, 2022. Economic Policy Uncertainty and the Yield Curve. *Review of Finance* 26(4), 751–797. 4

Lin, Leming, Atanas Mihov, and Leandro Sanz, 2025. Economic policy uncertainty and multinational companies. *Review of Finance* 29(6), 1769–1807. 4

Liu, Li and Tao Zhang, 2015. Economic Policy Uncertainty and Stock Market Volatility. *Finance Research Letters* 15, 99–105. 3, 5, 6, 21, 23

Makrychoriti, Panagiota, and Emmanouil G. Pyrgiotakis, 2024. Firm-level Political Risk and Stock Price Crashes. *Journal of Financial Stability* 74, 101303. 6, 21

Nguyen, Nam H., and Hieu V. Phan, 2017. Policy Uncertainty and Mergers and Acquisitions. *Journal of Financial and Quantitative Analysis* 52(2), 613–644. 4

Pastor, Lubos, and Pietro Veronesi, 2012. Uncertainty about Government Policy and Stock Prices. *The Journal of Finance* 67(4), 1219–1264. 6

Figure 1: Original and Revised PRisk Measures vs. EPU Over Time

This figure displays the broad comovement patterns between the quarterly EPU and the quarterly aggregates of two political risk measures: original Hassan ($PRisk_{\text{ref}}$) and revised Hassan ($PRisk_{\text{revised}}$).

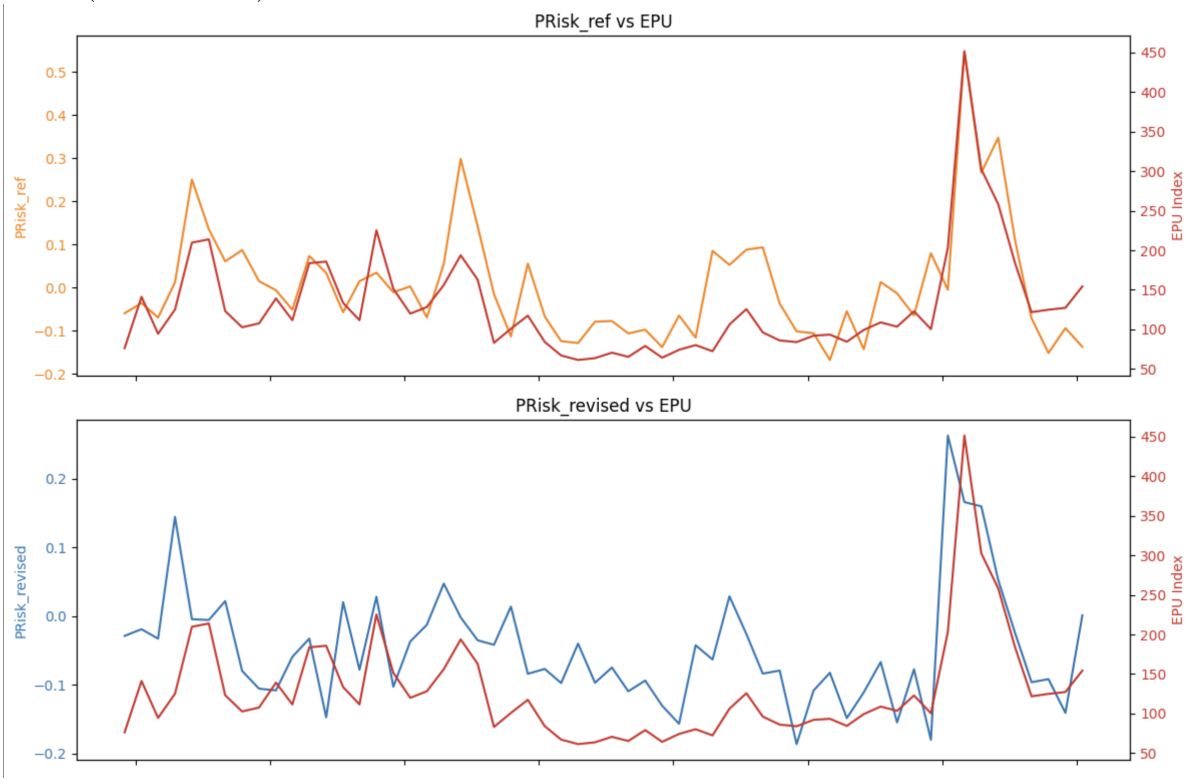

Figure 2: Embedding-Based PRis kMeasures vs. EPU Over Time

This figure provides a macro-level validation of our political risk measures ($PRisk_{\text{emb}}$) by comparing its quarterly aggregates with the EPU index.

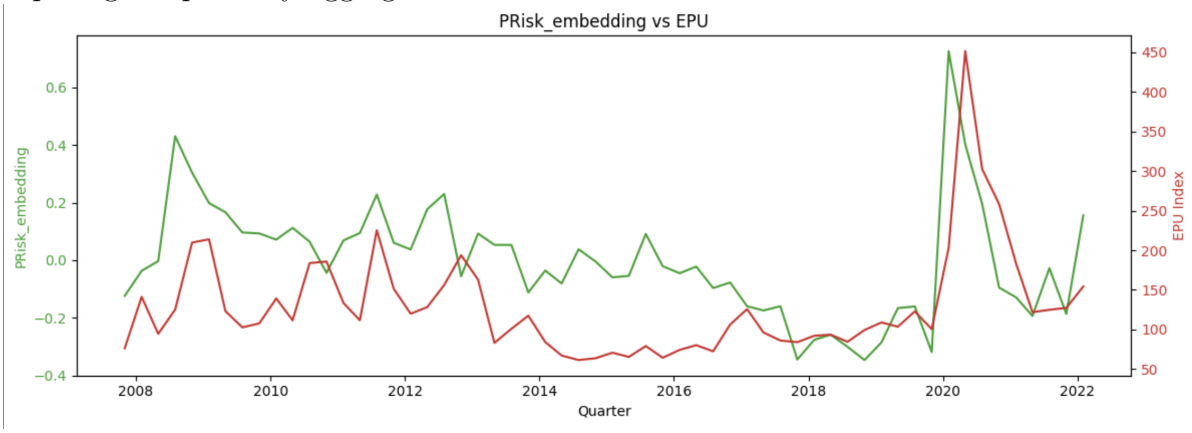

Table 1: Summary Statistics

Variable Name	Mean	Std. Dev.	Min	Max	N
$\text{PRisk}_{\text{ref}}$	533.94	316.75	0	5,942.00	25,709
$\text{PRisk}_{\text{revised}}$	419.42	300.77	0	5788.52	25709
$\text{PRisk}_{\text{emb}}$	0.65	0.24	0	2.16	25,585
Quarterly realized volatility	0.31	0.20	0.01	4.15	25,709
Quarterly implied volatility	0.31	0.14	0.07	1.75	25,709
Intraday volatility	1×10^{-6}	1.81×10^{-4}	0.00	0.14	1,544,453
Assets	65,884.79	221,123.1	692.94	3,757,576	25,703

Table 2: Replication: Correlation between Political risk and Volatility

This table reports our replication of the correlation between political risk and volatility. Columns (1) and (2) report our replicated results, and Columns (4) and (5) reproduce the original estimates from Columns (1) and (5) of Table 4 in Hassan et al. (2019). We use 91-day realized volatility of firm i in quarter t in Panel A and the implied volatility measured from 91-day at-the-money options in Panel B as the dependent variable. Columns (1) and (3) do not control for fixed effects. Columns (2) and (4) control for both time and firm fixed effects. All regressions control for the log of total assets of firm i in quarter t. Realized and implied volatility and PRisk are winsorized at the first and last percentile. All variables are standardized. Parentheses report standard errors clustered at firm level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

	Replication		**Original Paper**	
	(1)	(2)	(3)	(4)
Panel A: Realized vol$_{i,t}$ (std)				
PRisk$_{i,t}$ (std)	0.046***	0.017***	0.048***	0.014***
	(0.011)	(0.006)	(0.005)	(0.002)
R^2	0.007	0.672	0.140	0.621
N	25,703	25,562	162,153	162,153
Panel B: Implied vol$_{i,t}$ (std)				
PRisk$_{i,t}$ (std)	0.045***	0.015***	0.056***	0.013***
	(0.012)	(0.005)	(0.006)	(0.003)
R^2	0.018	0.744	0.214	0.711
N	25,703	25,562	115,059	115,059
Time FE	no	yes	no	yes
Firm FE	no	yes	no	yes

Table 3: Correlation Between Quarterly PRisk and Next-Period EPU

This table reports the correlation between quarterly averages of our firm-level political risk measures and the next-quarter value of the Economic Policy Uncertainty (EPU) index. $\text{PRisk}_{\text{ref}}$ denotes the original Hassan measure, $\text{PRisk}_{\text{revised}}$ is our dictionary-based revision, and $\text{PRisk}_{\text{embedding}}$ is the embedding-based measure constructed from earnings call transcripts.

PRisk Measure	**Correlation with next-period EPU**
$\text{PRisk}_{\text{ref}}$	0.325
$\text{PRisk}_{\text{revised}}$	0.355
$\text{PRisk}_{\text{emb}}$	**0.383**

Overall, the embedding-based PRisk exhibits the strongest macro-level alignment with policy uncertainty and provides the highest predictive correlation with future EPU.

Table 4: Correlation between Political risk and Volatility

This table reports the correlation between political risk and volatility. The dependent variable is 91-day realized volatility of firm i in quarter t in Panel A and the implied volatility measured from 91-day at-the-money options in Panel B. The independent variable is $PRisk_{emb}$ of firm i in quarter t in Columns (1) and (2) and $PRisk_{ref}$ from Hassan et al. (2019) in Columns (3) and (4). Columns (1) and (3) do not control for fixed effects. Columns (2) and (4) control for both time and firm fixed effects. All regressions control for the log of total assets of firm i in quarter t. Realized and implied volatility and PRisk are winsorized at the first and last percentile. All variables are standardized. Parentheses report standard errors clustered at firm level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

	PRisk$_{emb}$		**PRisk$_{ref}$**	
	(1)	(2)	(3)	(4)
Panel A: Realized vol$_{i,t}$ (std)				
PRisk$_{i,t}$ (std)	0.135***	0.072***	0.046***	0.017***
	(0.016)	(0.009)	(0.011)	(0.006)
R^2	0.018	0.674	0.007	0.672
N	25,703	25,562	25,703	25,562
Panel B: Implied vol$_{i,t}$ (std)				
PRisk$_{i,t}$ (std)	0.113***	0.061***	0.045***	0.015***
	(0.019)	(0.009)	(0.012)	(0.005)
R^2	0.025	0.746	0.018	0.744
N	25,703	25,562	25,703	25,562
Time FE	no	yes	no	yes
Firm FE	no	yes	no	yes

Table 5: HMSE of Volatility Forecasting Models

This table reports the HMSE from estimating HAR-RV model under different specifications. The dependent variable is the simple average of realized volatility from day $t+1$ to $t+h$. The predictors in Column (1) include realized volatility at day t ($RV_{i,t}$), lagged weekly realized volatility ($RV_{i,w,t}$), and lagged monthly realized volatility ($RV_{i,m,t}$). The predictors in Column (2) include $RV_{i,t}$, $RV_{i,w,t}$, $RV_{i,m,t}$, and the value of the political risk measure from Hassan et al. (2019) constructed from the most recent quarterly earnings conference call held on or before day t for firm i ($PRisk_{\text{ref}}$). The predictors in Column (3) include $RV_{i,t}$, $RV_{i,w,t}$, $RV_{i,m,t}$, and the value of the our newly constructed political risk measure from the most recent quarterly earnings conference call held on or before day t for firm i ($PRisk_{\text{emb}}$). N-days reports the number of unique trading dates in the out-of-sample period. All PRisk measures are winsorized at the first and last percentile.

	w/o PRisk	**with PRisk**$_{ref}$	**with PRisk**$_{emb}$
	(1)	(2)	(3)
$h = 1$	2,653.20	3,318.85	2,593.21
N-days	2,004	2,004	2,004
$h = 63$	7.27	9.22	7.16
N-days	1,985	1,985	1,985