HOW DOES GEOQUANT WORK?

- Historically, the lack of high-frequency political risk data has impeded analysts' ability to assess and forecast the outcome of political developments and their impact on markets. Until now, data has usually been annual, survey-based, and backwards looking.

- GeoQuant overcomes this limitation by systematically fusing accepted industry data with news media, quantifying country-level political risk on a real-time basis using natural language processing, machine learning algorithms and a team of PhD-trained political economists.

GeoQuant's risk indicators are derived by aggregating two parallel streams of data – **structural** and **high-frequency** data. GeoQuant then uses a proprietary weighting scheme to derive a composite, real-time risk score for each GeoQuant country risk indicator. Each country risk indicator is assigned a 0-100 score; a higher score equals greater risk.

MACRO POLITICAL RISK DATA

STRUCTURAL DATA:
We use 200-250 sources of low-frequency structural data to derive a **baseline** for each GeoQuant risk indicator.

Our **low frequency structural data** is drawn from multilateral institutions, government institutions, NGOs, polling and other reputable country databases.

HIGH-FREQUENCY DATA:
Generated via **natural language processing** of credible media sources identified by GeoQuant's team of PhD political economists and country experts.

High-frequency data is derived from a fusion of accepted industry data and news media sources covering 146 countries over the 2013-2024 time horizon.

Institutional Data:
Legislative seat share and structure, the nature of executive-legislative relations, popular support for the incumbent government.

Security Data:
Size of the military, number of terrorist attacks over the past year, participation in military alliances.

Demographic Data:
Literacy and education rates, population growth rates, income inequality, ethnic diversity.

Policy Environment Data:
Level of corruption, strength of rule of law, ease of doing business.

6 (Re)Train:
Human-generated inputs into the risk assessment are subsequently used to retrain and enhance our algorithms.

5 Merge:
Human scores are then merged with machine scores using a proprietary weighting algorithm, yielding a single high-frequency risk score for each country and risk indicator.

4 Human Review:
Analysts vet a random sample of machine-generated risk assessments. Analysts accept, modify, or delete machine-generated scores, as well as generate additional scores for reinforcement learning.

3 Machine Score:
Articles exceeding country/indicator relevance thresholds are then assigned an impact (e.g. negative/positive) and a duration (e.g. 1-3 months, 2-5 years) using a pre-defined, multi-level impact and duration scale.

2 Sort:
Machine learning algorithms evaluate the text and metadata by relevance to each country and Tier 3 risk indicator.

1 Scrape:
GeoQuant algorithms scrape the web hourly for articles drawn from local, regional and global media across languages.

FREQUENTLY ASKED QUESTIONS

CENTER PANEL VISUALS

What does the score/number mean? GeoQuant's Risk indicators are scaled between 0 and 100 (higher scores, higher Risk). Political Risk is inherently country-specific and GeoQuant's Risk scores reflect relative Risk across countries and within countries over time. A specific score, say 50, does not indicate a specific form or level of Risk. It should be contextualized and interpreted with reference to the level (against other countries) and change in a country's Risk trends over time. Score levels, presented as (five-part) quintiles, reflect differing intensities of Political Risk. Risk scores on specific indicators are interpreted as contributing to these aggregate political risks as follows:

- **0-20:** Low risk of regime change and/or arbitrary policy change, sustained social instability or general physical insecurity.
- **20-40:** Low-Medium risk
- **40-60:** Medium risk
- **60-80:** Medium-High risk
- **80-100:** High risk

What do the colors mean? On the landing page and the heatmap, the color reflects that country's risk trend (literally the slope of the trend line) for the selected time frame. Bright red circles reflect larger positive slopes because higher scores reflect higher Risk. Conversely, bright green circles reflect larger negative slopes.

What is the difference between the solid and dotted lines on a line graph? The two lines correspond to two different types of risk data: composite (solid line) and pulse (volatile, dotted line). The composite trend lines combine our structural and high frequency scores, hence composite. Risk scores range from 0 to 100 with the scale shown on the left y-axis. The pulse trend lines present daily change in risk. These scores center around zero (right y-axis) where scores above zero correspond to increasing risk and scores below zero reflect decreasing risk. A score of zero indicates no change in risk.

PULSE SCORES

What is the pulse score? The pulse measures daily change in risk apart from any trend. It is visualized as the dotted line on the webapp graphs. These scores center around zero (visible in the webapp on the right y-axis). Scores above zero correspond to increasing risk and scores below zero reflect decreasing risk. A score of zero indicates no change in risk.

What does the pulse measure on days when there are no news events? The pulse jumps when there are influential news events, as described in a media article, on a country-specific fundamental risk indicator. When there are no news events, the pulse will trend back towards zero.

How is the pulse different from daily changes in the composite score? A pulse score isolates point-in-time changes in risk based solely on our media feed, whereas our composite score includes both the pulse and a structural baseline in a roughly balanced weighting. As a result, the first difference of the composite – today's score minus yesterday's score – includes (1) the change in pulse score, (2) the change in structural baseline, and (3) the change in the residual influence of the structural baseline on the composite.

What's the best use case for the pulse? The pulse can serve as an early warning signal to monitor political shocks. It offers an independent assessment of an event, apart from any structural influence on how that event will evolve. Thus, it provides a leading indicator of political shifts subsequently revealed by the slower-moving composite risk scores.

How do you validate the pulse? We validate the pulse based on its ability to accurately forecast asset movements, election outcomes, and geopolitical scenarios. The pulse score provides independent explanatory power apart from the composite indicators in these forecast models. It is also validated via our Retrospective Insight Series which tracks how changes in the pulse were leading indicators of political shocks, such as coup d'etats.

NEWS/HIGH-FREQUENCY DATA

How many media sources are fed into the natural language processing (NLP) model? As of June 2024, GeoQuant is scraping approximately 300 media sources worldwide including globally distributed newswires and newspapers and local coverage using country and region-specific media sources.

Do you use languages other than English in the model? Yes, we use six languages at present. This project was developed with an English language model but since 2018 performs bulk translations via Google Translate applied to vernacular sources in Arabic, French, Portuguese, Spanish, and Russian. We are in the process of testing and integrating additional vernacular sources, covering a broader set of languages, including Mandarin.

How do you control for media bias? First, we do not use state-controlled media. Beyond that, we are very careful about curating our country-specific media sources to capture a range of sources from across the political spectrum. Media feed selection is informed and updated according to ongoing discussion with our country-specific political scientists.

Do you use social media? Although we have experimented with using social media feeds in the past, we do not at present use it in our data generation process. We have found social media to have too high a noise-to-signal ratio for our current application.

How often are articles/scores updated in your model system? Our NLP model checks media sources every hour of the day to scrape new articles and machine score them. Machine scores are reviewed and, if necessary, adjusted (to facilitate machine learning) by country analysts daily for larger economies and two-to-three times a week for all other countries. We can provide frequency measures upon request.

STRUCTURAL DATA

How often is structural data updated? Structural data is updated annually based on when international institutional and academic databases release updates. From 2024, we will be updating the structural data more than once annually, depending on the frequency of data releases, For example, when UCDP releases annual conflict data in the spring, this will be updated in the model.

Do your structural data sources include higher frequency economic and financial data? First, GeoQuant does not use financial market or economic data that reflects observable outcomes because we purposefully estimate dimensions of country risk independent of those outcomes to minimize confounding with the sorts of economic outcomes for which our data provides explanatory leverage. Second, observed structural data is typically updated once (or occasionally a few times) a year.

How does the data compare with other country risk benchmarks (e.g., World Bank (WB) WGI, MSCI Governance Quality, etc.)? Conceptually similar GeoQuant indicators correlate highly with such providers on an annual basis, but differ because the GeoQuant data provide more frequent and up-to-date estimates of these risks. Structural data drawn from these sources mean the composite risk scores correlate with their WB and MSCI analogues on an annual basis.

How are varying structural data inputs standardized to provide apples-to-apples comparison? Each risk indicator's structural baseline is grounded in a common set of observable data inputs from various cross-national datasets. We standardize the inputs through a normalization process and then aggregate these normalized values to create a common baseline score.

CONTINGENT RISK SCORES

What is contingent risk? Contingent Risk scores estimate bilateral sentiment between pairs of countries (21,170 potential pairs among the 146 countries in our data generation system) by isolating instances when each country in a pair is scored for the same indicator, either International Relations (IR) Risk or Investment/Trade Risk.

What is the difference between IR Risk and Investment/ Trade Risk? Contingent IR Risk scores reflect diplomatic and security interactions between a country pair and Contingent Investment/Trade Risk scores reflect interactions specific to their bilateral economic relationship including interactions over tariff rates, sanctions, and regulatory alignment.

What does the score mean? Contingent Risk uses the same scale as the fundamental risk indicators; higher scores correspond to higher bilateral risk. For visualization in the web application, we set initial, structural baselines for relative sentiment – reflecting an estimated historical level of positive or negative sentiment for each country pair – on both Contingent IR and Investment/Trade Risks to reflect relative levels of Risk when making visual comparisons.

RISK TAXONOMY

How did you design your taxonomy? The taxonomy is constructed to reflect a comprehensive range of mostly mutually exclusive (even if interactive) political risks. Since risk can be highly subjective, our taxonomy tries to take a step back and provide a more holistic assessment of different dimensions of Political Risk (e.g., Governance, Social, Security). Ultimately the taxonomy is modular, allowing for a large degree of versatility and flexibility so that users can isolate different types of risk or mix-and-match different dimensions of risk to create their own taxonomies, focus indicators, or to explore complex multi-dimensional events.

How are the lower tier risk components weighted and aggregated? Tier 3 Risk indicators (e.g., Elite Support Risk, Ethno-Religious Risk, Political Violence Risk) are custom weighted based on the relative explanatory leverage each indicator exhibits against a battery of politically relevant dependent variables such as GDP per capita, Foreign Direct Investment, Human Development, etc.

MODEL QUESTIONS

How are structural and high-frequency components combined? Each composite Risk indicator is a custom-weighted combination of structural and high-frequency newsflow data. These are combined according to a set of supervised machine learning algorithms and integrated as time-based deviation from structural trend.

How frequently does GeoQuant update algorithms? We perform annual reviews and updates of our machine learning model that are typically implemented on 1 Jan. each year. We update the scoring engine less frequently, but still perform annual reviews and algorithm updates. In practice, it takes time to advance the machine learning models – particularly on large, multidimensional datasets like ours – and so our effort to pay attention to and adopt new techniques is continuous.

How is the machine model trained? Since our data extends back ten years, we have a series of back testable historical data which includes both machine and analyst scores. We train the data on historical events (proxied by a large feed of news articles) and apply standard machine learning methods to estimate an event's relevance for a given risk dimension, the event's impact on that risk, and the duration of that event's impact. We optimize our scores against a given evaluation metric and perform a series of robustness checks before deploying on incoming data.

How do you test the performance and accuracy of the outputs? We test the performance and accuracy of the outputs in several ways:

- Internal validation:
  - We check data quality by iterating through the data, using our political science expertise to identify errors/inconsistencies, updating the algorithm to minimize misclassifications, and then retrain the model to improve future performance.
  - We test the internal accuracy of our data by ensuring scores are measured consistently across different events. Human scorers go through a systematic training process with a standardized codebook, training period, and manager reviews to ensure coding consistency.
  - We evaluate the internal accuracy of our model by comparing its performance on a hold-out set of observations (i.e., observations not used in model estimation) to avoid overfitting.

- External validation (by event or asset type): We test the external accuracy of our data by making sure that the data explains real-world events accurately and in a sensible way. We externally validate our results through a series of analytical models that establish relationships between GeoQuant Risk indicators and financial markets (e.g., foreign exchange rates, bond yields) and through our prediction board (which provides forecasts of complex events such as elections or policy outcomes).

DELIVERY, QUANTITY, POINT IN TIME

How far back does the data go? Our structural and high-frequency newsflow datasets begins in Jan. 2013 for all countries. Our initial data began in July 2016 covering the G20 and increased incrementally to 127 countries by July 2021. From mid-2018 onward we have generated machine scores back to Jan. 2013 for all new countries as they are added to the system.

What are the volumes of data processed on a daily basis? In the 31 days of March 2023 our model logged more than 75,000 machine scores and over 22,000 analyst scores. Between April 2022 and March 2023, our system read an average of 132,000 events/articles per month (4,300 events/articles per day).

Does GeoQuant provide "point-in-time" data coverage? Yes, we preserve daily point-in-time Country Risk scores for both the Composite and High Frequency components of our indicators.

FORECASTS

How do you generate the forecasts? Our forecasts are based on our annual estimates of the structural trend extrapolated forward for two years, and then influenced by the combination of the impact and duration of our daily model-driven, analyst-reviewed scores.

How long are the forecasts? We forecast two years ahead. The high-frequency component of our forecasts fades over the course of each score's estimated duration. As a result, the long-run forecast becomes increasingly structural over time.

How much do the forecasts change over time?

The volatility of the forecasts is low. If you look at the overall composite risk scores you would see that aggregate risk scores do not change much from day-to-day, implying a high degree of stability and consistency in our results. There is higher volatility in the pulse scores which capture day-to-day news flow.

Country forecasts are typically more volatile when a country is added to the data generation system, since a broad set of analyst assessments across all fundamental indicators is needed as the system initially calibrates a country-specific model. For example, there are fewer pulse scores in Myanmar pre-2015. Because there are naturally fewer data points during a country's integration into our system, there is more noise and uncertainty in these estimates.

PREDICTIONS

How are the prediction topics determined? We make formal calls on political events that can be stated in binary terms, typically national election outcomes, governance outcomes (e.g., leadership survival), major policy outcomes, or geopolitical events. We identify emerging events of broad magnitude and significant interest which have substantial uncertainty associated with their outcome. This ensures that we do not artificially boost our prediction rate accuracy by making calls on high likelihood, minor magnitude events. We then have our analysts evaluate the data to make a prediction (formal call) based on the context provided by our data relative to their subject matter expertise.

Are the predictions all binary outcomes and do they have a probability score Yes, we assign binary outcomes ("correct" or "incorrect") to our predictions. We assign a confidence assessment to indicate our relative uncertainty around a prediction. High confidence is comparable to a higher expected probability and vice versa. We encourage clients to ask us questions on issues on which they would like us to make calls but will only make explicit calls when the outcomes can be framed in binary terms. If we cannot frame the question's outcomes in binary terms, we can still provide our ranking of alternative outcomes (but if the most likely outcome has a 30% probability, we would not make an explicit call on that event).

WHAT IS THE ACCURACY OF GEOQUANT PREDICTIONS? HOW MANY PREDICTIONS ARE INCLUDED?

Since 1 July 2016, GeoQuant has correctly predicted 76% of electoral outcomes, including whether there has been an ideological or ruling coalition change following a legislative or presidential election. When we include forecast predictions for a broader set of issues, including policy reforms, government formation, snap election, and geopolitical events, our running accuracy rises to 77% across the sample of 285 calls that have already come to pass.

We also have machine learning forecast models with similar accuracy results. These models forecast specific events, using a more systematic series of machine learning algorithms. For example, our forecast model of electoral outcomes correctly predicts 76% of 359 elections that have occurred since 1 Jan. 2015 in our coverage of 127 countries using data available at least 30 days in advance of an election date. The election model is not yet available on the web application, but we are happy to provide election forecasts to clients upon request.

Please contact us to learn more: **info@geoquant.com**