**Part 1: Keyword Selection**

Step 1: We identified the root words related to geopolitics from following sources: Step 1.1) Caldera and Iacoviello: Measuring Geopolitics Risk and 1.2) by google search and perplexity.ai search, prompt - top words related to geopolitics over the past 20 years and identified the root words from the sources.

Step 2: We input these words in Google trends and identified top 25 related queries to these words over the time period 2004 to 2024.

Step 3: From top 25 related queries we manually selected keywords related to geopolitics from the list.

Step 4: We ran Step 2 again for the words selected from step 3 and ran this loop again and again till the top 25 related queries started repeating words.

Step 5: We got set of 248 words from the search.

**Part 2: Data and methodology – Weekly Index**

**Data:**

Step 1: We downloaded weekly SVI (covering the search volume from Sunday to Saturday) time series from Google Trends for each search term for the sample period from January 2004 to December 2024.

Step 2: Download S&P 500 total return index weekly data from Refinitiv. Calculate weekly returns and 4 week rolling volatility of the calculated returns. The reason for calculating volatility is because risk triggers volatility.

**Methodology:**

Step 3: We calculate the weekly change in SVI (ΔSVI) for each search term. Equation

Step 4: We winsorize each keyword timeseries at 5% level(2.5% at each tail) for extreme observations.

Step 5: To eliminate seasonality from winsorized we regress ΔSVI on month dummies and keep the residuals.

Step 6: To address hetroscedasticity and make time series comparables, we standardize the ΔSVI residuals for each keyword. Equation :

where mean residual and std dev of residual is for each keyword time series.

*From above steps we got adjusted weekly change in SVI which is winsorized, deseasonalized and standaridzed.*

Step 7: We did backward rolling regression of *ΔASVI* on market returns volatility for every six months, i.e. most recent June and December. Regression equation:

βΔASVI+ µ

Step 8: We calculate t-statistics for each beta from above regression and rank the keyword by t-statistics for each year, month and week.

Step 9: We forward fill the rankings for (t + 5) months.

Step 10: To construct sentiment index we did simple average of top 30 search terms and bottom 30 search terms as per t-stat ranks and calculate the difference to get weekly sentiment index values for each year, month and week combination.

**Part 3: Data and methodology – Monthly Index**

**Data:**

Step 1: We downloaded monthly SVI (covering the search volume from Sunday to Saturday) time series data from Google Trends for each search term for the sample period from January 2004 to December 2024.

Step 2: Download S&P 500 index data adjusted close price from yahoo finance. Calculate monthly returns and 2 month rolling volatility of the calculated returns. The reason for calculating volatility is because risk triggers volatility.

**Methodology:**

Step 3: We calculate the monthly change in SVI (ΔSVI) for each search term. Equation

Step 4: We winsorize each keyword timeseries at 5% level(2.5% at each tail) for extreme observations.

Step 5: To eliminate seasonality from winsorized we regress ΔSVI on month dummies and keep the residual value.

Step 6: To address hetroscedasticity and make time series comparables, we standardize the ΔSVI residuals for each keyword. Equation :

where mean residual and std dev of residual is for each keyword time series.

*From above steps we got adjusted weekly change in SVI which is winsorized, deseasonalized and standaridzed.*

Step 7: We did backward rolling regression of *ΔASVI* on market returns volatility for every six months, i.e. most recent June and December. Regression equation:

βΔASVI+ µ

Step 8: We calculate t-statistics for each beta from above regression and rank the keywords by t-statistics for each year and month.

Step 9: We forward fill the rankings for (t + 5) months.

Step 10: To construct sentiment index we did simple average of top 30 search terms and bottom 30 search terms as per t-stat ranks and calculate the difference to get monthly sentiment index values for each year and month combination.