When Words Move Money: Diplomatic Sentiment and International Capital Flows

Cunyi Yang[†]

Faculty of Business and Economics, The University of Hong Kong

Shuchi Zhang[‡]

School of Mathematics, Sun Yat-sen University

Ioannis Kyriakou[§]

Bayes Business School, City St George's, University of London

Nikos C. Papapostolou[¶]

Bayes Business School, City St George's, University of London

March 18, 2026

[†]E-mail: yangcunyi@connect.hku.hk

[‡]E-mail: zhangshch27@mail2.sysu.edu.cn

[§]Corresponding author. E-mail: Ioannis.Kyriakou.2@citystgeorges.ac.uk

[¶]E-mail: n.papapostolou-1@citystgeorges.ac.uk

When Words Move Money: Diplomatic Sentiment and International Capital Flows

ABSTRACT

We study how war-related diplomatic sentiment, often preceding geopolitical risk indices, shapes international equity flows. Using press-conference transcripts from China's Ministry of Foreign Affairs and official statements from the U.S. State Department, we construct daily war-related diplomatic sentiment indices (WDSI) with human validation and extend the same large language model (LLM) pipeline to the United Kingdom, Japan, and South Korea. In China, more negative war-related rhetoric predicts larger southbound purchases of Hong Kong equities and lower net Stock Connect flows: a one-unit decline in China's WDSI corresponds to about a $35.6 million shift towards daily net outflows. In the United States, negative war-related sentiment coincides with higher foreign holdings of U.S. equities, consistent with safe-haven demand, and since the Russia-Ukraine war, more negative U.S. rhetoric is associated with weaker northbound inflows into China. Beyond the China-U.S. pair, negative war-related diplomatic sentiment in other major developed economies also spills over to Chinese equity flows, tilting transactions towards net capital outflows from China. Additional evidence suggests that war-related diplomatic sentiment is associated with significant stock-return responses and higher market activity.

Keywords: diplomatic sentiment; capital flows; geopolitical risk

JEL classification: F36, G15, F51

1 Introduction

In recent years, the global economy has operated under a persistent cloud of geopolitical tension. Great-power rivalry, recurrent regional conflicts, and the weaponisation of supply chains have all contributed to an environment in which international investors must continuously reassess geopolitical risk (Caldara and Iacoviello, 2022). Financial markets react sharply not only to the outbreak of war or the announcement of sanctions, but also to more subtle shifts in diplomatic language that shape expectations about future conflict and cooperation (Pástor and Veronesi, 2012). Yet most empirical work still treats international relations as a coarse set of states, peace versus war or sanctions versus no sanctions, rather than as a continuum of evolving diplomatic sentiment. As such, while theory offers many hypotheses on how political relations affect markets, what is missing is a reliable high-frequency measurement tool of diplomatic sentiment. This paper examines such latent diplomatic sentiment, constructs a high-frequency, continuous index of war-related diplomatic sentiment, and studies how it affects international capital flows.

Our point of departure is a large literature in international relations and political economy that seeks to measure the state of bilateral or multilateral relationships. Classic approaches rely on discrete indicators such as formal alliances, war onsets, militarised interstate disputes, or sanctions episodes, often drawing on event-based datasets such as the Correlates of War project or conflict event codings (Gowa, 1994; Hufbauer et al., 2007; Maoz, 2010). Other studies measure political affinity using revealed behaviour, including UN voting alignment, diplomatic recognition, or the intensity of high-level visits and summits (Gartzke, 2006; Fuchs and Klann, 2013; Bailey et al., 2017). Parallel work in macro-finance examines the pricing of geopolitical risk and policy uncertainty using news-based indices and event counts, such as the geopolitical risk and the economic policy uncertainty indices (Baker et al., 2016; Caldara and Iacoviello, 2022). A common feature of these measures is that they are typically low-frequency and highly discrete: they move mainly when conflict escalates into war, when sanctions are imposed, or when formal alliances change. They are, therefore, well suited for identifying large regime shifts, but ill equipped to capture the gradual cooling or warming of diplomatic relations and the more nuanced emotional tone of official discourse.

Recent text-as-data work has begun to infer foreign-policy positions and diplomatic narratives from political speeches and official communications (Baturo et al., 2017; Mochtak and Turcsanyi, 2021). In finance, official policy documents have also been shown to contain economically meaningful sentiment: Correa et al. (2021) construct sentiment measures from central banks' financial stability reports and show that they help explain the financial cycle and predict banking crises. Complementing the academic literature, several recent index-building initiatives by think tanks and private data providers have operationalised diplomatic engagement or conflict dynamics with NLP-based indicators – for example, HCSS's GINA Diplomatic, which combines dyadic UN voting similarity with sentiment and tone measures extracted from UN General Debate speeches (Kommandeur and van Dam, 2024), and Permutable AI's Global Conflict Tracker and War Index, which use large-scale AI-driven news analytics to monitor armed-conflict risk in (near) real time (Permutable AI, 2024). These initiatives underscore the feasibility of constructing continuously updated geopolitical indicators. Yet, for academic research, high-frequency measures built directly from the day-to-day language of official diplomacy (e.g., foreign-ministry briefings and statements), rather than annual multilateral speeches or ex post media coverage, remain relatively limited, especially when they are transparently constructed, systematically validated, and informative before geopolitical conditions are fully reflected in newspaper-based risk indices. Motivated by this gap, we construct daily, country-level war-related diplomatic sentiment indices from official foreign-ministry communications using a multi-stage large language model (LLM) pipeline with built-in cross-validation and human audits, and study their implications for international equity flows.

We introduce the *War-related Diplomatic Sentiment Index* (WDSI), which aims to quantify the emotional tone of a country's official statements on matters related to war, military tensions, and security threats. Conceptually, diplomatic sentiment is a latent, continuous variable that summarises how friendly, neutral, or hostile one state is in its official communication towards other states at a given point in time. Rather than relying on discrete events, we use the full text of foreign-policy communication to infer this sentiment at a daily frequency. As such, war-related diplomatic sentiment is treated as a structured text feature that can be extracted using modern language models. Starting from high-frequency official texts, we classify and score each statement according to whether it conveys concern, support, condemnation, or threats, and aggregate these into a daily index. The resulting WDSI serves as a continuous, high-frequency instrument for measuring war-related diplomatic tone, facilitating temporal analysis and cross-country comparisons.

We construct daily war-related diplomatic sentiment indices for China and the United States using a multi-stage LLM pipeline with human validation. The indices build on a large corpus of official foreign-ministry communication. For China, the primary source is the transcript of regular press conferences of the Ministry of Foreign Affairs (MFA), which provide high-frequency, policy-relevant commentary on current events. For the United States, we use press statements and official remarks by the Department of State. We also extend the index construction to the United Kingdom, Japan, and South Korea, thereby enabling a unified cross-country analysis of war-related diplomatic rhetoric and financial spillovers. The construction pipeline consists of several key steps: (1) we clean and standardise the raw text, including segmentation, removal of boilerplate, and harmonisation across Chinese and English language content, (2) we identify segments that are genuinely related to war, military tensions, or security threats, as opposed to purely economic or cultural topics, (3) we apply a multi-stage sentiment classification procedure, implemented with a large language model, to assign each war-related segment a sentiment score on an ordered scale ranging from strongly negative to strongly positive, and (4) we aggregate these scores to the day-country level and apply a short-term moving average to reduce idiosyncratic noise. The end result is a continuous, high-frequency index that is explicitly anchored in the language of official diplomacy rather than media coverage or realised events.

A natural concern is whether such an index truly captures meaningful variation in war-related diplomatic sentiment rather than model artefacts. We therefore devote substantial effort to validating the WDSI. First, we benchmark it against well-known geopolitical events, i.e., major military confrontations and ceasefire agreements, and verify that the index reacts in the expected direction and timing. Second, we compare it to the widely used geopolitical risk index. The WDSI exhibits only modest contemporaneous co-movement with GPR, retains substantial independent variation, and, as we later show in daily VAR evidence, temporally leads newspaper-based GPR measures, indicating earlier policy-signal content rather than simple duplication. Third, we conduct manual audits using human coders trained in international relations to evaluate whether our automated labels correspond to their qualitative assessments of the same texts. Finally, we show that the WDSI is predictive in settings where war-related sentiment should matter, including cross-border capital flows as well as stock market returns and market activity, even after controlling for standard macro-financial drivers. Rather than positing sharp theoretical hypotheses, we view these empirical relationships as tests of whether the WDSI behaves like a credible measurement tool.

Our main empirical tests focus on international capital flows, with China as the primary setting by design. This choice is motivated by both institutional structure and data quality. Because China maintains capital-account management, cross-border equity transactions are channelled through clearly defined programs (northbound and southbound Stock Connect), which creates a rare setting where daily cross-border flows are both observable and directionally decomposable. This institutional architecture allows us to distinguish purchases of A-shares by offshore investors (northbound) from mainland purchases of Hong Kong equities (southbound), and to map changes in diplomatic tone to high-frequency reallocations in a tightly identified market interface. In many other countries, by contrast, publicly available international-flow measures are typically monthly, more aggregated, and less directly aligned with bilateral trading channels, which weakens comparability and attenuates short-horizon inference. More broadly, our focus on cross-border capital reallocation is consistent with evidence that political uncertainty reshapes other forms of international investment, including cross-border acquisitions (Cao et al., 2019).

We therefore use China as the main empirical laboratory and then extend externally along two dimensions. First, we complement the China evidence with U.S. data on cross-border portfolio positions and flows from the Treasury International Capital system, using the U.S. WDSI as the main explanatory variable. Second, we go beyond the China-U.S. pair and deploy UK, Japanese, and South Korean WDSI series to test whether war-related diplomatic sentiment in other major developed economies spills over to Chinese equity flows traded through Stock Connect. We also analyse stock returns, trading volume, and return volatility in major equity indices to assess whether war-related sentiment affects not only capital reallocation but also trading intensity and risk absorption. Our empirical framework is based on time-series regressions with a broad control set for global and domestic financial conditions, including equity returns, geopolitical risk indices, economic policy uncertainty, exchange rates, and safe-haven demand.

A central design choice is to treat cross-border capital flows as the primary dependent variable and stock returns as a complementary outcome. In semi-strong efficient markets, publicly available text signals can be reflected in prices quickly, so close-to-close return predictability is often weak even when the information is economically relevant. By contrast, text-based signals are frequently more robust in quantity and risk dimensions (trading activity, volatility, and portfolio reallocation) than in aggregate daily returns. This emphasis on quantity-based outcomes is also consistent with recent evidence from China showing that foreign order flow contains predictive information about future returns and news at the firm level (Lundblad et al., 2026). This perspective is also consistent with international-finance evidence that treats equity flows as a first-order object with their own persistence and adjustment dynamics (Froot et al., 2001; Brennan and Cao, 1997; Portes and Rey, 2005; Griffin et al., 2004; Hau and Rey, 2006). To complement the flow-based baseline, Section 4 also presents additional price-side evidence from Chinese equity-index returns under matched control specifications.

The main findings can be summarised in two dimensions: measurement and economic impact. On the measurement side, the WDSI provides a finely grained representation of the evolution of war-related diplomacy. It responds in a timely manner to major security events, distinguishes between different phases of escalation and de-escalation, and reveals gradual drifts in diplomatic tone that are invisible to standard discrete indicators. Compared with existing proxies such as war dummy variables, sanctions episodes, or newspaper-based geopolitical risk measures, the WDSI offers higher temporal resolution and earlier, more forward-looking information about evolving geopolitical conditions. On the economic side, we find that more negative Chinese war-related diplomatic sentiment is associated with net capital outflows from China through the Stock Connect southbound channel, consistent with domestic investors reallocating towards cheaper Hong Kong-listed shares when perceived war risk rises[1]. In contrast, more negative U.S. war-related sentiment coincides with increased foreign holdings of U.S. equities and a more positive U.S. net equity position, in line with the safe-haven role of U.S. financial assets (Krishnamurthy and Vissing-Jorgensen, 2012; Caballero et al., 2017; Maggiori et al., 2020). We also provide evidence of cross-country spillovers beyond the China-U.S. pair, where more negative war-related diplomatic sentiment in other major developed economies (United Kingdom, Japan, and South Korea) propagates to Chinese equity flows, tilting cross-border transactions towards net capital outflows from China, with heterogeneous strength across countries. We also find that war-related diplomatic shocks raise trading volume in both Chinese and U.S. markets, while increasing volatility in the U.S. and dampening short-run volatility in mainland Chinese indices, suggesting different institutional mechanisms for absorbing war risk.

The paper contributes to three strands of literature. First, it adds to international re-

[1]Here, "cheaper" refers to the relative valuation of firms that are dually listed in mainland China and Hong Kong. In particular, it captures the well-documented A-H share pricing differential, whereby H-shares of the same firm often trade at a discount relative to their corresponding A-shares.

lations and international political economy by moving beyond discrete conflict or alliance indicators towards a continuous, text-based measure of diplomatic sentiment. By transforming official diplomatic language into a time-varying index, we provide a tool that can be used to study the dynamics of interstate relations at much higher frequency and granularity than traditional event data (see Baum and Zhukov (2019) and Gentzkow et al. (2019) for related efforts to quantify foreign-policy stances from text). Second, building on the literature on political risk and international investment (Papaioannou, 2009; Julio and Yook, 2012; Kelly et al., 2016; Cao et al., 2019), it contributes to international macro-finance by introducing a bilateral political-sentiment measure that can be linked to trading, investment, and portfolio allocation decisions. Our results show that war-related diplomatic sentiment has explanatory power for cross-border capital flows and market behaviour (Benhima and Cordonier, 2022; Feng et al., 2023; Choi and Havel, 2025) even after controlling for conventional measures of geopolitical risk and macroeconomic conditions, thereby helping to disentangle expectation-driven channels from realised policy actions. Third, we contribute to the growing literature on text-based measurement in economics and finance by demonstrating how large language models can be used to construct structured indices of "soft" variables such as sentiment, hostility, or trust from official documents (Buehlmaier and Whited, 2018; Gentzkow et al., 2019; Hassan et al., 2019; Correa et al., 2021). Our work establishes a general framework for turning qualitative diplomatic communication into quantitative data that can be deployed in empirical models.

The remainder of the paper is organised as follows. Section 2 introduces the concept of war-related diplomatic sentiment, describes the data sources, and details the construction of the WDSI for China and the United States. Section 3 presents the main empirical design and baseline results on cross-border equity flows. Section 4 extends the analysis to the U.S. case, examines cross-country spillovers (including evidence from the United Kingdom, Japan, and South Korea), and provides additional price-side and market-response evidence, including the A-H share premium, complementary return evidence, and market activity. Section 5 concludes.

2 Construction of the War-Related Diplomatic Sentiment Index

The construction of the WDSI involves four main steps: definition, measurement, audit and presentation. We first set out the definition of war-related diplomatic sentiment used in this study. We then explain how the index is built, including methodological design and implementation. Next, we examine its reliability through a rigorous audit. Finally, we present the core features of the index and discuss the insights that follow.

2.1 Definition: Diplomatic Sentiment and War-Related Diplomatic Sentiment

Diplomatic sentiment, as used in this paper, captures the emotional tone and attitudinal orientation conveyed in a state's official diplomatic statements concerning specific international events, the conduct of other nations, or the state of bilateral and multilateral relations. Unlike personal emotion, it is expressed through formal and strategic official language, such as that used in foreign ministry press briefings, and may thus be regarded as a policy-relevant soft signal that reveals national preferences, risk tolerance, and willingness to escalate or de-escalate a situation.

On this basis, we define war-related diplomatic sentiment as the emotional colouring and attitudinal stance embedded in a state's official foreign policy discourse when addressing issues of war and armed conflict, military confrontation and security crises, as well as the possible outbreak, escalation or easing of such situations. In contrast to broader notions of geopolitical risk or general security tension, this concept highlights both the emotional response and evaluative judgement concerning the use of force and its anticipated consequences, rather than a simple record of events. We therefore focus not only on whether war or conflict is present, but also on the emotional tone through which the foreign ministry signals its position.

In terms of scope, war-related diplomatic sentiment covers at least three categories. First, in response to ongoing or highly salient armed conflicts, such as regional wars, invasions, or large-scale military operations, statements expressing grave concern, sympathy, solidarity, or calls for ceasefire and negotiation all constitute war-related sentiment. Second, when addressing latent military frictions and security risks – such as joint exercises, arms sales, alliances, border or maritime disputes, nuclear and missile programs, adjustments to defence treaties or base deployments – official warnings, objections, firm positions and conciliatory signals likewise reflect judgements about war risk. Third, in cases where political events may trigger military escalation, such as coups, violent clashes, terrorist attacks or the breakdown of major security treaties, any statement framed in terms of international security or armed confrontation is also included. By contrast, matters limited to economic, cultural or technological affairs – without explicit elevation to security or military relevance – fall outside the scope of this definition.

2.2 Measurement: Data, Methodology, and Index Construction

2.2.1 Data

We use transcripts of regular press conferences of the Ministry of Foreign Affairs of China from 1 January 2010 to 30 September 2025. All texts released on the same day are merged to construct a daily WDSI. A detailed description of the data is provided in Appendix A.1. The processed country-level WDSI series used in the empirical analysis are available on a companion website.[2]

Our reliance on MFA regular press conferences is motivated by both institutional and econometric considerations. Institutionally, these texts are issued by the central foreign-policy authority and disseminated through official and mainstream media channels, so they combine high formal status with wide external exposure and agenda-setting power. Substantively, the limited number of questions in each briefing means that journalists concentrate

[2] https://yangcy0126.github.io/wdsi-index-site/.

on the most salient war, conflict and security topics of the day, so that the discourse is tightly aligned with contemporaneous shocks and explicitly addresses the use of force and its risks. Econometrically, questions are largely selected by the media, while the spokesperson's lines are pre-cleared within the bureaucracy. Briefings usually take place before the market closes on the same day, which helps mitigate concerns that short-term financial conditions feed back into the tone of the statements. As a result, the constructed WDSI more closely approximates an exogenous policy signal than indices derived from ex post news coverage.

2.2.2 Methodology: LLM-Based Multi-Stage Sentiment Extraction

Given our research objective and the properties of the data, we construct an evaluation pipeline that combines large language models with task-specific procedures. The core idea is to treat sentiment detection as a structured process. Rather than relying on a single opaque judgement, we decompose the task into a sequence of simple stages, improving the accuracy of sentiment extraction. Existing work usually adopts a horizontal decomposition where several related sub-indices are aggregated into a composite measure. Our approach is instead vertical. The extraction of a single sentiment label is divided into successive stages, each implemented by an AI agent designed for a specific subtask. This vertically decomposed, multi-agent pipeline improves reliability and interpretability relative to one-shot scoring by simplifying the inference task at each stage and adding a consensus check, in line with the text-as-data measurement literature (Buehlmaier and Whited, 2018; Gentzkow et al., 2019; Hassan et al., 2019).[3]

At each stage we also employ cross-validation. Multiple independent judgements are generated, and the pipeline proceeds only when these judgements coincide. This rule strengthens the transparency and reliability of the final classifications. Overall, compared with existing applications of large language models to sentiment extraction, our framework achieves

[3]Consistent with this rationale, Appendix A.3.3 reports that the multi-step pipeline outperforms single-step direct scoring by about 5-10 percentage points in overall accuracy and reduces occasional reasoning errors.

stronger interpretability and robustness, while remaining flexible enough to accommodate diverse linguistic styles and capture sentiment across multiple levels.

Detailed procedures and results are provided in Appendix A.2, and Figure 1 presents a simplified schematic. As Figure 1 illustrates, the process begins with screening the questions, followed by identifying and classifying military-related sentiment. We then apply custom-designed AI agents to quantify the intensity of each sentiment category, which yields the final sentiment index.

[Figure 1 about here.]

2.2.3 Index Construction

Building on the procedures described above, the initial Chinese WDSI assigns to each press conference a sentiment score on a discrete scale from -3 to 3, where lower values indicate stronger negative sentiment, and vice versa. For missing values we adopt a forward-filling procedure, using the most recent past press conference to proxy the sentiment indicator for the current day. Since press conferences are approximately evenly distributed over time and the share of trading days without a conference does not exceed 20% of all working days, and given that the influence of sentiment is inherently subject to delay, this treatment is reasonable. The distribution of scores, together with several illustrative examples, is presented in Table 1.

[Table 1 about here.]

To mitigate the excessive volatility of single conference readings and the influence of random disturbances, we compute a seven-day rolling average of the initial series to obtain the final WDSI, as shown in Figure 2. Figure 2 also reports the evolution of the index's mean and volatility under an annual rolling window. Over time, the volatility declines and the mean gradually drifts downwards, which is consistent with the fact that China has been conveying increasingly negative sentiment in the military domain in recent years.

[Figure 2 about here.]

2.3 *Audit: Validity and Reliability Verification*

Indicators constructed from text data often raise widespread concerns, primarily in two respects: first, whether the model can accurately identify military-related content and its sentiment; and second, whether the index is excessively sensitive to specific technical settings, thereby introducing systematic bias. To address these concerns, we conduct a comprehensive validation of the WDSI from three perspectives: manual auditing, large-model comparison and selection, and robustness tests.

First, we conduct manual annotation and consistency checks. From the full sample we draw, at equal monthly intervals, a random subset of press conferences and official statements, yielding approximately 1,200 question-answer units. Two research assistants trained in international relations independently code these materials, without access to the model outputs. They label (i) whether the text concerns war or military security, (ii) the sentiment category, and (iii) the sentiment intensity. Across annotators, the concordance rate for the "war- or military-related" indicator is about 88 percent, agreement on sentiment category is around 80 percent, and agreement on sentiment intensity is roughly 83 percent.

Second, we compare large language models with more traditional approaches. Prior to formally constructing the WDSI, we assess the performance of three families of methods on the manually labelled sample described above: (i) a traditional approach based on hand-crafted dictionaries and keyword counts, (ii) shallow models that combine word embeddings with supervised learning, and (iii) several large language models capable of handling mixed Chinese and English input. The results show that the Qwen-3 large language model adopted in this paper exhibits a clear advantage at every stage of the pipeline.

Third, we contrast a "single-step direct scoring" strategy with a "multi step decomposition combined with multiple agents that must agree before proceeding" workflow. The evidence indicates that for both human coders and AI agents, the latter approach substantially reduces error rates by simplifying the logical burden at each individual step.

Detailed process and statistics of the above tests are reported in Appendix A.3.

2.4 Understanding the War-Related Diplomatic Sentiment Index

2.4.1 Diplomatic Sentiment Across Categories and Regions

Using the same procedure described in Section 2.2, we also construct category specific indices of diplomatic sentiment, including an economically related diplomatic sentiment index (EDSI) and an index of other diplomatic sentiment (ODSI). We then compute the corresponding three types of diplomatic sentiment indices for the United States and compare the two countries along these dimensions[4]. The U.S. WDSI is presented in Figure 3.

[Figure 3 about here.]

We first find that the Chinese and American WDSI series exhibit only a modest overall correlation, with a coefficient of 0.09. This largely reflects the absence of sustained direct military confrontation between China and the United States. The Chinese WDSI is more closely tied to rhetoric on East Asia and China's immediate periphery, particularly references to Japan, the Republic of Korea, India, and Taiwan-related cross-strait issues, whereas U.S. war-related diplomatic sentiment is more often driven by statements concerning Middle Eastern conflicts and Russia, among other theatres. A clear piece of evidence is the marked increase in the correlation of the two WDSI series to 0.18 in 2016, when China and the United States engaged in an intense stand-off in the South China Sea. Another piece of evidence is that, after Donald Trump assumed the presidency on 20 January 2025, the relationship between the two indices turned distinctly negative, with a correlation of -0.17. Following Trump's re-election, Sino-American rivalry intensified, the United States adopted a more restrained military posture, and Chinese official rhetoric grew markedly more assertive.

We further find that China's WDSI is positively correlated with both the EDSI and the ODSI, with correlation coefficients of 0.18 and 0.25, respectively, whereas the corresponding

[4] The American textual corpus is drawn from the press statements of the United States Department of State from 1 January 2010 to 30 September 2025. Detailed description of the data is provided in Appendix A.1.

correlations for the United States are substantially stronger at 0.48 and 0.44. The divergence between the two countries may arise from differences in how military sentiment is expressed. A plausible interpretation is that China tends to be more restrained than the United States when articulating military-related sentiment, and only signals such sentiment when military issues are directly implicated and perceived threats are salient. The United States, by contrast, is more inclined to embed war-related sentiment in a broader set of economic or other non-military diplomatic issues. One indication is that, since Trump's second term began, the correlations between the American WDSI and the EDSI and ODSI declined to 0.14 and 0.40, respectively[5].

2.4.2 Comparison with Geopolitical Risk (GPR)

In the existing literature, the benchmark index most closely related to the WDSI and most widely used is the Geopolitical Risk Index (GPR) proposed by Caldara and Iacoviello (2022). Although both series focus on war, conflict, and security, they capture different layers of geopolitical information.

The GPR is essentially a newspaper-based event-risk measure. It tracks reports of objective developments such as war threats, terrorist attacks, and military confrontations, and quantifies the frequency and intensity of geopolitical events together with the uncertainty they generate. By contrast, the WDSI is a sentiment-based measure extracted from official diplomatic communication. It captures the tone and policy stance embedded in government statements regarding war and the use of force. Put differently, the GPR is closer to the realised salience of geopolitical events in the news, whereas the WDSI is closer to the evolving policy signal conveyed through official rhetoric. The two indices are therefore related, but not redundant.

This distinction is reinforced by their data-generating processes. The GPR is constructed from newspaper coverage and is thus shaped by media attention and the news cycle. The

[5] Since these patterns lie somewhat beyond the central focus of this study, we do not explore them in greater depth.

WDSI, by construction, is derived from routine press conferences and official statements issued by central foreign-policy authorities. As a result, the WDSI can move when official rhetoric hardens or softens even before geopolitical developments are fully reflected in media-based risk measures, whereas the GPR tends to respond more strongly once conflicts become salient as realised events. This timing difference suggests that the WDSI may contain earlier information about evolving geopolitical conditions.

In the data, the two indices display only modest contemporaneous co-movement. The correlation between China's baseline WDSI and the global GPR index is about -0.10. We also find that the correlations between the WDSI and the China- and U.S.-specific GPR indices are about -0.13 and -0.09, respectively. These relatively low contemporaneous correlations indicate that the WDSI does not simply replicate the information already embedded in existing GPR measures. At the same time, low contemporaneous correlation does not imply the absence of a relationship; rather, it is consistent with the two indices being linked through a lead-lag structure. In particular, the GPR tends to exhibit sharper spikes around realised conflict events, whereas the WDSI varies more smoothly as a measure of official sentiment.

To examine this timing channel directly, we estimate bivariate daily VAR models between China's thirty-day sentiment index $WDSI_t^{CN,30}$ and the China-specific geopolitical risk index GPR_t^{CN} in the full sample over the full lag grid, $p = 1, \dots, 22$. For each lag order, we run Granger-causality tests in both directions. Table 2 and Figure 4 report the full profile of the Granger-causality test P-Values. The pattern is clear. At very short lags (1-6), the predictive relation from WDSI to GPR is limited. From lag 7 onwards, however, the null that WDSI does not predict GPR is rejected at conventional levels, whereas the reverse-direction null (GPR $\rightarrow$ WDSI) is not rejected at any lag order. We interpret this asymmetric lag profile as evidence that the WDSI temporally leads the GPR: changes in official war-related diplomatic tone appear in our sentiment measure before they are fully codified in newspaper-based geopolitical risk indices.

[Table 2 about here.]

[Figure 4 about here.]

This lead-lag structure is important for both measurement and empirical use. Relative to conventional GPR measures, the WDSI appears to provide a more forward-looking signal of evolving geopolitical conditions. For investors and other forward-looking agents, this timing advantage implies that the WDSI can serve as an earlier indicator of geopolitical deterioration and of the repricing and portfolio adjustment that may follow. In this sense, the value of the WDSI is not that it replaces the GPR, but that it complements it by delivering earlier information about geopolitical conditions before those conditions become fully reflected in media-based risk measures.

3 War-related Diplomatic Sentiment and Capital Flows

3.1 Baseline Results: The Case of China

Consistent with the measurement framework in Section 2, we begin by examining how Chinese war-related diplomatic sentiment may affect cross-border equity flows through the Stock Connect programs. These programs record daily trading in Chinese equities between mainland Chinese and international investors via both northbound and southbound channels[6].

Our baseline specification relates daily capital flows to war-related diplomatic sentiment, controlling for domestic and global financial conditions. According to He et al. (2023), for the Chinese market we estimate

$$Flow_{k,t}^{CN} = \alpha_k^{CN} + \beta_k^{CN} WDSI_t^{CN} + \boldsymbol{\gamma}_k^{CN\prime}\mathbf{R}_t + \boldsymbol{\delta}_k^{CN\prime}\mathbf{Z}_t + \mu_{m(t)} + \varepsilon_{k,t}^{CN}, \tag{1}$$

[6]The Stock Connect schemes link the Shanghai and Shenzhen exchanges with the Hong Kong Stock Exchange through two distinct channels. Northbound trading allows overseas and Hong Kong-based investors to trade eligible A-shares listed on the mainland exchanges, whereas southbound trading allows mainland Chinese investors to trade eligible Hong Kong-listed equities. Accordingly, our daily flow measures capture the joint behaviour of foreign investors (primarily in the northbound segment) and domestic investors reallocating wealth into offshore Chinese listings (through the southbound segment).

where t indexes trading days and $k \in \{\text{Inflow}, \text{Outflow}, \text{Net flow}\}$. The dependent variable $Flow^{CN}_{k,t}$ is constructed from Stock Connect trading as follows. Inflow is the net northbound trading volume, capturing net purchases of mainland A-shares by investors using the northbound channel. Outflow is the net southbound trading volume, capturing net purchases of Hong Kong-listed equities by investors using the southbound channel. Net flow is the aggregate net cross-border flow combining northbound and southbound trades, and summarises the overall net capital movement associated with Stock Connect activity. All three series are winsorised at the 5th and 95th percentiles in levels (He et al., 2023).

The key explanatory variable $WDSI^{CN}_t$ is the seven-day moving average of the Chinese war-related diplomatic sentiment index constructed from Ministry of Foreign Affairs press-conference transcripts using a large language model, as described in Section 2. The index is constructed so that higher values reflect more positive war-related diplomatic sentiment. Accordingly, when the diplomatic tone becomes more negative, the index value falls. Thus, a negative shock to war-related sentiment is captured as a decline in the $WDSI^{CN}_t$.

The control vector $\mathbf{R}_t$ comprises three stock market returns series: U.S. (CRSP Ret), all other developed countries (Developed ex US Ret), and China (A-share Ret).[7] These controls are in line with the international finance literature documenting that cross-border equity flows co-move strongly with both domestic and global equity market performance (e.g., Froot et al., 2001; Griffin et al., 2004; Portes and Rey, 2005; He et al., 2023). The control vector $\mathbf{Z}_t$ includes the Chinese and U.S. geopolitical risk indices (Caldara and Iacoviello, 2022)[8], the Chinese and U.S. economic policy uncertainty indices (Baker et al., 2016)[9], and gold returns (defined as log first differences of gold prices), which capture short-run adjustments

[7]The U.S. and developed-market equity returns are obtained from the data library of Kenneth R. French: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french`. The Chinese equity market return is from the CSMAR database. The developed-market return corresponds to a broad developed-market portfolio that excludes the United States, so as to avoid mechanically double counting U.S. market movements when both R^{US} and the developed-market return are included as controls. All three market-return series are winsorised at the 5th and 95th percentiles to mitigate the influence of extreme market conditions (e.g., the global outbreak of COVID-19 in March 2020) (He et al., 2023).

[8]Data downloaded from `https://www.matteoiacoviello.com/gpr.htm` on 23 October 2025.

[9]Data downloaded from `https://www.policyuncertainty.com/index.html` on 23 October 2025.

in safe-haven asset dynamics (Baur and McDermott, 2010), and the bilateral RMB-USD exchange rate, which is tightly linked to international portfolio rebalancing (Caporale et al., 2015; Gelman et al., 2015). Monthly GPR/EPU series are mapped to daily frequency by assigning the monthly value to each trading day within the month (last observation carried forward). All control variables are winsorised at the 5th and 95th percentiles. Month fixed effects $\mu_{m(t)}$ absorb within-year seasonality, and the error term $\varepsilon_{k,t}^{CN}$ captures idiosyncratic shocks. Equation (1) is estimated separately for the three flow measures using ordinary least squares (OLS) with heteroskedasticity-robust standard errors. Descriptive statistics for all variables used in the baseline regressions are reported in Table 3.

[Table 3 about here.]

Table 4 reports the baseline results for daily Stock Connect flows. Columns (1)-(3) correspond to net northbound inflows, net southbound flows, and the combined net flow, respectively. The coefficient on Chinese war-related diplomatic sentiment, $WDSI_t^{CN}$, is -0.025 and statistically insignificant in the inflow regression, indicating that changes in war-related rhetoric do not have an impact on net northbound purchases of A-shares by overseas and Hong Kong-based investors. In contrast, the outflow regression in column (2) yields a statistically significant coefficient of -0.265 on $WDSI_t^{CN}$, implying that a deterioration in China's war-related diplomatic tone is associated with larger net southbound purchases of Hong Kong-listed equities by mainland investors; therefore, larger capital outflows from the mainland through Stock Connect.

[Table 4 about here.]

The net-flow regression in column (3) shows a statistically significant coefficient of 0.356 on $WDSI_t^{CN}$. Given the definition of net flow as net northbound purchases minus net southbound purchases, and the fact that the inflow response in column (1) is statistically insignificant, the net flow response largely mirrors the southbound channel, with net northbound flows remaining unaffected. More positive war-related sentiment is associated with reduced (increased) net capital outflows (inflows) from the mainland (International/Hong Kong) investors, whereas a more negative diplomatic tone tilts the balance of Stock Connect activity towards net capital outflow. Quantitatively, a one-unit decline in the seven-day moving average of $WDSI_t^{CN}$ corresponds to roughly \$35.6 million reduction in daily net Stock Connect flows, reflecting a shift towards larger net outflows.

Consistent with a large international finance literature showing that cross-border equity flows are strongly procyclical and co-move with equity-market returns, reflecting positive-feedback (return-chasing) trading and portfolio-rebalancing motives (e.g., Bohn and Tesar, 1996; Brennan and Cao, 1997; Froot et al., 2001; Griffin et al., 2004), the Chinese market return enters positively and very precisely in all three flow regressions. Higher A-share returns are associated with stronger northbound inflows and stronger southbound trading, indicating procyclical position adjustments on both sides of Stock Connect. More broadly, the determinants of international equity transactions also reflect information frictions and other cross-border impediments (Portes and Rey, 2005).

The coefficients on the U.S. and developed-market return controls also point to systematic rebalancing across major markets. Importantly, because net flow is defined as northbound inflow minus southbound outflow, any regressor that increases southbound outflows mechanically reduces net flows unless it is offset by a sufficiently positive inflow response. In line with this arithmetic, higher U.S. market returns are associated with larger southbound outflows and, together with a modest decline in northbound inflows, with lower net flows. By contrast, higher returns in the rest of developed markets (excluding the United States) are associated with both higher northbound inflows and lower southbound outflows, yielding a positive association with net flows.

The RMB-USD exchange rate enters with large and statistically significant coefficients: RMB depreciation is associated with weaker northbound inflows and stronger southbound outflows, consistent with the role of exchange-rate movements and currency risk in cross-border portfolio adjustment (Hau and Rey, 2006; Camanho et al., 2022). Geopolitical risk displays a more differentiated pattern. Higher Chinese geopolitical risk is associated with stronger northbound inflows, weaker southbound outflows, and higher net flows, whereas higher U.S. geopolitical risk is mainly linked to weaker northbound inflows and lower net flows. Finally, policy uncertainty and gold returns are included to proxy for broader risk conditions and short-run adjustments in safe-haven demand (Baker et al., 2016; Baur and McDermott, 2010). In this baseline specification with month fixed effects, the China EPU index and gold returns are not precisely estimated, whereas the U.S. EPU index remains statistically significant.

3.2 Robustness Checks

We conduct a series of robustness checks to verify that the baseline results for Chinese Stock Connect flows are not driven by the construction process of our war-related diplomatic sentiment index. First, we examine whether the estimated effects are specific to war-related rhetoric or instead reflect a generic response to other types of diplomatic sentiment (Section 3.2.1). Second, we vary the time-aggregation window used to smooth the Chinese sentiment series and show that the baseline results remain unchanged (Section 3.2.2). Third, we use a residualised VAR framework to test dynamic consistency in the sentiment-flow relation (Section 3.2.3). In addition, supplementary inference checks based on HAC and moving-block bootstrap procedures are reported in Appendix A.3.5.

3.2.1 Replacing War-Related Diplomatic Sentiment Index

A natural concern is that the baseline results might be capturing a generic response to negative diplomatic tone rather than to war-related rhetoric per se. To address this issue, we re-estimate Equation (1) by replacing the war-related sentiment index $WDSI_t^{CN}$ with alternative sentiment measures extracted from the MFA of China press conference corpus. Specifically, we construct (i) an economic-related sentiment index, $EDSI_t^{CN}$, capturing rhetoric related to economic and trade disputes, and (ii) an index of other diplomatic sentiment, $ODSI_t^{CN}$, capturing non-military, non-economic diplomatic issues such as human rights, climate, public health, consular affairs, and cultural or educational exchanges. In addition, we consider a truncated version of the war-related index that focuses exclusively on negative episodes. We set $WDSI_t^{CN,-}$ equal to $WDSI_t^{CN}$ when the daily index is negative and to zero otherwise, and re-estimate Equation (1) using $WDSI_t^{CN,-}$ as the main regressor. This specification allows us to check whether the baseline results are driven by negative-war rhetoric rather than by fluctuations around positive or neutral statements. Table 5 summarises the nine resulting regressions. Columns (1)-(3) use the economic-conflict sentiment index; columns (4)-(6) use the other diplomatic sentiment index; and columns (7)-(9) use the negative-part war index.

[Table 5 about here.]

For the economic-conflict sentiment in columns (1)-(3), the coefficients are statistically insignificant in the inflow regression, positive and statistically significant in the outflow regression, and negative but statistically insignificant in the net-flow regression. The results imply that more negative economic-conflict rhetoric (a decline in $EDSI_t^{CN}$) is associated with reduced net capital outflows, i.e., the opposite pattern we document for the war-related diplomatic sentiment.

A different reversal appears for the other diplomatic sentiment index in columns (4)-(6). Unlike the baseline war-related index, whose inflow coefficient is small and statistically insignificant, the coefficient on $ODSI_t^{CN}$ is negative and statistically significant in both the inflow and net-flow specifications, while the coefficient in the outflow specification is positive but not precisely estimated. Hence, more negative non-military, non-economic diplomatic rhetoric is associated with stronger northbound inflows and higher net flows, indicating that the response pattern differs from the war-related baseline not only in sign but also in the margin through which flows adjust. Overall, the estimates on $EDSI_t^{CN}$ and $ODSI_t^{CN}$ show that cross-border equity flows do not respond mechanically to a generic deterioration in diplomatic tone; rather, once attention shifts away from war-related rhetoric, the signs either reverse or the transmission channel changes.

By contrast, columns (7)-(9), which use the negative-part war index $WDSI_t^{CN,-}$, reproduce the qualitative conclusion of our baseline results. The coefficient on $WDSI_t^{CN,-}$ is statistically insignificant in the inflow specification, and negative (positive) and statistically significant in the outflow (net flow) specification. Since $WDSI_t^{CN,-}$ only varies on days of negative war-related sentiment and takes more negative values when the rhetoric becomes more hawkish, the positive net-flow coefficient implies that negative war-related rhetoric continues to be associated with lower net capital inflows (or larger net outflows), even when we ignore positive war-related statements entirely.

Overall, the replacement exercise results in Table 5 support the interpretation that Chinese Stock Connect flows react specifically to war-related diplomatic sentiment rather than to generic negative or other types of conflict rhetoric.

3.2.2 War-Related Diplomatic Sentiment: Alternative Parametrisations

Our baseline Chinese sentiment index $WDSI_t^{CN}$ is computed as a seven-day moving average of daily war-related diplomatic sentiment. To ensure that the results are not an artefact of this particular smoothing window, we construct $WDSI_t^{CN}$ with different parameterisations, a three-day moving average $WDSI_t^{CN,3}$ and a thirty-day moving average $WDSI_t^{CN,30}$, and re-estimate Equation (1).

Table 6 reports results, where columns (1)-(3) include the three-day average; columns (4)-(6) include the thirty-day average. In both cases, the qualitative conclusions closely mirror those in the baseline regressions. For $WDSI_t^{CN,3}$, the coefficient is small and statistically insignificant in the inflow regression, negative and highly significant in the outflow regression, and positive and significant in the net-flow regression. For $WDSI_t^{CN,30}$, the same configuration arises with somewhat larger magnitudes: the inflow coefficient remains insignificant, while the outflow and net-flow coefficients are -0.649 and 1.017, both significant at the 1% level. In sum, the estimates based on the three-day moving average are slightly smaller than those from the seven-day benchmark, while the coefficients increase substantially when the sentiment index is aggregated over a thirty-day window, indicating that more persistent deteriorations in war-related diplomatic sentiment are associated with larger cross-border capital outflows from China.

[Table 6 about here.]

These findings confirm our baseline conclusion on the effect of Chinese war-related diplomatic sentiment on cross-border equity flows, which is robust to reasonable changes in the temporal aggregation of the sentiment index. Whether we look at relatively short-term (three-day) or more persistent (thirty-day) windows of diplomatic tone, the evidence continues to point to stronger southbound outflows and lower net capital inflows when Chinese war-related rhetoric turns more negative.

3.2.3 VAR-Based Dynamic Consistency Check

To complement the OLS evidence, we estimate residualised bivariate VAR models to test whether the sentiment-flow relation survives dynamic feedback. Following Equation (1), both sentiment and each flow channel are first residualised on the full control block (China/U.S./developed returns, China/U.S. GPR, China/U.S. EPU, gold, FX, and month fixed effects), and VARs are then estimated on residual series.

To keep the setup concise and directly comparable, we report four residualised VAR(1) specifications. Panels A-B use daily data with $WDSI_t^{CN}$ (seven-day moving average) and $WDSI_t^{CN,30}$ (thirty-day moving average), respectively. Panels C-D provide weekly-frequency counterparts constructed from Monday-anchored weeks, where flow variables are summed within week and sentiment/control variables are averaged. Table 7 reports all three flow channels (inflow, outflow, net flow), including VAR coefficients and Granger-causality tests.

The channel pattern is consistent across these specifications. In daily VARs, sentiment significantly predicts outflow under both sentiment definitions, and predicts net flow strongly under $WDSI_t^{CN,30}$. In weekly VARs, the same directional pattern remains: under weekly $WDSI_t^{CN,30}$, the sentiment coefficient is significant for outflow (-0.079, P-Value = 0.040) and net flow (0.105, P-Value = 0.011), while inflow remains insignificant; under weekly

$WDSI_t^{CN}$, outflow remains significant (-0.091, P-Value = 0.016) and net flow is marginally significant (0.080, P-Value = 0.052). Figure 5 reports impulse responses with 95% confidence bands for all four specifications and confirms the same dynamic ordering of channels.

[Table 7 about here.]

[Figure 5 about here.]

3.3 Addressing Endogeneity Concerns

Our baseline specification in Equation (1) treats war-related diplomatic sentiment as an exogenous driver of cross-border equity flows. In practice, however, several endogeneity concerns may arise. First, Stock Connect flows could themselves feed back into the tone of official diplomacy if policymakers react to market conditions, generating reverse causality from capital flows to $WDSI_t^{CN}$. Second, both war-related rhetoric and capital flows may be jointly influenced by common, slowly evolving underlying factors, such as longer-run geopolitical realignments or domestic political cycles, which could induce spurious co-movement even in the absence of a direct causal link. Third, our sentiment measure may still contain residual noise correlated with omitted determinants of capital flows.

To address these concerns, we implement three complementary checks. We begin with placebo regressions that use future sentiment as the dependent variable, and examine whether current flows help predict subsequent war-related rhetoric. We then estimate dynamic specifications with lagged flows and controls to absorb serial correlation and slow-moving co-movements. Finally, we employ an instrumental-variable (IV) strategy that exploits cross-spokesperson variation in average diplomatic tone as a source of exogenous shifts in the sentiment index.

3.3.1 Placebo Tests and Detrended Sentiment

We first examine whether Stock Connect flows forecast future war-related diplomatic sentiment. Specifically, we run placebo regressions in which the dependent variable is the one-day-ahead sentiment index $WDSI^{CN}_{t+1}$, and the key regressors are contemporaneous and lagged flows:

$$WDSI^{CN}_{t+1} = \alpha + \rho_0 Flow^{CN}_{k,t} + \rho_1 Flow^{CN}_{k,t-1} + \boldsymbol{\gamma}'\mathbf{R}_t + \boldsymbol{\delta}'\mathbf{Z}_t + \mu_{m(t)} + u_{k,t}, \tag{2}$$

for $k \in \{\text{Inflow}, \text{Outflow}, \text{Net flow}\}$, where $\mathbf{R}_t$ and $\mathbf{Z}_t$ denote the same sets of control variables as in the baseline regressions, and $u_{k,t}$ is an error term. If capital flows systematically anticipate changes in war-related diplomacy, we would expect significant coefficients on $Flow^{CN}_{k,t}$ and/or $Flow^{CN}_{k,t-1}$.

Panel A of Table 8 reports the results, where the coefficients on contemporaneous and lagged capital flows are statistically insignificant across all three specifications (columns 1-3). As such, we find no economically or statistically meaningful predictive power of capital flows for next-day war-related sentiment. Therefore, the placebo test mitigates concerns that our baseline results are mainly driven by reverse causality from Stock Connect activity to the tone of Chinese diplomatic rhetoric.

[Table 8 about here.]

We also investigate whether slow-moving trends in war-related diplomacy might spuriously correlate with capital flows. To this end, we construct a detrended sentiment index $WDSI^{CN}_t$ by regressing the raw series on a linear time trend and obtaining the residuals; we then re-estimate Equation (1) with $WDSI^{CN}_t$ replacing the raw index. Panel B of Table 8 reports the results. For inflows, the coefficient on detrended sentiment remains close to zero and statistically insignificant. On the other hand, the coefficients for outflows and net flows remain large, negative (for outflows) and positive (for net flows) respectively, and statistically significant. Therefore, these results, both in magnitude and statistical significance terms, are in line with the baseline regressions, suggesting that the main results are not driven by common linear trends.

3.3.2 Lag Specification Tests

Next, we allow for a richer dynamic specification in capital flows and macro-financial controls. If flows and war-related diplomatic sentiment are jointly influenced by persistent underlying factors, omitting lagged dependent variables and lagged controls could bias the estimated contemporaneous effect of $WDSI_t^{CN}$. We therefore estimate augmented versions of Equation (1) that include one or two lags of the dependent variable, together with lagged control variables:

$$
\begin{aligned}
Flow_{k,t}^{CN} = \alpha_k + \beta_k WDSI_t^{CN} + \phi_{1,k} Flow_{k,t-1}^{CN} + \phi_{2,k} Flow_{k,t-2}^{CN} + \boldsymbol{\gamma}_{0,k}' \mathbf{R}_t + \boldsymbol{\gamma}_{1,k}' \mathbf{R}_{t-1} \\
+ \boldsymbol{\gamma}_{2,k}' \mathbf{R}_{t-2} + \boldsymbol{\delta}_{0,k}' \mathbf{Z}_t + \boldsymbol{\delta}_{1,k}' \mathbf{Z}_{t-1} + \boldsymbol{\delta}_{2,k}' \mathbf{Z}_{t-2} + \mu_{m(t)} + \varepsilon_{k,t},
\end{aligned}
\tag{3}
$$

where the first-order specification sets $\phi_{2,k} = 0$, $\boldsymbol{\gamma}_{2,k} = \mathbf{0}$, and $\boldsymbol{\delta}_{2,k} = \mathbf{0}$ (i.e., omits second lags), while the second-order specification estimates these parameters freely.

Panel A of Table 9 reports the results of the first-order specification. The coefficient on $WDSI_t^{CN}$ is again small and insignificant for inflows, but remains negative and highly significant for outflows and positive and significant for net flows. Panel B of Table 9 reports the results of the second-order specification. The estimated coefficients on $WDSI_t^{CN}$ are still insignificant for inflows, statistically significant and negative for outflows, and statistically significant and positive for net flows. In all specifications, capital flow dynamics are strongly persistent, as evidenced by the highly significant lagged coefficients, and the R^2 values increase relative to the baseline regressions. However, the association between war-related diplomatic sentiment and capital flows remains unchanged, indicating that our baseline results are robust to a broad class of dynamic specifications.

[Table 9 about here.]

3.3.3 Instrumental Variable Strategy

Finally, we implement an instrumental variable (IV) strategy to address potential omitted-variable bias or measurement error in the war-related sentiment index. Our approach is motivated by a growing literature showing that political and policy elites exhibit persistent, individually specific communication styles that are only imperfectly related to underlying fundamentals. For example, Hansen et al. (2018) use computational linguistics to document stable heterogeneity in the language used by the Federal Open Market Committee (FOMC) members, while Dietrich et al. (2019) show that the emotional intensity of U.S. congressional speeches varies systematically across legislators. A related strand of work in political science and social psychology finds that gender and other personal characteristics are associated with systematic differences in rhetorical style, including the use of emotional or negative language (Carli, 1990; Haselmayer et al., 2022; Hargrave and Blumenau, 2022). These studies provide external validation for our key assumption that Ministry of Foreign Affairs spokespeople differ in their average "tone" in a way that is persistent over time.

Building on this evidence, we exploit cross-spokesperson differences in communication style as a source of exogenous variation in war-related diplomatic sentiment. Specifically, for each spokesperson we compute the average value of the Chinese sentiment index across all answers they provide, regardless of topic, and use this spokesperson-specific mean, denoted "Spokesperson fixed sentiment", as a proxy for their inherent communication style. To align this cross-sectional measure with the daily time series, we then construct a seven-day moving average of the spokesperson fixed-sentiment variable based on the spokespersons who appear on each day. Intuitively, days on which an inherently more negative (or more positive) spokesperson happens to brief the press are associated with predictably lower (higher) values of $WDSI_t^{CN}$, even conditional on the same underlying news environment.

The exclusion restriction is that these stylistic differences across spokespeople affect Stock Connect flows only through their impact on war-related diplomatic sentiment, not through any direct effect on fundamentals. This assumption is plausible in our setting. Individual spokespeople have no formal authority over geopolitical decisions, macroeconomic policies, or capital-account regulation, and their assignment to particular briefings is largely administrative. Conditional on the content of the underlying events (which enters through controls and global risk proxies), it is difficult to think of a channel through which investors would respond directly to who is delivering the message rather than to how war-related rhetoric is framed[10]. Under this exclusion restriction, the spokesperson fixed-sentiment measure is a valid instrument for $WDSI_t^{CN}$ in the Stock Connect regressions.

Table 10 reports the IV strategy results. Column (1) presents the first-stage regression of $WDSI_t^{CN}$ on the spokesperson fixed-sentiment instrument and the full set of controls. The coefficient on the instrument is 1.000 with a robust standard error of 0.089, implying a t-statistic of about 11.25 and confirming strong relevance. Across the three second-stage samples, the excluded-instrument first-stage Wald χ^2 statistics range from 124.884 to 128.741. The Kleibergen–Paap rk Wald F statistics range from 123.629 to 127.468, and the rk LM statistics reject under-identification at P-Value < 0.001 in all cases. Columns (2)-(4) report the second-stage 2SLS estimates for inflows, outflows, and net flows, respectively. The IV coefficient on $WDSI_t^{CN}$ is positive but statistically insignificant for inflows, negative and strongly significant for outflows, and positive and strongly significant for net flows. Relative to the baseline OLS estimates, the IV coefficients retain the same sign but are substantially larger in magnitude, especially in the outflow and net-flow regressions, a pattern consistent with attenuation bias from classical measurement error in $WDSI_t^{CN}$ biasing the OLS estimates towards zero.

[Table 10 about here.]

Taken together with the placebo and detrended sentiment tests, and dynamic specifications, these IV estimates provide consistent evidence that our main findings are not driven by reverse causality, common trends, or omitted-variable bias in the war-related sentiment index. The effect of Chinese war-related diplomatic sentiment on Stock Connect flows is consistent with a causal interpretation: more negative rhetoric is systematically associated

[10]This identification logic is closely related to the use of speaker-specific language measures in the FOMC and congressional-speech literatures, where individual communication styles are treated as a source of exogenous variation in the tone of public communication (Hansen et al., 2018; Hargrave and Blumenau, 2022).

with stronger southbound outflows and larger net capital outflows from the mainland.

4 Additional Analysis

4.1 Evidence from the U.S., U.K., Japan, and South Korea

We examine whether war-related diplomatic sentiment in the U.S., U.K., Japan, and South Korea predicts Chinese Stock Connect flows. This cross-country spillover perspective is also related to evidence that foreign investors' attention to emerging markets varies with distraction from risk factors in their home markets (Marmora, 2023). Following the daily specification in Equation (1), we regress Chinese northbound inflow, southbound outflow, and net flow on the seven-day moving average of each foreign sentiment index, with the same control block and month fixed effects.

Table 11 reports full-sample and split-sample evidence. In the full sample, the most stable common pattern is on the southbound margin: coefficients on foreign sentiment are uniformly negative in the outflow equation and statistically significant for all four countries, although the South Korean estimate is only marginally significant relative to the more precisely estimated U.S., U.K., and Japanese coefficients. This complements the baseline China result in Section 3.1 and the relative-price evidence based on the A-H share premium developed below in Section 4.2, where heightened war-related risk is associated with reallocation through the southbound channel. Once the sample is split at the start of 2022, the transmission channel shifts. Through 2021, Korean sentiment is the dominant predictor across Chinese flow equations; from 2022 onward, U.S. sentiment becomes strongly associated with both northbound inflow and net flow, while U.K. and Japanese sentiment also load on the inflow and net-flow specifications. This post-2021 re-weighting is consistent with the paper's core narrative that cross-country sentiment transmission into Chinese equity flows intensified after the Russia-Ukraine war and became more visible through the foreign-investor-facing northbound channel.

[Table 11 about here.]

We then examine each country's own international equity position response at the monthly frequency. Because comparable monthly own-flow series are not available for the U.K. in the current sample, Table 12 focuses on the U.S., Japan, and South Korea. To preserve strict cross-country comparability, all three country-specific regressions are estimated on the common monthly window for which the own-position variables and country-specific controls are jointly available for all three countries, namely January 2015 to June 2025, yielding 126 monthly observations. This common window begins when the Japanese foreign equity flow series becomes available after monthly aggregation and ends with the last month covered by the U.S. monthly international equity position series. The U.S. coefficients remain strongly negative across inflow, outflow, and net-flow equations, reinforcing the safe-haven asymmetry highlighted above; Japanese estimates are statistically weak, while Korean estimates are positive in selected specifications.

[Table 12 about here.]

Figure 6 provides a compact directional summary of the same cross-country monthly system. Three patterns stand out. First, Korean sentiment emerges as the most connected outward transmitter in this four-country network: in Figure A it is positively associated with both Chinese and U.S. inflows, and in Figure C it is positively associated with Chinese net flow. Second, negative links dominate the outflow and net-flow dimensions. In Figure B, most significant edges are negative (including the strong U.S.-Korea and Japan-Korea links), with only one positive edge from Korea to U.S. outflow. In Figure C, the structure is sparse and asymmetric: Chinese sentiment loads negatively on U.S. and Japanese net-flow equations, while Korean sentiment loads positively on China's net-flow equation. Third, the network view complements Tables 11 and 12 by making directional asymmetry visually transparent: statistically significant transmission is concentrated in a few bilateral channels rather than being diffuse across all country pairs.

[Figure 6 about here.]

4.2 Relative-Price Evidence: The A-H Share Premium

Inspired by He et al. (2023), a potential relative-price channel through which war-related diplomatic sentiment may affect cross-border equity flows is the A-H share premium[11]. When geopolitical and war-related risks increase, international investors may reduce their exposure to Hong Kong-listed Chinese firms, putting downward pressure on H-share prices. If mainland A-shares remain relatively supported by domestic investors, the A-H premium widens, creating an arbitrage opportunity that induces mainland investors to buy cheaper H-shares via the southbound Stock Connect channel. This relative-price evidence helps explain why the Chinese war-related diplomatic sentiment effect is strongest in southbound reallocation. It also provides a bridge to the unobservable behaviour of international investors in the Hong Kong market: even if we cannot directly observe their trades, their reaction to war-related rhetoric can be inferred from movements in the A-H premium and the induced response of mainland investors.

To examine this relative-price channel formally, we regress the daily A-H premium on the Chinese war-related diplomatic sentiment index $WDSI^{CN}$, controlling for Chinese, U.S., and developed-market equity returns, Chinese and U.S. geopolitical risk, Chinese and U.S. economic policy uncertainty, gold returns, the RMB-USD exchange rate, and month fixed effects. Column (1) of Table 13 shows that the coefficient on $WDSI^{CN}$ is negative and highly significant; a one-unit increase in $WDSI^{CN}$ is associated with a reduction of about 1.43 percentage points in the A-H premium. Since lower values of $WDSI^{CN}$ correspond to more negative war-related rhetoric, this implies that heightened war risk (a decline in $WDSI^{CN}$) significantly raises the A-H premium, making H-shares cheaper relative to their A-share counterparts.

[11] The A-H premium is defined as the percentage price difference between mainland A-shares and their corresponding H-shares listed in Hong Kong (with A-share prices converted into Hong Kong dollars). A higher value indicates that A-shares trade at a larger premium over H-shares.

[Table 13 about here.]

The A-H premium is highly persistent, so columns (2) and (3) augment the specification with one and two lags of the premium. In column (3), the two lag coefficients sum to 0.994, indicating near-unit-root persistence in the level series. Untabulated ADF and KPSS tests are consistent with nonstationary level behavior for the A-H premium. For the dynamic specification, however, the residual from column (3) strongly rejects a unit root in the ADF test and does not reject stationarity in the KPSS test at the 5% level, and the negative coefficient on $WDSI^{CN}$ remains statistically significant at the 10% level under Newey–West HAC inference. Column (4) shows that this pattern also survives in first differences: when the dependent variable is the daily change in the A-H premium, the coefficient on $WDSI^{CN}$ remains negative at -0.071 and statistically significant at the 10% level. These checks alleviate the concern that the result is driven by non-stationary residuals rather than by a genuine shift in relative pricing. In other words, negative shocks to Chinese war-related diplomatic sentiment shift the level of the A-H premium over and above its own past history. This pattern is precisely what we would expect if war-related rhetoric triggers a repricing between mainland and Hong Kong markets that opens up arbitrage opportunities for mainland investors.

Importantly, this relative-price evidence also alleviates concerns about reverse causality from capital flows to the A-H premium. If capital flows were the primary driver of the premium, war-related risk would probably lead to capital outflows from the mainland and an A-share discount; instead, we find that war-related risk is associated with a higher A-H premium. A more plausible interpretation is that changes in war-related diplomatic sentiment move the A-H premium, which in turn encourages additional southbound investment into cheaper H-shares, linking our earlier capital-flow results to a concrete relative-price channel.

To further verify directionality, we run supplementary daily Granger-causality tests between the A-H premium and Stock Connect flows in a residualised VAR setup (same control block and month fixed effects as the baseline). The lag-1/3/5 results reported in Appendix

Table A.4 show that the A-H premium significantly predicts southbound outflow (A-H premium $\rightarrow$ outflow), while reverse-direction evidence is weak for outflow, supporting the view that the premium is an active transmission channel for flow reallocation.

While the A-H premium captures the relative-valuation wedge most directly linked to southbound reallocation, it is also useful to examine whether Chinese war-related sentiment is associated with broader equity-market repricing. We turn to this evidence next.

4.3 Complementary Return Evidence: Beyond an Absolute-Price-Decline Channel

Our main empirical object is cross-border equity flows rather than stock returns. As complementary price-side evidence, we examine whether the same war-related diplomatic sentiment shifts are also associated with repricing in broad Chinese equity-index returns. Using the same control block as Equation (1), we estimate three return specifications: (i) weekly baseline sentiment $WDSI_{t-1}^{CN}$ based on the seven-day moving average, (ii) weekly alternative sentiment $WDSI_{t-1}^{CN,30}$ based on the thirty-day moving average, and (iii) monthly contemporaneous sentiment $WDSI_{t}^{CN,30}$.

Table 14 shows that the return evidence does not support a simple absolute-price channel in which the flow effects are driven merely by falling stock prices. While the daily return response is not robust, the weekly coefficients are uniformly negative across indices but not always precisely estimated, and the negative relation becomes systematic and more precisely estimated at the monthly horizon. Because a lower $WDSI$ indicates more adverse war-related diplomatic sentiment, the uniformly negative coefficients imply the opposite pattern: worsening sentiment is associated with higher, rather than lower, Chinese equity-index returns. This inverse relation is already visible at the weekly horizon in both Panel A and Panel B, and becomes stronger and more precisely estimated at the monthly horizon in Panel C. Figure 7 conveys the same result graphically: the estimated coefficients lie systematically on the negative side, with the monthly estimates farthest from zero. Taken together with the baseline flow results, these findings indicate that the capital-flow response is not a mechanical reaction to broad market declines, but is more consistent with portfolio reallocation and relative-pricing adjustment.

[Table 14 about here.]

[Figure 7 about here.]

4.4 Market Activity and Risk Absorption

The preceding analysis has focused on the direction and magnitude of cross-border capital flows. However, war-related diplomatic sentiment may also affect the trading intensity and the way risk is absorbed by financial markets. Market activity – in particular trading volume and return volatility – is a complementary dimension of geopolitical risk transmission: war-related news can trigger portfolio rebalancing, hedging, or speculative trading even when aggregate flows remain relatively modest (Bollerslev et al., 2018). To capture this, we examine how daily war-related diplomatic sentiment indices for China and the U.S. are associated with trading volumes and realized volatility in the respective equity markets.

Panel A of Table 15 reports regressions of the log of daily trading volume on war-related diplomatic sentiment. Columns (1) and (2) show that the coefficient on the Chinese war-related sentiment index $WDSI^{CN}$ is negative and highly significant for both the CSI 300 and CSI 1000. A one-unit decrease in $WDSI^{CN}$, corresponding to more negative Chinese war-related diplomatic rhetoric, is associated with approximately a 3.3% increase in CSI 300 trading volume and a 9.3% increase in CSI 1000 trading volume. Consistent with this pattern, more negative Chinese war-related sentiment is accompanied by a pronounced rise in onshore trading activity. Columns (3) and (4) use the U.S. war-related sentiment index $WDSI^{US}$ and show a similar pattern for U.S. markets. A one-unit decrease in $WDSI^{US}$ (more negative war-related rhetoric) is associated with a 14.4% and 11.3% increase in trading volume on Nasdaq and the Dow Jones Industrial Average, respectively. In all cases, the sign of the effect is consistent with the idea that war-related diplomatic tensions induce more active trading and portfolio rebalancing on both sides of the Pacific.

[Table 15 about here.]

Panel B of Table 15 examines realized volatility. For the U.S. market (columns (7)-(8)), the coefficients on $WDSI^{US}$ are negative and highly significant, i.e., more negative U.S. war-related sentiment (a decline in $WDSI^{US}$) is associated with higher equity return volatility on both Nasdaq and the Dow Jones Industrial Average. In contrast, the $WDSI^{CN}$ coefficients in the Chinese volatility regressions (columns (5)-(6)) are positive and statistically significant. This implies that a deterioration in Chinese war-related diplomatic sentiment is associated with lower realized volatility on CSI 300 and CSI 1000, even as trading volumes increase. Taken together, these results suggest that adverse war-related rhetoric is reflected in both higher turnover and increased price volatility in U.S. markets. On the other hand, in the onshore Chinese market, adverse war-related rhetoric is associated with heavier trading but somewhat dampened daily volatility, possibly reflecting a combination of risk-absorbing flows and policy-driven volatility smoothing. Overall, war-related diplomatic sentiment influences not only cross-border capital flows but also the microstructure of equity markets.

5 Conclusion

This paper develops a high-frequency war-related diplomatic sentiment index from official foreign-policy communications and shows that diplomatic language contains market-relevant information beyond standard geopolitical risk measures. Relative to newspaper-based geopolitical risk indices, the WDSI moves earlier in time and provides a more forward-looking signal of geopolitical deterioration. Empirically, more negative Chinese war-related diplomatic sentiment predicts larger southbound purchases of Hong Kong equities and lower net Stock Connect flows, while more negative U.S. war-related sentiment is associated with stronger foreign demand for U.S. equities, consistent with the safe-haven role of U.S. assets. We also document spillovers from the United Kingdom, Japan, and South Korea to Chinese equity flows, together with complementary evidence for returns and market activity. Overall, the results show that official diplomatic tone is not merely symbolic: it helps predict portfolio reallocation and broader capital-market adjustment.

These findings have direct implications for both investors and policymakers. For investors, because the WDSI leads conventional geopolitical risk measures and predicts subsequent market movements, diplomatic sentiment should be incorporated into portfolio construction, risk management, and tactical asset allocation. For policymakers, official diplomatic communication is itself a market-relevant signal. Governments should therefore take into account the potential stock-market consequences of diplomatic statements, especially during periods of elevated geopolitical tension. The cross-country evidence further suggests that similar shifts in diplomatic tone can generate different capital-market responses across countries. In our setting, more negative war-related rhetoric is associated with stronger foreign demand for U.S. equities but with larger southbound purchases and weaker net flows in China, underscoring that the design and timing of official communication should be evaluated in light of each country's role in global capital markets and the channels through which investors reallocate risk.

References

Bailey, Michael A., Anton Strezhnev, and Erik Voeten (2017). Estimating dynamic state preferences from united nations voting data. *International Organization 71*(1), 1–31.

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

Baturo, Alexander, Niheer Dasandi, and Slava J. Mikhaylov (2017). Understanding state preferences with text as data: Introducing the un general debate corpus. *Research & Politics 4*(2), 1–9.

Baum, Matthew A. and Yuri M. Zhukov (2019). Media ownership and news coverage of international conflict. *Political Communication 36*(1), 1–22.

Baur, Dirk G. and Thomas K. McDermott (2010). Is gold a safe haven? international evidence. *Journal of Banking & Finance 34*(8), 1886–1898.

Benhima, Kenza and Romain Cordonier (2022). News, sentiment and capital flows. *Journal of International Economics 137*, 103621.

Bohn, Henning and Linda L. Tesar (1996). U.s. equity investment in foreign markets: Portfolio rebalancing or return chasing? *American Economic Review 86*(2), 77–81.

Bollerslev, Tim, Jia Li, and Yuan Xue (2018). Volume, volatility, and public news announcements. *Review of Economic Studies 85*(4), 2005–2041.

Brennan, Michael J. and Henry H. Cao (1997). International portfolio investment flows. *Journal of Finance 52*(5), 1851–1880.

Buehlmaier, Christoph and Toni M. Whited (2018). Are financial constraints priced? evidence from textual analysis. *Review of Financial Studies 31*(7), 2693–2728.

Caballero, Ricardo J., Emmanuel Farhi, and Pierre-Olivier Gourinchas (2017). The safe assets shortage conundrum. *Journal of Economic Perspectives 31*(3), 29–46.

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

Camanho, Nelson, Harald Hau, and Hélène Rey (2022). Global portfolio rebalancing and exchange rates. *Review of Financial Studies 35*(11), 5228–5274.

Cao, Chunfang, Xiaoyang Li, and Guilin Liu (2019). Political uncertainty and cross-border acquisitions. *Review of Finance 23*(2), 439–470.

Caporale, Guglielmo Maria, Faek Menla Ali, and Nicola Spagnolo (2015). Exchange rate uncertainty and international portfolio flows: A multivariate garch-in-mean approach. *Journal of International Money and Finance 54*, 70–92.

Carli, Linda L. (1990). Gender, language, and influence. *Journal of Personality and Social Psychology 59*(5), 941–951.

Choi, Seungjin and Jan Havel (2025). Geopolitical risk and U.S. foreign portfolio investment: A tale of advanced and emerging markets. *Journal of International Money and Finance 151*, 103253.

Correa, Ricardo, Keshav Garud, Juan M. Londono, and Nathan Mislang (2021). Sentiment in central banks' financial stability reports. *Review of Finance 25*(1), 85–120.

Dietrich, Bryce J., Thomas J. Hayes, and Diana Z. O'Brien (2019). Pitch perfect: Vocal pitch and the emotional intensity of congressional speech. *American Political Science Review 113*(4), 941–962.

Feng, Chen, Lei Han, Samuel A. Vigne, and Yanan Xu (2023). Geopolitical risk and the dynamics of international capital flows. *Journal of International Financial Markets, Institutions & Money 82*, 101693.

Froot, Kenneth A., Paul G. J. O'Connell, and Mark S. Seasholes (2001). The portfolio flows of international investors. *Journal of Financial Economics 59*(2), 151–193.

Fuchs, Andreas and Nils-Hendrik Klann (2013). Paying a visit: The Dalai Lama effect on international trade. *Journal of Development Economics 101*, 55–63.

Gartzke, Erik (2006). The affinity of nations: Similarity of state voting positions in the UN general assembly, 1946–2005. Dataset, Version 4.0. University of California, San Diego.

Gelman, Maria, Axel Jochem, Stefan Reitz, and Mark P. Taylor (2015). Real financial market exchange rates and capital flows. *Journal of International Money and Finance 54*, 50–69.

Gentzkow, Matthew, Bryan Kelly, and Matt Taddy (2019). Text as data. *Journal of Economic Literature 57*(3), 535–574.

Gowa, Joanne (1994). *Allies, Adversaries, and International Trade*. Princeton, NJ: Princeton University Press.

Griffin, John M., Federico Nardari, and René M. Stulz (2004). Are daily cross-border equity flows pushed or pulled? *Review of Economics and Statistics 86*(3), 641–657.

Hansen, Stephen, Michael McMahon, and Andrea Prat (2018). Transparency and deliberation within the FOMC: A computational linguistics approach. *Quarterly Journal of Economics 133*(2), 801–870.

Hargrave, Eleanor and Jack Blumenau (2022). No longer conforming to stereotypes? gender, political style and parliamentary debate in the UK. *British Journal of Political Science 52*(4), 1584–1603.

Haselmayer, Martin, Sarah C. Dingler, and Marcelo Jenny (2022). How women shape negativity in parliamentary speeches—a sentiment analysis of debates in the Austrian parliament. *Parliamentary Affairs 75*(4), 867–886.

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

Hau, Harald and Hélène Rey (2006). Exchange rates, equity prices, and capital flows. *Review of Financial Studies 19*(1), 273–317.

He, Zhiguo, Yao Wang, and Xiao Zhu (2023). The stock connect to china. In *AEA Papers and Proceedings*, Volume 113, Nashville, TN, pp. 125–130. American Economic Association.

Hufbauer, Gary Clyde, Jeffrey J. Schott, Kimberly Ann Elliott, and Barbara Oegg (2007). *Economic Sanctions Reconsidered* (3 ed.). Washington, DC: Peterson Institute for International Economics.

Julio, Brandon and Youngsuk Yook (2012). Political uncertainty and corporate investment cycles. *Journal of Finance 67*(1), 45–83.

Kelly, Bryan, Lubos Pastor, and Pietro Veronesi (2016). The price of political uncertainty: Theory and evidence from the option market. *Journal of Finance 71*(5), 2417–2480.

Kommandeur, Jesse and Rens van Dam (2024, July). Gina diplomatic: Methodological notes. Technical report, The Hague Centre for Strategic Studies (HCSS). HCSS Datalab. Created/updated July 2024.

Krishnamurthy, Arvind and Annette Vissing-Jorgensen (2012). The aggregate demand for treasury debt. *Journal of Political Economy 120*(2), 233–267.

Lundblad, Christian T., Donghui Shi, Xiaoyan Zhang, and Zijian Zhang (2026). Foreign capital in the Chinese stock market: A firm-level study. *Journal of Financial and Quantitative Analysis 61*(2), 799–840.

Maggiori, Matteo, Brent Neiman, and Jesse Schreger (2020). International currencies and capital allocation. *Journal of Political Economy 128*(6), 2019–2066.

Maoz, Zeev (2010). *Networks of Nations: The Evolution, Structure, and Impact of International Networks, 1816–2001.* Cambridge: Cambridge University Press.

Marmora, Paul (2023). Identifying the effects of macroeconomic attention through foreign investor distraction. *Journal of Financial and Quantitative Analysis 58*(8), 3644–3671.

Mochtak, Michal and Richard Q. Turcsanyi (2021). Studying chinese foreign policy narratives: Introducing the ministry of foreign affairs press conferences corpus. *Journal of Chinese Political Science 26*(4), 743–761.

Papaioannou, Elias (2009). What drives international bank flows? politics, institutions and other determinants. *Journal of Development Economics 88*(2), 269–281.

Pástor, Lubos and Pietro Veronesi (2012). Uncertainty about government policy and stock prices. *Journal of Finance 67*(4), 1219–1264.

Permutable AI (2024). War index: What wars are happening now 2025 update. Web page (shows original date Jul 22, 2024; later updates shown on page). Accessed 2025-12-29.

Portes, Richard and Hélène Rey (2005). The determinants of cross-border equity flows. *Journal of International Economics 65*(2), 269–296.

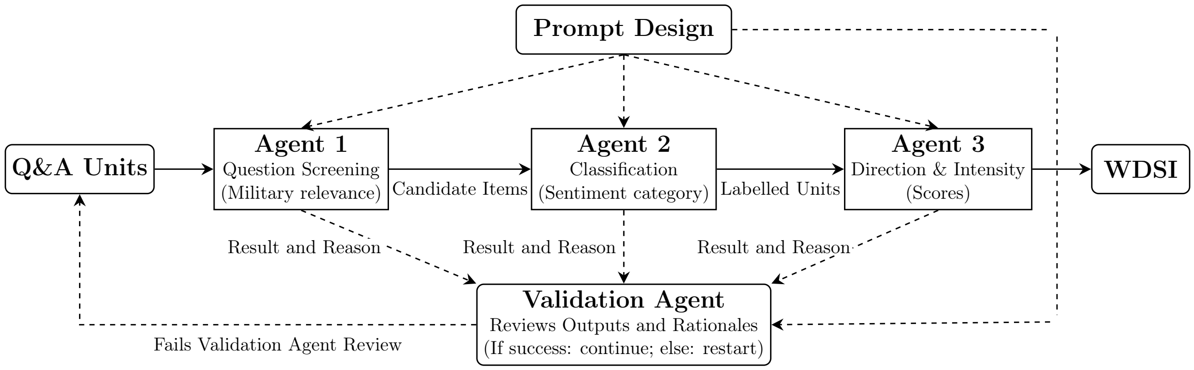

Figure 1: Process of constructing the War-Related Diplomatic Sentiment Index.

Note: The workflow consists of three sequential LLM stages (war-related screening, sentiment classification, and intensity scoring), followed by a validation agent. Solid arrows indicate the primary processing sequence, while dashed arrows transmit intermediate outputs and rationales to the validator. If validation fails, the item is routed back to the input stage for re-evaluation.

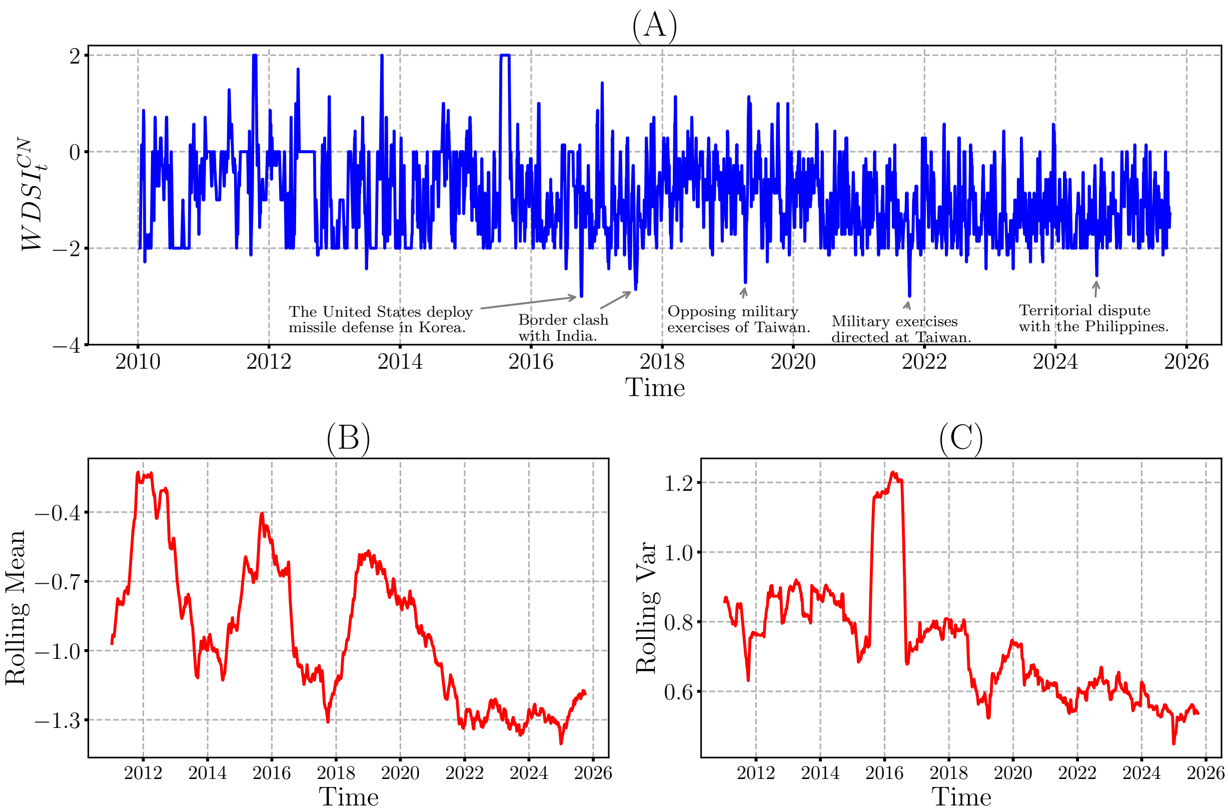

Figure 2: Chinese War-Related Diplomatic Sentiment Index (WDSI^{CN}): Daily Series, Rolling Mean, and Rolling Volatility

Panel A plots the daily Chinese WDSI over the sample period (2010-2025), constructed as a seven-day rolling average of the raw per-conference sentiment scores on the [-3, 3] scale. Several major events in the military and security domain (for example the deployment of U.S. missile defence in Korea, border clashes with India, large-scale military exercises around Taiwan and maritime disputes in the South China Sea) are annotated along the timeline. Panel B reports the one-year rolling mean of the WDSI, while Panel C shows the corresponding one-year rolling variance. The decline in volatility and the gradual downward drift of the rolling mean indicate that China's official diplomatic language has become progressively more negative on military and security issues over time.

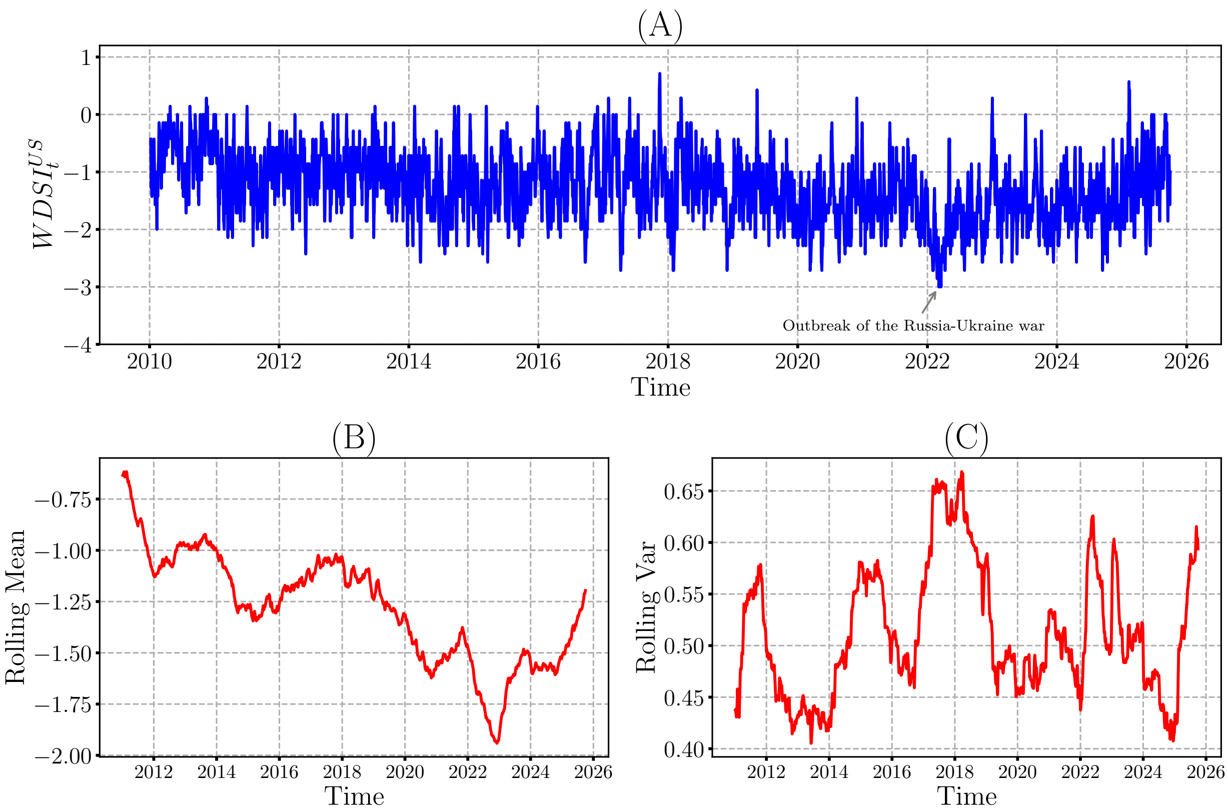

Figure 3: U.S. War-Related Diplomatic Sentiment Index (WDSI^{US}): Daily Series, Rolling Mean, and Rolling Volatility

Panel A plots the daily U.S. WDSI over the sample period (2010-2025), constructed as a seven-day rolling average of the raw sentiment scores on the [-3, 3] scale derived from Department of State press statements. Several major security events, such as the outbreak of the Russia-Ukraine war, are annotated along the timeline. Panel B reports the one-year rolling mean of the U.S. WDSI, which fluctuates around a moderately negative level and becomes more negative during episodes of heightened military involvement and geopolitical tension. Panel C shows the corresponding one-year rolling variance. Volatility is clearly time-varying, with pronounced spikes around periods of sustained conflict and crisis, indicating that the intensity of U.S. war-related diplomatic sentiment is clustered in time rather than evenly distributed over the sample.

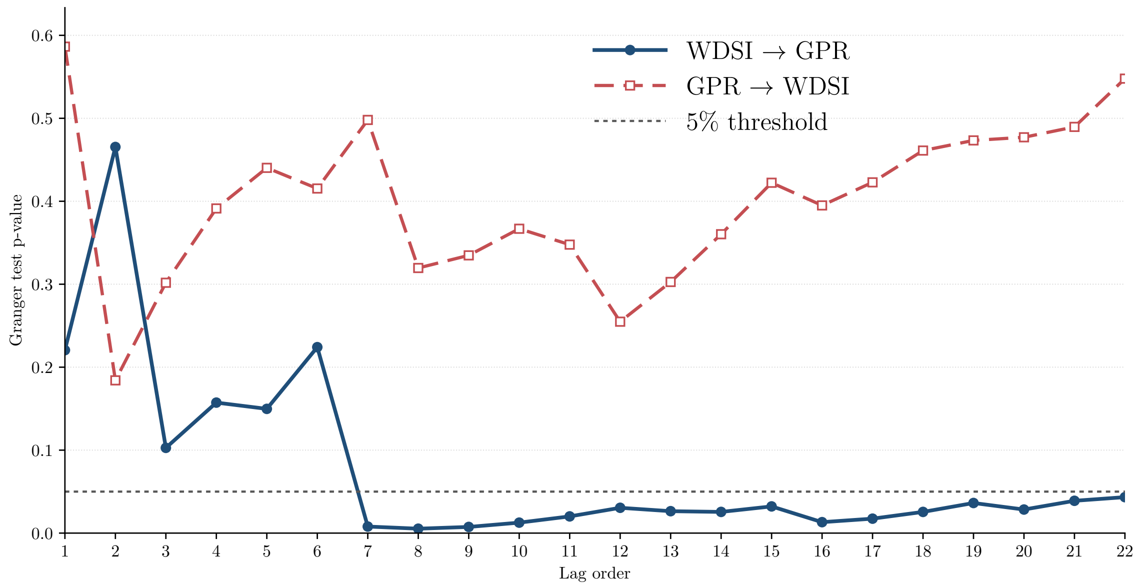

Figure 4: Directional Granger Predictability Across the Full VAR Lag Grid

The figure plots Granger-causality test P-Values from the bivariate daily VAR models in Table 2 for lag orders $p = 1, \ldots, 22$. The underlying system is a raw-level (non-residualised) VAR on $(WDSI_t^{CN,30}, GPR_t^{CN})$ estimated on the full daily sample (2010-2025), without additional controls inside the VAR. The black line corresponds to the null "WDSI does not predict GPR", and the blue line corresponds to the reverse-direction null "GPR does not predict WDSI". The dashed horizontal line marks the 5% significance threshold. The WDSI→GPR line falls below 5% from lag 7 onwards, while reverse-direction P-Values remain above conventional significance thresholds throughout.

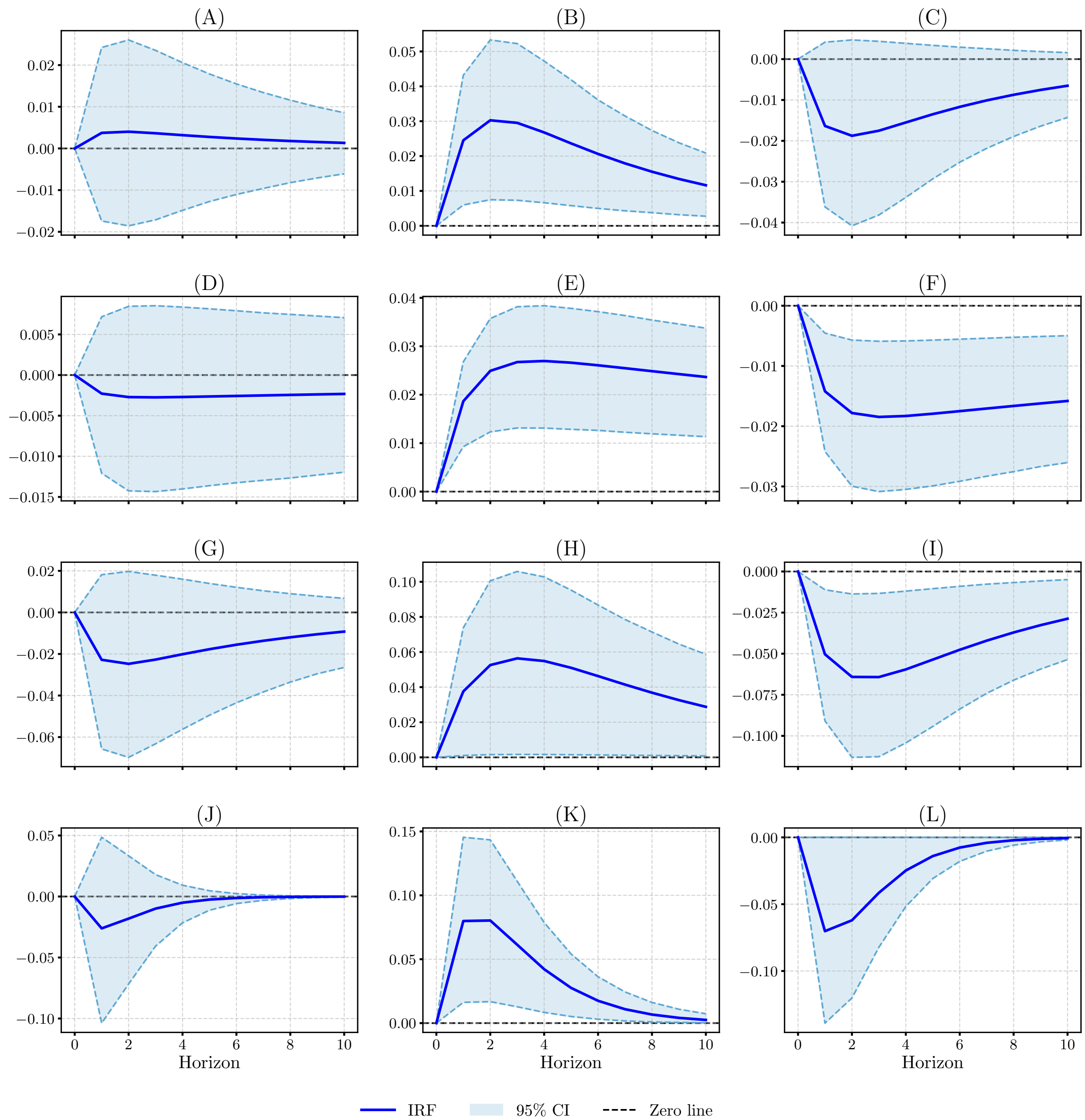

Figure 5: Impulse Responses to a Negative Sentiment Shock: VAR(1), Daily and Weekly Specifications

The figure reports orthogonalised impulse responses for inflow, outflow, and net flow under four residualised VAR(1) specifications. Panels A-C use daily baseline $WDSI_t^{CN}$ (seven-day moving average). Panels D-F use daily $WDSI_t^{CN,30}$ (thirty-day moving average). Panels G-I use weekly $WDSI_t^{CN,30}$, and Panels J-L use weekly $WDSI_t^{CN}$. Weekly observations are built from Monday-anchored weeks, with flow variables summed and sentiment/control variables averaged within week. Shaded areas and dashed bounds indicate 95% confidence intervals, and the dashed horizontal line marks zero. Responses are for a one-standard-deviation negative sentiment shock (a decline in the sentiment index).

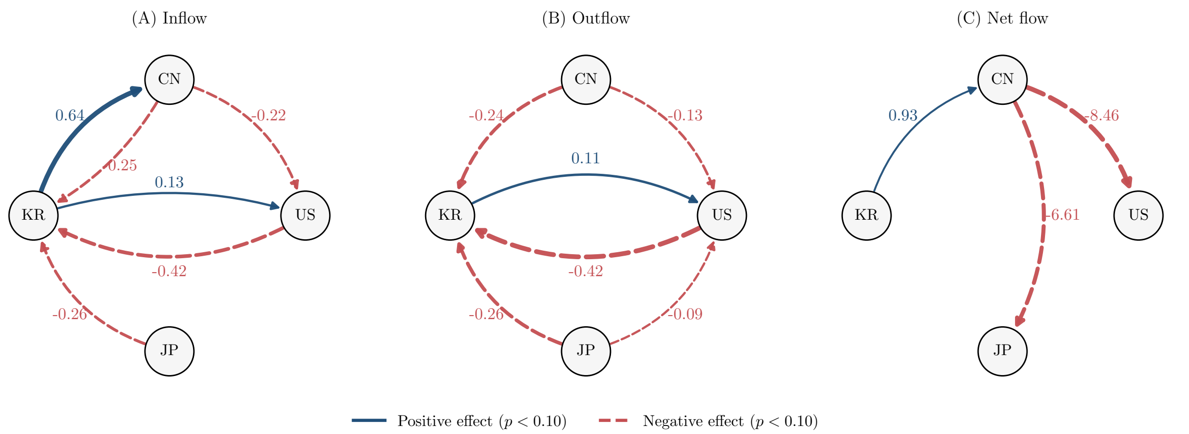

Figure 6: Directional Cross-Country Sentiment-Flow Network (CN, US, JP, KR)

This figure shows directional cross-country links from war-related diplomatic sentiment to equity flow outcomes among China (CN), the United States (US), Japan (JP), and South Korea (KR). Panels A-C report results for monthly Inflow, Outflow, and Net flow, respectively. For each target country and each flow outcome, we estimate a monthly regression where the dependent variable is the target-country flow measure and the key regressors are the four countries' sentiment series; the control variables and fixed-effects structure are the same as in the main baseline specification, and standard errors are heteroskedasticity-robust (HC1). A directed arrow from country j to country i is plotted only when the corresponding sentiment coefficient is significant at the 10% level (P-Value < 0.10); self-links are omitted. Blue solid arrows denote positive coefficients and red dashed arrows denote negative coefficients, and the numeric labels on arrows are coefficient estimates. Arrow thickness scales with the absolute coefficient magnitude within each panel. Node positions and curve shapes are chosen for visual clarity only and do not carry econometric interpretation.

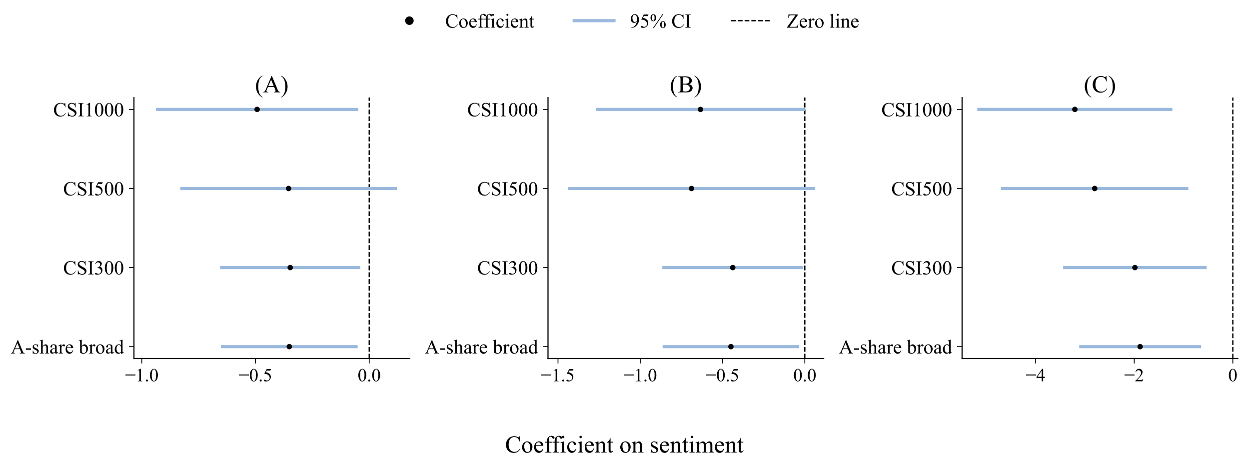

Figure 7: Sentiment Coefficients for Chinese Equity Returns

The figure plots sentiment coefficients and 95% confidence intervals from Table 14. Panel A reports weekly baseline $WDSI_{t-1}^{CN}$ (7-day MA). Panel B reports weekly alternative $WDSI_{t-1}^{CN,30}$ (30-day MA). Panel C reports monthly contemporaneous $WDSI_{t}^{CN,30}$.

Table 1: Distribution of Chinese WDSI Scores and Example of Model Output

Panel A reports the distribution of Chinese WDSI scores assigned to individual press-conference days on the [-3, 3] scale, before any forward-filling or temporal interpolation. Panels B and C present illustrative negative (-3) and positive (3) cases, respectively, including the input text, date and press-conference title, assigned score, and the AI-generated rationale.

Panel A. Distribution of Raw Discrete Chinese WDSI Scores

	-3	-2	-1	0	1	2	3
Number of press conferences	48	1470	481	638	206	123	4
Share (%)	1.62	49.49	16.20	21.48	6.84	4.14	0.13

Panel B. Example of Negative War-related Diplomatic Sentiment (Score: -3; Date: 2022-12-09)

Title	Regular Press Conference of Foreign Ministry Spokesperson Mao Ning, 9 December 2022
Q&A Unit	**Agence France-Presse Reporter:** The U.S. House of Representatives has passed the "National Defense Authorization Act", which includes provisions for up to US$10 billion in military loans to Taiwan over the next five years. What is your comment? **Spokesperson Ning Mao:** China firmly opposes the inclusion of negative China-related content in the NDAA. Taiwan is China's Taiwan, and the Taiwan question is purely China's internal affair. The U.S. should abide by the one-China principle and the three China-U.S. joint communiqués, delete the negative China-related content in the bill, stop playing the "Taiwan card" and using Taiwan to contain China, and stop U.S.-Taiwan military contacts so as to avoid further damage to China-U.S. relations. China will take all necessary measures to firmly safeguard its sovereignty and security interests.
Model Rationale	The exchange condemns the U.S. "National Defense Authorization Act" for authorising large-scale military loans to Taiwan, reaffirms that Taiwan is part of China and that the Taiwan issue is a purely internal matter, urges the U.S. to stop "playing the Taiwan card" and all U.S.-Taiwan military contacts, and warns that China will take "all necessary measures" to defend its sovereignty and security. The combination of strong negative evaluation, opposition to foreign military involvement and threats of countermeasures leads the pipeline to classify this unit as high-intensity negative war-related sentiment and to assign the lowest score of -3.

Panel C. Example of Positive War-related Diplomatic Sentiment (Score: 3; Date: 2021-04-15)

Title	Regular Press Conference of Foreign Ministry Spokesperson Lijian Zhao, 15 April 2021
Q&A Unit	**China National Radio Reporter:** It is reported that Russian Foreign Minister Lavrov said in an interview on the 14th that the United States has cast diplomatic principles aside and publicly claimed that Russia should "help the U.S. contain China," which is despicable and shameless. Russia has no reason to "distance itself from China to please the U.S." and would not be so foolish as to abandon existing Russia-China consensus. What is China's comment? **Spokesperson Lijian Zhao:** On attempts by certain countries to drive a wedge between China and Russia, Foreign Minister Lavrov has repeatedly made China's position clear, expressing opposition and strongly refuting such provocations that violate basic norms of international relations and are unworthy of a major country. Certain countries try to fabricate issues out of thin air and stir up discord between China and Russia; this is exactly the outcome they deserve. China fully agrees with Russia's just position...
Model Rationale	The exchange rejects attempts by a third party to provoke confrontation between China and Russia, explicitly endorses Russia's position, and emphasizes mutual trust, strategic coordination, and the stability and resilience of the bilateral relationship. By framing external provocations as illegitimate and stressing cooperation and restraint rather than escalation or the use of force, the statement conveys a clear preference for peace and stability. The combination of strong affirmation of partnership, opposition to conflict, and de-escalatory signalling leads the pipeline to classify this unit as high-intensity positive war-related diplomatic sentiment and to assign the highest score of 3.

Table 2: Directional VAR Evidence Between WDSI and GPR: Full Lag Grid

This table reports bivariate daily VAR(p) results for China's 30-day war-related diplomatic sentiment index $WDSI_t^{CN,30}$ and the China-specific geopolitical risk index GPR_t^{CN} in the full sample over lag orders $p = 1, \ldots, 22$. The estimation uses raw (non-residualised) level series from the baseline integrated daily dataset (full sample, 2010-2025), with no additional controls included inside the VAR system. "P-Value: WDSI→GPR" is the Granger-causality test for the null that lagged WDSI does not help predict GPR. "P-Value: GPR→WDSI" reports the reverse-direction test. Lower AIC/BIC indicate better in-sample fit. All estimated VAR systems are stable. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Lag order p	Observations	P-Value: WDSI→GPR	P-Value: GPR→WDSI	AIC	BIC
1	5,718	0.221	0.586	-12.096	-12.089
2	5,717	0.465	0.184	-12.335	-12.324
3	5,716	0.103	0.302	-12.348	-12.332
4	5,715	0.157	0.391	-12.346	-12.325
5	5,714	0.150	0.440	-12.350	-12.325
6	5,713	0.224	0.415	-12.352	-12.322
7	5,712	0.008***	0.498	-12.356	-12.321
8	5,711	0.005***	0.320	-12.358	-12.319
9	5,710	0.007***	0.335	-12.357	-12.313
10	5,709	0.013**	0.367	-12.355	-12.306
11	5,708	0.020**	0.348	-12.358	-12.305
12	5,707	0.031**	0.255	-12.358	-12.300
13	5,706	0.026**	0.303	-12.357	-12.294
14	5,705	0.026**	0.360	-12.356	-12.288
15	5,704	0.032**	0.422	-12.359	-12.286
16	5,703	0.013**	0.395	-12.359	-12.282
17	5,702	0.017**	0.423	-12.358	-12.276
18	5,701	0.026**	0.461	-12.358	-12.272
19	5,700	0.036**	0.473	-12.357	-12.266
20	5,699	0.028**	0.477	-12.356	-12.260
21	5,698	0.039**	0.490	-12.355	-12.255
22	5,697	0.043**	0.548	-12.354	-12.249

Table 3: Descriptive Statistics

Variable (Full Name)	Symbol	Unit	Obs.	Mean	Std. Dev.	P10	P90
Net Northbound Inflows	Inflow	100 million USD	2,205	0.988	4.907	-5.750	7.529
Net Southbound Flows	Outflow	100 million USD	2,224	1.815	2.834	-1.372	6.184
Net Stock Connect Flow	NetFlow	100 million USD	2,190	-0.782	6.182	-9.455	7.304
War-Related Diplomatic Sentiment	$WDSI^{CN}$	Index	3,618	-0.934	0.808	-2.000	0.000
Chinese Equity Market Return	R^{CN}	%	3,618	0.037	0.939	-1.225	1.355
U.S. Equity Market Return	R^{US}	%	3,618	0.060	0.845	-1.160	1.210
Developed Market Equity Return	R^{DEV}	%	3,618	0.030	0.703	-0.980	0.980
China Geopolitical Risk Index	GPR^{CN}	Index	3,618	0.638	0.260	0.329	1.035
U.S. Geopolitical Risk Index	GPR^{US}	Index	3,618	2.858	0.654	2.107	3.842
China Economic Policy Uncertainty	EPU^{CN}	Index	3,618	144.469	26.290	110.929	183.329
U.S. Economic Policy Uncertainty	EPU^{US}	Index	3,618	141.532	43.677	93.204	201.032
Gold Return	$Gold$	%	3,618	0.000	0.008	-0.011	0.011
RMB-USD Exchange Rate	FX	RMB/USD	3,618	6.647	0.351	6.204	7.180

Table 4: Baseline Results: Chinese War-Related Diplomatic Sentiment and Daily Stock Connect Flows

The table reports OLS estimates of Equation (1) with heteroskedasticity-robust standard errors in parentheses. The dependent variables are daily net northbound inflows (column 1), net southbound flows (column 2), and their combined net flow (column 3) through the Stock Connect programs. Flow variables are measured in hundred million USD. Equity returns are expressed in percentage points. All flow variables, equity returns, and control variables are winsorised at the 5th and 95th percentiles; additional winsorisation-threshold robustness checks, not separately tabulated, using the 1st/99th and 10th/90th percentiles yield very similar coefficient estimates and significance levels. Month fixed effects are included in all specifications but their coefficients are omitted for brevity. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1) Inflow	(2) Outflow	(3) Net flow
$WDSI^{CN}$	-0.025	-0.265***	0.356**
	(0.126)	(0.067)	(0.157)
R^{CN}	1.959***	0.397***	1.558***
	(0.113)	(0.060)	(0.141)
R^{US}	-0.242*	0.166*	-0.442**
	(0.140)	(0.085)	(0.184)
$R_{Developed}$	0.972***	-0.362***	1.378***
	(0.183)	(0.109)	(0.239)
GPR^{CN}	2.464***	-0.899***	3.367***
	(0.464)	(0.285)	(0.621)
GPR^{US}	-1.192***	0.159	-1.257***
	(0.191)	(0.114)	(0.254)
EPU^{CN}	0.000	-0.003	0.002
	(0.004)	(0.002)	(0.006)
EPU^{US}	-0.002	0.014***	-0.017***
	(0.003)	(0.002)	(0.004)
$Gold$	-4.859	5.761	-16.421
	(13.258)	(7.744)	(17.564)
FX	-1.148***	2.207***	-2.887***
	(0.377)	(0.218)	(0.497)
Month FE	Yes	Yes	Yes
Observations	2,205	2,224	2,190
R^2	0.223	0.136	0.143

Table 5: Robustness to Alternative Sentiment Indicators: Chinese Stock Connect Flows

This table reports OLS estimates of Equation (1) with heteroskedasticity-robust standard errors in parentheses. The dependent variables are daily net northbound inflows (Inflow), net southbound flows (Outflow), and their combined net flow (Net flow) through the Stock Connect programs, measured in hundred million USD. Columns (1)-(3) replace the baseline war-related sentiment index $WDSI_t^{CN}$ with an economic-related sentiment index $EDSI_t^{CN}$; columns (4)-(6) use an index of other diplomatic sentiment $ODSI_t^{CN}$; columns (7)-(9) use the negative-part war sentiment index $WDSI_t^{CN,-}$, which equals $WDSI_t^{CN}$ when the index is negative and zero otherwise. All sentiment indices are expressed by their respective seven-day moving averages. Equity returns are expressed in percentage points. All flow variables, returns, and controls are winsorised at the 5th and 95th percentiles. Month fixed effects are included in all specifications but their coefficients are omitted for brevity. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Economic conflict sentiment			Other diplomatic sentiment			War sentiment, negative only		
	(1) Inflow	(2) Outflow	(3) Net flow	(4) Inflow	(5) Outflow	(6) Net flow	(7) Inflow	(8) Outflow	(9) Net flow
Sentiment index	-0.019	0.141*	-0.247	-0.206**	0.071	-0.256*	-0.063	-0.363***	0.444**
	(0.134)	(0.077)	(0.177)	(0.104)	(0.060)	(0.134)	(0.144)	(0.084)	(0.190)
R^{CN}	1.960***	0.395***	1.562***	1.957***	0.398***	1.556***	1.960***	0.401***	1.554***
	(0.113)	(0.060)	(0.141)	(0.113)	(0.060)	(0.141)	(0.113)	(0.059)	(0.141)
R^{US}	-0.243*	0.172**	-0.453**	-0.239*	0.164*	-0.437**	-0.241*	0.170**	-0.446**
	(0.140)	(0.086)	(0.185)	(0.140)	(0.086)	(0.185)	(0.140)	(0.085)	(0.184)
$R_{Developed}$	0.974***	-0.361***	1.379***	0.976***	-0.357***	1.372***	0.971***	-0.364***	1.380***
	(0.183)	(0.109)	(0.240)	(0.183)	(0.109)	(0.240)	(0.183)	(0.109)	(0.239)
GPR^{CN}	2.455***	-0.804***	3.203***	2.396***	-0.856***	3.249***	2.462***	-0.903***	3.371***
	(0.474)	(0.291)	(0.634)	(0.466)	(0.286)	(0.625)	(0.464)	(0.285)	(0.621)
GPR^{US}	-1.185***	0.174	-1.269***	-1.139***	0.178	-1.244***	-1.196***	0.152	-1.253***
	(0.191)	(0.113)	(0.253)	(0.190)	(0.112)	(0.252)	(0.190)	(0.113)	(0.253)
EPU^{CN}	0.000	-0.005**	0.004	0.001	-0.005**	0.004	0.000	-0.004*	0.003
	(0.004)	(0.002)	(0.006)	(0.004)	(0.002)	(0.006)	(0.004)	(0.002)	(0.005)
EPU^{US}	-0.002	0.015***	-0.018***	-0.003	0.015***	-0.019***	-0.002	0.014***	-0.017***
	(0.003)	(0.002)	(0.004)	(0.003)	(0.002)	(0.004)	(0.003)	(0.002)	(0.004)
$Gold$	-4.835	5.130	-15.443	-4.972	5.545	-16.242	-4.834	5.818	-16.474
	(13.262)	(7.772)	(17.599)	(13.225)	(7.767)	(17.560)	(13.255)	(7.735)	(17.568)
FX	-1.136***	2.289***	-2.987***	-1.140***	2.305***	-3.016***	-1.152***	2.226***	-2.924***
	(0.372)	(0.215)	(0.490)	(0.371)	(0.216)	(0.489)	(0.374)	(0.216)	(0.492)
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,205	2,224	2,190	2,205	2,224	2,190	2,205	2,224	2,190
R^2	0.223	0.133	0.142	0.224	0.132	0.142	0.223	0.138	0.143

Table 6: Robustness to Alternative Smoothing Windows for Chinese War-Related Sentiment

The table reports OLS estimates of Equation (1) with heteroskedasticity-robust standard errors in parentheses. The dependent variables are daily net northbound inflows (Inflow), net southbound flows (Outflow), and their combined net flow (Net flow) through the Stock Connect programs, measured in hundred million USD. Columns (1)-(3) use a three-day moving average of the Chinese war-related diplomatic sentiment index $WDSI_t^{CN,3}$, while columns (4)-(6) use a thirty-day moving average $WDSI_t^{CN,30}$. Equity returns are expressed in percentage points. All flow variables, returns, and controls are winsorised at the 5th and 95th percentiles. Month fixed effects are included in all specifications but their coefficients are omitted for brevity. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	3-day average			30-day average		
	(1) Inflow	(2) Outflow	(3) Net flow	(4) Inflow	(5) Outflow	(6) Net flow
$WDSI^{CN}$	0.022	-0.228***	0.307**	0.198	-0.649***	1.017***
	(0.109)	(0.062)	(0.140)	(0.202)	(0.098)	(0.245)
R^{CN}	1.966***	0.394***	1.567***	1.968***	0.389***	1.578***
	(0.113)	(0.060)	(0.141)	(0.112)	(0.060)	(0.141)
R^{US}	-0.211	0.150*	-0.393**	-0.212	0.155*	-0.401**
	(0.141)	(0.086)	(0.185)	(0.141)	(0.085)	(0.184)
$R_{Developed}$	0.915***	-0.338***	1.297***	0.918***	-0.343***	1.308***
	(0.185)	(0.110)	(0.241)	(0.184)	(0.109)	(0.239)
GPR^{CN}	2.463***	-0.874***	3.332***	2.500***	-0.987***	3.524***
	(0.465)	(0.286)	(0.622)	(0.462)	(0.286)	(0.621)
GPR^{US}	-1.184***	0.158	-1.256***	-1.174***	0.145	-1.232***
	(0.191)	(0.114)	(0.254)	(0.191)	(0.113)	(0.254)
EPU^{CN}	0.000	-0.004	0.002	-0.001	-0.002	0.000
	(0.004)	(0.002)	(0.006)	(0.004)	(0.002)	(0.006)
EPU^{US}	-0.002	0.014***	-0.017***	-0.002	0.014***	-0.017***
	(0.003)	(0.002)	(0.004)	(0.003)	(0.002)	(0.004)
$Gold$	-4.889	5.735	-16.375	-4.899	5.750	-16.407
	(13.254)	(7.743)	(17.559)	(13.247)	(7.711)	(17.524)
FX	-1.120***	2.198***	-2.861***	-1.069***	2.097***	-2.682***
	(0.375)	(0.219)	(0.496)	(0.376)	(0.216)	(0.494)
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,205	2,224	2,190	2,205	2,224	2,190
R^2	0.222	0.135	0.140	0.222	0.144	0.145

Table 7: VAR Dynamic Robustness: Daily and Weekly VAR(1) Specifications

The table reports residualised VAR(1) results for the three Stock Connect flow channels. Panel A uses daily $WDSI_t^{CN}$ (seven-day moving average), Panel B uses daily $WDSI_t^{CN,30}$ (thirty-day moving average), Panel C uses weekly $WDSI_t^{CN,30}$, and Panel D uses weekly $WDSI_t^{CN}$. Weekly observations are built from Monday-anchored weeks, with flow variables summed and sentiment/control variables averaged within week. For each channel, both the sentiment index and the flow variable are first residualised on the full baseline control set and month fixed effects from Equation (1), and the VAR is then estimated on the residualised series. Coefficients are reported with standard errors in parentheses. Granger tests are one-lag tests. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

	Inflow / Sentiment	Outflow / Sentiment	Net flow / Sentiment
Panel A: Daily VAR(1), baseline $WDSI_t^{CN}$ (7-day MA)			
$L1.$ Sentiment	-0.007	-0.049**	0.032
	(0.021)	(0.020)	(0.021)
$L1.$ Flow	0.213***	0.370***	0.283***
	(0.021)	(0.020)	(0.021)
p(Sentiment $\rightarrow$ Flow)	0.722	0.013	0.114
p(Flow $\rightarrow$ Sentiment)	0.491	0.141	0.789
Observations	2,204	2,223	2,189
Panel B: Daily VAR(1), alternative $WDSI_t^{CN,30}$ (30-day MA)			
$L1.$ Sentiment	0.010	-0.085***	0.064***
	(0.021)	(0.020)	(0.021)
$L1.$ Flow	0.213***	0.362***	0.279***
	(0.021)	(0.020)	(0.021)
p(Sentiment $\rightarrow$ Flow)	0.623	0.000	0.002
p(Flow $\rightarrow$ Sentiment)	0.825	0.036	0.308
Observations	2,204	2,223	2,189
Panel C: Weekly VAR(1), alternative $WDSI_t^{CN,30}$ (30-day MA)			
$L1.$ Sentiment	0.048	-0.079**	0.105**
	(0.044)	(0.038)	(0.041)
$L1.$ Flow	0.207***	0.533***	0.398***
	(0.044)	(0.038)	(0.041)
p(Sentiment $\rightarrow$ Flow)	0.278	0.040	0.011
p(Flow $\rightarrow$ Sentiment)	0.451	0.010	0.415
Observations	497	496	496
Panel D: Weekly VAR(1), baseline $WDSI_t^{CN}$ (7-day MA)			
$L1.$ Sentiment	0.030	-0.091**	0.080*
	(0.044)	(0.038)	(0.041)
$L1.$ Flow	0.208***	0.537***	0.406***
	(0.044)	(0.038)	(0.041)
p(Sentiment $\rightarrow$ Flow)	0.498	0.016	0.052
p(Flow $\rightarrow$ Sentiment)	0.859	0.012	0.341
Observations	497	496	496

Table 8: Endogeneity Checks: Placebo Tests and Detrended Sentiment

Panel A reports placebo regressions in which the dependent variable is the one-day-ahead Chinese war-related diplomatic sentiment index $WDSI^{CN}_{t+1}$ and the key regressors are contemporaneous and lagged Stock Connect flows (inflow, outflow, or net flow). Panel B re-estimates Equation (1) using a detrended sentiment index $WDSI^{CN}_t$ – the residuals from regressing $WDSI^{CN}_t$ on a linear time trend – in place of the raw index. All specifications include the same set of controls as in Table 4 as well as month fixed effects; their coefficients are omitted for brevity. Robust standard errors are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Panel A: Placebo tests			Panel B: Detrended sentiment		
	(1) Inflow	(2) Outflow	(3) Net flow	(4) Inflow	(5) Outflow	(6) Net flow
$Flow^{CN}_{k,t}$	-0.00060	-0.00824	0.00375			
	(0.00606)	(0.00988)	(0.00453)			
$Flow^{CN}_{k,t-1}$	0.04738	-0.18904	0.06394			
	(0.07301)	(0.14258)	(0.05988)			
$WDSI^{CN}_t$				-0.010	-0.228***	0.339**
(detrended)				(0.126)	(0.067)	(0.158)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,204	2,223	2,189	2,205	2,224	2,190
R^2	0.047	0.045	0.047	0.221	0.134	0.140

Table 9: Dynamic Specifications with Lagged Flows and Controls

The table augments the baseline Stock Connect regressions in Equation (1) with lagged dependent variables. Panel A includes the first lag of each flow variable; Panel B includes both first and second lags. All specifications retain the full set of controls as in Table 4 as well as month fixed effects; coefficients on these controls and on lagged controls are omitted for brevity. Robust standard errors are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Panel A: First-order lags			Panel B: Second-order lags		
	(1) Inflow	(2) Outflow	(3) Net flow	(4) Inflow	(5) Outflow	(6) Net flow
$WDSI_t^{CN}$	0.010	-0.188***	0.312**	0.039	-0.144**	0.298**
	(0.120)	(0.061)	(0.147)	(0.118)	(0.058)	(0.143)
$Flow_{k,t-1}^{CN}$	2.986***	5.333***	4.030***	2.522***	3.945***	3.259***
	(0.358)	(0.390)	(0.360)	(0.369)	(0.414)	(0.375)
$Flow_{k,t-2}^{CN}$				1.872***	3.846***	2.534***
				(0.347)	(0.405)	(0.370)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Controls($t-1$)	Yes	Yes	Yes	Yes	Yes	Yes
Controls($t-2$)	No	No	No	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,204	2,223	2,189	2,203	2,222	2,188
R^2	0.287	0.280	0.252	0.328	0.336	0.295

Table 10: Instrumental Variable Estimates

Column (1) reports the first-stage regression of the Chinese war-related diplomatic sentiment index $WDSI_t^{CN}$ on the spokesperson fixed-sentiment measure, defined as the average sentiment across all answers provided by each spokesperson, along with the full set of controls used in Table 4 and month fixed effects. Columns (2)-(4) report second-stage 2SLS estimates of Equation (1), where $WDSI_t^{CN}$ is instrumented with the spokesperson fixed-sentiment variable. The dependent variables are daily net northbound inflows (column 2), net southbound flows (column 3), and their combined net flow (column 4) through the Stock Connect programs. All specifications include the same controls and fixed effects as the baseline OLS regressions; their coefficients are omitted for brevity. Robust standard errors are reported in parentheses. "First-stage Wald χ^2 (excluded IV)" reports the robust first-stage relevance test in each equation sample. "KP rk Wald F" and "KP rk LM P-Value" are the Kleibergen–Paap weak-identification and under-identification diagnostics (from `ivreg2`, robust option). *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1) First stage	(2) Inflow	(3) Outflow	(4) Net flow
Dependent variable	$WDSI_t^{CN}$	$Flow^{CN}_{\text{Inflow},t}$	$Flow^{CN}_{\text{Outflow},t}$	$Flow^{CN}_{\text{Net},t}$
Spokesperson fixed sentiment	1.000***			
	(0.089)			
$WDSI_t^{CN}$		0.308	-2.342***	2.483***
		(0.621)	(0.413)	(0.848)
Controls	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
First-stage Wald χ^2 (excluded IV)	127.812	127.812	128.741	124.884
KP rk Wald F		126.537	127.468	123.629
KP rk LM P-Value		< 0.001	< 0.001	< 0.001
Observations	2,205	2,205	2,224	2,190

Table 11: War-Related Sentiment in the U.S., U.K., Japan, and South Korea and Chinese Stock Connect Flows

The table reports daily OLS estimates with heteroskedasticity-robust standard errors in parentheses. Columns correspond to Chinese Stock Connect flow equations (inflow, outflow, net flow), and rows report coefficients on each foreign war-related sentiment index. Dependent variables are Chinese Stock Connect flows measured in hundred million USD. All specifications include the same control block (Chinese/U.S./developed-market returns, China/U.S. GPR and EPU, gold returns, RMB-USD exchange rates) and month fixed effects. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1) Inflow	(2) Outflow	(3) Net flow
Panel A: Full sample			
U.S. WDSI	0.075	-0.217**	0.315
	(0.163)	(0.105)	(0.220)
U.K. WDSI	0.105	-0.209***	0.287*
	(0.120)	(0.075)	(0.162)
Japan WDSI	0.021	-0.191**	0.270
	(0.124)	(0.077)	(0.165)
South Korea WDSI	0.355***	-0.150*	0.491***
	(0.132)	(0.080)	(0.174)
Controls	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Observations	2,205	2,224	2,190
Panel B: Pre-2022 (year $\leq$ 2021)			
U.S. WDSI	-0.157	-0.119	-0.034
	(0.169)	(0.109)	(0.226)
U.K. WDSI	-0.243*	0.055	-0.379**
	(0.128)	(0.077)	(0.169)
Japan WDSI	-0.124	-0.194**	0.113
	(0.129)	(0.077)	(0.170)
South Korea WDSI	0.463***	-0.287***	0.757***
	(0.130)	(0.076)	(0.165)
Controls	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Observations	1,612	1,626	1,597
Panel C: Post-2021 (year $>$ 2021)			
U.S. WDSI	1.405***	-0.221	1.665**
	(0.516)	(0.323)	(0.699)
U.K. WDSI	0.612*	-0.169	0.866**
	(0.325)	(0.197)	(0.439)
Japan WDSI	0.639**	-0.056	0.742*
	(0.309)	(0.183)	(0.387)
South Korea WDSI	0.114	-0.002	0.067
	(0.296)	(0.187)	(0.397)
Controls	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Observations	593	598	593

Table 12: Domestic War-Related Sentiment and Own-Country International Equity Positions

The table reports monthly OLS estimates with heteroskedasticity-robust standard errors in parentheses. Columns correspond to own-country flow equations (inflow, outflow, net flow), and rows report coefficients on each country's domestic war-related sentiment index. Dependent variables are inverse hyperbolic sine transformations of own-country inward, outward, and net equity positions. To keep the three country blocks directly comparable, all specifications are estimated on the common monthly sample from January 2015 to June 2025, which yields 126 observations in each column; the start date is determined by the availability of the Japanese foreign equity flow series after monthly aggregation, and the end date by the availability of the U.S. monthly international equity position series. All specifications include country-level controls (U.S./developed-market returns, local GPR, local EPU, local exchange-rate control) and month fixed effects. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1) Inflow_{own}	(2) Outflow_{own}	(3) Net flow_{own}
U.S. WDSI	-0.283***	-0.158***	-14.576***
	(0.075)	(0.048)	(2.610)
Japan WDSI	0.018	0.013	5.429
	(0.039)	(0.040)	(3.941)
South Korea WDSI	0.143*	0.135*	8.214***
	(0.077)	(0.076)	(2.271)
Controls	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Observations	126	126	126

Table 13: Chinese War-Related Sentiment and the A-H Share Premium

The table reports OLS estimates using the same control block and month fixed effects as Equation (1), with the A-H share premium as the dependent variable. The key explanatory variable is the Chinese war-related sentiment index $WDSI^{CN}$ (seven-day moving average). Column (1) includes only contemporaneous regressors; columns (2) and (3) additionally include one and two lags of the A-H premium, respectively, to capture persistence. Columns (1)–(3) use the level of the A-H premium as the dependent variable, whereas column (4) uses its first difference and includes one and two lags of the differenced dependent variable. The A-H premium is measured in percentage points. Standard errors are heteroskedasticity-robust and reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1) A-H Share Premium	(2) A-H Share Premium	(3) A-H Share Premium	(4) Δ A-H Share Premium
$WDSI^{CN}$	-1.433***	-0.070*	-0.073*	-0.071*
	(0.294)	(0.041)	(0.041)	(0.041)
A-H Share Premium$_{t-1}$		0.993***	0.910***	
		(0.002)	(0.037)	
A-H Share Premium$_{t-2}$			0.084**	
			(0.037)	
Δ A-H Share Premium$_{t-1}$				-0.094**
				(0.037)
Δ A-H Share Premium$_{t-2}$				-0.068**
				(0.027)
R^{CN}	-0.051	0.024	0.018	0.013
	(0.239)	(0.035)	(0.034)	(0.035)
R^{US}	0.169	0.162***	0.161***	0.161***
	(0.298)	(0.041)	(0.042)	(0.043)
$R_{Developed}$	-0.542	-0.525***	-0.524***	-0.525***
	(0.387)	(0.054)	(0.054)	(0.056)
GPR^{CN}	4.379***	0.046	0.046	0.027
	(0.951)	(0.120)	(0.120)	(0.119)
GPR^{US}	3.264***	0.048	0.047	0.029
	(0.404)	(0.040)	(0.040)	(0.040)
EPU^{CN}	0.040***	-0.000	-0.000	-0.000
	(0.009)	(0.001)	(0.001)	(0.001)
EPU^{US}	0.002	-0.000	-0.000	-0.000
	(0.006)	(0.001)	(0.001)	(0.001)
Gold	13.279	9.548***	9.375***	10.497***
	(18.330)	(2.231)	(2.229)	(3.376)
FX	21.481***	0.076	0.056	-0.090
	(0.645)	(0.090)	(0.091)	(0.073)
Month FE	Yes	Yes	Yes	Yes
Observations	3,242	3,241	3,240	3,239
R^2	0.425	0.993	0.993	0.061

Table 14: Chinese War-Related Sentiment and Equity-Index Returns

The table reports OLS regressions of Chinese equity-index returns (in percentage points) on war-related diplomatic sentiment. Panel A uses weekly baseline sentiment $WDSI_{t-1}^{CN}$ (seven-day moving average, lag 1). Panel B uses weekly alternative sentiment $WDSI_{t-1}^{CN,30}$ (thirty-day moving average, lag 1). Panel C uses monthly contemporaneous sentiment $WDSI_{t}^{CN,30}$. All specifications include the same control block as Equation (1): U.S./developed-market returns, China/U.S. GPR, China/U.S. EPU, gold, FX, and month fixed effects. Robust standard errors are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1) A-share broad	(2) CSI 300	(3) CSI 500	(4) CSI 1000
Panel A: Weekly baseline $WDSI_{t-1}^{CN}$ (7-day MA)				
$WDSI_{t-1}^{CN}$	-0.351**	-0.347**	-0.354	-0.493**
	(0.153)	(0.157)	(0.243)	(0.227)
Controls	Yes	Yes	Yes	Yes
Observations	790	800	632	799
R^2	0.132	0.136	0.106	0.098
Panel B: Weekly alternative $WDSI_{t-1}^{CN,30}$ (30-day MA)				
$WDSI_{t-1}^{CN,30}$	-0.449**	-0.438**	-0.689*	-0.634*
	(0.212)	(0.218)	(0.383)	(0.325)
Controls	Yes	Yes	Yes	Yes
Observations	787	797	632	796
R^2	0.128	0.133	0.111	0.095
Panel C: Monthly $WDSI_{t}^{CN,30}$ (contemporaneous)				
$WDSI_{t}^{CN,30}$	-1.881***	-1.986***	-2.801***	-3.204***
	(0.630)	(0.743)	(0.969)	(1.009)
Controls	Yes	Yes	Yes	Yes
Observations	185	187	148	187
R^2	0.280	0.277	0.333	0.299

Table 15: War-Related Sentiment and Market Activity

The table reports OLS estimates of Equation (1) with measures of market activity as dependent variables. Panel A uses the log of daily trading volume for the CSI 300, CSI 1000, Nasdaq, and Dow Jones Industrial Average. Panel B uses daily realized volatility of the same indices as dependent variables. In columns (1), (2), (5), and (6) the key explanatory variable is the Chinese war-related sentiment index $WDSI^{CN}$ (seven-day moving average); in columns (3), (4), (7), and (8) it is the U.S. war-related sentiment index $WDSI^{US}$. Standard errors are heteroskedasticity-robust and reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A: Log trading volume				
	(1) CSI 300	(2) CSI 1000	(3) Nasdaq	(4) Dow Jones Industrial Average
$WDSI^{CN}$	-0.033**	-0.093***		
	(0.013)	(0.015)		
$WDSI^{US}$			-0.144***	-0.113***
			(0.009)	(0.012)
Controls	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Observations	3,618	3,618	3,618	3,618
R^2	0.128	0.274	0.473	0.600
Panel B: Realized volatility				
	(5) CSI 300	(6) CSI 1000	(7) Nasdaq	(8) Dow Jones Industrial Average
$WDSI^{CN}$	0.000013**	0.000033***		
	(0.000005)	(0.000008)		
$WDSI^{US}$			-0.000026***	-0.000010***
			(0.000005)	(0.000003)
Controls	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes
Observations	3,736	3,736	3,618	3,618
R^2	0.042	0.036	0.106	0.103

Appendix A Supplementary Materials on the Measurement of WDSI

Appendix A.1 Data Sources

Table A.1 presents the complete set of textual sources used to construct the WDSI. For the baseline regressions and the comparative analysis, we draw on transcripts from the press conferences of China's Ministry of Foreign Affairs and press releases issued by the U.S. Department of State (a detailed description of the main characteristics of these two datasets is provided below), while the supplementary analysis incorporates data from the United Kingdom, Japan and the Republic of Korea.

Appendix A.1.1 Chinese Ministry of Foreign Affairs Press Conferences

We assemble a corpus of transcripts from the regular press conferences of the Ministry of Foreign Affairs of China between 1 January 2010 and 30 September 2025. Each press conference is an official briefing at which a spokesperson answers questions from domestic and foreign journalists. All Q&A units released on the same day are merged into a single daily document to construct the Chinese WDSI series. This yields several thousand daily observations, with coverage roughly even over time (press conferences are held on most weekdays, with fewer sessions during major holidays). The transcripts are published on the MFA's official website and disseminated through state media, and can therefore be regarded as authoritative expressions of China's diplomatic position on current international affairs.

The Chinese press conferences have several salient features. First, the limited number of questions per session induces journalists to focus on the most important war, conflict and security issues of the day, so the content is closely aligned with contemporaneous geopolitical shocks. Regional wars, military confrontations, security threats and other high-profile incidents are repeatedly discussed. Second, the spokesperson's answers are carefully scripted and pre-cleared within the foreign-policy bureaucracy, and thus reflect formal state positions.

Table A.1: Diplomatic Text Corpora Used to Construct the WDSI

The table summarises the diplomatic text corpora used to construct the war-related diplomatic sentiment indices. Data sources: China - regular press-conference transcripts from the Ministry of Foreign Affairs of the People's Republic of China (https://www.fmprc.gov.cn); United States - press statements and official remarks from the U.S. Department of State (https://www.state.gov/briefings-statements/); United Kingdom - foreign policy news releases and statements from the UK Foreign, Commonwealth and Development Office (https://www.gov.uk/government/organisations/foreign-commonwealth-development-office); Japan - press conferences and written statements from the Ministry of Foreign Affairs of Japan (https://www.mofa.go.jp); South Korea - press releases and briefings from the Ministry of Foreign Affairs of the Republic of Korea (https://www.mofa.go.kr).

Countries	Text source	Issuing institution	Time coverage	Number of texts
China (WDSI^{CN})	Regular press-conference transcripts	Ministry of Foreign Affairs of the PRC	2010-01-01 to 2025-09-30	2,970 press conferences & 17,874 Q&A units
United States (WDSI^{US})	Press statements and official remarks	U.S. Department of State	2010-01-01 to 2025-09-30	34,317 statements
United Kingdom (WDSI^{UK})	Foreign policy news releases and statements	UK Foreign, Commonwealth and Development Office	2010-01-01 to 2025-09-30	17,401 documents
Japan (WDSI^{JP})	Press conferences and written statements	Ministry of Foreign Affairs of Japan	2013-03-01 to 2025-09-30	14,140 documents
South Korea (WDSI^{KR})	Press releases and briefings on foreign and security policy	Ministry of Foreign Affairs of the Republic of Korea	2010-01-01 to 2025-09-30	10,966 documents

As a result, the emotional tone of the transcripts provides a clear signal of China's war-related diplomatic sentiment. For example, at the 9 December 2022 press conference, MFA Spokesperson Mao Ning responded to a question on U.S. military aid to Taiwan by "firmly opposing" the legislation, urging the United States to "stop playing the Taiwan card", and warning that China would "take all necessary measures to firmly safeguard its sovereignty". As illustrated in Table 1, Panel B, the pipeline classifies this exchange as high-intensity negative sentiment and assigns the lowest score of -3. Finally, these briefings typically take place in the afternoon before Asian markets close, and are driven by news events rather than market conditions, which makes the resulting sentiment measures closer to exogenous policy signals for our subsequent asset-pricing analysis.

Appendix A.1.2 United States Department of State Press Statements

For the United States, we construct a parallel corpus of official press statements issued by the U.S. Department of State between 1 January 2010 and 30 September 2025. These statements are formal announcements or comments, usually in the name of the Secretary of State or the Department Spokesperson, that set out the U.S. government's position on international developments. All such statements available in the State Department's online archives over the sample period are collected and used as the American textual corpus for constructing the U.S. WDSI. Each document is a stand-alone statement, typically prompted by discrete events such as the outbreak or escalation of conflict, terrorist attacks, military operations, peace agreements, or other security-related developments. Statements are issued on an event-driven rather than a fixed schedule, so the daily frequency varies from zero to several releases. The content spans a broad range of war and security topics reflecting the global reach of U.S. foreign policy, and frequently expresses condemnation, concern, support for allies, or calls for de-escalation. All statements are processed through the same multi-stage evaluation pipeline as the Chinese data to obtain daily measures of U.S. war-related diplomatic sentiment.

While complementary, the two corpora also differ systematically. Substantively, U.S. statements more often address conflicts and security issues involving third countries (for example, civil wars, interventions, terrorist incidents and peacekeeping operations) in line with the broader external orientation of U.S. security policy. Chinese press conferences, by contrast, concentrate on issues directly related to China's own security environment, including regional disputes with neighbouring states, sovereignty questions such as Taiwan, border clashes and military activities in East Asia. Accordingly, the war-related discourse in the Chinese corpus is more heavily weighted towards East Asian and near-abroad tensions, whereas the U.S. corpus contains a larger share of statements on conflicts in the Middle

East, Africa and Europe, alongside occasional commentary on China-U.S. military frictions. For example, in an 8 July 2025 press statement, the Department Spokesperson "condemn[ed] the unprovoked Houthi terror attack" on a civilian cargo ship in the Red Sea, using strong negative language to denounce the perpetrators and express solidarity with affected parties. Such releases provide clear instances of strongly negative war-related sentiment analogous to the Chinese example above.

Taken together, the Chinese MFA press conferences and U.S. Department of State press statements offer high signal-to-noise measures of official diplomatic sentiment. By relying on central government communications rather than news reports, the WDSI is anchored in deliberate and strategic policy rhetoric rather than media framing or ex post commentary. The combined dataset spans more than fifteen years of high-frequency observations and underpins the construction of the bilateral war-related diplomatic sentiment indices analysed in the main text.

Appendix A.2 Detailed procedure for constructing the WDSI

This appendix provides a concise description of the technical procedure used to construct the war-related diplomatic sentiment index from the Chinese Ministry of Foreign Affairs press-conference transcripts.

Appendix A.2.1 Data preprocessing

We collect full transcripts of MFA regular press conferences from 1 January 2010 to 30 September 2025. Each transcript is segmented into question-answer (Q&A) units, defined as one journalist question and the corresponding spokesperson reply. We remove boilerplate text (greetings, housekeeping remarks, purely logistical announcements) and discard non-substantive fragments (for example, incomplete sentences or purely procedural exchanges). All cleaned Q&A units that occur on the same calendar day are later aggregated to a daily sentiment series. Days without a regular press conference are initially treated as missing and are filled forward at the index-construction stage.

Appendix A.2.2   Three-stage LLM pipeline

Sentiment extraction is implemented as a vertically decomposed three-stage pipeline, with one LLM "agent" responsible for each subtask:

(i) Military relevance screening (Agent 1). Agent 1 reads each Q&A unit and determines whether it concerns war, armed conflict, military security or closely related escalation risks. It returns a binary label indicating whether the unit is war related, together with a concise natural-language rationale. Only units classified as war related proceed to the subsequent steps; all remaining units are coded as having no war related sentiment for that day.

(ii) Sentiment category classification (Agent 2). For war related units, Agent 2 assigns a sentiment direction with respect to war and the use of force, choosing among negative, neutral and positive. Negative sentiment includes condemnation, explicit or implicit threats, strong warnings and highly critical language. Positive sentiment captures expressions of support for de-escalation, sympathy for conflict resolution efforts and approval of peaceful outcomes. Neutral sentiment refers to factual or balanced statements that do not convey a clear evaluative tone. Agent 2 outputs the assigned category together with a brief rationale that highlights the key phrases on which the judgement is based.

(iii) Intensity scoring (Agent 3). Conditional on the category chosen in Stage (ii), Agent 3 evaluates sentiment intensity on a discrete scale from -3 to 3. For negative sentiment, scores -1, -2 and -3 represent mildly, moderately and strongly negative tones, while for positive sentiment, +1, +2 and +3 analogously indicate increasing strength. Neutral sentiment is coded as 0. Agent 3 is instructed to distinguish, for instance, between mild discomfort, firm opposition and explicit threats of countermeasures, and to provide a short justification for the assigned score.

Appendix A.2.3 Prompt design and implementation

All three agents are implemented using the same underlying LLM, with task-specific prompts. Each prompt (i) defines the task in plain language, (ii) provides the relevant label set and definitions, and (iii) requests both a machine-readable output (label or score) and a concise rationale. Prompts for Stage 2 and Stage 3 explicitly condition on the previous-stage output to avoid inconsistent combinations of direction and intensity. This design shortens the reasoning chain required at each step and improves interpretability, since every classification decision is accompanied by an explicit explanation.

Appendix A.2.4 Validation and error control

To enhance reliability, we adopt a simple consensus mechanism. For each Q&A unit and each stage, the corresponding agent is queried multiple times with slightly perturbed prompts or random seeds. A separate validation agent then compares the independent outputs and their rationales. (i) If all runs agree on the same label or score, the result is accepted. (ii) If there is disagreement, the validation agent reviews the rationales, identifies the most coherent interpretation, and either selects that label or instructs the pipeline to rerun the stage with a clarified prompt.

This consensus procedure reduces random misclassifications and is particularly helpful in edge cases such as indirect references to military issues, rhetorical questions, ironic or highly coded language and multi-country comparisons. Manual audits on a stratified subsample (described in Appendix A.3) show that the pipeline attains accuracy above 85% across all three individual stages and surpasses 75% overall, markedly outperforming traditional dictionary-based and shallow vector-based approaches.

Appendix A.2.5 From scores to WDSI

Each press conference day is assigned a single conference-level raw sentiment score in {-3, -2, -1, 0, 1, 2, 3}. When multiple war-related Q&A units appear on the same day, the conference-day score is not computed as the arithmetic average of unit-level scores.

Instead, the full conference transcript is evaluated at the conference level: relevant war-related Q&A units are identified and scored, and their scores are then aggregated using a weighted average that assigns a weight of 1.2 to negative scores and an additional multiplier of 1.5 to exchanges involving major powers. The weighted average is then clipped to the $[-3, 3]$ range and rounded to the nearest integer. Days without any war-related Q&A are scored as 0. To generate a continuous daily series, we apply forward filling: a missing day inherits the most recent past score, reflecting the idea that official sentiment is sticky between briefings. Finally, we compute a 7-day rolling average of the daily scores, which smooths idiosyncratic noise and yields the final WDSI used in the empirical analysis.

Appendix A.3 Validation and Robustness of the WDSI

Appendix A.3.1 Manual annotation validation

We begin by benchmarking the pipeline against human-coded labels. From the full set of press conferences, we draw a stratified random sample of approximately 1,200 Q&A units (at roughly equal monthly intervals over 2010-2025) and have them independently annotated by two research assistants with background in international relations.

Each annotator, with no access to the model outputs, labels (i) whether the Q&A pertains to war or military security, (ii) the sentiment category (negative, neutral, positive) if war-related, and (iii) the sentiment intensity. The inter-annotator agreement is high: the concordance rate for the binary war-related classification is about 88%, the average accuracy for categorizing sentiment direction is around 80%, and agreement on sentiment intensity (directional tone positive/negative and strength) is roughly 83%. We treat the modal human label in each case as the ground truth. Comparing the pipeline's outputs to this benchmark, we find that our multi-stage approach correctly identifies war-related content and sentiment with high accuracy. In fact, at each of the three stages the pipeline achieves over 85% agreement with the human ground truth, and the overall accuracy (i.e., the share of Q&A units for which the entire sentiment extraction is correct) exceeds 75%. This indicates that the

WDSI construction procedure is largely faithful to expert human judgement. The misclassifications that do occur tend to involve borderline cases or ambiguous language (for example, indirect allusions to conflicts or mixed diplomatic messages), which are also the instances where human coders most frequently disagreed.

Notably, we do not observe any systematic bias in the pipeline's errors – it is not, for instance, consistently overestimating negativity or misclassifying economic disputes as military ones – suggesting that the model's outputs are well calibrated to the task.

Appendix A.3.2 Model comparison and selection

We next evaluate the performance of our large language model approach against two more traditional text-analysis methods.

The first baseline is a dictionary-based classifier, constructed with a curated lexicon of military terms and sentiment-laden keywords: this method flags war-related content by the presence of words like "war", "military", "attack", etc., and gauges sentiment polarity by counting negative versus positive phrases (e.g., "condemn", "warn" vs. "welcome", "commend"). The second baseline is a shallow machine-learning model. Specifically, we train a logistic classifier on word embedding features (pre-trained Word2Vec representations) to predict war relevance and sentiment, using the manually labelled subset for supervision. Finally, we compare these with several state-of-the-art LLMs capable of Chinese-English text understanding, and ultimately adopt the *Qwen-3* model for our pipeline due to its superior performance.

As summarised in Table A.2, the dictionary and shallow learning approaches perform notably worse than the LLM-based methods. The dictionary approach has particular difficulty with nuanced sentiment and intensity, yielding an overall accuracy of around 50-60%. The shallow classifier performs better, but still reaches only about two-thirds accuracy. By contrast, the Qwen-3 LLM outperforms both baselines in a single-step setup, whereas the multi-step pipeline delivers the best results, with war detection and overall accuracies in excess of 90% and close to 80% respectively. These comparisons indicate that our LLM pipeline captures sentiment information that simpler text-analysis techniques systematically miss.

Table A.2: Classification Accuracy of Different Sentiment Extraction Methods

The table reports the classification accuracy of various methods at key stages of the sentiment extraction task, evaluated on a human-labelled sample of approximately 1,200 Q&A units. *War detection* refers to correctly identifying whether a unit is war/military-related. *Sentiment category* accuracy is the percentage of war-related units for which the method correctly classifies the sentiment as negative, neutral, or positive. *Sentiment intensity* accuracy is the percentage of war-related units where the strength of sentiment (on the -3 to +3 scale) is correctly assigned. *Overall* accuracy is the share of all units (war-related or not) for which the method's complete output matches the ground-truth label (including correctly outputting "no war content" for non-war units). The LLM methods use the Qwen-3 model. The multi-step pipeline corresponds to our full procedure with staged prompts and consensus checks, whereas the direct single-step LLM approach attempts to produce the sentiment score in one shot. The LLM pipeline exhibits the highest accuracy at each stage and overall.

Method	War detection (%)	Sentiment category (%)	Sentiment intensity (%)	Overall (%)
Dictionary-based approach	81.4	65.0	57.2	52.3
Shallow ML model (embeddings + LR)	86.5	74.1	70.3	64.1
LLM (direct single-step)	90.2	82.0	75.6	72.5
LLM (multi-step pipeline)	93.3	88.4	87.1	79.8

Appendix A.3.3 Single-step versus multi-step strategy

An important finding from the above experiments is that breaking the task into multiple stages significantly improves reliability compared to a single-step approach. In the "direct scoring" strategy, we prompted the same LLM to output a final sentiment score (or a "no-war" designation) in one go, without intermediate prompts. This one-shot method, while faster, often resulted in reasoning errors. For example, the model occasionally overlooked implicit military context or conflated mildly negative remarks with neutral tone. As seen in Table A.2, the single-step LLM's overall accuracy was about 5-10 percentage points lower than the structured multi-step pipeline. The multi-step framework forces the model to focus on one aspect at a time (first topical relevance, then sentiment direction, then intensity), which simplifies the reasoning at each step. It also enables us to apply a consistency check. Specifically, we generate multiple independent runs and require agreement before finalising a label. In our implementation, each agent was run three times per Q&A unit with slight prompt variations; if the three outputs did not coincide, a validation routine automatically reviewed the rationales and, if needed, re-queried the model with a clarified prompt. We find that this consensus procedure further reduces misclassifications. In the validation sample, about 12% of war-related units initially yielded inconsistent outputs across the three runs at one of the stages; in most of these cases, a single automated retry resolved the discrepancy, and fewer than 3% of units required a second retry or manual inspection. By contrast, a single-pass approach has no such safeguard and is prone to occasional hallucinations or subjective leaps in logic. With the multi-step pipeline, we also observe a parallel benefit for human coders. When the research assistants were asked to assign an overall sentiment score directly to each unit (ignoring the stepwise protocol), their inter-coder agreement dropped noticeably (by roughly 10 percentage points) relative to the structured labelling approach. This highlights that even for human analysts, a decomposed, disciplined process leads to more consistent and accurate sentiment assessments. Overall, the evidence strongly favours the vertical decomposition strategy in terms of both accuracy and interpretability.

Appendix A.3.4 Robustness checks

Finally, we verify that the WDSI is not unduly sensitive to specific technical choices. First, we experiment with alternative smoothing windows and missing-value treatments. The baseline index is computed as a 7-day rolling average with forward-filling of non-conference days. Shortening or lengthening the window (to 5 or 10 days) produces very similar trajectories, and the pairwise correlation coefficients between these variants and the baseline WDSI are in excess of 0.95. Likewise, if we adopt an alternative imputation (e.g., carrying the last available value backwards, or leaving gaps and interpolating monthly means), the resulting index remains virtually unchanged in terms of trends and key turning points. Second, we confirm that using a different LLM yields consistent signals. For example, we re-ran the sentiment extraction pipeline on a subset of the data using GPT-4 (via the OpenAI API) in place of Qwen-3. GPT-4's classifications correlated strongly with the original pipeline outputs (with an overall Pearson $r \approx 0.8$ and no systematic shift in the mean sentiment level).

The Qwen-3 model was retained mainly because it slightly outperformed others on our validation metrics and is domain-tuned for bilingual texts, but the substantive conclusions do not depend on this particular model. Third, we probe the possibility of systematic bias by examining error patterns more closely. We find no evidence that the pipeline consistently leans towards a particular sentiment. For instance, among the false negatives in war detection, some are genuinely subtle cases (such as indirect references to security alliances) rather than clear misses, and the frequency of false positives is low. Errors in sentiment categorization are roughly evenly split between confusion of neutral vs. mild negative tone and confusion of mild vs. moderate negativity, suggesting the model is not simply over-classifying everything as strongly negative. In sum, these robustness checks support the validity of the WDSI. The index accurately reflects the intended outcome, that is, official war-related diplomatic sentiment, and is stable under a range of methodological variations.

Appendix A.3.5 Supplementary Inference Checks for Baseline Flow Regressions

Table A.3 reports supplementary inference checks for the baseline China flow regressions in Table 4. We report Newey–West HAC standard errors and moving-block bootstrap inference using two block/bandwidth choices (5 and 22 trading days).

Appendix A.3.6 Supplementary Granger-Causality Check: A-H Premium and Stock Connect Flows

As an additional mechanism check, we estimate residualised daily VAR(p) systems between the A-H share premium and each Stock Connect flow series, where both variables are first residualised on the full baseline control set and month fixed effects. Table A.4 reports Granger-causality P-Values for lag orders $p = 1, 3, 5$. The strongest and most stable pattern is A-H premium $\rightarrow$ outflow across all three lag choices, while net-flow predictability is present at shorter lags.

Table A.3: Supplementary Inference Checks: HAC and Moving-Block Bootstrap

This table re-estimates the baseline China regressions in Table 4 and reports the coefficient on the Chinese war-related diplomatic sentiment index $WDSI_t^{CN}$ under alternative inference methods. Panels A-B use Newey–West HAC standard errors with lag lengths 5 and 22. Panels C-D use moving-block bootstrap (MBB) with block lengths 5 and 22, respectively, based on 999 bootstrap replications; bootstrap standard errors are the standard deviation of bootstrap coefficients. For the bootstrap panels, significance stars are based on the empirical bootstrap P-Value rather than on a normal approximation using the reported bootstrap standard error, so star assignments need not coincide exactly with coefficient-to-standard-error ratios. All specifications include the same control block and month fixed effects as the baseline model. Standard errors are in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	(1) Inflow	(2) Outflow	(3) Net flow
Panel A: Newey–West HAC			
$WDSI^{CN}$	-0.023	-0.266***	0.361*
	(0.164)	(0.102)	(0.218)
Lag / block length	5	5	5
Controls	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Observations	2,205	2,224	2,190
Panel B: Newey–West HAC			
$WDSI^{CN}$	-0.023	-0.266**	0.361
	(0.166)	(0.132)	(0.254)
Lag / block length	22	22	22
Controls	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Observations	2,205	2,224	2,190
Panel C: Moving-block bootstrap			
$WDSI^{CN}$	-0.023	-0.266**	0.361
	(0.165)	(0.102)	(0.217)
Lag / block length	5	5	5
Controls	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Observations	2,205	2,224	2,190
Panel D: Moving-block bootstrap			
$WDSI^{CN}$	-0.023	-0.266**	0.361
	(0.178)	(0.133)	(0.274)
Lag / block length	22	22	22
Controls	Yes	Yes	Yes
Month FE	Yes	Yes	Yes
Observations	2,205	2,224	2,190

Table A.4: Supplementary Granger-Causality Results: A-H Premium and Stock Connect Flows

This table reports Granger-causality tests from residualised daily VAR(p) models between the A-H premium and Stock Connect flows (Inflow, Outflow, Net flow), for lag orders $p = 1, 3, 5$. Both series are residualised using the same baseline control block and month fixed effects as Equation (1). "P-Value: A-H→Flow" tests whether lagged A-H premium predicts the flow variable; "P-Value: Flow→A-H" reports the reverse-direction test.

Flow channel	Lag order	P-Value: A-H → Flow	P-Value: Flow → A-H	Direction at 5%
Inflow	1	0.0552	0.00000744	Flow → A-H
Inflow	3	0.3441	0.00000663	Flow → A-H
Inflow	5	0.6300	0.0000598	Flow → A-H
Outflow	1	0.000000901	0.2978	A-H → Flow
Outflow	3	0.0000314	0.3453	A-H → Flow
Outflow	5	0.000239	0.5001	A-H → Flow
Net flow	1	0.000345	0.00638	Bidirectional
Net flow	3	0.0470	0.00284	Bidirectional
Net flow	5	0.1424	0.0139	Flow → A-H