Measuring Geopolitical Risk

Dario Caldara Matteo Iacoviello*

November 15, 2021

Abstract

We present a news-based measure of adverse geopolitical events and associated risks. The geopolitical risk (GPR) index spikes around the two world wars, at the beginning of the Korean War, during the Cuban Missile Crisis, and after 9/11. Higher geopolitical risk foreshadows lower investment and employment and is associated with higher disaster probability and larger downside risks. The adverse consequences of the GPR index are driven by both the threat and the realization of adverse geopolitical events. We complement our aggregate measures with industry- and firm-level indicators of geopolitical risk. Investment drops more in industries that are exposed to aggregate geopolitical risk. Higher firm-level geopolitical risk is associated with lower firm-level investment.

KEYWORDS: Geopolitical Risk; War; Terrorism; Business Cycles; Disaster Risk; Firm-level investment; Textual Analysis; Earnings Calls; Quantile Regressions.

JEL CLASSIFICATION: C1. D80. E32. H56.

*Caldara: Board of Governors of the Federal Reserve (email: dario.caldara@frb.gov); Iacoviello: Board of Governors of the Federal Reserve (email: matteo.iacoviello@frb.gov). We thank Alessandra Bonfiglioli, Andrea Prestipino, Andrea Raffo, Bo Sun, Chris Erceg, Colin Flint, Nick Bloom, Nils Gornemann, Ricardo Correa, Robert Engle, Steve Davis, as well as seminar and conference audiences. We are grateful for the support from the GRUV (Global Risk, Uncertainty, and Volatility) network at the Federal Reserve Board and for the help from the staff of the Federal Reserve Board Research Library. We are grateful to the editor and four referees for their helpful and constructive comments. Andrew Kane, Bethel Cole-Smith, Charlotte Singer, Erin Markiewitz, Fatima Choudhary, Joshua Herman, Lucas Husted, Maddie Penn, Patrick Molligo, Sarah Conlisk, and Theresa Dhin provided outstanding research assistance. All errors and omissions are our own responsibility. The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of anyone else associated with the Federal Reserve System. The monthly data used in this paper and the replication codes can be found at https://www.matteoiacoviello.com/gpr_replication.htm. Updated data on geopolitical risk can be found at https://www.matteoiacoviello.com/gpr.htm.

Entrepreneurs, market participants, and central bank officials view geopolitical risks as key determinants of investment decisions and stock market dynamics. The Bank of England includes geopolitical risk—together with economic and policy uncertainty—among an 'uncertainty trinity' that could have significant adverse economic effects (Carney, 2016). In recent years, the European Central Bank, the International Monetary Fund, and the World Bank have routinely highlighted and monitored the risks to the outlook posed by geopolitical tensions.[1] In a 2017 Gallup survey of more than 1,000 investors, 75 percent of respondents expressed worries about the economic impact of the various military and diplomatic conflicts happening around the world.[2]

From the standpoint of many economic models, adverse geopolitical events and threats can impact macroeconomic variables through several channels, such as loss of human life, destruction of capital stock, higher military spending, or increased precautionary behavior. However, the importance of geopolitical factors in shaping macroeconomic outcomes has not been the subject of systematic empirical analysis. The main limitation has been the lack of an indicator that is consistent over time, and that measures real-time geopolitical tensions as perceived by the press, the public, global investors, and policymakers. This is the perspective we adopt here. We construct newspaper-based indexes of geopolitical risk (GPR), daily and monthly, global and country-specific, and examine their evolution since 1900. Using aggregate macroeconomic data, we then show that higher GPR increases the probability of an economic disaster and predicts lower investment and employment. Using firm-level data, we document that the adverse implications of geopolitical risk are stronger for firms in more exposed industries, and that high firm-level GPR is associated with lower firm-level investment.

The construction of our index consists of definition, measurement, and validation. Section I presents definition and measurement. We define geopolitical risk as the threat, realization, and escalation of adverse events associated with wars, terrorism, and any tensions among states and political actors that affect the peaceful course of international relations.[3] In the measurement step, we draw on Saiz and Simonsohn (2013) and Baker, Bloom, and Davis (2016), and construct the GPR index with an algorithm that computes the share of articles mentioning adverse geopolitical events in leading newspapers published in the United States, the United Kingdom, and Canada. These newspapers cover geopolitical events of global interest, often implying an involvement of the United States. That said, while the GPR index can be viewed either as a measure that is relevant for major companies, investors, and policymakers, or as a measure that is mostly relevant from

[1] These institutions keep track of geopolitical risks using our index presented here. The index is updated monthly and is available at https://www.matteoiacoviello.com/gpr.htm.

[2] See http://www.businesswire.com/news/home/20170613005348/en/.

[3] The term 'risk' is a bit of a misnomer, since it includes both the threat and the realization of adverse events. Section I explains the rationale for our naming convention.

a North American and British perspective, our validation analysis shows that our index can be further sliced into separate country-specific components, likely reflecting the different geographic imprint of major geopolitical events.

We plot the recent index, dating back to 1985, in Figure 1. The three largest spikes are recorded during the Gulf War, after 9/11, and during the 2003 invasion of Iraq. More recently, the index spikes after the Paris terrorist attacks and during the 2017-18 North Korea crisis. We also construct the daily GPR index (Figure 2) as well as the historical GPR index, dating back to 1900, which spikes at the beginning of the two world wars, as well as around D-Day, the Korean War, and the Cuban Missile Crisis (Figure 3). Elevated readings of the index reflect the realization or escalation of *current* adverse events, as well as expectations and threats about *future* adverse geopolitical events. To quantify these two components, we construct the geopolitical acts index and the geopolitical threats index, shown in Figure 4.

In Section II we present a variety of checks that verify the plausibility of the GPR index and compare the index with related economic and geopolitical indicators. In addition to performing a formal audit of a sample of 7,000 newspaper articles, we verify that our automated index is highly correlated with a narrative counterpart constructed by manually scoring the 44,000 front pages of *The New York Times* published from 1900 through 2019. Moreover, we show that spikes in our index and its components highlight well-known historical episodes associated with wars, terrorism, or international crises. Based on these exercises and other robustness checks, we conclude that the GPR index is meaningful and accurate.

In Sections III and IV, we look at the macroeconomic effects of geopolitical risk. For the United States, using vector autoregressive (VAR) models for the period 1985 to 2019, we find that a shock to geopolitical risk induces persistent declines in investment, employment, and stock prices, with the decline in activity due to both the threat and the realization of adverse geopolitical events. In addition, using cross-country data and country-specific indexes spanning 120 years, we find that higher values of the GPR index are associated with: (1) higher probability of economic disasters, (2) lower expected GDP growth, and (3) higher downside risks to GDP growth.

In Section V, we provide further evidence on the implications of geopolitical risk using industry and firm-level data. The aggregate GPR index correlates well with listed firms' own perceptions of geopolitical risks, which we construct from mentions of geopolitical risks in 135,000 firms' earnings calls, inspired by Hassan et al. (2019). We study the dynamic effect of industry- and firm-specific geopolitical risk on firm-level investment. Industries that are positively exposed to geopolitical risks suffer a decline in investment that is larger than the aggregate effect. Idiosyncratic geopolitical risk—constructed using the transcripts of firms' earnings calls, and purged of aggregate and industry-specific components—is associated with lower investment at the firm level, with effects that accumulate and persist over time.

Our paper makes three contributions. First, we develop a new measure of adverse geopolitical events. Around some key dates, the GPR index shares some of its spikes with the military spending news variable of Ramey (2011), with indicators of the human cost of conflicts, with the Economic Policy Uncertainty (EPU) Index of Baker, Bloom, and Davis (2016), and with financial volatility. However, the GPR index also captures important information about geopolitical events that is not reflected in these indicators. Second, we distinguish the threats of adverse geopolitical events from their actual realization.[4] We do so because our methodology pinpoints the timing of different types of geopolitical events, thus allowing measurement of their effects.[5] Third, we present new systematic evidence on the role of adverse geopolitical events in business fluctuations, using quarterly VARs, cross-country historical data, and firm-level data.

I Construction of the Geopolitical Risk Indexes

The construction of geopolitical risk indexes involves definition, measurement, and validation. We first describe the definitions of geopolitics and geopolitical risk adopted in our paper. We then discuss how we measure geopolitical risk and describe the key features of the resulting indexes.

I.A Definition of Geopolitical Risk

Formally, geopolitics is the study of how geography affects politics and the relations among states (Foster, 2006 and Dijkink, 2009). By contrast, the popular usage of the term geopolitics is more complex and contested, ranging from narrow to broad definitions of what constitutes geography and who the relevant political actors are. In *A Dictionary of Human Geography*, Rogers, Castree, and Kitchin (2013) state that the media often refer to geopolitical concerns to describe the impact of international crises and international violence. This is the perspective we adopt here.

We define geopolitical risk as the threat, realization, and escalation of adverse events associated with wars, terrorism, and any tensions among states and political actors that affect the peaceful

[4] A growing literature studies the distinction between expectations and realizations of macroeconomic and financial phenomena. Bloom (2009) controls for the level of the stock market when identifying shocks to financial uncertainty. Berger, Dew-Becker, and Giglio (2019) find that expectations about future volatility are not contractionary after controlling for current volatility.

[5] Ludvigson, Ma, and Ng (2021) and Caldara et al. (2016) study the relationship between economic uncertainty and the business cycle by controlling for financial and economic activity when identifying uncertainty shocks. Our emphasis on geopolitical risk also links our paper to the literature on disaster risk. See for instance Barro (2006), Gourio (2008), Berkman, Jacobsen, and Lee (2011), Pindyck and Wang (2013), and Nakamura et al. (2013).

course of international relations.

Two considerations about our definition are in order. First, our definition of geopolitical builds on the historical usage of the term—to describe the practice of states to control and compete for territory (Flint, 2016). However, in line with recent assessments of modern international relations, our definition also includes power struggles that do not involve acts of violence and competition over territories, such as the Cuban Missile Crisis or recent tensions between the U.S. and Iran, or the U.S. and North Korea. Our definition also includes terrorism. In recent decades, terrorist acts have generated political tensions among states and, in some instances, have led to full-fledged wars.

Second, our definition of geopolitical risk captures—with a slight abuse of the word 'risk'—a wide range of adverse geopolitical events, from their threat, to their realization, to their escalation. This choice is dictated by journalistic practices and measurement considerations. Regarding journalistic practices, in naming our index, we followed a tradition in the media that refers to geopolitical risks as a catchall phrase to describe the effects of international crises and violence, actual or perceived (Rogers, Castree, and Kitchin, 2013). Regarding measurement considerations, our extensive reading of news coverage on wars, terrorism, and international crises over the past 120 years revealed that the threat, realization, and escalation of international violence are often intertwined, so that a headline measure that abstracts from one of these components may not capture the range of events that could be of interest to researchers. That said, we break the headline index into separate 'acts' and 'threats' components, so that interested researchers can choose their preferred components for downstream empirical applications.

I.B Measurement

Our sample is the text contained in about 25 million news articles published in the print edition of leading English-language newspapers from 1900 through the present, corresponding to about 30,000 and 10,000 articles per month in the recent and historical sample, respectively. We construct the GPR index by counting, each month, the share of articles discussing adverse geopolitical events and associated threats. The recent GPR index starts in 1985 and is based on automated text-searches on the electronic archives of 10 newspapers: the *Chicago Tribune*, the *Daily Telegraph*, the *Financial Times*, *The Globe and Mail*, *The Guardian*, the *Los Angeles Times*, *The New York Times*, *USA Today*, *The Wall Street Journal*, and *The Washington Post*. The choice of six newspapers from the U.S., three from the United Kingdom, and one from Canada reflects our intention to capture events that have global dimension and repercussions.[6] The index counts, each month, the number

[6] These newspapers have high circulation throughout the sample, consistent coverage of international political events, and digital archives that span a long period. In Section II we verify that an index that excludes non-U.S. newspapers is very similar to the benchmark index.

of articles discussing rising geopolitical risks, divided by the total number of published articles. By the same token, the historical GPR index, dating back to 1900, is based on searches of the historical archives of the *Chicago Tribune*, *The New York Times*, and *The Washington Post*.

To construct our outcome of interest, we use a dictionary-based method, specifying a dictionary of words whose occurrence in newspaper articles is associated with coverage of geopolitical events and threats. Such a method organizes prior information about how features of a text (e.g. the occurrence in newspaper articles of the words 'war' and 'threat' within close proximity) map into the outcome of interest (e.g. news coverage of geopolitical risks). The use of supervised or unsupervised algorithms or pre-specified dictionaries are less applicable to our case as the outcome of interest is not directly observed and there are no readily available data to train a supervised model.[7]

How do we specify the information that guides the construction of the dictionary? First, we build directly on the definition of geopolitical risk adopted in this paper, selecting words that closely align with our definition. Second, we use information from two geopolitical textbooks and from the Corpus of Historical American English to isolate themes that are more likely to be associated with geopolitical events (such as 'war [on] terror' or 'nuclear weapon') or words that are more likely to be used in conjunction with war-related words (such as 'declare'). Third, we organize the search around high-frequency words and their synonyms that are more likely to appear in newspapers on days of high geopolitical tensions (see Tables A.1 and A.2 in the Appendix). For instance, the word 'crisis' has a relative term frequency of 0.25 percent on days of high geopolitical tensions compared to 0.04 percent on an average day. Words very likely to appear in newspapers on days of high geopolitical tensions include 'terror,' 'blockade,' 'invasion,' 'troops,' and 'war.'

Our goal is to provide an index that can highlight distinct aspects of geopolitical risk, and that can be sliced conceptually and geographically. Doing so exclusively with one-word searches would likely lead to misclassification and measurement error. These considerations lead to our search query, which specifies two words or phrases whose joint occurrence likely indicates adverse geopolitical events. The query is described in Table 1, and is organized in eight categories (see Panel A). Each category is captured by a search query comprising two sets of words, the first set containing topic words (e.g. 'war,' 'nuclear,' or 'terrorism'), the second set containing 'threat' words for categories 1 through 5 and 'act' words for categories 6 through 8. For six of our categories, we run proximity searches (e.g., searching for 'terrorist' and 'risk' appearing within two words of each other). For two categories, we search for either two words appearing in the same article ('weapons' and 'blockade') or for one bigram and one word appearing in the same article ('nuclear war' AND 'threat'). We do plenty of robustness analysis around this search strategy (discussed in Section II)

[7] See Gentzkow, Kelly, and Taddy (2019) for a detailed comparison of methods for text analysis.

and verify that, in our application, this approach yields better outcomes relative to a search using bigrams only, as in Hassan et al. (2019), or using Boolean operators only, as in Baker, Bloom, and Davis (2016), who search 'economic' and 'policy' and 'uncertainty' terms.

Panel B of Table 1 describes the sets of words constituting our dictionary. For each category, we started from a minimal set of 'core words,' denoted in red. For instance, for category 1 the two core words are 'war' and 'conflict.' For category 2, the core word is 'peace.' For category 3, the core words are 'military' and 'troops.' Core words that indicate threats are 'threat,' 'warn,' 'fear,' 'risk,' and 'concern.' These sets of words are the most common words used in news coverage to discuss war-related threats. As shown in Section II, exclusive reliance on these core words, while resulting in an index that shares a similar contour to our final index, would lead to searches that fail to capture several articles that discuss geopolitical events and risks. For this reason, we add words that are used throughout our historical sample to cover multiple episodes. For instance, news coverage of military buildups, embargoes, and sanctions (such as during the Cold War, the Cuban Missile Crisis, or the run-up to the Gulf War) relies on words that are not included in the core set. Threats to peace are often referred to as 'disruptions' of peace, a word that is not used to directly indicate war threats. For the nuclear threats category, we use bigrams to reduce the possibility that articles related to civilian usage of nuclear technologies would slip into our search. Finally, the bottom panel lists 'excluded words' that our audit revealed to be more frequently associated with false positives. Articles that mention these words cover a diverse set of topics, such as movies and books, sport events, war anniversaries, and obituaries of famous generals and politicians. The excluded words do not affect the spikes in our index. Nonetheless, accounting for these words mitigates spurious trends and reduces the share of false positive articles in the index (see Table A.3 in the Appendix).

I.C The Recent GPR Index

Figure 1 presents the GPR index from 1985 through 2020 based on 10 newspapers. The index is characterized by several spikes corresponding to key adverse geopolitical events. The first spike is recorded in April 1986 and corresponds to the terrorist escalation that led to the U.S. bombing of Libya. The second spike happens around the Iraq invasion of Kuwait and the subsequent Gulf War. The index surges at the beginning of 1993, during a period of escalating tensions between the United States and Iraq. It then trends downwards until 2001 when it surges after the 9/11 events, before spiking again during the 2003 invasion of Iraq. In recent years, the index is high during the 2011 military intervention in Libya, around the 2014 Russian annexation of the Crimea peninsula, and after the 2015 Paris terrorist attacks. The index displays a break in its mean after

2001. The 9/11 terrorist attacks saw a shift in news coverage of geopolitical events, driven by increased reporting on terrorist threats and on the war on terror.[8]

Figure 2 shows the GPR index at daily frequency. The daily index is noisier than its monthly counterpart but provides a detailed view of a larger set of episodes, including those that may seem to be missed by the monthly index. For instance, in August 1991, the daily index captures the escalation of ethnic violence in the former Yugoslavia, and the attempted coup in the Soviet Union. In March 1999, the index spikes at the beginning of the NATO air strikes in Kosovo. These events have a low bearing on the monthly index, as the associated news coverage was short-lived.

The daily GPR index illustrates how the unfolding of geopolitical tensions can add up to elevated values in its monthly counterpart. In a first scenario, a protracted buildup in tensions leads to a defining event causing a big spike in the index, as in the case of the Gulf War. In a second scenario, one climactic event causes a large spike in daily geopolitical risk and is followed by readings that are persistently higher than the average, as in the aftermath of the 9/11 terrorist attacks. In a third scenario, slow-moving geopolitical tensions persistently remain in the news cycle, averaging out to elevated values in the monthly GPR. Examples include the Syrian Civil War and the 2017-18 North Korea crisis. In all these scenarios, spikes in the daily index correctly point to when tensions materialized, thus bolstering evidence of the informative content that the index produces at daily frequencies. That said, it is possible that our index may not appropriately measure episodes that slowly unfold over multiple years, such as the fall of Communism in the Soviet Union and Eastern Europe, and are recognized as geopolitical risks only with the benefit of hindsight.

I.D The Historical GPR Index

Figure 3 displays the historical GPR index from 1900 onward. The historical index closely mimics the recent index during the period 1985 to 2020 when their coverage overlaps, with a correlation of 0.95. The historical GPR index is higher, on average, during the first half of the 20th century (see summary statistics in Table A.3 in the Appendix).

Perhaps unsurprisingly, the highest readings of the index coincide with the two world wars. The index spikes at the onset of World War I and World War II and remains persistently high during each war. The index declines rapidly at the end of World War II only to rise again during the Korean War. The second half of the 20th century witnessed several geopolitical threats and

[8] We perform a supremum Wald test for structural break at an unknown date using symmetric trimming of 15%. We reject the null of no break in the log of the GPR index (p-value of < 0.001) and find a break in September 2001. Higher news coverage of geopolitical risks after 9/11 may indicate either an increase in actual risks of wars and terrorism, or an increase in the public perception of these risks. An important question for future research would be to study the relative importance of perceived versus actual geopolitical risks for economic outcomes.

crises. For instance, the index spikes during the Suez Crisis, the Cuban Missile Crisis, the Six-Day War, and the Falklands War. The index stays at relatively high levels from the 1950s through the mid-1980s, a time when the threat of nuclear war and geopolitical tensions between countries were more prevalent than actual wars. As discussed, since the 2000s, terrorism, the Iraq War, and rising bilateral tensions dominate the index.

I.E Geopolitical Threats and Geopolitical Acts

Throughout history, the realization of adverse geopolitical events has often been the catalyst for increased fears about future adverse events. For instance, terrorist attacks may increase the threat of future attacks or of a war. Our search query and the resulting GPR index capture both the realization of adverse geopolitical events (a terrorist attack or the outbreak of a war), and threats about the future adverse events.

We construct two components of the GPR index, the geopolitical threats (GPT) and the geopolitical acts (GPA) indexes. The GPT index searches articles including phrases related to threats and military buildups (categories 1 through 5 in Table 1), while the GPA index searches phrases referring to the realization or the escalation of adverse events (categories 6 through 8 in Table 1). Figure 4 plots the two indexes since 1900. The GPT and GPA indexes have a correlation of 0.59 over the full sample, and of 0.45 from 1985 onward. Even if some spikes in the two indexes coincide, there is also independent variation that is better highlighted when examining particular historical episodes. The beginning of World War I appears largely unexpected. Throughout the war, the GPA index remains elevated while the GPT index remains subdued, although a spike in threats when the U.S. severs diplomatic relations with Germany in February 1917 is followed by the American entry into World War I two months later. The buildup to World War II sees the GPT index rise amid news coverage of the risk of war, for instance during the annexation of Czechoslovakia by Nazi Germany, whereas the GPA index spikes at the beginning of the war, after Pearl Harbor, and around D-Day. By contrast, the 1960s witnessed international crises captured by spikes in the GPT index that did not lead to wars such as the Berlin crisis and the Cuban Missile Crisis. The GPT index surges in 1990 in the run-up Gulf war. The GPA index spikes after 9/11 and at the beginning of the Gulf War. Finally, the GPT index is high relative to its historical average during the recent tensions between the U.S. and North Korea and Iran.

II Validation of the Index

This section presents three exercises aimed at ensuring the validity of our indexes. First, we verify that the GPR indexes provide a plausible quantification of the historical and geographical evolution of geopolitical risks. Second, we compare the indexes with similar economic and geopolitical data. Third, we summarize the audit process and additional accuracy checks.

II.A Plausibility

Largest Spikes in the Historical Index. Our first plausibility test relies on the logic that jumps in the index must capture the most important geopolitical risks of the past 120 years, in the way these risks were perceived by the contemporaries.[9] We calculate surprises in the index and in its two main subcomponents as the residuals of a regression of the relevant monthly indexes on three of their own lags.

Table 2 illustrates that the relative magnitude of the historical jumps in the index is reasonable. The largest shocks capture well-known episodes of sizable increases in the risk associated with wars, terrorism, or international crises. The five largest shocks are the beginning of both world wars, 9/11, Pearl Harbor, and the onset of the Korean War. Some of these events illustrate examples of shocks to both the threat and act components of the index. Other shocks, such as the Cuban Missile Crisis or the Gulf War, weigh more heavily on either component, showcasing the independent role played by threats and acts in the construction of the index. For example, the Cuban Missile Crisis ranks fourth among the largest threats within the past 120 years despite its official duration of only 13 days and the lack of public attention that it garnered within its first week.

Comparison with a Narrative GPR Index. Traditionally, a newspaper's front page gives the reader a summary of the most important news event of the day in order of importance, with editors always ready to break out big headlines for the most important stories. As a second check for the plausibility of the index, we compare it with a 'narrative' index of adverse geopolitical events that we constructed by reading and scoring the headlines of 44,000 front pages of the print edition of *The New York Times* from 1900 through 2019.[10]

[9] One example of a possible discrepancy between contemporaries' perception of risks and ex post perception is given by the Cuban Missile Crisis. With hindsight, it is reasonable to claim that the dangers posed by the crisis were far greater than the contemporaries understood. See for instance Sherwin (2012).

[10] The front page of *The New York Times* has changed dramatically over time. A typical front page in 1900 had four times as much text as today, as well as more articles. Early on, the subject in the front page was mostly domestic and international politics. Today, the front page covers a larger variety of topics including finance, family, technology, and medicine. See The New York Times Public Editor (2004). That said, the front page and its headlines have always directed the reader to the most important issues of the day.

Together with a team of research assistants, we read all headlines above the fold of the front page of *The New York Times*, and assign to each day a score of a 0, 1, 2, or 5 depending on whether: no headline features rising or existing geopolitical tensions (score: 0); one headline, but not the lead headline, features GPR (score: 1); the lead headline, but not a banner headline, features GPR (score: 2); the banner headline features GPR (score: 5).[11] The resulting narrative index places heavy weight on the importance of the article, as reflected by its placement in the newspaper, and adequately captures the tone of the event. Additionally, the narrative index, not relying on a preset list of words, is unlikely to be affected by changes in language over time.

The narrative index is plotted in Figure 5 alongside our automated one. The two indexes share very similar long-run trends and display a very high correlation of 0.86, sharing very similar spikes during the world wars and in the wake of the Korean War, the Gulf War, and 9/11. This positive correlation bolsters our confidence that the automated index is an accurate measure of geopolitical risks. We consider the automated index to be a better benchmark relative to the narrative for three main reasons. First, the automatic index enhances transparency and replicability. Second, the narrative index relies only on the front page articles of one newspaper thereby rendering scaling up and maintenance costly. Third, the narrative index may suffer more from mismeasurement due to limited front-page space (e.g. major concurrent events crowd out front-page space so other relevant events are pushed elsewhere in the newspaper) and ambiguity of historical records (thereby requiring difficult judgment calls).

Country-Specific Measures of Geopolitical Risk. We construct country-specific measures of geopolitical risk by counting joint occurrences in newspapers of geopolitical terms and the name of the country (or its capital or main city) in question. For instance, the GPR index for Japan is the share of articles that meet the criterion for inclusion in the GPR index and that contain the words 'Japan' or 'Tokyo.' The geographical disaggregation permits a more granular assessment of the index, quantifying exposure of countries to global risks and highlighting geopolitical episodes that, while relevant for individual countries or regions, receive little weight in the aggregate index. Importantly, the resulting indexes, being constructed using three U.S. newspapers, capture the U.S. perspective on risks posed by, or involving, the country in question.

Figure 6 plots country-specific GPR indexes for selected countries. Most countries share exposure to common geopolitical events, most notably the two world wars and, more recently, the Gulf War and Iraq War. That said, a few spikes are isolated to specific countries or regions. After World War II, the U.K. was involved in several international crises, ranging from the dispute with Egypt over the Suez Canal to the war against Argentina for control over the Falkland Islands. Germany faced

[11] The weights are chosen to be roughly proportional to the space taken by the headline across the page.

a major crisis that culminated in the construction of the Berlin Wall in 1961. Japan, Russia, and China were opposed in regional wars in the first half of the 20th century. Mexico and Korea were each embroiled in two major wars that saw the direct involvement of the United States.

II.B Comparison with Related Economic and Geopolitical Data

Comparison with News about Military Spending. The top panel of Figure 7 compares the historical GPR index with Ramey's (2011) measure of news about U.S. military expenditures constructed from historical records. Ramey's series reports the present discounted value of expected changes in defense expenditures constructed, akin to our measure, using news from *Business Week* and other newspaper sources. The two measures are clearly related, with a correlation of 0.29 over the period 1900:Q2 to 2016:Q4. The GPR index is above its historical mean in 15 out of the 16 instances in which the military spending news variable is larger than 5 percent of GDP. The two measures also display independent variation driven by spikes in the GPR index unrelated to U.S. military spending (see Figure A.1 in the Appendix), such as during both world wars, throughout the Korean War, and in the years following 9/11.

Comparison with War Deaths. Our index assumes that the propensity to discuss a phenomenon in newspapers can be seen as an ordinal measure of the intensity of that phenomenon, and is monotonically increasing in the phenomenon itself. Figure 7 shows that the GPR index is positively correlated with worldwide deaths from conflicts, a cardinal—albeit crude—measure of the risks posed by armed conflicts. The correlation coefficient between the two measures is 0.82. War deaths correlate more with GPR acts (0.83) than with GPR threats (0.46). The GPR index and deaths from conflict surge together during the two world wars, but their correlation weakens after the 1950s. Of note, the level of the GPR index has been higher almost every year since the end of World War II compared to any year during the interwar period, whereas deaths have stayed at relatively low levels. It is no surprise that the level of the index appears permanently higher after the world wars made humanity more attentive to the risks posed by armed conflicts.

Comparison with Proxies for Uncertainty and Granger Causality Tests. Figure 8 compares the recent GPR index with two popular measures of uncertainty: the VIX—a measure of stock market volatility—and the news-based EPU index of Baker, Bloom, and Davis (2016). There are two periods where all three indexes rise simultaneously: in 1990-91, around the time of the Gulf War, and in 2001, after the 9/11 terrorist attacks. However, in both cases it seems plausible to argue that the causation runs from geopolitical events to stock market volatility and policy uncertainty. The three indexes also exhibit sizable independent variation. The GPR index does not move during periods of economic and financial distress or around presidential elections, periods characterized by elevated policy uncertainty. By contrast, rises in the EPU index and VIX do not coincide with the Russian annexation of Crimea or with terrorist events other than 9/11. In sum, the graphical evidence indicates that, compared to the VIX and the EPU index, the GPR index appear to capture—because of its own nature—events that (i) are less likely to have an economic origin, and (ii) could give rise to heightened financial volatility and policy uncertainty.[12]

Appendix B.5 shows that the GPR index is not Granger caused by news related to recent developments in the United States. We regress the log of the GPR index on macroeconomic variables (change in U.S. industrial production, private employment, and the log of the WTI price of oil deflated by U.S. CPI); financial variables (real returns on the S&P500 index and the two-year Treasury yield); and proxies for uncertainty (the VIX and the log of the EPU index). Macroeconomic, financial, and uncertainty developments do not Granger-cause the GPR index.

II.C Additional Checks

Audit. We evaluate the GPR index against alternatives based on different search queries and we perform an extensive human audit of newspaper articles likely discussing geopolitical risks.

In the first exercise, we use the narrative index—constructed using *The New York Times* front pages as discussed in Subsection II.A—as a reference point for assessing the accuracy of the benchmark index. Specifically, we compare the benchmark index with three alternatives based on slight modifications of the search query of Table 1. The alternative indexes: (1) do not remove the 'excluded words' from the query; (2) are based on a smaller set of 'core words'; (3) use the Boolean operator 'AND' for all search categories (as opposed to a search of terms within two words from each other). We find that the GPR index exhibits a higher correlation with the narrative index than the three alternative indexes (see Table A.3 for details). Additionally, for each index, we randomly sample a large number of articles, read each of them, and manually code them as either discussing high or rising geopolitical tensions or not. We find the GPR index has a lower Type-I error rate relative to all alternatives.[13]

In the second exercise, we follow the approach of Baker, Bloom, and Davis (2016) and evaluate the GPR index through a human audit that further confirms the validity of the article selection process. The GPR index has a correlation of 0.93—at an annual frequency—with a 'human' GPR

[12] In Appendix B.10, we compare the GPR index to other quantitative proxies: the ICB database, the National Security EPU Subindex, and the U.S. External Conflict Rating Index.

[13] The GPR index trends slightly downward from 1900 onward, a plausible feature given the two world wars and the Korean War in the early part of the sample.

index that is constructed by manually reading and coding a sample of more than 7,000 newspaper articles (see Appendix B.6 for additional details).[14]

Are Results Sensitive to the Use of Different Newspapers? The recent and historical GPR indexes rely on 10 and 3 newspapers, respectively. This choice avoids reliance on one particular news source and provides a robust and stable account of geopolitical risks. We find that the exact number of newspapers has only a modest effect on the index (see also Appendix A.2). The correlation between the historical index and the recent index is 0.95 for the period in which the two indexes overlap. Additionally, the correlation between non-U.S. and U.S. newspapers GPR is 0.88, thus suggesting that the global nature of most geopolitical events receives similar coverage across U.S. and non-U.S. newspapers. Finally, the Cronbach alpha, a measure of internal consistency across indexes based on the 10 individual newspapers, is 0.96, a number that indicates an excellent degree of reliability of our measure.

Does War Language Change over Time? The construction of our index relies on an extensive analysis of the most common words and sentences used in newspapers over time to describe risks of war and risks to peace, and acts of war and terror. We offer a detailed description of this analysis in Appendix B.7, where we confirm that we neither ignore nor over-rely on words used relatively more often in some historical periods. First, we verify that we do not omit any crucial, war-related words that are used *relatively* frequently in newspapers during selected episodes of elevated geopolitical tensions. In particular, words such as terrorism, blockade, invasion, war, crisis, troops, and threat, among others, have odds of appearing in newspapers on days of high geopolitical risk that are at least five times higher relative to any average day (see Tables A.1). Second, we analyze term frequency for the words and word combinations used to construct the index and study their evolution over time. Tables A.4 and A.5 confirm that our query includes both words that are more frequent in the early part of the 20th century, such as 'menace' or 'peril', and words that are more frequent in recent decades, such as 'risk' or 'tension'.

As a final consideration, we recognize that newspapers appear to have devoted increasingly more space to arts, history, sports, and entertainment, often borrowing some of their language from warfare and military terminology. For this reason, our search ignores the articles containing the 'excluded words' of Table 1. Without these words, the index would have a slight upward trend throughout the historical period, and slightly higher measurement error (see Table A.3).

[14] Saiz and Simonsohn (2013) list a number of formal conditions that must hold to obtain useful document frequency-based proxies for variables and concepts that are otherwise elusive to measure, such as ours. In Appendix B.9, we show that our index satisfies the Saiz and Simonsohn (2013) conditions.

Does Media Attention Measure the Underlying Risk? An implicit hypothesis of our analysis is that the propensity to mention geopolitical risks in newspapers is representative of such propensity in the wider population. While a formal test of this hypothesis would be beyond the scope of this paper, our Appendix provides evidence that the GPR index is not unduly affected by issues related to how the media reports the news. First, we show that the index is not prone to spurious fluctuations when geopolitical events could be crowded out by unpredictable or predictable newsworthy events—from natural disasters to inflation to Olympic Games to presidential elections (see Figure A.2 and Table A.6 in the Appendix). Second, we verify that our index is not impacted by the political orientation of the newspapers used in the analysis (see Appendix Figure A.3). Finally, we show that there is a high correlation between occurrence and extent of murders, hijackings, and nuclear tests on the one hand, and the media coverage of these events on the other. This correlation suggests that, even if these events share with geopolitical news an alarmist message that may sell more newspapers, their occurrence is in line with the media coverage (see Appendix Figure A.4).

III VAR Evidence on the Effects of Geopolitical Risk

In this section, we present our investigation of the relationship between the GPR index and aggregate economic activity in the United States using VAR models for the period 1985 to 2019.

III.A Aggregate Economic Effects

We examine the macroeconomic consequences of innovations to geopolitical risk using a structural VAR model (details and robustness analysis are in Appendix C). Our main specification—which we estimate using two lags and quarterly data from 1986:Q1 through 2019:Q4—consists of eight variables: (1) the log of the GPR index; (2) the VIX; (3) the log of real business fixed investment per capita; (4) the log of private hours per capita; (5) the log of the Standard and Poor's 500 index; (6) the log of the West Texas Intermediate price of oil; (7) the yield on two-year U.S. Treasuries; (8) the Chicago Fed's National Financial Conditions Index (NFCI).[15]

We identify a GPR shock by using a Cholesky decomposition of the covariance matrix of the VAR reduced-form residuals, ordering the GPR index first. The ordering implies that any contemporaneous correlation between economic variables and the GPR index reflects the effect of the GPR index on the economic variables, rather than the other way around. The characteristics of the GPR index discussed in the previous two sections lend support to this assumption. We explore robustness to alternative identification assumptions and VAR specifications in the Appendix.

[15] The stock market index and oil prices are divided by the Consumer Price Index for All Urban Consumers.

The solid lines in Figure 9 show the median impulse responses to a two standard deviation shock to the GPR index.[16] The size of the shock reflects the average of the innovations in the right 10 percent tail of the GPR shock distribution. The GPR index rises persistently and remains elevated for nearly two years. High geopolitical risk is followed by a short-lived increase in financial uncertainty as measured by the VIX, by a decline in stock prices and oil prices, and by a modest decrease in the two-year yield. Fixed investment gradually declines, bottoming out at negative 1.5 percent after about one year, before slowly reverting to trend. Labor market conditions deteriorate, with hours declining 0.6 percent one year after the shock. The decline in investment and hours in the wake of a GPR shock is broadly consistent both with models that emphasize the contractionary effects of future negative news about the future—as in Beaudry and Portier (2006) and Jaimovich and Rebelo (2009)—and with models where recessions are driven by shocks with a negative first moment and a positive second moment—such as Bloom et al. (2018).[17]

III.B Acts and Threats

Next, we evaluate the difference between innovations in the two broad components of the GPR index, the GPA index (geopolitical acts) and the GPT index (geopolitical threats). We modify the benchmark VAR by replacing the GPR index with the GPA and GPT indexes, using a Cholesky ordering with the GPA and GPT indexes ordered first and second, respectively. This ordering captures a specific configuration of shocks such that 'GPA shocks' can prompt a contemporaneous comovement in acts and threats, whereas 'GPT shocks' capture threats that do not immediately materialize, leaving acts unchanged within the month.[18]

The solid lines in Figure 10 plot the median responses to the GPA and GPT shocks. A shock to acts leads to a sharp and significant increase in threats, whereas shocks to threats lead to a small and short-lived increase in acts. GPA and GPT shocks induce similar declines on investment and hours, though the effects of GPA shocks are more persistent.

To better quantify the role of acts and threats in affecting macroeconomic variables, we construct a counterfactual set of impulse responses for the two VAR shocks in which threats are held constant in response to act shocks, and vice versa. Specifically, in response to the GPA and GPT shocks, we

[16] Figure A.6 in the Appendix plots the estimated shocks to the GPR index and to its components both for the VAR specification of this subsection and the VAR specification of subsection III.B.

[17] When we add GDP to the VAR, we find that GDP drops 0.3 percent over the first year in response to a two standard deviation geopolitical risk shock (see Appendix Figure A.7).

[18] An alternative identification scheme in which 'threats' are ordered before 'acts' would have the unpalatable property that both GPT and GPA shocks move the GPA on impact, thus making it difficult to isolate historical events when the threat component of the index moves substantially without a contemporaneous movement in acts, such as the Cuban Missile Crisis or the recent U.S.–North Korea and U.S.–Iran tensions.

select a sequence of GPT and GPA shocks that hold GPT and GPA constant, respectively. The dashed lines in Figure 10 illustrate that both acts and threats in isolation produce contractionary effects. Were threats to remain unchanged in response to an acts shock, the response of investment and hours would be smaller, thus supporting the notion that unrealized threats about future events could have contractionary effects. This result is corroborated by the decline in activity associated with increases in threats, keeping acts unchanged.

The contractionary consequences of the threats of adverse events support the insights of theoretical models where agents form expectations using a worst case probability, as in Ilut and Schneider (2014), or models where the threat of adverse events leads agents to reassess macroeconomic tail risks, as in Kozlowski, Veldkamp, and Venkateswaran (2018). Of course, these findings may well depend on the country and the period that are studied in our VAR. With the notable exception of 9/11, most adverse geopolitical events in the sample did not directly hit the United States. By contrast, it is well-known that countries experiencing adverse geopolitical events, wars in particular, on their soil suffer very large drops in economic activity, as documented by Barro (2006) and Glick and Taylor (2010). We return to this theme in the next section.

IV Tail Effects of Geopolitical Risk

In this section, we quantify the relationship between geopolitical risk—a non-economic risk—and risks to economic activity. We first show that high geopolitical risk is associated with a higher probability of economic disasters. We then show, using quantile regressions, that elevated geopolitical risk is associated with lower expected GDP growth and higher downside risks to GDP growth. We exploit variation in geopolitical risks and economic activity over time and across nations, using annual data for 26 countries for the period 1900 to 2019. We measure geopolitical risk using both the historical GPR index and the country-specific indexes described above. The main advantage of using the country-specific indexes is to exploit episodes of higher geopolitical risk that are important for individual countries but that receive a low weight in the aggregate index. For instance, country-specific geopolitical risk is extraordinarily high for Korea in the 1950s, for Chile in 1973, and for Argentina and Peru in 1982, all of which are episodes that saw foreign involvements and that contributed to geopolitical tensions in Asia and South America.

IV.A Effects on Disaster Probability

We model the occurrence of disaster $D_{i,t}$ in country i in year t as given by:

$$D_{i,t} = \alpha_i + \beta \ GPR_t + \gamma \ GPRC_{i,t} + \delta \ \Delta GDP_{i,t-1} + controls + u_{i,t}, \tag{1}$$

where $D_{i,t}$ is a 0/1 dummy for an economic disaster, α_i is a country-fixed effect, GPR is the 'global' GPR index, $GPRC$ is the country-specific index, and ΔGDP is real GDP growth. To measure $D_{i,t}$, we use the disaster dummy constructed in Nakamura et al. (2013) using an approach that generates endogenous estimates of the timing and length of an economic disaster. We update their estimation with data through 2019.[19]

The first five columns of Table 3 show results from different specifications of equation (1). All models are estimated using a linear probability specification to simplify the interpretation of the coefficients, but the results are largely unchanged when using a logistic specification. The simplest specification in column 1 has no country-fixed effects and does not control for country-specific risk. The coefficient on global GPR is economically large. It indicates that a one standard deviation increase in global geopolitical risk increases the probability of disaster by 18 percentage points.[20] Column 2 adds country fixed effects as well as country-specific GPR. After controlling for global factors, a one-standard deviation rise in country-specific GPR increases the disaster probability by 9 percentage points. Column 3 illustrates the important role played by the two world wars in driving the relationship between the (global) GPR and disaster probability. When the world war dummies are added to the specification, the coefficients on both (global) GPR index and war dummies are positive but not statistically significant, while the impact of country-specific GPR remains large and significant. While many economic disasters of the 20th century took place during the two world wars, geopolitical risks and the associated economic consequences materialized through history and across countries.

Column 4 replaces GPR with a variable measuring spikes in the index with nearly unchanged results. Column 5 controls for U.S. military spending news and allows for a common shift in the disaster probability across three subsamples, as in Nakamura et al. (2013): one before 1946, one for the period 1946 to 1972, and one for the period since 1973. The association of geopolitical risk

[19] We use the codes in Nakamura et al. (2013) to extend the estimation of the disaster events through 2019. Our procedure reproduces their disaster dates almost exactly, with a tetrachoric correlation coefficient between our disaster dummy and theirs of 0.99. China and Russia are not part of their sample, but we include them for their role in the geopolitical events of the period. We define disaster years in China the periods 1940-1946 and 1960-1968. We define disaster years in Russia the periods 1914-1920, 1941-1945, and 1990-1995.

[20] The share of disaster events in the sample is 17 percent. Sample average GDP growth is 2.9 percent in the non-disaster state, -0.2 percent in the disaster state.

with occurrence of disaster is only slightly attenuated. Finally, in columns 6 and 7 we follow the approach in Bazzi and Blattman (2014), replacing the disaster dummy with a dummy equal to 1 either at the onset or at the end of a disaster, and 0 otherwise.[21] Column 6 shows that disasters are more likely to start, rather than occur and persist, at times of high geopolitical risk. A one-standard deviation increase in country-specific geopolitical risk brings the probability of disaster onset from its historical mean of about 2.2 percent to 9 percent, an increase of 6.8 percentage points. Column 7 shows that high geopolitical risk also reduces the probability of the ending of a disaster, though the effects are smaller and more imprecise.

The evidence in this subsection supports the idea that, historically, changes in geopolitical risk are associated with substantial variations in the probability of large declines in economic activity. Many economic disasters of the 20th century took place during the world wars, the two global events in our sample. However, our estimates also demonstrate that regional and country-specific geopolitical events were associated with major economic crises.

IV.B Quantile Effects of Geopolitical Risk

Throughout history, wars have at times destroyed human and physical capital, shifted resources from productive to less productive uses, and diverted international trade. At other times, wars have enabled larger labor force participation, better technological diffusion, and larger infrastructure spending (see Stein and Russett, 1980). We use cross-country data and quantile regressions to evaluate how geopolitical risk is associated with the distribution of future economic growth. Suppose for instance that conflict is followed in some cases by faster, in some cases by slower growth, like in the United States and Germany during World War II, respectively. If that is the case, geopolitical risks may be associated with different outcomes at the low and high ends of the GDP growth distribution. To test this hypothesis, we run quantile regressions of the following form:

$$\mathcal{Q}_{\tau}(\Delta y_{i,t+1}|x_{i,t}) = \alpha_{\tau} + \beta_{\tau} GPRC_{i,t}. \quad (2)$$

Above, we estimate the best linear predictor of the quantile τ of variable $\Delta y_{i,t+1}$ one-year ahead, conditional on values of country-specific geopolitical risk, denoted by $GPRC_{i,t}$ (the regressions also control for global geopolitical risk). As dependent variables, we consider GDP growth, total factor productivity (TFP) growth, and military spending as a share of GDP. We estimate equation (2) at different quantiles.

[21] The onset disaster dummy is one when $D_{i,t} - D_{i,t-1} = 1$ and $D_{i,t-1} = 0$, zero in non-disaster years, and missing when both $D_{i,t} = 1$ and $D_{i,t-1} = 1$ The ending of a disaster dummy treats all disaster years as zero, the year of the ending of a disaster as one, and all other years as missing.

Table 4 shows the results. The OLS estimates show that a rise in country-specific GPR predicts lower expected GDP growth, lower expected TFP growth, and higher expected military spending. The median effects (row labeled *q50*) have the same sign as the OLS estimates, though they are slightly smaller in magnitude, suggesting that the effects of GPR are somewhat larger during a crisis. The rows labeled *q10* and *q90* estimate equation (2) at the 10th and 90th quantiles. In line with the findings from the disaster risk regressions, a rise in the GPR index increases the probability of particularly adverse economic outcomes. The left tail of the GDP distribution—measured by the 10th quantile coefficient—shows a decline that is four times larger than the OLS effect, whereas the right tail of the distribution—measured by the 90th quantile—slightly increases. The conditional distributions of one-year-ahead TFP growth displays higher uncertainty, with both positive and negative tail events becoming more likely. Finally, the right tail of military spending moves disproportionally: elevated GPR predicts a risk of a large military buildup.

V Geopolitical Risk and Firm-Level Investment

In our last step, we provide evidence on the effects of geopolitical risk on investment using firm-level data. There are two questions that we are interested in. First, do firms in industries more exposed to aggregate geopolitical risks experience a larger decline in investment? Second, are idiosyncratic geopolitical events at the level of the firm associated with fluctuations in investment?

V.A Measuring Geopolitical Risk across Firms and Industries

It is useful to think of firm-level geopolitical risk as embedding three components:

$$GPR_{i,t} = GPR_t + GPR_t\Lambda_k + Z_{i,t}, \tag{3}$$

where the subscripts i and k denote firms and industries, respectively. The first component in equation (3) is aggregate GPR. The second component interacts aggregate GPR with industry exposure Λ_k, capturing the idea that some industries may be disproportionately affected by aggregate geopolitical risks. For instance, defense or petroleum companies may be particularly affected by geopolitical tensions in the Middle East, while airlines may be highly exposed to the fallout from terrorist attacks. The third component, $Z_{i,t}$, is idiosyncratic and isolates firm-level geopolitical risks that are not reflected at the aggregate and industry levels.

We first describe how we calculate industry exposure Λ_k. We regress daily portfolio returns in the 49 industry groups of Fama and French (1997) on changes in the daily geopolitical risk index:

$$R_{k,t} = \alpha_k + \beta_k \Delta GPR_t + \varepsilon_{k,t}, \tag{4}$$

where $R_{k,t}$ is the annualized daily excess return in industry k over the one-month T-bill rate and ΔGPR_t is the change in the daily geopolitical risk index. The sample runs from 1985 through 2019. Our idea is that stock returns in sectors with higher exposure drop relatively more than the aggregate market in response to spikes in the GPR index. By contrast, sectors with lower exposure tend to gain from geopolitical risks relative to the market. For instance, on September 17, 2001, the day the stock market reopened after 9/11, the returns in the transportation and precious metals sectors were -13 and $+7.4$ percent, respectively. This example underscores the importance of using daily data. Stock prices quickly react to news. Daily data also allow for a more granular taxonomy of geopolitical risks that—for episodes that do not dominate the news cycle for a prolonged period—is partly lost by aggregating data to monthly or quarterly frequencies.

We estimate the β_k coefficients in equation (4), demean them and change their sign so that positive values indicate high exposure. Figure A.8 in the Appendix plots the average exposure by industry. Precious metals, petroleum, and defense are among the industries negatively exposed to increases in geopolitical risk. Shipping and transportation are among the industries with positive exposure. For our empirical application below, the exposure measure Λ_k is a dummy that equals one for industries with above-median exposure, and zero otherwise.[22]

Next, we turn to the measurement of idiosyncratic geopolitical risk $Z_{i,t}$. A company might face elevated geopolitical risks because it operates in countries whose events are not reflected in the aggregate and industry measure (e.g., an oil company operating in Gabon). Alternatively, a company could have unique and time-varying exposure to aggregate geopolitical events, due to its location, political connections, trade exposure, or risk-management strategies.

Following Hassan et al. (2019), we perform text analysis on the transcripts of quarterly earnings calls of U.S.–listed firms. The sample runs from 2005:Q1 through 2019:Q4. We construct firm-level geopolitical risk by counting mentions of adverse geopolitical events and risks in the earnings calls. Specifically, we count the joint occurrences of 'risk' words within 10 words of 'geopolitical' words, normalizing the counts by the total number of words in the transcript.[23] In Appendix Figure A.9, we plot the GPR index alongside the index obtained by aggregating across firms, each quarter, the transcripts that discuss concerns about geopolitical risk. The correlation between the two indexes is 0.19. The positive correlation, albeit calculated on a short sample, bolsters our confidence that

[22] The use of a dummy makes the estimation more robust to the exact quantification of exposure. Results using the beta coefficients as a measure of exposure are similar and are shown in the Appendix (Table A.7).

[23] See Appendix E.3 for details. Examples of geopolitical words include 'war,' 'military,' 'terror,' 'conflict,' 'coup,' and 'embargo'. Examples of risk words include 'risk,' 'potential,' 'danger,' 'dispute,' 'incident,' and 'attack.'

investors' and newspapers concerns about geopolitical events are aligned.

V.B Dynamic Effects of Industry-Specific Geopolitical Risk

We quantify the differential effects of geopolitical risk on investment across industries. Using Compustat data, we measure investment as the ratio of capital expenditures to previous-period property, plant, and equipment, and denote it by ik. We regress firm-level investment at various horizons against aggregate GPR interacted with industry exposure. Our baseline strategy follows the local projection approach developed by Jorda (2005). We estimate

$$\log ik_{i,t+h} = \alpha_{i,h} + \beta_h \left(\mathbb{D}_{\Bbbk} \Delta \log GPR_t\right) + \mathbf{d_h} \, \mathbf{X}_{i,t} + \varepsilon_{i,t+h}, \tag{5}$$

where $h \geq 0$ indexes current and future quarters. The goal is to estimate, for each horizon h, the sequence of regression coefficients β_h associated with the interaction between aggregate geopolitical risk and industry exposure. In the equation above, α_i denotes firm fixed effects. The term $\mathbb{D}_{\Bbbk} \Delta \log GPR_t$ is the product of the industry exposure dummy times log changes in aggregate geopolitical risk. The term $\mathbf{X}_{i,t}$ denotes control variables—namely firm-level cash flows, firm-level Tobin's Q, and the lagged value of $\log ik_{i,t}$.

The top panel of Figure 11 shows the differential response of firm-level investment to a two-standard deviation aggregate GPR shock, for a firm belonging to an industry with high exposure to GPR. In the first year after the shock, an exposed firm experiences a decline in investment that is about 1 percentage point larger than its non-exposed counterpart. These estimates indicate that the negative repercussions of a typical spike in geopolitical risk on the investment rate vary depending on the industry of operation.

We conclude with a cautionary note on how to interpret our industry regressions. Our approach can be interpreted through the lens of a two-stage regression. In the first stage, we extract industry exposure by regressing stock returns on daily geopolitical risk industry-by-industry. In the second stage, we look at how investment responds to geopolitical risk depending on industry exposure. Accordingly, our second regression has the flavor of an IV regression of industry investment on industry stock returns where the instruments are industry dummies interacted with GPR. That said, our regression does not merely confirm that investment and stock prices are positively correlated, but also shows that movements in geopolitical risk affect some industries more than others, and that the differential effect is captured by the differential response of stock prices.[24]

[24] Alfaro, Bloom, and Lin (2018) look at differential firms exposure to energy prices, exchange rates, and economic uncertainty shocks and use the differential exposures to draw conclusions about the effects of uncertainty.

V.C Dynamic Effects of Firm-Specific Geopolitical Risk

To assess the dynamic relationship between investment and geopolitical risk at the firm level, we estimate

$$\log ik_{i,t+h} = \alpha_{i,h} + \alpha_{k,t,h} + \gamma_h Z_{i,t} + \mathbf{d_h} \mathbf{X}_{i,t} + \varepsilon_{i,t+h}. \tag{6}$$

The goal is to estimate, for each horizon $h \geq 0$, the coefficient γ_h which measures the dynamic effect on investment of changes in firm-level geopolitical risk. The regression includes firm fixed effects (α_i) and sector-by-quarter dummies ($\alpha_{k,t}$). Firm-control variables $\mathbf{X}_{i,t}$ include firm-level cash flows, firm-level Tobin's Q, and $\log ik_{i,t-1}$.

Mentions of geopolitical risks in the text of the earnings calls are a proxy for $GPR_{i,t}$, as the typical earnings call of a firm contains references to idiosyncratic as well as aggregate and industry-specific geopolitical risks. To isolate the firm-specific component $Z_{i,t}$, we absorb the aggregate and industry-specific components by including in equation (6) sector-by-quarter dummies. Our sample runs from 2005:Q1 through 2019:Q4 and is dictated by the availability of the earnings calls data.

The bottom panel of Figure 11 plots the response of firm-level investment (the sequence of coefficients γ_h at different horizons) after an increase in firm-level GPR of two standard deviations. Firms gradually reduce their investment over the two quarters after the shock, with investment declining more than 1 percent at the trough and staying below the baseline for up to one year.

V.D Summary of Firm-Level Evidence

Table 5 summarizes the analysis, tabulating the investment response to changes in geopolitical risk at the firm and industry levels. We focus on the response of investment two quarters ahead, in line with the results from the local projections that show that changes in geopolitical risk materialize with a delay of one to two quarters. In columns 1 and 2, investment responds to changes in geopolitical risk more for industries with above-average exposure. In column 3, investment at the firm-level is negatively associated with changes in firm-level geopolitical risk. Of note, in column 4, the response estimated with our firm-level variable is similar in sign and magnitude to the response of firm-level investment to firm-level *political* risk as measured by Hassan et al. (2019).[25] Overall, changes in geopolitical risks are associated with heterogeneous effects on firm investment, depending on the industry of operation and on firm-specific risks. The link between geopolitical risk and firm-level activity is significant, economically meaningful, and persistent over time.

[25] The measure by Hassan et al. (2019) is a broader concept of risk at the firm level encompassing concerns for instance about the government budget, health care, trade, and national security.

VI Conclusions

We propose and implement indicators of geopolitical risk that measure the threat, realization, and escalation of adverse geopolitical events. A detailed set of validation exercises confirm that our GPR indexes accurately capture the timing and intensity of adverse geopolitical events, both across countries and over time. Higher geopolitical risk foreshadows lower investment and is associated with higher disaster probability and larger downside risks to GDP growth. The adverse consequences of geopolitical risk are stronger for firms in more exposed industries, and high firm-level geopolitical risk is associated with lower firm-level investment.[26]

We conclude highlighting three areas for future research.

First, an implicit hypothesis underlying the construction of our indexes is that newspapers' attention to geopolitical events is an accurate measure of the perceptions of investors, economic agents, and policymakers. It would be useful in the future to extend our measurement exercise using additional sources, such as foreign-language publications, periodical country reports, or the transcripts of parliamentary debates.

Second, an important extension would be to investigate the international ramifications of geopolitical risks. Geopolitical risks can impact the price of risky assets and the flow of capital across countries. In a similar vein, tensions among countries can be an important force shaping trade flows and global supply chains through firms' actions and government policies.

Finally, in the empirical analysis, we have treated geopolitical risk as a driver of business fluctuations, highlighting a new force and a new set of shocks that economists have not traditionally emphasized. That said, an active literature in economics and political science has worked to better understand the causes of internal conflict and interstate warfare (see e.g. Blattman and Miguel, 2010 and Jackson and Morelli, 2011, among others). We hope that our measures can help researchers to better address these questions as well.

[26] While we find that higher geopolitical risk is associated with adverse economic outcomes, we caution that our empirical analysis is limited to analyzing past historical events. Future geopolitical risks could take different forms and yield different economic effects than in the past.

References

Alfaro, I., N. Bloom, and X. Lin (2018): "The Finance Uncertainty Multiplier," Tech. rep., National Bureau of Economic Research.

Baker, S. R., N. Bloom, and S. J. Davis (2016): "Measuring Economic Policy Uncertainty," *The Quarterly Journal of Economics*, 131, 1593–1636.

Barro, R. J. (2006): "Rare Disasters and Asset Markets in the Twentieth Century," *The Quarterly Journal of Economics*, 121, 823–66.

Barro, R. J. and J. F. Ursúa (2012): "Rare Macroeconomic Disasters," *Annual Review of Economics*, 4, 83–109.

Bazzi, S. and C. Blattman (2014): "Economic Shocks and Conflict: Evidence from Commodity Prices," *American Economic Journal: Macroeconomics*, 6, 1–38.

Beaudry, P. and F. Portier (2006): "Stock Prices, News, and Economic Fluctuations," *American Economic Review*, 96, 1293–1307.

Bergeaud, A., G. Cette, and R. Lecat (2016): "Productivity Trends in Advanced Countries Between 1890 and 2012," *Review of Income and Wealth*, 62, 420–44.

Berger, D., I. Dew-Becker, and S. Giglio (2019): "Uncertainty Shocks as Second-Moment News Shocks," *The Review of Economic Studies*, 87, 40–76.

Berkman, H., B. Jacobsen, and J. B. Lee (2011): "Time-Varying Rare Disaster Risk and Stock Returns," *Journal of Financial Economics*, 101, 313–32.

Blattman, C. and E. Miguel (2010): "Civil War," *Journal of Economic literature*, 48, 3–57.

Bloom, N. (2009): "The Impact of Uncertainty Shocks," *Econometrica*, 77, 623–85.

Bloom, N., M. Floetotto, N. Jaimovich, I. Saporta-Eksten, and S. J. Terry (2018): "Really Uncertain Business Cycles," *Econometrica*, 86, 1031–65.

Caldara, D., C. Fuentes-Albero, S. Gilchrist, and E. Zakrajšek (2016): "The Macroeconomic Impact of Financial and Uncertainty Shocks," *European Economic Review*, 88, 185–207.

Carney, M. (2016): "Uncertainty, the Economy and Policy," *Bank of England*.

Dijkink, G. (2009): "Geopolitics and Religion," in *International Encyclopedia of Human Geography*, ed. by R. Kitchin and N. Thrift, Oxford: Elsevier, 453–57.

Eckberg, D. L. (1995): "Estimates of Early Twentieth-Century US Homicide Rates: an Econometric Forecasting Approach," *Demography*, 32, 1–16.

Fama, E. F. and K. R. French (1997): "Industry Costs of Equity," *Journal of Financial Economics*, 43, 153–93.

Flint, C. (2016): *Introduction to geopolitics*, Routledge.

Foster, J. B. (2006): "The New Geopolitics of Empire," *Monthly Review*, 57, 1.

Gentzkow, M., B. Kelly, and M. Taddy (2019): "Text as Data," *Journal of Economic Literature*, 57, 535–74.

Glick, R. and A. M. Taylor (2010): "Collateral Damage: Trade Disruption and the Economic Impact of War," *The Review of Economics and Statistics*, 92, 102–27.

Gourio, F. (2008): "Disasters and Recoveries," *American Economic Review*, 98, 68–73.

Hassan, T. A., S. Hollander, L. van Lent, and A. Tahoun (2019): "Firm-Level Political Risk: Measurement and Effects," *The Quarterly Journal of Economics*, 134, 2135–2202.

Ilut, C. L. and M. Schneider (2014): "Ambiguous Business Cycles," *American Economic Review*, 104, 2368–99.

Jackson, M. O. and M. Morelli (2011): "The reasons for Wars: an Updated Survey," *The Handbook on the Political Economy of War*, 34.

Jaimovich, N. and S. Rebelo (2009): "Can News about the Future Drive the Business Cycle?" *American Economic Review*, 99, 1097–1118.

Jorda, O. (2005): "Estimation and Inference of Impulse Responses by Local Projections," *American Economic Review*, 95, 161–82.

Kozlowski, J., L. Veldkamp, and V. Venkateswaran (2018): "The Tail that Keeps the Riskless Rate Low," Working Paper 24362, National Bureau of Economic Research.

Ludvigson, S. C., S. Ma, and S. Ng (2021): "Uncertainty and Business Cycles: Exogenous Impulse or Endogenous Response?" *American Economic Journal: Macroeconomics*, 13, 369–410.

Nakamura, E., J. Steinsson, R. Barro, and J. Ursúa (2013): "Crises and Recoveries in an Empirical Model of Consumption Disasters," *American Economic Journal: Macroeconomics*, 5, 35–74.

Pindyck, R. S. and N. Wang (2013): "The Economic and Policy Consequences of Catastrophes," *American Economic Journal: Economic Policy*, 5, 306–39.

Ramey, V. A. (2011): "Identifying Government Spending Shocks: It's All in the Timing," *The Quarterly Journal of Economics*, 126, 1–50.

Ramey, V. A. and S. Zubairy (2018): "Government Spending Multipliers in Good Times and in Bad: Evidence from US Historical Data," *Journal of Political Economy*, 126, 850–901.

Rogers, A., N. Castree, and R. Kitchin (2013): *A Dictionary of Human Geography*, Oxford University Press.

Roser, M. and M. Nagdy (2013): "Military Spending," *Our World in Data*, https://ourworldindata.org/military-spending.

Saiz, A. and U. Simonsohn (2013): "Proxying for Unobservable Variables with Internet Document-Frequency," *Journal of the European Economic Association*, 11, 137–65.

Sherwin, M. J. (2012): "The Cuban Missile Crisis at 50: In Search of Historical Perspective," *National Archives and Records Administration*, 44, 6–16.

Stein, A. A. and B. Russett (1980): "Evaluating War: Outcomes and Consequences," *Handbook of political conflict: theory and research*, 399–422.

The New York Times Public Editor (2004): "What Belongs on the Front Page of The New York Times," August 22, 2004.

Table 1: Search Query for the Geopolitical Risk Index

A. Search Categories and Search Queries

	Category	Search Query	Peak (Month)	Contribution to Index% Full sample	1900-1959	1960-2019
Threats	1. War Threats	War_words N/2 Threat_words	Germany Invades Czech. (September 1938)	13.5	17.9	9.2
	2. Peace Threats	Peace_words N/2 Peace_disruption_words	Iran Crisis of 1946 (April 1946)	3.5	4.3	2.7
	3. Military Buildup	Military_words AND buildup_words	Cuban Missile Crisis (October 1962)	23.5	21.3	25.8
	4. Nuclear Threats	Nuclear_bigrams AND Threat_words	Nuclear Ban Negotiations (August 1963)	10.1	4.2	16
	5. Terrorist Threats	Terrorism_words N/2 Threat_words	9/11 (October 2001)	2.7	0.3	5
Acts	6. Beginning of War	War_words N/2 War_begin_words	WWII Begins (September 1939)	18.8	26.8	10.7
	7. Escalation of War	Actors_words N/2 Actors_fight_words	D-Day (June 1944)	19.6	23.9	15.3
	8. Terrorist Acts	Terrorism_words N/2 Terrorism_act_words	9/11 (September 2001)	8.3	1.3	15.2

B. Search Words

Topic Sets	Phrases
War_words	war OR conflict OR hostilities OR revolution* OR insurrection OR uprising OR revolt OR coup OR geopolitical
Peace_words	peace OR truce OR armistice OR treaty OR parley
Military_words	military OR troops OR missile* OR "arms" OR weapon* OR bomb* OR warhead*
Nuclear_bigrams	"nuclear war*" OR "atomic war*" OR "nuclear missile*" OR "nuclear bomb*" OR "atomic bomb*" OR "h-bomb*" OR "hydrogen bomb*" OR "nuclear test" OR "nuclear weapon*"
Terrorism_words	terror* OR guerrilla* OR hostage*
Actor_words	allie* OR enem* OR insurgen* OR foe* OR army OR navy OR aerial OR troops OR rebels

Threat/Act Sets	Phrases
Threat_words	threat* OR warn* OR fear* OR risk* OR concern* OR danger* OR doubt* OR crisis OR troubl* OR disput* OR tension* OR imminen* OR inevitable OR footing OR menace* OR brink OR scare OR peril*
Peace_disruption_words	threat* OR menace* OR reject* OR peril* OR boycott* OR disrupt*
Buildup_words	buildup* OR build-up* OR sanction* OR blockad* OR embargo OR quarantine OR ultimatum OR mobiliz*
War_begin_words	begin* OR start* OR declar* OR begun OR began OR outbreak OR "broke out" OR breakout OR proclamation OR launch*
Actor_fight_words	advance* OR attack* OR strike* OR drive* OR shell* OR offensive OR invasion OR invad* OR clash* OR raid* OR launch*
Terrorism_act_words	attack OR act OR bomb* OR kill* OR strike* OR hijack*

C. Excluded Words

Exclusion words	movie* OR film* OR museum* OR anniversar* OR obituar* OR memorial* OR arts OR book OR books OR memoir* OR "price war" OR game OR story OR history OR veteran* OR tribute* OR sport OR music OR racing OR cancer OR "real estate" OR mafia OR trial OR tax

Note: In panel A, the contribution to the index is the percent of articles in each category satisfying the condition for inclusion in the GPR index, as a share of all articles satisfying that condition. In panel B, 'core words' for each category are highlighted in red.

Table 2: Largest Geopolitical Shocks since 1900

Shocks to the GPR Index

Month	Rank	GPR	Shock to GPR	Event
1914m4	15	145.2	84.5	Occupation of Vera Cruz
1914m8	1	472.3	341.5	WWI Begins
1916m6	14	318.3	93.2	WWI Escalation
1917m2	6	350.2	141.9	U.S. Severs Germany Relations
1938m9	11	210.7	109.9	Germany occupies Czechia
1939m9	2	484.2	318.6	WWII Begins
1941m12	3	447.5	245.7	Pearl Harbor
1944m6	12	473.2	107.9	D-Day
1950m7	5	242.4	143.5	Korean War
1962m10	8	228.1	121.2	Cuban Missile Crisis
1973m10	13	161.1	94.3	Yom Kippur War
1990m8	9	191.9	115.5	Iraq invades Kuwait
1991m1	7	250.4	126.4	Gulf War
2001m9	4	289.9	238.2	September 11
2003m3	10	244.6	110.2	Iraq War

Shocks to the Threats Component of the GPR Index

Month	Rank	GPR Threats	Shock to GPR Threats	Event
1914m8	1	432.6	279.2	WWI Begins
1938m9	5	316.1	217.1	Germany occupies Czechia
1939m9	2	480.0	246.8	WWII Begins
1962m10	3	376.6	234.0	Cuban Missile Crisis
1990m8	4	314.1	225.7	Iraq invades Kuwait

Shocks to the Acts Component of the GPR Index

Month	Rank	GPR Acts	Shock to GPR Acts	Event
1914m8	2	571.5	456.9	WWI Begins
1939m9	1	560.0	463.0	WWII Begins
1941m12	4	665.7	391.5	Pearl Harbor
1991m1	5	273.1	196.9	Gulf War
2001m9	3	457.5	403.4	September 11

Note: The table lists the largest shocks to the GPR index (and its components) in the 1900-2019 sample. For this table, the shocks are constructed as the residuals of a regression of the level of the relevant monthly index against its first three lags.

Table 3: Geopolitical Risk and Economic Disasters

	(1) Disaster	(2) Disaster	(3) Disaster	(4) Disaster	(5) Disaster	(6) Onset	(7) Ending
GDP Growth t-1	-0.0071	-0.0062	-0.0056	-0.0065	-0.0056	-0.0009	0.0012
	(0.0030)	(0.0030)	(0.0030)	(0.0032)	(0.0026)	(0.0010)	(0.0010)
GPR	0.1753	0.1144	0.0337		0.1001	0.0180	-0.0175
	(0.0223)	(0.0241)	(0.0469)		(0.0236)	(0.0237)	(0.0094)
Country GPR		0.0940	0.0842		0.0794	0.0664	-0.0090
		(0.0160)	(0.0170)		(0.0175)	(0.0295)	(0.0105)
Dummy WWI/WWII			0.3328				
			(0.1761)				
GPR Spikes				0.1692			
				(0.0246)			
Country GPR Spikes				0.0821			
				(0.0122)			
Dummy Pre-1946					0.2437		
					(0.0490)		
Dummy 1946-1972					0.1152		
					(0.0467)		
Constant	0.2309	0.2289	0.1947	0.1762	0.1112	0.0401	0.1180
	(0.0252)	(0.0273)	(0.0341)	(0.0302)	(0.0320)	(0.0185)	(0.0130)
Observations	3,056	3,056	3,056	3,056	3,056	2,447	609
R^2	0.20	0.20	0.21	0.18	0.26	0.13	0.02
Countries	26	26	26	26	26	26	26
Country Fixed Effects	No	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses clustered by country and year.

Note: Effects of global and country-specific geopolitical risk on the probability of economic disaster in a panel of countries from 1900 through 2019. GDP growth is expressed in percent. GPR is standardized. Country GPR is standardized by country. Country GPR is the number of GPR articles mentioning the country divided by total number of newspaper articles. The GPR Spikes variable equals GPR in the 10 observations with the highest value of the GPR relative to a 20-year lagged moving average, and zero otherwise. The Country GPR Spikes variable equals Country GPR when Country GPR is larger than 2 standard deviations relative to a 20-year lagged moving average, and zero otherwise. The war dummy equals 1 in the years 1914-1918 and 1939-1945. See Appendix D for the list of countries.
Disaster episode data were constructed using updated GDP and consumption per capita data from Barro and Ursúa (2012) and the methodology described in Nakamura et al. (2013).

Table 4: Quantile Regression Effects of Country-Specific Geopolitical Risk

	(1) GDP Growth(t+1)	(2) TFP Growth(t+1)	(3) Military Exp.(t+1)
OLS	-0.35	-0.22	2.15
	(0.22)	(0.27)	(0.39)
Quantile			
q50	-0.24	-0.04	0.63
	(0.22)	(0.14)	(0.19)
q10	-1.44	-1.86	0.16
	(0.63)	(0.45)	(0.03)
q90	0.30	1.53	7.08
	(0.30)	(0.55)	(0.55)
Observations	3082	2261	2681
Countries	26	19	26

Standard errors in parentheses

Note: Quantile regression effects of geopolitical risk in a panel of countries from 1900 through 2019. In each specification, the right-hand side variable in country-specific GPR in year t (standardized by country). The dependent variables are GDP growth, TFP growth, and military expenditures in year $t+1$, respectively. GDP growth and TFP growth are expressed in percent units. Military expenditures are expressed as a share of GDP. The OLS coefficients are reported in the top row. The quantile coefficients report the effects at the 50th, 10th and 90th percentile of the distribution of the dependent variable. All regressions include an intercept and control for global geopolitical risk. Standard errors are bootstrapped using 500 replications.

Real GDP per capita data are from Barro and Ursúa (2012), extended through 2019 using the World Bank World Development Indicators. TFP data are from Bergeaud, Cette, and Lecat (2016). Military expenditures are taken from Roser and Nagdy (2013).

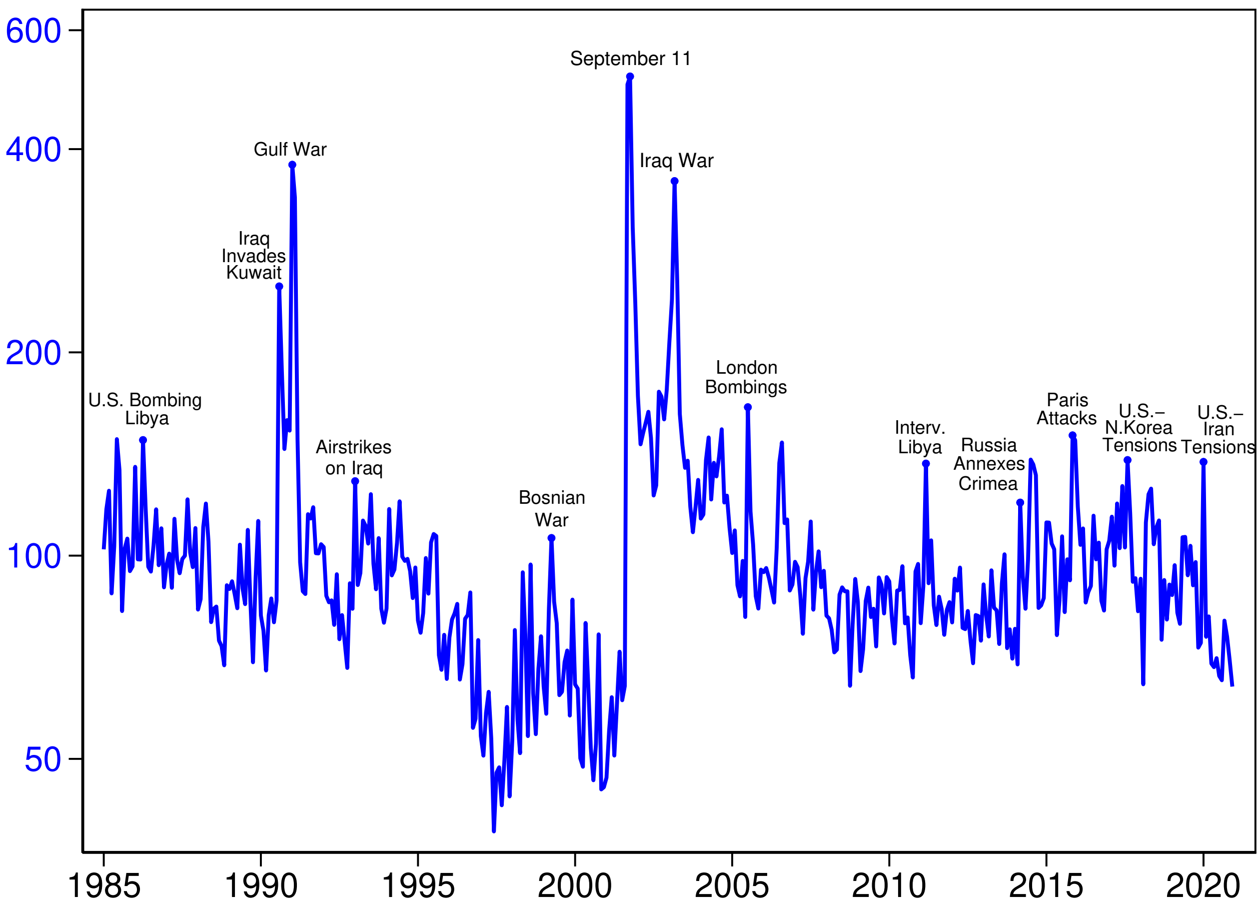
Table 5: Geopolitical Risk and Firm-Level Investment

$IK(t+2)$	(1)	(2)	(3)	(4)
Δ GPR $\times$ Dummy Industry Exposure	-0.63	-0.64		
	(0.29)	(0.27)		
GPR Firm-Level			-0.67	
			(0.30)	
Δ GPR	-1.39			
	(1.19)			
Political Risk Hassan et al.				-0.75
				(0.25)
Cash Flow	2.72	2.78	2.67	2.48
	(0.46)	(0.46)	(0.38)	(0.30)
Tobin's Q	8.91	7.93	9.31	9.47
	(1.68)	(1.56)	(0.92)	(0.90)
$IK(t-1)$	0.31	0.30	0.24	0.26
	(0.01)	(0.01)	(0.01)	(0.01)
Observations	374,727	374,727	95,073	112,161
Firm Fixed Effects	Yes	Yes	Yes	Yes
Time Effects	No	Yes	Yes	Yes
R-squared	0.45	0.47	0.59	0.58
Sample	85Q1-19Q4	85Q1-19Q4	05Q1-19Q4	05Q1-19Q4

Standard errors in parentheses

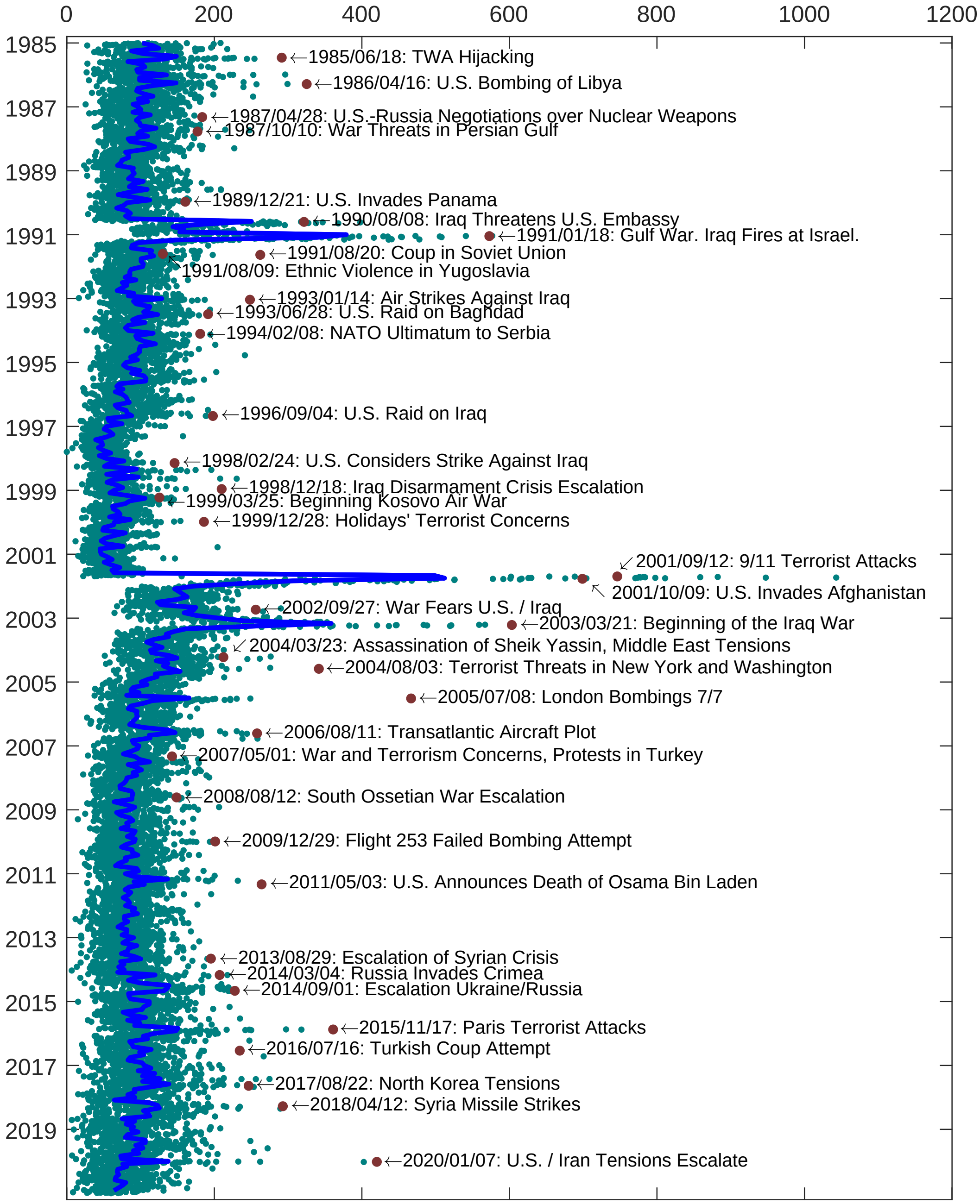
Note: The table shows results from regressions of firm-level investment on geopolitical risk at the industry or at the firm level. The dependent variable is IK (100 times the log of the investment rate) two quarters ahead.
All variables (except the dummy exposure variable) are standardized.
The standard errors are clustered by industry and quarter in columns 1 and 2, by firm and quarter-industry in columns 3 and 4.

Figure 1: Recent Geopolitical Risk Index from 1985



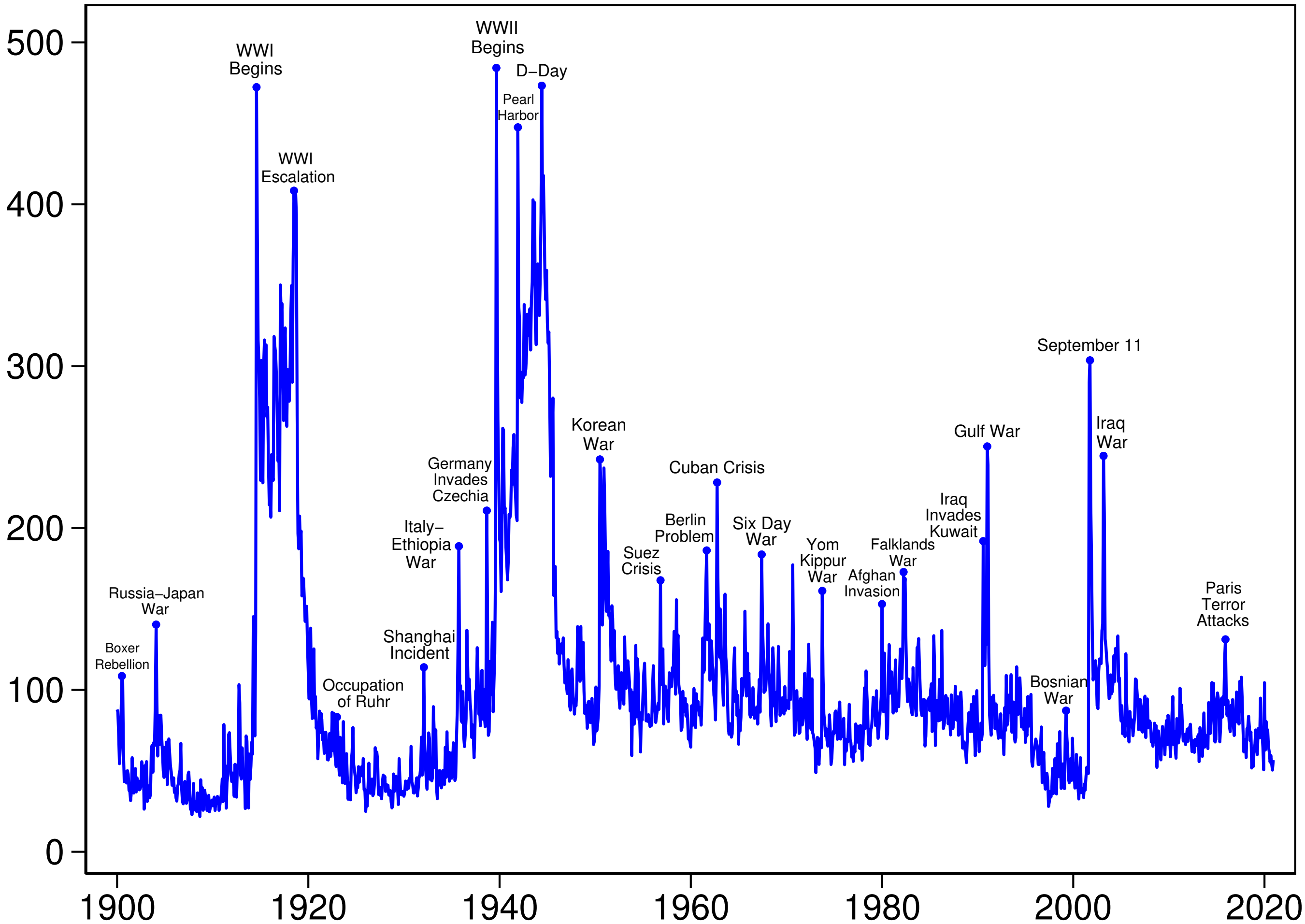
Note: Recent Geopolitical Risk Index from 1985 through 2020. Index is normalized 100 throughout the 1985-2019 period.

Figure 2: Daily Geopolitical Risk



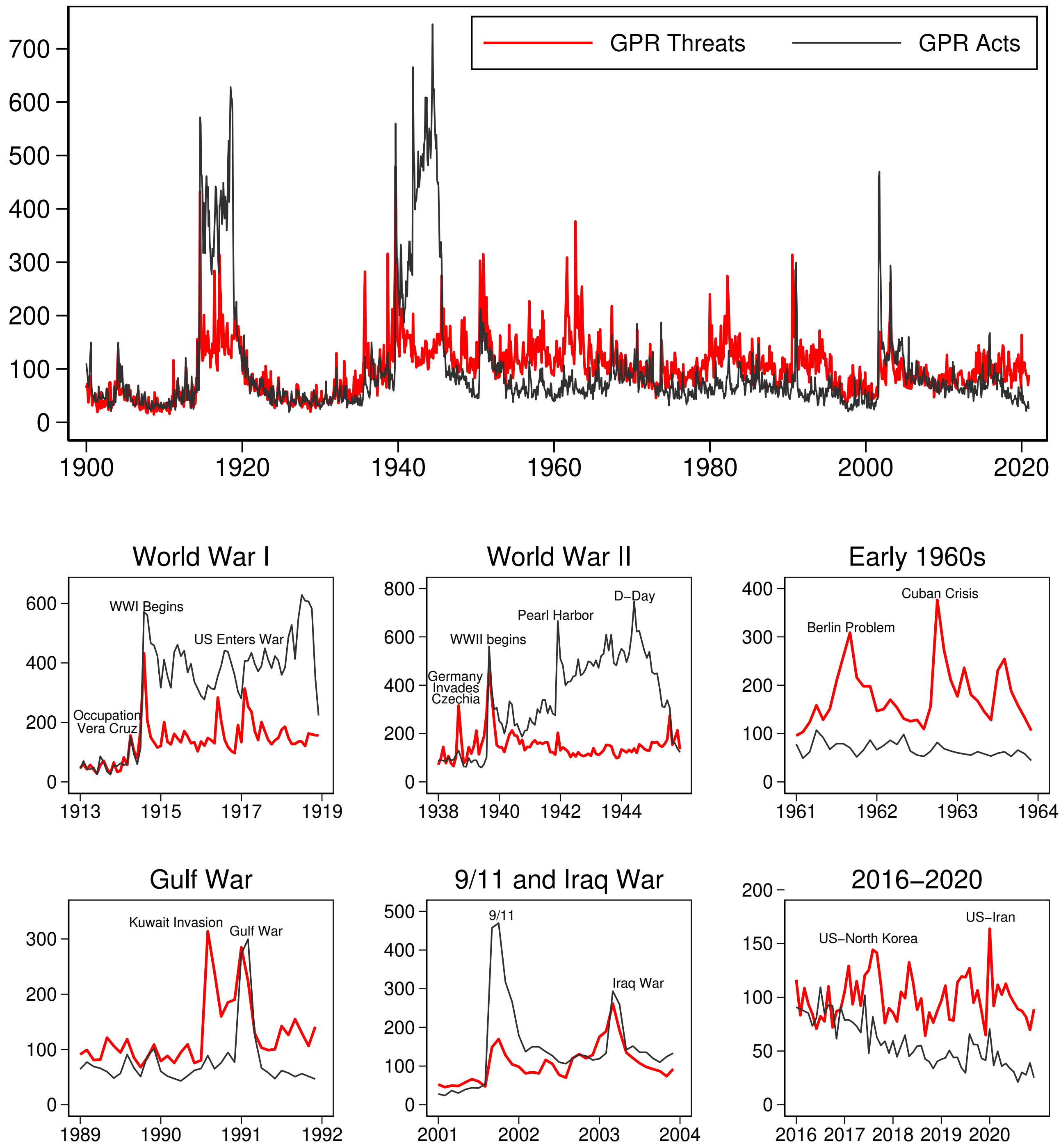
Note: Timeline of the daily GPR index from 1985 through end-2020. The solid blue line plots the monthly index. The green dots show the daily observations, including descriptions of the events reported by the newspapers on selected days featuring spikes in the index (shown by the large red dots). Index is normalized to 100 in the 1985-2019 period.

Figure 3: Historical Geopolitical Risk Index from 1900



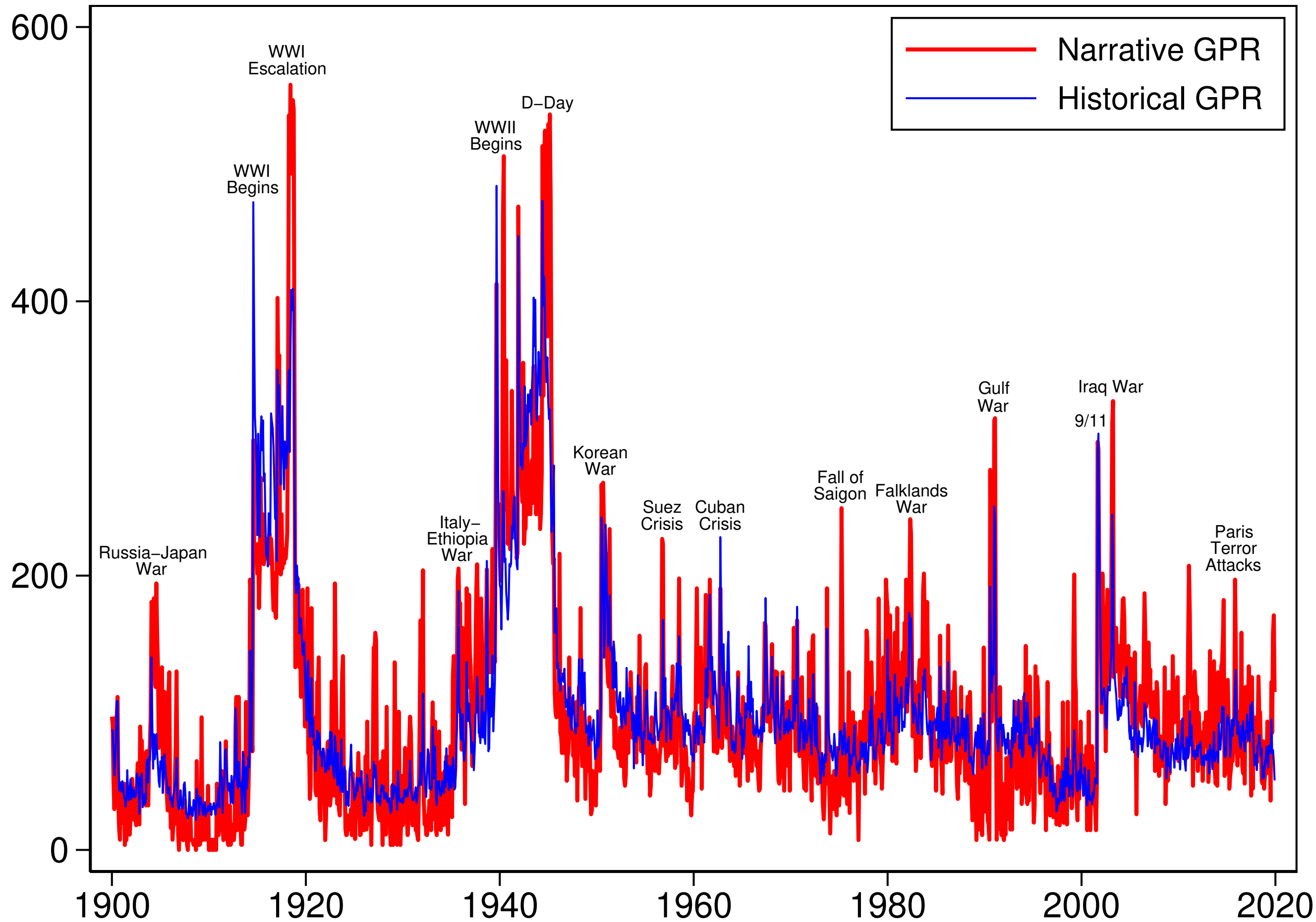
Note: Historical Geopolitical Risk Index from January 1900 through December 2020. Index is normalized to 100 throughout the 1900-2019 period.

Figure 4: Geopolitical Threats and Geopolitical Acts



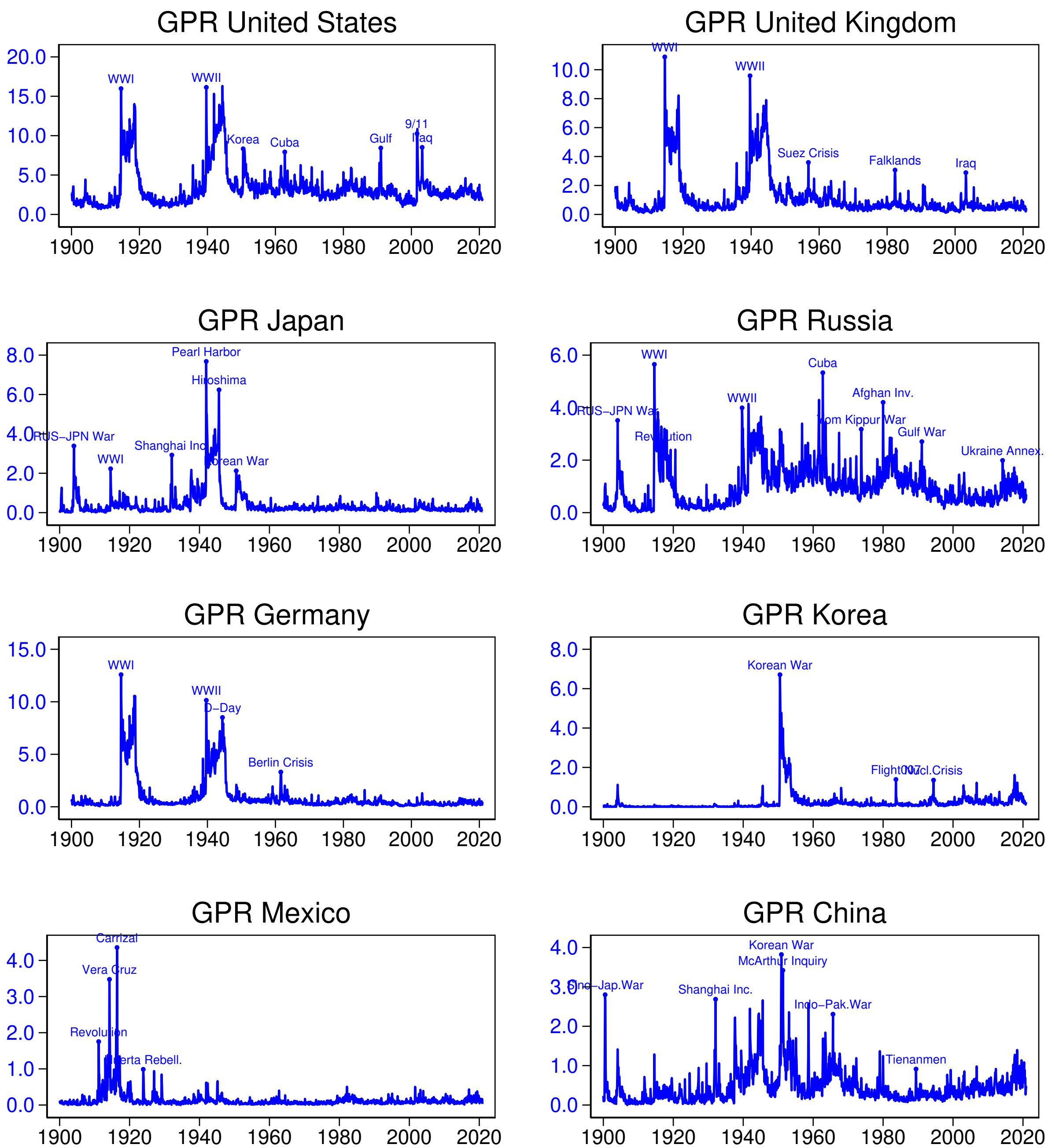
Note: Geopolitical threats (GPT) and the geopolitical acts (GPA) indexes. The GPT index is constructed by searching articles in categories 1 to 5 in Table 1. The GPA index is constructed by searching articles in categories 6 to 8 in Table 1. Both indexes are normalized to 100 in the 1900-2019 period.

Figure 5: Narrative GPR Index



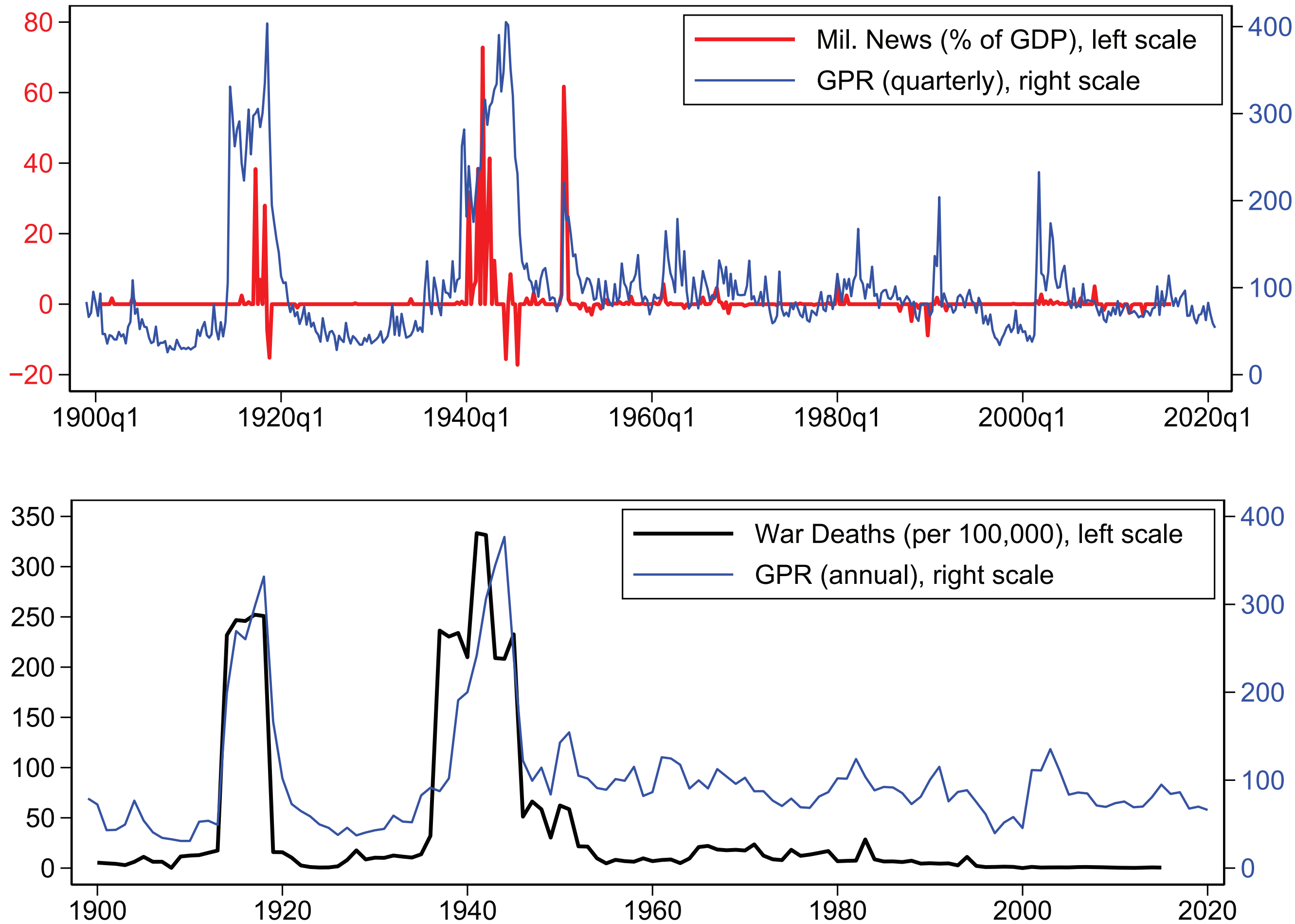
Note: The narrative GPR index is constructed by reading all daily front pages of *The New York Times* since 1900 and scoring them as 0, 1, 2, or 5 depending on the intensity of news about adverse geopolitical events. Both indexes are normalized to 100 in the 1900-2019 period.

Figure 6: Country-Specific Geopolitical Risk



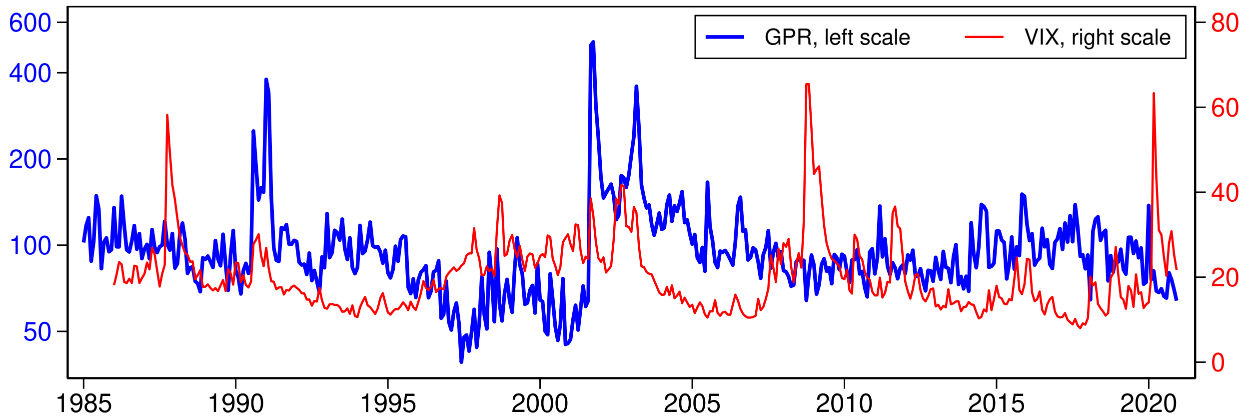
For each country, the country-specific GPR index measures share of articles simultaneously mentioning geopolitical risks together with the name of the country (or its capital or main city) in question.

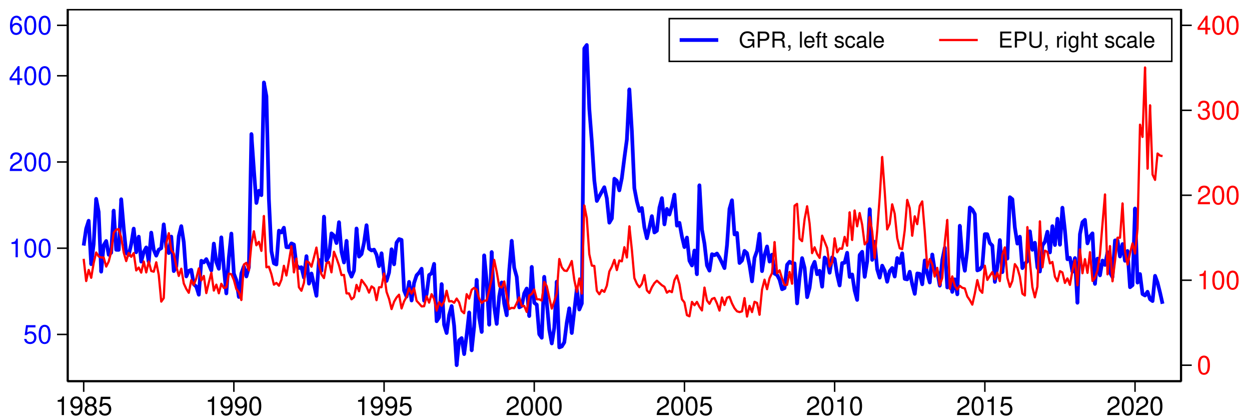
Figure 7: Comparisons with Military Spending News and War Deaths



Note: In the top panel, comparison of quarterly GPR Index with the expected military spending news variable from Ramey (2011), updated in Ramey and Zubairy (2018). In the bottom panel, comparison of the annual historical GPR index with worldwide military and civilian death rate from conflicts and terrorism (see Appendix B.4 for data sources).

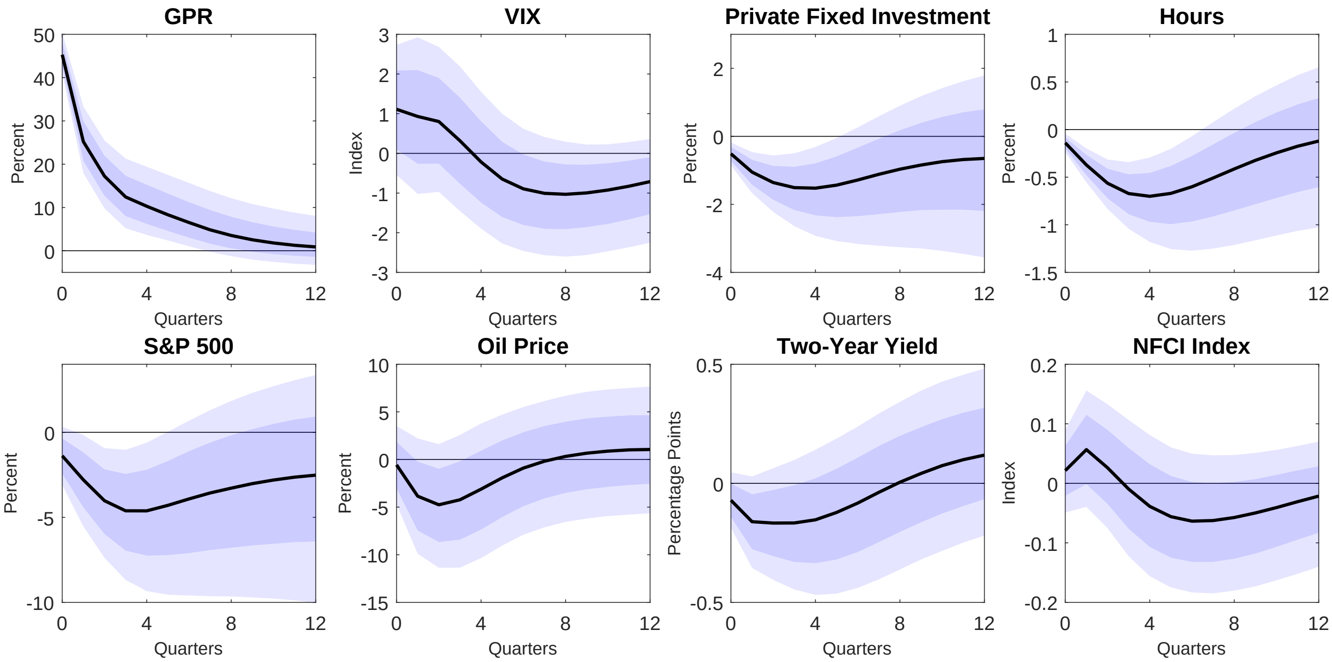
Figure 8: Comparison with Financial and Economic Uncertainty Measures





Note: Comparison of the GPR index (plotted on a log scale) with financial volatility as measured by the CBOE Volatility Index (old VIX—VXO) and with the economic policy uncertainty (EPU) index constructed by Baker, Bloom, and Davis (2016).

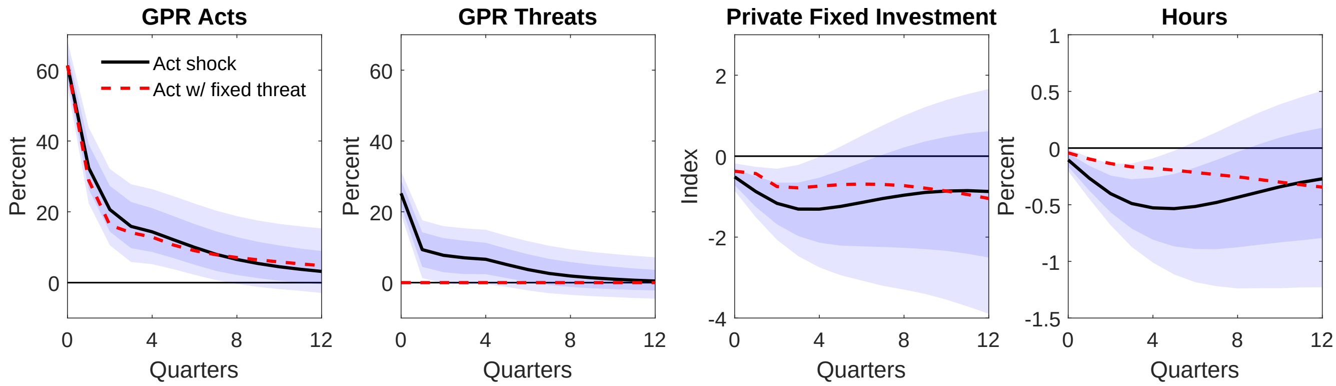
Figure 9: The Impact of Increased Geopolitical Risk



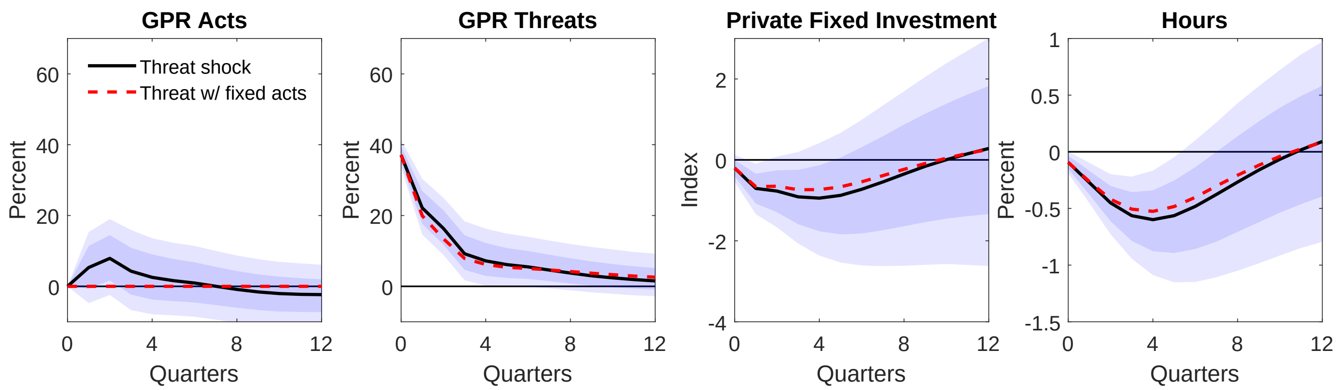
Note: The black solid line depicts the median impulse response of the specified variable to a two standard deviation increase in the GPR index. The dark and light shaded bands represent the 68 percent and 90 percent pointwise credible sets, respectively.

Figure 10: The Impact of Increased Geopolitical Risk: Acts vs. Threats

(a) Impulse Responses to a GPA Shock

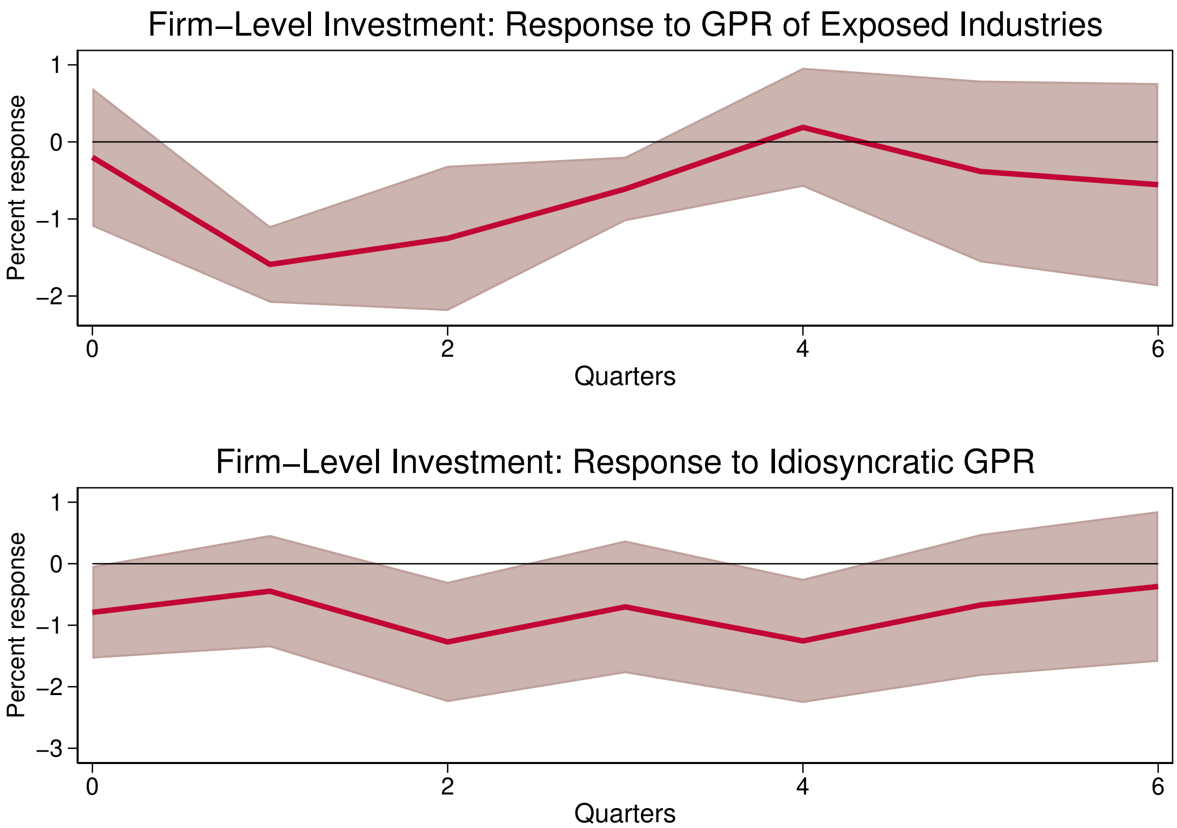


(b) Impulse Responses to a GPT Shock



Note: The black line depicts the median impulse response of the specified variable to a two-standard-deviations exogenous increase in the GPA index (panel a) and in the GPT index (panel b). The red dashed line depicts the outcome of a counterfactual simulation that keeps GPT (panel a) and GPA (panel b) constant. The dark and light shaded bands represent the 68 percent and 90 percent pointwise credible sets, respectively.

Figure 11: Response of Firm-Level Investment to Geopolitical Risk



Note: The top panel plots the dynamic response of investment following a two-standard deviation increase in aggregate GPR for a firm in an industry with positive exposure to geopolitical risk. The bottom panel plots the dynamic response of investment following a two-standard deviation increase in firm-level GPR. The shaded areas denote 90 percent confidence intervals. Standard errors are two-way clustered by firm and quarter-industry.

Appendix

A Appendix on: Construction of the Geopolitical Risk Indexes

A.1 Selection of the Words Entering the Index

As discussed in the main text, we use standard textual analysis techniques to inform the construction and organization of the dictionary of geopolitical terms used in the search query. Here we provide some additional details.

- We analyze the most frequent unigrams and bigrams found in two recent geopolitics books in order to get an idea for the range of topics covered by geopolitics. The book *Introduction to Geopolitics* (Flint, 2016) contains 48,759 bigrams, of which among the most common ones are 'geopolit code', 'war [on] terror', 'geopolit agent', 'cold war', 'soviet union', 'world war', and 'foreign polic[y].' The volume *The Geopolitics Reader*—edited by Simon Dalby, Paul Routledge, and Gearóid Tuathail—which is a compendium of 39 geopolitics essays written by different authors, contains 91,210 bigrams, of which the most common ones are 'unit[ed] states', 'cold war', 'foreign polic[y]', 'nation secur[ity]', 'world war', 'world order', 'nation[al] state', 'gulf war', 'war II', and 'nuclear weapon.'

- We search in the corpus of Historical American English the most common collocates of the words war, military, conflict, terrorism, nuclear, peace and battle in order to set up and refine our dictionary. The first ten collocates of the word 'war' are world, civil, II, during, department, secretary, cold, declare, peace, Vietnam. Except for the word 'declare', we do not use the first ten words because they refer to a particular historical period or situation, rather than to obvious war risks or beginning of wars. Scrolling down the list of the first 100 collocates, we encounter words such as 'outbreak', 'inevitable', and 'imminent', that we include in the final dictionary of either risk or action words. When we repeat the same process for the word 'military', we add words such as 'threat', 'coup', 'buildup.' The ten most frequent collocates of the word 'terrorism' are war, against, act, international, fight, threat, political, expert, support, campaign. We single act and threat from this list. We repeat this procedure for other key words mostly to ensure there are no glaring omissions in our query.

- To organize the search query, we use high frequency words and their corresponding synonyms, as well as words that are more likely to appear on the front page of newspapers on days of high geopolitical tensions. To this end, we compare the text of the front pages of newspapers on days of high geopolitical tensions with the text of the front pages of newspapers on random days. The days of high geopolitical tensions are the days in which adverse geopolitical events are featured in a banner headline on the front page of *The New York Times*. The more likely (lemmatized) words on days of high geopolitical tensions are reported in Table A.1. For instance, 'crisis' has a term frequency of 0.254 percent on days of high geopolitical tensions, whereas its term frequency on average is 0.042 percent. Accordingly, an article containing the word 'crisis' is about 6.1 times more likely to belong to the high-GPR category.

Note that some potentially likely words are not included in the search in spite of their relatively high odds. Words such as ‘communique’ or ‘neutral’ or ‘civilian’ or ‘command’ are left out because of their ambiguous meaning. Words such as ‘combat’ or ‘tank’ or ‘submarine’ have a low average term frequency.

A.2 Newspapers Coverage and Contribution of Search Categories

The recent geopolitical risk index is constructed by running a search query in the ProQuest Newsstand Database. We search the archives of the following newspapers (start date availability in parentheses): *Chicago Tribune* (1/1/1985); *The Daily Telegraph* (4/1/1991); *Financial Times* (5/31/1996); *The Globe and Mail* (1/1/1985); *The Guardian* (8/18/1992); *Los Angeles Times* (1/1/1985); *The New York Times* (1/1/1985); *USA Today* (4/1/1987); *The Wall Street Journal* (1/1/1985); and the *Washington Post* (1/1/1985).

For the historical index, we search the historical archives of the *Chicago Tribune*; *The New York Times*; and the *Washington Post*, starting on January 1, 1900.

To construct the numerator of the GPR index, we run one joint query across all categories. Note that a single article could belong to more than one category. The sum of the hits across categories is 15 percent higher than the number of articles belonging to the GPR index, thus suggesting some overlap across categories.

To construct the denominator of the GPR index, we search news articles that simultaneously contain the words ‘the’, ‘be’, ‘to’, ‘of’, ‘and’, ‘at’, and ‘in.’ These words are among the 20 most common words found in the historical archives since 1900. By searching for articles that simultaneously include several of the most frequent words in English, we exclude from the count one-line news, articles that are too short, or titles of articles that are sometimes erroneously classified as full articles.

For the recent period, newspaper-specific indexes are shown in Figure A.10, expressed as a share of news articles for each of the newspapers.[2] As the top left panel shows, coverage of geopolitical risks aligns with the benchmark GPR index for the three general interest newspapers that we use in the construction of the historical index. As the top right panel shows, coverage of geopolitical risks is slightly higher than the average for the two business newspapers in the sample, *The Wall Street Journal* and the *Financial Times*. Coverage also lines up with the average for the two U.S. newspapers not included in the historical index (middle left panel). Coverage of geopolitical events by non-U.S. general interest newspapers lines up with the average, but is slightly more volatile (middle right and bottom left panels).

Figure A.11 elaborates on the contribution of each search category of the index, this time focusing on the historical period. Nuclear threats are disproportionately important during the Cold War. Terror threats and acts trend higher over the sample period, spiking after 9/11, and remaining at elevated levels ever since. The categories relating to beginning and escalation of war exhibit two large spikes in corresponding of the world war.

[2] The actual indexes are a normalization of the articles’ share.

A.3 The Actual Search Query

For the interested researchers, here are the actual search queries as they are entered in the ProQuest database.

The articles mentioning geopolitical risks are found with the following query:

```
DTYPE(article OR commentary OR editorial OR feature OR front page article OR front page/cover
story OR news OR report OR review) AND (((war OR conflict OR hostilities OR revolution*
OR insurrection OR uprising OR revolt OR coup OR geopolitical) NEAR/2 (risk* OR warn*
OR fear* OR danger* OR threat* OR doubt* OR crisis OR troubl* OR disput* OR concern* OR
tension* OR imminen* OR inevitable OR footing OR menace* OR brink OR scare OR peril*))
OR ((peace OR truce OR armistice OR treaty OR parley) NEAR/2 (menace* OR reject* OR threat*
OR peril* OR boycott* OR disrupt*)) OR ((military OR troops OR missile* OR "arms" OR weapon*
OR bomb* OR warhead*) AND (buildup* OR build-up* OR blockad* OR sanction* OR embargo OR
quarantine OR ultimatum OR mobiliz* OR offensive)) OR ((("nuclear war" OR "nuclear warfare"
OR "nuclear warhead" OR "nuclear warheads" OR "nuclear wars") OR ("atomic war" OR "atomic
warfare" OR "atomic warheads" OR "atomic wars") OR ("nuclear missile" OR "nuclear missiles")
OR ("nuclear bomb" OR "nuclear bombardment" OR "nuclear bomber" OR "nuclear bombers" OR
"nuclear bombing" OR "nuclear bombs") OR ("atomic bomb" OR "atomic bombing" OR "atomic
bombings" OR "atomic bombs") OR "h-bomb*" OR ("hydrogen bomb" OR "hydrogen bombs") OR
"nuclear test") AND (risk* OR warn* OR fear* OR danger* OR threat* OR doubt* OR crisis
OR troubl* OR disput* OR concern* OR tension* OR imminen* OR inevitable OR footing OR
menace* OR brink OR scare OR peril*)) OR ((terroris* OR guerrilla* OR hostage*) NEAR/2
(risk* OR warn* OR fear* OR danger* OR threat* OR doubt* OR crisis OR troubl* OR disput*
OR concern* OR tension* OR imminen* OR inevitable OR footing OR menace* OR brink OR scare
OR peril*)) OR ((war OR conflict OR hostilities OR revolution* OR insurrection OR uprising
OR revolt OR coup OR geopolitical) NEAR/2 (begin* OR begun OR began OR outbreak OR "broke
out" OR breakout OR start* OR declar* OR proclamation OR launch* OR wage*)) OR ((allie*
OR enem* OR foe* OR army OR navy OR aerial OR troops OR rebels OR insurgen*) NEAR/2 (drive*
OR shell* OR advance* OR invasion OR invad* OR clash* OR attack* OR raid* OR launch* OR
strike*)) OR ((terroris* OR guerrilla* OR hostage*) NEAR/2 (act OR attack OR bomb* OR
kill* OR strike* OR hijack*)) NOT (movie* OR film* OR museum* OR anniversar* OR obituar*
OR memorial* OR arts OR book OR books OR memoir* OR "price war" OR game OR story OR history
OR veteran* OR tribute* OR sport OR music OR racing OR cancer))
```

The articles mentioning geopolitical risks are normalized by total number of articles. The total number of articles are found with the following query:

```
DTYPE(article OR commentary OR editorial OR feature OR front page article OR front page/cover
story OR news OR report OR review) AND ("THE" AND "BE" AND "TO" AND "OF" AND "AND" AND
"AT" AND "IN")
```

B Appendix on: Validation of the Index

B.1 Appendix on: Comparison with a Narrative GPR Index

We construct a 'narrative' GPR index by reading and scoring 44,000 daily front pages of *The New York Times* from 1900 through 2019.

Together with a team of research assistants, we read all headlines above the fold of the daily front pages of *The New York Times* and score each day with a 0, 1, 2, or 5 depending on whether: no headline features rising or existing geopolitical tensions (score: 0); one headline—but not the lead headline—features geopolitical tensions (score: 1); the lead headline—but not a banner headline—features geopolitical tensions (score: 2); the banner headline features geopolitical tensions (score: 5).

The guide we used to implement uniform coding of the articles is available at https://www.matteoiacoviello.com/gpr_replication.htm.

In 1978, *The New York Times* was on strike from August 10 through November 4. We replace it with the Washington Post throughout that period.

To verify uniform coding, we select about 1,000 front pages that are coded simultaneously by more than one research assistant. We find that the Cronbach alpha for the articles coded by more than one research assistant is 0.9329, a number that indicates a very strong overlap among coding practices across research assistants. In particular, 85 percent of articles are given the same narrative rating by two different research assistants.

B.2 Appendix on: Country-Specific GPR Measures

Country-specific GPR indexes are constructed for each country by counting the number of articles satisfying two criteria: (1) the article must satisfy conditions for inclusion in the GPR index; (2) the article must contain the name of the country (including any names and/or spelling variants from the past) or its capital or its main city. For instance, any article satisfying the conditions for inclusion in the GPR index and containing 'Japan' OR 'Japanese' OR 'Tokyo' OR 'Tokio' counts for inclusion in Japan's country-specific GPR index.

B.3 Appendix on: Comparison with News about Military Spending

The bottom four panels of Figure A.1 compare military spending news (Ramey, 2011) with surprises in the GPR index during selected historical episodes. During both world wars, some of the spikes in geopolitical risk align with jumps in the military spending news measure. Yet, the military spending news variable only spikes in the middle of the wars when U.S. intervention appears increasingly likely. At the onset of the Korean War, the largest jump in the GPR index coincides with a large shock to military spending news. Lastly, following 9/11, the largest spikes in GPR take place in 2001:Q3 and 2001:Q4, whereas Ramey's variable increases in 2002:Q1 and 2007:Q4. In particular, the 2007 spike occurs on news of higher-than-projected costs of the Afghan and Iraq wars, while the GPR index barely moves.

B.4 Appendix on: Comparison with War Deaths

Data on global war deaths measure death rate from conflicts (military and civilian, deaths per 100,000 people). Data on conflict deaths from 1900 through 1969 are from Peter Brecke's Conflict Catalog (https://brecke.inta.gatech.edu/research/conflict/). Data from 1970 through 2015 combine deaths from conflict and terrorism and are constructed using data from the National Consortium for the Study of Terrorism (https://start.umd.edu/), the UCDP One-sided Violence Dataset (https://ucdp.uu.se/downloads/), the PRIO Battle Deaths Dataset https://www.prio.org/Data/Armed-Conflict/Battle-Deaths/The-Battle-Deaths-Dataset-version-30/, and the Defense Casualty Analysis System (https://dcas.dmdc.osd.mil/dcas/pages/casualties.xhtml). Data on population are from Our World in Data (https://ourworldindata.org/world-population-growth).

B.5 Appendix on: Tests of Granger Causality

We run Granger-causality tests based on the following regression:

$$LGPR_t = \alpha + \sum_{i=1}^{p} \beta_i LGPR_{t-i} + \sum_{i=1}^{p} \Gamma'_{M,i}\mathbf{M}_{t-i} + \sum_{i=1}^{p} \Gamma'_{F,i}\mathbf{F}_{t-i} + \sum_{i=1}^{p} \Gamma'_{U,i}\mathbf{U}_{t-i} + \varepsilon_{LGPR,t},$$

where $LGPR$ is the log benchmark GPR index; $\mathbf{M}$ denotes a vector of macroeconomic variables; $\mathbf{F}$ denotes a vector of financial variables; and $\mathbf{U}$ denotes a vector of proxies for uncertainty. In our application, $\mathbf{M}$ consists of the log-difference of U.S. industrial production, the log-difference of private employment, and the log of the WTI price of oil deflated by U.S. CPI; $\mathbf{F}$ consists of the real return on the S&P500 index and the 2-year Treasury yield; and $\mathbf{U}$ includes the VIX and the log the EPU index. We include in the regression a constant term and set $p = 3$. The sample runs from 1986:M1 through 2019:M12.

Column (1) of Table A.8 tabulates the results of the exogeneity test that we run for the log GPR index. Columns (2) and (3) show the results of the Granger causality tests when we replace the log GPR index with the log of GPA and GPT indexes, respectively. Both regressions include lags of both log GPA and log GPT as independent variables. As for the baseline GPR index, we do not find any significant impact of macroeconomic, financial, and uncertainty variables on the GPA and GPT indexes.

B.6 Appendix on: Audit of the GPR Index

The full-scale audit consists of the construction of a human-generated GPR index and the evaluation of the computer-generated GPR index.

The set of newspaper articles used to construct the historical index—denoted by $\mathcal{U}$—contains about $10,000$ articles, on average, each month. The audit underlying the construction of the historical human index was conducted from a subset of $\mathcal{U}$—denoted by $\mathcal{E}$—consisting of articles that contain any of the following words: *geopolitics*, *war*, *military*, *terrorism/t*. The subset $\mathcal{E}$ contains about $1,600$ articles per month, about 15 percent of the articles in $\mathcal{U}$. We focus on a subset of articles containing the words above to make our audit more efficient and less prone to sampling error. Indeed, as shown by Table A.1, words such as *war* and *military* are very popular words both in days of high geopolitical risks and on days of low geopolitical risks. Among the articles belonging to set $\mathcal{E}$, 33 percent only refer to rising geopolitical tensions. This fraction would be much lower in the universe $\mathcal{U}$, thus requiring a much bigger audit sample. Of course, there remains the possibility that articles that do not mention these words also mention geopolitical risks. However, in a random sample of 585 articles not containing any of these words, the fraction of articles mentioning high geopolitical risks was only 2.6 percent, thus allaying our concerns.

To construct a human-generated GPR index, we randomly sampled 7,365 articles from $\mathcal{E}$—on average about 60 articles per year. For each year, we calculated the fraction of articles assigned to $\mathcal{E}^1$, multiplied this fraction by the quarterly rate $\mathcal{E}/\mathcal{U}$, and normalized the resulting index to 100 over the entire sample. In Figure A.12, we show the human-generated GPR index. The historical, computer-generated, index lines up well with an index that could be constructed by humans. The correlation between the two series is 0.93.

To evaluate the computer-generated GPR index, we randomly sampled 2,400 articles from the set of articles selected by the automated text-search algorithm, and classified them as either discussing high or rising geopolitical tensions or not. The fraction of articles that constitute the computer-generated GPR and mention high or rising geopolitical risks is 79 percent. Of the remaining articles, less than 1 percent mention low or decreasing geopolitical tensions. The remaining 20 percent false positives fall under various categories, for instance discussions of past geopolitical events and related personal experiences (e.g. trauma) without an immediate connection to current developments. The low incidence of articles discussing favorable geopolitical developments supports our claim that our choice of words captures negative risks to the geopolitical outlook.

Table A.3 presents additional results and lists the three alternative search queries that: do not remove the 'excluded words' from the query (GPRNOEW); use smaller sets of basic words highlighted in red in Table 1 (GPRBASIC); use the Boolean operator 'AND' for all search categories—as opposed to a search of two terms within two words from each other (GPRAND). For each alternative index, we randomly sample 500 articles. We code manually each article as either discussing high or rising geopolitical tensions or not. As shown in the table, the alternative indexes have a higher error rate.

B.7 Appendix on: Does War Language Change over Time?

Any text search covering a long period of time must be flexible enough to accommodate neologisms and obsolete words, as well as semantic, syntactic, and spelling changes. The construction of our index reflects an extensive analysis of the most common words and sentences used in newspaper articles over time to describe risks to war and to peace, and acts of war and terror. In particular, our dictionary includes enough words to rule out the possibility that the inclusion or exclusion of a few words may bias the index in particular periods. Additionally, we verify that changes in language over time do not affect our automated index as follows:

1. We verify that there are no divergent trends between the narrative indexes, which is constructed

by human reading of actual newspaper front pages, and our automated index.

2. When we compare the period 1900-1959 with the 1960-2019 period—see Table A.2—we verify that commonly appearing words of the period are in our search and the search is focused on words that have a high signal to noise ratio. Words such as communique, league, fleet, and men were relatively more frequent in the past and words such as television, vow, jet, target and oil are relatively more frequent in recent decades, but in both instances their inclusion would have worsened the accuracy of the index.

3. We analyze term frequency for the words and word combinations used to construct the index. Tables A.4 and A.5 tabulate the results for the entire sample and across subperiods. We find that our query includes words that are more frequent in early part of the 20th century (such as menace or peril) as well as words that are more common in recent decades. We also show how our search focuses on words with high signal to noise ratio and excluded words that would have worsened the accuracy of the index. For instance, many of the words to indicate risk that are included in our search, such as 'menace', 'peril', and 'scare', are rarely used in the second half of the 20th Century.

4. In initial checks, we noticed that over time newspapers appear to have devoted increasingly more space to arts, history, sports, and entertainment, often borrowing some of their language from warfare and military terminology. For this reason, our search does not count as articles measuring geopolitical risks any of the articles containing any of the 'excluded words' listed in Table 1. Without these words, the index would have a slight upward trend throughout the historical period.

B.8 Appendix on: Does Media Attention Measure the Underlying Risk?

In this subsection, we elaborate on the evidence that the GPR index is not unduly affected by issues related to media reporting of news such as unpredictable or seasonal newsworthy events, political slant of the media, or changes in societal norms.[3]

First, we show that there is little evidence that unpredictable and predictable newsworthy events can explain fluctuations in the index. In the top panel of Figure A.2, we show that there is little correlation between the GPR index and a news index of natural disasters, even if news about natural disasters attracts significant media attention. In the bottom panel, we confirm that the irrelevance of other newsworthy events still applies when we look at an index capturing newspapers' attention towards recurring and predictable sport events, such as the Olympics or the World Series. More in general, we find little correlation between the GPR index and a swath of newsworthy events, as shown in Table A.6 in the Appendix.

[3] There are many reasons that the GPR may fluctuate for reasons unrelated to latent geopolitical risks. For instance, the high levels of the index in the years following 9/11 may reflect public fear towards geopolitical tensions more than actual risk. Additionally, geopolitical issues may receive more or less coverage in the news depending on the attention of the press to other newsworthy events. Finally, the use of war and terrorism-related words may reflect issues that a newspaper likes to report on and that readers are passionate about, rather than objective geopolitical risks.

The political bias or slant of newspapers may also induce measurement error in our index. In our second check, we verify that there is little slant in the newspapers' coverage of geopolitical risks. For the recent sample, when we split our 10 newspapers into five left-leaning and five right-leaning. The left-leaning newspapers are *The Globe and Mail*, *The Guardian*, *Los Angeles Times*, *The New York Times*, and *The Washington Post*. The right-leaning newspapers are *The Daily Telegraph*, the *Chicago Tribune*, the *Financial Times*, *USA Today*, and *The Wall Street Journal*. As shown in Figure A.3, the 'left" and 'right" versions of our GPR index move together closely, with a correlation of 0.87, suggesting that while different media outlets may cover geopolitical events with different intensity, the broad time-series properties of the index are robust to the political slant of newspapers.

War reporting could have changed over time as norms of patriotism or censorship shifted. Similarly, competition in the media industry may have encouraged some newspapers to cover more unconventional topics—from family to sports to technology to climate change—at the expense of the traditional events of the day. In our final check, we ask whether long-run shifts in the newspapers' coverage of particular events may induce spurious trends in our index. We do so by checking whether there is a substantial divergence between news coverage and actual occurrence of 'fear-based' phenomena that are somewhat easier to count and quantify, such as murders or hijackings or nuclear tests. Figure A.4 shows a remarkably good correlation between occurrence and extent of murders, hijackings and nuclear tests on the one hand, and the media coverage of these events on the other.

In sum, there are many historical trends that may have affected the information content of our index, and researchers should be aware of these issues. However, we believe that these trends are unlikely to significantly affect its usefulness for economic analysis.

B.9 Appendix on: Checking the Saiz and Simonsohn (2013) Conditions

Saiz and Simonsohn (2013) state a number of conditions that must hold to obtain useful document-frequency based proxies for variables, such as geopolitical risk, that are otherwise difficult to measure. Our audit, among other things, makes sure that these conditions are indeed satisfied in our application. We provide a point-by-point discussion on how we perform these data checks below.

1. We verify that our search terms are more likely to be used when geopolitical risk is high than when it is low (*Data check 1: Do the different queries maintain the phenomenon and keyword constant?*, and *Data check 3: Is the keyword employed predominately to discuss the occurrence rather than non-occurrence of phenomenon?*). Across all the documents in our human audit, we found that 79 percent of articles measure high geopolitical risk, whereas only a smaller fraction of these articles measure declining tensions. We therefore conclude that increases in GPR are far more likely to lead to the use of our preferred search terms.

2. The GPR index is a frequency, thus satisfying data check 2 (*Data check 2. Is the variable being proxied a frequency?*).

3. We verify that the average number of documents found is large enough for variation to be driven by factors other than sampling error (*Data check 4: Is the average number of documents found large enough [...]?*). In particular, we verify that spikes in GPR are easily attributable to well-defined historical events at both a monthly and at a daily frequency. For instance, the first spike in monthly data since 1985 is in April 1986, reflecting the events that culminated with U.S. air strikes against Libya on April 15. However, the index also spikes, within the month, on April 8, when the United States accused Muammar el-Qaddafi of sponsoring terrorist acts aimed at Americans (such as the Berlin discotheque bombing which occurred on April 5). It also spikes on April 18, when British police found a bomb in a bag that was taken onto an El Al aircraft.

4. We verify that measurement error is low enough (*Data check 3*, and *Data check 5: Is the expected variance in the occurrence-frequency of interest high enough to overcome the noise associated with document-frequency proxying?*), by choosing combinations of search terms that—unlike with a single keyword or a bi-gram—are unlikely to be used outside of the realm of rising geopolitical risk. For instance, a naïve geopolitical risk index that merely counts the share of articles containing *geopolitics*, *war*, *military*, or *terrorism/t* is nearly as high in March 1991 as in January 1991, whereas the benchmark GPR index is much lower in March 1991. This occurs because the naïve index fails to account for the fact that many articles comment on the aftermath of the Gulf War, but do not explicitly mention rising threats or risks, something that our index takes into account.

5. We have constructed and examined broader (GPRAND) and narrower (GPRBASIC) versions of the index around the benchmark index (see Table A.3), thus satisfying data check number 5.

6. We construct a version of the GPR index that excludes articles containing economics and finance related words. The resulting index, plotted in Figure A.13, is nearly identical to the benchmark index, with a correlation of 0.99 (*Data check 6: [...] Does the chosen keyword have as its primary or only meaning the occurrence of the phenomenon of interest?*, and *Data check 7: [...] Does the chosen keyword also result in documents related to the covariates of the occurrence of interest?*).

7. In robustness checks, we use the naïve index as a placebo document-frequency variable in our vector autoregression (VAR) analysis. In particular, there is the possibility that it is not geopolitical risks per se that are bad, but that the overall tendency to discuss geopolitical events rises during recessions. We verify that adding the naïve index to the VAR does not change the predictive power of GPR in the VAR. (*Data check 8: Are there plausible omitted variables that may be correlated both with the document-frequency and its covariates?*)

B.10 Appendix on: Comparison with Other Indicators of Geopolitical Risk

Several studies have constructed quantitative proxies of war intensity or terrorism-related events. One widely used source is the ICB (International Crisis Behavior) database, which provides detailed information on 476 major international crises that occurred during the period from 1918 to 2015. This database has been used in the political science literature as well as in studies on war and economics. The proxy, which counts the number of international crises per month, is plotted alongside the GPR index in the top panel of Figure A.14. The ICB crisis index and the GPR index display some comovement in various historical periods, such as the aftermath of World War I, the Cold War in the early 1960s and late 1970s, the Gulf War, and the Iraq War. But there are also some remarkable differences, such as during World War II, when the ICB crisis index is remarkably low, or during the mid-1990s, when the ICB crisis index is higher than the GPR index. Some differences are due to the different nature of the indexes—the ICB index counts international crises, including those that might receive little press coverage. Moreover, the GPR index displays substantially more high-frequency variation.

The second panel of Figure A.14 compares the GPR index with the national security component of the economic policy uncertainty index (EPU) constructed by Baker, Bloom, and Davis (2016). Like our measure, the national security EPU spikes during the Gulf War, after 9/11, and during the Iraq War. However, the GPR index seems to better capture other spikes in geopolitical risk that are missed by the national security EPU. The correlation between the two measures is 0.69, a plausible value because the national security component of the EPU captures uncertainty about policy responses to events associated with national security (of which geopolitical events are a subset), which is not the same concept as the uncertainty generated by geopolitical events.

Finally, the third panel of Figure A.14 compares the GPR index with an outside measure of political risk related to wars, the U.S. External Conflict Rating (ECR) constructed by the International Country Risk Guide (ICRG). The ratings constructed by the ICRG are largely subjective, as they are based on the insights of various analysts following developments in a particular country or region. The ECR measure moves only occasionally over the sample, changing on average once a year, with more pronounced movements and more frequent changes around 9/11, when both the GPR index and ICRG index spike. The correlation between the two series is 0.41.

C Appendix on: VAR Evidence on the Effects of Geopolitical Risk

C.1 Data Sources

We describe the macroeconomic series used in the VAR first. They are:

- the VIX (CBOE Market Volatility Index; Haver mnemonics: SPVXO@USECON);
- the log of real business fixed investment per capita (FH@USECON—Real Private Fixed Investment—divided by LN16N@USECON—Civilian Noninstitutional Population: 16 Years and Over)
- the log of private hours per capita (LHTPRIVA@USECON—Nonfarm Payrolls in Private Sector—divided by LN16N@USECON)
- the log of the Standard and Poor's 500 index, divided by the Consumer Price Index for All Urban Consumers (SP500@USECON—Stock Price Index: Standard Poor's 500 Composite—divided by PCUN@USECON—CPI-U: All Items)
- the log of the West Texas Intermediate price of oil, divided by the Consumer Price Index for All Urban Consumers (PZTEXP@USECON—Spot Oil Price: West Texas Intermediate—, divided by PCUN@USECON)
- the yield on two-year U.S. Treasuries (FCM2@USECON, 2-Year Treasury Note Yield at Constant Maturity)
- the Chicago Fed National Financial Conditions Index (NFCI: source Chicago Fed National Financial Conditions Index, source: FRED database)
- EPU index, used for the robustness analysis of Figure A.5 (SEPUI@USECON—Economic Policy Uncertainty Index)
- real GDP per capita, used for the exercise described in Figure A.7 (GDPH@USECON, Real Gross Domestic Product—divided by LN16N@USECON)

C.2 Estimation

All VAR models presented in the paper are estimated using Bayesian techniques by imposing an inverse-Wishart prior on the reduced-form VAR parameters. All the results reported in the paper are based on 20,000 draws from the posterior distribution of the structural parameters, where the first 4,000 draws were used as a burn-in period.

C.3 Robustness.

The result from the VAR analysis that changes in geopolitical risk have substantial and significant effects on investment and hours is robust to a variety of alternative specifications. Figure A.5 illustrates our findings. We modify our baseline specification in five alternative ways: (1) we replace the GPR index with a variable that equals the GPR index in the case of the ten largest spikes in the index, and zero otherwise;[4] (2) we increase the number of lags in the VAR from two to four; (3) we estimate a small-scale VAR with only GPR, the two-year yield, the financial conditions index, and either investment or hours; (4) we order the GPR in the Cholesky factorization of the VAR residuals after the financial variables; (5) we add the EPU Index to the VAR. Across all specifications, the effects of a shock to the GPR index are within the 68 percent credible sets of the baseline specification.

[4] The spikes are identified as the ten largest observations of the index divided by its lagged three-year moving average. The impulse response to the shock in the GPR-spikes variable are virtually identical to those that obtain using a dummy indicator variable in place of the spikes.

D Appendix on: Tail Effects of Geopolitical Risk

D.1 Appendix on: Effects on Disaster Probability

The list of countries included in the sample is Argentina, Australia, Belgium, Brazil, Canada, Chile, China, Denmark, Finland, France, Germany, Italy, Japan, Korea, Mexico, Netherlands, Norway, Peru, Portugal, Russia, Spain, Sweden, Switzerland, Taiwan, United Kingdom, and the United States.

For 24 of the 26 countries in our panel, we follow the procedure in Nakamura et al. (2013) and data on consumption to construct the disaster episodes. For China and Russia, the two countries in our panel that are not part of the sample used by Nakamura et al. (2013), we defined disasters as windows of years for which GDP growth is consistently in the bottom fifth of all GDP for that country in our sample. Under our expanded definition of disaster, disaster occur 17.6% of the time, compared to 17.9% in the sample of 24 countries in Nakamura et al. (2013).

The real per capita GDP data are from Barro and Ursúa (2012), extended through 2019 using the World Bank World Development Indicators (WDI) for all countries except Taiwan, for which real per capita GDP is taken from Haver Analytics based on underlying data from national statistical offices (series mnemonics A528GCPC@EMERGE). Growth is calculated using Barro and Ursúa's data until 2005, and the WDI data from 2006 through 2019.

D.2 Appendix on: Quantile Effects of Geopolitical Risk

The initial sample includes the 26 countries used in the disaster probability regressions and listed in subsection D.1.

Data on TFP growth are taken from the Long-Term Productivity Database (version v2.4, updated on October 2020) described in Bergeaud, Cette, and Lecat (2016). The data were retrieved from `http://www.longtermproductivity.com/`. The countries included in the regression are Australia, Belgium, Canada, Chile, Denmark, Finland, France, Germany, Italy, Japan, Mexico, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom, and the United States.

Data on military expenditures as a share of GDP are taken from Roser and Nagdy (2013) and extend through 2016 for the 26 countries in the panel. The data were retrieved from `https://ourworldindata.org/military-spending`. Coverage for six countries is available as early as 1900. The average number of observations per country is 103.

E Appendix on: Geopolitical Risk and Firm-Level Investment

E.1 Appendix on: Details on Industry Exposure Regressions

To compute industry exposure, we use daily stock market returns data for the 49 Industry Portfolios from the Kenneth French data library, which groups NYSE, AMEX, and NASDAQ stocks based on four-digit SIC codes. We also incorporate the daily excess return of the market over the risk-free rate, taken to be the one-month T-bill rate

Stock market-based exposure is measured using the estimated coefficient on GPR from regressions of daily industry portfolio excess returns—calculated as market returns minus daily returns on a T-bill—on daily GPR. Because printed newspapers report event with one day delay, the relevant GPR for stock returns for day t is the geopolitical risk index reported in newspapers for day $t+1$.

E.2 Appendix on: Firm-Level Variables from Compustat

Our firm-level data source is the Compustat North America database. Our firm-level variables are investment rate, cash flows, and Tobin's Q.

1. We construct the investment rate as the ratio of quarterly capital expenditures (DCAPXY, defined as the first difference of CAPXY withi a firm's fiscal year) to the beginning-of-period stock of property, plants, and equipment (lag of PPENTQ). We consider only firms with headquarters located in the United States (Compustat variable LOC is "USA"). We drop the observations where DCAPXY is negative and all observations where PPENTQ is less than \$5 million in chained 2009 dollars. We drop observations where the capital stock (PPENTQ) decreases and then increases (or vice versa) more than fifty percent between two successive quarters. We only include a firm if it has at least ten quarters of nonempty observations. We winsorize the variable at the 1st and 99th percentile.

2. We measure Tobin's Q as the market value of equity plus the book value of assets minus book value of equity plus deferred taxes, all divided by the book value of assets. We normalize cash flows by beginning of period assets.

   We construct Tobin's Q using the quarterly Compustat items PRCCQ (share price at close), CSHOQ (common shares), ATQ (total assets), and CEQQ (common equity). The measure is equal to $\frac{(PRCCQ*CSHOQ)+ATQ-CEQQ}{ATQ}$. We winsorize the variable at the 1st and 99th percentile.

   We construct cash flows using the ratio of Compustat item CHEQ (cash and short-term investments) to beginning-of-period PPE, which is the first lag of PPENTQ in our sample. The variable is winsorized at the 1st and 99th percentile.

3. We match Compustat firms to the 49 Fama-French industries using each firm's unique SIC Code and following the industry definitions in http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/changes_ind.html.

E.3 Appendix on: Search Terms for Firm-Level Geopolitical Risk

We perform text analysis on 157,000 transcripts of quarterly earnings calls of firms listed in U.S. stock markets for the sample 2005-2019 that we are able to match with the corresponding quarterly firm-level Compustat data. We obtain the conference call transcripts from the Fair Disclosure Wire and from Standard & Poor. We construct a firm-quarter variable that counts the occurrence of mentions of geopolitical risks in the earnings call. Specifically, we count the joint occurrences of risk and adverse event words within ten words of 'geopolitical' words, and normalize the number of joint occurrences by the total number of words in the transcript. For instance, if a firm's earnings call reads like "We have been *worried* because of the *war*. Additionally, we have scaled back our investment plans because of *concerns* about *war-related* sanctions," its firm-specific geopolitical risk index will equal $\frac{2}{22}$, where 2 are the instances of mentions, and 22 are the total words in the transcript.

The number of joint occurrences is zero for 81.5 percent of the firm-quarter observations, one for 12.1 percent of observations, two for 3.6 percent of observations, and greater than two for 2.5 percent of observations.

The list of 'geopolitical' terms in the earnings calls is: *war*, *military*, *terror**, *geopolitical*, *conflict*, *"Middle East"*, *Iraq*, *Afghanistan*, *Iran*, *Syria*, *Libya*, *Ukrain**, *Russia**, *"North Korea"*, *Venezuela*, *coup*, *expropriation*, *confiscation*, *nationalism*, *security*, *protest**, *country*, *countries*, *political*, *retaliation*, *unrest*, *geograph**, *troop**, *sanction*, *sanctions*, *embargo*, *wars*, *warfare*, *army*, *navy*, *weapon**, *combat*, *missile**, *immigration*, *diplomacy*.

We require the 'risk-related' terms to be within ten words of one following risk/adverse event terms. The list of risk/act terms is: *risk**, *uncertain**, *variab**, *chance**, *possib**, *pending*, *doubt**, *prospect**, *bet*, *bets*, *betting*, *exposed*, *likel**, *threat**, *probab**, *unknown**, *potential*, *concern**, *tension**, *issue**, *instability*, *cautio**, *fear**, *volatil**, *varying*, *unclear*, *speculative*, *hesitant*, *headwind**, *backlog**, *dispute*, *disrupt**, *worry**, *worries*, *hurdle**, *obstacle**, *disturbance**, *hostil**, *unrest*, *conflict*, *pressure**, *crisis*, *trigger**, *impact*, *peril**, *effect**, *acts*, *attack**, *incident**.

In an earlier version of this paper (see https://www.matteoiacoviello.com/gpr_files/GPR_PAPER_DEC_2019.pdf), the list of geopolitical terms included *China*. We removed references to China in later versions since many firm-level 'GPR hits' in 2019 appeared to refer to the 'trade war' between United States and China.

Table A.1: Most and Least Likely Words in Newspapers on Days of High Geopolitical Tensions, Relative to any Day in the Sample.

Word	**Word Type**	**Term Frequency High-GPR Days**	**Term Frequency Average**	**Odds**
blockad	Event	0.069	0.006	11.6
terror	Event	0.186	0.019	9.7
invas	Act	0.117	0.015	8.0
communiqu		0.064	0.008	7.9
war	Event	1.710	0.237	7.2
terrorist	Event	0.116	0.016	7.2
militari	Event	0.599	0.090	6.6
coup	Event	0.053	0.008	6.3
crisi	Threat	0.254	0.042	6.1
troop	Actor	0.605	0.104	5.8
threat	Threat	0.183	0.034	5.4
armi	Actor	0.744	0.146	5.1
attack	Act	0.720	0.143	5.0
alli	Actor	0.403	0.081	5.0
peac	Event	0.540	0.111	4.8
neutral		0.068	0.015	4.6
combat		0.057	0.012	4.6
invad	Act	0.063	0.014	4.4
bomber	Event	0.057	0.013	4.3
enemi	Actor	0.121	0.028	4.3
missil	Event	0.089	0.021	4.2
fear	Threat	0.259	0.063	4.1
command		0.183	0.046	4.0
conflict	Event	0.077	0.020	3.9
tank		0.069	0.018	3.8
submarin		0.064	0.017	3.8
strike	Act	0.416	0.119	3.5
forc		0.662	0.189	3.5
civilian		0.062	0.018	3.5
soldier		0.145	0.042	3.4
...	...	...	...	...
season		0.006	0.064	0.1
percent		0.007	0.084	0.1
photo		0.004	0.074	0.1
sport		0.008	0.181	0.0
ms		0.001	0.058	0.0

Note: Top 30 words (and bottom 5 words) on days of extremely high geopolitical tensions, relative to any day in the sample (The days of high geopolitical tensions are the days in which adverse geopolitical events are featured in a banner headline on the front page of *The New York Times*). The term frequency on high-GPR days (TFH) is the relative occurrence of the word (expressed in percent) on days in which adverse geopolitical events are featured in a banner headline on the front page of *The New York Times*. The average term frequency (TFA) is the relative occurrence of the word on any day. The 'Odds' column reports TFH/TFA. The sample runs from 1900 and 2019. Proper nouns, stop-words and uncommon (not in the top-400 either on days of high tensions or on any other day) words are excluded from the list. Words that are featured in the headline GPR index are labeled in the 'Word Type' column.

Table A.2: Most Likely Words in Newspapers on Days of High Geopolitical Tensions: Comparing the 1900-59 sample with the 1960-2019 sample.

	Relatively more frequent words in Early Sample				Relatively more frequent words in Late Sample			
Rank	Word	Term Frequency Early	Term Frequency Late	Odds Early	Word	Term Frequency Late	Term Frequency Early	Odds Late
1	communiqu	0.093	0.004	24.8	terrorist	0.352	0.004	87.8
2	submarin	0.091	0.007	12.9	televis	0.103	0.001	77.3
3	leagu	0.146	0.012	12.4	nuclear	0.138	0.003	41.5
4	marshal	0.096	0.008	12.0	missil	0.261	0.007	39.1
5	neutral	0.095	0.010	9.7	terror	0.532	0.023	23.4
6	labor	0.180	0.020	9.1	vow	0.089	0.004	21.1
7	fleet	0.093	0.013	7.3	senior	0.077	0.004	20.5
8	committe	0.215	0.032	6.7	ceasefir	0.088	0.006	15.2
9	red	0.113	0.018	6.3	jet	0.080	0.006	12.9
10	treati	0.175	0.031	5.7	target	0.145	0.020	7.4
11	repli	0.089	0.016	5.4	weapon	0.137	0.022	6.3
12	men	0.344	0.065	5.3	alert	0.084	0.014	6.0
13	deleg	0.094	0.021	4.5	role	0.114	0.021	5.4
14	receiv	0.120	0.028	4.3	sanction	0.077	0.015	5.1
15	confer	0.225	0.059	3.8	coup	0.112	0.024	4.6
16	situat	0.139	0.037	3.8	ground	0.182	0.040	4.5
17	board	0.129	0.034	3.8	respons	0.190	0.042	4.5
18	navi	0.216	0.058	3.7	led	0.086	0.022	4.0
19	present	0.118	0.034	3.5	famili	0.087	0.023	3.8
20	session	0.092	0.027	3.4	region	0.101	0.028	3.7
21	result	0.100	0.030	3.4	combat	0.109	0.032	3.4
22	food	0.085	0.026	3.3	launch	0.106	0.033	3.3
23	servic	0.186	0.059	3.1	just	0.145	0.046	3.1
24	coast	0.084	0.027	3.1	support	0.295	0.097	3.0
25	ship	0.257	0.089	2.9	militari	1.098	0.363	3.0
26	hear	0.086	0.030	2.9	past	0.077	0.026	3.0
27	statement	0.149	0.053	2.8	oil	0.125	0.043	2.9
28	naval	0.117	0.042	2.8	emerg	0.111	0.043	2.6
29	appeal	0.114	0.042	2.7	school	0.100	0.039	2.6
30	repres	0.090	0.034	2.7	began	0.125	0.049	2.5
31	plant	0.087	0.033	2.6	conflict	0.131	0.051	2.5
32	assert	0.093	0.036	2.6	mission	0.080	0.032	2.5
33	discuss	0.101	0.040	2.5	fear	0.436	0.174	2.5
34	gun	0.102	0.041	2.5	try	0.085	0.035	2.4
35	foe	0.105	0.043	2.5	week	0.096	0.040	2.4
36	drive	0.145	0.060	2.4	threat	0.302	0.126	2.4
37	declar	0.295	0.122	2.4	polit	0.197	0.083	2.4
38	vote	0.223	0.093	2.4	attack	1.173	0.506	2.3
39	pass	0.095	0.040	2.4	tank	0.111	0.049	2.3
40	taken	0.109	0.046	2.4	blockad	0.112	0.049	2.3
41	depart	0.192	0.083	2.3	turn	0.140	0.064	2.2
42	headquart	0.099	0.043	2.3	crisi	0.397	0.186	2.1
43	train	0.125	0.054	2.3	assault	0.089	0.042	2.1
44	port	0.090	0.039	2.3	administr	0.220	0.106	2.1
45	sea	0.120	0.053	2.3	withdraw	0.104	0.051	2.0
46	worker	0.130	0.058	2.3	begin	0.152	0.076	2.0
47	enemi	0.147	0.066	2.2	southern	0.077	0.039	2.0
48	line	0.207	0.095	2.2	look	0.080	0.040	2.0
49	point	0.136	0.063	2.2	thousand	0.122	0.062	2.0
50	sent	0.112	0.053	2.1	leader	0.369	0.190	1.9

Note: Top 50 words ranked by relative odds on days of extremely high geopolitical tensions in two subsamples. The early (late) sample runs from 1900 (1960) through 1959 (2019). Proper nouns, stop-words, and uncommon words (not in the top 200) are excluded from the list. The 'Odds Early' ('Odds Late') column is calculated as TFE/TFL (TFL/TFE), where TFE is 'Term Frequency Early', and TFL is 'Term Frequency Late'.

Table A.3: Validation of GPR Index: Subsample Averages and Correlations with Narrative Index

Index	Share of Articles	Index: 1900-1959	Index: 1960-2019	Corr.with Narrative	Corr.with Narrative, 1900-1959	Corr.with Narrative, 1960-2019	Type I error (%)
NARRATIVE	17.9	109.0	91.0				—
GPR	3.6	112.7	87.3	0.76	0.86	0.58	21
GPRNOEW	6.9	102.4	97.6	0.73	0.86	0.51	27
GPRBASIC	3.1	98.9	101.1	0.72	0.84	0.50	31
GPRAND	12.7	123.5	76.5	0.53	0.81	0.35	40

Note: All indexes are normalized to have mean equal to 100 in the sample 1900-2019.
The NARRATIVE GPR is hand-coded scoring articles above the fold in *The New York Times*. The article share for the narrative index is constructed so that a 100 percent share would indicate that every day there is a banner article on geopolitical risks in the print edition of the *New York Times*.
The GPRNOEW index does not exclude from the search the excluded words listed in Table 1.
The GPRBASIC index only searches for the most frequent words highlighted in red in Table 1 and does not exclude from the search the words listed in Table 1.
The GPRAND index replaces the 'N/2' proximity operator in Table 1 with the Boolean operator 'AND'.

Table A.4: Validation of Geopolitical Risk: Historical Frequency of Individual Words

Rank	**Variable**	**Word Type**	**Correlation with GPR**	**Share of Articles**	**Share 1900-1939**	**Share 1940-1979**	**Share 1980-2019**
1	War	WAR	0.86	25.0	25.4	31.8	17.9
2	Revolution	WAR	-0.14	3.9	4.0	3.8	3.9
3	Conflict	WAR	0.13	3.5	2.4	3.3	4.9
4	Hostilities	WAR	0.46	1.0	1.2	1.0	0.8
5	Revolt	WAR	0.13	0.9	1.1	1.0	0.5
6	Coup	WAR	-0.08	0.9	0.5	1.0	1.1
7	Uprising	WAR	-0.03	0.6	0.5	0.5	0.7
8	Insurrection	WAR	-0.10	0.2	0.3	0.1	0.1
9	Geopolitical	WAR	-0.09	0.1	0.0	0.0	0.2
1	Concern	RISK	-0.20	16.2	14.2	15.0	19.5
2	Threat	RISK	-0.03	10.5	8.0	10.0	13.7
3	Trouble	RISK	-0.37	10.5	10.8	9.2	11.6
4	Fear	RISK	-0.22	9.9	9.6	8.2	11.9
5	Warning	RISK	0.17	9.8	7.5	10.2	11.6
6	Doubt	RISK	-0.07	9.0	11.7	7.7	7.7
7	Danger	RISK	0.02	7.8	8.6	6.8	8.0
8	Risk	RISK	-0.15	5.6	2.4	3.5	10.8
9	Dispute	RISK	-0.17	4.7	3.2	5.0	5.9
10	Crisis	RISK	0.00	4.4	2.3	4.2	6.6
11	Tension	RISK	-0.12	2.0	0.6	2.1	3.3
12	Inevitable	RISK	0.06	1.5	1.5	1.4	1.6
13	Peril	RISK	0.37	1.4	1.7	1.3	1.1
14	Menace	RISK	0.33	1.1	1.8	1.0	0.4
15	Scare	RISK	-0.19	0.8	0.6	0.6	1.1
16	Imminent	RISK	0.29	0.7	0.6	0.6	0.9
17	Footing	RISK	0.08	0.4	0.5	0.3	0.5
18	Brink	RISK	-0.13	0.4	0.3	0.4	0.7
1	Begin	WARBEGIN	-0.16	34.5	28.2	32.1	43.1
2	Start	WARBEGIN	-0.16	34.3	26.7	32.8	43.3
3	Declare	WARBEGIN	0.14	14.6	23.1	13.4	7.3
4	Launch	WARBEGIN	0.13	4.0	2.5	3.8	5.8
5	Outbreak	WARBEGIN	0.58	1.0	1.3	0.9	0.7
6	Breakout	WARBEGIN	0.17	0.8	0.6	0.9	1.0
7	Proclamation	WARBEGIN	0.39	0.5	0.8	0.5	0.3
1	Army	ACTOR	0.79	12.6	13.2	17.1	7.5
2	Navy	ACTOR	0.65	6.9	7.9	9.2	3.5
3	Allied	ACTOR	0.83	4.7	4.4	5.9	3.8
4	Enemy	ACTOR	0.87	4.4	4.8	5.5	2.8
5	Foe	ACTOR	0.59	2.0	2.6	2.2	1.2
6	Rebels	ACTOR	-0.18	1.1	0.8	0.8	1.6
7	Aerial	ACTOR	0.61	0.9	1.0	1.2	0.5
8	Insurgents	ACTOR	-0.26	0.7	0.8	0.4	1.1
1	Sanction	BUILDUP	-0.13	1.4	1.2	0.9	2.2
2	Buildup	BUILDUP	0.13	1.0	0.6	1.3	1.2
3	Mobilize	BUILDUP	0.72	0.8	0.8	1.1	0.7
4	Blockade	BUILDUP	0.44	0.5	0.6	0.6	0.3
5	Embargo	BUILDUP	0.26	0.5	0.5	0.4	0.5
6	Ultimatum	BUILDUP	0.28	0.3	0.5	0.3	0.2
7	Quarantine	BUILDUP	-0.03	0.2	0.4	0.1	0.1
1	Drive	FIGHT	0.14	17.0	13.5	17.4	20.2
2	Attack	FIGHT	0.73	13.2	11.8	13.7	14.2
3	Advance	FIGHT	0.46	11.8	13.7	11.7	9.8
4	Strike	FIGHT	0.07	7.5	6.9	8.3	7.2
5	Launch	FIGHT	0.13	4.0	2.5	3.8	5.8
6	Raid	FIGHT	0.72	3.1	3.0	3.6	2.7
7	Shell	FIGHT	0.78	3.1	2.7	3.5	3.1
8	Invasion	FIGHT	0.73	3.0	2.8	3.5	2.8
9	Offensive	FIGHT	0.41	2.8	1.5	2.6	4.4
10	Clash	FIGHT	0.01	2.1	2.1	2.1	2.2
1	Military	MILITARY	0.81	11.3	8.2	13.9	11.9
2	Troops	MILITARY	0.87	6.0	6.0	7.2	4.6
3	Bomb	MILITARY	0.70	5.5	2.5	7.9	6.1
4	Arms	MILITARY	0.35	5.2	5.1	5.0	5.3
5	Weapon	MILITARY	0.03	4.1	1.8	4.2	6.3
6	Missile	MILITARY	-0.02	1.3	0.2	1.7	2.1
7	Warhead	MILITARY	-0.06	0.2	0.0	0.2	0.3
1	Peace	PEACE	0.58	7.6	8.2	8.7	5.8
2	Treaty	PEACE	-0.01	2.7	3.5	2.9	1.7
3	Parley	PEACE	0.16	0.7	0.9	1.1	0.0
4	Truce	PEACE	0.06	0.5	0.4	0.8	0.5
5	Armistice	PEACE	0.25	0.5	0.8	0.6	0.1
1	Threat	PEACEDISRUPT	-0.03	10.5	8.0	10.0	13.7
2	Reject	PEACEDISRUPT	-0.12	5.1	2.7	5.6	6.9
3	Peril	PEACEDISRUPT	0.37	1.4	1.7	1.3	1.1
4	Disrupt	PEACEDISRUPT	-0.06	1.3	0.4	1.2	2.4
5	Menace	PEACEDISRUPT	0.33	1.1	1.8	1.0	0.4
6	Boycott	PEACEDISRUPT	-0.19	0.8	0.5	0.9	0.9
1	Terrorism/t	TERROR	0.00	2.0	0.3	0.9	4.8
2	Guerrilla	TERROR	0.00	1.0	0.1	1.3	1.7
3	Hostage	TERROR	-0.02	0.6	0.1	0.4	1.3
1	Kill	TERRORACT	-0.05	14.8	13.1	13.0	18.2
2	Act	TERRORACT	-0.03	14.5	16.0	13.8	13.7
3	Attack	TERRORACT	0.70	10.8	9.4	11.2	11.8
4	Strike	TERRORACT	0.07	7.5	6.9	8.3	7.2
5	Bomb	TERRORACT	0.70	5.5	2.5	7.9	6.1
6	Hijack	TERRORACT	0.02	0.3	0.0	0.3	0.6

Note: The table shows key summary statistics for the risk-related, act-related, and war-related words used in the construction of the GPR index.

Table A.5: Validation of Geopolitical Risk: Historical Frequency of Selected Word Combinations

Rank	**Search Terms**	**Bigram Type**	**Correlation with GPR**	**Share of Articles**	**Share 1900-1939**	**Share 1940-1979**	**Share 1980-2019**
1	War Escalation Terms	—	0.87	0.97	0.93	1.45	0.52
2	Military Buildup Terms	—	0.62	0.93	0.73	1.07	0.98
3	War Begin Terms	—	0.84	0.82	1.14	0.93	0.40
4	War Risk Terms	—	0.73	0.52	0.63	0.63	0.30
5	Nuclear Risk Terms	—	0.00	0.39	0.00	0.61	0.58
6	Terror Act Terms	—	0.06	0.32	0.02	0.25	0.71
7	Peace Risk Terms	—	0.23	0.13	0.12	0.19	0.07
8	Terror Risk Terms	—	0.02	0.11	0.01	0.05	0.26
1	Risk Words N/2 War	War Threat	0.75	0.79	0.89	0.97	0.52
2	Risk Words N/2 Conflict	War Threat	0.03	0.11	0.08	0.11	0.16
3	Risk Words N/2 Revolution	War Threat	0.05	0.07	0.11	0.05	0.04
4	Risk Words N/2 Revolt	War Threat	0.11	0.03	0.05	0.03	0.01
5	Risk Words N/2 Hostilities	War Threat	0.09	0.03	0.03	0.03	0.03
6	Risk Words N/2 Coup	War Threat	-0.03	0.02	0.01	0.02	0.03
7	Risk Words N/2 Uprising	War Threat	0.02	0.02	0.03	0.01	0.01
8	Risk Words N/2 Geopolitical	War Threat	-0.06	0.01	0.00	0.00	0.02
9	Risk Words N/2 Insurrection	War Threat	-0.00	0.00	0.01	0.00	0.00
1	Military Words AND Sanction	Military Buildups	-0.03	0.52	0.28	0.25	1.05
2	Military Words AND Mobilize	Military Buildups	0.71	0.42	0.46	0.52	0.29
3	Military Words AND Buildup	Military Buildups	0.22	0.38	0.12	0.57	0.45
4	Military Words AND Blockade	Military Buildups	0.46	0.26	0.24	0.33	0.21
5	Military Words AND Embargo	Military Buildups	0.20	0.23	0.17	0.19	0.33
6	Military Words AND Ultimatum	Military Buildups	0.39	0.13	0.17	0.13	0.09
7	Military Words AND Quarantine	Military Buildups	0.10	0.05	0.08	0.04	0.03
1	Nuclear Weapons AND Risk Words	Nuclear Threat	-0.09	0.43	0.00	0.40	0.90
2	Nuclear War AND Risk Words	Nuclear Threat	-0.03	0.16	0.00	0.19	0.29
3	Atom Bomb AND Risk Words	Nuclear Threat	0.08	0.15	0.00	0.32	0.12
4	Nuclear Test AND Risk Words	Nuclear Threat	-0.00	0.06	0.00	0.11	0.07
5	Nuclear Bomb AND Risk Words	Nuclear Threat	-0.07	0.06	0.00	0.06	0.12
6	Nuclear Missile AND Risk Words	Nuclear Threat	-0.05	0.06	0.00	0.03	0.14
7	Hydrogen Bomb AND Risk Words	Nuclear Threat	0.00	0.05	0.00	0.12	0.03
8	Atomic War AND Risk Words	Nuclear Threat	0.04	0.03	0.00	0.08	0.01
9	H Bomb AND Risk Words	Nuclear Threat	0.01	0.03	0.00	0.07	0.01
1	War-Begin Words N/2 War	War Begin	0.86	1.53	1.92	1.68	0.98
2	War-Begin Words N/2 Revolution	War Begin	-0.03	0.13	0.17	0.10	0.11
3	War-Begin Words N/2 Hostilities	War Begin	0.49	0.08	0.14	0.07	0.02
4	War-Begin Words N/2 Conflict	War Begin	0.28	0.06	0.04	0.04	0.08
5	War-Begin Words N/2 Revolt	War Begin	0.07	0.04	0.05	0.04	0.02
6	War-Begin Words N/2 Uprising	War Begin	-0.06	0.03	0.02	0.02	0.07
7	War-Begin Words N/2 Coup	War Begin	-0.04	0.01	0.01	0.01	0.02
8	War-Begin Words N/2 Insurrection	War Begin	-0.04	0.01	0.01	0.01	0.00
9	War-Begin Words N/2 Geopolitical	War Begin	-0.01	0.00	0.00	0.00	0.00
1	Actor Words N/2 Attack	War Escalation	0.85	0.72	0.68	1.00	0.47
2	Actor Words N/2 Advance	War Escalation	0.77	0.20	0.27	0.25	0.07
3	Actor Words N/2 Drive	War Escalation	0.79	0.18	0.21	0.25	0.07
4	Actor Words N/2 Invasion	War Escalation	0.68	0.17	0.14	0.26	0.12
5	Actor Words N/2 Raid	War Escalation	0.75	0.13	0.11	0.21	0.07
6	Actor Words N/2 Offensive	War Escalation	0.74	0.12	0.09	0.19	0.08
7	Actor Words N/2 Launch	War Escalation	0.71	0.09	0.06	0.13	0.08
8	Actor Words N/2 Shell	War Escalation	0.76	0.09	0.09	0.12	0.04
9	Actor Words N/2 Strike	War Escalation	0.62	0.08	0.07	0.11	0.06
10	Actor Words N/2 Clash	War Escalation	0.21	0.08	0.05	0.09	0.08
1	Terror Words N/2 Attack	Terror Act	0.07	0.44	0.01	0.11	1.22
2	Terror Words N/2 Kill	Terror Act	-0.05	0.13	0.01	0.15	0.24
3	Terror Words N/2 Act	Terror Act	-0.01	0.11	0.01	0.06	0.25
4	Terror Words N/2 Bomb	Terror Act	-0.07	0.08	0.01	0.06	0.16
5	Terror Words N/2 Strike	Terror Act	0.06	0.03	0.00	0.01	0.06
6	Terror Words N/2 Hijack	Terror Act	0.03	0.02	0.00	0.02	0.04

Note: The table shows key summary statistics for selected two-word combinations used in the construction of the GPR index. The first 8 entries in the table correspond to the eight meta-categories behind the construction of the index. The next entries are selected slices of the meta-categories, calculated without removing from the search the excluded words listed in Table 1: for this reason, the entries in a category may be larger than the meta-categories themselves.

Table A.6: Correlations of Geopolitical Risk with Selected News-based Indexes of Other Phenomena

Keyword	**Correlation**	**Share:1900-2019**	**Share:1900-59**	**Share:1960-2019**
GPR		3.61	4.07	3.15
INFLATION	0.30	7.48	6.49	8.47
SPORT	0.07	2.30	0.87	3.74
OLYMPIC	-0.03	1.17	0.42	1.91
DISASTER	-0.16	1.74	1.08	2.40
FLU	0.01	0.73	0.76	0.70
PRESIDENT	-0.03	1.47	0.57	2.37
CAMPUS	0.05	0.13	0.07	0.19
MURDER	-0.38	5.94	4.90	6.99
COALSTRIKE	-0.06	0.53	0.87	0.19
WEDDING	0.16	3.95	4.56	3.33

Note: Pairwise correlations of GPR Index with selected indexes capturing selected news. Specifically, we construct news-based indexes of the phenomena above calculating the share of articles containing any of the following words or word combinations:
INFLATION: inflation* OR ((price* OR wage* OR cost*) N/2 (rise OR rising OR high* OR increas*))
SPORT: (olympics OR olympiad OR "olympic games" OR "world cup" OR "world series").
OLYMPIC: (olympics OR olympiad OR "olympic games").
DISASTER: (hurricane* OR earthquake* OR tsunami* OR wildfire* OR tornado*).
FLU: (flu OR influenza).
PRESIDENTELECTION: (president* N/2 election*).
CAMPUSPROTEST: (campus OR college* OR university* OR school*) N/2 (riot* OR protest*).
MURDER: (murder* OR homicide*).
COALSTRIKE: (coal AND strike).
WEDDING: (wedding).

Table A.7: Geopolitical Risk and Firm-Level Investment: Robustness Analysis

$IK(t+2)$	(1)	(2)
ΔGPR $\times$ Industry Exposure	-0.19	-0.18
	(0.08)	(0.17)
ΔGPR	-1.72	
	(1.32)	
Cash Flow	2.72	2.78
	(0.46)	(0.46)
Tobin's Q	8.91	7.93
	(1.68)	(1.56)
$IK(t-1)$	0.31	0.30
	(0.01)	(0.01)
Observations	374,727	374,727
Firm Fixed Effects	Yes	Yes
Time Effects	No	Yes
R-squared	0.45	0.47
Sample	85Q1-19Q4	85Q1-19Q4

Standard errors in parentheses

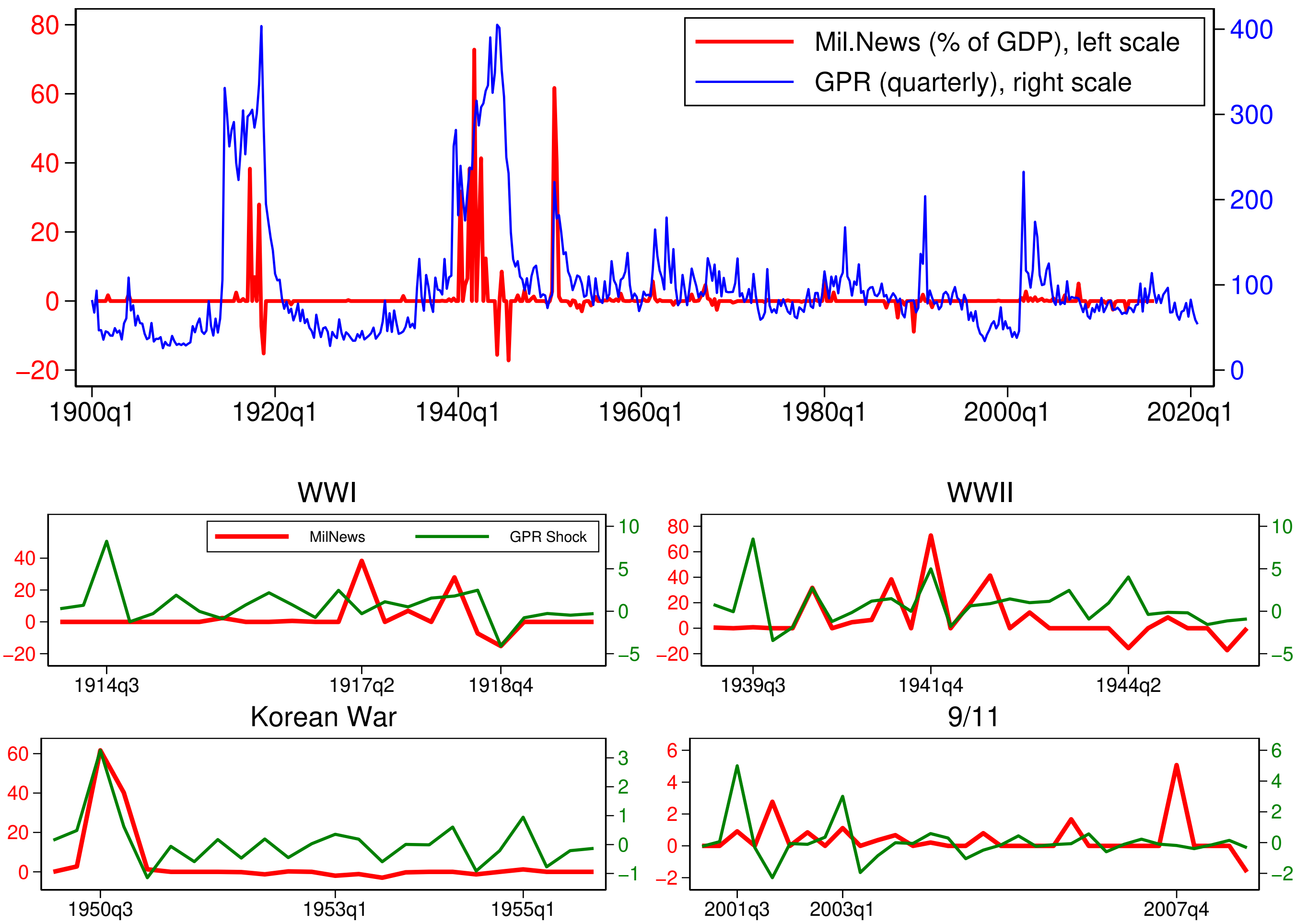
Note: The table shows robustness results from regressions of firm-level investment on geopolitical risk at the industry level. In the main text, the industry exposure measure is a dummy variable equal to one for industries with above-median exposure, and zero otherwise. Here we replace the dummy variable used in columns 1 and 2 of Table 5 with the beta coefficients estimated from equation 4 in the main text, with the sign switched so that positive values indicate high exposure.
The dependent variable is IK (100 times the log of the investment rate) two quarters ahead. All variables (except the dummy exposure variable) are standardized. The standard errors are clustered by industry and quarter.

Table A.8: Granger Causality Tests

Variable Groups	(1) LGPR	(2) LGPRA	(3) LGPRT
Macro	1.02	0.87	0.91
	(0.42)	(0.55)	(0.52)
Financial	1.33	1.34	1.71
	(0.24)	(0.24)	(0.12)
Uncertainty	1.09	0.37	1.50
	(0.36)	(0.90)	(0.18)
LGPR	106.88		
	(0.00)		
LGPRA		136.90	0.91
		(0.00)	(0.44)
LGPRT		0.30	49.40
		(0.83)	(0.00)
Adj. R^2	0.60	0.63	0.52

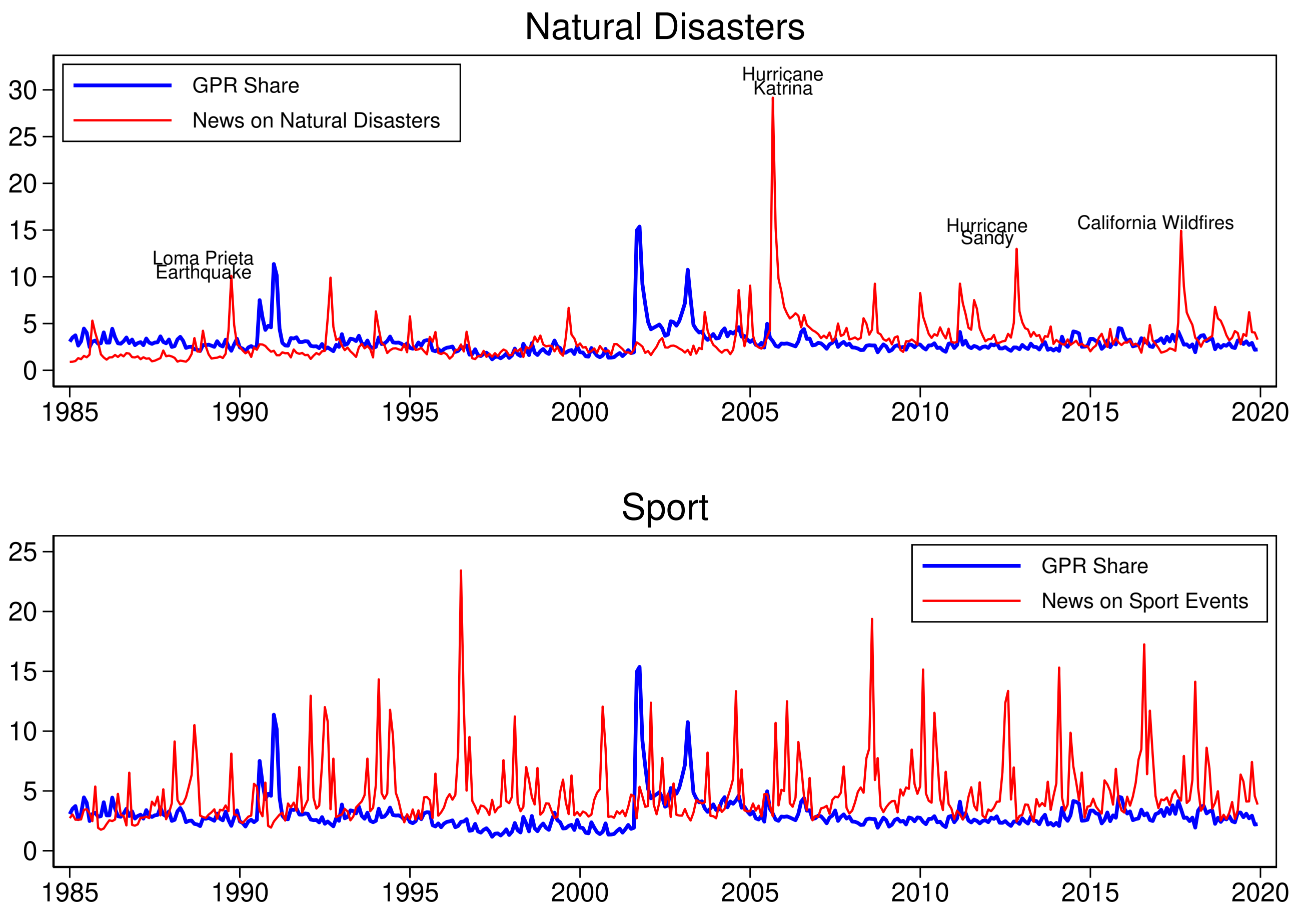
Note: The entries in the table are the test statistics and p-values (in parentheses) for the joint hypothesis that all lags of the variables included in each group are equal to zero. See Appendix B.5 for additional details.

Figure A.1: Comparison with Military Spending News Variable: Detail on Specific Events



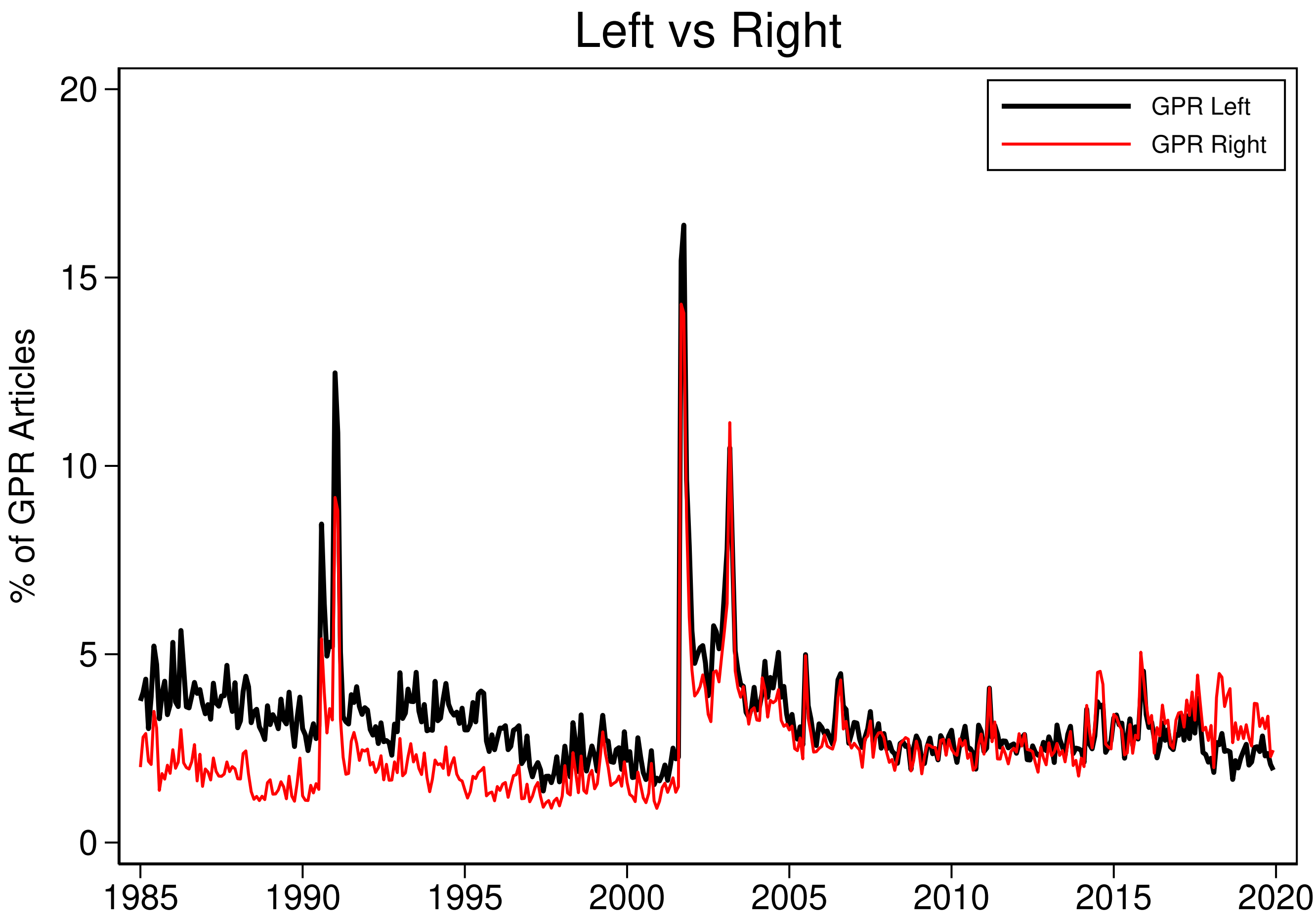
Note: Detailed Time-Series Comparison of the quarterly GPR Index (top panel) and selected quarterly GPR shocks (bottom four panels) with Military Spending News variable from Ramey (2011). The GPR shocks are calculated as the residual of a monthly autoregression of geopolitical risk on three lags, averaged over the quarter, and standardized.

Figure A.2: GPR and News on Natural Disasters and Sport Events



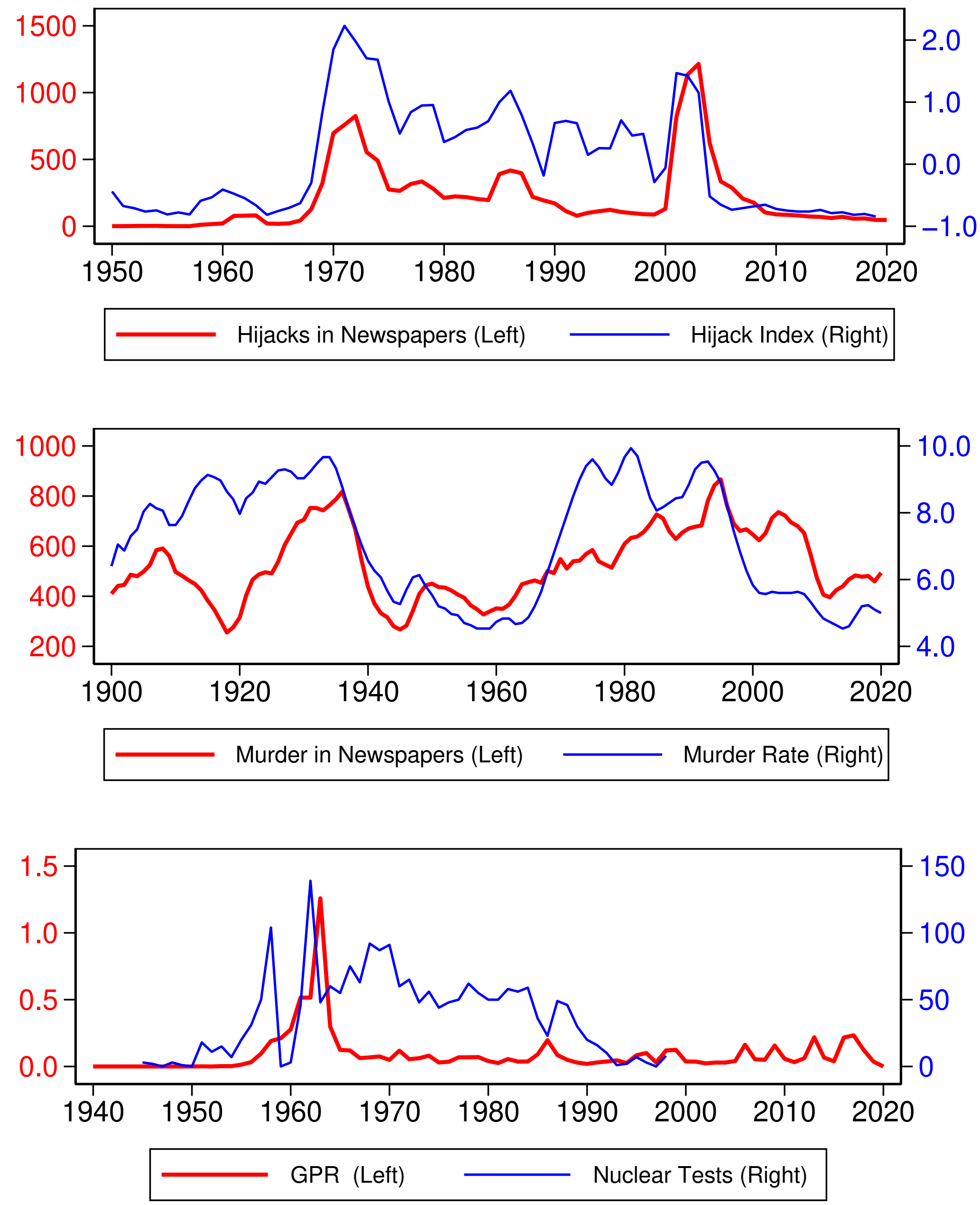
Note: The top panel of the figure compares the recent GPR index with a news-based index of natural disasters, constructed by counting the share of newspapers articles mentioning any of the following words: earthquake(s), hurricane(s), tornado(es), tsunami(s), or wildfire(s).
The bottom panel compares the GPR index with a news-based index of sport popularity, constructed by counting the share of articles mentioning: ‘Olympics” OR ‘olympiad’ OR ‘Olympic Games” OR ‘World Cup” OR ‘World Series.”

Figure A.3: GPR and Political Slant



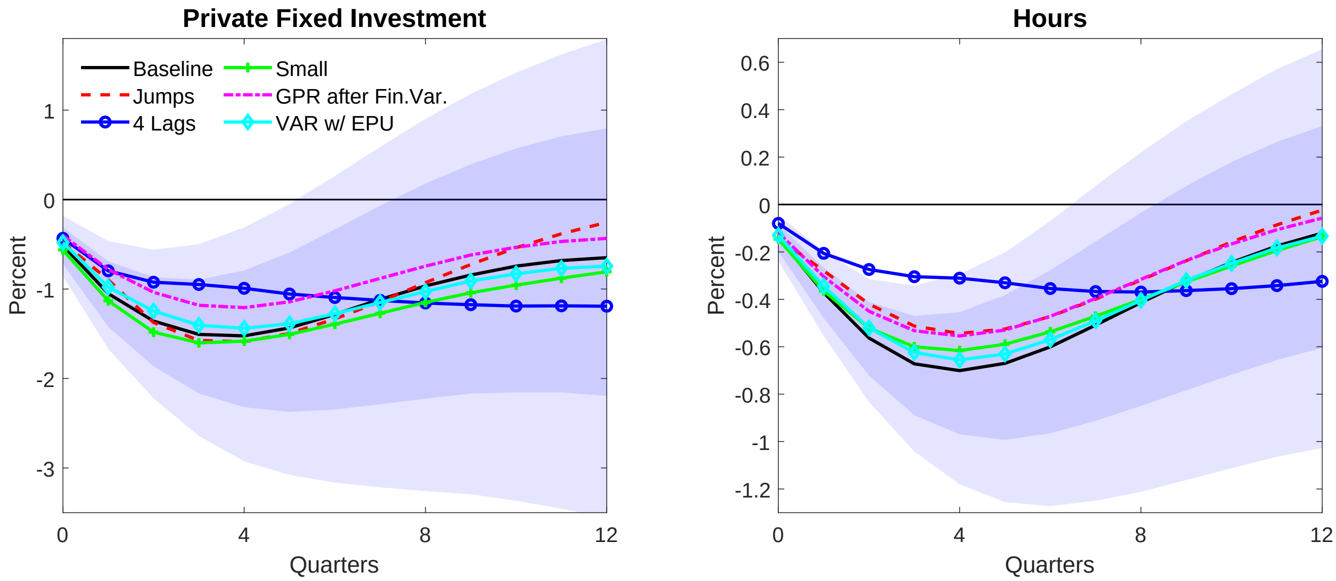
Note: Geopolitical Risk Index for left-leaning and right-leaning newspapers. See text for list of newspapers.

Figure A.4: Hijackings, Murders, Nuclear Tests and Media Mentions



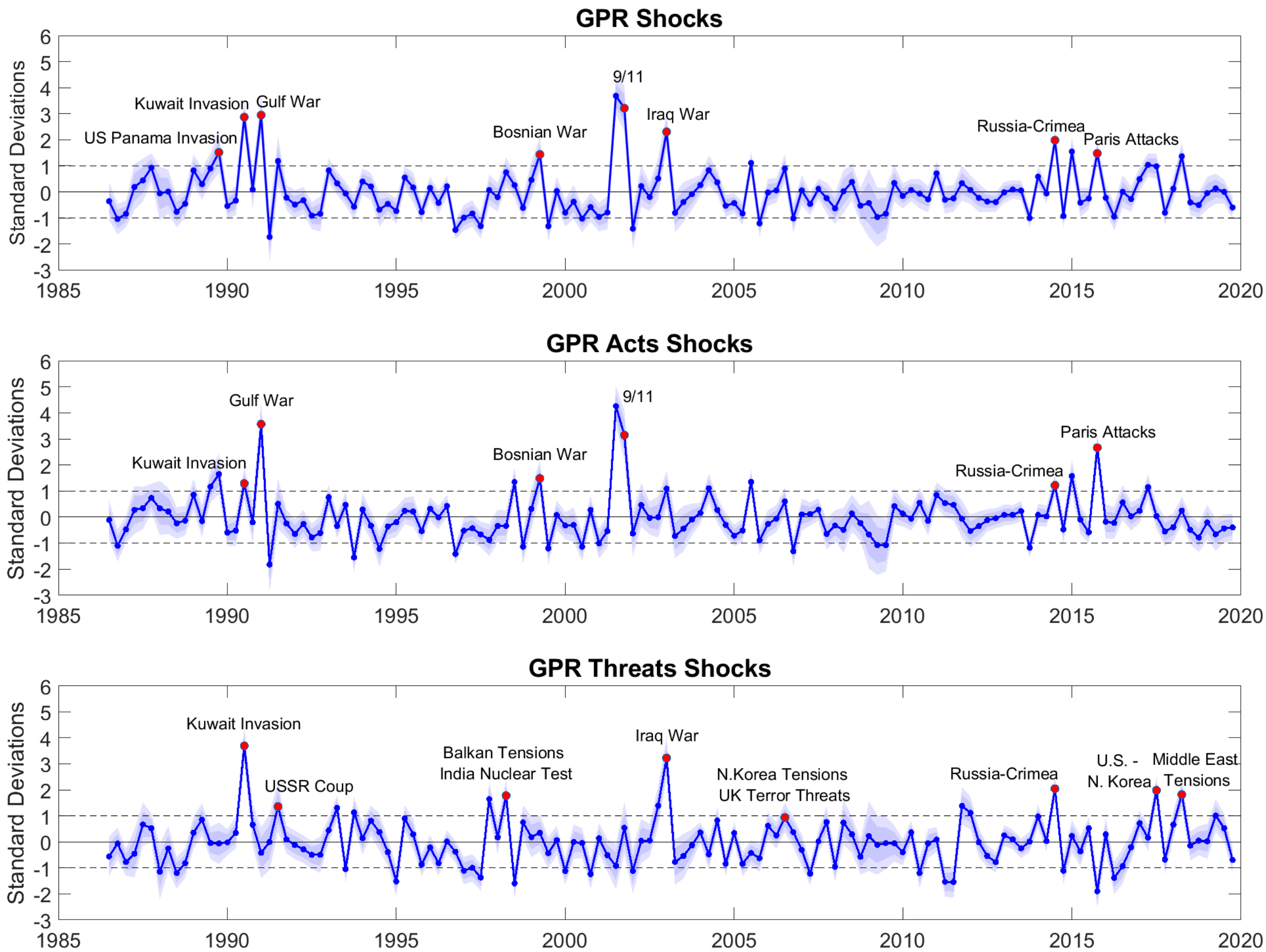
Note: Top panel: Comparison between monthly number of newspaper articles containing the expression hijack* N/2 (plane OR airplane OR air OR aircraft OR airline*) and the first principal component of (1) global number of hijacking incidents and (2) fatalities from hijacking incidents (source: Aviation Safety Network).
Middle panel: Comparison between number of newspaper articles containing the expressions 'was murdered" OR 'was slain" OR 'was shot and killed" and the U.S. murder rate (sources: Eckberg (1995) and http://www.disastercenter.com/).
Bottom panel: Comparison between share of number of newspaper articles containing the expression 'nuclear test" and one risk-related word, and total nuclear tests in the world (source: https://ourworldindata.org/nuclear-weapons).

Figure A.5: The Impact of Increased Geopolitical Risk: Robustness



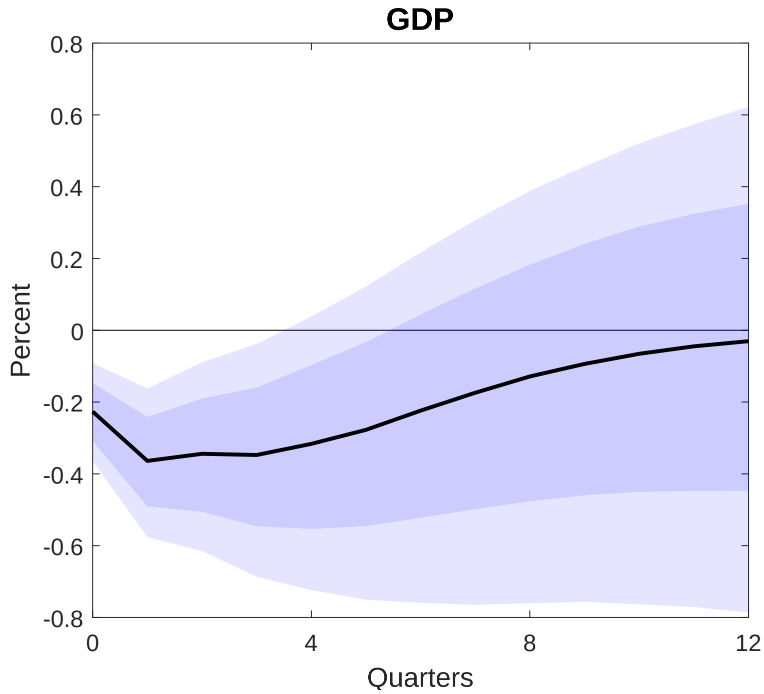
Note: The black solid line depicts the median impulse response of private fixed investment and hours to a two-standard-deviations increase in the GPR index under the benchmark model, with their respective 68 and 90 percent pointwise credible sets. The other lines depict median responses from alternative specifications of the VAR discussed in the main text.

Figure A.6: Time Series of the VAR-identified Geopolitical Shocks



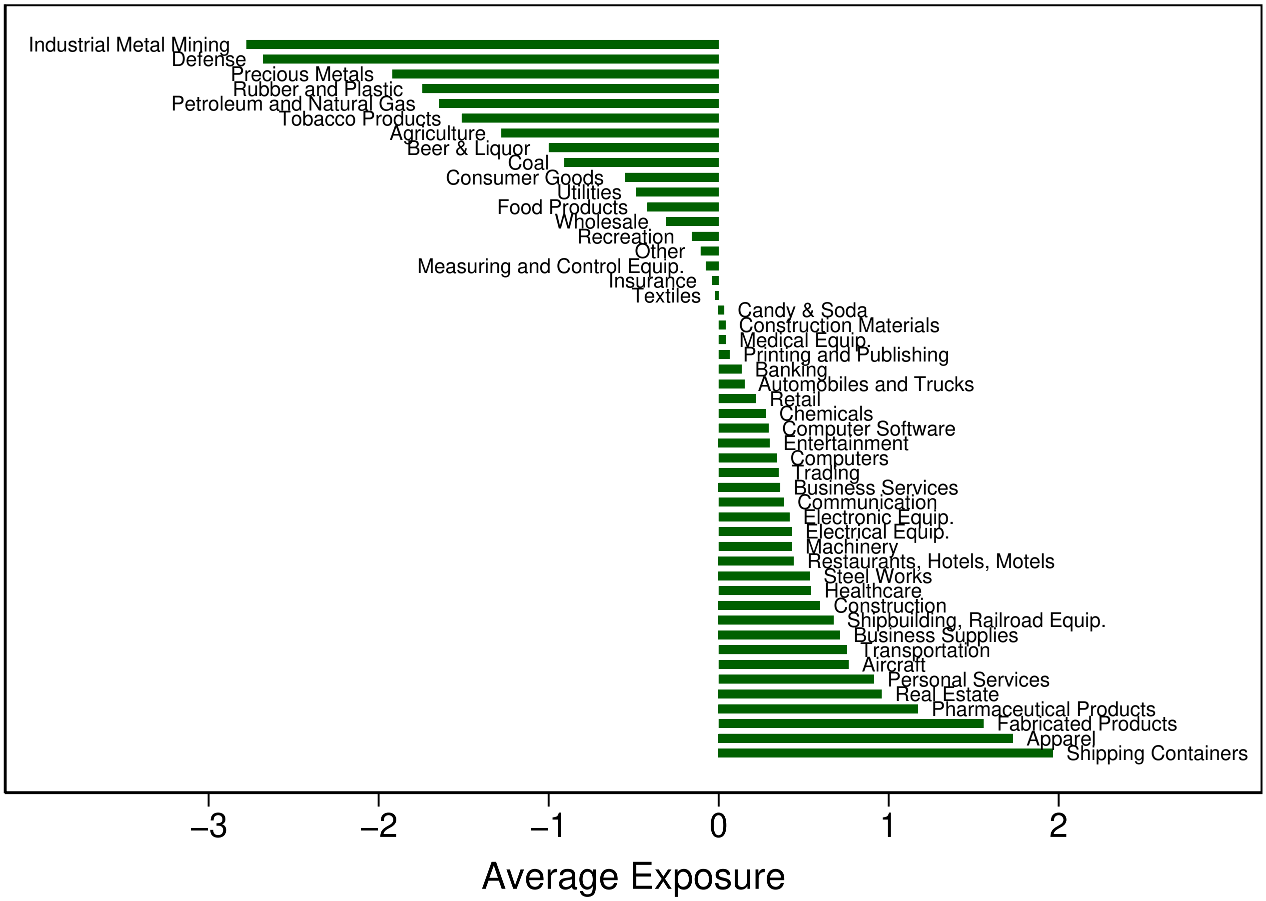
Note: The solid-dotted lines plot the median time series of GPR shocks (top panel, VAR described in subsection III.A) and of GPR Acts and GPR Threats shocks (middle and low panel, VAR described in subsection III.B). Shaded areas are 68 percent and 90 percent credible sets.

Figure A.7: The Impact of Increased Geopolitical Risk on GDP



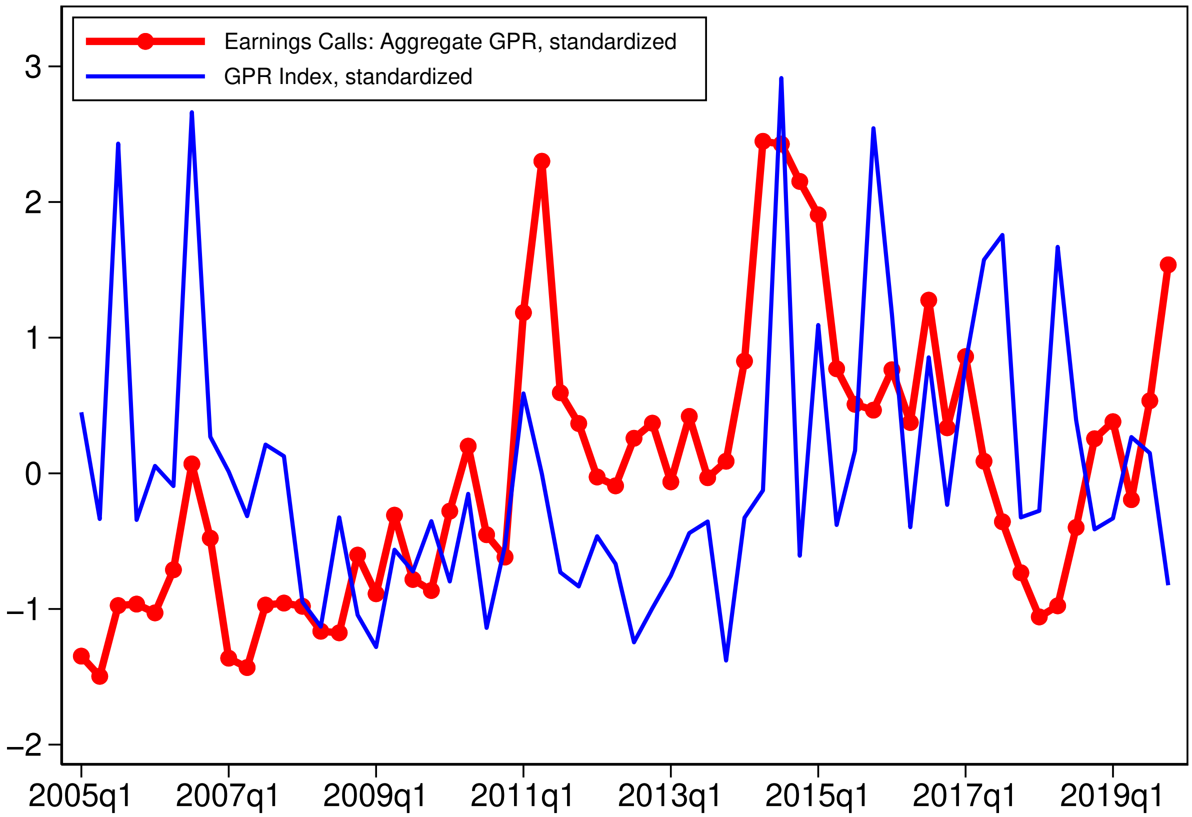
Note: The black solid line depicts the median impulse response of GDP to a two-standard-deviations increase in the GPR index. The VAR model is the same as the baseline specification of Section III.A with the addition of GDP (ordered after GPR). The dark and light shaded bands represent the 68 percent and 90 percent pointwise credible sets, respectively.

Figure A.8: Exposure to GPR by Industry Using Stock Returns



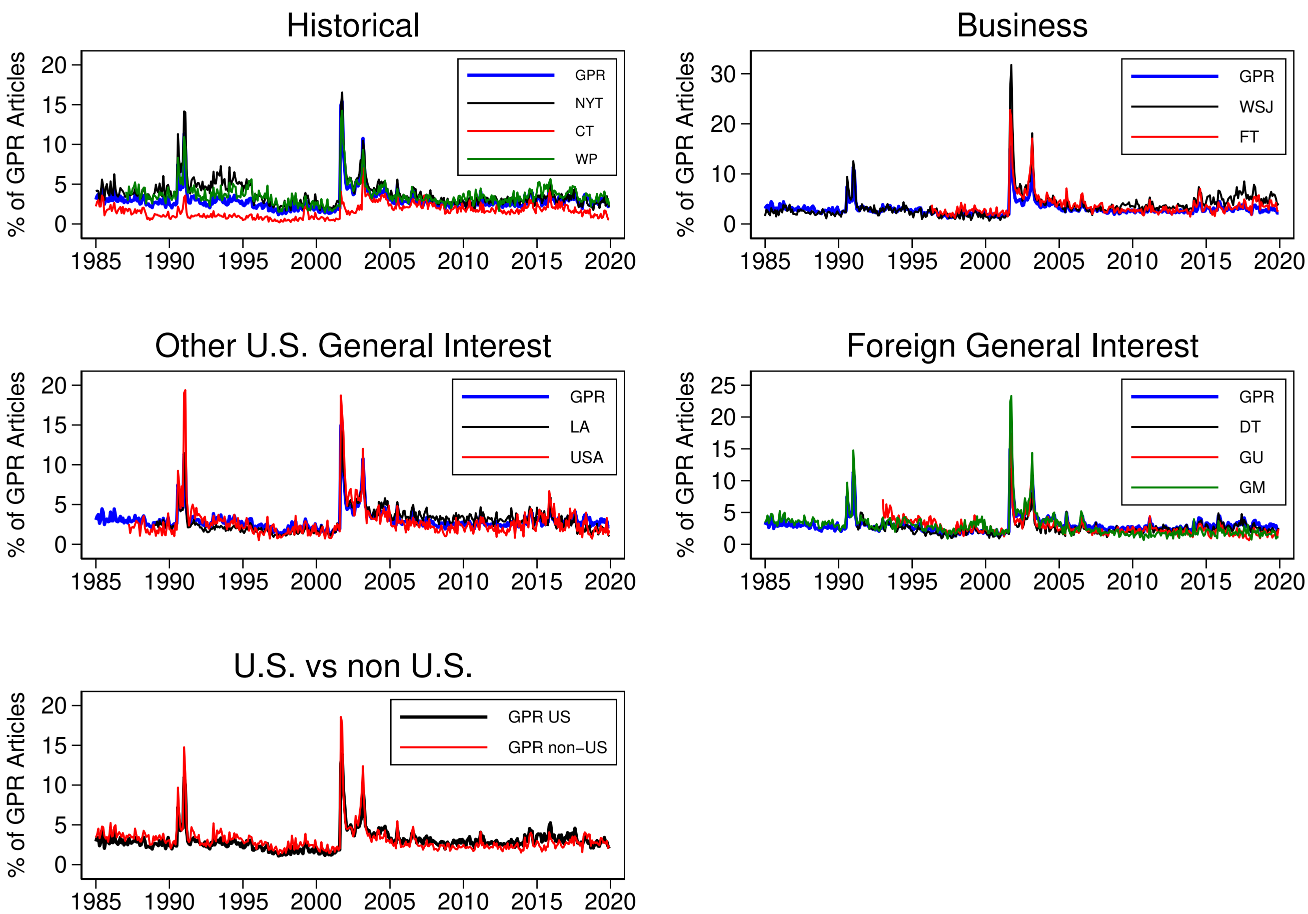
Note: Industry exposure to adverse geopolitical risk: values estimated on sample from 1985 through 2019, standardized to have zero mean and unit standard deviation. Higher values indicate a larger decline in industry daily stock returns after an increase in daily geopolitical risk.

Figure A.9: GPR Index and Firms' Perception of Geopolitical Risk



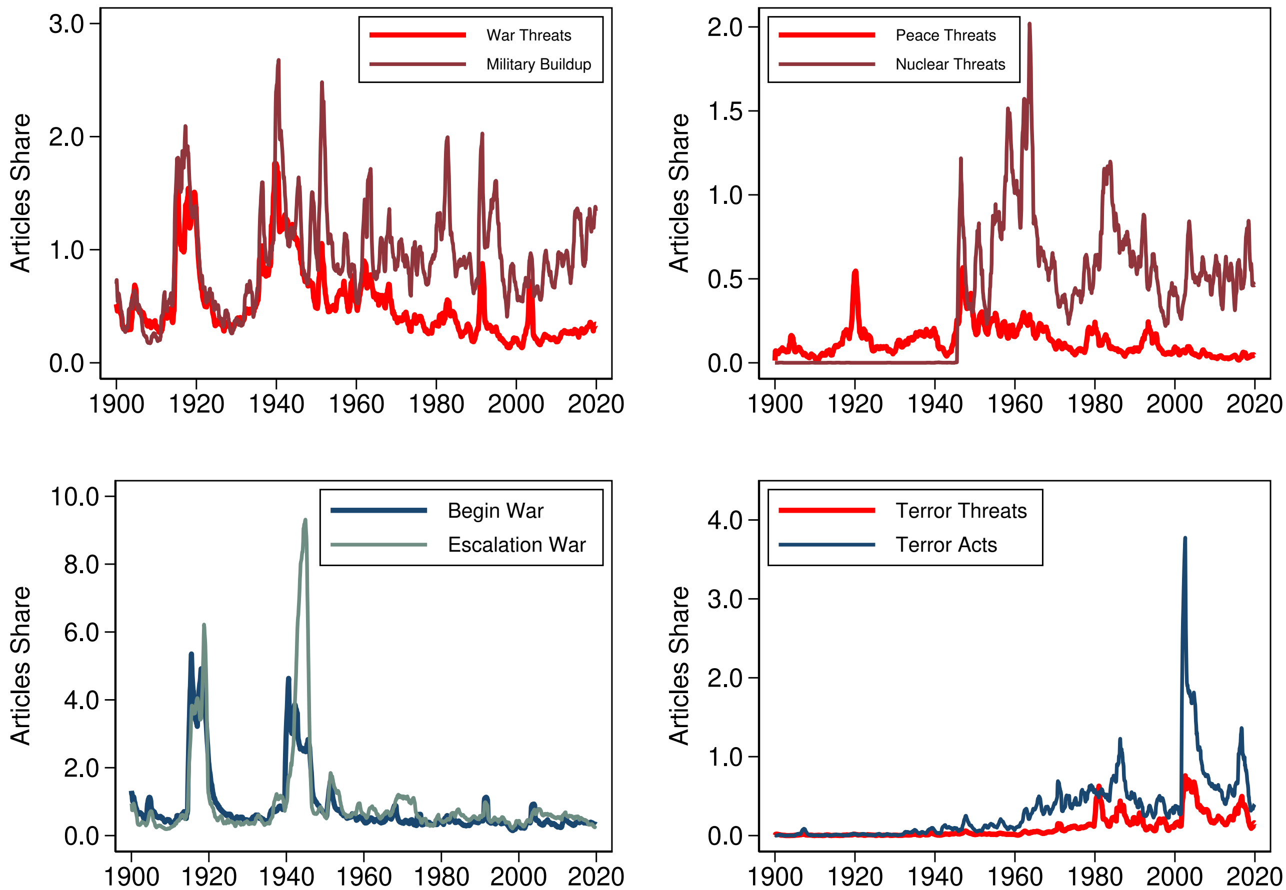
Note: The GPR Index from the listed firms' earnings calls transcripts is the average share of phrases mentioning geopolitical risks. Both measures have been transformed to have zero mean and unit standard deviation in the 2005-2019 period.

Figure A.10: Share of GPR Articles by Individual Newspapers



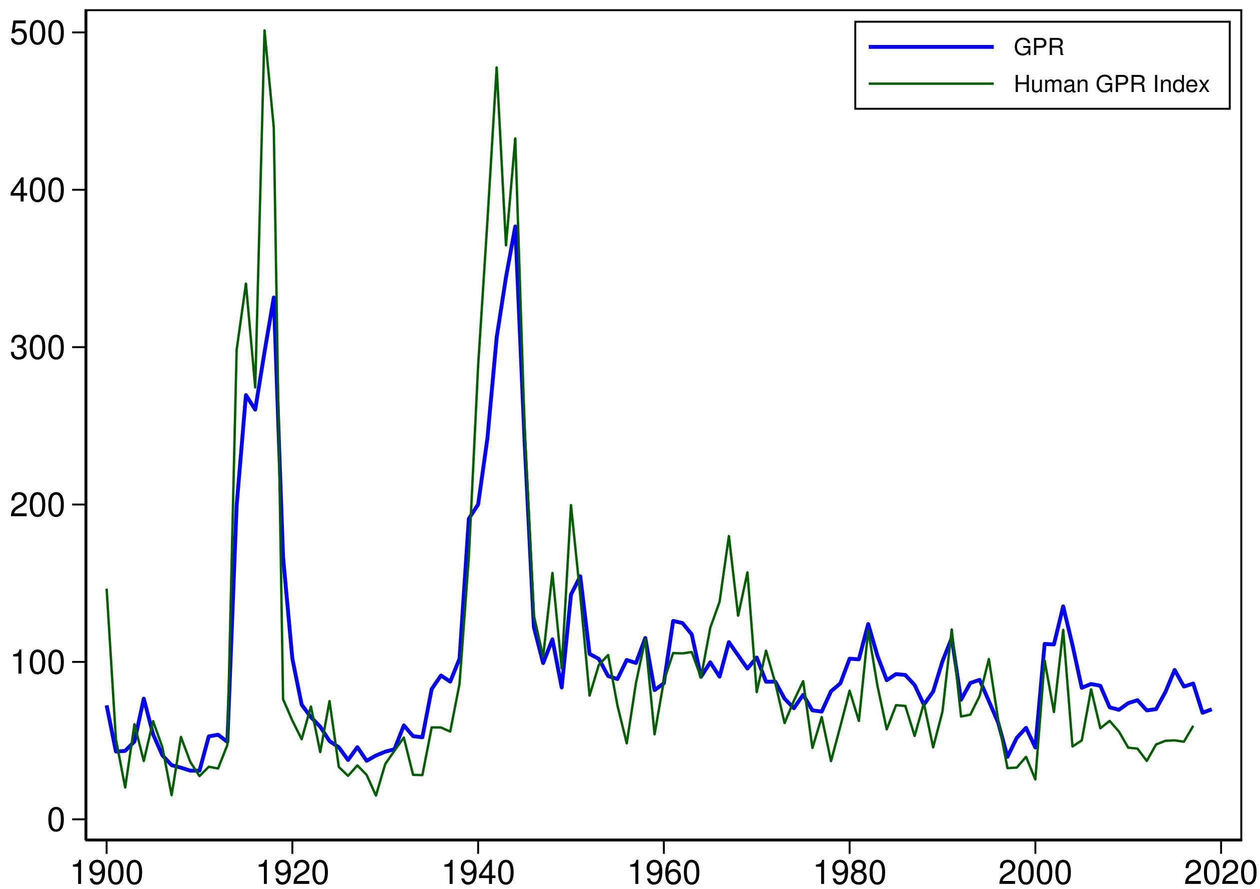
Note: Each panel plots the share of articles containing words related to geopolitical risk for each of the 10 newspapers used to construct the baseline GPR index.

Figure A.11: The Geopolitical Risk Index:
Contribution of the Search Categories



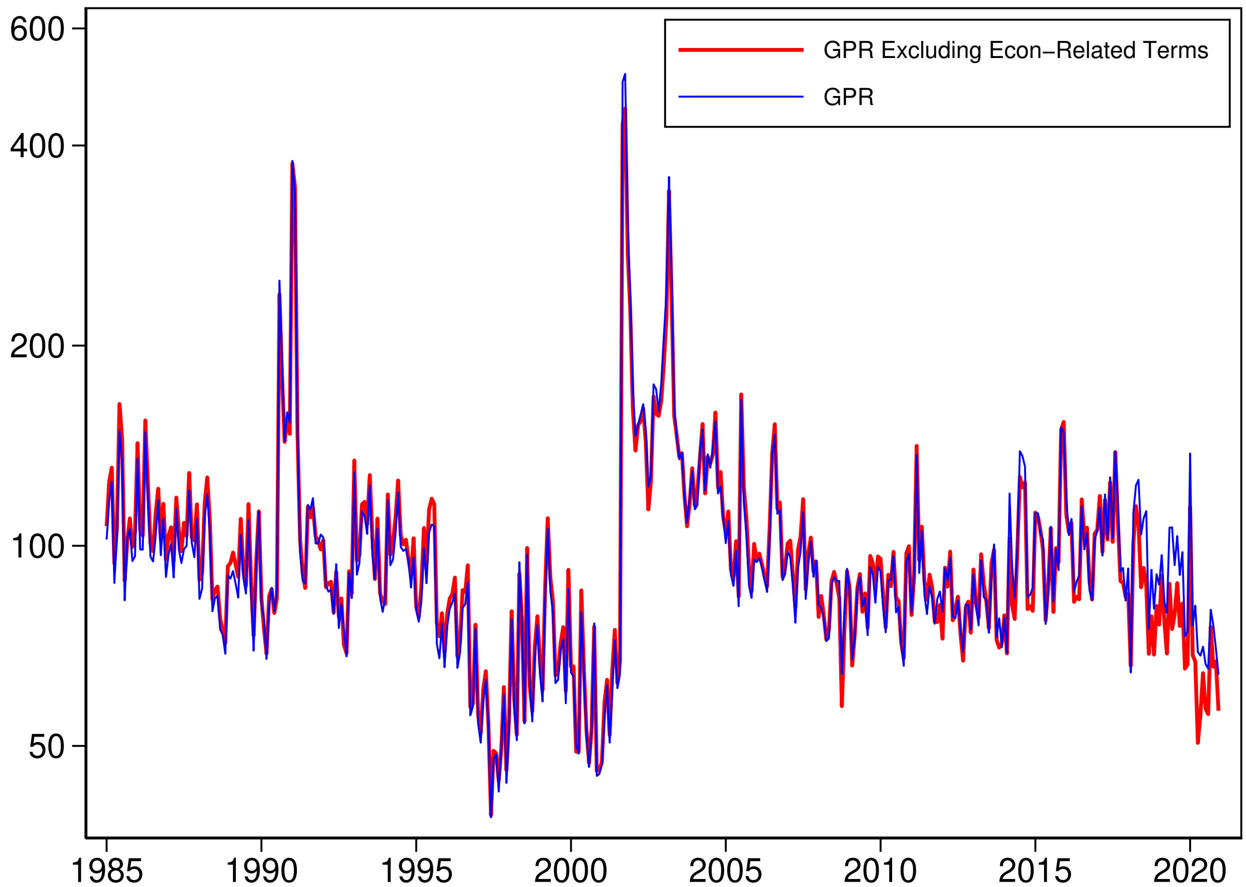
Note: The chart plots the relative contributions to the GPR index of the articles associated with the search categories described in Table 1. Each monthly series is plotted as a 12-month moving average. Each series plots the share of articles belonging to each of the categories listed in Table 1.

Figure A.12: Human Index



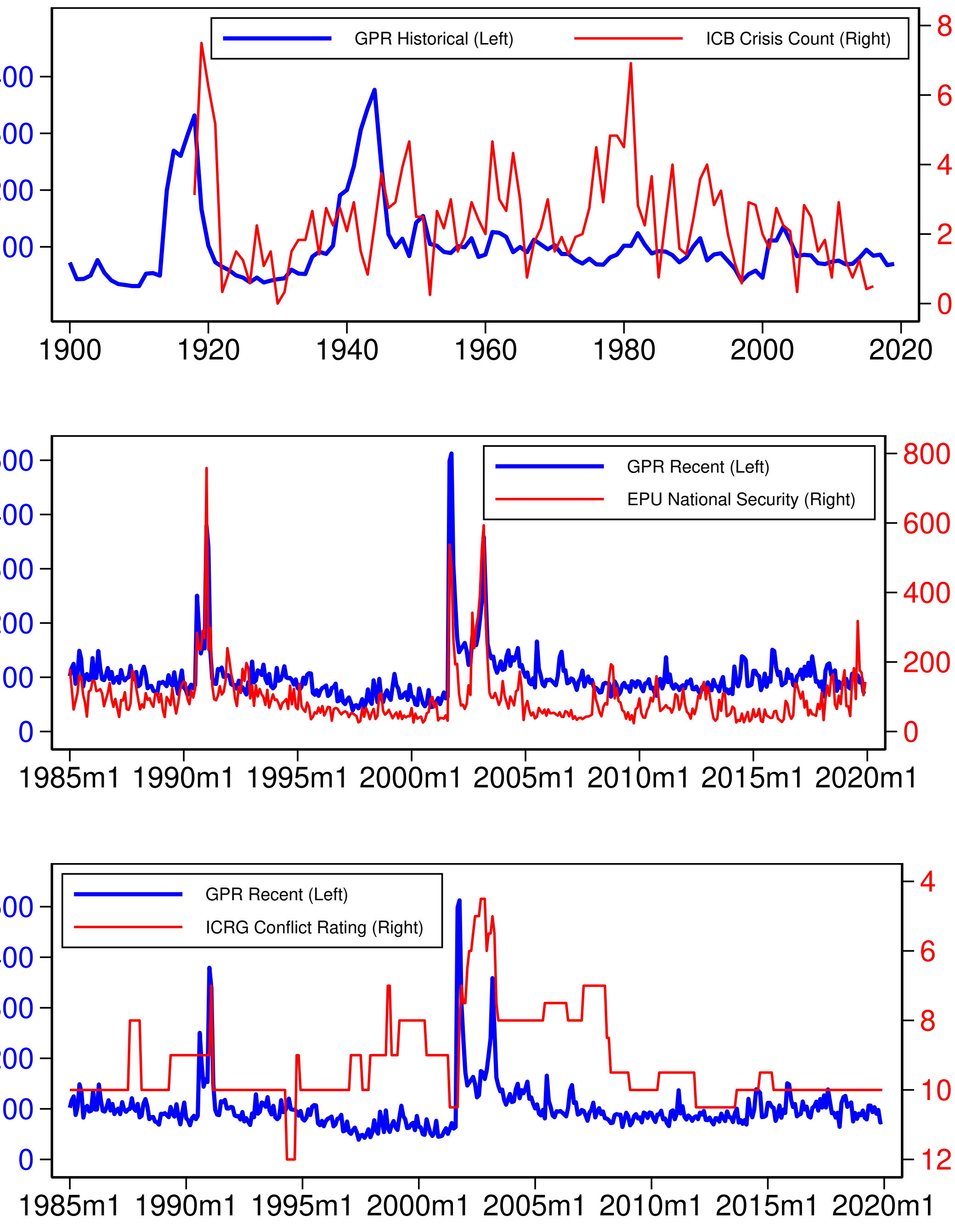
Note: Comparison of the historical GPR index (blue thick line) with the human index constructed by reading 7,416 articles (green thin line). Both series are plotted at yearly frequency and indexed to equal 100, on average, throughout the sample.

Figure A.13: The GPR Index Excluding Economics-Related Terms



Note: The figure compares the recent GPR index with a version of the index constructed excluding the search terms 'economy" OR 'stock market*" OR 'financial market*" OR 'stock price*." The correlation between the resulting index and the GPR index is 0.989. Both indexes are normalized to equal 100 in the 1985-2019 period and are plotted on a log scale.

Figure A.14: The Geopolitical Risk Index and Other Proxies for Geopolitical Risk



Note: In the top panel, the historical GPR and the ICB Crisis Count are annualized.