Testing Power Transition Theory Using Time Series Trends in Chinese Diplomatic Event Data

Prepared for the 2022 Western Political Science Association Annual Conference
Working Paper; Please do not cite

Max Plithides* and Bruce Ming Fuong Tsai†

Word Count: 9,149

Abstract

China's rapid rise, concurrent with the erosion of the U.S. unipolar moment, has motivated a renaissance in the study of Power Transition Theory (PTT). Yet modern PTT scholars still rely largely on traditional formal and qualitative methods, limiting their ability to perform rigorous statistical tests of their hypotheses. By harnessing 52,189 separate observations of Chinese actions directed towards the U.S. since 1994, we break from this traditional orthodoxy and construct a novel time series dataset of mean weekly Chinese-initiated interactions with the U.S. Analysis of this dataset identifies trends in Chinese behavior consistent with those of a dissatisfied challenger. In fact, the data demonstrates that Chinese-initiated interactions with the U.S. took on a significant negative trajectory when China reached power "parity," confirming PTT scholars' expectations. This finding and the time series methodology harnessed to arrive at it have significant empirical and methodological implications for the study of U.S.-China relations narrowly and PTT broadly.

*PhD Candidate, Department of Political Science, UCLA. Email: mplithid@ucla.edu.
†PhD Candidate, Department of Political Science, UCLA. Email: brucetsaimf@ucla.edu.

"The great turning points in world history have been provided by these hegemonic struggles among political rivals; these periodic conflicts have reordered the international system and propelled history in new and uncharted directions. They resolve the question of which state will govern the system, as well as what ideas and values will predominate, thereby determining the ethos of successive ages. . . . International society cannot and does not stand still. War and violence remain serious possibilities as the world moves from the decay of one international system toward the creation of another."

– *Robert Gilpin, War and Change in World Politics (1981)*

"We cannot predict with certainty what the future will bring, but we can be certain about the issues that will define our times. And we know this: The relationship between the United States and China will shape the 21st century."

– *President Barack Obama (2009)*

1 Introduction

Eminent scholars (Mearsheimer 2010, 2014; Allison 2015, 2017) expect China's rise to precipitate a new era of instability and great power conflict. Many compare China to the Kaiser's Germany and the United States (U.S.) to Britain, standing together on the precipice of disaster (Allison 2017, chap. 4; Brunnermeier, Doshi, and James 2018; Layne 2020). Yet there is precedent for peaceful power transitions. Through accommodation and sometimes painful adjustment, the U.S. and Britain, beginning in 1895, expertly avoided conflict (Kugler, Tammen, and Swaminathan 2001, 13; Wenckus 2015; Rauch 2017a, 210–11). The puzzle then is: Will China and the United States avoid war as Britain and America did? Or will they go down the same path to ruin as Britain and Germany in 1914?

U.S.-China War may seem unthinkable – so devastating would be its consequences. But 1914 offers a sobering reminder of mankind's capacity for folly. Few in 1914 could imagine the devastation that would wrack Europe over the next generation. In 1914, most European powers were global powers. States as small as Belgium, the Netherlands, and Portugal ruled huge swathes of territory in Africa and Asia. After a generation of systemic conflict over Germany's place in the world, the European empires were gone or teetering. Even the mighty British Empire, shorn of her youth and treasure, dissolved a mere two decades after the end of the World Wars. Five hundred years in which Europe was the center of the world came to a crashing halt. Whether China's rise will trigger a similar global catastrophe is thus the defining issue for a generation.

Of particular interest to scholars is whether China and America can avoid what Graham Allison (2017) terms the "Thucydides Trap." The Thucydides Trap refers to "the natural, inevitable discombobulation that occurs when a rising power threatens to displace a ruling power" (Allison 2017, xv), and derives in part from Power Transition Theory (PTT). PTT is a structural theory whose foundational works (Organski 1968; Organski and Kugler 1981; Tammen et al. 2000) address the central issue of international relations (IR): great power relations (Lemke and Tammen 2003). PTT predicts that when a dissatisfied power reaches 80% of the power level of the dominant power (parity), tensions will begin to emerge (Organski and Kugler 1981, 44; Tammen et al. 2000, 7). A large and vocal group of scholars (e.g., Xuetong 2011; Mead 2014; Lim 2015; Zhao 2018; Layne 2020) has long suggested that China is a dissatisfied power. But only in 2017, did the power ratio reach 80% (World Bank 2020b, 2020a).[1] PTT, thus, predicts that China should have begun to challenge the U.S. in 2017, and that tensions should have mounted afterwards.

This article focuses on the period from January 1, 1994 to January 22, 2022, looking for contemporaneous evidence of PTT. Harnessing 52,189 separate observations of Chinese actions directed towards the United States and its allies, we construct several time series datasets to test whether the large-N empirical data supports the notion that the U.S. and China are headed towards a cataclysm. Consistent with the predictions of PTT scholars, we identify trends in Chinese signal messaging since August 2017 consistent with those of a dissatisfied challenger. Chinese activities towards the U.S. become more hostile towards the U.S. from August 2017 to January 2022, just as Germany's behavior towards Britain did in the decade before 1914 (Tammen et al. 2000, 51–53). Furthermore, we demonstrate that – consistent with the expectations of PTT – Chinese actions towards the United States became significantly more hostile in 2017, representing a significant statistical break point in the data. These results demonstrate both novel empirical support for the predictions of PTT and the practicality of using time series methods to analyze signals data in interstate relations.

The following article is organized as follows: We briefly explore PTT and derive several hypotheses from it about the expected shape of Chinese-initiated interactions with the U.S. since 2017. Then

[1]There is some disagreement over whether nominal gross domestic product (GDP) or purchasing power parity (PPP) GDP should be used to measure relative power in the case of China, which manufactures most weapons systems domestically but still imports high-end technical components (Larson 2020; Peck 2020). This article splits the difference by using the average of these two measures of Chinese economic activity.

we construct a new weekly time series dataset from the 26,564 Chinese-initiated interactions with the U.S. logged by Lockheed Martin's Integrated Crisis Early Warning System (ICEWS). We next explore this dataset, addressing challenges and demonstrating that time series techniques can can be used to model Chinese-initiated interactions with the U.S. as a function of time. Finally, we apply breakpoint analysis, segmented OLS trend analysis, and autoregressive stochastic time series modeling to demonstrate that large-N empirical signals data supports PTT theory in the context of U.S.-China relations. The manuscript concludes by discussing its limitations and future directions for research.

2 Theory and Hypotheses

2.1 Power Transition Theory

While PTT has been used to explain a growing number of phenomena in International Relations (IR) (Organski and Kugler 1981; Kugler and Lemke 2000), the central issue lying at the heart of the theory remains great power relations (Lemke and Tammen 2003). Power transition theorists see the world as inherently hierarchical. And at the top of the hierarchy, sits a dominant power "by virtue of the common recognition of its preeminent position in world politics" (Tammen 2008, 319).[2]

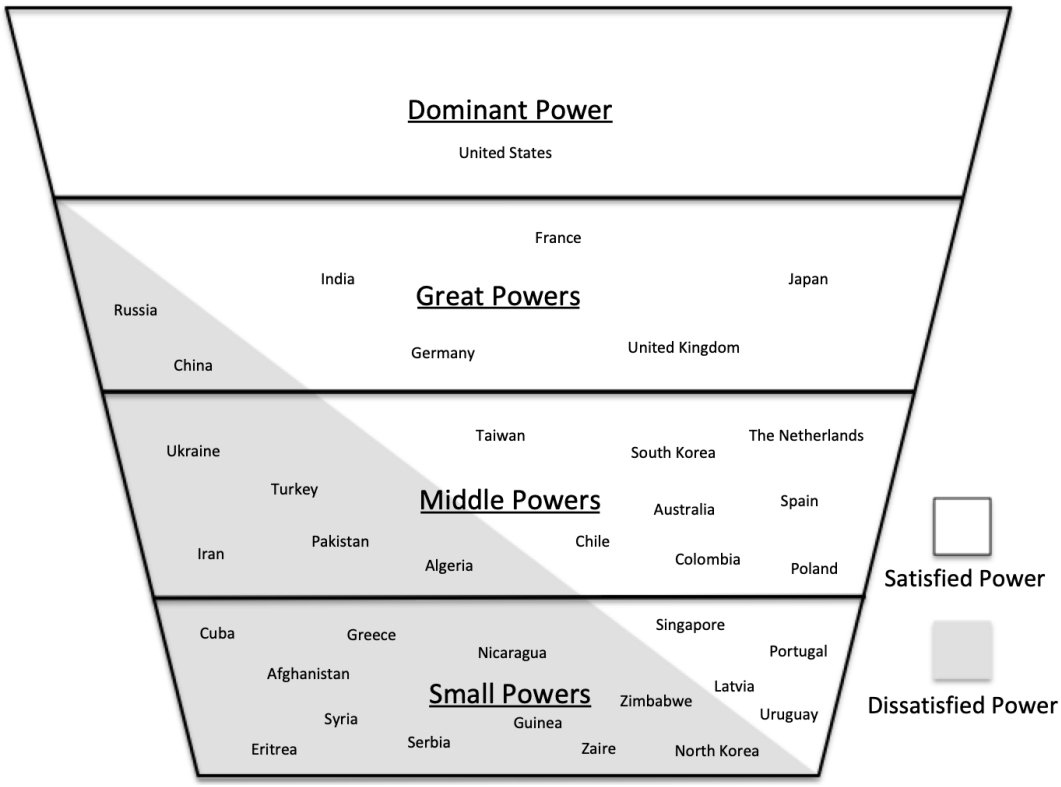
The dominant power has incentive to spread satisfaction as broadly as possible among those "great powers"[3] who sit beneath it and might challenge it (Tammen 2008, 321). The proportion of great powers that are dissatisfied with the system at any given time thus tends to be small. (See Figure 1.) The problem, however, is that state systems are not static hierarchies (Organski 1968, 371). Power dynamics are likely to eventually favor the emergence of some new great powers (e.g., India), the decline of others (e.g., the Soviet Union), and the dramatic strengthening of some (e.g., China). Troubles emerge when one dissatisfied great power manages to attain 80% of the dominant nation's power – i.e. when it reaches "parity" (Organski and Kugler 1981, 44; Tammen et al. 2000, 7).[4]

[2]Its position atop the order of course necessarily implies its satisfaction with that order (Organski 1968, 326).

[3]"Great powers" in this context refers to those states that sit just one rung below the dominant state in the status hierarchy.

[4]Power is traditionally measured in terms of gross national product (GNP) (Organski and Kugler 1981, 44). This is roughly equivalent to GDP, and the terms are often used interchangeably.

Figure 1: PTT's Contemporary Inverted Pyramid of Power, Hierarchy, and Satisfaction

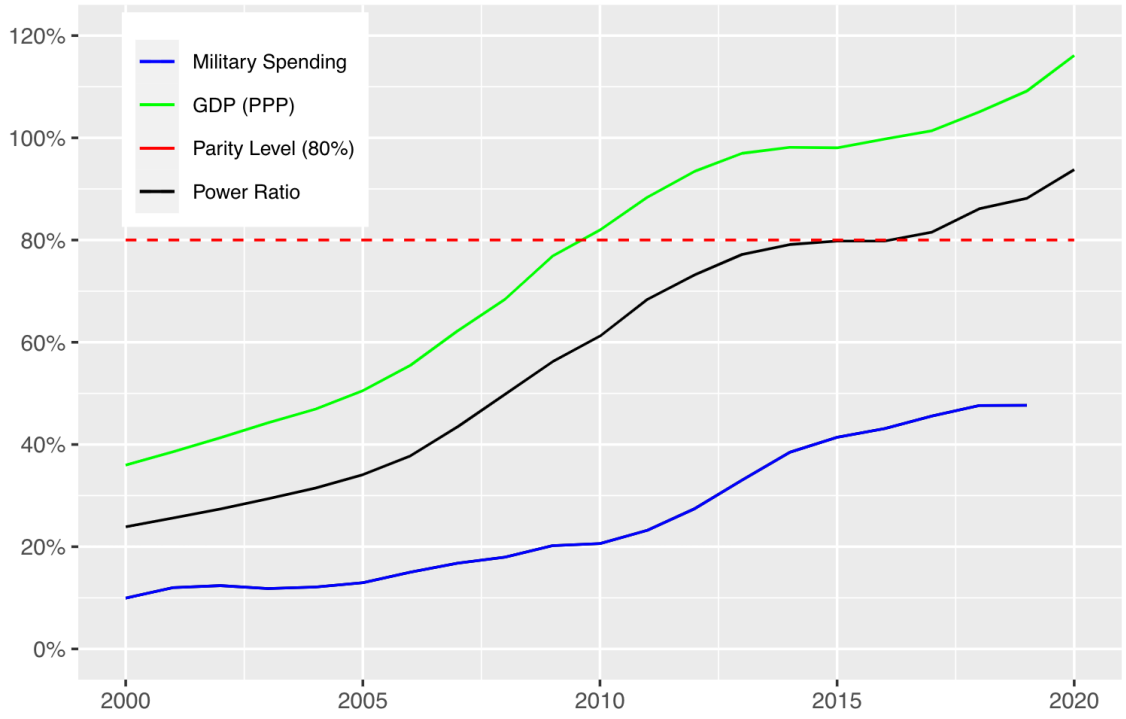


2.2 China's Rising Challenge

China's growth over the past twenty years reflects the dynamic nature of international power. Since 2000, the Chinese economy has grown by a factor of twelve (World Bank 2020b). Even more impressively, during this time it has gone from being a fifth the size of the U.S. economy to being slightly larger in purchasing power parity (PPP) terms (World Bank 2020b). Yet China still relies on Russian imports for a limited number of high-end weapons systems (Larson 2020; Peck 2020). Thus, this paper, in the interest of taking a conservative view of Chinese power, splits the difference between the nominal and PPP measures of Chinese power. Using this combined average, it finds that Chinese power reached parity with the U.S. in 2017 (World Bank 2020b, 2020a). Figure 2 illustrates the Chinese long march to parity.

China's dissatisfaction with the U.S.-led international order meanwhile precedes its attainment of parity (e.g., Xuetong 2011; Mead 2014; Lim 2015; Zhao 2018; Layne 2020). The first indicator of dissatisfaction with a system according to PTT is a rapidly growing military budget (Lemke and Werner 1996; Lim 2015, 285–88). Growing military budgets axiomatically reflect a belief on the part

Figure 2: Chinese Power Relative to the U.S. Level



of the challenger that only force may compel the dominant state to share the benefits of leadership. China's military budget has exploded from 10% of the U.S. budget in 2000 to nearly 50% today (SIPRI 2020; World Bank 2020c).[5] Figure 2 demonstrates Chinese military spending relative to that of the U.S.

Furthermore, Lim (2015) points to two other sources of Chinese dissatisfaction. For one, China remains an authoritarian and state capitalist regime in an international system that privileges American notions of governance and economic liberalization (Lim 2015, 288–92). Second, China consistently seeks to preclude America from the regional security architecture (Lim 2015, 292–96). One can infer from this that China wishes the United States to have a less significant role in Asian regional affairs. In cases such as this, PTT expects some significant deterioration in relations when dealing with a dissatisfied challenger, even if war could possibly be avoided (Organski 1968, 375).

[5]This is even more impressive when one considers the Chinese military's much more limited mission set. Military spending is adjusted for PPP prices in line with the recommendations of Robertson (2021).

2.3 Declining Relations

The decline in Anglo-German relations before 1914 illustrates this expectation (Allison 2017, chap. 4; Brunnermeier, Doshi, and James 2018). In 1884, Britain produced more than twice as much steel as Germany (Stafford 1982, 250–51), and the British Conservative Party's weekly journal, *England*, praised Bismarck as "the greatest man in the world" (Stafford 1982, 250). But by the end of the 19th century, as the German economy surged to parity, sentiments gradually shifted. The period after 1900 saw a steady deterioration in German-initiated diplomatic overtures towards Britain. Diplomatic affronts such as the Kruger Telegram, the Muller memorandum, and Tirpitz's risk theory were all part and parcel of this steady decline in relations. And Germany's dissatisfaction was given weight by Germany's major fleet construction program (Stafford 1982, 254). In other words, Sarajevo may have been the spark for war, but German diplomacy steadily loaded the powder keg for decades.

China's rise closely – and purposefully (Brunnermeier, Doshi, and James 2018, 3) – mirrors that of imperial Germany. Unification struggles left both modern China and imperial Germany as latecomers to modernization, and each believed that catching up would require centralized state-directed economic programs (Brunnermeier, Doshi, and James 2018, 3–9). Both felt constrained by Anglo-American systems that privileged laissez-faire economics and constitutional government. And both placed a premium on international status and regional hegemony (Layne 2020). Little imagination then is required to expect that Chinese diplomacy will become steadily more hostile towards the U.S., just as German diplomacy towards Britain did. This leads to Hypothesis 1.

Hypothesis 1 (Declining Relations) The longer the U.S. and China spend at parity, the more hostile Chinese-initiated interactions will become.

We, thus, have built a case for the direction that Chinese-initiated interactions will trend after parity is reached. Yet we are equally interested in understanding the shape that deteriorating relations should take. Again, we turn to the Anglo-German analogy for guidance.

2.4 The Parity Breakpoint

To provide data on the comparative economic strengths of England and Germany c. 1900, we turn to steel production figures. Although it is impossible to get exact nineteenth century data, the goal is to locate generally the moment in which Germany reached parity with Britain in steel production. Based on reliable export figures (one million tons for Germany; four million tons for Britain), Britain in 1895 maintained a significant economic advantage over Germany (Allen 1979) in the industry (steel) most crucial for contemporary war making. But by 1907, German steel production had risen to 97.5% of the British level (Allen 1979, 919), all but eviscerating the British economic advantage. We deduce from these figures and the growth rate estimates in Allen (1979) that parity was reached c. 1900. Significantly, this date overlaps with a sharp break in German-initiated diplomatic overtures towards Britain.

In 1895, Germans paid Britain no special amount of attention. Germany's primary military rivals were France and Russia. Most Germans regarded Britain simply as a "weary titan" (Kennedy 1980, 313), generally detached from continental affairs. Yet as Germany's transformation into Britain's economic peer took place, German opinions towards Britain shifted at lightning speed. In 1899, German admiral Tirpitz first began to speculate privately that "[England] wishes to ruin us before our fleet emerges from the egg shell" (Kennedy 1980, 314). And Tirpitz's sentiment – buoyed by economic competition and framed in economic terms – was spreading in parallel among German nationalists (Kennedy 1980, chaps. 16–20).

The outcome was a sharp break in German policy towards Britain. In 1900, Germany passed a major naval expansion bill, directly challenging Britain's core security interest just as it was reaching parity in steel production. Although there had been previous points of tension in South Africa and Samoa, the naval bill represented Germany's first challenge to the security of the British isles themselves. By 1902, the British admiralty was forced to consider countermeasures (Kennedy 1980, 418). Though relations steadily declined after 1900, the fleet buildup remained – according to the German ambassador in London – the "cardinal point" of Anglo-German tensions (Kydd 1997, 371; Woodward 1935, 168–69). This leads to a second hypothesis, which emerges from the Wilhelmian historical analogy.

Hypothesis 2 (Parity Breakpoint) Chinese-initiated interactions will demonstrate a significant negative breakpoint shortly after parity is reached.

The Anglo-German analogy is, however, subject to some caveats and scope conditions. Assumptions about observed behavior assume that we can accurately gauge when parity is reached.[6] To do so, we rely on publicly available GDP data. But GDP data is commonly manipulated for strategic reasons (Michalski and Stoltz 2013), and Chinese data is notoriously unreliable (Wallace 2016; Owyang and Shell 2017; Clark, Pinkovskiy, and Sala-i-Martin 2017).[7] Furthermore, even small changes in a study's power indicator can have a significant effect on the results (Rauch 2017b). We, thus, defend in great detail our choice of model specification in Appendix A.

Omitted from our Wilhelmian analogy is mention of Britain's actions towards Germany. The omission is intentional. The goal of this paper is to empirically test whether the structural underpinnings of PTT are sound. However, we consider in Appendix B the possibility that American misperceptions of Chinese actions post-parity could be driving the negative relational dynamics. We find no evidence for such an alternative hypothesis. We, thus, take as a given that the U.S. would only initiate a preventative war if China first showed concerted signs of deep dissatisfaction with the current world order (Yoder 2019).

2.5 Alternative Hypothesis: Satisfaction and Accomidation

The satisfaction of the challenger is an important concept in PTT. In fact, PTT posits that a deterioration in relations is avoidable, contingent upon the satisfaction of the challenger. The leading power may – if sufficiently flexible – accommodate a satisfied upstart (Chan 2004). This outcome is particularly likely if the challenger's size practically guarantees its eventual dominance (Organski 1968, 372). Thus, under the right conditions, the leading power may "socialize" the challenger into the existing system of rules and norms through limited accommodation (Tammen 2008, 321; Deudney and Ikenberry 2009). Prominent scholars (e.g., Tammen and Kugler 2006) have speculated that this may be one possible direction for Sino-American relations.

[6] Not all authors take this assumption for granted. However, a formal model that assume informational asymmetry about whether parity has been reached actually finds that belligerence is *more* likely than it would be in a perfect information environment. See Reed (2003).

[7] In 2007, Li Keqiang, China's Premier, allegedly told an American diplomatic official that GDP figures are "man-made" and therefore unreliable (Randt 2007).

British appeasement of the U.S. serves as the classic case in point of successful accommodation (Rock 2014, chap. 2). For most of the 19th century, British policy makers considered the U.S. an afterthought.[8] Yet by the late 1890s, American GDP was approaching parity with that of the British empire (Bolt and Zanden 2020).[9] And America was beginning to chafe at British influence in the Western Hemisphere. But from the start, American revisionist aims were limited, and politicians in London were quick to notice America's finite appetite for change. British ministers were only too happy to settle disputes in Venezuela, the Yukon, and Central America to ensure broader American satisfaction. As Bradford Perkins writes: "British ministries from Salisbury to Asquith made important concessions of substance and form to the United States. These statesmen gained . . . above all a transformation of American attitudes" (Perkins 1968; Rock 2014, 25). And as Anglophobia declined in the U.S., relations warmed, differences were settled, and ultimately, a "special relationship" emerged (Orde 1996).

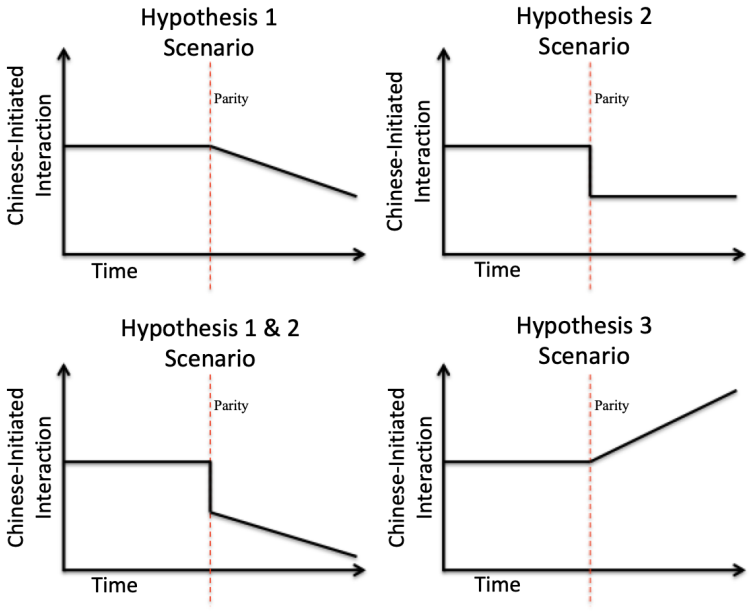
Could China's rise stimulate a similar improvement in relations? The majority of scholars are skeptical, but a sizable minority (e.g., F. Roberts 2014; Epstein 2019; Shifrinson 2020) believes that it potentially could. Contemporary Sino-American relations do share some similarities with 19th century Anglo-British ties. For example, China – similar to the United States in the 19th century – has the potential to substantially surpass all other contemporary great powers (with the possible exception of India) in sheer economic heft. China has an imperfect human rights record and a long record of intellectual property theft, but so did the U.S. in 1850 (Morris 2012; T. Roberts 2014). As the U.S. and Britain demonstrated in the second half of the nineteenth century, significant hurdles to improving relations can be overcome under the right circumstances. Thus, it seems prudent to consider a possible alternative hypothesis built on this analogy.

Hypothesis 3 (Ascendant Relations) The longer the U.S. and China spend at parity, the more amiable Chinese-initiated diplomatic interactions with the U.S. will become.

[8] Not that they would have had a great deal of information with which to craft detailed policy had they wanted; in 1895, only one British newspaper, *The Times*, had a permanent correspondent stationed in the U.S. (Orde 1996, 13).

[9] For comparisons between Britain and the U.S., we use the GDP of the entire British empire because the Anglo-American relationship is extra-regional for both powers. For comparisons between Britain and Germany, we use both countries' Europe-only GDP's. This is because competition between Britain and Germany occurred within the Western European subsystem, even if the second-order repercussions were global. For more on regional subsystems within PTT see Lemke (2002).

Figure 3: Hypotheses Visualized



The Alternative Hypothesis as well as Hypotheses 1 and 2 are visualized in Figure 3. The actual data contains significant noise, but we expect Chinese-initiated interactions over time to generally cohere to one of the four trends shown. The remainder of the manuscript is dedicated to the task of demonstrating empirical support for one of the theoretically established scenarios.

3 Data

3.1 Data Creation

To test our hypotheses, we construct a series of time series datasets of bilateral interactions between states. To capture every recorded interaction, we turn to Lockheed Martin's Integrated Crisis Early Warning System (ICEWS). ICEWS uses automated text analysis of more than 100 data sources and 250 news sources to capture transnational interactions, and contains records of more than ten million interactions. From ICEWS, we extract all Chinese actions directed at the United States beginning on January 1, 1994 and ending on January 22, 2022. This subset of ICEWS data constitutes 26,564 separate Chinese-initiated interactions and constitutes the primary time series dataset of interest.

Each of these data points comes with an "intensity score" scaled ordinally from -10 to 10.

Combatant commanders traditionally use this intensity score to forecast the outbreak of conflicts. As such -10 represents the threat or use of conventional military force and 10 represents the signing of a peace agreement. Since China and the U.S. have not been at war since 1994, the highest-scored interaction they have ever engaged in is an 8, "signing a formal agreement" (e.g., a trade agreement). Chinese employment of convention military force for diplomatic leverage during the Taiwan Strait crisis represents the only -10 in the dataset with regards to the United States. In general, higher-numbered interactions represent more positive interactions, and lower-ranked interactions represent less positive interactions. Interactions are rated ordinally; for example, to "confiscate property" is scored -9.2, to "reject" is scored -4, and to "accuse" is scored -2.

To create the primary dataset of interest, we calculate the mean intensity score of each week's Chinese-initiated interactions from January 1, 1994 to January 22, 2022. This yielded a time series dataset, of Chinese-initiated actions with "weekly mean intensity scores." Our primary purpose in creating this dataset is to observe if automated large-N analysis can detect empirical evidence of the presence or absence of PTT in Chinese actions.

We also create a series of datasets to use for comparison's sake. These datasets ensure that the Chinese actions towards the U.S. do not reflect a generalized animosity towards foreign states but specifically reflect of Chinese attitudes targeted at the United States. The subjects of Chinese-initiated interactions chosen for these controls are Japan, Australia, and Taiwan. We make the decision to use Chinese-initiated interactions with these three countries as controls for several reasons. First, using Chinese-initiated exchanges with three American allies in Asia as controls makes this a hard test case for Chinese actions directed towards the United States. One would intuitively expect that, as go Chinese relations with American allies, so would go Chinese relations with the U.S. and vice versa. Second, concerns have also been raised with our approach's inability to detect "triangulated" communications (i.e. – messages sent from Beijing to Washington through third party states). In particular, scholars commonly perceive China as deferring to the U.S. while showing no such deference to America's allies. As Peter Martin (2021, 208) explains:

> "The United States was powerful enough that Chinese officials were forced to treat it with caution, even if they deplored its behavior. Chinese diplomats and state media generally refrained from criticizing Trump personally for fear of provoking the president,

and even continued to tout Trump and Xi's personal 'friendship.' They did not afford the same deference to America's allies."

In 2017 and 2018 alone, China intervened in Australian domestic politics, declared something akin to economic warfare against South Korea, and detained several Canadian citizens. We, therefore, wish to ensure that our dataset of Chinese actions aimed towards Washington does not present the relationship through rose-tinted glasses. Third, choosing Japan and Taiwan offer secondary insights into China's recent regional ascendancy both in the context of a period of local parity (2004 – 2008)[10] (World Bank 2020b, 2020a), and in the context of an enduring regional rivalry. These secondary insights offer directions for future research that are discussed in the conclusion.

3.2 Challenges and Limitations of the Dataset

All event data based on public sources are susceptible to selection issues and reporting bias (McClelland 1983; Schrodt and Gerner 1994; Weidmann 2016; Colaresi et al. 2018). Diplomatic event data collected via ICEWS face an additional complication because a nontrivial amount of everyday and crisis diplomacy is conducted in a manner that most likely eludes ICEWS automated collection algorithm. Classified and confidential communications, for instance, conveyed to American officials from Beijing are most likely absent from the dataset.

We know though that China, more than most countries, conducts both state-to-state and public diplomacy online and in publicly available media. Beijing's use of "Twitter diplomacy" has become routine since 2019 (Alden and Chan 2021), and the state commonly harnesses Chinese media to conduct public diplomacy (Thussu, De Burgh, and Shi 2018). Chinese media also commonly applies pressure on Chinese policy-makers creating a feedback loop. The Chinese media helps to shape the agenda for foreign policy makers, to narrow down the set of policy options, to change the pace of policy making and implementation, and to influence the direction of final decision making. The government, meanwhile, controls and regulates the media (Wang and Wang 2014). The tight interaction between Chinese media and government provides us with confidence that public event data broadly reflects private Chinese diplomatic activity.

Weekly-level time series data also does not distinguish between periods of high-frequency

[10] This period saw China rapidly surpass Japan according to our preferred power metric. See Appendix I.

communication and low-frequency contact. Suppose the U.S. and China plan a week-long series of negotiations. Our data will average all news reports that emerge from the series of negotiations by week. Even if the outcome of the negotiations is a large-scale trade agreement covering multiple commercial sectors, the mean intensity score to emerge from the week of negotiations may be significantly lower than it might have been if the two countries had simply signed a single, narrow commercial agreement sans negotiation. We recognize this shortcoming and partially palliate it in two ways. First, we calculate weekly sums of event data in lieu of average event intensity scores in the Appendix D as a robustness check on our analysis here. Second, we calculate daily mean intensity scores then exclude all days without an event for the purposes of model generation. This second check provides us further confidence that our findings are robust against the issue raised here.

Finally, automated text analysis creates a number of issues, which we address in Appendix E. But in short, to soothe concerns over coverage based on language, algorithmic parsing techniques, and event duplication, we have hand parsed all 26,564 observations. We have identified roughly 600 duplicate events[11], found that Chinese language sources are extremely common, and confirmed the algorithmic parsing to be accurate above the 95% threshold among a random sample of 300 events. A cursory exploration of the weekly data provides further reassurance.

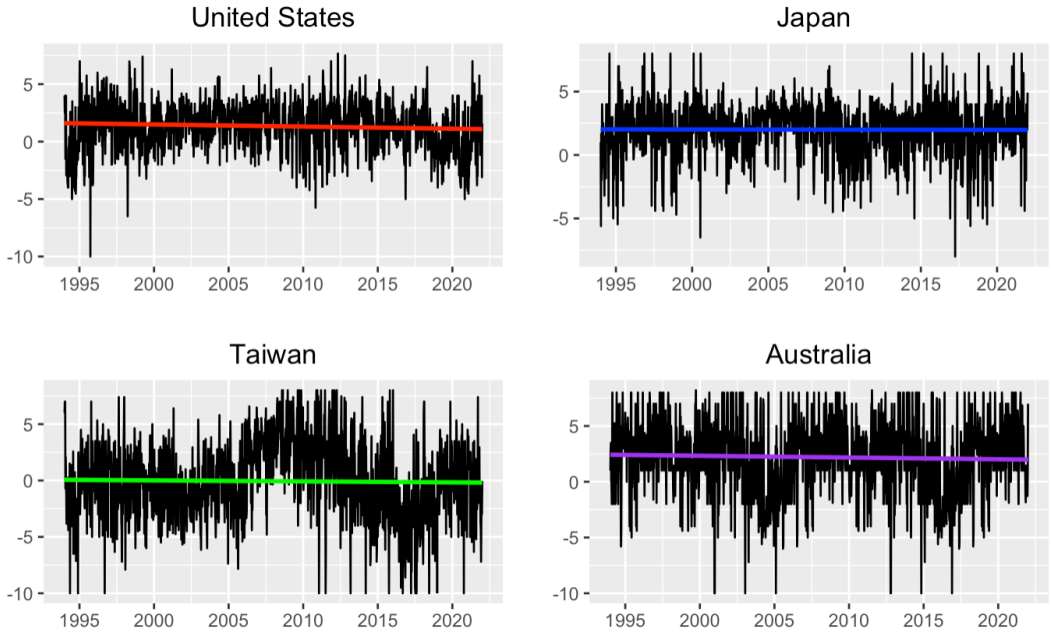
3.3 Data Exploration

As can be seen in figure 4, the time series datasets are all very choppy. The mean of the primary dataset of interest is 1.42; however, there are no obvious trends in the data because of white noise. The same is true of all of the datasets. To see if there is a linear trend in the data, we regress mean weekly intensity scores on time. The ordinary least squared lines of best fit are shown on each of the four time series datasets. While each is significant, each has a coefficient close to 0, indicating that diplomatic relations generally remain stable over time. This provides some basic intuitive sense that our time series data accurately reflect overall relations, since foreign policies have a general stickiness to them both due to systemic realities and bureaucratic inertias.

We also examine some outlier points in our time series data in order to confirm its general

[11] We leave these since they represent less than 3% of all events. Their inclusion has no effect on our overall finding, and including them avoids controversy over our preferred definition of what constitutes a "duplicate."

Figure 4: Mean Weekly Chinese Action Towards States over Time



validity. To begin with, we seek to understand the prevalence of -10 mean intensity scores in our data given the absence of kinetic conflict in the Pacific. We find that these represent specific instances in which China has either used or threatened to use conventional military force against its regional neighbors. In particular, we find a high volume of negative ten events clustered around statements regarding Taiwanese elections and territorial disputes. In the main dataset of interest, we observe one -10 as a result of Chinese missile tests during the Taiwan Strait crisis, but that this event should be assigned a score of -10 passes the metaphorical "gut check." Therefore, we are confident that the extreme values in the data accurately reflect Chinese rhetoric.

4 Methods

To analyze the time series datasets in question we consider a broad range of statistical methodologies. Ultimately, we reject traditional ordinary least squared (OLS) methods as ill-fitting for the task at hand. In what follows we discuss this decision. We then explain how we intend to conduct breakpoint analysis on the data before segmenting the time series and secular trend data. Finally, we apply an autoregressive time series model to test for the existence of "momentum" in Chinese-initiated interactions.

To our knowledge this is the first time in the U.S.-China literature that anybody has seen fit to combine automated text analysis data with time series decomposition in a large-N empirical analysis of diplomatic relations. It is also the first time that simultaneous estimation of multiple breakpoints will be conducted using a dynamic programming approach based on the Bellman Principle to model relations between states.[12] And the first time that an autoregressive model has been used to measure momentum in a state's foreign policy behavior.

4.1 Appropriate Model Selection

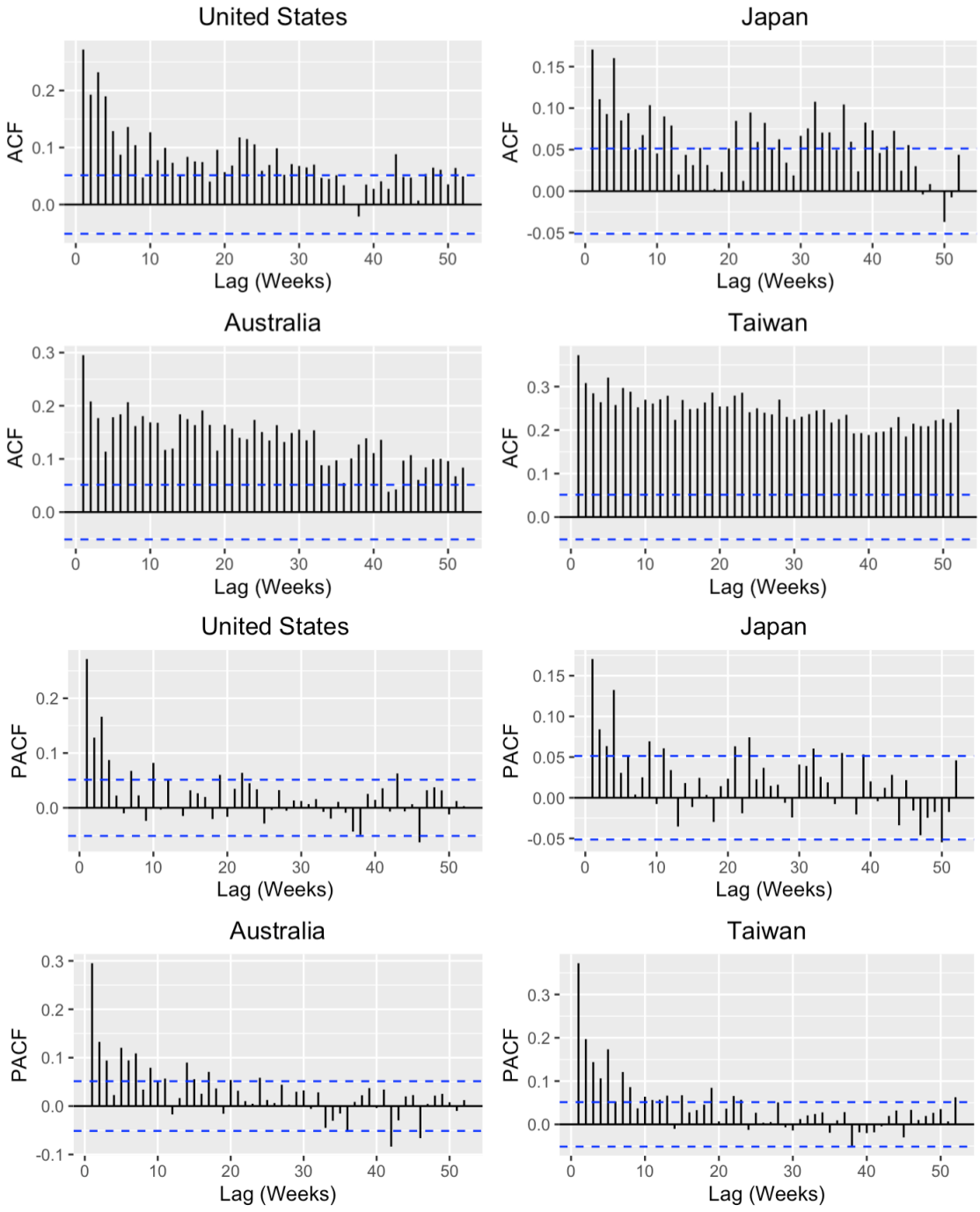
Before advancing further, we discuss why we decide not to precede with a traditional OLS analysis of the data. The decision evolves out of several explorations of the initial time series data. First, as Figure 5 clearly demonstrates, Chinese attitudes towards each of the states in question are extremely "sticky" with high levels of autocorrelation between weekly mean intensity scores. We thus dismiss the assumption that observations are independent and identically distributed. The partial autocorrelation function tells a similar story. (See Figure 5.) As is common with time series data, these observations influence each other, and this can oft-times require scholars to turn to a first-differenced model conditional on the data proving nonstationary.[13]

Further assumptions necessary for proper application of OLS are also definitively violated by our time series dataset. Axiomatically some periods are likely to see more volatility in relations than others. Consider the difference between Presidents Obama, Trump, and their Twitter habits. It would be unreasonable to make assumptions about the homoskedasticity of error terms were a single linear model fitted to the entirety of the data. In addition, we doubt linear assumptions due to the complex nature of diplomatic relations. So many possible predictors could enter into Chinese calculations that a researcher selecting covariates for such a model would be akin to a monkey throwing darts. And these covariates themselves would be shifting as functions of time, creating a messy model. The ability to control for error terms in previous periods in this context can actually be a benefit as it allows us to soak up some of that unknown stochastic variation in a systematic way. We thus apply methods of time series analysis which incorporate moving-average components for part of our analysis.

[12] Admittedly, it does not represent the first time the term "breakpoint" has been used in a paper about China. See, for example, Chubb (2020). However, it has never been used it in a mathematical or programming sense.

[13] We apply the Dickey-Fuller test to determine stationarity in the Results section.

Figure 5: Autocorrelation and Partial Autocorrelation Functions of Mean Intensities



Finally, a fundamental assumption of OLS is that the true coefficients of the specified model are linear. Given that we seek to answer the question if there is indeed a change in state relations and interactions regarding PTT, it would be odd to expect the coefficients to remain linear throughout the entire longitudinal span of the data. Forcing a standard OLS regression with linear coefficients throughout would impose upon an assumption violation that would certainly result in a consistent null in hypotheses testing. As such, we adopt breakpoint analysis and modify the standard OLS to fit piecewise iteratively through a dynamic programming procedure, identifying if and when coefficients for the mean weekly intensity of Chinese-initiated interactions shift. If there is no change, then the piecewise iterative fitting process should return a result lacking breakpoints (effectively a standard OLS model). If there is a shift, however, then the piecewise iterative fitting process would indicate the relevant periods of change.

4.2 Time Series Analysis

We apply three methods to test our hypotheses regarding PTT and Chinese-initiated interactions with the U.S.: break date analysis, traditional OLS and data segmentation combined with trend decomposition, and extended autoregressive moving average (ARMAX) modeling and its "integrated" counterpart. Yet before applying these three methods, we must first decompose and explicate the mathematics behind them. What follows is the mathematical "nuts and bolts" of the tools, which we will later apply.

4.2.1 Breakpoint Analysis

We first need to understand breakpoint analysis. We can safely reject the null – that there is no breakpoint at parity – if Chinese behavior towards the U.S. coheres to the expectations of Power Transition theorists by demonstrating a sharp breakpoint in 2017. Furthermore, we can further evince that this breakpoint can be attributed to PTT if we can demonstrate that no other Chinese relationship involves a similar break.

For break date analysis, we use the procedure pioneered by Bai (1994), extended by Bai (1997, 2010), and implemented by Bai and Perron (2003), Zeileis et al (2002), and Zeileis, Shah, and Patnaik (2010). This procedure compares dates of significant parametric change across time series datasets. For the sake of parsimony, we choose to test or assess deviations from stability in the

classic linear model

$$y_t = x_t'\beta_t + u_t \qquad (t = 1, ..., n), \tag{1}$$

where x_t is a $[k \times 1]$ vector.

Since our time series data extends back to 1995, we reasonably assume that in most cases there are m break points (where the best-fit linear coefficients shift from one stable relationship to another) in any given time series dataset of Chinese behavior towards another state such that $m \geq 1$. Thus, for Chinese actions towards another state there should be $m + 1$ segments in which the regression coefficients are constant, and the model can be rewritten as:

$$y_t = x_t'\beta_j + u_t \qquad (t = t_{j-1} + 1, ..., t_j, j = 1, ..., m + 1), \tag{2}$$

where j is the segment index, $T_{m,n} = t_1, ..., t_m$ denotes the set of the breakpoints, and by convention $t_0 = 0$ and $t_{m+1} = n$. In this model, the break points $(t_1, \ldots, t_m)$ are explicitly unknown. The goal is to estimate these breakpoints using the known date observations (x_t) and interaction intensity observations (y_t).

We estimate breakpoints by minimizing the residual sum of squares (RSS). Given an m-partition $t_1, \ldots, t_m$, The least-squares estimates for the β_j can easily be obtained. The resulting RSS is given by

$$RSS(t_1, ..., t_m) = \sum_{j=1}^{m+1} RSS(t_{j-1} + 1, t_j), \tag{3}$$

where $RSS(t_{j-1} + 1, t_j)$ is the usual minimal residual sum of squares in the jth segment. Finding break dates $(\widehat{t}_1, ..., \widehat{t}_m)$ is then an issue of minimizing the objective function

$$(\widehat{t}_1, ..., \widehat{t}_m) = \text{argmin}_{t_1,...,t_m} \; RSS(t_1, ..., t_m), \tag{4}$$

over all partitions $(t_1, ..., t_m)$ with $t_j - t_{j-1} \geq n_h \geq k$ where n_h is a trimming parameter set to six months. We next appropriate a dynamic programming algorithm from Bai and Pearson (2003) to obtain the global minimizers. The basic idea is rooted in the Bellman Principle: the optimal

segmentation satisfies the recursion

$$RSS(T_{m,n}) = \min_{mn_h \leq t \leq n-n_h} [RSS(T_{m-1,t}) + RSS(t+1,n)]. \tag{5}$$

Therefore, it suffices to know the "optimal previous partner" for each point t if t was the last breakpoint in an m-partition. This can be derived from a triangular matrix as demonstrated by (Bai and Perron 2003).

We also construct confidence intervals for the break dates. In general, the narrower the confidence interval around a calculated break date, then the higher the degree of confidence in that estimate. To get an asymptotic distribution for the break dates, the strategy considered is to adopt an asymptotic framework where the magnitudes of the shifts converge to zero as the sample size increases. The resulting limiting distribution is then independent of the specific distribution of u_t. The confidence intervals can be constructed from the methods outlined in Bai (1997) and Bai (2003), with the $100(1-\alpha)\%$ confidence interval being given as:

$$[\widehat{k} - [\frac{c}{\widehat{\widehat{L}}}] - 1, \widehat{k} + [\frac{c}{\widehat{\widehat{L}}}]], \tag{6}$$

where $\widehat{L} = \frac{(\widehat{\delta_T}' \widehat{Q} \widehat{\delta_T})^2}{\widehat{\delta_T}' \widehat{\Omega} \widehat{\delta_T}}$ and $[\frac{c}{\widehat{\widehat{L}}}]$ represents the integer portion of $\frac{c}{\widehat{\widehat{L}}}$. c, meanwhile, represents the $(1-\alpha)$-th quantile of the random variable $\operatorname{argmax} W(s) - |s|/2$, where $W(s) = W_1(-s)$ when ≤ 0 and $W(s) = W_2(-s)$ when > 0 – $W_i(s),\ i = 1, 2$ being two independent standard Wiener processes defined on $[0, \infty)$, starting at origin, $s = 0$. The derivations of the estimators $\widehat{k}, \widehat{\delta_T}, \widehat{\Omega}$ can be obtained from Bai (1997). And the quantile c is obtained from the cumulative distribution function derived in Yao (1988) and Bai (1997):

$$G(x) = 1 + (2\pi)^{-\frac{1}{2}} \sqrt{x} e^{-\frac{x}{8}} - (\frac{1}{2})(x+5) \times N(-(\frac{\sqrt{x}}{2})) + \frac{3}{2} e^x N(-\frac{3\sqrt{x}}{x}), \tag{7}$$

where $N(\cdot)$ is the cumulative distribution function of the standard normal random variable.

4.2.2 Extracted Trend Data, OLS, and Data Segmentation

We next use a basic additive decomposition to extract the trend in relations from the white noise, cyclical, and seasonal components. This will allow us to get a basic sense of the secular change

in the direction of Chinese-initiated interactions since the hypothetical 2017 breakpoint. Further, we also extract the trend from Chinese treatment of Taiwan, Japan, and Australia. We then apply basic OLS analysis to the extracted trends from before and after 2017 to demonstrate that Chinese action towards the U.S. *uniquely* reflect the PTT hypotheses.

To analyze the shape of Chinese interactions with the U.S. relative to those China initiates with other countries, we use additive time series decomposition and OLS time series segmentation. The decomposition of time series is a statistical task that deconstructs a time series into several components, each representing one of the underlying categories of patterns. We are primarily interested in the underlying trend (T_t) in Chinese-initiated relations with other countries. Using an additive model we can conceive of any given time series dataset as,

$$y_t = T_t + S_t + C_t + I_t \tag{8}$$

where T_t represents a trend component which reflects the long-term progression of the series (secular variation). We can then compare trends between our hypothesized graphs, our main relationship of interest (China-U.S.), and our control relationships (China-Japan, China-Australia, and China-Taiwan).

To derive our trend estimate, we must remove periodicity from our data. To do so, we use a nonparametric spectral estimation method based on embedding a time series in a vector space of dimension M to estimate the seasonal and cyclical periodicity. The $M \times M$ lag-covariance matrix $\mathbf{C_X}$ of $X(t)$ allows us to obtain this information. The matrix $\mathbf{C_X}$ can be estimated directly with constant diagonals (c_{ij}) depending only on the lag $|i-j|$:

$$c_{ij} = \frac{1}{N-|i-j|} \sum_{t=1}^{N-|i-j|} X(t)X(t+|i-j|), \tag{9}$$

as is demonstrated in Vautard and Ghil (1989). This allows us to identify cyclical and seasonal patterns in the data.

After deriving the secular trend, we calculate which combination of the theorized hypotheses best fits the extracted trend of our primary dataset of interest. To do this, we first calculate a OLS line(s) of best fit for the detrended data proceeding the 2017 break point (if one is detected).

We then calculate an OLS model for all trend data on and after the break point. If the slope of the post-2017 data model is statically distinguishable from that of the segment proceeding 2017 then that will constitute strong evidence in support or either Hypothesis 1 or 3. Additionally, we calculate the probability that the intercept term for the model of best fit post-2017 is statistically distinguishable from the fitted value at the observation proceeding it using the pre-2017 data. If the intercept term is nontrivially smaller, that will constitute evidence of hypothesis 2. Finally, we wish to demonstrate that the coefficient on any OLS model applied to the U.S. data since 2017 is unique. We, thus, compare Chinese treatment of the U.S. to its treatment of other regional states since 2017.

4.2.3 Creating ARMAX and ARIMAX Models

Finally, based on the (possible) need to first-difference our data and soak up unknown variation in a systematic way, we turn to the autoregressive moving average model extended (ARMAX) and the autoregressive integrated moving average model extended (ARIMAX).[14] We construct each model for our primary dataset of interest. We begin with a discussion of the ARMAX model before turning to its first-differenced counterpart.

An ARMAX model expresses the expected value of y_t at any time, t, as a function of the past observations of y_t, $(y_{t-1}, y_{t-2}, ...)$, the past "innovations" (white noise error terms $u_{t-1}, u_{t-2},$ *etc.*) of the model (allowing it to systematically soak up unexplained stochastic variation), and the past observations of some identifiable control variable x_t. In this case, American treatment of China is included in the models as the known time-dependent control variable x_t to acknowledge the bilateral nature of the relationships modeled. Chinese behavior towards America, meanwhile, represents y_t in each respective bilateral time-dependent models. Our ARMAX model also includes nonzero time-invariant constants (α's) to better reflect the steady shifts that inevitably take place over time in the international system.[15] Our ARMAX model, thus, can be expressed as follows:

$$\widehat{y}_t = \alpha_1 + \sum_{i=1}^{p} \Phi_i y_{t-i} + \sum_{k=1}^{r} \Theta_k u_{t-k} + u_t + \sum_{j=1}^{q} \beta_j x_{t-j}. \tag{10}$$

We fix $p = 1$ for the sake of parsimony.[16] This allows us to derive a single coefficient of interest

[14] The ARIMAX model serves as the ARMAX model's first-differenced counterpart.
[15] As discussed above the international system is not defined by any long-term mean state. See Organski (1968).
[16] We explore alternative values for p in Appendix G.

Φ for each of the ARMAX models. This number represents the momentum in Chinese-initiated interactions as a function of past interactions (i.e., whether relations have a "direction" over time). By limiting our ARMAX models to the segmented periods before and after the 2017 breakpoint in Chinese-initiated interactions with the U.S., we can, thus, estimate the post-parity momentum in relations using a methodology that soaks up unknown stochastic variance in a systematic way. r and q are set by minimizing the Akaike information criterion (AIC) of the ARMAX models.

Yet the ARMAX model still suffers from the serial autocorrelation problem identified above. This problem risks detectioN of a spurious regression trap. We, thus, first-difference ($d = 1$) our ARMAX model, creating an ARIMAX model.[17] This ARIMAX model can be expressed as:

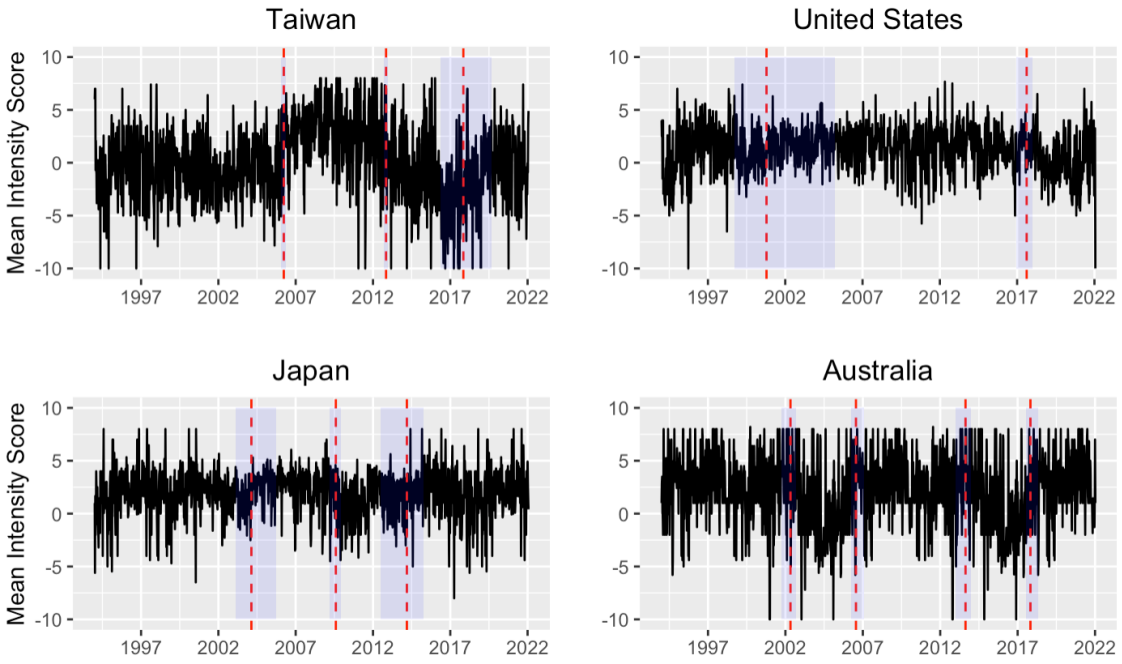
$$\widehat{y}_t = \alpha_2 + \sum_{\ell=1}^{d=1} y_{t-\ell} + \sum_{i=1}^{p=1} \Gamma_i (y_{t-i} - y_{t-i-1}) + \sum_{k=1}^{r} \Psi_k u_{t-k} + u_t + \sum_{j=1}^{q} \Omega_j \nabla^1 x_{t-j}, \tag{11}$$

where Γ_i is our primary coefficient of interest and where Θ_i and β_i are still set by minimizing the AIC of the model. This model solves the autocorrelation issue avoiding spurious regression. However, the interpretation of Φ_i has now changed to represent the expected effect the difference between y_{t-1} and y_{t-2} will have on y_t. Hence, the coefficient of interest now represents the existence of momentum from between period changes in the relationship. Essentially, if relations move up or down, does that create a self-reinforcing trend that draws strength from its own elan? Answering this question will further evince that post-parity relations are "trending" *either* up or down, as opposed to simply moving randomly.[18]

5 Results

We structure our results around the three methodological tests outlined above (break date analysis, OLS segment and OLS trend subset analysis, and ARMAX/ARIMAX analysis). In particular, we address each in turn demonstrating which hypotheses they evince and which they contradict. Generally, the empirical findings evince Hypotheses 1 and 2.

Figure 6: Breakpoints (Red) and 95 Percent Confidence Intervals (Blue) by Relationship



5.1 Confirming a Unique 2017 Break Date

We begin our analysis by checking for break dates in our four time series data sets that we have constructed. To reject the null hypothesis (that there is no breakpoint at parity), we need to observe a significant breakpoint in Chinese behavior towards the U.S. in 2017. This, in fact, is exactly what we observe with a sharp break detected in week 33 (August) of 2017. (Figure 6 illustrates this with breakpoints colored red and confidence intervals shaded blue.) Furthermore, the confidence for this breakpoint runs from the 48th week of 2016 to the first week of 2018. This provides strong evidence for a substantial shift in Chinese treatment of the U.S. during 2017 – directly after parity was reached. Furthermore, this breakpoint in Chinese-initiated interactions stands as one of only two in the previous twenty-eight years, and has a much narrower confidence interval than its counterpart in 2002, which has a confidence interval running from 1998 to 2005 (nearly a fourth of the period in question). In other words, 2017 was clearly a seminal year in U.S.-China relations.

Yet the evidence that 2017 was a unique year for the U.S. in terms of its treatment by China is a bit more mixed. While Chinese-initiated interactions with Japan show no evidence of a break date in 2017, the same cannot be said for Chinese interactions with Taiwan and Australia. In fact, China

[17] p is still set to $p = 1$.
[18] Although it will do so in a much less parsimonious way than any of the more simple models above.

appears to have shifted tact towards both smaller American allies in 2017 raising questions about the uniqueness of the American experience. We, therefore, pursue this trend further by performing a simple segmented OLS analysis.

5.2 Segmented OLS Analysis and Segmented OLS Trend Analysis

Having established that Chinese behavior towards the U.S. demonstrates a breakpoint in 2017, we next apply segmented OLS analysis to each of the four time series datasets. We are interested both in the direction of the OLS slope coefficients for the datasets in each segmented period, as well as the OLS slope coefficients for the secular trend during each break date-separated segment. We begin by investigating the segmented slope coefficients for the general data. These are shown in Figure 7 with their associated confidence intervals. Coefficients are shaded according to their size, conditional on their being statistically distinguishable from zero at the 90% level of confidence. The table clearly shows that each state's most recent segmented OLS slope coefficient is positive or statically indistinguishable from zero with the exception of the United States. This strongly evinces both the uniqueness of Chinese actions towards the U.S. since 2017, and it provides empirical supporting evidence in the large-N data for Hypothesis 1 (The Declining Relations Hypothesis).

Figure 7: OLS Regression Slope Coefficients by Between-Breakpoint Segment

Chinese Target State	*Segment 1*	*Segment 2*	*Segment 3*	*Segment 4*	*Segment 5*
United States	0.002692**	−0.000481**	−0.004105*	N/A	N/A
	(0.00127)	(0.000240)	(0.002241)		
Japan	0.0002722	−0.002078*	0.004556**	−0.0009280	N/A
	(0.0006091)	(0.001272)	(0.001816)	(0.0009812)	
Taiwan	−0.0009482	−0.002349	−0.016396***	0.010796***	N/A
	(0.0006095)	(0.001596)	(0.002777)	(0.003334)	
Australia	−0.0001087	−0.006501*	−0.003554**	−0.020984***	0.004068
	(0.0011430)	(0.003790)	(0.001404)	(0.003732)	(0.002976)
β est. < −0.01	β est. < −0.001	β est. < −0.0001	β est. > 0.0001	β est. > 0.001	β est. > 0.01

Note: * p < 0.1; ** p < 0.05; *** p < 0.01.

Particularly noteworthy is the cases of recent Chinese interactions with Taiwan and Chinese actions towards Japan from the beginning of 2004 until 2009. In the large-N data, the 2017 break date in Chinese-initiated interactions with Taiwan is mathematically indicative of the beginning of a period of (generally) warming mean weekly Chinese-initiated interactions. This finding though is at odds with press reports and the general scholarly consensus. We, therefore, suspect that the large-N

data could demonstrate that there are cognitive biases in how scholars and commentators think about Chinese interactions with Taiwan. Alternatively, it may reflect a very low baseline of Chinese behavior towards Taiwan before 2017. China could simply be treating Taiwan progressively "less terribly" while treating the U.S. progressively "less well" over time. Regardless, this large-N result – the first of its kind – should invite further large-N studies of Chinese interactions with Taiwan.

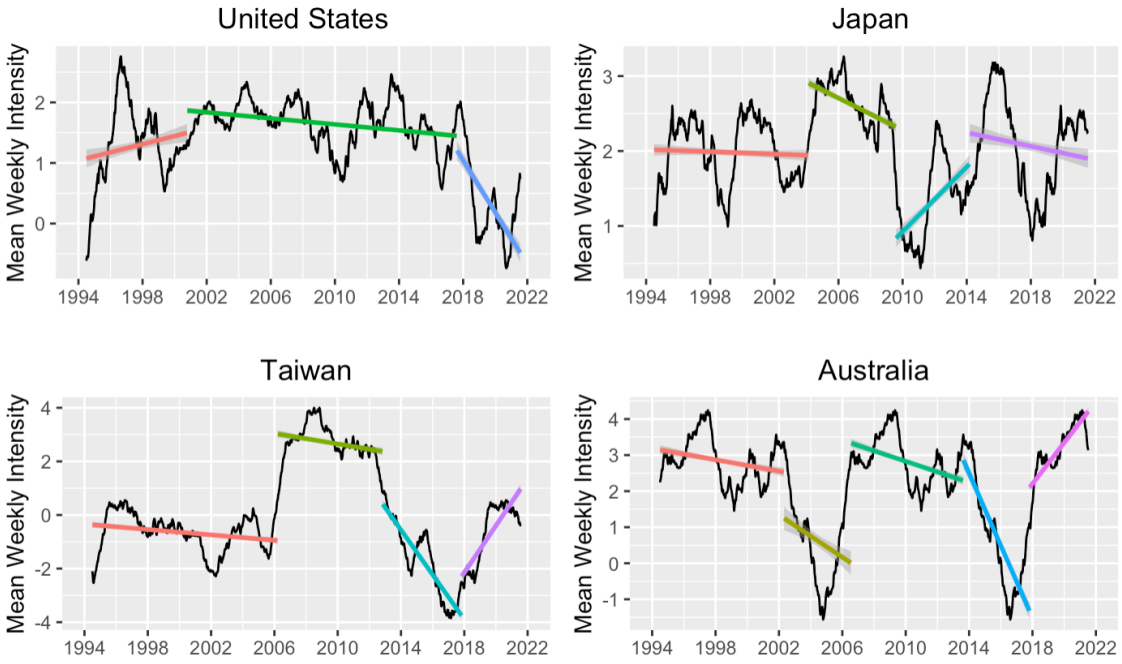
The record of Chinese interactions with Japan from 2004 to 2009, meanwhile, evince a local power transition causing a temporary discombobulation in relations, as regional parity was reach in 2003. Yet the data also demonstrates that as China surpassed 120% of Japan's power level (World Bank 2020b, 2020a), relations recovered rapidly until 2014 before leveling out. We, thus, recommend regional power transitions as another promising research area for large-N studies.

Returning to the primary subject of the paper, we now turn to the secular trends for evidence of our hypotheses. We begin by additively extracting the secular trends from our time series datasets. The secular trends as shown in Figure 8 clearly demonstrate that only Chinese-initiated interactions the U.S. have degenerated since 2017. By fitting segmented OLS lines onto the secular trend of the Chinese-initiated interactions with the U.S., we can observe further evidence for Hypotheses 1 and 2. In fact, the negative trend in Chinese-initiated interactions with the U.S. since August 2017 becomes pronounced once seasonal and cyclical elements are removed from the data, strongly evincing Hypothesis 1. The final fitted value of Segment 2's secular trend OLS model, moreover, is nontrivially larger than the first fitted value of Segment 3's secular trend model. This demonstrates a sharp deterioration in Chinese-initiated interactions with the U.S. at the parity breakpoint. We, thus, find evidence for Hypotheses 1 and 2 in the segmented secular trend data.

5.3 ARMAX and ARIMAX: Stochastic Error Mitigation, Controls for Bilateral Diplomacy, and Differencing

For our next two pieces of evidence we turn to ARMAX and and ARIMAX modeling where $p = 1$ is assumed. A brief application of the augmented Dickey-Fuller test generates a p-value of 0.01, indicating that the time series data is stationary. Conveniently, $p = 1$ is, thus, not only parsimonious but also mathematically appropriate. This unanticipated finding of stationarity also throws into doubt the value of differencing, but we proceed with the analysis plan already laid out in our Methods section for the sake of transparency.

Figure 8: Segmented OLS Models for Extracted Secular Trend Data



Having already demonstrated the primary dataset's uniqueness and direction, we focus this section exclusively on it and ignore the three control datasets. Yet before we can begin generating estimates for the momentum in the relationship via the coefficients of the autoregression terms, we must first create a control dataset to account for the bilateral nature of the U.S.-China relationship. We do this by applying the same methods already used to create the first four datasets. This final dataset – created as a control – represents U.S.-initiated interactions towards China modeled as time series data. We then select lagged values for the control and past innovations following a strategy of minimizing AIC values. AIC and BIC values for different model specifications are calculated in Appendix H. The final ARMAX model yielded by this modeling criteria is:

$$\widehat{y}_t = \alpha_1 + \Phi_1 y_{t-1} + \Theta_1 u_{t-1} + \Theta_2 u_{t-2} + \Theta_3 u_{t-3} + u_t + \beta_1 x_{t-1}, \tag{12}$$

where y_t is the primary time series dataset of interest (Chinese-initiated interactions with the U.S.), x_t is the new control dataset of U.S. actions towards China, and Φ_1 is our primary coefficient of interest.

To evince Hypothesis 1, we must demonstrate that relations are not just trending down, but that this downward trend has a "momentum" to it (i.e. – that past actions influence

future ones). This momentum is a characteristic of states caught in the Thucydides trap (Allison 2017). We calculate the best-fit coefficients of this model before and after week 33 of 2017, to determine if the parity breakpoint had a substantial effect on the momentum in the relationship. This yields the two estimates for Φ_1 before and after the breakpoint shown in Figure 9.[19]

Figure 9: Autocorrelation Coefficients Before and After Parity

Value of the Autocorrelation Coefficient	*ARMAX(1,0,3)*	*ARIMAX(1,1,2)*
Before Parity	0.2743	0.7296***
	(0.1723)	(0.0764)
After Parity	0.7585***	0.8404***
	(0.1323)	(0.0861)

Note: * p < 0.05; ** p < 0.01; *** p < 0.001.

These demonstrate that momentum only emerged in the relationship after parity was reached. Relations vacillated before parity, and Chinese actions one week rarely affected Chinese actions in the next week as demonstrated by $\widehat{\Phi}_{1,a}$, which is not statistically distinguishable from zero. Yet after parity $\widehat{\Phi}_{1,b}$ demonstrates that a substantial degree of momentum emerged in the relationship. Based on the results of the trend distillation and OLS segmentation above, we assume this momentum was negative. This reinforces our strong belief in the validity of Hypothesis 1.

Next, we turn to the best-fit ARIMAX model:

$$\widehat{y}_t = \alpha_2 + y_{t-1} + \Gamma_1(y_{t-1} - y_{t-2}) + \Psi_1 u_{t-1} + \Psi_2 u_{t-2} + u_t + \Omega_1(x_{t-1} - x_{t-2}). \tag{13}$$

This model yields coefficient estimates for how previous changes in the tone of China-initiated interactions with the U.S. affect current relations. The ARIMAX model results fly completely in the face of all of our other modeling results. We attribute this incongruence to differencing and view our ARIMAX results with extreme skepticism as quintessential outliers.

The rationale is relatively straight-forward: While differencing helps to somewhat overcome the autocorrelation issues identified in the ACF charts shown in Figure 5, it also muddies any interpretations of coefficients. In particular, differencing raises two concerns that lead us to discount

[19]For the remainder of this section these two estimates of Φ_1 shall be referred to as $\widehat{\Phi}_{1,a}$ and $\widehat{\Phi}_{1,b}$.

the ARIMAX findings. First, because we do not know any information about the size of the average change before and after parity, this differenced result could potentially produce two similar coefficients with radically different interpretations. Second, first-differencing the data also provides less benefit than the ACF plots would lead a casual observer to initially anticipate since the data is stationary. Essentially, differencing trades parsimony for little benefit. Hence, we view our outlier ARIMAX results with extreme skepticism.

6 Conclusion

The rise of China has led to an outgrowth of interest in PTT, and a surge of scholarship exploring the empirical prospects for a peaceful power transition. Yet modern PTT scholars still rely largely on traditional formal and qualitative methods, limiting their ability to perform rigorous statistical tests of their empirical hypotheses. Given the the high stakes involved, bringing novel and hetero-orthodox methodologies to bear on the problem is, therefore, an axiomatically useful endeavor.

By harnessing 52,189 separate observations of Chinese actions directed towards the US, we introduce one such unorthodox methodology, and in doing so, identify time series trends in Chinese behavior consistent with those of a dissatisfied challenger. Our large-N time series analysis statistically identifies a significant negative deterioration in Chinese-initiated interactions with the U.S. when parity was reached in 2017. Furthermore, it evinces a self-reinforcing negative trend in Chinese behavior towards the U.S. after parity is reached as a function of time. In fact, our segmented OLS analysis predicts that if current trends continue, Chinese-initiated "diplomacy" with the U.S. will consist entirely of the threat or use of military force by 2070.[20] Extracted secular trend analysis confirms this finding, and ARMAX analysis – which incorporates American initiated-interactions with China – evinces a self-reinforcing momentum (indicative of a Thucydides Trap) contained within this negative trend. These findings highlight the extreme risks inherent in the current Chinese diplomatic trajectory.

Yet secondary findings offer a glimmer of hope. Chinese-initiated interactions with U.S. regional allies have been trending in a positive direction since 2017, decreasing the volatility of potential

[20]This assumes that an individual disagreement does not spiral into war sooner.

flashpoints for kinetic conflict. China's peaceful displacement of Japan also provides a glimmer of hope for U.S.-China relations. While Chinese-initiated interactions with Japan did become more belligerent during their period of regional parity (2004–2008), the displacement episode was brief, and relations quickly recovered once China became dominant. If China surpasses the U.S. in a similarly rapid fashion, it might palliate the potential for system-wide discombobulation.[21]

Our study also suggests some skepticism about some pervasive assumptions in scholarly writings about modern Chinese diplomacy. First, we find that China's interactions with the United States are not necessarily predictive of its interactions with American allies or vice-versa. In fact, China commonly seeks to improve relations with America's allies while treating the U.S. worse or vice versa. Hence, purposefully – or inadvertently – studying Chinese diplomacy through an Ameri-centric lens (or any other national lens) may provide an inaccurate view of empirical events. Second, our large-N results suggest that focusing on major empirical episodes in Chinese diplomacy may render scholars unable to see the empirical forest through the episodic trees. In particular, our finding that Chinese-initiated interactions with Taiwan have actually warmed *on average* since 2017 demonstrates that a focus on individual negative episodes in a relationship may obscure general improvement, leading scholars to incorrect readings of recent events.

Finally, in demonstrating the ability of large-N data to aid in studies of diplomacy, these findings suggest several directions for future research. For one, scholars (e.g., Lemke 2002) have long theorized as to the existence to regional subsystems within the international system. Time series analysis of automated signals data provides a novel methodology that could allow scholars to study less commonly documented substate systems (e.g., the manifold regional African systems). An analysis of these systems could provide insight into whether IR's (mostly) Euro-centric understandings of the causes of systemic conflict (Zvobgo and Loken 2020) are broadly applicable. Our analysis also demonstrates a novel methodological alternative to qualitative analysis for empirical scholars. This alternative could inform future studies of time-limited discrete events. Third, we largely side-step the question here of what specific events are causing our observed break dates. Future work should dig deeper into signalling data to better understand the discrepancies between our large-N findings and the findings of qualitative scholars. This study, thus, points to a number of possible future

[21] This is in line with Organski's supposition that systemic conflict is less likely when a challenger's size practically guarantees its eventual dominance (Organski 1968, 372).

research avenues.

7 References

Alden, Chris, and Kenddrick Chan. 2021. "Twitter and Digital Diplomacy: China and COVID-19." *LSE Covid-19*, June. https://blogs.lse.ac.uk/covid19/2021/06/09/twitter-and-digital-diplomacy-china-and-covid-19/.

Allen, Robert C. 1979. "International Competition in Iron and Steel, 1850–1913." *The Journal of Economic History* 39 (4): 911–37.

Allison, Graham. 2015. "The Thucydides Trap: Are the US and China Headed for War?" *The Atlantic* 24 (9).

———. 2017. *Destined for War: Can America and China Escape Thucydides's Trap?* Boston, MA: Houghton Mifflin Harcourt.

Bai, Jushan. 1994. "Least Squares Estimation of a Shift in Linear Processes." *Journal of Time Series Analysis* 15 (5): 453–72.

———. 1997. "Estimation of a Change Point in Multiple Regression Models." *Review of Economics and Statistics* 79 (4): 551–63.

———. 2010. "Common Breaks in Means and Variances for Panel Data." *Journal of Econometrics* 157 (1): 78–92.

Bai, Jushan, and Pierre Perron. 2003. "Computation and Analysis of Multiple Structural Change Models." *Journal of Applied Econometrics* 18 (1): 1–22.

Bolt, Jutta, and Jan Luiten van Zanden. 2020. "Maddison Style Estimates of the Evolution of the World Economy." Maddison Project Database, version 2020. https://www.rug.nl/ggdc/historicaldevelopment/maddison/releases/maddison-project-database-2020.

Brunnermeier, Markus, Rush Doshi, and Harold James. 2018. "Beijing's Bismarckian Ghosts: How Great Powers Compete Economically." *The Washington Quarterly* 41 (3): 161–76.

Chan, Steve. 2004. "Exploring Puzzles in Power-Transition Theory: Implications for Sino-American Relations." *Security Studies* 13 (3): 103–41.

Chubb, Andrew. 2020. "PRC Assertiveness in the South China Sea: Measuring Continuity and Change, 1970–2015." *International Security* 45 (3): 79–121.

Clark, Hunter, Maxim Pinkovskiy, and Xavier Sala-i-Martin. 2017. "China's GDP Growth May Be Understated." w23323. Cambridge, MA: National Bureau of Economic Research.

Colaresi, Michael, David Lindsey, Eric Min, and Paul Poast. 2018. "Advancing the Scientific Study of Diplomacy: Events Data." *Conference of the Peace Science Society*, 7.

Deudney, Daniel, and G John Ikenberry. 2009. "The Myth of the Autocratic Revival: Why Liberal Democracy Will Prevail." *Foreign Affairs* 88 (1): 77–93.

Epstein, Katherine C. 2019. "To Understand China, Look to America's History." *The Wall Street Journal*, March.

Gilpin, Robert. 1981. *War and Change in World Politics*. Cambridge University Press.

Kennedy, Paul. 1980. *The Rise of the Anglo-German Antagonism, 1860-1914*. London: George Allen & Unwin.

Kugler, Jacek, and Douglas Lemke. 2000. "The Power Transition Research Program: Assessing Theoretical and Empirical Advances." *Handbook of War Studies II*, 129–63.

Kugler, Jacek, Ronald L. Tammen, and Siddharth Swaminathan. 2001. "Power Transitions and Alliances in the 21st Century." *Asian Perspective*, 5–29.

Kydd, Andrew. 1997. "Game Theory and the Spiral Model." *World Politics* 49 (3): 371–400.

Larson, Caleb. 2020. "China's J20 Stealth Fighter Is Getting an Upgrade—Thanks to Russia." *The National Interest*.

Layne, Christopher. 2020. "Coming Storms: The Return of Great-Power War." *Foreign Affairs* 99 (6): 42–53.

Lemke, Douglas. 2002. *Regions of War and Peace*. Vol. 80. Cambridge University Press.

Lemke, Douglas, and Ronald L. Tammen. 2003. "Power Transition Theory and the Rise of China." *International Interactions* 29 (4): 269–71.

Lemke, Douglas, and Suzanne Werner. 1996. "Power Parity, Commitment to Change, and War." *International Studies Quarterly* 40 (2): 235–60.

Lim, Yves-Heng. 2015. "How (Dis) Satisfied Is China? A Power Transition Theory Perspective." *Journal of Contemporary China* 24 (92): 280–97.

Martin, Peter. 2021. *China's Civilian Army: The Making of Wolf Warrior Diplomacy*. New York, NY: Oxford University Press.

McClelland, Charles A. 1983. "Let the User Beware." *International Studies Quarterly* 27 (2): 169–77.

Mead, Walter Russell. 2014. "The Return of Geopolitics: The Revenge of the Revisionist Powers." *Foreign Affairs* 93 (3): 69.

Mearsheimer, John J. 2010. "The Gathering Storm: China's Challenge to US Power in Asia." *The Chinese Journal of International Politics* 3 (4): 381–96.

———. 2014. "Can China Rise Peacefully?" *The National Interest* 25 (1): 1–40.

Michalski, Tomasz, and Gilles Stoltz. 2013. "Do Countries Falsify Economic Data Strategically? Some Evidence That They Might." *Review of Economics and Statistics* 95 (2): 591–616.

Morris, Charles R. 2012. "We Were Pirates, Too: Why America Was the China of the 19th Century." *Foreign Policy*, December.

Obama, Barack H. 2009. Remarks by the President at the U.S./China Strategic & Economic Dialogue. Washington, D.C. https://obamawhitehouse.archives.gov/realitycheck/the-press-office/remarks-president-uschina-strategic-and-economic-dialogue.

Orde, Anne. 1996. *The Eclipse of Great Britain: The United States and British Imperial Decline, 1895–1956*. New York, NY: St. Martin's Press.

Organski, Abramo F. K. 1968. *World Politics*. 2nd ed. [rev.]. Alfred A. Knopf.

Organski, Abramo F. K., and Jacek Kugler. 1981. *The War Ledger*. University of Chicago Press.

Owyang, Michael T., and Hannah Shell. 2017. "China's Economic Data: An Accurate Reflection, or Just Smoke and Mirrors?" Federal Reserve Bank of St. Louis.

Peck, Michael. 2020. "China's Purchase of Russia's Su-35 Fighter Shows It Isn't Ready for Great Power Status." *The National Interest*.

Perkins, Bradford. 1968. *The Great Rapprochement: England and the United States, 1895-1914*. New York: Atheneum.

Randt, Clark T. 2007. "Fifth Generation Star Li Keqiang Discusses Domestic Challenges, Trade Relations, with Ambassador." Washington, D.C.: U.S. Department of State, Confidential diplomatic cable, Paragraph 4.

Rauch, Carsten. 2017a. "A Tale of Two Power Transitions: Capabilities, Satisfaction, and the Will to Power in the Relations Between the United Kingdom, the United States, and Imperial Germany." *International Area Studies Review* 20 (3): 201–22.

———. 2017b. "Challenging the Power Consensus: GDP, CINC, and Power Transition Theory." *Security Studies* 26 (4): 642–64.

Reed, William. 2003. "Information, Power, and War." *American Political Science Review* 97 (4): 633–41.

Roberts, Farley. 2014. "China Might Just Be an Assertive Status Quo Power." *The Diplomat*, August.

Roberts, Tim. 2014. "US-China: 19th Anglo-American Rivalry Redux?" *The Diplomat*, February.

Robertson, Peter E. 2021. "The Real Military Balance: International Comparisons of Defense Spending." *Review of Income and Wealth* forthcoming.

Rock, Stephen R. 2014. "British Appeasement of the United States, 1896–1903." In *Appeasement in International Politics*, 25–48. Lexington, KY: University Press of Kentucky.

Schrodt, Philip A, and Deborah J Gerner. 1994. "Validity Assessment of a Machine-Coded Event Data Set for the Middle East, 1982-92." *American Journal of Political Science*, 825–54.

Shifrinson, Joshua. 2020. "The Rise of China, Balance of Power Theory and US National Security: Reasons for Optimism?" *Journal of Strategic Studies* 43 (2): 175–216.

SIPRI. 2020. "SIPRI Military Expenditure Database." Stockholm, Sweden: Stockholm International Peace Research Institute. https://www.sipri.org/databases/milex.

Stafford, David. 1982. "A Moral Tale: Anglo-German Relations, 1860–1914." *The International History Review* 4 (2): 249–63.

Tammen, Ronald L. 2008. "The Organski Legacy: A Fifty-Year Research Program." *International Interactions* 34 (4): 314–32.

Tammen, Ronald L., and Jacek Kugler. 2006. "Power Transition and China–US Conflicts." *The Chinese Journal of International Politics* 1 (1): 35–55.

Tammen, Ronald L., Jacek Kugler, Douglas Lemke, and Allan C. Stam III. 2000. *Power Transitions: Strategies for the 21st Century*. CQ Press.

Thussu, Daya Kishan, Hugo De Burgh, and Anbin Shi. 2018. *China's Media Go Global*. London, UK: Routledge.

Vautard, Robert, and Michael Ghil. 1989. "Singular Spectrum Analysis in Nonlinear Dynamics, with Applications to Paleoclimatic Time Series." *Physica D: Nonlinear Phenomena* 35 (3): 395–424.

Wallace, Jeremy L. 2016. "Juking the Stats? Authoritarian Information Problems in China." *British Journal of Political Science* 46 (1): 11–29.

Wang, Jianwei, and Xiaojie Wang. 2014. "Media and Chinese Foreign Policy." *Journal of Contemporary China* 23 (86): 216–35.

Weidmann, Nils B. 2016. "A Closer Look at Reporting Bias in Conflict Event Data." *American Journal of Political Science* 60 (1): 206–18.

Wenckus, Joseph H. 2015. "Sino-American Relations in the 21st Century: Taking a Page from the Venezuelan Crisis of 1895." National Defense University Norfolk, VA Joint Advanced Warfighting School.

Woodward, Ernest Llewellyn. 1935. *Great Britain and the German Navy*. Oxford: Clarendon Press.

World Bank. 2020a. "GDP (Current US$)." Washington, DC: The World Bank Group. https://data.worldbank.org/indicator/NY.GDP.MKTP.CD.

———. 2020b. "GDP, PPP (Current International $)." Washington, DC: The World Bank Group. https://data.worldbank.org/indicator/NY.GDP.MKTP.PP.CD.

———. 2020c. "PPP Conversion Factor, GDP (LCU Per International $)." Washington, DC: The World Bank Group. https://data.worldbank.org/indicator/PA.NUS.PPP.

Xuetong, Yan. 2011. "How China Can Defeat America." *The New York Times*, November.

Yao, Yi-Ching. 1988. "Estimating the Number of Change-Points via Schwarz' Criterion." *Statistics & Probability Letters* 6 (3): 181–89.

Yoder, Brandon K. 2019. "Uncertainty, Shifting Power and Credible Signals in US-China Relations: Why the 'Thucydides Trap' Is Real, but Limited." *Journal of Chinese Political Science* 24 (1): 87–104.

Zeileis, Achim, Friedrich Leisch, Kurt Hornik, and Christian Kleiber. 2002. "Strucchange: An r Package for Testing for Structural Change in Linear Regression Models." *Journal of Statistical Software* 7 (1): 1–38.

Zeileis, Achim, Ajay Shah, and Ila Patnaik. 2010. "Testing, Monitoring, and Dating Structural Changes in Exchange Rate Regimes." *Computational Statistics & Data Analysis* 54 (6): 1696–706.

Zhao, Suisheng. 2018. "A Revisionist Stakeholder: China and the Post-World War II World Order." *Journal of Contemporary China* 27 (113): 643–58.

Zvobgo, Kelebogile, and Meredith Loken. 2020. "Why Race Matters in International Relations." *Foreign Policy* 237: 11–13.