Spheres of Influence*

Matteo Camboni† Michael Porcellacchia‡

March 2, 2025

Abstract

We propose and test a theory on how geopolitical risk and the economic balance between great powers (e.g., the US and China) influence geoeconomic fragmentation. In our model, countries design foreign policies to pursue their goals while managing the risk of coercion by two great powers. As geopolitical risk rises, some countries remain non-aligned, while others optimally choose to align with one power, making significant concessions to that great power only. This alignment strategy endogenously creates two distinct *spheres of influence* gravitating around the two great powers. Changes in the balance of power affect geoeconomic fragmentation in a non-trivial way. As the world becomes less unipolar (e.g., with China's rise), countries deepen their respective alignments, pushing the two spheres of influence farther apart in terms of foreign policies. Paradoxically, China's economic growth may drive countries within the US sphere of influence to offer even greater concessions to the US. Empirically, we test our model and show that it correctly predicts the shifts in alignment with the US that followed two exogenous shocks to the balance of power: the Soviet Union's collapse and China's rise.

Keywords: Geopolitics, Alignment, Non-Alignment, Geoeconomic Fragmentation, Geopolitical Competition, Spheres of Influence, Foreign Influence, Rise of China, Cold War

*For useful comments and suggestions we thank Sandeep Baliga, Pietro Bomprezzi, Matthias Doepke, Georgy Egorov, Dana Foarta, Walker Hanlon, Leander Heldring, Giampaolo Lecce, Matteo Magnaricotte, Johannes Marzian, Ralph Meisenzhal, Joel Mokyr, Massimo Morelli, Ameet Morjaria, Alessandro Pavan, Harry Pei, Davide Porcellacchia, Nancy Qian, Stephen J. Redding, Michele Rosenberg, Paola Sapienza, Konstantin Sonin, Lorenzo Stanca, Christoph Trebesch, and participants to the history seminars at Northwestern University, the XI Ibeo Workshop on Political Economy, the 2022 Annual Meeting of the Canadian Economic Association, and the 2022 European Meeting of the Econometric Society. The usual disclaimer applies.

†Economics Department, University of Wisconsin-Madison, camboni@wisc.edu

‡Kiel Institute for the World Economy, michael.porcellacchia@ifw-kiel.de

1 Introduction

China's growing foreign influence and escalating tensions with the United States have raised concerns about a potential return to a geopolitically fragmented global economy (Gopinath et al., 2025). These concerns echo historical patterns, most notably during the Cold War, when tensions between great powers gave rise to *spheres of influence*—blocs of countries adopting policies closely aligned with one great power or the other. Given that most nations today are deeply integrated with both the US and China, and both great powers appear disinterested in forcing such dichotomous alignments, these concerns might seem unwarranted.[1] Nevertheless, the process of geoeconomic fragmentation appears to be well underway, posing an important puzzle: why would countries still choose to align with one particular great power rather than remaining neutral and pursue their own domestic goals? And how is this choice affected by geopolitical risk and the balance of power between great powers?

To answer these questions, we propose and test a new bargaining model in which every country determines its foreign policies toward two great powers (e.g., offering economic cooperation or diplomatic concessions) by trading off its policy preferences against the geopolitical risk of losing their support or facing their coercion. Our paper offers three main contributions: (i) We show how spheres of influence may endogenously emerge regardless of the great powers' will: when geopolitical risk is sufficiently high, some countries optimally choose to align, making significant concessions to *one* great power only (typically the one holding the greater economic/military leverage); (ii) We show how geoeconomic fragmentation increases as the power gap between the great powers closes. For example, our model predicts that the rise of China should push countries within the US sphere of influence to make even more concessions to the US, while having the opposite effect on others—a counterintuitive pattern already documented in both Asia and Europe.[2] (iii) We develop a new empirical test of geoeconomic fragmentation based on our model and confirm its predictions. Using UN voting data to measure alignments and trade dependence for foreign influence, we analyze how countries adjusted their relations with the US in response to two exogenous shocks to their exposure to the other great power: the Soviet Union's collapse and China's rise.

Our main theoretical contribution is the development of a new bargaining model to examine how countries optimally manage geopolitical risk and foreign pressures. In our model, a domestic country designs policies that affect both its own payoffs and those of two great powers. To capture a setting where each country can maintain cooperative relations

[1] For example, Xi Jinping has repeatedly stated that "China does not seek spheres of influence," and Antony Blinken remarked that such notions "should be relegated to the dustbin of history."

[2] See, for example, Selden (2013) and Koga (2018). We discuss the anecdotal evidence in Section 4.1.

with both great powers, we assume that they value different policy dimensions.[3] Moreover, each great power has some *foreign influence* over the domestic country—a bilateral parameter capturing its military and economic leverage over the domestic country.

Once the country announces its policies, each great power decides whether to support or oppose/coerce the country. The payoff of a supporting great power depends solely on the announced policies, while an opposing power's payoff depends on whether its attempt to coerce the country into revising its policies in its favor succeeds. The probability that a great power's coercive action succeeds, which naturally increases with its foreign influence, is lower when the other great power supports the country. This feature introduces strategic complementarity between great powers, potentially leading to multiple equilibria. The same announced policy may result in either a good equilibrium, where both great powers support the domestic government, or a bad equilibrium, where both resort to coercion, depending on their expectations of each other's actions. We interpret the probability of great powers coordinating on the bad equilibrium (when present) as the *geopolitical risk* faced by the domestic country. To formalize this idea, we introduce a sunspot: a binary public signal that determines whether great powers coordinate on the good or bad equilibrium.

Our first main result is that, in equilibrium, foreign policies must follow one of two strategies: (i) secure the support of *both* great powers irrespective of the sunspot (full-insurance), insulating the country from geopolitical risk, or (ii) secure their support only contingent on a favorable sunspot realization (partial-insurance), thus remaining exposed to geopolitical risk. The optimal partial-insurance strategy plays one great power against the other by offering policy concessions that discourage opposition from each, but only when they expect the domestic government to be supported by the other. This strategy involves minor concessions and treats the two great powers symmetrically (*non-alignment*), but it exposes the country to geopolitical risk by triggering the coercion from both great powers after an unfavorable sunspot realization. Conversely, the optimal full-insurance policy entails securing the unconditional support of *one* great power while averting the opposition of the second by making it indifferent between supporting and opposing a domestic government empowered by the first great power's support. Since obtaining unconditional support requires much larger policy concessions than conditional support, this strategy endogenously creates an asymmetry in the country's relations with the two great powers (*alignment*). We then fully characterize this second strategy (*alignment strategy*), showing that a domestic government optimally chooses to align with—securing the unconditional support of—the great power holding greater influence in the country.[4]

[3] Assuming a direct conflict of interest between the great powers would only reinforce our main results.

[4] In Section 3.2, we allow for preference asymmetries and show that, unsurprisingly, the domestic govern-

Our final theoretical results examine how geopolitical risk and the balance of power shape geoeconomic fragmentation. First, our model predicts a world divided into three blocs: a bloc of non-aligned countries and two *spheres of influence*, each comprising countries aligned with one of the two great powers. Second, we show that a country's placement within these blocs depends on its relative exposure to foreign influence: countries sufficiently more exposed to one great power optimally enter its sphere of influence, while those with a more balanced exposure remain non-aligned. Third, we show that, as geopolitical risk rises, countries gradually transition from non-alignment to one of the two spheres of influence until all countries choose to align. Finally, we show that countries aligning with a great power must compensate it for their exposure to the rival power (*balancing effect*). As a result, when the rival's influence grows, a great power obtains more concessions from countries within its sphere and fewer from the others. Thus, as the global balance of power becomes more bipolar (as with China's rise), geoeconomic fragmentation intensifies, with countries in opposing spheres adopting increasingly divergent policies. Conversely, in a more unipolar world (as with the Soviet Union's collapse), fragmentation subsides.

We begin our empirical analysis by examining a well-documented puzzle in international relations through the lens of our model. During the 1990s, the United States improved its bilateral relations with countries historically within Russia's sphere of influence (e.g., Eastern Europe) while simultaneously experiencing a deterioration in relations with countries traditionally considered within its own sphere (e.g., Latin America). While initially counterintuitive, this pattern aligns with our theoretical finding that the collapse of a rival benefits the United States outside its sphere of influence but harms it within its own sphere. Similarly, in line with our model's predictions, we discuss how the international system has become increasingly polarized as China has started catching up with the United States.

Finally, we conduct a new test of geoeconomic fragmentation, evaluating the main empirical predictions of our model: (i) countries more exposed to a great power are more likely to align with it, and (ii) countries aligning with a great power must compensate it for their exposure to the other great power. We rely on standard data on bilateral relations and foreign influence from 1960 to 2014. For the first, we use established measures of how a country's votes in the United Nations General Assembly align with those of the United States. For foreign influence, we consider a range of variables, from trade dependence to more sophisticated indices such as the formal bilateral influence capacity index (Moyer et al., 2018).

A key empirical challenge is the potential endogeneity of alignment and foreign influence. For example, a country's intrinsic preferences may simultaneously shape both its bilateral

ment aligns with the strategically/ideologically closer great power unless the other holds sufficiently greater influence.

relations with the US and its exposure to Russia or China. To address this issue, we exploit two major geopolitical shocks as instrumental variables: the collapse of the Soviet Union and China's rise. These events generated plausibly exogenous shifts in foreign influence by reshaping the global political and economic landscape independently of any individual country's pre-existing alignments or preferences. Thus, our empirical analysis amounts to assessing how countries more exposed to the United States adjust their international relations as their exposure to Russia or China changes. The empirical results strongly support our theoretical predictions across multiple measures of foreign influence and specifications, underscoring the robustness and empirical relevance of our theoretical analysis.

Related Literature. Our paper contributes to the emerging literature on geoeconomics by moving beyond the traditional focus on a single great power (hegemon) leveraging its foreign influence to coerce other countries' policies. While previous studies have illustrated how hegemonic coercion can enhance national welfare and expand global trade (Antràs and Padró i Miquel, 2011; Broner et al., 2024) or how it may extract geopolitical rents and distort trade and capital flows (Clayton, Maggiori and Schreger, 2023, 2024), little is known about how the simultaneous influence of *multiple* great powers jointly shapes other countries' foreign policies. We address this gap by examining the trade-off between alignment and non-alignment strategies in settings where countries face pressure from more than one great power; our empirical analysis underscores the importance of accounting for these multiple influences to understand the dynamics of bilateral relations, and our novel approach allows us to explore the mechanisms driving geoeconomic fragmentation and the formation of alignment blocs (spheres of influence). In doing so, our model provides a theoretical foundation for the empirical literature on the economic costs of geoeconomic fragmentation (Aiyar, Presbitero and Ruta, 2023), the drivers of alignments and de-globalization (Kleinman, Liu and Redding, 2020; Gopinath et al., 2025; Aiyar et al., 2023; Aiyar and Ohnsorge, 2024), and the broader impact of alignment behavior on various economic outcomes.[5]

Our paper also contributes to the political economy literature on bargaining in the shadow of coercion/conflict (reviewed in Jackson and Morelli (2011), Baliga and Sjöström (2013), and Egorov and Sonin (2024)). Within this literature, one strand focuses on key drivers of bargaining failures and, consequently, conflicts (Fearon, 1995; Powell, 1999; Jackson and Morelli, 2009; Yared, 2010; Chassang and Miquel, 2010; Morelli and Rohner, 2015; Meirowitz et al., 2019; Baliga and Sjöström, 2020), while another focuses on settings where bargaining always succeeds in equilibrium (Acemoglu and Robinson, 2000; Acemoglu, Robinson and

[5]For example, trade (Yeats, 1990; Berger et al., 2013; Fuchs and Klann, 2013; Du et al., 2017; Davis, Fuchs and Johnson, 2019), aid (Kuziemko and Werker, 2006; Faye and Niehaus, 2012), loans (Ambrocio and Hasan, 2021), and investments (Aiyar, Malacrino and Presbitero, 2024).

Verdier, 2004; Padró i Miquel, 2007; Acemoglu, Egorov and Sonin, 2008; Diermeier, Egorov and Sonin, 2017). While bargaining may fail in our equilibria, our primary focus is on characterizing the bargaining outcomes that emerge in the presence of strategic complementarities and how they give rise to alignment strategies.

From a technical standpoint, our paper relates to the growing literature on contracting with externalities, pioneered by Segal (1999) and Winter (2004), that examines how a principal simultaneously deals with multiple agents when concerned that they might coordinate on her least-preferred equilibrium outcomes as opposed to her most-preferred outcome (e.g., Halac, Kremer and Winter (2020); Halac, Lipnowski and Rappoport (2021); Kapon, Del Carpio and Chassang (2024)). With respect of this literature, we adopt an intermediate approach to principal pessimism by explicitly introducing a parameter, geopolitical risk, that explicitly captures the probability that the agents (great powers) coordinate on the principal's (domestic country's) least-preferred equilibrium.

Finally, our paper contributes to the extensive literature in international relations that examines the determinants of foreign policy biases (*alignment*). The cornerstone of this literature is Waltz (1979), which introduced the concepts of *bandwagoning* (aligning with the strongest) and *balancing* (aligning with the weakest to offset exposure to the strongest). Since the early 2000s, scholars have explained more nuanced post-Cold War alignment patterns using the concept of *hedging*, wherein countries avoid picking sides and maintain good relations with all great powers (Roy, 2005; Wohlforth et al., 2007; Goh, 2006; Wivel, 2008; Tessman, 2012; Selden, 2013; Guzansky, 2015; Koga, 2018; Lim and Mukherjee, 2019; Strating, 2020; Hamdi and Salman, 2020; Haacke and Ciorciari, 2022). Our contribution is to show that the optimal strategy combines elements of all these strategies thus rationalizing the seemingly contradictory evidence that is discussed in this literature.

Structure of the paper. Section 2 presents the set-up of the model and a discussion of the main assumptions. Section 3 includes our theoretical analysis. Section 4 describes our empirical exercises. Section 5 concludes.

2 Model Set-Up

We study a world where H countries design their foreign policies constrained by the influence of two *great powers* $f \in \{F, G\}$, e.g. the United States and the Soviet Union or China. Our objective is to show why and when such countries should be expected to adopt policies skewed in favor of one or the other great power (thus forming "spheres of influence") even in a model where they can design foreign policies that appease both f and $-f$.

2.1 Foreign Influence and Bargaining

For every *home* country $h \in H$ and great power f, denote by $X_{hf} \in \mathbb{R}_+$ the foreign policy adopted by h toward f, affecting the payoffs from their bilateral relation. From an empirical perspective, we can think of $X_{hf} \geq 0$ as specifying how much focus or attention h dedicates to its relation with f, e.g., making a state visit to f, signing cooperation or trade agreements with f, or any type of diplomatic concession to f. To capture the idea that each home government $h \in H$ is sovereign but constrained by foreign influence, we assume that foreign policies are determined through a bargaining protocol characterized by the following parameters: the influences $w_{Fh}, w_{Gh} \geq 0$ that the two great powers have on country h and country h's ability to withstand foreign influence $w_h \geq 0$. From an applied perspective, $\frac{w_{fh}}{w_{Fh}+w_{Gh}+w_h}$ represents f's ability to obtain a favorable outcome if it attempts to coerce h, it can thus be thought of as an index summarizing f's ability to destabilize h, e.g., h's economic or financial dependence on f.

The bargaining protocol is as follows

- Government h publicly announces its (intended) policies $(\hat{X}_{hF}, \hat{X}_{hG}) \in \mathbb{R}^2_+$
- Each f simultaneously chooses whether to oppose or support h's policies, $r_{fh} \in \{0, 1\}$.
- If f supports h, i.e., if $r_{fh} = 1$, policy $\hat{X}_{hf}$ is implemented.
- If f opposes h, $r_{fh} = 0$,
  - with probability $\frac{w_{fh}+(1-r_{-f,h})w_{-f,h}}{w_{Fh}+w_{Gh}+w_h}$, f makes a "take-it-or-leave-it" offer to h, $\overline{X}^f_{hf} \in \mathbb{R}_+$;
  - with probability $\frac{w_h+r_{-f,h}w_{-f,h}}{w_{Fh}+w_{Gh}+w_h}$, h makes a "take-it-or-leave-it" offer to f, $\overline{X}^h_{hf} \in \mathbb{R}_+$.

At all times, both f and h can break their bilateral relations, resulting in a zero outside option for both. Given the implemented policies $\mathbf{X} = (X_{hF}, X_{hG})$ and the great power responses $\mathbf{r} = (r_{fh}, r_{Gh})$, when relations are not broken, the payoffs of h and f are, respectively,

$$u_h(\mathbf{X}, \mathbf{r}) = v_{hF}(X_{hF}) - (1 - r_{Fh})\delta_{hF} + v_{hG}(X_{hG}) - (1 - r_{Gh})\delta_{hG}$$
$$u_{fh}(\mathbf{X}, \mathbf{r}) = v_{fh}(X_{hf})$$

where $\delta_{hf} > 0$ is the disutility associated with f's opposition, and $v_{hf}(\cdot)$ and $v_{fh}(\cdot)$ are two continuous, and concave functions of the implemented foreign policy X_{hf}. To avoid trivialities, we assume that the foreign goals of the home country h and the great power f are not fully antagonistic: there exists $X_{hf} \in \mathbb{R}_+$ such that $v_{f,h}(X_{hf}) \geq 0$ $v_{h,f}(X_{hf}) \geq 0$ with at least one inequality being strict. To simplify the analysis and ensure the existence

of a solution, we also assume that $v_{hf}(\cdot)$ and $v_{fh}(\cdot)$ are strictly quasi concave and such that $v_{hf}(0) < 0$ and $\lim_{K\to+\infty} v_{fh}(K) < 0$. A conflict of interest between h and f arises whenever they would make two distinct take-it-or-leave-it offers, i.e., whenever $\bar{X}^h_{fh} \neq \bar{X}^f_{fh}$.

2.2 Strategic interdependence and equilibrium multiplicity

To avoid trivialities, we focus on cases where h and f preferences are misaligned: $\bar{X}^h_{hf} \neq \bar{X}^f_{hf}$. Moreover, since there exists a foreign policy that both parties prefer to breaking relations, relations will never be broken. Given the proposed policies $\hat{\mathbf{X}} = (\hat{X}_{hF}, \hat{X}_{hG})$, and the great power responses, $\mathbf{r} = (r_{fh}, r_{-f,h})$, we can write h's expected payoff as $U_h(\hat{\mathbf{X}}, \mathbf{r}) = \sum_{f\in\{F,G\}}[U_{hf}(\hat{\mathbf{X}}, \mathbf{r}) - (1 - r_{fh})\delta_{hf}]$ where $U_{hf}(\hat{\mathbf{X}}, \mathbf{r})$ is

$$r_{fh}v_{hf}(\hat{X}_{hf}) + (1 - r_{fh})\left[\frac{w_h + r_{-f,h}w_{-f,h}}{w_{Fh} + w_{Gh} + w_h}v_{hf}(\bar{X}^h_{hf}) + \frac{w_{fh} + (1 - r_{-f,h})w_{-f,h}}{w_{Fh} + w_{Gh} + w_h}v_{hf}(\bar{X}^f_{hf})\right]$$

Instead, for each f, the expected payoffs from $r_{fh} = 1$ and from $r_{fh} = 0$ are, respectively,

$$U_{fh}(\hat{\mathbf{X}}, 1, r_{-f,h}) = v_{fh}(\hat{X}_{hf})$$
$$U_{fh}(\hat{\mathbf{X}}, 0, r_{-f,h}) = \left[\frac{w_h + r_{-f,h}w_{-f,h}}{w_{Fh} + w_{Gh} + w_h}v_{fh}(\bar{X}^h_{hf}) + \frac{w_{fh} + (1 - r_{-f,h})w_{-f,h}}{w_{Fh} + w_{Gh} + w_h}v_{fh}(\bar{X}^f_{hf})\right]$$

Note that f's expected payoff from supporting h depends on $-f$'s choice only when f opposes h. After $r_{fh} = 0$, the probability of successfully imposing f's preferred foreign policy is higher when h is weakened by the simultaneous opposition of the other great power. Consequently, in the subgame following offers $\hat{\mathbf{X}}$, multiple equilibria may arise depending on the expectations of the great powers. There may exist an equilibrium where both great powers support h, expecting the other to do the same, and another equilibrium where both great powers oppose h, expecting the other to do likewise.

2.3 Equilibrium refinement and geopolitical risk

To discipline the equilibrium multiplicity emerging in our setting, we introduce a sunspot. Specifically, we assume that after $\hat{\mathbf{X}}$ are proposed, both great powers observe a binary public signal $\zeta_h \in \{0, 1\}$, with $Pr(\zeta_h = 0) = p_\zeta$, and coordinate on one equilibrium or the other based on the realization of ζ. This approach provides a straightforward way to capture the *geopolitical risk* faced by the home country h: as we denote with $\zeta = 0$ the "bad" realization of the signal and with $\zeta = 1$ the "good" one, a larger p_ζ corresponds to a higher likelihood that the great powers coordinate on an unfavorable equilibrium. Letting $r_{fh}(\mathbf{X}, \zeta)$ denote

f's strategy for any announced policy $\mathbf{X} = (X_{hF}, X_{hG})$ and any realization $\zeta_h = \zeta$, we focus on p_ζ subgame perfect equilibria (p_ζ-SPE), defined as follows:

Definition 1 *We say that an equilibrium of the subgame following* $\mathbf{X}$ *is sunspot-consistent if* $U_{fh}(\mathbf{X}, \zeta, \zeta) > U_{fh}(\mathbf{X}, 1-\zeta, \zeta)$ *for both* $f = F$ *and* $f = G$ *implies* $r_{Fh}(\mathbf{X}, \zeta) = r_{Gh}(\mathbf{X}, \zeta) = \zeta$. *We call* p_ζ*-SPE any subgame perfect equilibrium of the game* $(\mathbf{X}, \mathbf{r}(\mathbf{X}, \zeta))$ *such that* $\mathbf{r}(\mathbf{X}, \zeta)$ *is sunspot-consistent for every* $\mathbf{X}$.

To understand the role of this solution concept, suppose h considers an offer $\hat{\mathbf{X}}$ that may induce two equilibria in the responses' subgame: one where $r_{Fh} = r_{Gh} = 1$ and another where $r_{Fh} = r_{Gh} = 0$. The payoff that h expects to obtain from offering $\hat{\mathbf{X}}$ depends critically on which equilibrium is selected. A common approach in the mechanism design literature is to focus on principal-preferred or efficient equilibria: the principal, h, evaluates $\hat{\mathbf{X}}$ under the assumption that $r_{Fh} = r_{Gh} = 1$ will obtain in response. Another standard approach is to consider adversarial equilibria: h evaluates $\hat{\mathbf{X}}$ assuming $r_{Fh} = r_{Gh} = 0$, i.e., the worst-case scenario from h's perspective. While we allow for these two extreme approaches with $p_\zeta = 0$ and $p_\zeta = 1$ respectively, in a p_ζ-SPE, h evaluates $\hat{\mathbf{X}}$ expecting it to induce $r_{Fh} = r_{Gh} = 1$ with probability $1-p_\zeta$ and the $r_{Fh} = r_{Gh} = 0$ with probability p_ζ. This enables us to capture, in a stylized way, the gradual impact of strategic uncertainty on h's foreign policy decisions through the parameter $p_\zeta \in [0, 1]$. As we will show, higher values of p_ζ (more geopolitical risk) lead h to prefer "safe policies," averting the opposition of both great powers in every equilibrium of the subgame, while lower values of p_ζ encourage h to choose "risky policies," taking the risk of facing the great powers' opposition after an unfavorable realization of signal ζ.

2.4 Discussion of the assumptions

Bilateral foreign policies. A crucial aspect of our model is that we allow for bilateral foreign policies: every $h \in H$ can cultivate its relation with both F and G, without necessarily picking sides. This is the empirically relevant case for the modern era, as most countries today, and even during the Cold War, have relations with both great powers. From a theoretical perspective, a country's ability to have relations with both F and G is linked to the assumption that each f derives its utility from X_{hf} and not from $X_{h,-f}$. Consider, for example, an alternative specification such as $u_{fh}(\mathbf{X}, \mathbf{r}) = v_{fh}(X_{hf}) - \eta\, v_{-f,h}(X_{h,-f})$ for some $\eta \in [0, 1]$. Such case reduces h's ability to generate value for both F and G, creating an extra force that pushes f towards having relations with only one great power. The extreme case where $\eta = 1$ is effectively the same as restricting h to unilateral foreign policies (either

favoring F or G), a common assumption in the geoeconomics literature. It can be shown that this extension would not change the main insights of the theory or the empirical analysis.

Foreign influence. In our model, foreign influences are parameters determining the great powers' probability of making a take-it-or-leave-it offer to h after rejecting h's policy. This stylized approach captures the idea that f can leverage its foreign influence to destabilize h and force it to the bargaining table on f's terms (policy $\overline{X}^f_{hf}$). Conversely, when f's influence operation fails, it is h that can dictate the terms of the bargain (policy $\overline{X}^h_{hf}$). This modeling of foreign influence aligns with the contest-success-function approach prevalent in political economy, particularly within the conflict and lobbying literature (Garfinkel and Skaperdas, 2007). Indeed, our bargaining protocol can be reinterpreted as a setting where selecting $r_{fh} = 0$ initiates a contest between f and h, with f's probability of success depending on the foreign influence exerted by the great powers opposing h.

Strategic interdependence. A key novelty of our model is the strategic interdependence that emerges between F and G: the choice of one great power depends on what the other does. Indeed, the probability that f gains from opposing h increases when also the other great power, $-f$, opposes h. This assumption is not tied to the idea that f and $-f$ join forces against h: it is sufficient that h is more inclined to make concessions when pressured by two great powers rather than one. An alternative interpretation could be that f's attempt to coerce h is less likely to succeed when h can count on $-f$'s support. From a theoretical perspective, this interdependence emerges in other bargaining and mechanism design settings such as the sale of goods with network effects or how to incentivize a team of complementary workers (Segal, 1999; Winter, 2004).

Bargaining protocol. Relative to the geoeconomics literature, an innovation of our approach is to study a bargaining model where the domestic government $h \in H$ has a proposer advantage relative to the great powers. This captures the idea that internationally recognized governments derive part of their bargaining power from institutions, e.g., the principle of sovereignty, rather than just the balance of coercive capabilities. As it will be clearer later, such proposer advantage is a crucial driver of our main results as it enables each h to play great powers off each other even when they are overwhelmingly weaker than either of them. Besides this feature, our results are robust to many alternative bargaining protocols. For instance, we could introduce the possibility that $r_{-fh} = 0$ destabilizes the government and comes with the risk that h will not be able to deliver the intended $\hat{X}_{hf}$ to any great power. This feature would introduce another form of interdependence between great powers, potentially reinforcing our results.

Payoffs. To fix ideas on the type of payoff specifications allowed by our assumptions, we discuss some compatible examples that are relevant for international relations:

- *Costly Cooperation.* Our setting can capture positive-sum policies whereby the utilities of f and h are both increasing in X_{hf} for some levels of X_{hf}. A simple specification could be $v_{hf}(X_{hf}) = X_{hf} - X_{hf}^{\gamma_f}$ and $v_{hf}(X_{hf}) = X_{hf}$, where $\gamma_f > 1$ might be f-specific. If $\gamma_F < \gamma_G$, increasing the cooperation with F beyond h's optimum is less costly than for G, suggesting that h has more complementarities with F than with G.

- *Extractive Tariffs.* The assumptions also allow for negative-sum policies, the standard approach in the political economy of tariffs. For instance, we could consider $v_{hf}(X_{hf}) = X_{hf}A_{hf}e^{-b_{hf}X_{hf}}$ and $v_{hf}(X_{hf}) = A_{hf}(1 - X_{hf})$, where $A_{hf} > 0$ and $b_{hf} \geq 1$ might be f-specific. In this setting, tariffs are subject to a Laffer curve, as increasing the tariff rate X_{hf} reduces the trade base to $A_{hf}e^{-b_{hf}X_{hf}}$. If $A_{hF} > A_{hG}$, h trades more with F than G, if $b_{hF} > b_{hG}$, then F can more easily redirect trade away from h than G.

- *Ideal Policies.* A particularly simple specification is with $v_{hf}(X_{hf}) = \overline{X}_{hf}^{h} - |X_{hf} - \overline{X}_{hf}^{h}|$ and $v_{fh}(X_{hf}) = \overline{X}_{hf}^{f} - |X_{hf} - \overline{X}_{hf}^{f}|$ where $\overline{X}_{hf}^{h}$ and $\overline{X}_{hf}^{f} \in [0, 2\overline{X}_{hf}^{h}]$ might be f-specific. For this case, we could interpret X_{hf} as h's typical vote regarding resolutions that affects f at the UN General Assembly. We could interpret $|\overline{X}_{hF}^{F} - \overline{X}_{hF}^{h}| < |\overline{X}_{hG}^{G} - \overline{X}_{hG}^{h}|$ as h having a greater ideological affinity with F than with G, as h would more often find itself voting in favor of F compared to what it would do with G.

3 Theoretical Analysis

In this section, we show how h determines its policies by balancing the risk of great power opposition against its foreign policy preferences. Our theoretical analysis proceeds in three key steps. First, we show that only two types of optimal foreign policies exist: one ensuring the support of both great powers after both the good and the bad realization of ζ (*full insurance*) and another where h is supported only when $\zeta = 1$ (*partial insurance*). Second, we show that the optimal full insurance strategy requires h to *align* with one great power, whereas the optimal partial insurance policy does not. Lastly, we introduce the concept of *spheres of influence* showing how the decision between non-alignment and alignment depends on the distribution of foreign influence.

3.1 Foreign Policies and Self-Insurance

Proceeding by backward induction, we start by noting that when $i \in \{f, h\}$ can make a take-it-or-leave-it offer to $-i$, it will choose the foreign policy that maximizes its bilateral

payoff while keeping $-i$ willing to maintain the relation. Formally,[6]

$$\overline{X}^{s}_{hf} = \arg\max_{X} v_{i,-i}(X) \quad \text{subject to} \quad v_{-i,i}(X) \geq 0.$$

Moreover, under our assumptions, with $X(\alpha) = \alpha\overline{X}^{f}_{hf} + (1-\alpha)\overline{X}^{h}_{hf}$, we have that $v_{hf}(X(\alpha))$ is decreasing in $\alpha \in [0,1]$ (strictly so in $\alpha \in (0,1]$), and $v_{fh}(X(\alpha))$ is increasing in $\alpha \in [0,1]$ (strictly so in $\alpha \in [0,1)$). The conflict of interest thus concerns whether the implemented foreign policy is closer to h's optimum, $\overline{X}^{h}_{hf}$, or f's optimum, $\overline{X}^{f}_{hf}$.

Following this logic, and to simplify the exposition, we define f's (geopolitical) rent as

$$\tilde{v}_{fh}(X_{fh}) := \frac{v_{fh}(X_{fh}) - v_{fh}(\overline{X}^{h}_{fh})}{v_{fh}(\overline{X}^{f}_{fh}) - v_{fh}(\overline{X}^{h}_{fh})}.$$

This measure is particularly convenient since, regardless of whether $\overline{X}^{f}_{fh} > \overline{X}^{h}_{fh}$ or vice versa, higher values of $\tilde{v}_{fh}(X_{fh})$ correspond to a lower payoff for h and a higher payoff for f. Consequently, h's foreign policies offers will aim to minimize the rent earned by each great power f, while ensuring the desired level of support from the great powers.

Note that foreign policies $\mathbf{X}$ may induce two "types" of support from each f:

- **Unconditional support:** X_{fh} induces f to support h regardless of $-f$'s response if $U_{fh}(\mathbf{X}, 1, r_{-f,h}) > U_{fh}(\mathbf{X}, 0, r_{-f,h})$ for all $r_{-f,h} \in \{0,1\}$, which simplifies to:

$$\tilde{v}_{fh}(X_{fh}) \geq \frac{w_{fh} + w_{-f,h}}{w_h + w_{fh} + w_{-f,h}}. \tag{1}$$

- **Conditional support:** X_{fh} induces f to support h *conditionally* on $-f$'s support if $U_{fh}(\mathbf{X}, 1, 1) > U_{fh}(\mathbf{X}, 0, 1)$, which simplifies to:

$$\tilde{v}_{fh}(X_{fh}) \geq \frac{w_{fh}}{w_h + w_{fh} + w_{-f,h}}. \tag{2}$$

In a p_ζ-SPE, a foreign policy X_{fh} that ensures unconditional support guarantees f's support for any $X_{-f,h}$ and realization of ζ. In contrast, a foreign policy X_{fh} that provides only conditional support may or may not result in f's equilibrium support, depending on both $X_{-f,h}$ and ζ. Specifically, in the subgame following $(\mathbf{X}, \zeta = 0)$, f supports h only if $X_{h,-f}$ induces the unconditional support of $-f$. In the subgame following $(\mathbf{X}, \zeta = 1)$, f supports h even if $X_{h,-f}$ induces only the conditional support of $-f$. As a result, the great powers' equilibrium responses $\hat{\mathbf{r}}(\mathbf{X}, \zeta) = (\hat{r}_{Fh}(\mathbf{X}, \zeta), \hat{r}_{Gh}(\mathbf{X}, \zeta))$ depend on $\mathbf{X}$ and ζ in a straightforward way:

[6] Under our assumptions, i's maximization problem is well defined and admits a unique solution.

- $\hat{r}_f(\mathbf{X}, 0) = 1$ if and only if $\mathbf{X}$ is such that either (i) X_{hf} satisfies condition 1, regardless of $X_{h,-f}$ or (ii.a) X_{hf} satisfies condition 2 and $X_{h,-f}$ satisfies condition 1.

- $\hat{r}_f(\mathbf{X}, 1) = 1$ if and only if $\mathbf{X}$ is such that either (i) X_{hf} satisfies condition 1, regardless of $X_{h,-f}$ or (ii.b) both X_{hf} and X_{hf} satisfy condition 2.

From the perspective of the home government, inducing f's support is costly, as it requires distorting its foreign policy away from h's optimum, $\overline{X}^h_{hf}$, and towards f's optimum, $\overline{X}^f_{hf}$. Inducing unconditional support is even more expensive. This naturally raises the question of whether h would ever seek f's support at all. In general, since opposition leads to a disutility, $\delta_{hf} > 0$, inducing $r_{fh} = 0$ results in an inefficient bargaining outcome. A superficial view of Coase's theorem might then suggest that h will always want to induce participation. Our next result shows that this is only partially true in our model: while h's offer ensures the support of both great powers after the good realization of the signal, $\zeta = 1$, this might not hold when $\zeta = 0$.

Theorem 1 *In every p_ζ-SPE, h's offer vector $\mathbf{X}$ satisfies one of the following:*

- ***Full Insurance (FI):*** $\mathbf{X}$ *ensures the support of both F and G irrespective of the realization of ζ, i.e., $\hat{\mathbf{r}}(\hat{\mathbf{X}}^{FI}, \zeta) = \mathbf{1}$ for all $\zeta \in \{0, 1\}$.*

- ***Partial Insurance (PI):*** $\mathbf{X}$ *ensures the support of F and G only after the "good" realization of the signal ($\zeta = 1$), i.e., $\hat{\mathbf{r}}(\hat{\mathbf{X}}^{PI}, 1) = \mathbf{1}$ but $\hat{\mathbf{r}}(\hat{\mathbf{X}}^{PI}, 0) = \mathbf{0}$.*

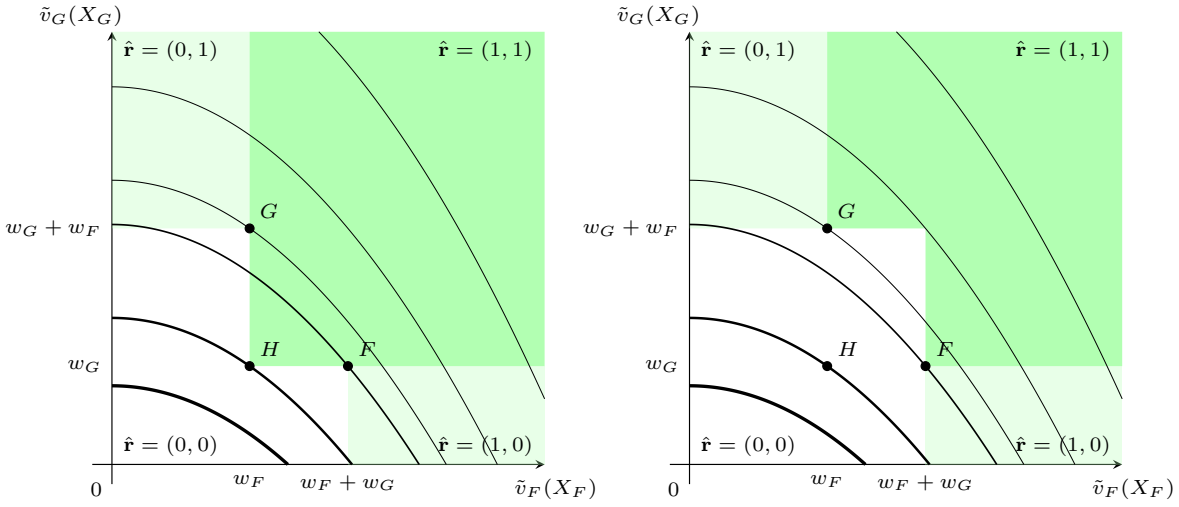
Intuitively, the inefficiency associated with $r_{fh} = 0$ implies that, in the absence of uncertainty about ζ ($p_\zeta \in \{0, 1\}$), h prefers to avert the opposition of the great powers. However, since unconditional support is significantly more expensive than conditional support, ensuring the great powers' support in equilibrium is much costlier when $p_\zeta = 1$ than when $p_\zeta = 0$. Thus, when geopolitical risk $p_\zeta \in (0, 1)$ is sufficiently low, it might be optimal for h to inducing only the conditional support of F and G, accepting the risk of triggering their opposition after an unfavorable realization of the signal, $\zeta = 0$.

Theorem 1 provides a powerful tool for characterizing every p_ζ-SPE: it suffices to find the optimal policies that implement full and partial insurance, respectively denoted by $\hat{\mathbf{X}}^{FI}$ and $\hat{\mathbf{X}}^{PI}$, and then compare which one yields the highest expected payoff for h. This is particularly convenient, as finding $\hat{\mathbf{X}}^{FI}$ and $\hat{\mathbf{X}}^{PI}$ is straightforward:

$$\hat{\mathbf{X}}^{PI} = \arg\max_{\mathbf{X}} U_h(\mathbf{X}, \hat{\mathbf{r}}(\mathbf{X}, 1) = \mathbf{1}), \qquad \hat{\mathbf{X}}^{FI} = \arg\max_{\mathbf{X}} U_h(\mathbf{X}, \hat{\mathbf{r}}(\mathbf{X}, 0) = \mathbf{1})$$

3.1.1 Graphical Analysis

Figure 1 provides a side-by-side representation of the two optimization problems leading to the FI and PI strategy, for a given h such that $w_{Fh} > w_{Gh}$ and $A_{Fh} = A_{Gh}$. The shaded areas represent the great powers' responses for any $(\tilde{v}_{Fh}(X_{hF}), \tilde{v}_{Gh}(X_{hG})) \in [0,1]^2$, thus any relevant $\mathbf{X} \in \mathbb{R}^2_+$. Darker colors represent more support; Panel (A) is for $\hat{\mathbf{r}}(\mathbf{X}, 1)$ and Panel (B) is for $\hat{\mathbf{r}}(\mathbf{X}, 0)$. The black lines represent h's indifference curves for $\mathbf{r} = \mathbf{1}$, i.e., $\{\mathbf{X} \in \mathbb{R}^2_+ : U_h(\mathbf{X}, 1, 1) = u\}$ for different values of $u \geq 0$. Thicker curves correspond to higher values of h's utility. From the analysis of the indifference curves, we see that the thickest indifference curve that lies in each dark green region passes through point H for Panel (A) and through point F for Panel (B). So, H is the optimal PI strategy and F is the optimal FI strategy: the p_ζ-SPE can be found by comparing h's payoff at F with its expected payoff at H.



Panel (A): $\zeta = 1$ Panel (B): $\zeta = 0$

Figure 1: Graphical representation of the optimization problem

3.2 Alignment as full insurance

We now focus on characterizing the optimal fully insuring (FI) strategy $\hat{\mathbf{X}}^{FI}$ for each h, i.e., the solution to $\max_{\mathbf{X}} U_h(\mathbf{X}, \hat{\mathbf{r}}(\mathbf{X}, 0) = \mathbf{1})$, represented graphically by Figure 1, Panel (B).

From before, we know that $\hat{r}_f(\mathbf{X}, 0) = \mathbf{1}$ if and only if X_{hf} satisfies condition 1, or X_{hf} satisfies condition 2 and $X_{h,-f}$ satisfies condition 1. Given that our assumptions on payoffs imply that h wants to minimize rents, $\tilde{v}_{fh}(X_{hf})$ and $\tilde{v}_{-f,h}(X_{hf})$, we conclude:

Lemma 1 *The optimal fully insuring strategy satisfies one of the following:*

- *F-alignment:* $\hat{\mathbf{X}}^F$ *such that* $\tilde{v}_{Fh}(X^F_{hF}) = \frac{w_{Fh}+w_{Gh}}{w_h+w_{Fh}+w_{Gh}}$ *and* $\tilde{v}_{Gh}(X^F_{hG}) = \frac{w_{Gh}}{w_h+w_{Fh}+w_{Gh}}$
- *G-alignment:* $\hat{\mathbf{X}}^G$ *such that* $\tilde{v}_{Fh}(X^G_{hF}) = \frac{w_{Fh}}{w_h+w_{Fh}+w_{Gh}}$ *and* $\tilde{v}_{Gh}(X^G_{hG}) = \frac{w_{Gh}+w_{Fh}}{w_h+w_{Fh}+w_{Gh}}$.

An immediate consequence of the Lemma is that any FI strategy $\hat{\mathbf{X}}^{FI}$ will never have $\tilde{v}_{Fh}(X^{FI}_{hF}) = \tilde{v}_{Gh}(X^{FI}_{hG})$, even if F and G were identical in all dimensions. Thus, any h that opts for a fully insuring strategy will adopt policies that are skewed in favor of one or the other great power, as full insurance requires h to offer a foreign policy that obtains the unconditional support of either F or G. A second observation is that h can determine its optimal $\hat{\mathbf{X}}^{FI}$ by simply comparing the value of the F-alignment and the G-alignment policies: a choice between F and G. In this sense, fully insuring policies capture the concept of *geopolitical alignment*, the observation that countries tend to "choose" between favoring one or the other great power. Our model explains why we should expect to see such pattern even in a world where countries are exposed to both great powers and can cultivate their relations with both of them.

Role of foreign influence. In terms of Figure 1, Lemma 1 is telling us that, for every h, the optimal FI strategy will either correspond to point F (F-alignment) or point G (G-alignment). In the figure, we see that h prefers F-alignment to G-alignment. As the next result will show, this feature depends on the fact that we depicted a case where $w_{Fh} > w_{Gh}$. To see the role of foreign influence, we must build some notation. First, note that, by definition of $\tilde{v}_{fh}(\cdot)$, the policy X_{hf} such that $\tilde{v}_{fh}(X_{hf}) = w$ can be written as $\tilde{v}^{-1}_{fh}(w) = v^{-1}_{fh}(wv_{fh}(\overline{X}^f_{hf}) + (1-w)v_{fh}(\overline{X}^h_{hf}))$ which is strictly increasing in w, as an increase in w moves the policy away from $\overline{X}^h_{hf}$ and closer to $\overline{X}^f_{hf}$. While this is beneficial for f, it is costly for h, and indeed h's value for that policy, $v_{hf}(\tilde{v}^{-1}_{fh}(w))$, is strictly decreasing in w. To understand when f-alignment is preferred to $-f$-alignment, note that the cost of shifting from $-f$ to f is that h must increase the offer that it makes to f from $\tilde{v}^{-1}_{fh}(\frac{w_{fh}}{w_h+w_{fh}+w_{-f,h}})$ to the more costly policy $\tilde{v}^{-1}_{fh}(\frac{w_{fh}+w_{-f,h}}{w_h+w_{fh}+w_{-f,h}})$. This has the following impact on h's payoffs

$$C_{hf}(w_{-f,h}) := \underbrace{v_{hf}\left(\tilde{v}^{-1}_{fh}\left(\frac{w_{fh}}{w_h+w_{fh}+w_{-f,h}}\right)\right) - v_{hf}\left(\tilde{v}^{-1}_{fh}\left(\frac{w_{fh}+w_{-f,h}}{w_h+w_{fh}+w_{-f,h}}\right)\right)}_{\text{increasing in } w_{-f,h}}$$

But the benefit of shifting from $-f$ to f is that h can reduce the cost of the policy it offers to $-f$. This benefit is exactly $C_{h,-f}(w_{fh})$. Thus, h's preference between F-alignment and

G-alignment is as follows

$$\mathbf{X}^F \succ_h \mathbf{X}^G \quad \text{if and only if} \quad C_{h,-f}(w_{fh}) - C_{hf}(w_{-f,h}) > 0 \tag{3}$$

Given that the benefit, $C_{hf}(w_{-f,h})$, is increasing in $w_{f,h}$ and the cost, $C_{h,-f}(w_{fh})$, is increasing in $w_{-f,h}$, the incentive to choose f-alignment is increasing in the balance of power $\frac{w_{fh}}{w_{fh}+w_{-f,h}}$. Indeed, we can establish the following result:

Theorem 2 *Given h's overall exposure to the great powers, $w_{fh}+w_{-f,h}$, there is a threshold $\bar{W} \in [0,1]$ such that h prefers f-alignment to $-f$-alignment if and only if $\frac{w_{fh}}{w_{fh}+w_{-f,h}} \geq \bar{W}$. In particular, if $v_{hf}(\cdot) = v_{h,-f}(\cdot)$ and $v_{fh}(\cdot) = v_{-f,h}(\cdot)$, $\bar{W} = \frac{1}{2}$.*

This result points to the importance of foreign influence in determining alignments. The key element to keep in mind is that when h aligns with f it must guarantee f's unconditional support: f's support even when $-f$ opposes h. The greater h's exposure to $-f$, the more costly it will be to induce f's support under $-f$'s opposition. Thus, the cost of aligning with f is increasing in $w_{-f,h}$, making it cheaper to aligning with the strongest great power.

Role of preferences. As we can see from $C_{hf}(\cdot)$, alignment decisions will also depend on the preferences of h and each f. Specifically, h will be more interested in aligning with F rather than G if (i) h is hurt less from a policy shift towards $\bar{X}^F_{hF}$ rather than towards $\bar{X}^G_{hG}$, and if (ii) F gains more from such a movement than G. As a result, we should expect alignments to reflect whether h's preferences are closer to F rather than G, as well as h's relative exposure to their foreign influence. For example, a payoff specification that delivers a sharp characterization is $v_{fh}(X_{hf}) = \bar{X}^f_{hf} - |X_{hf} - \bar{X}^f_{hf}|$ and $v_{hf}(X_{hf}) = \bar{X}^h_{hf} - |X_{hf} - \bar{X}^h_{hf}|$. In this case, country h prefers to align with F rather than G if and only if

$$\frac{w_F}{|\bar{X}^h_{hF} - \bar{X}^F_{hF}|} \geq \frac{w_G}{|\bar{X}^h_{hG} - \bar{X}^G_{hG}|}$$

This case emphasizes how similar policy preferences (smaller denominators) act as a multiplier for foreign influence, leading countries to potentially align with the weaker great power if they find their preferences sufficiently more similar.

3.3 Non-Alignment as partial insurance

In this section we characterize the optimal partially insuring (PI) strategy and study when h should prefer it to a fully insuring (FI) one.

The optimal PI strategy solves $\max_{\mathbf{X}} U_h(\mathbf{X}, \hat{\mathbf{r}}(\mathbf{X}, 1) = \mathbf{1})$, represented in Figure 1, Panel (A). Recall that $\hat{r}_f(\mathbf{X}, 0) = \mathbf{1}$ if and only if X_{hf} satisfies condition 1, or X_{hf} and $X_{h,-f}$

satisfy 2. Given that conditional support (condition 2) is cheaper than unconditional support (condition 1), we have:

Lemma 2 *The optimal partially insuring strategy satisfies Non-Alignment, i.e.,* $\hat{\mathbf{X}}^{PI}$ *is such that* $\tilde{v}_{fh}(\hat{X}_f^{PI}) = \frac{w_{fh}}{w_h+w_{fh}+w_{-f,h}}$ *for both* f.

Note that this does not imply that h offers the same policies to F and G. First, we see that, if $w_{Fh} > w_{Gh}$, h's policies will offer a higher rent to F than with G. Still, we use the term *non-alignment* because, in contrast to the FI strategies of the previous section, if the two great powers were identical, h would treat them in the same way.

A comparison between Lemma 1 and 2 clarifies that when $\zeta = 1$, non-alignment is less costly for country h than alignment. However, when $\zeta = 0$, non-alignment leads to an inefficiency, as h would lose the support of both great powers. So, the decision between alignment and non-alignment will trade-off the gains from non-alignment obtained when $\zeta = 1$ and the cost of losing support when $\zeta = 0$. To measure such inefficiency, we let $\pi_{hf}(X_{hf}) := v_{hf}(X_{hf}) - \left[\frac{w_h}{w_h+w_{fh}+w_{-f,h}} v_{hf}(\bar{X}^h_{hf}) + \frac{w_{fh}+w_{-f,h}}{w_h+w_{fh}+w_{-f,h}} v_{hf}(\bar{X}^f_{hf}) - \delta_{hf}\right]$. This function captures the gains that h derives from its relation with f under policy X_{hf} when both great powers support compared to when both oppose. With this, we define:

$$\Pi_h(w_h) = \pi_{hf}\left(\tilde{v}_{fh}^{-1}\left(\frac{w_{fh}+w_{-f,h}}{w_h+w_{fh}+w_{-f,h}}\right)\right) + \pi_{h,-f}\left(\tilde{v}_{-f,h}^{-1}\left(\frac{w_{fh}+w_{-f,h}}{w_h+w_{fh}+w_{-f,h}}\right)\right)$$

Using this expression, we can show that country h's preference between non-alignment, $\mathbf{X}^{PI}$, and f-alignment, $\mathbf{X}^f$, is such that

$$\mathbf{X}^{PI} \succeq_h \mathbf{X}^f \quad \text{if and only if} \quad \frac{C_{hf}(w_{-f,h})}{\Pi_h(w_h) + C_{h,-f}(w_{fh})} \geq \frac{p_\zeta}{1-p_\zeta} \tag{4}$$

As we would expect, when geopolitical risk is low (small p_ζ), h is more likely to find non-alignment more appealing than alignment, as the risks associated with non-alignment are negligible. Given that $C_{hf}(w_{-f,h}) > 0$ for every $\mathbf{w} > 0$, the left-hand side is strictly greater than zero. Thus, when p_ζ is small enough, every country will prefer non-alignment to alignment, implying that in times of very low geopolitical risk we should not observe any alignment. On the other hand, recall from condition 3 that h prefers F-alignment over G-alignment if $C_{hF}(w_{Gh}) > C_{hG}(w_{Fh})$. We thus see that for every given w_h, the incentive to choose non-alignment over any kind of alignment is greatest when $C_{hF}(w_{Gh}) = C_{hG}(w_{Fh}) = C$. But then we conclude that the left-hand side of condition 4 is bounded above by 1, as $\Pi(w_h) > 0$, implying that, for high enough geopolitical risk ($p_z \geq \frac{1}{2}$), every home country will eventually resort to aligning with either F or G.

Theorem 3 *For every country h there is a a level of geopolitical risk $\overline{p}_\zeta^h \in (0, \frac{1}{2})$ such that h prefers alignment over non-alignment if and only if $p_\zeta \geq \overline{p}_\zeta^h$.*

This result predicts that in times of increasing geopolitical risk, we should expect the world to become increasingly polarized along two alignment blocs. Specifically, countries with a clear preference in favor of one great power (a larger $|C_{hf}(w_{-f,h}) - C_{h,-f}(w_{fh})|$) will tend to choose alignment even for relatively low levels of geopolitical risk. Instead, countries who are caught "in the middle" of the great powers – with similar preferences towards them and a similar exposure to their foreign influence – should be expected to exhibit some hesitancy in picking sides between them. Eventually, however, as geopolitical risk is large enough, our result suggests that more and more countries will enter the orbit of one or the other great power.

3.4 Spheres of Influence and Geoeconomic Fragmentation

We conclude our theoretical analysis introducing the concept of spheres of influence and establishing the link between geoeconomic fragmentation and the distribution of power between the two great powers.

From the previous discussion, we know that we can collect every home government $h \in H$ into one of the following mutually exclusive groups:

- **Spheres of influence:** for each $f \in \{F, G\}$, f's sphere of influence is the set of home countries choosing f-alignment in every p_ζ-SPE, i.e., $S_f := \{h \in H : \hat{\mathbf{X}}_h = \hat{\mathbf{X}}_h^f\}$.
- **Non-aligned:** the non-aligned group is the set of home countries choosing non-alignment in every p_ζ-SPE, i.e., $N := \{h \in H : \hat{\mathbf{X}}_h = \hat{\mathbf{X}}_h^{PI}\}$

Every h will be in one group or the other depending on its parameters, in particular its exposure to the great powers' foreign influence. From condition 3 we see that when we increase the balance of power $\frac{w_{fh}}{w_{fh}+w_{-f,h}}$ in favor of f, countries should become more willing to align with f than $-f$. Instead, from condition 4 we see that when $\frac{w_{fh}}{w_{fh}+w_{-f,h}}$ moves closer to $\frac{1}{2}$, countries should be more willing to choose non-alignment over alignment. We thus conclude that, for fixed w_h, the balance of power $\frac{w_{fh}}{w_{fh}+w_{-f,h}}$ determines whether h will be part of the non-aligned group or one of the two spheres of influence:

Proposition 1 *For every h, there are thresholds $\overline{W}_h^G \leq \overline{W}_h^F$ such that*

- *Country h is in the non-aligned group if its exposure to the great powers is sufficiently balanced, i.e., for given $w_{Fh} + w_{Gh}$, $h \in N$ if $\frac{w_{Fh}}{w_{Fh}+w_{Gh}} \in (\overline{W}_h^G, \overline{W}_h^F)$.*

- *Country h is in f's sphere of influence if it is sufficiently more exposed to f than $-f$. In particular, given $w_{Fh}+w_{Gh}$, $h \in S_G$ if $\frac{w_{Fh}}{w_{Fh}+w_{Gh}} < \bar{W}_h^G$, and $h \in S_F$ if $\frac{w_{Fh}}{w_{Fh}+w_{Gh}} > \bar{W}_h^F$.*

The thresholds determine the switching points from one group to the other. At such points, h is indifferent and can be in one group or the other depending on the p_ζ-SPE considered. When p_ζ is small enough, then $\bar{W}_h^G < \bar{W}_h^F$ and we can derive the thresholds by equalizing condition 4 for each f. Otherwise, we have $\bar{W}_h^G = \bar{W}_h^F$, the non-aligned group is empty, and we can find the unique threshold by equalizing condition 3. The result tells us that when the balance of power swings in favor of country F, then we should expect F's sphere of influence to expand (attracting countries from G's sphere of influence, from the non-aligned group, or both). In this sense, we find that great powers "compete" over spheres of influence, even if our model allows countries to foster their relations with both great powers.

Geoeconomic fragmentation. To conclude the theoretical analysis, we explore the link between geoeconomic fragmentation and the great powers' relative influences. The idea of fragmentation we want to capture is that the great powers obtain different foreign policy rents from different countries in the international system. For example, the Cold War was a time of large fragmentation from the perspective of the United States because some countries adopted policies very favorable towards it, whereas others did not. Similarly for the Soviet Union. With this in mind, we want to measure geoeconomic fragmentation as the extent of variability in the foreign policy rents that a given great power is expected to receive. For concreteness, we calculate the level of geoeconomic fragmentation as the sum of the Gini coefficients of the distribution of foreign policy rents received by F and G:

$$\sum_{f\in\{F,G\}} \mathcal{G}_f(\tilde{\mathbf{v}}_f) := \sum_{f\in\{F,G\}} \frac{1}{2|H|^2} \sum_{h\in H}\sum_{h'\in H} |\tilde{v}_{hf}(X_{hf}) - \tilde{v}_{h'f}(X_{hf})|$$

Because great powers have different foreign influences over different countries, it should not be surprising that they also receive heterogeneous foreign policy rents. This implies that an heterogeneous distribution of foreign influence will trivially lead to some geoeconomic fragmentation. The next result, however, shows that in our model geoeconomic fragmentation emerges even when the great powers have the same foreign influence on any country as long as not all countries remain non-aligned and they are not all in the same sphere of influence.

Proposition 2 *Suppose that $w_h = w$, $w_{fh} = w_f$ and $w_{-f,h} = w_{-f}$ for every $h \in H$, the level of geoeconomic fragmentation is strictly positive if and only if there is f such that $S_f \neq \emptyset$*

and $S_f \neq H$

$$\sum_{f\in\{F,G\}} \mathcal{G}_f(\tilde{\mathbf{v}}_f) = \sum_{f\in\{F,G\}} \frac{|S_f||H\setminus S_f|}{|H|^2}\frac{w_{-f}}{w+w_f+w_{-f}} > 0$$

Moreover, if there is no preference heterogeneity, i.e., $v_{Fh}(\cdot) = v_{Gh}(\cdot)$ *and* $v_{hF}(\cdot) = v_{hG}(\cdot)$*, then the level of geoeconomic fragmentation is maximized at* $\frac{|S_F|}{|H|} = \frac{|S_G|}{|H|} = \frac{1}{2}$ *and* $\frac{w_F}{w_F+w_G} = \frac{1}{2}$*.*

The proposition highlights that there are two sources of geoeconomic fragmentation that are specific to our model: (i) the relative sizes of the two spheres of influence and (ii) the level of influence that each great power has on the competitor's sphere of influence. The former force has straightforward implications: when more countries opt for an alignment rather than a non-alignment strategy, then the level of fragmentation increases. For the latter, we find that when great power G becomes stronger (for fixed S_F and S_G), the distribution of foreign policy rents received by F becomes more uneven, whereas the foreign policy rents received by G become less uneven. When $|S_F| > |S_G|$, the former effect prevails and geoeconomic fragmentation increases. On the other hand, the latter effect might prevail when F's sphere of influence is sufficiently larger than G's.

Clearly, these forces cannot be separated as a change in the balance of power should also affect S_F and S_G. To build intuition on the interaction between these forces, the last part of the proposition considers the simple case where $v_{Fh} = v_{Gh}$ and $v_{hF} = v_{hG}$, so that $h \notin S_f$ if $w_{-f} > w_f$. In this case, we can see that the extent of fragmentation is maximized when all countries are in either one or the other sphere of influence and the two great powers have the same influence everywhere. All in all, this proposition supports the idea that when the distribution of power becomes less (more) unipolar then geoeconomic fragmentation should increase (decrease). However, the result also emphasizes the key role of geopolitical risk: a change in geopolitical risk affects the sizes of $|S_F|$ and $|S_G|$ and thus the extent of geoeconomic fragmentation.

4 Empirical Analysis

The main prediction of our model is that the geopolitical rent $\tilde{v}_{fh}(X_{fh})$, which a domestic country h grants to a great power f through its equilibrium policy X_{fh}, depends on h's exposure to *all* great powers and whether h is in f's sphere of influence S_f:

$$\tilde{v}_{fh}(X_{fh}) = \frac{w_{fh}}{w_h + w_{fh} + w_{-f,h}} + \mathbf{1}\{h \in S_f\} \times \frac{w_{-f,h}}{w_h + w_{fh} + w_{-f,h}} \tag{5}$$

The first term represents the 'classical' geopolitical rent that a great power f obtains from all countries due to its relative foreign influence, $\frac{w_{fh}}{w_h+w_{fh}+w_{-f,h}}$. The second term is instead novel and represents an additional rent that f obtains from countries within its sphere of influence ($h \in S_f$). This second component increases with the relative foreign influence of the other great power, $\frac{w_{-f,h}}{w_h+w_{fh}+w_{-f,h}}$, reflecting that countries in f's sphere must further compensate f to maintain its unconditional support when $-f$ poses a greater threat to their internal stability (*balancing effect*). This balancing effect is what sets our model apart from simpler alternative models, and it is the driver of our results on geoeconomic fragmentation as it induces a non-monotonicity in how a great power's geopolitical rents should respond to changes in the balance of power. For example, we see that the weakening of the Soviet Union (a reduction in $w_{-f,h}$) should negatively impact the United States within its sphere of influence but positively outside of it. Conversely for the rise of China (an increase in $w_{-f,h}$).

4.1 Case studies

We begin our empirical analysis summarizing the anecdotal evidence discussing how countries changed their bilateral relations with the United States (i) after the collapse of the Soviet Union, and (ii) during China's rise. These examples highlight the non-monotonic patterns captured by Equation 5.

Collapse of the Soviet Union. The end of the Cold War marked a significant shift in global power dynamics, prompting countries to recalibrate their foreign policies toward the United States. Many countries who previously belonged to the Eastern bloc pivoted toward the US starting a process of democratization and socioeconomic reforms to increase ties with the West. As early as 1992, Poland, Hungary, and Czechoslovakia signed a joint declaration announcing their intention to join NATO. In 1999, Poland, Hungary, and Czechia were officially admitted, and ten more Eastern European countries had already initiated the necessary reforms to be eligible for accession (NATO, 2024). In contrast to this, during the same period, traditional allies of the United States began asserting greater foreign policy autonomy. The acceleration in the process of European integration of the 1990s, especially in the security domain, is often viewed as an attempt from France and Germany to gain more strategic autonomy from the United States after the Cold War (Art, 2004; Posen, 2006). This ambition emerged clearly in 2003, when Germany and France strongly opposed the US-led invasion of Iraq (Paul, 2005; Pape, 2005). Similarly, the end of the Cold War coincided with the rise of leftist governments in Latin America adopting explicitly anti-US stances. This shift led to a significant deterioration of US relations in an area of the world that was closely aligned with it during the Cold War (Levitsky, 2011).

Rise of China. The emergence of China as a global power has similarly led to varied responses in countries' bilateral relations with the United States. As it is natural to expect, the rise of China eroded American influence and led many countries to distance themselves from the United States. In Africa, for example, countries who expanded their cooperation with China often decreased their cooperation with the United States (Whitaker, 2010). Another example of this pattern is the relation between the United States and Cambodia, formerly in the Soviet bloc. Cambodia and the US normalized their relation after the Cold War. However, in the last decade Cambodia increased its cooperation with China, while adopting a defiant stance towards the United States that led to sanctions and a suspension from US aid programs (BBC, 2018; Pich and Aun, 2023). In contrast to these patterns, the rise of China had the paradoxical effect of leading many countries closer to the United States. Since the late 2000s, South Korea and Japan increased their alignments and security ties with the US in response to China's rise (Selden, 2013; Wong, 2015; Koga, 2018). Similarly, China's growing assertiveness pushed Australia closer to the United States as exemplified by its growing number of security partnerships with the US such as the AUKUS or the Quadrilateral Security Dialogue (DFAT, 2024; ASA, 2024).

4.2 Data and Measurement

Our model makes new predictions on how a country's bilateral relation with a great power depends on the distribution of foreign influence that it is subject to. Clearly, this is not a new question, and indeed, we can rely on previous work to find measures that we can use to test our predictions.

4.2.1 Geopolitical rents

A variable that is commonly used in the literature to measure the extent of a country's policy bias in favor of the United States is based on its voting behavior at the United Nations General Assembly and how it compares with the voting behavior of the United States. In particular, the standard approach is to use a measure of voting similarity provided by Bailey, Strezhnev and Voeten (2017): the absolute distance between each country's estimated preference (ideal point) and the ideal point estimated for the US. Just as the geopolitical rent that we considered in our model, this distance expresses the result of the trade-off faced by a country between remaining on the good side of the United States and staying close to its intrinsic foreign policy preferences.

4.2.2 Measuring Foreign influence

To measure a great power's foreign influence on a country, the simplest variable that we could consider is trade dependence – the value of a country's exports towards a great power scaled by its own GDP. While this measure is crude, it has the advantage of capturing the idea that becoming more economically exposed to a great power enhances its coercive capabilities. To make sure that our analysis is robust, we also use a more complex but realistic measure of foreign influence – the formal bilateral influence capability index developed by Moyer et al. (2018). This measure is a composite measure of foreign influence which is based on arms transfers, aid dependence, security relations, and more.[7]

4.3 Empirical Strategy

We start the discussion of our empirical strategy from the empirical specification and its interpretation. Letting $|\hat{Y}_{ft} - \hat{Y}_{ht}|$ denote the absolute distance between the ideal point of great power f (the United States) and country h in year t and letting $w_{f',ht}$ measure the log of country h's trade dependence on great power f' in year t. We study the most easily interpretable empirical specification that is compatible with our model's predictions:

$$\ln(|\hat{Y}_{ft} - \hat{Y}_{ht}|) = \alpha w_{fh,t-1} + \beta w_{-f,h,t-1} \times \mathbf{1}\{w_{fh,t-1} > w_{-f,h,t-1}\} + Cntrl_{fht} + \varepsilon_{fht} \qquad (6)$$

Ignoring, for now, any endogeneity concerns, we interpret $\beta < 0$ as evidence in favor of our predicted balancing effect, because (i) countries more exposed to great power f than $-f$, i.e., $w_{fh,t-1} > w_{-f,h,t-1}$, are more likely to be in f's sphere of influence (Theorem 2), and (ii) countries in f's sphere of influence should offer a geopolitical rent to f that compensates it for their exposure to $-f$ (less distance from f's ideal point). Within our model, $\beta = 0$ would imply that we cannot reject the hypothesis that f's sphere of influence is empty. More in general, we can think of this as a test of our model: as all our results are conditional on $w_{fh,t-1}$, finding $\alpha < 0$ and $\beta = 0$ would imply that our predictions cannot be empirically disentangled from the simpler case where geopolitical rents are simply proportional to a great power's foreign influence. Finally, in our baseline specification, we also condition on $w_{-f,h,t-1}$, the log of h's trade with f in year t, h's overall trade-openness in year t, the log of h's population and GDP in year t, as well as a fixed effect for each h and each t. Conditioning on $w_{-f,h,t-1}$ captures the possibility that when countries become more exposed to $-f$ their intrinsic preferences may shift farther from f, which may be especially relevant during times of ideological competition like the Cold War. The presence of year fixed effects absorbs all

[7] This measure can be accessed from https://korbel.du.edu/fbic.

factors that may affect how the ideal point of the United States may have shifted over time. The other variables account for the idea that when countries trade more with the United States, or they become richer and more open to trade (so more similar to the US), then their intrinsic preference may move closer to the US, thus reducing the observed difference between the two ideal points.

4.3.1 Endogeneity and Instrumental Variables

The main concern in Equation 6 is that a country may choose to change its exposure to foreign influence today in anticipation of changes to its future foreign policy preferences. For example, countries who want to reduce their future "alignment" with great power f may strategically reduce their dependence on f. If the policies implemented to reduce dependence on f also reduced the country's dependence on $-f$, then we could find that a reduced exposure to $-f$ is associated with a worsened relation with f (thus $\beta < 0$). If such policies instead increased the country's dependence on $-f$, then β would be biased upwards.

We partially address this concern by conditioning on both $w_{f,h,t-1}$ and $w_{-f,h,t-1}$. However, if such an effect were more likely (or larger) for countries with $w_{f,h,t-1} > w_{-f,h,t-1}$, then we could still find $\beta < 0$ even in the absence of any balancing effect. To deal with this, we must find instrumental variables that are plausibly uncorrelated with a country's policy choices but that affect the distribution of foreign influence. A natural idea is to exploit the global reduction in Russia's foreign influence during and after the collapse of the Soviet Union or China's rise, building instruments of the type $\mathbf{1}\{T_0 \leq t \leq T_1\}$, where T_0 is the starting period and T_1 is the final one.[8] However, we cannot use these variables in a specification with year fixed effects. Fortunately, given that we are interested in the differential impact of $\ln(w_{-f,h,t-1})$ on countries within f's sphere of influence relative to the other countries, we can still use this idea by interacting $\mathbf{1}\{T_0 \leq t \leq T_1\}$ with a measure determining the probability that h is in h's sphere of influence. To minimize endogeneity concerns, our baseline specification will instrument $w_{-f,h,t-1} \times \mathbf{1}\{w_{fh,t-1} > w_{-f,h,t-1}\}$ using $\mathbf{1}\{T_0 \leq t \leq T_1\} \times \mathbf{1}\{w_{f,h,T_0-1} > w_{-f,h,T_0-1}\}$, which only uses the observed distribution of foreign influence before T_0, the beginning of the global shock.

4.4 Empirical Results

Table 1 shows the reduced-form regression results based on Equation 6. The three columns differ in terms of which country is considered as the competitor of the United States: Column

[8] For the case of Russia, our baseline assumes $T_0 = 1990$ and $T_1 = 1994$. Instead, for China, $T_0 = 2007$ and T_1 is the end of the sample.

(1) focuses on Russia, Column (2) focuses on China, and Column (3) assumes that the relevant competitor can be either Russia or China, depending on which one has the greatest amount of foreign influence on country h in year t. The first row of the table reports the coefficient capturing the effect of the foreign influence of the US (h's trade dependence on the US). Unsurprising, this coefficient is negative, meaning that the US tends to receive more geopolitical rents (lower ideal point distance) from countries on which it has more foreign influence. The other coefficients are instead the test of the balancing effect (β in terms of Equation 6). Across all three specifications, we see that the US receives higher geopolitical rents from countries within its sphere of influence when such countries become more exposed to the foreign influence of the competitor of the United States. Looking at the magnitude of the balancing effect we see that it is around 1/3 of the direct effect of US foreign influence.

Table 1: Ideal point distance and trade dependence: reduced form evidence

	$\ln(\lvert\hat{Y}_{USA,t+1} - \hat{Y}_{h,t+1}\rvert)$		
	(1)	(2)	(3)
$w_{USA,t}$	-0.052***	-0.050***	-0.047***
	[0.0186]	[0.015]	[0.016]
$w_{RUS,t} \times \mathbf{1}\{w_{USA,t} > w_{RUS,t}\}$	-0.021*		
	[0.014]		
$w_{CHN,t} \times \mathbf{1}\{w_{USA,t} > w_{CHN,t}\}$		-0.015**	
		[0.006]	
$w_{MAX,t} \times \mathbf{1}\{w_{USA,t} > w_{MAX,t}\}$			-0.013**
			[0.006]
Num. Countries	183	187	187
Observations	6,623	7,089	7,389

Note: the unit of observation is a country h in a given year t. The outcome is the log of the absolute difference in the ideal points of the US and ideal points of country h as measured from their voting behavior at the United Nations General Assembly in year t. Foreign influence $w_{f,t}$ is measured as the log of trade dependence, defined as the value of h's export at time t towards f divided by h's GDP at time t. $w_{MAX,t} := max_{j \in \{RUS,CHN\}}\{w_{j,t}\}$. All estimates are conditional on country h fixed effect, year t fixed effects, foreign influence of the competitor of the US ($w_{MAX,t}$, $w_{RUS,t}$, or $w_{CHN,t}$), the log of h's GDP, the log of h's population, the log of h's trade with the rest of the world, the log of h's GDP per capita. Standard errors clustered at the country level are reported in brackets. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

As discussed in the previous section, however, the coefficients of the reduced form regression are likely biased as countries strategically determine their exposure to foreign influence as a function of their foreign policy preferences, which are only partly accounted for by our conditioning variables. Table 2 reports the regression results from our 2SLS estimation. Just

as in Table 1, the three columns correspond to a different assumption about which country is considered to be the relevant competitor of the US. In this specification, we find that the coefficients associated with the balancing effect are as large as the direct effect of US foreign influence. This observation suggests that the coefficient of the reduced-form regressions are biased downwards, further reinforcing the finding that the balancing effect has a relevant empirical role in determining the countries' foreign policies.

Table 2: Ideal point distance and trade dependence: 2SLS estimation leveraging the collapse of Soviet Union (1990-1994) and China's rise (2007-)

	$\ln(\lvert\hat{Y}_{USA,t+1} - \hat{Y}_{h,t+1}\rvert)$		
	(1)	(2)	(3)
$w_{USA,t}$	-0.17***	-0.15***	-0.16***
	[0.056]	[0.044]	[0.047]
$w_{RUS,t} \times \mathbf{1}\{w_{USA,t} > w_{RUS,t}\}$	-0.27**		
	[0.107]		
$w_{CHN,t} \times \mathbf{1}\{w_{USA,t} > w_{CHN,t}\}$		-0.13***	
		[0.043]	
$w_{MAX,t} \times \mathbf{1}\{w_{USA,t} > w_{MAX,t}\}$			-0.15***
			[0.054]
Num. Countries	155	186	157
Observations	5,955	7,086	6,835
Kleibergen-Paap F statistics	15.23	25.55	20.83
Num. excluded instruments	1	1	2
Hansen J test			4.251**

Note: the unit of observation is a country h in a given year t. The outcome is the log of the absolute difference in the ideal points of the US and ideal points of country h as measured from their voting behavior at the United Nations General Assembly in year t. Foreign influence $w_{f,t}$ is measured as the log of trade dependence, defined as the value of h's export at time t towards f divided by h's GDP at time t. $w_{MAX,t} := max_{j \in \{RUS,CHN\}}\{w_{j,t}\}$. Column (1) is based on instrument $\mathbf{1}\{1990 \leq t \leq 1994\} \times \mathbf{1}\{w_{USA,1989} > w_{RUS,1989}\}$, column (2) on $\mathbf{1}\{t \geq 2007\} \times \mathbf{1}\{w_{USA,2006} > w_{CHN,2006}\}$, column (3) is based on both. All estimates are conditional on country h fixed effect, year t fixed effects, foreign influence of the competitor of the US ($w_{MAX,t}$, $w_{RUS,t}$, or $w_{CHN,t}$), the log of h's GDP, the log of h's population, the log of h's trade with the rest of the world, the log of h's GDP per capita. Standard errors clustered at the country level are reported in brackets. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

4.5 Robustness checks

We include all our robustness checks in Appendix B. The results are remarkably robust to a variety of deviations from the specification used for Table 2. In Sections B.1 we confirm our findings by using an alternative specification, in Section B.4 we do so by considering alternative measures of foreign influence, and we study alternative outcome variables in Section B.5. In Sections B.2 we include or omit alternative controls and fixed effects, and in Section B.3 we consider alternative instrumental variables. Across all specifications, we find that our estimates for the balancing effect remain close to the coefficients displayed in Table 2 and to the direct effect of US foreign influence. As we show, the results hold even when we exclude the foreign influence of the competitor of the US from the controls. This implies that when countries in the US sphere of influence become more exposed to US rivals, they increase their alignment with the US not only relative to other countries but also in absolute terms. Another interesting finding emerges when excluding year fixed effects: we find that in the period 1990-1994 countries have on average distanced themselves from the ideal point of the United States, whereas in the post-2007 period, they have on average moved closer to the US. This can be explained within our model by noting that in 1989 the United States' sphere of influence was larger than the Soviet sphere, and in 2006, the US had a larger sphere of influence than China. Finally, our results are also robust to alternative specifications of the instrumental variables based on different time periods. For the case of Russia, we consider an instrument based on the period 1989-1999 and find even larger coefficients. Instead, for the case of China, our results weaken when considering instruments based on the post-2001 rather than the post-2007 period. This might reflect the fact that it took a few years before China's rise started emerging as a force capable of shaping the alignment decisions of countries.

5 Conclusion

This paper investigates the strategic foundations of geopolitical fragmentation, analyzing why countries align with one great power rather than remaining non-aligned, even in a setting where they are deeply integrated with both great powers and great powers cannot directly force countries to pick sides.

Our game-theoretic model reveals a central trade-off: aligning with one great power guarantees the support of that great power but requires adopting policies that are very costly, i.e., farther from the policy that the country would autonomously implement. In contrast, non-alignment allows countries to save on the costs of alignments but exposes them to the

risk of losing support entirely under adverse geopolitical conditions. Thus, geopolitical risk emerges as a first driver of geopolitical fragmentation: the greater the likelihood of coercive actions by great powers, the more non-alignment becomes a risky choice and alignment a comparatively safer strategy. A second driver of geopolitical fragmentation is the balance of power between great powers. We show that when the distribution of power becomes less unipolar (as with China's rise) the countries aligning with one or the other great power should be expected to adopt increasingly diverging foreign policies. The key reason for this is a central prediction of our model, the *balancing effect*: when a country chooses to align with one great power it must compensate it for its exposure to the other great power.

Given the central importance of this balancing effect in our model, we test it empirically using UN General Assembly voting data and measures of foreign influence, leveraging the Soviet Union's collapse and China's WTO accession as exogenous shocks. The results confirm the role of this force in shaping countries' observed foreign policies, underscoring the empirical relevance of our theoretical framework and its predictions.

In a world that is becoming increasingly less unipolar and characterized by rising geopolitical tensions, our analysis suggests that we should expect geopolitical fragmentation to increase rather than decrease, with spheres of influence becoming more pronounced and policy divergence between blocs intensifying. These dynamics have far-reaching implications, suggesting heightened challenges for global governance, increased barriers to international cooperation, and a geopolitical landscape where strategic alignments play an increasingly central role in shaping economic and political outcomes.

References

Acemoglu, Daron, and James A Robinson. 2000. "Why did the West extend the franchise? Democracy, inequality, and growth in historical perspective." Quarterly Journal of Economics, 115(4): 1167–1199.

Acemoglu, Daron, Georgy Egorov, and Konstantin Sonin. 2008. "Coalition formation in non-democracies." Review of Economic Studies, 75(4): 987–1009.

Acemoglu, Daron, James A. Robinson, and Thierry Verdier. 2004. "Kleptocracy and Divide-and-Rule: A Model of Personal Rule." Journal of the European Economic Association, 2(2-3): 162–192.

Aiyar, Mr Shekhar, Mr Jiaqian Chen, Christian Ebeke, Mr Christian H Ebeke, Mr Roberto Garcia-Saltos, Tryggvi Gudmundsson, Ms Anna Ilyina, Mr Alvar Kangur, Tansaya Kunaratskul, Mr Sergio L Rodriguez, et al. 2023. Geoeconomic fragmentation and the future of multilateralism. International Monetary Fund.

Aiyar, Shekhar, and Franziska Ohnsorge. 2024. "Geoeconomic Fragmentation and 'Connector' Countries." Working Paper.

Aiyar, Shekhar, Andrea Presbitero, and Michele Ruta. 2023. Geoeconomic Fragmentation: The Economic Risks from a Fractured World Economy. CEPR Press.

Aiyar, Shekhar, Davide Malacrino, and Andrea F Presbitero. 2024. "Investing in friends: The role of geopolitical alignment in FDI flows." European Journal of Political Economy, 83.

Ambrocio, Gene, and Iftekhar Hasan. 2021. "Quid pro quo? Political ties and sovereign borrowing." Journal of International Economics, 133: 103523.

Antràs, Pol, and Gerard Padró i Miquel. 2011. "Foreign influence and welfare." Journal of International Economics, 84(2): 135–148.

Art, Robert. 2004. "Europe Hedges Its Security Bets." Balance of Power: Theory and Practice in the 21st Century, (editors: Paul, Thazha Varkey, James J. Wirtz, and Michel Fortmann).

ASA, Australian Government. 2024. "AUKUS." Accessed: January 29, 2025: URL: https://www.asa.gov.au/aukus.

Bailey, Michael A, Anton Strezhnev, and Erik Voeten. 2017. "Estimating dynamic state preferences from United Nations voting data." Journal of Conflict Resolution, 61(2): 430–456.

Baliga, Sandeep, and Tomas Sjöström. 2013. "Bargaining and war: A review of some formal models." Korean Economic Review, 29(2): 235–266.

Baliga, Sandeep, and Tomas Sjöström. 2020. "The strategy and technology of conflict." Journal of Political Economy, 128(8): 3186–3219.

BBC. 2018. "US cuts Cambodia aid over democracy concerns." Accessed: January 29, 2025: URL: https://www.bbc.com/news/world–asia–43217043.

Berger, Daniel, William Easterly, Nathan Nunn, and Shanker Satyanath. 2013. "Commercial imperialism? Political influence and trade during the Cold War." American Economic Review, 103(2): 863–96.

Broner, Fernando, Alberto Martin, Josefin Meyer, and Christoph Trebesch. 2024. "Hegemonic globalization." Working Paper.

Chassang, Sylvain, and Gerard Padró I Miquel. 2010. "Conflict and deterrence under strategic risk." Quarterly Journal of Economics, 125(4): 1821–1858.

Clayton, Christopher, Matteo Maggiori, and Jesse Schreger. 2023. "A Theory of Economic Coercion and Fragmentation."

Clayton, Christopher, Matteo Maggiori, and Jesse Schreger. 2024. "A Theory of Economic Coercion and Fragmentation."

Davis, Christina L, Andreas Fuchs, and Kristina Johnson. 2019. "State control and the effects of foreign relations on bilateral trade." Journal of Conflict Resolution, 63(2): 405–438.

DFAT, Australian Government. 2024. "The Quad." Accessed: January 29, 2025: URL: https://www.dfat.gov.au/international–relations/regional–architecture/quad.

Diermeier, Daniel, Georgy Egorov, and Konstantin Sonin. 2017. "Political Economy of Redistribution." Econometrica, 85(3): 851–870.

Du, Yingxin, Jiandong Ju, Carlos D Ramirez, and Xi Yao. 2017. "Bilateral trade and shocks in political relations: Evidence from China and some of its major trading partners, 1990–2013." Journal of International Economics, 108: 211–225.

Egorov, Georgy, and Konstantin Sonin. 2024. "The political economics of non-democracy." Journal of Economic Literature, 62(2): 594–636.

Faye, Michael, and Paul Niehaus. 2012. "Political aid cycles." American Economic Review, 102(7): 3516–30.

Fearon, James D. 1995. "Rationalist Explanations for War." International Organization, 49(3): 379–414.

Fuchs, Andreas, and Nils-Hendrik Klann. 2013. "Paying a visit: The Dalai Lama effect on international trade." Journal of International Economics, 91(1): 164–177.

Garfinkel, Michelle R., and Stergios Skaperdas. 2007. "Economics of Conflict: An Overview." In Handbook of defense economics. Vol. 2, 649–709.

Goh, Evelyn. 2006. "Understanding "hedging" in Asia-Pacific security." PacNet, 43: 31.

Gopinath, Gita, Pierre-Olivier Gourinchas, Andrea F Presbitero, and Petia Topalova. 2025. "Changing global linkages: A new Cold War?" Journal of International Economics, 153: 104042.

Guzansky, Yoel. 2015. "Strategic Hedging by Non-Great Powers in the Persian Gulf." In Great Powers and Geopolitics. 231–252. Springer.

Haacke, Jürgen, and John D Ciorciari. 2022. "Hedging as Risk Management: Insights from Works on Alignment, Riskification, and Strategy."

Halac, Marina, Elliot Lipnowski, and Daniel Rappoport. 2021. "Rank uncertainty in organizations." American Economic Review, 111(3): 757–86.

Halac, Marina, Ilan Kremer, and Eyal Winter. 2020. "Raising capital from heterogeneous investors." American Economic Review, 110(3): 889–921.

Hamdi, Sofie, and Mohammad Salman. 2020. "The Hedging Strategy of Small Arab Gulf States." Asian Politics & Policy, 12(2): 127–152.

Jackson, Matthew O., and Massimo Morelli. 2009. "Strategic Militarization, Deterrence and Wars." Quarterly Journal of Political Science, 4(4): 279–313.

Jackson, Matthew O., and Massimo Morelli. 2011. "The Reasons for War: an Updated Survey." In Handbook on the Political Economy of War. , ed. Edward Elgar Oublishing.

Kapon, Samuel, Lucia Del Carpio, and Sylvain Chassang. 2024. "Using divide-and-conquer to improve tax collection." The Quarterly Journal of Economics, 139(4): 2475–2523.

Kleinman, Benny, Ernest Liu, and Stephen J Redding. 2020. "International friends and enemies." National Bureau of Economic Research.

Koga, Kei. 2018. "The concept of "hedging" revisited: the case of Japan's foreign policy strategy in East Asia's power shift." International Studies Review, 20(4): 633–660.

Kuziemko, Ilyana, and Eric Werker. 2006. "How much is a seat on the Security Council worth? Foreign aid and bribery at the United Nations." Journal of Political Economy, 114(5): 905–930.

Levitsky, Steven. 2011. The Resurgence of the Latin American Left. Cambridge University Press.

Lim, Darren J, and Rohan Mukherjee. 2019. "Hedging in South Asia: balancing economic and security interests amid Sino-Indian competition." International Relations of the Asia-Pacific, 19(3): 493–522.

Meirowitz, Adam, Massimo Morelli, Kristopher W Ramsay, and Francesco Squintani. 2019. "Dispute resolution institutions and strategic militarization." Journal of Political Economy, 127(1): 378–418.

Morelli, Massimo, and Dominic Rohner. 2015. "Resource concentration and civil wars." Journal of Development Economics, 117: 32–47.

Moyer, Jonathan D., Tim Sweijs, Mathew J. Burrows, and Hugo Van Manen. 2018. "Power and Influence in a Globalized World." The Atlantic Council.

NATO. 2024. "NATO member countries." Accessed: January 29, 2025: URL: https://www.nato.int/cps/en/natohq/topics$_5$2044.*htm*.

Padró i Miquel, Gerard. 2007. "The control of politicians in divided societies: the politics of fear." Review of Economic Studies, 74(4): 1259–1274.

Pape, Robert A. 2005. "Soft balancing against the United States." International security, 30(1): 7–45.

Paul, Thazha V. 2005. "Soft balancing in the age of US primacy." International security, 30(1): 46–71.

Pich, Charadine, and Chhengpor Aun. 2023. "How Small States Navigate U.S.-China Rivalry: The Case of Cambodia." URL: https://www.usip.org/publications/2023/09/how–small–states–navigate–us–china–rivalry–case–cambodia.

Posen, Barry R. 2006. "European Union security and defense policy: response to unipolarity?" Security studies, 15(2): 149–186.

Powell, Robert. 1999. "In the Shadow of Power: States and Strategies in International Politics." Princeton University Press.

Roy, Denny. 2005. "Southeast Asia and China: balancing or bandwagoning?" Contemporary Southeast Asia: A Journal of International and Strategic Affairs, 27(2): 305–322.

Segal, Ilya. 1999. "Contracting with externalities." Quarterly Journal of Economics, 114(2): 337–388.

Selden, Zachary. 2013. "Balancing against or balancing with? The spectrum of alignment and the endurance of American hegemony." Security studies, 22(2): 330–364.

Strating, Rebecca. 2020. "Small power hedging in an era of great-power politics: Southeast Asian responses to China's pursuit of energy security." Asian Studies Review, 44(1): 97–116.

Tessman, Brock F. 2012. "System structure and state strategy: Adding hedging to the menu." Security Studies, 21(2): 192–231.

Waltz, Kenneth N. 1979. Theory of international politics. McGraw-Hill.

Whitaker, Beth Elise. 2010. "Soft balancing among weak states? Evidence from Africa." International Affairs, 86(5): 1109–1127.

Winter, Eyal. 2004. "Incentives and discrimination." American Economic Review, 94(3): 764–773.

Wivel, Anders. 2008. "Balancing against threats or bandwagoning with power? Europe and the transatlantic relationship after the Cold War." Cambridge Review of International

Affairs, 21(3): 289–305.

Wohlforth, William C, Richard Little, Stuart J Kaufman, David Kang, Charles A Jones, Victoria Tin-Bor Hui, Arthur Eckstein, Daniel Deudney, and William L Brenner. 2007. "Testing balance-of-power theory in world history." European Journal of International Relations, 13(2): 155–185.

Wong, Audrye Y. 2015. "Comparing Japanese and South Korean strategies toward China and the United States: All politics is local." Asian survey, 55(6): 1241–1269.

Yared, Pierre. 2010. "A dynamic theory of war and peace." Journal of Economic Theory, 145(5): 1921–1950.

Yeats, Alexander J. 1990. "Do African countries pay more for imports? Yes." The World Bank Economic Review, 4(1): 1–20.

Appendices

A Proofs **34**

B Robustness checks **47**

B.1 Alternative functional form 47

B.2 Alternative controls and fixed effects 47

B.3 Alternative instrumental variables 52

B.4 Alternative measures of foreign influence 54

B.5 Alternative outcomes 57

A Proofs

Proof of Theorem 1. To prove this result we proceed in four steps:

Step 1: The continuation value of $i \in \{f,h\}$ and $-i \in \{f,h\} \setminus \{i\}$ from reaching the subgame where player $i \in \{f,h\}$ can make a take-it-or-leave-it offer to $-i \in \{f,h\} \setminus \{i\}$ is $v_{i,-i}(\overline{X}^i)$ and $v_{-i,i}(\overline{X}^i)$, where $\overline{X}^i = argmax_X v_{i,-i}(X)$ subject to $v_{-i,i}(X) \geq 0$.

Proof of Step 1. Starting from the subgame where $-i$ chooses whether to accept i's take-it-or-leave-it offer X^i obtaining $v_{-i,i}(X^i)$ or to break relations with i and obtain a payoff of 0. The only equilibria of this subgame are (a) $-i$ does not break its relation with i if and only if $v_{-i,i}(X^i) \geq 0$ or (b) $-i$ does not break its relation with i if and only if $v_{-i,i}(X^i) > 0$. Moving up the game tree we reach the node where i chooses what take-it-or-leave-it offer $X^i \in \mathbb{R}_+$ it will make to $-i$ anticipating $-i$'s potential responses: (a) or (b). Considering case (a), i obtains a payoff of 0 when offering X^i such that $v_{-i,i}(X^i) < 0$ and a payoff of $v_{i,-i}(X^i)$ for every X^i such that $v_{-i,i}(X^i) \geq 0$. Given that for every $f \in \{F,G\}$ there exists $X_{hf} \in \mathbb{R}_+$ such that $v_{hf}(X_{hf}) \geq 0$ and $v_{fh}(X_{hf}) \geq 0$, then the optimal $X^i \in \mathbb{R}_+$ must be such that $v_{-i,i}(X^i) \geq 0$. Thus, any equilibrium take-it-or-leave-it offer must belong to $\text{argmax}_X v_{i,-i}(X)$ subject to $v_{-i,i}(X) \geq 0$. Given our assumptions on the payoff functions, we know that such set is a singleton and thus i offers $\overline{X}^i = \text{argmax}_X v_{i,-i}(X)$ subject to $v_{-i,i}(X) \geq 0$. The continuation payoffs of i and $-i$ from reaching this subgame is thus $v_{i,-i}(\overline{X}^i)$ and $v_{-i,i}(\overline{X}^i)$. For what concerns case (b), note that either the set $\text{argmax}_X v_{i,-i}(X)$ subject to $v_{-i,i}(X) > 0$ is a singleton (in which case it coincides with $\text{argmax}_X v_{i,-i}(X)$ subject to $v_{-i,i}(X) \geq 0$), or it is empty and strategy (b) cannot be part of an equilibrium. In either cases, case (b) and case (a) lead to the same continuation values: $v_{i,-i}(\overline{X}^i)$ and $v_{-i,i}(\overline{X}^i)$. □

Step 2: We say that an equilibrium of the subgame following $\mathbf{X}$ and ζ satisfies the proposer-preferred tie-breaking rule if $U_{fh}(\mathbf{X},1,\zeta) = U_{fh}(\mathbf{X},0,\zeta)$ implies $r_{Fh} = 1$. Every subgame starting with offers $\mathbf{X}$ and realization ζ has a unique sunspot-consistent equilibrium which also satisfies the proposer-preferred tie-breaking rule. In such equilibrium, for each combination of $\mathbf{X}$ and ζ, every f plays according to function $r_{fh}(\mathbf{X},\zeta)$, where

$$r_{fh}(\mathbf{X},0) = \begin{cases} 1 & \text{if} \quad \tilde{v}_{fh}(X_{hf}) \geq \frac{w_{fh}+w_{-f,h}}{w_h+w_{fh}+w_{-f,h}} \quad \text{and} \quad \tilde{v}_{-f,h}(X_{h,-f}) < \frac{w_{fh}+w_{-f,h}}{w_h+w_{fh}+w_{-f,h}} \\ 1 & \text{if} \quad \tilde{v}_{fh}(X_{hf}) \geq \frac{w_{fh}}{w_h+w_{fh}+w_{-f,h}} \quad \text{and} \quad \tilde{v}_{-f,h}(X_{h,-f}) \geq \frac{w_{fh}+w_{-f,h}}{w_h+w_{fh}+w_{-f,h}} \\ 0 & \text{otherwise} \end{cases}$$

$$r_{fh}(\mathbf{X},1) = \begin{cases} 1 & \text{if} \quad \tilde{v}_{fh}(X_{hf}) \geq \frac{w_{fh}+w_{-f,h}}{w_h+w_{fh}+w_{-f,h}} \quad \text{and} \quad \tilde{v}_{-f,h}(X_{h,-f}) < \frac{w_{-f,h}}{w_h+w_{fh}+w_{-f,h}} \\ 1 & \text{if} \quad \tilde{v}_{fh}(X_{hf}) \geq \frac{w_{fh}}{w_h+w_{fh}+w_{-f,h}} \quad \text{and} \quad \tilde{v}_{-f,h}(X_{h,-f}) \geq \frac{w_{-f,h}}{w_h+w_{fh}+w_{-f,h}} \\ 0 & \text{otherwise} \end{cases}$$

Proof of Step 2. Given offers $\mathbf{X}$, the expected payoff $U_{fh}(\mathbf{X}, r_{fh}, r_{-f,h})$ of each f conditional on choice r_{fh}, $r_{-f,h}$, and the continuation values characterized in Step 1 is

$$U_{fh}(\mathbf{X}, 1, r_{-f,h}) = v_{fh}(X_{hf})$$
$$U_{fh}(\mathbf{X}, 0, r_{-f,h}) = \left[\frac{w_h + r_{-f,h}w_{-f,h}}{w_{Fh} + w_{Gh} + w_h} v_{fh}(\bar{X}^h_{hf}) + \frac{w_{fh} + (1 - r_{-f,h})w_{-f,h}}{w_{Fh} + w_{Gh} + w_h} v_{fh}(\bar{X}^f_{hf})\right]$$

Thus, f's gain from choosing $r_{fh} = 1$ over $r_{fh} = 0$, i.e., $U_{fh}(\mathbf{X}, 1, r_{-f,h}) - U_{fh}(\mathbf{X}, 0, r_{-f,h})$, is greater or equal to zero if and only if

$$v_{fh}(X_{hf}) - \left[1 - \frac{w_{fh} + (1 - r_{-f,h})w_{-f,h}}{w_{Fh} + w_{Gh} + w_h}\right] v_{fh}(\bar{X}^h_{hf}) - \frac{w_{fh} + (1 - r_{-f,h})w_{-f,h}}{w_{Fh} + w_{Gh} + w_h} v_{fh}(\bar{X}^f_{hf}) \geq 0$$
$$\iff \tilde{v}_{fh}(X_{hf}) := \frac{v_{fh}(X_{hf}) - v_{fh}(\bar{X}^h_{hf})}{v_{fh}(\bar{X}^f_{hf}) - v_{fh}(\bar{X}^h_{hf})} \geq \frac{w_{fh} + (1 - r_{-f,h})w_{-f,h}}{w_{Fh} + w_{Gh} + w_h}$$

Where the last inequality relies on the fact that $v_{fh}(\bar{X}^f_{hf}) > v_{fh}(\bar{X}^h_{hf})$ for both f.

With this, note that there are offers $\mathbf{X}$ for which the corresponding subgame has a unique rationalizable strategy profile and thus a unique equilibrium: (i) $\hat{r}_{fh} = 1$ and $\hat{r}_{-f,h} = 1$ if $\tilde{v}_{fh}(X_{hf}) > \frac{w_{fh}+w_{-f,h}}{w_{Fh}+w_{Gh}+w_h}$ and $\tilde{v}_{-f,h}(X_{hf}) > \frac{w_{-f,h}}{w_{Fh}+w_{Gh}+w_h}$, (ii) $\hat{r}_{fh} = 0$ and $\hat{r}_{-f,h} = 0$ if $\tilde{v}_{fh}(X_{hf}) < \frac{w_{fh}}{w_{Fh}+w_{Gh}+w_h}$ and $\tilde{v}_{-f,h}(X_{hf}) < \frac{w_{-f,h}}{w_{Fh}+w_{Gh}+w_h}$, and (iii) $\hat{r}_{fh} = 1$ and $\hat{r}_{-f,h} = 0$ if $\tilde{v}_{fh}(X_{hf}) > \frac{w_{fh}+w_{-f,h}}{w_{Fh}+w_{Gh}+w_h}$ and $\tilde{v}_{-f,h}(X_{hf}) < \frac{w_{-f,h}}{w_{Fh}+w_{Gh}+w_h}$. For every $\mathbf{X}$ covered by the cases above, we can see that our proposed function implies $r_{fh}(\mathbf{X}, 1) = r_{fh}(\mathbf{X}, 0) = \hat{r}_{fh}$ and $r_{-f,h}(\mathbf{X}, 1) = r_{-f,h}(\mathbf{X}, 0) = \hat{r}_{-f,h}$ for both f, as required. Now consider the subgame following $\mathbf{X}$ such that for both f we have $\frac{w_{fh}+w_{-f,h}}{w_{Fh}+w_{Gh}+w_h} > \tilde{v}_{fh}(X_{hf}) > \frac{w_{fh}}{w_{Fh}+w_{Gh}+w_h}$. Note that in these cases both $\hat{r}_{fh} = \hat{r}_{-f,h} = 1$ and $\hat{r}_{fh} = \hat{r}_{-f,h} = 0$ are equilibria. Additionally, note that any sunspot-consistent equilibrium must select $\hat{r}_{fh} = \hat{r}_{-f,h} = 1$ after $\zeta = 1$ because for every such $\mathbf{X}$ we have $U_{fh}(\mathbf{X}, 1, 1) - U_{fh}(\mathbf{X}, 0, 1) > 0$ for both f. Moreover, any sunspot-consistent equilibrium must select $\hat{r}_{fh} = \hat{r}_{-f,h} = 0$ after $\zeta = 0$ because for every such $\mathbf{X}$ we have $U_{fh}(\mathbf{X}, 1, 0) - U_{fh}(\mathbf{X}, 0, 0) < 0$ for both f. As required, for these $\mathbf{X}$, our proposed function is such that $r_{fh}(\mathbf{X}, 1) = r_{-f,h}(\mathbf{X}, 1) = \hat{r}_{fh} = \hat{r}_{-f,h} = 1$ and $r_{fh}(\mathbf{X}, 0) = r_{-f,h}(\mathbf{X}, 0) = \hat{r}_{fh} = \hat{r}_{-f,h} = 0$. To conclude we have some knife-edge cases to consider. First, consider symmetric offers $\mathbf{X}$ such that, for both f, $\tilde{v}_{fh}(X_{hf}) = \frac{w_{fh}}{w_{Fh}+w_{Gh}+w_h}$, then any sunspot-consistent equilibrium must select $\hat{r}_{fh} = \hat{r}_{-f,h} = 0$ after $\zeta = 0$ because for every such $\mathbf{X}$ we have $U_{fh}(\mathbf{X}, 1, 0) - U_{fh}(\mathbf{X}, 0, 0) < 0$ for both f. As required, for such $\mathbf{X}$, our function is $r_{fh}(\mathbf{X}, 0) = r_{-f,h}(\mathbf{X}, 0) = 0$. On the other hand, we must have either $\hat{r}_{fh} = \hat{r}_{-f,h} = 1$ or $\hat{r}_{fh} = \hat{r}_{-f,h} = 0$ after $\zeta = 1$ as both are equilibria and $U_{fh}(\mathbf{X}, 1, 1) - U_{fh}(\mathbf{X}, 0, 1) = 0$ for both f. However, by the proposer-preferred tie-breaking rule we still select a unique pair

of responses after $\zeta = 1$, i.e., $r_{fh}(\mathbf{X}, 1) = r_{-f,h}(\mathbf{X}, 1) = 1$. Similarly, consider offers $\mathbf{X}$ such that, for both f, we have $\tilde{v}_{fh}(X_{hf}) = \frac{w_{fh}+w_{-f,h}}{w_{Fh}+w_{Gh}+w_h}$. In such case, any sunspot-consistent equilibrium must select $\hat{r}_{fh} = \hat{r}_{-f,h} = 1$ after $\zeta = 1$ because for every such $\mathbf{X}$ we have $U_{fh}(\mathbf{X}, 1, 1) - U_{fh}(\mathbf{X}, 0, 1) > 0$ for both f. And indeed our function evaluated at such $\mathbf{X}$ and $\zeta = 1$ is $r_{fh}(\mathbf{X}, 1) = r_{-f,h}(\mathbf{X}, 1) = 1$. Instead, we must have either $\hat{r}_{fh} = \hat{r}_{-f,h} = 1$ or $\hat{r}_{fh} = \hat{r}_{-f,h} = 0$ after $\zeta = 0$ as both are equilibria and $U_{fh}(\hat{\mathbf{X}}, 1, 0) - U_{fh}(\hat{\mathbf{X}}, 0, 0) = 0$ for both f. Just as before, our tie-breaking rule solves this multiplicity and we get $r_{fh}(\mathbf{X}, 0) = r_{-f,h}(\mathbf{X}, 0) = 1$. Finally, consider asymmetric offers $\mathbf{X}$ where there is f receiving $\tilde{v}_{fh}(X_{hf}) = \frac{w_{fh}+w_{-f,h}}{w_{Fh}+w_{Gh}+w_h}$ and the other great power receives $\tilde{v}_{-f,h}(X_{hf}) = \frac{w_{-f,h}}{w_{Fh}+w_{Gh}+w_h}$, then both $\hat{r}_{fh} = \hat{r}_{-f,h} = 1$ and $\hat{r}_{fh} = \hat{r}_{-f,h} = 0$ are equilibria. Moreover, both are sunspot-consistent equilibria because we have $U_{fh}(\mathbf{X}, 1, 0) - U_{fh}(\mathbf{X}, 0, 0) = 0$ and $U_{-f,h}(\mathbf{X}, 1, 1) - U_{fh}(\mathbf{X}, 0, 1) = 0$. However, only $\hat{r}_{fh} = \hat{r}_{-f,h} = 1$ satisfy our tie-breaking condition. □

Step 3: In every p_ζ-SPE consistent with the proposer-preferred tie-breaking rule, $\mathbf{X}$ satisfies (i) $\mathbf{r}(\mathbf{X}, \zeta) = \mathbf{1}$ for all $\zeta \in \{0, 1\}$, or (ii) $\mathbf{r}(\mathbf{X}, 1) = \mathbf{1}$ but $\mathbf{r}(\mathbf{X}, 0) = \mathbf{0}$.

Proof of Step 3. To prove the statement we will show the converse: if there is f such that $r_{fh}(\mathbf{X}, 1) = 0$ or $r_{fh}(\mathbf{X}, 0) \neq r_{-f,h}(\mathbf{X}, 0)$, then $\mathbf{X}$ cannot be part of a p_ζ-SPE consistent with the proposer-preferred tie-breaking rule. From Step 2 we know that in any p_ζ-SPE consistent with the proposer-preferred tie-breaking rule the responses must satisfy function $r_{fh}(\mathbf{X}, \zeta)$ defined at Step 2. We now show that to prove our statement we just need to rule out the following possibilities: (Option 1) $\tilde{\mathbf{X}}$ such that $\tilde{v}_{hf}(\tilde{X}_{hf}) \geq \frac{w_{fh}+w_{-f,h}}{w_h+w_{fh}+w_{-f,h}}$ and $\tilde{v}_{h,-f}(\tilde{X}_{h,-f}) < \frac{w_{-f,h}}{w_h+w_{fh}+w_{-f,h}}$ (Option 2) $\tilde{\mathbf{X}}$ such that $\frac{w_{fh}+w_{-f,h}}{w_h+w_{fh}+w_{-f,h}} > \tilde{v}_{hf}(\tilde{X}_{hf}) \geq \frac{w_{fh}}{w_h+w_{fh}+w_{-f,h}}$ and $\tilde{v}_{h,-f}(\tilde{X}_{h,-f}) < \frac{w_{-f,h}}{w_h+w_{fh}+w_{-f,h}}$ and (Option 3) $\tilde{\mathbf{X}}$ such that $\tilde{v}_{hf}(\tilde{X}_{hf}) < \frac{w_{fh}}{w_h+w_{fh}+w_{-f,h}}$ and $\tilde{v}_{h,-f}(\tilde{X}_{h,-f}) < \frac{w_{-f,h}}{w_h+w_{fh}+w_{-f,h}}$. To see why these options cover all cases, note that changing the second condition in Option 1 into $\tilde{v}_{h,-f}(\tilde{X}_{hf}) \geq \frac{w_{-f,h}}{w_h+w_{fh}+w_{-f,h}}$ would lead to $\mathbf{r}(\mathbf{X}, \zeta) = \mathbf{1}$ for all $\zeta \in \{0, 1\}$, which is compatible with (i) in the claim. Similarly, changing the second condition in Option 2 with $\tilde{v}_{h,-f}(\tilde{X}_{hf}) \geq \frac{w_{-f,h}}{w_h+w_{fh}+w_{-f,h}}$ would either lead to $\mathbf{r}(\mathbf{X}, \zeta) = \mathbf{1}$ for all $\zeta \in \{0, 1\}$ or to $\mathbf{r}(\mathbf{X}, 1) = \mathbf{1}$ but $\mathbf{r}(\mathbf{X}, 0) = \mathbf{0}$, which are compatible with (i) and (ii) in the claim, respectively. Finally, it is true that changing the second condition in Option 3 with $\tilde{v}_{h,-f}(\tilde{X}_{h,-f}) \geq \frac{w_{fh}+w_{-f,h}}{w_h+w_{fh}+w_{-f,h}}$ or with $\frac{w_{fh}+w_{-f,h}}{w_h+w_{fh}+w_{-f,h}} > \tilde{v}_{h,-f}(\tilde{X}_{h,-f}) \geq \frac{w_{-f,h}}{w_h+w_{fh}+w_{-f,h}}$ would still lead to cases ruled out by the proposition. However, such cases correspond to Options 1 and 2 with the great powers re-labeled, so we do not have to consider them. Our objective is now to show that for any offer $\tilde{\mathbf{X}}$ compatible with Options 1, 2, or 3 there is at least another offer $\tilde{\mathbf{X}}'$ that yields a strictly greater expected payoff to h in any p_ζ-SPE consistent with the proposer-preferred tie-breaking rule and that satisfies either condition (i) or (ii) in the claim.

Using Step 1 and Step 2 we can write the expected payoff derived by every player

for any given offer in any p_ζ-SPE consistent with the proposer-preferred tie-breaking rule. For any offer satisfying Option 1, f obtains $v_{fh}(\tilde{X}_{hf})$, $-f$ gets $\frac{w_{-f,h}}{w_h+w_{fh}+w_{-f,h}}v_{-f,h}(\overline{X}^{-f}_{h,-f}) + \frac{w_h+w_f}{w_h+w_{fh}+w_{-f,h}}v_{-f,h}(\overline{X}^{h}_{h,-f})$, and h obtains $U_h(\tilde{\mathbf{X}},\mathbf{r}) = v_{hf}(\tilde{X}_{hf}) + U_{h,-f}(\tilde{\mathbf{X}},\mathbf{r}) - \delta_{h,-f}$, where $U_{h,-f}(\tilde{\mathbf{X}},\mathbf{r}) = \frac{w_{-f,h}}{w_h+w_{fh}+w_{-f,h}}v_{h,-f}(\overline{X}^{-f}_{h,-f}) + \frac{w_h+w_f}{w_h+w_{fh}+w_{-f,h}}v_{h,-f}(\overline{X}^{h}_{h,-f})$. Now consider an alternative offer $\tilde{\mathbf{X}}'$ such that $\tilde{X}'_{hf} = \tilde{X}_{hf}$ and $\tilde{X}'_{h,-f} = \frac{w_{-f,h}}{w_h+w_{fh}+w_{-f,h}}\overline{X}^{-f}_{h,-f} + \frac{w_h+w_{fh}}{w_h+w_{fh}+w_{-f,h}}\overline{X}^{h}_{h,-f}$. Because of the concavity of $v_{-f,h}(\cdot)$, we know that $v_{-f,h}(\tilde{X}'_{h,-f}) \geq \frac{w_{-f,h}}{w_h+w_{fh}+w_{-f,h}}v_{-f,h}(\overline{X}^{-f}_{h,-f}) + \frac{w_h+w_{fh}}{w_h+w_{fh}+w_{-f,h}}v_{-f,h}(\overline{X}^{h}_{h,-f})$. Thus, $\tilde{v}_{-f,h}(\tilde{X}'_{hf}) \geq \frac{w_{-f,h}}{w_h+w_{fh}+w_{-f,h}}$ which by Step 2 implies $r_{fh}(\tilde{X}'_{hf},\zeta) = r_{-f,h}(\tilde{X}'_{h,-f},\zeta) = 1$ for every ζ, thus showing that $\tilde{\mathbf{X}}'$ satisfies condition (i). Moreover, note that this alternative offer provides h with an expected payoff $U_h(\tilde{\mathbf{X}}',\mathbf{r}) = v_{hf}(\tilde{X}_{hf}) + v_{h,-f}(\tilde{X}'_{h,-f})$. Given $\delta_{h,-f} > 0$ and the concavity of $v_{h,-f}(\cdot)$, implying $v_{h,-f}(\tilde{X}'_{h,-f}) \geq \frac{w_{-f,h}}{w_h+w_{fh}+w_{-f,h}}v_{h,-f}(\overline{X}^{-f}_{h,-f}) + \frac{w_h+w_f}{w_h+w_{fh}+w_{-f,h}}v_{h,-f}(\overline{X}^{h}_{h,-f})$, we conclude $U_h(\tilde{\mathbf{X}}',\mathbf{r}) > U_h(\tilde{\mathbf{X}},\mathbf{r})$ as we wanted to show. For any offer satisfying Option 2 or 3, each $f \in \{F,G\}$ obtains $\frac{w_{fh}+w_{-f,h}}{w_h+w_{fh}+w_{-f,h}}v_{fh}(\overline{X}^{f}_{hf}) + \frac{w_h}{w_h+w_{fh}+w_{-f,h}}v_{fh}(\overline{X}^{h}_{hf})$ and h obtains $U_h(\tilde{\mathbf{X}},\mathbf{r}) = U_{hF}(\tilde{\mathbf{X}},\mathbf{r}) + U_{hG}(\tilde{\mathbf{X}},\mathbf{r}) - \delta_{hF} - \delta_{hG}$, where $U_{hf}(\tilde{\mathbf{X}},\mathbf{r}) = \frac{w_{fh}+w_{-f,h}}{w_h+w_{fh}+w_{-f,h}}v_{hf}(\overline{X}^{f}_{hf}) + \frac{w_h}{w_h+w_{fh}+w_{-f,h}}v_{hf}(\overline{X}^{h}_{hf})$ for both $f \in \{F,G\}$. Now consider an alternative offer $\tilde{\mathbf{X}}'$ such that $\tilde{X}'_{hf} = \frac{w_{fh}+w_{-f,h}}{w_h+w_{fh}+w_{-f,h}}\overline{X}^{f}_{hf} + \frac{w_h}{w_h+w_{fh}+w_{-f,h}}\overline{X}^{h}_{hf}$ for both $f \in \{F,G\}$. By concavity of $v_{fh}(\cdot)$ for both $f \in \{F,G\}$, we know that $v_{fh}(\tilde{X}'_{hf}) \geq \frac{w_{fh}+w_{-f,h}}{w_h+w_{fh}+w_{-f,h}}v_{fh}(\overline{X}^{f}_{hf}) + \frac{w_h}{w_h+w_{fh}+w_{-f,h}}v_{fh}(\overline{X}^{h}_{hf})$ for both $f \in \{F,G\}$. Thus, for both $f \in \{F,G\}$, $\tilde{v}_{fh}(\tilde{X}'_{hf}) \geq \frac{w_{fh}+w_{-f,h}}{w_h+w_{fh}+w_{-f,h}}$ which by Step 2 implies $r_{fh}(\tilde{X}'_{hf},\zeta) = r_{-f,h}(\tilde{X}'_{h,-f},\zeta) = 1$ for every ζ, thus showing that $\tilde{\mathbf{X}}'$ satisfies condition (i). Moreover, note that this alternative offer provides h with an expected payoff $U_h(\tilde{\mathbf{X}}',\mathbf{r}) = v_{hf}(\tilde{X}'_{hf}) + v_{hf}(\tilde{X}'_{hf})$. Because of the concavity of $v_{hf}(\cdot)$ for both $f \in \{F,G\}$, we see that $v_{hf}(\tilde{X}'_{hf}) \geq \frac{w_{-f,h}}{w_h+w_{fh}+w_{-f,h}}v_{h,-f}(\overline{X}^{-f}_{h,-f}) + \frac{w_h+w_f}{w_h+w_{fh}+w_{-f,h}}v_{h,-f}(\overline{X}^{h}_{h,-f})$ for both $f \in \{F,G\}$. Moreover, because $\delta_{hf} > 0$ for both $f \in \{F,G\}$, we conclude that $U_h(\tilde{\mathbf{X}}',\mathbf{r}) > U_h(\tilde{\mathbf{X}},\mathbf{r})$, as desired. □

Step 4: In every p_ζ-SPE, $\mathbf{X}$ satisfies (i) $\mathbf{r}(\mathbf{X},\zeta) = \mathbf{1}$ for all $\zeta \in \{0,1\}$, or (ii) $\mathbf{r}(\mathbf{X},1) = \mathbf{1}$ but $\mathbf{r}(\mathbf{X},0) = \mathbf{0}$.

Proof of Step 4. We conclude showing that Step 3 holds for any sunspot-consistent strategy and not just the one satisfying the proposer-preferred tie-breaking rule. From the proof of Step 2 we know that any sunspot-consistent equilibrium must have responses consistent with the functions $\mathbf{r}_{fh}(\mathbf{X},\zeta)$ reported in Step 2 except at some of the cut-off points. With this in mind, and recalling that $v_{hf}(\cdot)$ and $v_{fh}(\cdot)$ are continuous for both $f \in \{F,G\}$, we see that there is $\boldsymbol{\varepsilon} > 0$ such that, if we let $\tilde{\mathbf{X}}'' = \tilde{\mathbf{X}}' + \boldsymbol{\varepsilon}$, we can still find (a) $r_{fh}(\tilde{\mathbf{X}}'',\zeta) = r_{fh}(\tilde{\mathbf{X}}',\zeta)$ for every f and ζ, and (b) $U_h(\tilde{\mathbf{X}}',\mathbf{r}) > U_h(\tilde{\mathbf{X}}'',\mathbf{r}) > U_h(\tilde{\mathbf{X}},\mathbf{r})$, where $\tilde{\mathbf{X}}'$ is the strategy we used in the proof of Step 3. □

■

Lemma 3 *In every p_ζ-SPE, the optimal offers* $\mathbf{X}$ *are such that* $X_{hf} \in conv(\{\bar{X}^h_{hf}, \bar{X}^f_{hf}\}) = \{X \in \mathbb{R}_+ : \exists \lambda \in [0,1], X = \lambda \bar{X}^f_{hf} + (1-\lambda)\bar{X}^h_{hf}\}$ *for every* $f \in \{F, G\}$.

Proof of Lemma 3. Suppose not, then there is offer $X_{hf} \in \mathbb{R}_+$ such that either $X_{hf} > \max\{\bar{X}^h_{hf}, \bar{X}^f_{hf}\}$ or $X_{hf} < \min\{\bar{X}^h_{hf}, \bar{X}^f_{hf}\}$. WLOG, suppose that $\bar{X}^h_{hf} < \bar{X}^f_{hf}$.

First, consider $X_{hf} > \max\{\bar{X}^h_{hf}, \bar{X}^f_{hf}\} = \bar{X}^f_{hf}$. Recall that $\bar{X}^f_{hf}$ is the unique optimum of $v_{fh}(X)$ subject to $v_{hf}(X) \geq 0$. So, $X_{hf} > \bar{X}^f_{hf}$ implies $v_{hf}(X_{hf}) < 0$ or $v_{fh}(X_{hf}) < v_{fh}(\bar{X}^f_{hf})$. First, suppose that $X_{hf} > \bar{X}^f_{hf} > \bar{X}^h_{hf}$ and $v_{hf}(X_{hf}) < 0$, then X_{hf} cannot be optimal because $v_{hf}(\bar{X}^f_{hf}) \geq 0 > v_{hf}(X_{hf})$ and, from Theorem 1, for any $(X_{hf}, X_{h,-f})$, we have $r_{fh}(\bar{X}^f_{hf}, X_{h,-f}, \zeta) \geq r_{fh}(X_{hf}, X_{h,-f}, \zeta)$ and $r_{-f,h}(\bar{X}^f_{hf}, X_{h,-f}, \zeta) \geq r_{fh}(X_{hf}, X_{h,-f}, \zeta)$ for every $\zeta \in \{0,1\}$, implying that h's expected payoff is strictly greater when offering $(\bar{X}^f_{hf}, X_{h,-f})$ rather than $(X_{hf}, X_{h,-f})$. Second, suppose that $X_{hf} > \bar{X}^f_{hf} > \bar{X}^h_{hf}$ and $v_{fh}(\bar{X}^f_{hf}) > v_{fh}(X_{hf})$, then consider $X_{hf}(\lambda^f) = \lambda^f X_{hf} + (1-\lambda^f)\bar{X}^f_{hf}$ for some $\lambda^f \in (0,1)$. By concavity of $v_{hf}(\cdot)$, $v_{hf}(X_{hf}(\lambda^f)) \geq \lambda^f v_{hf}(X_{hf}) + (1-\lambda^f) v_{hf}(\bar{X}^f_{hf})$. Moreover, $\bar{X}^f_{hf} = \lambda X_{hf} + (1-\lambda)\bar{X}^h_{hf}$ for some $\lambda \in (0,1)$. Again, by concavity of $v_{hf}(\cdot)$, we have $v_{hf}(\bar{X}^f_{hf}) \geq \lambda v_{hf}(X_{hf}) + (1-\lambda) v_{hf}(\bar{X}^h_{hf})$. So, $v_{hf}(X_{hf}(\lambda^f)) \geq \lambda^f v_{hf}(X_{hf}) + (1-\lambda^f)[\lambda v_{hf}(X_{hf}) + (1-\lambda) v_{hf}(\bar{X}^h_{hf}) = [\lambda^f + (1-\lambda^f)\lambda] v_{hf}(X_{hf}) + (1-\lambda^f)(1-\lambda) v_{hf}(\bar{X}^h_{hf}) > v_{hf}(X_{hf})$ because $(1-\lambda^f)(1-\lambda) > 0$ and by definition of $\bar{X}^h_{hf}$ we have $v_{hf}(\bar{X}^h_{hf}) > v_{hf}(X_{hf})$. Similarly, for the concavity of $v_{fh}(\cdot)$ we have $v_{fh}(X_{hf}(\lambda^f)) \geq \lambda^f v_{fh}(X_{hf}) + (1-\lambda^f) v_{fh}(\bar{X}^f_{hf}) > v_{fh}(X_{hf})$. As above, these findings imply that h's expected payoff is strictly greater when offering $(X_{hf}(\lambda^f), X_{h,-f})$ rather than $(X_{hf}, X_{h,-f})$ for any $\lambda^f \in (0,1)$.

Second, consider $X_{hf} < \min\{\bar{X}^h_{hf}, \bar{X}^f_{hf}\} = \bar{X}^h_{hf}$. Recall that $\bar{X}^h_{hf}$ is the unique optimum of $v_{hf}(X)$ subject to $v_{fh}(X) \geq 0$. So, $X_{hf} < \bar{X}^h_{hf}$ implies $v_{fh}(X_{hf}) < 0$ or $v_{hf}(X_{hf}) < v_{hf}(\bar{X}^h_{hf})$. However, if $X_{hf} < \bar{X}^h_{hf} < \bar{X}^f_{hf}$ and $v_{fh}(X_{hf}) < 0$, then $r_{fh}(X_{hf}, X_{h,-f}, 1) = 0$ which cannot be optimal by Theorem 1. So, we are left with the case of $X_{hf} < \bar{X}^h_{hf} < \bar{X}^f_{hf}$ and $v_{hf}(X_{hf}) < v_{hf}(\bar{X}^h_{hf})$. Consider $X_{hf}(\lambda^h) = \lambda^h X_{hf} + (1-\lambda^h)\bar{X}^h_{hf}$ for some $\lambda^h \in (0,1)$. By concavity of $v_{fh}(\cdot)$, $v_{fh}(X_{hf}(\lambda^h)) \geq \lambda^h v_{fh}(X_{hf}) + (1-\lambda^h) v_{fh}(\bar{X}^h_{hf})$. Moreover, $\bar{X}^h_{hf} = \lambda X_{hf} + (1-\lambda)\bar{X}^f_{hf}$ for some $\lambda \in (0,1)$. Again, by concavity of $v_{fh}(\cdot)$, we have $v_{fh}(\bar{X}^h_{hf}) \geq \lambda v_{fh}(X_{hf}) + (1-\lambda) v_{fh}(\bar{X}^f_{hf})$. So, $v_{fh}(X_{hf}(\lambda^h)) \geq \lambda^h v_{fh}(X_{hf}) + (1-\lambda^h)[\lambda v_{fh}(X_{hf}) + (1-\lambda) v_{fh}(\bar{X}^f_{hf}) = [\lambda^h + (1-\lambda^h)\lambda] v_{fh}(X_{hf}) + (1-\lambda^h)(1-\lambda) v_{fh}(\bar{X}^f_{hf}) > v_{fh}(X_{hf})$ because $(1-\lambda^h)(1-\lambda) > 0$ and by definition of $\bar{X}^f_{hf}$ we have $v_{fh}(\bar{X}^f_{hf}) > v_{fh}(X_{hf})$. Similarly, for the concavity of $v_{hf}(\cdot)$ we have $v_{hf}(X_{hf}(\lambda^h)) \geq \lambda^h v_{hf}(X_{hf}) + (1-\lambda^h) v_{hf}(\bar{X}^h_{hf}) > v_{hf}(X_{hf})$. As above, these findings imply that h's expected payoff is strictly greater when offering $(X_{hf}(\lambda^h), X_{h,-f})$ rather than $(X_{hf}, X_{h,-f})$ for any $\lambda^h \in (0,1)$.

■

Lemma 4 *Let* $X_f(\lambda) = \lambda \bar{X}^f_{hf} + (1-\lambda)\bar{X}^h_{hf}$ *for every* $\lambda \in [0,1]$ *and* $f \in \{F, G\}$*, and recall that* $\tilde{v}_{fh}(X_{hf}) := \frac{v_{fh}(X_{hf}) - v_{fh}(\bar{X}^h_{hf})}{v_{fh}(\bar{X}^f_{hf}) - v_{fh}(\bar{X}^h_{hf})}$ *for every* $X_{hf} \in \mathbb{R}_+$

(A) $v_{fh}(X_f(\lambda))$ *is a strictly increasing continuous function of* $\lambda \in [0,1]$ *and it is such that* $v_{fh}(X_f(0)) = v_{fh}(\bar{X}^h_{hf})$ *and* $v_{fh}(X_f(1)) = v_{fh}(\bar{X}^f_{hf})$*;* $v_{hf}(X_f(\lambda))$ *is a strictly decreasing continuous function of* $\lambda \in [0,1]$ *and it is such that* $v_{hf}(X_f(1)) = v_{hf}(\bar{X}^f_{hf})$ *and* $v_{hf}(X_f(0)) = v_{hf}(\bar{X}^h_{hf})$*.*

(B) There exists a unique mapping $\tilde{v}^{-1}_{fh} : [0,1] \to conv(\{\bar{X}^h_{hf}, \bar{X}^f_{hf}\})$ *satisfying* $\tilde{v}_{fh}(\tilde{v}^{-1}_{fh}(w))) = w$*. This mapping* $\tilde{v}^{-1}_{fh}(\cdot)$ *is a continuous bijective function.*

(C) $v_{hf}(\tilde{v}^{-1}_{fh}(w))$ *is a strictly decreasing continuous function for every* $w \in [0,1]$*. In particular, we have* $v_{hf}(\tilde{v}^{-1}_{fh}(w)) \geq v_{hf}(X_f(w))$ *for every* $w \in [0,1]$*,* $v_{hf}(\tilde{v}^{-1}_{fh}(0)) = v_{hf}(X_f(0)) = v_{hf}(\bar{X}^h_{hf})$*, and* $v_{hf}(\tilde{v}^{-1}_{fh}(1)) = v_{hf}(X_f(1)) = v_{hf}(\bar{X}^f_{hf})$

(D) $v_{fh}(\tilde{v}^{-1}_{fh}(w))$ *is a strictly increasing continuous function for every* $w \in [0,1]$*. In particular, we have* $v_{fh}(\tilde{v}^{-1}_{fh}(w)) \leq v_{fh}(X_f(w))$ *for every* $w \in [0,1]$*,* $v_{fh}(\tilde{v}^{-1}_{fh}(0)) = v_{fh}(X_f(0)) = v_{fh}(\bar{X}^h_{hf})$*, and* $v_{fh}(\tilde{v}^{-1}_{fh}(1)) = v_{fh}(X_f(1)) = v_{fh}(\bar{X}^f_{hf})$

Proof of Lemma 4. Result (A): consider $\lambda_1 > \lambda_2$. There is $\tilde{\lambda} \in (0,1)$ such that $X_f(\lambda_1) = \tilde{\lambda}\bar{X}^f_{hf} + (1-\tilde{\lambda})X_f(\lambda_2)$. But then $v_{fh}(X_f(\lambda_1)) \geq \tilde{\lambda} v_{fh}(\bar{X}^f_{hf}) + (1-\tilde{\lambda}) v_{fh}(X_f(\lambda_2)) > v_{fh}(X_f(\lambda_2))$. The first inequality is due to the concavity of $v_{fh}(\cdot)$ and the last inequality follows from the observation that (i) $\bar{X}^f_{hf}$ is the unique maximizer of $v_{fh}(\cdot)$ for every X such that $v_{hf}(X) \geq 0$ and (ii) $v_{hf}(X_f(\lambda)) \geq 0$ for every $\lambda \in [0,1]$. Similarly, we show that $v_{hf}(X_f(\lambda))$ is strictly decreasing in $\lambda \in [0,1]$. Consider $\lambda_1 > \lambda_2$, there is $\tilde{\lambda} \in (0,1)$ such that $X_f(\lambda_2) = \tilde{\lambda} X_f(\lambda_1) + (1-\tilde{\lambda})\bar{X}^h_{hf}$. The result follows from $v_{hf}(X_f(\lambda_2)) \geq \tilde{\lambda} v_{hf}(X_f(\lambda_1)) + (1-\tilde{\lambda}) v_{hf}(\bar{X}^h_{hf}) > v_{hf}(X_f(\lambda_1))$, where the first inequality is due to the concavity of $v_{hf}(\cdot)$ and the last inequality to the observation that (i) $\bar{X}^h_{hf}$ is the unique maximizer of $v_{hf}(\cdot)$ for every X such that $v_{fh}(X) \geq 0$ and (ii) $v_{fh}(X_f(\lambda)) \geq 0$ for every $\lambda \in [0,1]$. Continuity follows from continuity of $v_{fh}(\cdot)$ and $v_{hf}(\cdot)$. The boundary conditions follow from the fact that $X_f(0) = X^h_{hf}$ and $X_f(1) = X^f_{hf}$. □

Step 1: Let $V_f(w) = w v_{fh}(\bar{X}^f_{hf}) + (1-w) v_{fh}(\bar{X}^h_{hf})$*. There exists a unique mapping* $\lambda_f : [v_{fh}(\bar{X}^h_{hf}), v_{fh}(\bar{X}^f_{hf})] \to [0,1]$ *satisfying the equation* $v_{fh}(X_f(\lambda_f(V_f(w)))) = V_f(w)$ *for every* $w \in [0,1]$*. This mapping is a strictly increasing, bijective, and continuous function and it is such that* $\lambda_f(V_f(w)) \leq w$*,* $\lambda_f(V_f(0)) = 0$*, and* $\lambda_f(V_f(1)) = 1$*.*

Note that $\lambda_f(\cdot)$ such that $v_{fh}(X_f(\lambda_f(V_f(w)))) = V_f(w)$ is the partial inverse of $v_{fh}(X_f(\cdot))$ restricted to its image $[v_{fh}(\bar{X}^h_{hf}), v_{fh}(\bar{X}^f_{hf})]$. From Result (A), $v_{fh}(X_f(\cdot))$ is a strictly increasing and continuous function with image $[v_{fh}(\bar{X}^h_{hf}), v_{fh}(\bar{X}^f_{hf})]$, it thus has a unique, strictly

increasing, continuous and bijective inverse mapping from $[v_{fh}(\bar{X}^h_{hf}), v_{fh}(\bar{X}^f_{hf})]$. To see why $\lambda_f(V_f(w)) \leq w$ for every $w \in [0,1]$, suppose that there is $w \in [0,1]$ such that $\lambda_f(V_f(w)) > w$. Because Result (A) shows that $v_{fh}(X_f(\lambda))$ is strictly increasing in $\lambda \in [0,1]$, this implies that $v_{fh}(X_f(\lambda_f(V_f(w)))) > v_{fh}(X_f(w))$ but by definition $v_{fh}(X_f(\lambda_f(V_f(w)))) = V_f(w)$ and we would thus have $V_f(w) = wv_{fh}(\bar{X}^f_{hf}) + (1-w)v_{fh}(\bar{X}^h_{hf}) > v_{fh}(w\bar{X}^f_{hf} + (1-w)\bar{X}^h_{hf})$ which contradicts the concavity of $v_{fh}(\cdot)$. Finally, note that $X_f(1) = \bar{X}^f_{hf}$ and $V_f(1) = \bar{V}^f_f = v_{fh}(\bar{X}^f_{hf})$, and $X_f(0) = \bar{X}^h_{hf}$ and $V_f(0) = \bar{V}^f_h = v_{fh}(\bar{X}^f_{hf})$, so $\lambda_f(V_f(1)) = 1$ and $\lambda_f(V_f(0)) = 0$. □

Result (B): From Step 1 we know that for every $w \in [0,1]$ there is a unique strictly increasing function $\lambda_f(\cdot)$ such that $v_{fh}(X_f(\lambda_f(V_f(w)))) = V_f(w) = wv_{fh}(\bar{X}^f_{hf}) + (1-w)v_{fh}(\bar{X}^h_{hf})$. Rearranging, we get $\tilde{v}_{fh}(X_f(\lambda_f(V_f(w)))) = w$. Thus, $\tilde{v}^{-1}_{fh}(w) = X_f(\lambda_f(V_f(w)))$ for every $w \in [0,1]$. Note that $V_f(\cdot)$, $\lambda_f(\cdot)$ and $X_f(\cdot)$ are all continuous functions, implying that $\tilde{v}^{-1}_{fh}(\cdot) = X_f(\lambda_f(V_f(\cdot)))$ is a continuous function. Additionally, by construction, $V_f(\cdot)$ and $X_f(\cdot)$ are strictly monotone, and from Step 1 we know that $\lambda_f(\cdot)$ is strictly monotone. Moreover, $X_f(\lambda_f(V_f(0))) = X_f(0) = \bar{X}^h_{hf}$ and $X_f(\lambda_f(V_f(1))) = X_f(1) = \bar{X}^f_{hf}$. We thus conclude that $\tilde{v}^{-1}_{fh} : [0,1] \to conv(\{\bar{X}^h_{hf}, \bar{X}^f_{hf}\})$ is a bijection.

Results (C) and (D): From Step 1 and Result (B) we know that $v_{hf}(\tilde{v}^{-1}_{fh}(w)) = v_{hf}(X_f(\lambda_f(V_f(w))))$ and $v_{fh}(\tilde{v}^{-1}_{fh}(w)) = v_{fh}(X_f(\lambda_f(V_f(w))))$. Since they are compositions of continuous functions, they are both continuous. By Step 1 we know that $\lambda_f(V_f(w)) \in [0,1]$ is strictly increasing for every $w \in [0,1]$. From Result (A) we know $v_{hf}(X_f(\lambda))$ is strictly decreasing for every $\lambda \in [0,1]$ and $v_{fh}(X_f(\lambda))$ is strictly increasing for every $\lambda \in [0,1]$. Combining the two, we conclude that $v_{hf}(\tilde{v}^{-1}_{fh}(w))$ is strictly decreasing for every $w \in [0,1]$ and that $v_{fh}(\tilde{v}^{-1}_{fh}(w))$ is strictly increasing for every $w \in [0,1]$. Finally, from Result (A) and Step 1, $v_{hf}(\tilde{v}^{-1}_{fh}(w)) \geq v_{hf}(X_f(w))$ and $v_{fh}(\tilde{v}^{-1}_{fh}(w)) \leq v_{fh}(X_f(w))$ for every $w \in [0,1]$. Moreover, $v_{hf}(\tilde{v}^{-1}_{fh}(0)) = v_{hf}(\bar{X}^h_{hf})$ and $v_{fh}(\tilde{v}^{-1}_{fh}(0)) = v_{fh}(\bar{X}^h_{hf})$; $v_{hf}(\tilde{v}^{-1}_{fh}(1)) = v_{hf}(\bar{X}^f_{hf})$ and $v_{fh}(\tilde{v}^{-1}_{fh}(1)) = v_{fh}(\bar{X}^f_{hf})$.

■

Proof of Lemma 1. By Step 2 in Theorem 1, we know that the only offers satisfying $r_{fh}(\mathbf{X}, \zeta) = r_{-f,h}(\mathbf{X}, \zeta) = 1$ for every ζ must be such that there is $f \in \{F, G\}$ receiving $\tilde{v}_{fh}(X_{hf}) \geq \frac{w_{fh}+w_{-f,h}}{w_h+w_{Fh}+w_{Gh}}$ and $-f \in \{F,G\} \setminus \{f\}$ receiving $\tilde{v}_{-f,h}(X_{h,-f}) \geq \frac{w_{-f,h}}{w_h+w_{Fh}+w_{Gh}}$. By Lemma 3, we can further restrict attention to offers $\mathbf{X}$ such that $X_{hf} \in conv(\{\bar{X}^h_{hf}, \bar{X}^f_{hf}\})$ for every f. For every $\mathbf{X}'$ such that $X'_{hf} \in conv(\{\bar{X}^h_{hf}, \bar{X}^f_{hf}\})$ and such that $X'_{hf} = \tilde{v}^{-1}_{fh}(w')$ with $w' > \frac{w_{fh}+w_{-f,h}}{w_h+w_{Fh}+w_{Gh}}$, h can obtain a strictly greater expected payoff by deviating to $\mathbf{X}''$ such that $X''_{hf} = \tilde{v}^{-1}_{fh}(w'')$ with $w' > w'' > \frac{w_{fh}+w_{-f,h}}{w_h+w_{Fh}+w_{Gh}}$. Indeed, by Lemma 4, we know that $v_{hf}(\tilde{v}^{-1}_{hf}(w'')) > v_{hf}(\tilde{v}^{-1}_{hf}(w'))$, thus $v_{hf}(X''_{hf}) > v_{hf}(X'_{hf})$. In the same way, for every $\mathbf{X}'$ such that $X'_{h,-f} \in conv(\{\bar{X}^h_{h,-f}, \bar{X}^f_{h,-f}\})$ and $X'_{h,-f} = \tilde{v}^{-1}_{fh}(w')$ with $w' > \frac{w_{-f,h}}{w_h+w_{Fh}+w_{Gh}}$, h

can obtain a strictly greater expected payoff by deviating to $\mathbf{X}''$ where $X''_{h,-f}$ is such that $X''_{h,-f} = \tilde{v}^{-1}_{-f,h}(w'')$ with $w' > w'' > \frac{w_{-f,h}}{w_h + w_{Fh} + w_{Gh}}$.
■

Proof of Theorem 2. From Lemma 1 we know that in every p_ζ-SPE where $r_{fh}(\mathbf{X}, \zeta) = 1$ for every ζ and f, offers $\mathbf{X}$ must be either (A) $\mathbf{X}^F$ or (B) $\mathbf{X}^G$, where $\mathbf{X}^f$ is such that for every $f \in \{F, G\}$ and $-f \in \{F, G\} \setminus \{f\}$,

$$X^f_{fh} = \tilde{v}^{-1}_{fh}\left(\frac{w_{fh} + w_{-f,h}}{w_{fh} + w_{-f,h} + w_h}\right)$$
$$X^{-f}_{fh} = \tilde{v}^{-1}_{fh}\left(\frac{w_{fh}}{w_{fh} + w_{-f,h} + w_h}\right)$$

Step 1: if h prefers $\mathbf{X}^f$ over $\mathbf{X}^{-f}$ for $\mathbf{w}$, then h prefers $\mathbf{X}^f$ over $\mathbf{X}^{-f}$ for $\mathbf{w}'$, where $w'_{fh} > w_{fh}$, $w'_{-f,h} + w'_{fh} = w_{-f,h} + w_{fh}$ and $w'_h = w_h$.

Proof of Step 1. h's payoff from $\mathbf{X}^f$ is greater or equal than h's payoff from $\mathbf{X}^{-f}$ if and only if

$$v_{hf}\left(X^f_f\right) + v_{h,-f}\left(X^f_{-f}\right) \geq v_{h,-f}\left(X^{-f}_{-f}\right) + v_{hf}\left(X^{-f}_f\right)$$

Re-arranging, this is true if and only if

$$v_{hf}\left(\tilde{v}^{-1}_{fh}\left(\frac{w_{fh} + w_{-f,h}}{w_{fh} + w_{-f,h} + w_h}\right)\right) - v_{hf}\left(\tilde{v}^{-1}_{fh}\left(\frac{w_{fh}}{w_{fh} + w_{-f,h} + w_h}\right)\right) \geq$$
$$v_{h,-f}\left(\tilde{v}^{-1}_{-f,h}\left(\frac{w_{fh} + w_{-f,h}}{w_{fh} + w_{-f,h} + w_h}\right)\right) - v_{h,-f}\left(\tilde{v}^{-1}_{-f,h}\left(\frac{w_{-f,h}}{w_{fh} + w_{-f,h} + w_h}\right)\right) \tag{7}$$

By definition of $\mathbf{w}$ and $\mathbf{w}'$ we have $v_{hf}\left(\tilde{v}^{-1}_{fh}\left(\frac{w_{fh}+w_{-f,h}}{w_{fh}+w_{-f,h}+w_h}\right)\right) = v_{hf}\left(\tilde{v}^{-1}_{fh}\left(\frac{w'_{fh}+w'_{-f,h}}{w'_{fh}+w'_{-f,h}+w_h}\right)\right)$ and, by Lemma 4, $v_{hf}\left(\tilde{v}^{-1}_{fh}\left(\frac{w_{fh}}{w_{fh}+w_{-f,h}+w_h}\right)\right) > v_{hf}\left(\tilde{v}^{-1}_{fh}\left(\frac{w'_{fh}}{w'_{fh}+w'_{-f,h}+w_h}\right)\right)$, we thus have

$$v_{hf}\left(\tilde{v}^{-1}_{fh}\left(\frac{w'_{fh} + w'_{-f,h}}{w'_{fh} + w'_{-f,h} + w_h}\right)\right) - v_{hf}\left(\tilde{v}^{-1}_{fh}\left(\frac{w'_{fh}}{w'_{fh} + w'_{-f,h} + w_h}\right)\right) > \tag{8}$$
$$v_{hf}\left(\tilde{v}^{-1}_{fh}\left(\frac{w_{fh} + w_{-f,h}}{w_{fh} + w_{-f,h} + w_h}\right)\right) - v_{hf}\left(\tilde{v}^{-1}_{fh}\left(\frac{w_{fh}}{w_{fh} + w_{-f,h} + w_h}\right)\right) \tag{9}$$

Similarly, by definition of $\mathbf{w}$ and $\mathbf{w}'$ we have $v_{h,-f}\left(\tilde{v}^{-1}_{-f,h}\left(\frac{w_{fh}+w_{-f,h}}{w_{fh}+w_{-f,h}+w_h}\right)\right) = v_{h,-f}\left(\tilde{v}^{-1}_{-f,h}\left(\frac{w'_{fh}+w'_{-f,h}}{w'_{fh}+w'_{-f,h}+w_h}\right)\right)$, and, by Lemma 4, $v_{h,-f}\left(\tilde{v}^{-1}_{-f,h}\left(\frac{w_{-f,h}}{w_{fh}+w_{-f,h}+w_h}\right)\right) < v_{h,-f}\left(\tilde{v}^{-1}_{-f,h}\left(\frac{w'_{-f,h}}{w'_{fh}+w'_{-f,h}+w_h}\right)\right)$, we thus

have

$$v_{h,-f}\left(\tilde{v}_{-f,h}^{-1}\left(\frac{w_{fh}+w_{-f,h}}{w_{fh}+w_{-f,h}+w_h}\right)\right)-v_{h,-f}\left(\tilde{v}_{-f,h}^{-1}\left(\frac{w_{-f,h}}{w_{fh}+w_{-f,h}+w_h}\right)\right)>$$
$$v_{h,-f}\left(\tilde{v}_{-f,h}^{-1}\left(\frac{w'_{fh}+w'_{-f,h}}{w'_{fh}+w'_{-f,h}+w_h}\right)\right)-v_{h,-f}\left(\tilde{v}_{-f,h}^{-1}\left(\frac{w'_{-f,h}}{w'_{fh}+w'_{-f,h}+w_h}\right)\right)$$

Together, the above inequalities establish Step 1. □

Step 2: Given $w_{fh}+w_{-f,h}$, for every h there is a threshold $\bar{W}\in(0,1)$ such that h prefers $\mathbf{X}^F$ over $\mathbf{X}^G$ if and only if $\frac{w_{fh}}{w_{fh}+w_{-f,h}}\geq\bar{W}$.

Proof of Step 2. As we have seen in Lemma 4, $v_{hf}(\tilde{v}_{fh}^{-1}(w))$ is a continuous function, so if $w_{fh}\to w_{fh}+w_{-f,h}$, then the LHS of Equation 7 converges to zero. By Lemma 4, if $w_{fh}\to w_{fh}+w_{-f,h}$ and $w_{fh}+w_{-f,h}$ is fixed, then the RHS of Equation 7 converges to $v_{h,-f}\left(\tilde{v}_{-f,h}^{-1}\left(\frac{w_{fh}+w_{-f,h}}{w_{fh}+w_{-f,h}+w_h}\right)\right)-v_{h,-f}\left(\bar{X}_{hf}^{h}\right)<0$. Taken together, these imply that there is $0<w_{fh}<w_{fh}+w_{-f,h}$ such that h prefers $\mathbf{X}^f$ over $\mathbf{X}^{-f}$ for $f\in\{F,G\}$. Together with Step 1, this implies that there is $\bar{W}\in(0,1)$ such that h prefers $\mathbf{X}^f$ over $\mathbf{X}^{-f}$ if and only if $\frac{w_{fh}}{w_{fh}+w_{-f,h}+w_h}\geq\bar{W}$ for $f\in\{F,G\}$. □

■

Proof of Lemma 2. As for Lemma 1, we show the contrapositive. By Step 2 in Theorem 1, we know that the only offers satisfying $r_{fh}(\mathbf{X},\zeta)=r_{-f,h}(\mathbf{X},\zeta)=\zeta$ for every ζ must be such that there is $f\in\{F,G\}$ receiving $\tilde{v}_{fh}(X_{hf})\geq\frac{w_{fh}}{w_h+w_{Fh}+w_{Gh}}$ for every f. By Lemma 3, we can further restrict attention to offers $\mathbf{X}$ such that $X_{hf}\in conv(\{\bar{X}_{hf}^{h},\bar{X}_{hf}^{f}\})$ for every f. For every $\mathbf{X}'$ such that $X'_{hf}\in conv(\{\bar{X}_{hf}^{h},\bar{X}_{hf}^{f}\})$ and $X'_{hf}=\tilde{v}_{fh}^{-1}(w')$ with $w'>\frac{w_{fh}}{w_h+w_{Fh}+w_{Gh}}$, h can obtain a strictly greater expected payoff by deviating to $\mathbf{X}''$ such that $X''_{hf}=\tilde{v}_{fh}^{-1}(w'')$ with $w'>w''>\frac{w_{fh}}{w_h+w_{Fh}+w_{Gh}}$. Indeed, by Lemma 4, we know that $v_{hf}(\tilde{v}_{hf}^{-1}(w''))>v_{hf}(\tilde{v}_{hf}^{-1}(w'))$, thus $v_{hf}(X''_{hf})>v_{hf}(X'_{hf})$.

■

Proof of Theorem 3. Let $U_h(\mathbf{X}^f)$ and $U_h(\mathbf{X}^{PI})=p_\zeta U_h(\mathbf{X}^{PI},0)+(1-p_\zeta)U_h(\mathbf{X}^{PI},1)$ be h's expected payoff from $\mathbf{X}^f$ and from $\mathbf{X}^{PI}$, respectively. From the results in Lemma 1, 2, and 4, we know that $U_h(\mathbf{X}^{PI},1)>U_h(\mathbf{X}^f)>U_h(\mathbf{X}^{PI},0)$ for every f. Without loss of

generality, assume that $U_h(\mathbf{X}^f) \geq U_h(\mathbf{X}^{-f})$. Country h prefers $\mathbf{X}^f$ over $\mathbf{X}^{PI}$ iff

$$\begin{aligned} U_h(\mathbf{X}^f) &\geq p_\zeta U_h(\mathbf{X}^{PI}, 0) + (1 - p_\zeta) U_h(\mathbf{X}^{PI}, 1) \\ \iff \frac{U_h(\mathbf{X}^f) - U_h(\mathbf{X}^{PI}, 0)}{U_h(\mathbf{X}^{PI}, 1) - U_h(\mathbf{X}^{PI}, 0)} &\geq 1 - p_\zeta \\ \iff \frac{U_h(\mathbf{X}^{PI}, 1) - U_h(\mathbf{X}^f)}{U_h(\mathbf{X}^{PI}, 1) - U_h(\mathbf{X}^f) + U_h(\mathbf{X}^f) - U_h(\mathbf{X}^{PI}, 0)} &\leq p_\zeta \end{aligned} \tag{10}$$

Only the RHS depends on p_ζ and the LHS is strictly positive for every $\mathbf{w} > 0$, so the existence of a threshold $\overline{p}_\zeta > 0$ such that the above inequality holds for every $p_\zeta \geq \overline{p}_\zeta$ follows trivially from $\infty > U_h(\mathbf{X}^{PI}, 1) > U_h(\mathbf{X}^f) > U_h(\mathbf{X}^{PI}, 0)$. We are thus left to show that $\overline{p}_\zeta < \frac{1}{2}$. In terms of the function in Lemma 4, we can write $U_h(\mathbf{X}^{PI}, 0)$ as

$$\begin{aligned} &\frac{w_{fh} + w_{-f,h}}{w_{fh} + w_{-f,h} + w_h} \left[v_{hF}\left(\tilde{v}_{Fh}^{-1}(1)\right) + v_{hG}\left(\tilde{v}_{Gh}^{-1}(1)\right) \right] + \\ &+ \left(1 - \frac{w_{fh} + w_{-f,h}}{w_{fh} + w_{-f,h} + w_h}\right) \left[v_{hF}\left(\tilde{v}_{Fh}^{-1}(0)\right) + v_{hG}\left(\tilde{v}_{Gh}^{-1}(0)\right) \right] - \delta_{Fh} - \delta_{Gh} \end{aligned}$$

From Lemma 4, we know that $v_{hf}(\tilde{v}_{fh}^{-1}(w)) \geq v_{hf}(X_f(w))$ for every $w \in [0,1]$ and $f \in \{F, G\}$. Moreover, recall that $v_{hf}(\cdot)$ is concave and $\delta_{fh} > 0$ for every $f \in \{F, G\}$. We thus have

$$\begin{aligned} &v_{hf}\left(\tilde{v}_{fh}^{-1}\left(\frac{w_{fh} + w_{-f,h}}{w_{fh} + w_{-f,h} + w_h}\right)\right) + v_{h,-f}\left(\tilde{v}_{-f,h}^{-1}\left(\frac{w_{fh} + w_{-f,h}}{w_{fh} + w_{-f,h} + w_h}\right)\right) \\ &\geq v_{hf}\left(X_f\left(\frac{w_{fh} + w_{-f,h}}{w_{fh} + w_{-f,h} + w_h}\right)\right) + v_{h,-f}\left(X_{-f}\left(\frac{w_{fh} + w_{-f,h}}{w_{fh} + w_{-f,h} + w_h}\right)\right) \\ &\geq \frac{w_{fh} + w_{-f,h}}{w_{fh} + w_{-f,h} + w_h} \left[v_{hF}\left(X_F(1)\right) + v_{hG}\left(X_G(1)\right) \right] + \\ &+ \left(1 - \frac{w_{fh} + w_{-f,h}}{w_{fh} + w_{-f,h} + w_h}\right) \left[v_{hF}\left(X_F(0)\right) + v_{hG}\left(X_G(0)\right) \right] \\ &> \frac{w_{fh} + w_{-f,h}}{w_{fh} + w_{-f,h} + w_h} \left[v_{hF}\left(X_F(1)\right) + v_{hG}\left(X_G(1)\right) \right] + \\ &+ \left(1 - \frac{w_{fh} + w_{-f,h}}{w_{fh} + w_{-f,h} + w_h}\right) \left[v_{hF}\left(X_F(0)\right) + v_{hG}\left(X_G(0)\right) \right] - \delta_{hF} + \delta_{Gh} \\ &= U_h(\mathbf{X}^{PI}, 0) \end{aligned}$$

Writing $U_h(\mathbf{X}^f)$ in terms of the functions of Lemma 4 and using the inequality above,

$$U_h(\mathbf{X}^f) - U_h(\mathbf{X}^{PI}, 0) > v_{h,-f}\left(\tilde{v}_{-f,h}^{-1}\left(\frac{w_{-f,h}}{w_{fh} + w_{-f,h} + w_h}\right)\right) - v_{h,-f}\left(\tilde{v}_{-f,h}^{-1}\left(\frac{w_{fh} + w_{-f,h}}{w_{fh} + w_{-f,h} + w_h}\right)\right) \tag{11}$$

Writing also $U_h(\mathbf{X}^{PI}, 1)$ in terms of the functions of Lemma 4, the other term in the denominator (also in the numerator) of Equation 10 is

$$U_h(\mathbf{X}^{PI}, 1) - U_h(\mathbf{X}^f) = v_{hf}\left(\tilde{v}_{fh}^{-1}\left(\frac{w_{fh}}{w_{fh} + w_{-f,h} + w_h}\right)\right) - v_{hf}\left(\tilde{v}_{fh}^{-1}\left(\frac{w_{fh} + w_{-f,h}}{w_{fh} + w_{-f,h} + w_h}\right)\right) \tag{12}$$

As we are assuming that f is the optimal alignment, by Equation 7 we know that

$$\begin{aligned} &v_{hf}\left(\tilde{v}_{fh}^{-1}\left(\frac{w_{fh}}{w_{fh} + w_{-f,h} + w_h}\right)\right) - v_{hf}\left(\tilde{v}_{fh}^{-1}\left(\frac{w_{fh} + w_{-f,h}}{w_{fh} + w_{-f,h} + w_h}\right)\right) \\ &\leq v_{h,-f}\left(\tilde{v}_{-f,h}^{-1}\left(\frac{w_{-f,h}}{w_{fh} + w_{-f,h} + w_h}\right)\right) - v_{h,-f}\left(\tilde{v}_{-f,h}^{-1}\left(\frac{w_{fh} + w_{-f,h}}{w_{fh} + w_{-f,h} + w_h}\right)\right) \end{aligned}$$

Combined with Equations 11 and 12, we deduce

$$U_h(\mathbf{X}^{PI}, 1) - U_h(\mathbf{X}^f) < U_h(\mathbf{X}^f) - U_h(\mathbf{X}^{PI}, 0)$$

And thus

$$\frac{U_h(\mathbf{X}^{PI}, 1) - U_h(\mathbf{X}^f)}{U_h(\mathbf{X}^{PI}, 1) - U_h(\mathbf{X}^f) + U_h(\mathbf{X}^f) - U_h(\mathbf{X}^{PI}, 0)} < \frac{U_h(\mathbf{X}^{PI}, 1) - U_h(\mathbf{X}^f)}{2\left[U_h(\mathbf{X}^{PI}, 1) - U_h(\mathbf{X}^f)\right]} = \frac{1}{2}$$

Leading to the conclusion that $\overline{p}_\zeta < \frac{1}{2}$.

■

Proof of Proposition 1. From Theorem 2 we know that for fixed $w_{Fh} + w_{Gh}$ there is $\bar{W} \in (0, 1)$ such that h prefers $\mathbf{X}^f$ over $\mathbf{X}^{-f}$ whenever $\frac{w_{fh}}{w_{Fh}+w_{Gh}} \geq \bar{W}$. Without loss of generality, assume that $\frac{w_{fh}}{w_{Fh}+w_{Gh}} \geq \bar{W}$.

Step 1: if h prefers $\mathbf{X}^f$ to $\mathbf{X}^{PI}$ for $\mathbf{w}$, then it will still prefer $\mathbf{X}^f$ to $\mathbf{X}^{PI}$ for $\mathbf{w}'$ such that $\frac{w'_{fh}}{w'_{Fh}+w'_{Gh}} > \frac{w_{fh}}{w_{Fh}+w_{Gh}}$ where $w'_{Fh} + w'_{Gh} = w_{Fh} + w_{Gh}$ and $w'_h = w_h$.

Re-arranging from Equation 10, country h prefers $\mathbf{X}^f$ over $\mathbf{X}^{PI}$ iff

$$U_h(\mathbf{X}^f) - (1 - p_\zeta)U_h(\mathbf{X}^{PI}, 1) \geq p_\zeta U_h(\mathbf{X}^{PI}, 0) \tag{13}$$

Starting from the LHS and using the functions defined in Lemma 4, we note that

$$
\begin{aligned}
&U_h(\mathbf{X}^f;\mathbf{w}) - (1-p_\zeta)U_h(\mathbf{X}^{PI},1;\mathbf{w}) = \\
&= v_{hf}\left(\tilde{v}_{fh}^{-1}\left(\frac{w_{fh}+w_{-f,h}}{w_{Fh}+w_{Gh}+w_h}\right)\right) - (1-p_\zeta)v_{hf}\left(\tilde{v}_{fh}^{-1}\left(\frac{w_{fh}}{w_{Fh}+w_{Gh}+w_h}\right)\right) + \\
&+ p_\zeta v_{h,-f}\left(\tilde{v}_{-f,h}^{-1}\left(\frac{w_{-f,h}}{w_{Fh}+w_{Gh}+w_h}\right)\right)
\end{aligned}
$$

As a consequence,

$$
\begin{aligned}
&\left[U_h(\mathbf{X}^f;\mathbf{w}') - (1-p_\zeta)U_h(\mathbf{X}^{PI},1;\mathbf{w}')\right] - \left[U_h(\mathbf{X}^f;\mathbf{w}) - (1-p_\zeta)U_h(\mathbf{X}^{PI},1;\mathbf{w})\right] = \\
&= (1-p_\zeta)v_{hf}\left(\tilde{v}_{fh}^{-1}\left(\frac{w_{fh}}{w_{Fh}+w_{Gh}+w_h}\right)\right) - (1-p_\zeta)v_{hf}\left(\tilde{v}_{fh}^{-1}\left(\frac{w'_{fh}}{w_{Fh}+w_{Gh}+w_h}\right)\right) + \\
&+ p_\zeta v_{h,-f}\left(\tilde{v}_{-f,h}^{-1}\left(\frac{w'_{-f,h}}{w_{Fh}+w_{Gh}+w_h}\right)\right) - p_\zeta v_{h,-f}\left(\tilde{v}_{-f,h}^{-1}\left(\frac{w_{-f,h}}{w_{Fh}+w_{Gh}+w_h}\right)\right) \\
&> 0
\end{aligned}
$$

Where the inequality follows from the result that $v_{hf}(\tilde{v}_{fh}^{-1}(\cdot))$ and $v_{h,-f}(\tilde{v}_{-f,h}^{-1}(\cdot))$ are strictly decreasing (Lemma 4), and the fact that the definitions of $\mathbf{w}'$ and $\mathbf{w}$ imply $\frac{w'_{fh}}{w'_{Fh}+w'_{Gh}+w'_h} > \frac{w_{fh}}{w_{Fh}+w_{Gh}+w_h}$ and $\frac{w'_{-f,h}}{w'_{Fh}+w'_{Gh}+w'_h} < \frac{w_{-f,h}}{w_{Fh}+w_{Gh}+w_h}$. Moving to the RHS of Equation 13, we see that

$$
\begin{aligned}
p_\zeta U_h(\mathbf{X}^{PI},0;\mathbf{w}) = &\frac{w_{fh}+w_{-f,h}}{w_{fh}+w_{-f,h}+w_h}\left[v_{hF}\left(X_F\left(1\right)\right)+v_{hG}\left(X_G\left(1\right)\right)\right]+ \\
&+\left(1-\frac{w_{fh}+w_{-f,h}}{w_{fh}+w_{-f,h}+w_h}\right)\left[v_{hF}\left(X_F\left(0\right)\right)+v_{hG}\left(X_G\left(0\right)\right)\right]-\delta_{hF}+\delta_{Gh}
\end{aligned}
$$

implying $p_\zeta U_h(\mathbf{X}^{PI},0;\mathbf{w}) = p_\zeta U_h(\mathbf{X}^{PI},0;\mathbf{w}')$. Thus, if $U_h(\mathbf{X}^f;\mathbf{w}) - (1-p_\zeta)U_h(\mathbf{X}^{PI},1;\mathbf{w}) \geq p_\zeta U_h(\mathbf{X}^{PI},0;\mathbf{w})$ then $U_h(\mathbf{X}^f;\mathbf{w}') - (1-p_\zeta)U_h(\mathbf{X}^{PI},1;\mathbf{w}') > p_\zeta U_h(\mathbf{X}^{PI},0;\mathbf{w}')$. □

To conclude the proof, we define the relevant thresholds in the following way. (i) If there is $\frac{\tilde{w}_{fh}}{w_{Fh}+w_{Gh}} \in [\bar{W},1]$ such that f is indifferent between $\mathbf{X}^f$ and $\mathbf{X}^{PI}$, set $\bar{W}^f = \frac{\tilde{w}_{fh}}{w_{Fh}+w_{Gh}}$. (ii) If, instead, for every $\frac{w_{fh}}{w_{Fh}+w_{Gh}} \in [\bar{W},1]$, f strictly prefers $\mathbf{X}^{PI}$ to $\mathbf{X}^f$, set $\bar{W}^f = 1$. Finally, (iii) if for every $\frac{w_{fh}}{w_{Fh}+w_{Gh}} \in [\bar{W},1]$, f strictly prefers $\mathbf{X}^f$ to $\mathbf{X}^{PI}$, set $\bar{W}^f = \bar{W}$. Based on these definitions and Step 1 above, we see that h chooses $\mathbf{X}^f$, thus $h \in S_f$, if $\frac{w_{fh}}{w_{Fh}+w_{Gh}} > \bar{W}^f$ for given $w_{Fh}+w_{Gh}$ and for every $f \in \{F,G\}$ and $h \in N$ if $1-\bar{W}^{-f} < \frac{w_{fh}}{w_{Fh}+w_{Gh}} < \bar{W}^f$.

■

Proof of Proposition 2. Suppose that $w_h = w$, $w_{fh} = w_f$ and $w_{-f,h} = w_{-f}$ for every

$h \in H$, and recall that

$$\sum_{f\in\{F,G\}} \mathcal{G}_f(\tilde{\mathbf{v}}_f) := \sum_{f\in\{F,G\}} \frac{1}{2|H|^2} \sum_{h\in H} \sum_{h'\in H} |\tilde{v}_{hf}(X_{hf}) - \tilde{v}_{h'f}(X_{hf})|$$

For every $h \in S_f$ we have $\tilde{v}_{hf}(X_{hf}) = \frac{w_f+w_{-f}}{w_f+w_{-f}+w_h}$ and for every $h' \notin S_f$, we have $\tilde{v}_{h'f}(X_{hf}) = \frac{w_f}{w_f+w_{-f}+w_h}$. We thus note the following:

$$\frac{1}{2|H|^2} \sum_{h\in S_f} \sum_{h'\in S_f} |\tilde{v}_{hf}(X_{hf}) - \tilde{v}_{h'f}(X_{hf})| = 0$$

$$\frac{1}{2|H|^2} \sum_{h\in S_f} \sum_{h'\notin S_f} |\tilde{v}_{hf}(X_{hf}) - \tilde{v}_{h'f}(X_{hf})| = \frac{|S_f||H\setminus S_f|}{2|H|^2} \frac{w_{-f}}{w_f+w_{-f}+w_h}$$

$$\frac{1}{2|H|^2} \sum_{h\notin S_f} \sum_{h'\notin S_f} |\tilde{v}_{hf}(X_{hf}) - \tilde{v}_{h'f}(X_{hf})| = 0$$

$$\frac{1}{2|H|^2} \sum_{h\notin S_f} \sum_{h'\in S_f} |\tilde{v}_{hf}(X_{hf}) - \tilde{v}_{h'f}(X_{hf})| = \frac{|H\setminus S_f||S_f|}{2|H|^2} \frac{w_{-f}}{w_f+w_{-f}+w_h}$$

We thus conclude that

$$\sum_{f\in\{F,G\}} \mathcal{G}_f(\tilde{\mathbf{v}}_f) = \sum_{f\in\{F,G\}} \frac{|S_f||H\setminus S_f|}{|H|^2} \frac{w_{-f}}{w_f+w_{-f}+w_h}$$

which implies that $\sum_{f\in\{F,G\}} \mathcal{G}_f(\tilde{\mathbf{v}}_f) > 0$ if and only if $|S_F||H\setminus S_F| + |S_G||H\setminus S_G| > 0$ if and only if there is $f \in \{F,G\}$ such that $S_f \neq \emptyset$ and $S_f \neq H$. To conclude the proof, note that if there is no preference heterogeneity, then by Theorem 2, h prefers S_f over S_{-f} if and only if $\frac{w_{-f}}{w_f+w_{-f}+w_h} \leq \frac{1}{2}$. Additionally, we know that $S_f + S_{-f} \leq H$. With these constraints, we see that $\sum_{f\in\{F,G\}} \mathcal{G}_f(\tilde{\mathbf{v}}_f) \leq \frac{1}{4}$, which is the level of geoeconomic fragmentation that corresponds to a situation where $\frac{w_{-f}}{w_f+w_{-f}} = \frac{1}{2}$, $\frac{|S_F|}{|H|} = \frac{|S_G|}{|H|} = \frac{1}{2}$, which by Theorem 2 is feasible for $\frac{w_{-f}}{w_f+w_{-f}} = \frac{1}{2}$ and by Theorem 3 is feasible when $p_\zeta > \frac{1}{2}$ so that $N = \emptyset$.

■

B Robustness checks

B.1 Alternative functional form

Table 3: Table 2 with an alternative functional to estimate the balancing effect

	$\ln(\lvert\hat{Y}_{USA,t+1} - \hat{Y}_{h,t+1}\rvert)$		
	(1)	(2)	(3)
$\ln(\tilde{w}_{USA,t})$	-0.32***	-0.42	-0.26**
	[0.121]	[0.282]	[0.091]
$\ln(\tilde{w}_{USA,t} + \tilde{w}_{RUS,t}) \times \mathbf{1}\{\tilde{w}_{USA,t} > \tilde{w}_{RUS,t}\}$	-0.47**		
	[0.221]		
$\ln(\tilde{w}_{USA,t} + \tilde{w}_{CHN,t}) \times \mathbf{1}\{\tilde{w}_{USA,t} > \tilde{w}_{CHN,t}\}$		-0.31	
		[0.198]	
$\ln(\tilde{w}_{USA,t} + \tilde{w}_{MAX,t}) \times \mathbf{1}\{\tilde{w}_{USA,t} > \tilde{w}_{MAX,t}\}$			-0.23***
			[0.081]
Num. Countries	157	186	157
Observations	7,135	7,687	6,835
Kleibergen-Paap F statistics	6.96	3.95	17.78
Num. excluded instruments	1	1	2
Hansen J test			1.220

Note: the unit of observation is a country h in a given year t. The outcome is the log of the absolute difference in the ideal points of the US and ideal points of country h as measured from their voting behavior at the United Nations General Assembly in year t. Foreign influence $w_{f,t}$ is measured as the log of trade dependence, defined as the value of h's export at time t towards f divided by h's GDP at time t. $w_{MAX,t} := max_{j \in \{RUS,CHN\}}\{w_{j,t}\}$. Column (1) is based on instrument $\mathbf{1}\{1990 \leq t \leq 1994\} \times \mathbf{1}\{w_{USA,1989} > w_{RUS,1989}\}$, column (2) on $\mathbf{1}\{t \geq 2007\} \times \mathbf{1}\{w_{USA,2006} > w_{CHN,2006}\}$, column (3) is based on both. All estimates are conditional on country h fixed effect, year t fixed effects, foreign influence of the competitor of the US ($w_{MAX,t}$, $w_{RUS,t}$, or $w_{CHN,t}$), the log of h's GDP, the log of h's population, the log of h's trade with the rest of the world, the log of h's GDP per capita. Standard errors clustered at the country level are reported in brackets. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

B.2 Alternative controls and fixed effects

Table 4: Table 2 omitting all controls

	$\ln(\lvert\hat{Y}_{USA,t+1} - \hat{Y}_{h,t+1}\rvert)$		
	(1)	(2)	(3)
$w_{USA,t}$	-0.11***	-0.19**	-0.12***
	[0.038]	[0.094]	[0.044]
$w_{RUS,t} \times \mathbf{1}\{w_{USA,t} > w_{RUS,t}\}$	-0.26***		
	[0.090]		
$w_{CHN,t} \times \mathbf{1}\{w_{USA,t} > w_{CHN,t}\}$		-0.25*	
		[0.144]	
$w_{MAX,t} \times \mathbf{1}\{w_{USA,t} > w_{MAX,t}\}$			-0.16***
			[0.053]
Num. Countries	158	188	158
Observations	6,384	7,327	7,064
Kleibergen-Paap F statistics	17.81	3.99	14.48
Num. excluded instruments	1	1	2
Hansen J test			4.251**

Note: the unit of observation is a country h in a given year t. The outcome is the log of the absolute difference in the ideal points of the US and ideal points of country h as measured from their voting behavior at the United Nations General Assembly in year t. Foreign influence $w_{f,t}$ is measured as the log of trade dependence, defined as the value of h's export at time t towards f divided by h's GDP at time t. $w_{MAX,t} := max_{j \in \{RUS,CHN\}}\{w_{j,t}\}$. Column (1) is based on instrument $\mathbf{1}\{1990 \leq t \leq 1994\} \times \mathbf{1}\{w_{USA,1989} > w_{RUS,1989}\}$, column (2) on $\mathbf{1}\{t \geq 2007\} \times \mathbf{1}\{w_{USA,2006} > w_{CHN,2006}\}$, column (3) is based on both. All estimates are conditional on country h fixed effect and year t fixed effects. Standard errors clustered at the country level are reported in brackets. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 5: Table 2 omitting the foreign influence of the competitor of the US

	$\ln(\lvert\hat{Y}_{USA,t+1} - \hat{Y}_{h,t+1}\rvert)$		
	(1)	(2)	(3)
$w_{USA,t}$	-0.17***	-0.27*	-0.17***
	[0.058]	[0.159]	[0.051]
$w_{RUS,t} \times \mathbf{1}\{w_{USA,t} > w_{RUS,t}\}$	-0.24**		
	[0.096]		
$w_{CHN,t} \times \mathbf{1}\{w_{USA,t} > w_{CHN,t}\}$		-0.29	
		[0.184]	
$w_{MAX,t} \times \mathbf{1}\{w_{USA,t} > w_{MAX,t}\}$			-0.16***
			[0.053]
Num. Countries	155	186	157
Observations	6,130	7,086	6,835
Kleibergen-Paap F statistics	11.98	3.23	11.80
Num. excluded instruments	1	1	2
Hansen J test			3.856**

Note: the unit of observation is a country h in a given year t. The outcome is the log of the absolute difference in the ideal points of the US and ideal points of country h as measured from their voting behavior at the United Nations General Assembly in year t. Foreign influence $w_{f,t}$ is measured as the log of trade dependence, defined as the value of h's export at time t towards f divided by h's GDP at time t. $w_{MAX,t} := max_{j \in \{RUS,CHN\}}\{w_{j,t}\}$. Column (1) is based on instrument $\mathbf{1}\{1990 \leq t \leq 1994\} \times \mathbf{1}\{w_{USA,1989} > w_{RUS,1989}\}$, column (2) on $\mathbf{1}\{t \geq 2007\} \times \mathbf{1}\{w_{USA,2006} > w_{CHN,2006}\}$, column (3) is based on both. All estimates are conditional on country h fixed effect, year t fixed effects, the log of h's GDP, the log of h's population, the log of h's trade with the rest of the world, the log of h's GDP per capita. Standard errors clustered at the country level are reported in brackets. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 6: Table 2 omitting year fixed effects

	$\ln(\lvert\hat{Y}_{USA,t+1} - \hat{Y}_{h,t+1}\rvert)$		
	(1)	(2)	(3)
$w_{USA,t}$	-0.18***	-0.14***	-0.15***
	[0.053]	[0.044]	[0.046]
$w_{RUS,t} \times \mathbf{1}\{w_{USA,t} > w_{RUS,t}\}$	-0.24**		
	[0.104]		
$w_{CHN,t} \times \mathbf{1}\{w_{USA,t} > w_{CHN,t}\}$		-0.11**	
		[0.045]	
$w_{MAX,t} \times \mathbf{1}\{w_{USA,t} > w_{MAX,t}\}$			-0.13**
			[0.056]
$\mathbf{1}\{1990 \leq t \leq 1994\}$	0.09***		0.03
	[0.032]		[0.026]
$\mathbf{1}\{t \geq 2007\}$		-0.21***	-0.19***
		[0.048]	[0.055]
Num. Countries	155	186	157
Observations	5,955	7,086	6,835
Kleibergen-Paap F statistics	15.04	24.33	11.80
Num. excluded instruments	1	1	2
Hansen J test			3.856**

Note: the unit of observation is a country h in a given year t. The outcome is the log of the absolute difference in the ideal points of the US and ideal points of country h as measured from their voting behavior at the United Nations General Assembly in year t. Foreign influence $w_{f,t}$ is measured as the log of trade dependence, defined as the value of h's export at time t towards f divided by h's GDP at time t. $w_{MAX,t} := max_{j \in \{RUS,CHN\}}\{w_{j,t}\}$. Column (1) is based on instrument $\mathbf{1}\{1990 \leq t \leq 1994\} \times \mathbf{1}\{w_{USA,1989} > w_{RUS,1989}\}$, column (2) on $\mathbf{1}\{t \geq 2007\} \times \mathbf{1}\{w_{USA,2006} > w_{CHN,2006}\}$, column (3) is based on both. All estimates are conditional on country h fixed effect, h's trade dependence on f's rival, the log of h's GDP, the log of h's population, the log of h's trade with the rest of the world, the log of h's GDP per capita. Standard errors clustered at the country level are reported in brackets. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 7: Table 2 omitting year fixed effects and the foreign influence of the competitor of the US

	$\ln(\lvert\hat{Y}_{USA,t+1} - \hat{Y}_{h,t+1}\rvert)$		
	(1)	(2)	(3)
$w_{USA,t}$	-0.18***	-0.26	-0.17***
	[0.055]	[0.158]	[0.046]
$w_{RUS,t} \times \mathbf{1}\{w_{USA,t} > w_{RUS,t}\}$	-0.22**		
	[0.094]		
$w_{CHN,t} \times \mathbf{1}\{w_{USA,t} > w_{CHN,t}\}$		-0.24	
		[0.179]	
$w_{MAX,t} \times \mathbf{1}\{w_{USA,t} > w_{MAX,t}\}$			-0.15***
			[0.051]
$\mathbf{1}\{1990 \leq t \leq 1994\}$	0.06*		-0.03
	[0.037]		[0.040]
$\mathbf{1}\{t \geq 2007\}$		0.28	-0.01
		[0.411]	[0.101]
Num. Countries	155	186	157
Observations	6,130	7,086	6,835
Kleibergen-Paap F statistics	11.87	2.84	12.15
Num. excluded instruments	1	1	2
Hansen J test			1.882

Note: the unit of observation is a country h in a given year t. The outcome is the log of the absolute difference in the ideal points of the US and ideal points of country h as measured from their voting behavior at the United Nations General Assembly in year t. Foreign influence $w_{f,t}$ is measured as the log of trade dependence, defined as the value of h's export at time t towards f divided by h's GDP at time t. $w_{MAX,t} := max_{j \in \{RUS,CHN\}}\{w_{j,t}\}$. Column (1) is based on instrument $\mathbf{1}\{1990 \leq t \leq 1994\} \times \mathbf{1}\{w_{USA,1989} > w_{RUS,1989}\}$, column (2) on $\mathbf{1}\{t \geq 2007\} \times \mathbf{1}\{w_{USA,2006} > w_{CHN,2006}\}$, column (3) is based on both. All estimates are conditional on country h fixed effect, the log of h's GDP, the log of h's population, the log of h's trade with the rest of the world, the log of h's GDP per capita. Standard errors clustered at the country level are reported in brackets. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

B.3 Alternative instrumental variables

Table 8: Table 2 with alternative instruments: (1) $\mathbf{1}\{1989 \leq t \leq 1999\} \times \mathbf{1}\{w_{USA,1988} > w_{RUS,1988}\}$ and (2) $\mathbf{1}\{t \geq 2001\} \times \mathbf{1}\{w_{USA,2000} > w_{CHN,2000}\}$

	$\ln(\lvert\hat{Y}_{USA,t+1} - \hat{Y}_{h,t+1}\rvert)$		
	(1)	(2)	(3)
$w_{USA,t}$	-0.20***	-0.44	-0.49***
	[0.058]	[0.31]	[0.171]
$w_{RUS,t} \times \mathbf{1}\{w_{USA,t} > w_{RUS,t}\}$	-0.32***		
	[0.117]		
$w_{CHN,t} \times \mathbf{1}\{w_{USA,t} > w_{CHN,t}\}$		-0.46	
		[0.354]	
$w_{MAX,t} \times \mathbf{1}\{w_{USA,t} > w_{MAX,t}\}$			-0.58***
			[0.208]
Num. Countries	154	184	148
Observations	5,942	7,069	6,611
Kleibergen-Paap F statistics	13.18	1.28	5.108
Num. excluded instruments	1	1	2
Hansen J test			0.459

Note: the unit of observation is a country h in a given year t. The outcome is the log of the absolute difference in the ideal points of the US and ideal points of country h as measured from their voting behavior at the United Nations General Assembly in year t. Foreign influence $w_{f,t}$ is measured as the log of trade dependence, defined as the value of h's export at time t towards f divided by h's GDP at time t. $w_{MAX,t} := max_{j \in \{RUS,CHN\}}\{w_{j,t}\}$. Column (1) is based on instrument $\mathbf{1}\{1989 \leq t \leq 1999\} \times \mathbf{1}\{w_{USA,1988} > w_{RUS,1988}\}$, column (2) on $\mathbf{1}\{t \geq 2001\} \times \mathbf{1}\{w_{USA,2000} > w_{CHN,2000}\}$, column (3) is based on both. All estimates are conditional on country h fixed effect, year t fixed effects, foreign influence of the competitor of the US. the log of h's GDP, the log of h's population, the log of h's trade with the rest of the world, the log of h's GDP per capita. Standard errors clustered at the country level are reported in brackets. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 9: Table 2 with alternative instruments: (1) $\mathbf{1}\{1989 \leq t \leq 1999\} \times \mathbf{1}\{w_{USA,1988} > w_{RUS,1988}\}$, (2) $\mathbf{1}\{t \geq 2001\} \times \mathbf{1}\{w_{USA,2000} > w_{CHN,2000}\}$, (3) $\mathbf{1}\{1990 \leq t \leq 1994\} \times \mathbf{1}\{w_{USA,1988} > w_{RUS,1988}\}$, and (4) $\mathbf{1}\{t \geq 2007\} \times \mathbf{1}\{w_{USA,2000} > w_{CHN,2000}\}$

	$\ln(\lvert\hat{Y}_{USA,t+1} - \hat{Y}_{h,t+1}\rvert)$		
	(1)	(2)	(3)
$w_{USA,t}$	-0.20***	-0.08***	-0.18***
	[0.059]	[0.024]	[0.042]
$w_{RUS,t} \times \mathbf{1}\{w_{USA,t} > w_{RUS,t}\}$	-0.32***		
	[0.118]		
$w_{CHN,t} \times \mathbf{1}\{w_{USA,t} > w_{CHN,t}\}$		-0.04**	
		[0.017]	
$w_{MAX,t} \times \mathbf{1}\{w_{USA,t} > w_{MAX,t}\}$			-0.18***
			[0.043]
Num. Countries	154	184	148
Observations	5,942	7,069	6,611
Kleibergen-Paap F statistics	7.28	43.65	19.71
Num. excluded instruments	2	2	4
Hansen J test	0.121	6.878***	15.748***

Note: the unit of observation is a country h in a given year t. The outcome is the log of the absolute difference in the ideal points of the US and ideal points of country h as measured from their voting behavior at the United Nations General Assembly in year t. Foreign influence $w_{f,t}$ is measured as the log of trade dependence, defined as the value of h's export at time t towards f divided by h's GDP at time t. $w_{MAX,t} := max_{j \in \{RUS,CHN\}}\{w_{j,t}\}$. Column (1) is based on instrument $\mathbf{1}\{1989 \leq t \leq 1999\} \times \mathbf{1}\{w_{USA,1988} > w_{RUS,1988}\}$ and $\mathbf{1}\{1990 \leq t \leq 1994\} \times \mathbf{1}\{w_{USA,1988} > w_{RUS,1988}\}$, column (2) on $\mathbf{1}\{t \geq 2001\} \times \mathbf{1}\{w_{USA,2000} > w_{CHN,2000}\}$ and $\mathbf{1}\{t \geq 2007\} \times \mathbf{1}\{w_{USA,2000} > w_{CHN,2000}\}$, column (3) is based on all four. All estimates are conditional on country h fixed effect, year t fixed effects, foreign influence of the competitor of the US ($w_{MAX,t}$, $w_{RUS,t}$, or $w_{CHN,t}$), the log of h's GDP, the log of h's population, the log of h's trade with the rest of the world, the log of h's GDP per capita. Standard errors clustered at the country level are reported in brackets. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

B.4 Alternative measures of foreign influence

Table 10: Table 2 measuring foreign influence from total trade dependence (imports + exports)

	$\ln(\lvert\hat{Y}_{USA,t+1} - \hat{Y}_{h,t+1}\rvert)$		
	(1)	(2)	(3)
$w_{USA,t}$	-0.18***	-0.18***	-0.20***
	[0.060]	[0.064]	[0.074]
$w_{RUS,t} \times \mathbf{1}\{w_{USA,t} > w_{RUS,t}\}$	-0.29**		
	[0.114]		
$w_{CHN,t} \times \mathbf{1}\{w_{USA,t} > w_{CHN,t}\}$		-0.19**	
		[0.080]	
$w_{MAX,t} \times \mathbf{1}\{w_{USA,t} > w_{MAX,t}\}$			-0.27**
			[0.104]
Num. Countries	157	186	158
Observations	6,133	7,180	6,937
Kleibergen-Paap F statistics	12.40	8.98	6.96
Num. excluded instruments	1	1	2
Hansen J test			8.17***

Note: the unit of observation is a country h in a given year t. The outcome is the log of the absolute difference in the ideal points of the US and ideal points of country h as measured from their voting behavior at the United Nations General Assembly in year t. Foreign influence $w_{f,t}$ is measured as the log of total trade dependence which is imports + exports divided by a country's GDP + its total imports. $w_{MAX,t} := max_{j \in \{RUS,CHN\}}\{w_{j,t}\}$. Column (1) is based on instrument $\mathbf{1}\{1990 \leq t \leq 1994\} \times \mathbf{1}\{w_{USA,1989} > w_{RUS,1989}\}$, column (2) on $\mathbf{1}\{t \geq 2007\} \times \mathbf{1}\{w_{USA,2006} > w_{CHN,2006}\}$, column (3) is based on both. All estimates are conditional on country h fixed effect, year t fixed effects, foreign influence of the competitor of the US ($w_{MAX,t}$, $w_{RUS,t}$, or $w_{CHN,t}$), the log of h's GDP, the log of h's population, the log of h's trade with the rest of the world, the log of h's GDP per capita. Standard errors clustered at the country level are reported in brackets. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 11: Table 2 measuring foreign influence from economic dependence (imports + exports + aid)

	$\ln(\lvert\hat{Y}_{USA,t+1} - \hat{Y}_{h,t+1}\rvert)$		
	(1)	(2)	(3)
$w_{USA,t}$	-0.19***	-0.13***	-0.16***
	[0.062]	[0.043]	[0.052]
$w_{RUS,t} \times \mathbf{1}\{w_{USA,t} > w_{RUS,t}\}$	-0.29**		
	[0.117]		
$w_{CHN,t} \times \mathbf{1}\{w_{USA,t} > w_{CHN,t}\}$		-0.14**	
		[0.053]	
$w_{MAX,t} \times \mathbf{1}\{w_{USA,t} > w_{MAX,t}\}$			-0.21***
			[0.073]
Num. Countries	157	186	158
Observations	6,137	7,200	6,955
Kleibergen-Paap F statistics	11.94	18.09	12.38
Num. excluded instruments	1	1	2
Hansen J test			7.98***

Note: the unit of observation is a country h in a given year t. The outcome is the log of the absolute difference in the ideal points of the US and ideal points of country h as measured from their voting behavior at the United Nations General Assembly in year t. Foreign influence $w_{f,t}$ is measured as the log of economic dependence which is imports + exports + received aid divided by a country's GDP + its total imports + received aid. $w_{MAX,t} := max_{j \in \{RUS,CHN\}}\{w_{j,t}\}$. Column (1) is based on instrument $\mathbf{1}\{1990 \leq t \leq 1994\} \times \mathbf{1}\{w_{USA,1989} > w_{RUS,1989}\}$, column (2) on $\mathbf{1}\{t \geq 2007\} \times \mathbf{1}\{w_{USA,2006} > w_{CHN,2006}\}$, column (3) is based on both. All estimates are conditional on country h fixed effect, year t fixed effects, foreign influence of the competitor of the US ($w_{MAX,t}$, $w_{RUS,t}$, or $w_{CHN,t}$), the log of h's GDP, the log of h's population, the log of h's trade with the rest of the world, the log of h's GDP per capita. Standard errors clustered at the country level are reported in brackets. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 12: Table 2 measuring foreign influence using Moyer et al. (2018)'s formal bilateral influence capabilities index

	$\ln(\lvert\hat{Y}_{USA,t+1} - \hat{Y}_{h,t+1}\rvert)$		
	(1)	(2)	(3)
$w_{USA,t}$	-0.18***	-0.12***	-0.14***
	[0.048]	[0.037]	[0.040]
$w_{RUS,t} \times \mathbf{1}\{w_{USA,t} > w_{RUS,t}\}$	-0.29***		
	[0.093]		
$w_{CHN,t} \times \mathbf{1}\{w_{USA,t} > w_{CHN,t}\}$		-0.06	
		[0.050]	
$w_{MAX,t} \times \mathbf{1}\{w_{USA,t} > w_{MAX,t}\}$			-0.18***
			[0.067]
Num. Countries	160	186	160
Observations	6,595	7,387	7,098
Kleibergen-Paap F statistics	30.90	15.30	16.00
Num. excluded instruments	1	1	2
Hansen J test			10.476***

Note: the unit of observation is a country h in a given year t. The outcome is the log of the absolute difference in the ideal points of the US and ideal points of country h as measured from their voting behavior at the United Nations General Assembly in year t. Foreign influence $w_{f,t}$ is measured as the log of the formal bilateral influence capabilities index defined in Moyer et al. (2018). $w_{MAX,t} := max_{j \in \{RUS,CHN\}}\{w_{j,t}\}$. Column (1) is based on instrument $\mathbf{1}\{1990 \leq t \leq 1994\} \times \mathbf{1}\{w_{USA,1989} > w_{RUS,1989}\}$, column (2) on $\mathbf{1}\{t \geq 2007\} \times \mathbf{1}\{w_{USA,2006} > w_{CHN,2006}\}$, column (3) is based on both. All estimates are conditional on country h fixed effect, year t fixed effects, foreign influence of the competitor of the US ($w_{MAX,t}$, $w_{RUS,t}$, or $w_{CHN,t}$), the log of h's GDP, the log of h's population, the log of h's trade with the rest of the world, the log of h's GDP per capita. Standard errors clustered at the country level are reported in brackets. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

B.5 Alternative outcomes

Table 13: Table 2 but the outcome is not logged

	$\lvert\hat{Y}_{USA,t+1} - \hat{Y}_{h,t+1}\rvert$		
	(1)	(2)	(3)
$w_{USA,t}$	-0.46***	-0.34***	-0.41***
	[0.120]	[0.081]	[0.097]
$w_{RUS,t} \times \mathbf{1}\{w_{USA,t} > w_{RUS,t}\}$	-0.65**		
	[0.248]		
$w_{CHN,t} \times \mathbf{1}\{w_{USA,t} > w_{CHN,t}\}$		-0.26***	
		[0.076]	
$w_{MAX,t} \times \mathbf{1}\{w_{USA,t} > w_{MAX,t}\}$			-0.37***
			[0.108]
Num. Countries	155	186	157
Observations	5,955	7,086	6,835
Kleibergen-Paap F statistics	15.23	25.55	20.83
Num. excluded instruments	1	1	2
Hansen J test			3.396*

Note: the unit of observation is a country h in a given year t. The outcome is the absolute value of the difference in the ideal points of the US and country h as measured from their voting behavior at the United Nations General Assembly in year t. Foreign influence $w_{f,t}$ is measured as the log of trade dependence, defined as the value of h's export at time t towards f divided by h's GDP at time t. $w_{MAX,t} := max_{j \in \{RUS,CHN\}}\{w_{j,t}\}$. Column (1) is based on instrument $\mathbf{1}\{1990 \leq t \leq 1994\} \times \mathbf{1}\{w_{USA,1989} > w_{RUS,1989}\}$, column (2) on $\mathbf{1}\{t \geq 2007\} \times \mathbf{1}\{w_{USA,2006} > w_{CHN,2006}\}$, column (3) is based on both. All estimates are conditional on country h fixed effect, year t fixed effects, foreign influence of the competitor of the US ($w_{MAX,t}$, $w_{RUS,t}$, or $w_{CHN,t}$), the log of h's GDP, the log of h's population, the log of h's trade with the rest of the world, the log of h's GDP per capita. Standard errors clustered at the country level are reported in brackets. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 14: Table 2 using the s2-score for voting similarity rather than the ideal point distance

	$S_2(Y_{USA,t+1}, Y_{h,t+1})$		
	(1)	(2)	(3)
$w_{USA,t}$	0.13***	0.08***	0.12***
	[0.04]	[0.026]	[0.032]
$w_{RUS,t} \times \mathbf{1}\{w_{USA,t} > w_{RUS,t}\}$	0.21**		
	[0.086]		
$w_{CHN,t} \times \mathbf{1}\{w_{USA,t} > w_{CHN,t}\}$		0.07***	
		[0.025]	
$w_{MAX,t} \times \mathbf{1}\{w_{USA,t} > w_{MAX,t}\}$			0.11***
			[0.036]
Num. Countries	155	186	157
Observations	5,896	7,032	6,764
Kleibergen-Paap F statistics	14.63	25.65	20.67
Num. excluded instruments	1	1	2
Hansen J test			5.351**

Note: the unit of observation is a country h in a given year t. The outcome is the s2 voting similarity score between country h and the United States as measured from their voting behavior at the United Nations General Assembly in year t. The s2 score assigns 0 each time a country abstains and the other does not, +1 when they agree, -1 when they disagree, and scales the yearly score by the number of votes in that year. Foreign influence $w_{f,t}$ is measured as the log of trade dependence, defined as the value of h's export at time t towards f divided by h's GDP at time t. $w_{MAX,t} := max_{j \in \{RUS,CHN\}}\{w_{j,t}\}$. Column (1) is based on instrument $\mathbf{1}\{1990 \leq t \leq 1994\} \times \mathbf{1}\{w_{USA,1989} > w_{RUS,1989}\}$, column (2) on $\mathbf{1}\{t \geq 2007\} \times \mathbf{1}\{w_{USA,2006} > w_{CHN,2006}\}$, column (3) is based on both. All estimates are conditional on country h fixed effect, year t fixed effects, foreign influence of the competitor of the US ($w_{MAX,t}$, $w_{RUS,t}$, or $w_{CHN,t}$), the log of h's GDP, the log of h's population, the log of h's trade with the rest of the world, the log of h's GDP per capita. Standard errors clustered at the country level are reported in brackets. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 15: Table 2 using the s3-score for voting similarity rather than the ideal point distance

	$S_3(Y_{USA,t+1}, Y_{h,t+1})$		
	(1)	(2)	(3)
$w_{USA,t}$	0.09***	0.06***	0.09***
	[0.028]	[0.020]	[0.025]
$w_{RUS,t} \times \mathbf{1}\{w_{USA,t} > w_{RUS,t}\}$	0.14**		
	[0.058]		
$w_{CHN,t} \times \mathbf{1}\{w_{USA,t} > w_{CHN,t}\}$		0.05**	
		[0.020]	
$w_{MAX,t} \times \mathbf{1}\{w_{USA,t} > w_{MAX,t}\}$			0.09***
			[0.027]
Num. Countries	155	186	157
Observations	5,896	7,032	6,764
Kleibergen-Paap F statistics	14.63	25.65	20.67
Num. excluded instruments	1	1	2
Hansen J test			4.051**

Note: the unit of observation is a country h in a given year t. The outcome is the s3 voting similarity score between country h and the United States as measured from their voting behavior at the United Nations General Assembly in year t. The s3 score ignores instances where a country abstains and the other does not, then assigns +1 when they agree, -1 when they disagree, and scales the yearly score by the number of votes in that year (excluding abstentions). Foreign influence $w_{f,t}$ is measured as the log of trade dependence, defined as the value of h's export at time t towards f divided by h's GDP at time t. $w_{MAX,t} := max_{j \in \{RUS,CHN\}}\{w_{j,t}\}$. Column (1) is based on instrument $\mathbf{1}\{1990 \leq t \leq 1994\} \times \mathbf{1}\{w_{USA,1989} > w_{RUS,1989}\}$, column (2) on $\mathbf{1}\{t \geq 2007\} \times \mathbf{1}\{w_{USA,2006} > w_{CHN,2006}\}$, column (3) is based on both. All estimates are conditional on country h fixed effect, year t fixed effects, foreign influence of the competitor of the US ($w_{MAX,t}$, $w_{RUS,t}$, or $w_{CHN,t}$), the log of h's GDP, the log of h's population, the log of h's trade with the rest of the world, the log of h's GDP per capita. Standard errors clustered at the country level are reported in brackets. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.