Supplemental Appendix

This Online Appendix provides technical details for the arguments in the main text and additional details about the calibration. It is organized by section, following the structure of the main text. Equation numbers are prefixed with “A” to indicate their location in the appendix.

3.2 Optimal MFN Tariff Conditional on Offering an FTA to Allies

To derive (7) in Section 3.2, we begin from the welfare function (5), reproduced here:

$$W = x + qy + (1-\alpha)\, m_S S\left(q_n, \tau\right) + \left(m + \alpha m_S\right) S\left(q, 1\right) + \beta \alpha m_S. \tag{A.1}$$

Differentiating this equation with respect to τ yields

$$\frac{W'\left(\tau\right)}{m_S} = (1-\alpha)\, \frac{dS\left(q_n, \tau\right)}{d\tau} + \left[\beta + S\left(q,1\right) - S\left(q_n,\tau\right)\right] \frac{d\alpha}{d\tau}. \tag{A.2}$$

The surplus function is defined by

$$\begin{aligned} S\left(q_n, \tau\right) &= u\left[c\left(\tau q_n\right)\right] - \tau q_n c\left(\tau q_n\right) + (\tau - 1)\, q_n c\left(\tau q_n\right) \\ &= u\left[c\left(p_n\right)\right] - p_n c\left(p_n\right) + (\tau-1)\, q_n c\left(p_n\right), \end{aligned}$$

where $p_n := \tau q_n$. Using the first-order condition

$$u'\left[c\left(p_n\right)\right] = p_n.$$

we can differentiate the surplus function, which implies

$$\begin{aligned} \frac{dS\left(q_n,\tau\right)}{d\tau} &= -c\left(p_n\right) \frac{dp_n}{d\tau} + (\tau-1)\, q_n c'\left(p_n\right) \frac{dp_n}{d\tau} + q_n c\left(p_n\right) + (\tau-1)\, c\left(p_n\right) \frac{dq_n}{d\tau} \\ &= -c\left(p_n\right)\left(q_n + \tau \frac{dq_n}{d\tau}\right) + (\tau-1)\, q_n c'\left(p_n\right) \frac{dp_n}{d\tau} + q_n c\left(p_n\right) + (\tau-1)\, c\left(p_n\right) \frac{dq_n}{d\tau} \\ &= -c\left(p_n\right) \frac{dq_n}{d\tau} + (\tau-1)\, q_n c'\left(p_n\right) \frac{dp_n}{d\tau}. \end{aligned}$$

Substituting into (A.2) we obtain

$$\frac{W'\left(\tau\right)}{m_S} = (1-\alpha)\left[-c\left(p_n\right)\frac{dq_n}{d\tau} + (\tau-1)\, q_n c'\left(p_n\right) \frac{dp_n}{d\tau}\right] + \left[\beta + S\left(q,1\right)\right) - S\left(q_n,\tau\right)\right] \frac{d\alpha}{d\tau},$$

which yields equation (7) in the main text.

From the market clearing condition for non-aligned goods (2), reproduced here as

$$mc\left(\tau q_n\right) + m_S c\left(q_n\right) = y, \tag{A.3}$$

we obtain $q_n = q(\tau)$, where $q(\tau)$ is the seller's price as a function of the tariff derived from this market clearing condition. We therefore have

$$\frac{dq_n}{d\tau} = q'(\tau) = -\frac{mc'(\tau q_n)\, q_n}{mc'(\tau q_n)\,\tau + m_S c'(q_n)} < 0. \tag{A.4}$$

Next, from (6), reproduced here as

$$\alpha = G\left[y(q - q_n)\right], \tag{A.5}$$

we obtain

$$\frac{d\alpha}{d\tau} = -G'\left[y(q-q_n)\right] y \frac{dq_n}{d\tau} > 0 \qquad \text{for} \qquad G'\left[y(q-q_n)\right] > 0. \tag{A.6}$$

That is, as the price of differentiated products from non-aligned countries falls with τ, the fraction of countries that align with the hegemon rises—so long as the density $G' > 0$, i.e., the set of swing states is non-empty.

3.3 To Offer FTAs to Allies or Not?

From the definition of $W^{FTA}(\beta)$ in Section 3.2 of the main text, we have

$$W^{FTA}(\beta) = \max_{\tau} \left[x + qy + (1-\alpha)\, m_S S(q_n, \tau) + (m + \alpha m_S)\, S(q,1) + \beta\alpha m_S\right] \tag{A.7}$$
$$\text{subject to equations (A.3) and (A.5).}$$

If the hegemon does not offer FTAs, its welfare is

$$W^{no-FTA}(\beta) = x + qy + m_S S(q_n^{\circ}, \tau^{\circ}) + mS(q,1) + \beta\alpha^{\circ} m_S,$$

where $\alpha^{\circ} = G(0)$, τ° is the Mill-Bickerdike tariff given by (8), and q_n° is the corresponding price, determined from (A.3). By the envelope theorem,

$$\frac{dW^{FTA}(\beta)}{d\beta} = \alpha(\beta)\, m_S > 0,$$

where $\alpha(\beta)$ is the fraction of countries that align with the hegemon in the solution to (A.7). Under the assumption that $G(\cdot)$ has positive density over a wide support, this fraction is strictly positive. The solution to (A.7) also yields a price function $q_n(\beta)$, which satisfies $\alpha(\beta) \equiv G\left[y(q - q_n(\beta))\right]$.

Define the normalized welfare as

$$\Delta(\beta) := \frac{W^{FTA}(\beta) - W^{no-FTA}(\beta)}{m_S} = \left[1 - \alpha(\beta)\right] \mathcal{S}(\beta) + \alpha(\beta)\, S(q,1) - S(q_n^{\circ}, \tau^{\circ}) + \beta\left[\alpha(\beta) - \alpha^{\circ}\right]$$

where $\mathcal{S}(\beta)$ denotes the surplus $S(q_n, \tau)$ from trade with non-aligned countries after accounting for the dependence of q_n and τ on β in the maximization of W^{FTA}, and $\alpha(\beta)$ is the fraction of countries that align with the hegemon when the MFN tariff is optimally chosen. Since $S(q_n^{\circ}, \tau^{\circ}) > S(q,1)$,

$S(q_n^\circ, \tau^\circ) > \mathcal{S}(\beta)$ and $\alpha(0) = \alpha^\circ > 0$, it follows that $\Delta(0) < 0$. Next note that the first-order condition $W'(\tau) = 0$—from which we derive $\tau(\beta)$—implies that as long as the second-order condition $W''[\tau(\beta)] < 0$ is satisfied, we have

$$\text{sign}\left[\tau'(\beta)\right] = \text{sign}\frac{\partial W'(\tau)}{\partial \beta}.$$

Using equation (A.2), we have

$$\frac{\partial W'(\tau)}{\partial \beta} = \frac{d\alpha}{d\tau} > 0,$$

and therefore $\tau'(\beta) > 0$. Thus, the tariff is strictly increasing in β and by (A.6), so is $\alpha(\beta)$. Since $\Delta'(\beta) = \alpha(\beta) - \alpha^\circ$ and $\alpha(0) = \alpha^\circ$, this implies $\Delta'(\beta) > 0$ for all $\beta > 0$ and $\lim_{\beta \to \infty} \Delta(\beta) > 0$. Hence, there exists a threshold value $\beta^* > 0$ such that $\Delta(\beta) > 0$—so that an FTA is offered— if and only if $\beta > \beta^*$.

3.4 Characterizing the Optimal Tariff

The first-order condition in (7) together with (A.6) and $p_n = \tau q_n$, can be restated as $W'(\tau) = 0$, where

$$\begin{aligned} \frac{W'(\tau)}{m_S} &= (1-\alpha)\left[-c(p_n)\frac{dq_n}{d\tau} + (\tau - 1)\, q_n c'(p_n)\left(q_n + \tau\frac{dq_n}{d\tau}\right)\right] \\ &\quad - \left[\beta + S(q,1)) - S(q_n, \tau)\right] G'\left[y(q - q_n)\right] y\frac{dq_n}{d\tau}. \end{aligned} \tag{A.8}$$

Using the definition of the foreign supply function, $e(q_n) := y - (1-m)\, c(q_n)$, and the market clearing condition for non-aligned goods (A.3), which can be expressed as $e(q_n) = mc(\tau q_n)$, we obtain

$$-c(p_n)\frac{dq_n}{d\tau} + (\tau - 1)\, q_n c'(p_n)\left(q_n + \tau\frac{dq_n}{d\tau}\right) = c(p_n)\frac{dq_n}{d\tau}\left[(\tau - 1)\,\varepsilon_{e(q_n)} - 1\right], \tag{A.9}$$

where $\varepsilon_{e(q_n)}$ is the elasticity of the supply function $e(q_n)$. Substituting this equation, along with (A.5), into (A.8) yields

$$\begin{aligned} \frac{W'(\tau)}{m_S} &= \{1 - G\left[y(q - q_n)\right]\}\, c(p_n)\left[(\tau - 1)\,\varepsilon_{e(q_n)} - 1\right]\frac{dq_n}{d\tau} \\ &\quad - \left[\beta + S(q,1)) - S(q_n, \tau)\right] G'\left[y(q - q_n)\right] y\frac{dq_n}{d\tau}. \end{aligned} \tag{A.10}$$

Since $dq_n/d\tau < 0$, the first-order condition $W'(\tau) = 0$ implies the following expression for the optimal tariff:

$$T(\tau^*) := \tau^* - 1 = \frac{1}{\varepsilon_{e(q_n^*)}}\left\{1 + \left[\beta + S(q,1) - S(q_n^*, \tau^*)\right]\frac{y}{c(\tau^* q_n^*)}\lambda(\eta^*)\right\}, \tag{A.11}$$

where

$$\lambda(\eta^*) := \frac{G'(\eta^*)}{1-G(\eta^*)}$$

is the hazard rate of $G(\cdot)$ evaluated at $\eta^* := y(q-q_n^*)$. This corresponds to (10) in the main text.

Now consider how the optimal tariff responds to a parameter ξ. The sign of $d\tau^*/d\,\xi$ is the same as the sign of $\partial W'(\tau^*)/\partial\xi$, which represents the shift in the marginal utility $W'(\tau^*)$ when τ is held constant at its optimal level, τ^*. From (A.10), we have

$$\begin{aligned}
\text{sign}\frac{\partial}{\partial\xi}\left[\frac{W'(\tau^*)}{m_S}\right] &= -\text{sign}\frac{\partial}{\partial\xi}\left[\frac{W'(\tau^*)}{m_S\left.\frac{dq_n}{d\tau}\right|_{\tau=\tau^*}\varepsilon_{e(q_n^*)}\{1-G[y(q-q_n^*)]\}c(p_n^*)}\right] \\
&= \text{sign}\frac{\partial}{\partial\xi}\frac{1}{\varepsilon_{e(q_n^*)}}\left\{1+[\beta+S(q,1)-S(q_n^*,\tau^*)]\frac{y}{c(\tau^* q_n^*)}\lambda(\eta^*)\right\} \\
&= \text{sign}\frac{\partial T(\tau^*)}{\partial\xi},
\end{aligned}$$

because $dq_n^*/d\tau < 0$. That is, the direction in which the optimal tariff responds to any parameter ζ corresponds to the sign of the partial derivative of the expression in equation (A.11).

The elasticity of the supply function $e(q_n) = y-(1-m)c(q_n)$ is

$$\varepsilon_{e(q_n)} = \frac{(1-m)c(q_n)}{y-(1-m)c(q_n)}\left[-\varepsilon_{c(q_n)}\right] > 0,$$

where $\varepsilon_{c(q_n)} < 0$ is the elasticity of the demand function $c(q_n)$. For a given tariff level τ, the impact of an increase in m on this elasticity operates directly through $(1-m)$ and indirectly through q_n. The direct effect is negative. The impact on q_n is obtained from (A.3), expressed as

$$mc(\tau q_n) + (1-m)c(q_n) = y.$$

Therefore

$$\frac{\partial q_n}{\partial m} = -\frac{c(\tau q_n)-c(q_n)}{mc'(\tau q_n)\tau+(1-m)c'(q_n)}.$$

For $\tau > 1$ this derivative is negative. Namely, q_n declines with m for constant τ. Therefore if the demand function has an elasticity that is increasing in price, $\left[-\varepsilon_{c(q_n)}\right]$ is declining with m. Next note that $\frac{(1-m)c(q_n)}{y-(1-m)c(q_n)}$ is declining in m if and only if $(1-m)c(q_n)$ is declining in m. However, holding τ constant,

$$\begin{aligned}
\frac{d(1-m)c(q_n)}{dm} &= -c(q_n)+(1-m)c'(q_n)\frac{\partial q_n}{\partial m} \\
&= -\frac{mc'(\tau q_n)\tau}{mc'(\tau q_n)\tau+(1-m)c'(q_n)}c(q_n) - \frac{(1-m)c'(q_n)}{mc'(\tau q_n)\tau+(1-m)c'(q_n)}c(\tau q_n) < 0.
\end{aligned}$$

It follows that a sufficient condition for $\varepsilon_{e(q_n)}$ to decline with m is that the demand function $c(q_n)$ satisfies Marshall's Second Law of Demand, i.e., that $\left[-\varepsilon_{c(q_n)}\right]$ is increasing in q_n.

3.5 More Lenient Interpretations of Article XXIV

We now consider the case in which a hegemon can discriminate between countries it offers a PTA in exchange for alignment by setting lower tariffs for some countries and a higher tariffs for others. We require all offers to be "preferential," in the sense that they provide for lower tariffs than the MFN tariff τ.

Let τ_η be one plus the tariff rate offered in the PTA to a small country with alignment cost η. To parameterize the "strictness" of the Article XXIV provisions, we introduce the constraint

$$\tau_\eta \leq 1 + \zeta(\tau - 1), \quad \zeta \in [0,1] \tag{A.12}$$

for countries that enter into a PTA. This constraint reduces to $\tau_\eta \leq \tau$ when $\zeta = 1$ (as in Section 3.5 of the main text) and is stricter the smaller is ζ. At the lower limit, when $\zeta = 0$, only PTAs with zero tariffs are allowed.

Recall that a small country with alignment cost η will accept the offer of a PTA in exchange for alignment only if

$$\eta \leq y\left[q(\tau_\eta) - q_n\right].$$

That is, its cost of alignment must not exceed the economic gain from joining the PTA. Since $q_n = q(\tau)$ and the function $q(\cdot)$ is strictly declining, we can express this constraint as

$$\tau_\eta \leq q^{-1}\left[q(\tau) + \frac{\eta}{y}\right], \tag{A.13}$$

where $q^{-1}[\cdot]$ is the inverse of $q(\cdot)$.

We do not allow import subsidies and thus impose

$$\tau_\eta \geq 1. \tag{A.14}$$

The hegemon chooses an MFN tariff τ, a set of countries with $\eta \leq \bar{\eta}$ to whom it offers participation in a PTA, and tariffs $\{\tau_\eta\}_{-\infty}^{\bar{\eta}}$ in order to maximize its objective function

$$\begin{aligned} W &= x + q(1)\,y + mS[q(1), 1] \\ &\quad + m_S \int_{-\infty}^{\bar{\eta}} S[q(\tau_\eta), \tau_\eta]\, g(\eta)\, d\eta + m_S[1 - G(\bar{\eta})]\, S[q(\tau), \tau] + \beta m_S G(\bar{\eta}) \end{aligned} \tag{A.15}$$

subject to constraints (A.12)-(A.14) for all $\eta \leq \bar{\eta}$.

Recall that the import surplus function, $S[q(\tau_\eta), \tau_\eta]$, is single-peaked and maximized at $\tau_\eta = \tau^\circ > 1$, i.e., the Mill-Bickerdike tariff. Since the import surplus is increasing for $\tau_\eta < \tau^\circ$, the hegemon offers a PTA with tariff τ° if that tariff does not violate the Article XXIV constraint (A.12) and if the country has an alignment cost η low enough that it would accept the offer. This set of countries includes all countries with $\eta \leq 0$ as well as countries with positive costs of alignments that satisfy $\eta \leq y[q(\tau^\circ) - q(\tau)]$. For countries with $\eta > y[q(\tau^\circ) - q(\tau)]$ and $\eta \leq \bar{\eta}$, it offers the tariff

$$\tau_\eta = q^{-1}\left[q(\tau) + \frac{\eta}{y}\right] \tag{A.16}$$

as long as it satisfies (A.14). If, alternatively, τ° violates (A.12), the hegemon offers $\tau_\eta = 1 + \zeta(\tau - 1)$ to all countries for which $\eta \leq y\left[q(1+\zeta(\tau-1)) - q(\tau)\right]$ and (A.16) for countries with $\eta > y\left[q(1+\zeta(\tau-1)) - q(\tau)\right]$ and $\eta \leq \bar{\eta}$ as long as (A.14) is satisfied. In short, the tariff τ_η in a PTA with for a country with alignment cost η has to satisfy

$$\tau_\eta = \tilde{\tau}_\eta(\tau) := \min\left\{\tau^\circ, 1 + \zeta(\tau-1), q^{-1}\left[q(\tau) + \frac{\eta}{y}\right]\right\}. \tag{A.17}$$

Note that $\tilde{\tau}_\eta(\cdot)$ is an increasing function. Using this function, we can rewrite the hegemon's objective function as

$$\begin{aligned} W &= x + q(1)\,y + mS\left[q(1), 1\right] \qquad \text{(A.18)} \\ &\quad + m_S \int_{-\infty}^{\bar{\eta}} S\left[q(\tilde{\tau}_\eta(\tau)), \tilde{\tau}_\eta(\tau)\right] g(\eta)\, d\eta + m_S\left[1 - G(\bar{\eta})\right] S\left[q(\tau), \tau\right] + \beta m_S G(\bar{\eta}). \end{aligned}$$

Now the hegemon's decision problem can be seen as choosing τ and $\bar{\eta}$ in order to maximize (A.18) subject to (A.14).

Suppose that constraint (A.14) does not bind. In this case, the first-order conditions are

$$\left[1 - G(\bar{\eta})\right] \frac{d}{d\tau} S\left[q(\tau), \tau\right] + \frac{d}{d\tau} \int_{-\infty}^{\bar{\eta}} S\left[q(\tilde{\tau}_\eta(\tau)), \tilde{\tau}_\eta(\tau)\right] g(\eta)\, d\eta = 0, \tag{A.19}$$

$$\beta + S\left[q(\tilde{\tau}_{\bar{\eta}}(\tau)), \tilde{\tau}_{\bar{\eta}}(\tau)\right] - S\left[q(\tau), \tau\right] = 0. \tag{A.20}$$

Since $\tilde{\tau}_\eta(\tau) \leq \tau^\circ$, with strict inequality for some small countries, the second term in (A.19) must be positive and thus the first term must be negative.[44] This implies that the optimal MFN tariff, which we denote by $\tau^\times$, exceeds τ°. That is, $\tau^\times > \tau^\circ$. Next note that (A.20) implies that the hegemon obtains no surplus from the marginal country that enters into a PTA. Namely, for $\bar{\eta} = \bar{\eta}^\times$, where $\bar{\eta}^\times$ denotes the optimal marginally aligning country, we have

$$\beta + S\left[q\left(\tilde{\tau}_{\bar{\eta}^\times}\left(\tau^\times\right)\right), \tilde{\tau}_{\bar{\eta}^\times}\left(\tau^\times\right)\right] - S\left[q\left(\tau^\times\right), \tau^\times\right] = 0.$$

Recall that, when the hegemon must offer a *free* trade agreement (FTA) to all that align, its optimal MFN tariff τ^* satisfies

$$\beta + S\left[q(1), 1\right] - S\left[q(\tau^*), \tau^*\right] > 0.$$

[44] To see why $\tilde{\tau}_\eta(\tau) < \tau^\circ$ for at least some small countries, note that lowering τ slightly below τ° only reduces the planner's economic welfare to second order. However, its effect on prices and hence alignment if first-order, so that the planner will always offer tariffs below the Mill-Bickerdike level to some small countries.

The last two equations imply

$$S\left[q\left(\tau^{\times}\right),\tau^{\times}\right]-S\left[q\left(\tau^{*}\right),\tau^{*}\right]>S\left[q\left(\tilde{\tau}_{\bar{\eta}^{\times}}\left(\tau^{\times}\right)\right),\tilde{\tau}_{\bar{\eta}^{\times}}\left(\tau^{\times}\right)\right]-S\left[q\left(1\right),1\right].$$

Since $\tilde{\tau}_{\bar{\eta}^{\times}}\left(\tau^{\times}\right)\leq\tau^{\circ}$, the right-hand side of this inequality is weakly positive, because the surplus function is increasing to the left of τ°. It follows that

$$S\left[q\left(\tau^{\times}\right),\tau^{\times}\right]>S\left[q\left(\tau^{*}\right),\tau^{*}\right].$$

And since both $\tau^{\times}$ and τ^{*} are to the right of τ°, it follows that $\tau^{\times}<\tau^{*}$.

We next compare the size of the alliances. In the benchmark case we have

$$\bar{\eta}^{*}=y\left[q\left(1\right)-q\left(\tau^{*}\right)\right]$$

while in the case discussed here, where (A.14) does not bind, we have

$$\bar{\eta}^{\times}=y\left[q\left(\tilde{\tau}_{\bar{\eta}^{\times}}\left(\tau^{\times}\right)\right)-q\left(\tau^{\times}\right)\right].$$

Since $q\left(\cdot\right)$ is a strictly decreasing function, $\tilde{\tau}_{\bar{\eta}^{\times}}\left(\tau^{\times}\right)\geq1$ and $\tau^{\times}<\tau^{*}$, it follows that $\bar{\eta}^{\times}<\bar{\eta}^{*}$. That is, fewer small countries align with the hegemon under the weaker version of Article XXIV when the constraint of no import subsidy does not bind and an FTA is offered.

We next study the case in which (A.14) binds. In this case the marginal country satisfies

$$\bar{\eta}=y\left[q\left(1\right)-q\left(\tau\right)\right]$$

and the hegemon chooses τ and $\bar{\eta}$ to maximize (A.18) subject to this constraint. This problem can be recast as choosing τ to maximize

$$\begin{aligned}W &= x+q\left(1\right)y+mS\left[q\left(1\right),1\right]\\ &\quad +m_{S}\int_{-\infty}^{y\left[q\left(1\right)-q\left(\tau\right)\right]}S\left[q\left(\tilde{\tau}_{\eta}\left(\tau\right)\right),\tilde{\tau}_{\eta}\left(\tau\right)\right]g\left(\eta\right)d\eta+m_{S}\left[1-G\left(y\left[q\left(1\right)-q\left(\tau\right)\right]\right)\right]S\left[q\left(\tau\right),\tau\right]\\ &\quad +\beta m_{S}G\left(y\left[q\left(1\right)-q\left(\tau\right)\right]\right).\end{aligned}$$

This yields the first-order condition

$$\begin{aligned}0 &= \frac{d}{d\tau}\int_{-\infty}^{y\left[q\left(1\right)-q\left(\tau\right)\right]}S\left[q\left(\tilde{\tau}_{\eta}\left(\tau\right)\right),\tilde{\tau}_{\eta}\left(\tau\right)\right]g\left(\eta\right)d\eta\\ &\quad +\left[1-G\left(y\left[q\left(1\right)-q\left(\tau\right)\right]\right)\right]\frac{d}{d\tau}S\left[q\left(\tau\right),\tau\right]-g\left(\bar{\eta}\right)\left\{\beta+S\left[q\left(1\right),1\right]-S\left[q\left(\tau\right),\tau\right]\right\}y\frac{dq\left(\tau\right)}{d\tau}.\end{aligned}$$

Let $\tau^{\times}$ be the solution of the MFN tariff from this equation. Note from (A.8) that the marginal utility $W'\left(\tau\right)$ in the baseline case has the same sign as the second line of this first-order condition. Moreover, the first line of this first-order condition is positive, as explained above. It follows that $W'(\tau^{\times}) < 0$. Therefore $\tau^{\times} > \tau^{*}$. That is, when the no-import-subsidy constraint binds, the hegemon chooses an MFN tariff that exceeds the MFN tariff in the baseline case. This then implies $\bar{\eta}^{\times} > \bar{\eta}^{*}$; more countries align with the hegemon when Article XXIV is relaxed and the no-import-subsidy constraint binds.

We now show that for $\beta > \beta^{*}$—when in the baseline case the hegemon prefers to offer an FTA instead of imposing τ° on all countries—the no-import-subsidy constraint (A.14) binds. This implies that if there exists an equilibrium in which (A.14) does not bind, it has to be for $\beta < \beta^{*}$. To see why the no-import-subsidy constraint binds when $\beta > \beta^{*}$, recall that we have shown in the main text that in this case[45]

$$\beta + S\left[q\left(1\right), 1\right] - S\left[q\left(\tau^{\circ}\right), \tau^{\circ}\right] > 0.$$

If there were a solution to the hegemon's problem in which the no-import-subsidy constraint does not bind, condition (A.20) would be satisfied. That is,

$$\beta + S\left[q\left(\tilde{\tau}_{\bar{\eta}}\left(\tau^{\times}\right)\right), \tilde{\tau}_{\bar{\eta}}\left(\tau^{\times}\right)\right] - S\left[q\left(\tau^{\times}\right), \tau^{\times}\right] = 0 \tag{A.21}$$

with $\tilde{\tau}_{\bar{\eta}}(\tau^{\times}) \in (1, \tau^{\circ})$ and $\tau^{\times} > \tau^{\circ}$. Therefore

$$\begin{aligned} 0 &= \beta + S\left[q\left(\tilde{\tau}_{\bar{\eta}}\left(\tau^{\times}\right)\right), \tilde{\tau}_{\bar{\eta}}\left(\tau^{\times}\right)\right] - S\left[q\left(\tau^{\times}\right), \tau^{\times}\right] \\ &> \beta + S\left[q\left(\tilde{\tau}_{\bar{\eta}}\left(\tau^{\times}\right)\right), \tilde{\tau}_{\bar{\eta}}\left(\tau^{\times}\right)\right] - S\left[q\left(\tau^{\circ}\right), \tau^{\circ}\right] \\ &> \beta + S\left[q\left(1\right), 1\right] - S\left[q\left(\tau^{\circ}\right), \tau^{\circ}\right], \end{aligned}$$

which is a contradiction. We conclude that for $\beta > \beta^{*}$ constraint (A.14) binds.

We now turn to comparative statics of $\tau^{\times}$ and $\eta^{\times}$ with respect to β. To this end, fix some $\beta < \beta'$ and suppose that $(\bar{\eta}^{\times}, \tau^{\times})$ is optimal under β and $(\bar{\eta}^{\times\prime}, \tau^{\times\prime})$ is optimal under β'.

We first argue that $\bar{\eta}^{\times} < \bar{\eta}^{\times\prime}$, i.e. alignment increases in β. To this end note that $(\bar{\eta}^{\times}, \tau^{\times})$ and $(\bar{\eta}^{\times\prime}, \tau^{\times\prime})$ are both feasible profiles (irrespective of the value of β). But then the facts that $(\bar{\eta}^{\times}, \tau^{\times})$ weakly dominates $(\bar{\eta}^{\times}, \tau^{\times})$ at β and $(\bar{\eta}^{\times\prime}, \tau^{\times\prime})$ weakly dominates $(\bar{\eta}^{\times}, \tau^{\times})$ at β' imply

$$\beta\left[G(\bar{\eta}^{\times\prime}) - G(\bar{\eta}^{\times})\right] \leq \beta'\left[G(\bar{\eta}^{\times\prime}) - G(\bar{\eta}^{\times})\right].$$

Since $G(\cdot)$ is strictly increasing, we may conclude that $\bar{\eta}^{\times} \leq \bar{\eta}^{\times\prime}$.

To demonstrate that moreover this inequality is strict, suppose not, i.e. $\bar{\eta}^{\times} = \bar{\eta}^{\times\prime}$. Then since β only affects the planner's problem through the term $\beta G(\bar{\eta})$, either value of τ—i.e. $\tau^{\times}$ or $\tau^{\times\prime}$ achieves the same value of the objective at β' and in particular $(\bar{\eta}^{\times\prime}, \tau^{\times}) = (\bar{\eta}^{\times}, \tau^{\times})$ is optimal at

[45] The difference between the hegemon's welfare with an FTA offer and without it, as a function of β, is

$$\begin{aligned} \Delta\left(\beta\right) &= \left(1-\alpha^{*}\right) S\left(q_n^{*}, \tau^{*}\right) + \alpha^{\circ} S\left(q, 1\right) - \left(1 - \alpha^{*} + \alpha^{\circ}\right) S\left(q_n^{\circ}, \tau^{\circ}\right) \\ &\quad + \left(\alpha^{*} - \alpha^{\circ}\right)\left[\beta + S\left(q, 1\right) - S\left(q_n^{\circ}, \tau^{\circ}\right)\right]. \end{aligned}$$

The first line on the right-hand side is weakly negative, since $S\left(q_n^{\circ}, \tau^{\circ}\right) \geq \max\left\{S\left(q_n^{*}, \tau^{*}\right), S\left(q, 1\right)\right\}$. Therefore, the condition $\Delta\left(\beta\right) > 0$, together with $\alpha^{*} > \alpha^{\circ}$, implies that $\beta + S\left(q, 1\right) - S\left(q_n^{\circ}, \tau^{\circ}\right) > 0$.

β' (in addition to, of course, β). Therefore, any feasible variation around $(\bar{\eta}^\times, \tau^\times)$ should have zero first-order effect on the objective at both β and β'. One such feasible variation increases $\bar{\eta}^\times$ by $d\bar{\eta}$ and $\tau^\times$ by $d\tau$ defined such that

$$\bar{\eta}^\times/y + q(\tau^\times) = (\bar{\eta}^\times + d\bar{\eta})/y + q(\tau^\times + d\tau).$$

However, this requires

$$\beta G'(\bar{\eta}^\times) = \beta' G'(\bar{\eta}^\times).$$

which is impossible given $\beta < \beta'$ and $G'(\cdot) > 0$.

We next argue that $\tau^\times \le \tau^{\times\prime}$, i.e. MFN tariffs are weakly increasing in β. Toward a contradiction now suppose $\tau^\times > \tau^{\times\prime}$. Since $(\bar{\eta}^{\times\prime}, \tau^{\times\prime})$ is feasible by assumption, $\tilde{\tau}_{\bar{\eta}^{\times\prime}}(\tau^{\times\prime}) \ge 1$. Since $\tilde{\tau}_\eta(\tau)$ is weakly decreasing in η and weakly increasing in τ, and as we have already established that $\bar{\eta}^\times < \bar{\eta}^{\times\prime}$, this implies $\tilde{\tau}_{\bar{\eta}^\times}(\tau^{\times\prime}) \ge 1$ and $\tilde{\tau}_{\bar{\eta}^{\times\prime}}(\tau^\times) \ge 1$, so both $(\bar{\eta}^\times, \tau^{\times\prime})$ and $(\bar{\eta}^{\times\prime}, \tau^\times)$ are also feasible (at either β). The facts that $(\bar{\eta}^\times, \tau^\times)$ dominates $(\bar{\eta}^\times, \tau^{\times\prime})$ and $(\bar{\eta}^{\times\prime}, \tau^{\times\prime})$ dominates $(\bar{\eta}^{\times\prime}, \tau^\times)$ then imply

$$\begin{aligned}
&\int_{-\infty}^{\bar{\eta}^\times} S\left[q(\tilde{\tau}_\eta(\tau^\times)), \tilde{\tau}_\eta(\tau^\times)\right] g(\eta)d\eta + [1 - G(\bar{\eta}^\times)] S[q(\tau^\times), \tau^\times] \\
\ge &\int_{-\infty}^{\bar{\eta}^\times} S\left[q(\tilde{\tau}_\eta(\tau^{\times\prime})), \tilde{\tau}_\eta(\tau^{\times\prime})\right] g(\eta)d\eta + [1 - G(\bar{\eta}^\times)] S[q\left(\tau^{\times\prime}\right), \tau^{\times\prime}]
\end{aligned}$$

and

$$\begin{aligned}
&\int_{-\infty}^{\bar{\eta}^{\times\prime}} S\left[q(\tilde{\tau}_\eta(\tau^\times)), \tilde{\tau}_\eta(\tau^\times)\right] g(\eta)d\eta + [1 - G(\bar{\eta}^{\times\prime})] S[q(\tau^\times), \tau^\times] \\
\le &\int_{-\infty}^{\bar{\eta}^{\times\prime}} S\left[q(\tilde{\tau}_\eta(\tau^{\times\prime}), \tilde{\tau}_\eta(\tau^{\times\prime})\right] g(\eta)d\eta + [1 - G(\bar{\eta}^{\times\prime})] S[q\left(\tau^{\times\prime}\right), \tau^{\times\prime}].
\end{aligned}$$

Combining these inequalities implies

$$\begin{aligned}
&[G(\bar{\eta}^{\times\prime}) - G(\bar{\eta}^\times)] \left\{S[q(\tau^\times), \tau^\times] - S[q(\tau^{\times\prime}), \tau^{\times\prime}]\right\} \\
\ge &\int_{\bar{\eta}^\times}^{\bar{\eta}^{\times\prime}} \left\{S\left[q(\tilde{\tau}_\eta(\tau^\times)), \tilde{\tau}_\eta(\tau^\times)\right] - S\left[q(\tilde{\tau}_\eta(\tau^{\times\prime})), \tilde{\tau}_\eta(\tau^{\times\prime})\right]\right\} g(\eta)d\eta.
\end{aligned}$$

The first line is strictly negative since $\bar{\eta}^\times < \bar{\eta}^{\times\prime}$, $G(\cdot)$ is strictly increasing, $\tau^\times > \tau^{\times\prime}$ (by assumption), and $S[q(\tau), \tau]$ is single-peaked at $\tau^\circ \le \tau^{\times\prime}$. The second line is strictly positive since (a) the integrand is strictly positive since $\tau^\times > \tau^{\times\prime}$, $\tilde{\tau}_\eta(\tau)$ is weakly below τ° and increasing in τ, and $S[q(\tau), \tau]$ is single-peaked at τ° and (b) $\bar{\eta}^\times < \bar{\eta}^{\times\prime}$. As we have arrived at a contradiction, it must be that $\tau^\times \le \tau^{\times\prime}$.

Next, we argue that there exists some critical $\beta^\times$ such that the positivity constraint binds if and only if $\beta > \beta^\times$. It suffices to show that (a) the constraint does not bind for sufficiently low

β, (b) the constraint does bind for sufficiently high β, and (c) it is impossible to have $\beta < \beta'$ for which the constraint binds at β but not β'. Claim (a) is immediate by considering $\beta = 0$. Claim (b) is implied by our earlier observation that the constraint binds for β above β^*. To see claim (c), suppose not. Suppose the constraint binds at β but does not bind as some $\beta' > \beta$. Take $(\bar{\eta}^\times, \tau^\times)$ to be optimal at β and $(\bar{\eta}^{\times\prime}, \tau^{\times\prime})$ to be optimal at β'. Recall that we have shown that $\bar{\eta}^\times < \bar{\eta}^{\times\prime}$ and $\tau^\times \leq \tau^{\times\prime}$. Then considering the first-order condition with respect to $\bar{\eta}$ implies

$$S\left[q(1), 1\right] + \beta \geq S[q(\tau^\times), \tau^\times] \qquad \text{and} \qquad S\left[q(\tau_{\bar{\eta}^{\times\prime}}), \tau_{\bar{\eta}^{\times\prime}}\right] + \beta' = S\left[q\left(\tau^{\times\prime}\right), \tau^{\times\prime}\right]$$

Subtracting these inequalities implies

$$S\left[q(1), 1\right] - S\left[q(\tau_{\bar{\eta}^{\times\prime}}), \tau_{\bar{\eta}^{\times\prime}}\right] + \beta - \beta' \geq S[q(\tau^\times), \tau^\times] - S[q\left(\tau^{\times\prime}\right), \tau^{\times\prime}]$$

where $S\left[q(1), 1\right] - S\left[q(\tau_{\bar{\eta}^{\times\prime}}), \tau_{\bar{\eta}^{\times\prime}}\right] < 0$, $\beta - \beta' < 0$ and $S[q(\tau^\times), \tau^\times] - S[q\left(\tau^{\times\prime}\right), \tau^{\times\prime}] \geq 0$. These signs use that (i) $S[q(\tau), \tau]$ is strictly single peaked around $\tau^\circ \geq \tau_{\bar{\eta}^{\times\prime}} > 1$ and (ii) $\tau^\circ \leq \tau^\times \leq \tau^{\times\prime}$. This is a contradiction and therefore claim (c) is satisfied.

Finally, we argue that $\tau^\times$ is weakly decreasing in β. Since (i) $\tau_{\bar{\eta}^\times} \geq 1$ in the region where the positivity constraint does not bind, (ii) $\tau_{\bar{\eta}^\times} = 1$ in the region where the positivity constraint does bind, and (iii) there is a threshold $\beta^\times$ for which the constraint binds if and only if $\beta > \beta^\times$, it suffices to show that $\tau_{\bar{\eta}^\times}$ is strictly decreasing in β within the region where the positivity constraint does not bind. To this end, suppose that for some $\beta < \beta'$, the positivity constraint does not bind, $(\bar{\eta}^\times, \tau^\times)$ is optimal under β, and $(\bar{\eta}^{\times\prime}, \tau^{\times\prime})$ is optimal under β'. The first-order conditions with respect to $\bar{\eta}$ imply

$$S\left[q(\tau_{\bar{\eta}^\times}), \tau_{\bar{\eta}^\times}\right] + \beta = S[q(\tau^\times), \tau^\times] \qquad \text{and} \qquad S\left[q(\tau_{\bar{\eta}^{\times\prime}}), \tau_{\bar{\eta}^{\times\prime}}\right] + \beta' = S\left[q\left(\tau^{\times\prime}\right), \tau^{\times\prime}\right].$$

Since $\tau^\circ \leq \tau^\times \leq \tau^{\times\prime}$ and $\beta < \beta'$, we have

$$S\left[q(\tau_{\bar{\eta}^\times}), \tau_{\bar{\eta}^\times}\right] > S\left[q(\tau_{\bar{\eta}^{\times\prime}}), \tau_{\bar{\eta}^{\times\prime}}\right].$$

Since $S\left[q\left(\tau\right), \tau\right]$ is single peaked with a maximum at τ° and since $\tau^\circ \geq \tau_{\bar{\eta}^\times}, \tau_{\bar{\eta}^{\times\prime}}$, this implies $\tau_{\bar{\eta}^\times} > \tau_{\bar{\eta}^{\times\prime}}$ as desired.

4 Nash Tariffs in a Bipolar World

In a bipolar world the market clearing condition are (12)-(15) in the main text, reproduced here as:

$$m_H c\left(q_h\right) + m_F c\left(\tau_F q_h\right) + m_S c\left(q_h\right) = y, \tag{A.22}$$

$$m_H c\left(\tau_H q_f\right) + m_F c\left(q_f\right) + m_S c\left(q_f\right) = y, \tag{A.23}$$

$$m_H c\left(\tau_H q_n\right) + m_F c\left(\tau_F q_n\right) + m_S c\left(q_n\right) = y, \tag{A.24}$$

$$m_J c(q_J) + m_{-J} c(\tau_{-J} q_J) + m_S c(q_J) = y, \quad J \in \{H, F\}, \tag{A.25}$$

where m_J is the size of large country J, q_J is the price in J of good y produced in country J and τ_J is the tariff in country J, $J \in \{H, F\}$; q_h is the price in a small country that aligns with H of y produced in this small country, q_f is the price in a small country that aligns with F of y produced in this small country and q_n is the price in a small unaligned country of y produced in this small country. The symbol $-J$ represents a large country that is not J. It follows from these market clearing conditions that $q_h = q_H$ and $q_f = q_F$, and that q_h depends on the tariff in F but not in H while q_f depends on the tariff in H but not in F. Price q_n is the only one that depends on both tariff levels. Since the demand functions $c(\cdot)$ are declining, it follows that q_h is declining in τ_F, q_f is declining in τ_H and q_n is declining in each one of the tariff levels. Therefore $q_n < \min\{q_H, q_F\}$ for $\tau_H > 1$ and $\tau_F > 1$.

In the bipolar world the objective function of the policy maker in country H is (18) in the main text, reproduced here as:

$$\begin{aligned} W_H(\tau_H, \tau_F) = {} & x + q_H y + m_H S(q_H, 1) + m_F S(q_F, \tau_H) \\ & + m_h S(q_h, 1) + m_f S(q_f, \tau_H) + m_n S(q_n, \tau_H) \\ & + \beta_H m_h - \delta_H m_f, \end{aligned} \tag{A.26}$$

where $m_h = \alpha_H m_S$, $m_f = \alpha_F m_S$ and $m_h = (1 - \alpha_H - \alpha_F) m_S$. Therefore

$$\frac{dm_h}{d\tau_H} = \frac{d\alpha_H}{d\tau_H} m_S, \quad \frac{dm_f}{d\tau_H} = \frac{d\alpha_F}{d\tau_H} m_S, \quad \frac{dm_n}{d\tau_H} = -\left(\frac{d\alpha_H}{d\tau_H} + \frac{d\alpha_F}{d\tau_H}\right) m_S.$$

Using properties of the surplus functions outlined in the previous section, this yields (19) in the main text, reproduced here as

$$\begin{aligned} \frac{\partial W_H}{\partial \tau_H} = {} & (m_F + m_f)\left[-c(p_F)\frac{dq_F}{d\tau_H} + (\tau_H - 1) q_F c'(p_F)\frac{dp_F}{d\tau_H}\right] \\ & + m_n \left[-c(p_n)\frac{dq_n}{d\tau_H} + (\tau_H - 1) q_n c'(p_n)\frac{dp_n}{d\tau_H}\right] \\ & + \left[\beta_H + S(q_h, 1) - S(q_n, \tau_H)\right]\frac{dm_h}{d\tau_H} \\ & + \left[S(q_f, \tau_H) - S(q_n, \tau_H) - \delta_H\right]\frac{dm_f}{d\tau_H}. \end{aligned} \tag{A.27}$$

Recall from (16)-(17) in the main text that

$$\alpha_H = \int \int^{\min\{(q_h - q_n)y, (q_h - q_f)y + \eta_{F,i}\}} \gamma(\eta_H, \eta_F)\, d\eta_H d\eta_F, \tag{A.28}$$

$$\alpha_F = \int \int^{\min\{(q_f - q_n)y, (q_f - q_h)y + \eta_{H,i}\}} \gamma(\eta_H, \eta_F)\, d\eta_F d\eta_H. \tag{A.29}$$

Therefore

$$\alpha_J = \left(\int_{(q_{-J}-q_n)y} \int^{(q_J-q_n)y} + \int^{(q_{-J}-q_n)y} \int^{(q_J-q_{-J})y+\eta_{-J}} \right) \gamma\left(\eta_H, \eta_F\right) d\eta_J d\eta_{-J}$$

and

$$\begin{aligned} \frac{d\alpha_H}{d\tau_H} &= \left\{ \int_{(q_F-q_n)y} \gamma\left[\left(q_H - q_n\right) y, \eta_F\right] d\eta_F \right\} \frac{d\left(q_H - q_n\right) y}{d\tau_H} \\ &\quad + \left\{ \int^{(q_F-q_n)y} \gamma\left[\left(q_H - q_F\right) y + \eta_F, \eta_F\right] d\eta_F \right\} \frac{d\left(q_H - q_F\right) y}{d\tau_H}, \end{aligned} \tag{A.30}$$

$$\begin{aligned} \frac{d\alpha_F}{d\tau_H} &= \left\{ \int_{(q_H-q_n)y} \gamma\left[\eta_H, \left(q_F - q_n\right) y\right] d\eta_H \right\} \frac{d\left(q_F - q_n\right) y}{d\tau_H} \\ &\quad + \left\{ \int^{(q_H-q_n)y} \gamma\left[\eta_H, \left(q_F - q_H\right) y + \eta_H\right] d\eta_H \right\} \frac{d\left(q_F - q_H\right) y}{d\tau_H}. \end{aligned} \tag{A.31}$$

4.1 An Optimal-Tariff Formula for a Great Power in a Bipolar World

In the main text we present the optimal tariff formula (20) for τ_H in a bipolar world, which we reproduce here as:

$$\tau_H^* - 1 = \frac{1}{\omega_n \varepsilon_{e_n(q_n^*)} + \left(1 - \omega_n\right) \varepsilon_{e_f\left(q_f^*\right)}} \left[1 + \omega_n \frac{y}{c\left(p_n^*\right)} \Omega_n + \left(1 - \omega_n\right) \frac{y}{c\left(p_f^*\right)} \Omega_f \right]. \tag{A.32}$$

The expressions for Ω_f and Ω_n are given in (21) and (22), respectively, in the main text. They are reproduced here as

$$\begin{aligned} \Omega_f &= \frac{\mu_{HF}}{m_F + m_f^*} \left\{ \beta_H + \delta_H - \left[S\left(q_f^*, \tau_H^*\right) - S\left(q_h^*, 1\right) \right] \right\} \\ &\quad + \frac{\mu_{nF}}{m_F + m_f^*} \left\{ \delta_H - \left[S\left(q_f^*, \tau_H^*\right) - S\left(q_n^*, \tau_H^*\right) \right] \right\}, \end{aligned} \tag{A.33}$$

$$\begin{aligned} \Omega_n &= \frac{\mu_{Hn}}{m_n^*} \left\{ \beta_H - \left[S\left(q_n^*, \tau_H^*\right) - S\left(q_h^*, 1\right) \right] \right\} \\ &\quad + \frac{\mu_{nF}}{m_n^*} \left\{ -\delta_H - \left[S\left(q_n^*, \tau_H^*\right) - S\left(q_f^*, \tau_H^*\right) \right] \right\}, \end{aligned} \tag{A.34}$$

where

$$\omega_n := \frac{m_n^* c\left(p_n^*\right) \frac{dq_n}{d\tau_H}}{\left(m_F + m_f^*\right) c\left(p_F^*\right) \frac{dq_f}{d\tau_H} + m_n^* c\left(p_n^*\right) \frac{dq_n}{d\tau_H}},$$

$dq_n/d\tau_H$ is evaluated at $\tau = \tau^*$ and

$$\begin{aligned} \mu_{HF} &= m_S \int^{(q_F^* - q_n^*)y} \gamma\left[\left(q_H^* - q_F^*\right)y + \eta_F, \eta_F\right] d\eta_F \\ &= m_S \int^{(q_H^* - q_n^*)y} \gamma\left[\eta_H, \left(q_F^* - q_H^*\right)y + \eta_H\right] d\eta_H, \end{aligned}$$

$$\mu_{nF} = m_S \int_{(q_H^* - q_n^*)y} \gamma\left[\eta_H, \left(q_F^* - q_n^*\right)y\right] d\eta_H,$$

$$\mu_{nH} = m_S \int_{(q_F^* - q_n^*)y} \gamma\left[\left(q_H^* - q_n^*\right)y, \eta_F\right] d\eta_F.$$

The terms Ω_f and Ω_n, evaluated at the optimal tariff, describe the marginal welfare effects of changes in alignment induced by a decline in q_n and q_F, respectively, and ω_n is the share of the total terms-of-trade effect of a change in the MFN tariff that reflects imports from non-aligned countries. Note that τ_H^* is the best response to a given tariff level τ_F in country F. As before, we have suppressed the functional relationship between variables on the right-hand side and the two tariff rates. The term μ_{HF} describes the density ("number") of small countries that are indifferent between aligning with H or with F, and similarly, μ_{nF} represents the density of countries that are indifferent between aligning with F and remaining nonaligned while μ_{Hn} represents the density of countries that are indifferent between aligning with H and remaining nonaligned. Using these densities, and recalling that q_H does not depend on τ_H, we can express the derivatives of α_H and α_F in (A.30) and (A.31), evaluated at $\tau_H = \tau_H^*$, as

$$m_S \frac{d\alpha_H}{d\tau_H} = -\left(\mu_{Hn}\frac{dq_n}{d\tau_H} + \mu_{HF}\frac{dq_F}{d\tau_H}\right)y, \tag{A.35}$$

$$m_S \frac{d\alpha_F}{d\tau_H} = -\mu_{nF}\, y \frac{dq_n}{d\tau_H} + \left(\mu_{HF} + \mu_{nF}\right) y \frac{dq_F}{d\tau_H}. \tag{A.36}$$

To derive (A.32), we use (A.27), $m_h = \alpha_H m_S$, $m_f = \alpha_F m_S$ and (A.35)-(A.36) to express the first order condition $\partial W_H / \partial \tau_H = 0$ as

$$\begin{aligned} 0 = \frac{\partial W_H}{\partial \tau_H} &= \left(m_F + m_f^*\right)\left[-c\left(p_f^*\right)\frac{dq_f}{d\tau_H} + \left(\tau_H^* - 1\right) q_f^* c'\left(p_f^*\right)\frac{dp_f}{d\tau_H}\right] \\ &+ m_n^*\left[-c\left(p_n^*\right)\frac{dq_n}{d\tau_H} + \left(\tau_H^* - 1\right) q_n^* c'\left(p_n^*\right)\frac{dp_n}{d\tau_H}\right] \\ &- \left[\beta_H + S\left(q_h^*, 1\right) - S\left(q_n^*, \tau_H^*\right)\right]\left(\mu_{Hn}\frac{dq_n}{d\tau_H} + \mu_{HF}\frac{dq_F}{d\tau_H}\right) y \\ &+ \left[S\left(q_f^*, \tau_H\right) - S\left(q_n^*, \tau_H^*\right) - \delta_H\right]\left[-\mu_{nF}\frac{dq_n}{d\tau_H} + \left(\mu_{HF} + \mu_{nF}\right)\frac{dq_F}{d\tau_H}\right] y. \end{aligned}$$

The first two lines on the right-hand side of this equation describe the marginal response of surpluses from imported goods from the powerful rival, small countries that align with this rival and small countries that are nonaligned. The last two lines describe the impact of changes in the measure of countries that align with H and the measure of countries that align with F. Using (A.9) and its comparable expression for changes in q_f, we obtain

$$\begin{aligned}
&\left(m_F + m_f^*\right)\left[-c\left(p_f^*\right)\frac{dq_f}{d\tau_H} + \left(\tau_H^* - 1\right) q_f^* c'\left(p_f^*\right)\frac{dp_f}{d\tau_H}\right] \\
&+ m_n^*\left[-c\left(p_n^*\right)\frac{dq_n}{d\tau_H} + \left(\tau_H^* - 1\right) q_n^* c'\left(p_n^*\right)\frac{dp_n}{d\tau_H}\right] \\
&= \left(m_F + m_f^*\right)\left[\left(\tau_H^* - 1\right)\varepsilon_{e\left(q_F^*\right)} - 1\right] c\left(p_f^*\right)\frac{dq_f}{d\tau_H} \\
&+ m_n^*\left[\left(\tau_H^* - 1\right)\varepsilon_{e(q_n^*)} - 1\right] c\left(p_n^*\right)\frac{dq_n}{d\tau_H},
\end{aligned}$$

where the derivatives $dq_f/d\tau_H$ and $dq_n/d\tau_H$ are evaluated at $\tau_H = \tau_H^*$. Therefore, recalling that $q_f = q_F$,

$$\begin{aligned}
0 = \frac{\partial W_H}{\partial \tau_H} &= \left(m_F + m_f^*\right)\left[\left(\tau_H^* - 1\right)\varepsilon_{e\left(q_F^*\right)} - 1\right] c\left(p_f^*\right)\frac{dq_f}{d\tau_H} \\
&+ m_n^*\left[\left(\tau_H^* - 1\right)\varepsilon_{e(q_n^*)} - 1\right] c\left(p_n^*\right)\frac{dq_n}{d\tau_H} \\
&- \left[\beta_H + S\left(q_h^*, 1\right) - S\left(q_n^*, \tau_H^*\right)\right]\left[\frac{\mu_{HF}}{m_F + m_f^*} c\left(p_f^*\right)\frac{dq_f}{d\tau_H}\frac{\left(m_F + m_f^*\right) y}{c\left(p_f^*\right)} + \frac{\mu_{Hn}}{m_n^*} c\left(p_n^*\right)\frac{dq_n}{d\tau_H}\frac{m_n^* y}{c\left(p_n^*\right)}\right] \\
&+ \left[S\left(q_f^*, \tau_H\right) - S\left(q_n^*, \tau_H^*\right) - \delta_H\right]\left[\frac{\mu_{HF} + \mu_{nF}}{m_F + m_f^*} c\left(p_f^*\right)\frac{dq_f}{d\tau_H}\frac{\left(m_F + m_f^*\right) y}{c\left(p_f^*\right)} - \frac{\mu_{nF}}{m_n^*} c\left(p_n^*\right)\frac{dq_n}{d\tau_H}\frac{m_n^* y}{c\left(p_n^*\right)}\right]
\end{aligned}$$

or

$$\begin{aligned}
0 &= \frac{\partial W_H}{\partial \tau_H}\frac{1}{\left(m_F + m_f^*\right) c\left(p_f^*\right)\frac{dq_f}{d\tau_H} + m_n^* c\left(p_n^*\right)\frac{dq_n}{d\tau_H}} \\
&= \left(\tau_H^* - 1\right)\left[\omega_n \varepsilon_{e(q_n^*)} + \left(1 - \omega_n\right)\varepsilon_{e\left(q_f^*\right)}\right] - 1 \\
&- \omega_n \frac{y}{c\left(p_n^*\right)}\Omega_n - \left(1 - \omega_n\right)\frac{y}{c\left(p_f^*\right)}\Omega_f.
\end{aligned}$$

This yields the optimal tariff formula (A.32).

Assuming that the second-order condition is satisfied, we have for every parameter ζ the relationship

$$\operatorname{sign}\frac{d\tau_H^*}{d\zeta} = \operatorname{sign}\frac{d}{d\zeta}\left(\frac{\partial W_H}{\partial \tau_H}\right).$$

Therefore, since $dq_f/d\tau_H < 0$ and $dq_n/d\tau_H < 0$, at $\tau = \tau^*$, we also have

$$\begin{aligned}
\operatorname{sign}\frac{d\tau_H^*}{d\zeta} &= -\operatorname{sign}\frac{d}{d\zeta}\left\{\frac{\partial W_H}{\partial \tau_H}\frac{\left[\omega_n \varepsilon_{e(q_n^*)} + (1-\omega_n)\,\varepsilon_{e\left(q_f^*\right)}\right]^{-1}}{\left(m_F + m_f^*\right) c\left(p_f^*\right)\frac{dq_f}{d\tau_H} + m_n^* c\left(p_n^*\right)\frac{dq_n}{d\tau_H}}\right\} \\
&= \operatorname{sign}\frac{d}{d\zeta}\left\{\frac{1}{\omega_n \varepsilon_{e_n(q_n^*)} + (1-\omega_n)\,\varepsilon_{e_f\left(q_f^*\right)}}\left[1 + \omega_n \frac{y}{c\left(p_n^*\right)}\Omega_n + (1-\omega_n)\frac{y}{c\left(p_f^*\right)}\Omega_f\right]\right\},
\end{aligned}$$

where the left-hand side holds τ_F fixed and the right-hand side holds both τ_H and τ_F fixed.

4.2 Are Rival's Tariffs Strategic Complements or Strategic Substitutes?

Consider the case in which there are no geopolitical considerations in both large countries. In this event every country J imposes its tariff τ_J on all small countries as well as on country $-J$. As a result, the market clearing conditions are (A.24) and (A.25), where q_n is the price of y produced in a small country and q_J is the price of y produced in large country J. Prices q_h and q_f do not exist. In this case $m_f^* = 0$, $m_n^* = m_S$ and the objective function of country H is

$$W_H(\tau_H, \tau_F) = x + q_H y + m_H S(q_H, 1) + m_F S(q_F, \tau_H) + m_S S(q_n, \tau_H),$$

which yields

$$\tau_H^* - 1 = \frac{1}{\omega_n \varepsilon_{e(q_n^*)} + (1-\omega_n)\,\varepsilon_{e\left(q_F^*\right)}}, \tag{A.37}$$

where $e_n(q_n^*) := y - m_F c(\tau_F q_n^*) - m_S c(q_n^*)$ and $e_F(q_F^*) := y - m_F c(q_F^*) - m_S c(q_F^*)$ and

$$\omega_n := \frac{m_S c\left(p_n^*\right)\frac{dq_n}{d\tau_H}}{m_F c\left(p_F^*\right)\frac{dq_F}{d\tau_H} + m_S c\left(p_n^*\right)\frac{dq_n}{d\tau_H}}. \tag{A.38}$$

In this expression $p_n^* = \tau_H^* q_n^*$ and $p_F^* = \tau_H^* q_F^*$. An increase in τ_F does not change the elasticity $\varepsilon_{e\left(q_F^*\right)}$, because the tariff of country F does not impact the market clearing condition (A.25) for $J = F$. Therefore the foreign tariff impacts the best response τ_H^* through two channels: the elasticity $\varepsilon_{e(q_n^*)}$ and the weight ω_n.

With a constant demand elasticity function $c(p) = p^{-\sigma}$,

$$\varepsilon_{e_n(q_n^*)} = \sigma \frac{m_F (\tau_F q_n^*)^{-\sigma} + m_S (q_n^*)^{-\sigma}}{y - m_F (\tau_F q_n^*)^{-\sigma} - m_S (q_n^*)^{-\sigma}}$$

and the market clearing condition (A.24) implies that

$$y - m_F (\tau_F q_n^*)^{-\sigma} - m_S (q_n^*)^{-\sigma} = m_H (\tau_H^* q_n^*)^{-\sigma}.$$

Therefore

$$\varepsilon_{e_n(q_n^*)} = \sigma \frac{m_F (\tau_F)^{-\sigma} + m_S}{m_H \left(\tau_H^*\right)^{-\sigma}}.$$

For constant τ_H^* this elasticity is declining in τ_F, which implies that through this channel an increase in τ_F raises the best response τ_H^*.

Next note that

$$\varepsilon_{e_F(q_F^*)} = \sigma \frac{m_F (q_F^*)^{-\sigma} + m_S (q_F^*)^{-\sigma}}{y - m_F \left(q_F^*\right)^{-\sigma} - m_S \left(q_F^*\right)^{-\sigma}}$$

and from the market clearing condition (A.25),

$$y - m_F (q_F^*)^{-\sigma} - m_S (q_F^*)^{-\sigma} = m_H (\tau_H^* q_F^*)^{-\sigma}.$$

Therefore

$$\varepsilon_{e_F(q_F^*)} = \sigma \frac{m_F + m_S}{m_H \left(\tau_H^*\right)^{-\sigma}}.$$

It follows that $\varepsilon_{e_n(q_n^*)} < \varepsilon_{e_f(q_f^*)}$ as long as $\tau_F > 1$. In this case τ_H^* also responds to a change in τ_F through the weight ω_n; it rises through this channel in response to an increase in τ_F if the increase in τ_F raises the weight ω_n. For a constant elasticity demand function this weight can be expressed as

$$\omega_n := \frac{m_S (q_n^*)^{-\sigma} \frac{dq_n}{d\tau_H}}{m_F \left(q_F^*\right)^{-\sigma} \frac{dq_F}{d\tau_H} + m_S (q_n^*)^{-\sigma} \frac{dq_n}{d\tau_H}}.$$

An increase in τ_F, holding constant τ_H^*, does not change q_F^*, but it reduces q_n^*, which increases this weight for given $dq_n/d\tau_H$ and $dq_f/d\tau_H$. However, (A.24) and (A.25) imply that

$$\frac{dq_F}{d\tau_H} = -\frac{m_H (\tau_H^*)^{-\sigma-1}}{m_H \left(\tau_H^*\right)^{-\sigma} + m_F + m_S} q_F^*,$$

$$\frac{dq_n}{d\tau_H} = -\frac{m_H (\tau_H^*)^{-\sigma-1}}{m_H \left(\tau_H^*\right)^{-\sigma} + m_F (\tau_F)^{-\sigma} + m_S} q_n^*,$$

and therefore

$$\omega_n = \frac{\frac{m_S (q_n^*)^{1-\sigma}}{m_H \left(\tau_H^*\right)^{-\sigma} + m_F (\tau_F)^{-\sigma} + m_S}}{\frac{m_F \left(q_F^*\right)^{1-\sigma}}{m_H \left(\tau_H^*\right)^{-\sigma} + m_F + m_S} + \frac{m_S (q_n^*)^{1-\sigma}}{m_H \left(\tau_H^*\right)^{-\sigma} + m_F (\tau_F)^{-\sigma} + m_S}}.$$

It follows that for constant τ_H^* this weight is increasing in the tariff τ_F if and only if

$$\frac{m_S (q_n^*)^{1-\sigma}}{m_H \left(\tau_H^*\right)^{-\sigma} + m_F (\tau_F)^{-\sigma} + m_S}$$

is increasing in τ_F. Next note that τ_F has a direct effect on this expression and an indirect effect through q_n^*. The direct effect raises ω_n when τ_F increases. An increase in τ_F also reduces q_n^*, which raises this expression when $\sigma > 1$. We therefore conclude that an increase in τ_F raises ω_n both through the direct and the indirect effects.

In summary, an increase in τ_F reduces the elasticity $\varepsilon_{e_n(q_n^*)}$ and raises the weight ω_n, and both these effects increase the best response τ_H^*.

5.1 Model Calibration

We obtain data from two sources. First, we obtain voting records for the U.N. General Assembly from 1946 through September 10, 2024 from the UN Digital Library.[46] Each entry in the dataset is a country-resolution pair that indicates how the country voted on the UNGA resolution (either Yes, No, Abstention, or Non-Voting). Second, we obtain annual GDP data for UN countries from 1970 through 2023 from the UN Statistics Division.[47]

The only substantive data cleaning step is to attach GDP data to the UN voting dataset. Our analysis only requires us to do this for the years 1997 and 2023; these are the years we use for GDP in the hegemon and bipolar calibrations, respectively, as we discussed in the main text. We perform this merge using a crosswalk from UN Statistical Division between M49 codes (used in GDP data) and ISO3 codes (used in voting data).[48] We manually correct this crosswalk in three cases:

1. We add GDP attributed to Greenland to that of Denmark. Greenland is an autonomous territory of Denmark that Denmark represents in the UN.

2. We set the GDP of Tanzania to the sum of GDPs of the sub-national entities of Mainland Tanzania and Zanzibar.

3. We reassign M49 code 736 (pre-2011 unified Sudan) to the code 729, which the crosswalk maps to pre-2011 Sudan in the UN voting data.

As described in the main text, we use UN voting data to calibrate the distribution of small country alignment costs in both the unipolar and bipolar models. We calibrate the distribution of alignment costs in the hegemon model, $G(\cdot)$, based on UN voting from 1995-1998 and GDP in 1997, and we calibrate the distribution of alignment costs in the bipolar model, $\Gamma(\cdot)$, based on UN voting data from 2021-2024 and GDP in 2023.

To calibrate these distributions, we compute, for each great power J and each small country i, the share of votes $f_{J,i}$ (in the relevant range of years) for which J and i cast the same vote, among resolutions on which neither is non-voting.[49] Letting V be the set of all votes v and letting

[46] Data are available at https://digitallibrary.un.org/record/4060887?ln=en#files.
[47] Data are available at https://data.un.org/Data.aspx?q=gdp+us+dollars&d=SNAAMA&f=grID%3a101%3bcurrID%3aUSD%3bpcFlag%3a0.
[48] The crosswalk is available at https://unstats.un.org/unsd/methodology/m49/overview/
[49] In the calibration of the unipolar model, we treat China as a small country. So $J = H$ and we consider $i = CHN$. In the calibration of the bipolar model, we treat China as a large country. So we consider $J = H$ and $J = F$, and we do not consider $i = CHN$.

	1995-1998 voting	2021-2024 voting
Number of Small Countries	180	191
Unweighted Average of $\boldsymbol{f_{H,i}}$	0.293	0.347
Unweighted Average of $\boldsymbol{f_{F,i}}$	n/a	0.636

Table A.1: Summary Statistics for Voting Similarity Measure

$v_j \in \{\text{Yes, No, Abstention, Non-Voting}\}$ be j's vote on v, we compute

$$f_{J,i} = \frac{\sum_{v \in V} \mathbf{1}_{v_J \neq \text{Non-Voting}} \mathbf{1}_{v_i \neq \text{Non-Voting}} \mathbf{1}_{v_J = v_i}}{\sum_{v \in V} \mathbf{1}_{v_J \neq \text{Non-Voting}} \mathbf{1}_{v_i \neq \text{Non-Voting}}}.$$

In the 1995-1998 period, there are four countries—Iraq, Somalia, São Tomé and Príncipe, and Yugoslavia—that appear in the UN voting data but whose votes are all "Non-Voting." We remove these countries from our analysis. In the 2021-2024 period, there are no such countries. All remaining countries vote on some (usually all, or almost all) of the same votes as the United States and China, allowing us to compute $f_{J,i}$.

Table A.1 reports summary statistics for our voting similarity measure.

We compute each country's alignment costs by setting

$$\eta_{J,i} = \kappa GDP_i (\bar{f}_J - f_{J,i})$$

as described in the main text. Finally, we set $G(\cdot)$ and $\Gamma(\cdot)$ by fitting one- and two-dimensional Gaussian mixture models to $\eta_{H,i}$ and $(\eta_{H,i}, \eta_{F,i})$, respectively. In doing so, we weight each small country by its GDP by fitting the Gaussian mixture models to an augmented empirical distribution of alignment costs that repeats each small country's costs a number of times that it proportional to its GDP. Guided by the data, we allow for three Gaussian mixture components when fitting $G(\cdot)$ and two when fitting $\Gamma(\cdot)$.

5.2 Alternative Calibration for Alignments with the Hegemon

The main text refers to an alternative calibration of the unipolar model based on UN voting in the 2021-2024 period. This section of the appendix details that alternative calibration and the corresponding results.

The alternative calibration uses the same values of all model parameters, except three: the size of Home, whose alternative calibration we denote $m_{H,\text{alt}}$, the complementary mass of all small countries, $m_{\text{S,alt}}$, and the distribution of alignment costs, $G_{\text{alt}}(\cdot)$. We set $m_{H,\text{alt}}$ to the US's share of 2023 world GDP, i.e., $m_{H,\text{alt}} = 0.265$, as in the bipolar calibration. We accordingly set $m_{S,\text{alt}} = 1 - m_{H,\text{alt}} = 0.735$. We set $G_{\text{alt}}(\cdot)$ equal to the marginal distribution over $\eta_{H,i}$ of the distribution

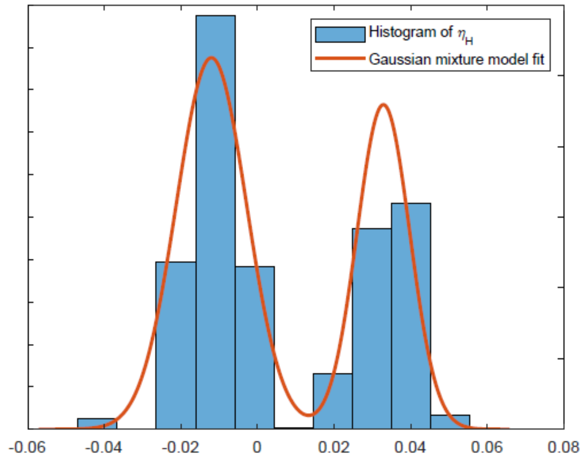
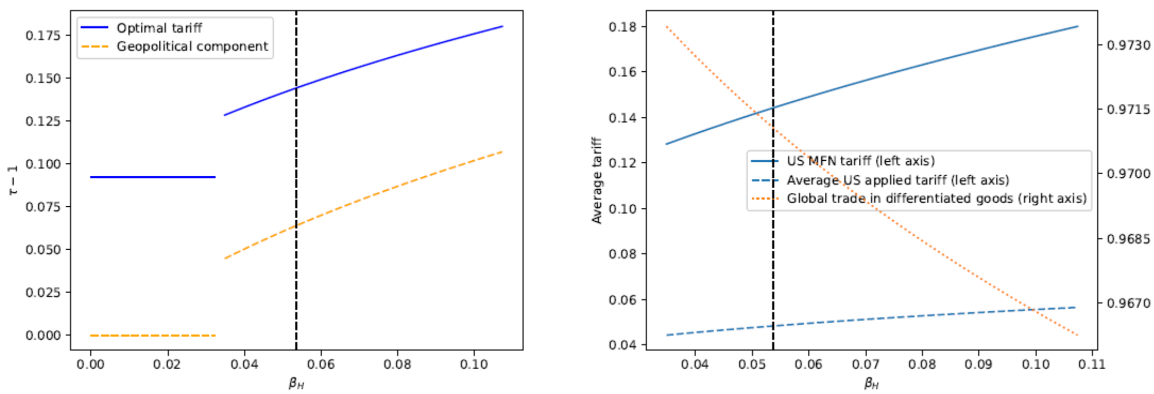


Figure A.1: Histogram and Estimated Distribution of Alignment Costs for Alternative Calibration



(a) (b)

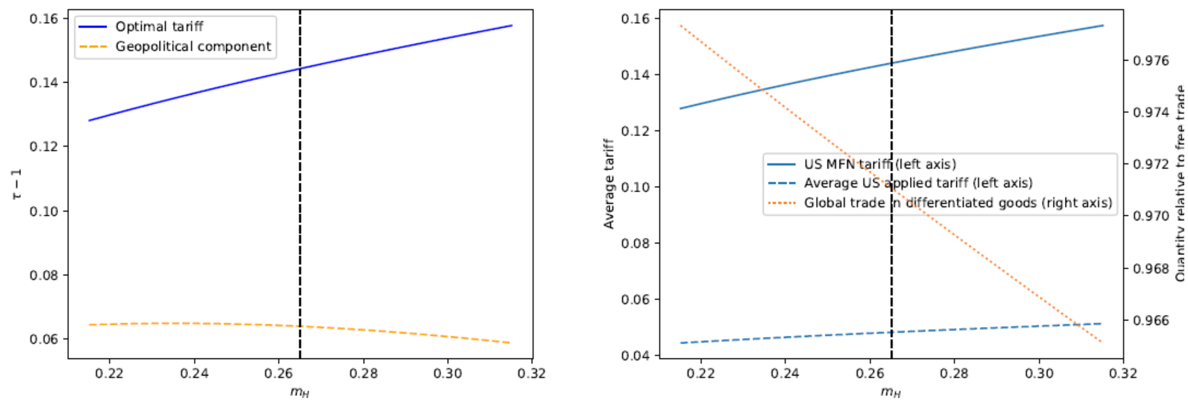
Figure A.2: Optimal Tariff as Function of β_H in Alternative Unipolar Calibration

of alignment costs $\Gamma(\cdot)$ used in the main text for the bipolar case. Formally, for all $\eta_{H,i}$,

$$G_{\text{alt}}(\eta_{H,i}) = \lim_{\eta_{F,i}\to\infty} \Gamma(\eta_{H,i}, \eta_{F,i}).$$

Because of this construction, the marginally aligning countries under $G_{\text{alt}}(\cdot)$ are the same as the marginally Home-aligning countries under $\Gamma(\cdot)$. These are Japan and San Marino, both of whom vote with the US 50.4% of the time in the UNGA. Figure A.1 displays a histogram of the estimates of $\eta_{H,i}$ that underlie our calibration of $\Gamma(\cdot)$ and, therefore, $G_{\text{alt}}(\cdot)$, as well as our estimate of $G_{\text{alt}}(\cdot)$ itself.

This calibration implies that the hegemon's optimal tariff is 14.4%. As can be seen in Figure A.2, the geopolitical component accounts for close to half of this tariff. Absent geopolitical considerations, the hegemon would impose a uniform tariff of 9.2% and offer no FTA to any partners. Under our calibration, however, the hegemon attracts as allies countries comprising 43.0% of world GDP, compared to the 40.2% that would align under free trade. This stick-and-carrot approach yields an average applied tariff of 4.8%, far lower than the MFN rate levied on non-aligners.



(a) (b)

Figure A.3: Optimal Tariff as Function of Size in Alternative Unipolar Calibration

Panel (a) of Figure A.2 also allows us to examine how the optimal MFN tariff varies with the hegemon's preference for alignment. When β_H is below roughly 0.033, the hegemon offers no FTAs and defaults to the Mill-Bickerdike tariff. Consistent with our theoretical analysis, the optimal tariff rises with β_H whenever the hegemon finds it advantageous to offer preferential access to encourage alignment. For example, our estimates suggest that doubling the value that the hegemon places on allies would push the optimal tariff to around 18%. As in the main text, panel (b) shows the MFN tariff rate and the average applied tariff (i.e., tariff revenues divided by the hegemon's total f.o.b. imports) relative to their estimated baseline levels for preference parameters large enough to induce FTAs. Predictably, stronger geopolitical preferences lead to larger tariffs and less trade.

Figure A.3 plots the optimal tariff as a function of the hegemon's size. The tariff rises with m_H, reflecting the enhanced market power and stronger terms-of-trade motive that come with increased size. Unlike in our primary unipolar calibration in the main text, here we find that the geopolitical component in the total tariff falls in the hegemon's size. Evidently, the hazard rate effect no longer counteracts the rising opportunity cost of offering preferential access and the expanding MFN tariff base associated with improved terms of trade. Still, panel (b) shows that, as in the main text, openness—whether measured by average applied tariffs or by the volume of trade relative to the free-trade benchmark—declines as the hegemon comes to dominate the global economy.

5.3 Optimal Tariffs in the Bipolar World

Best Response Functions

Within our calibrated bipolar model, we compute the best response functions for both H and F. As discussed in Section 4.2, there are many competing channels that can make either country's tariffs a complement or substitute to that of the other. Figure A.4 illustrates the net impact of these forces by plotting the best-response functions. The plot shows that from H's perspective, F's tariff is a modest substitute to its own, while from F's perspective, H's tariff is a very slight complement to its own.

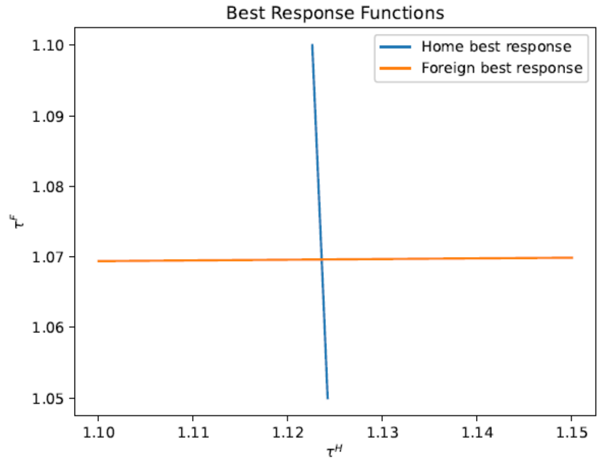


Figure A.4: Best Response Functions in the Bipolar Calibration

Alignment with and without FTAs

In our calibrated bipolar model, both large countries offer free trade agreements and impose tariffs on countries that do not align with them. In Figure A.5 we provide a visual representation of how these optimal policies affect the alignment decisions of small countries. The figure shows the empirical distribution of alignment costs inferred from UN voting data as well as contour lines of the Gaussian mixture model that we fit to it. The vertical, horizontal, and diagonal lines in the plot represent the boundaries between regions of alignment costs within which small countries align with either H, F, or neither. The black dashed lines represent these boundaries under free trade, i.e., when neither country uses trade as a carrot or a stick. In this case, 54.6% of small countries align with H, 9.3% of small countries align with F, and 36.1% remain non-aligned. The red dash-dotted lines, by contrast, represent the boundaries between alignment regions under H's and F's optimal (Nash) policies. Since each large country offers an FTA to aligners and imposes a positive tariff on others, more small countries align with each of them than under free trade: 58.2% of small countries align with H, 13.0% align with F, and 28.8% remain non-aligned.

Expansion of F at the Expense of H

In the main text, we considered the case where F expands at the expense of small countries. This mirrors the rise of China between 1990 and the present, which has come mostly at the expense of countries other than the United States. In this appendix, we instead consider the case where F's size m_F expands at the expense of H's size m_H, while the size of all other countries m_S is held fixed. One interpretation is that this reflects a possible future in which China grows faster than the rest of the world while the United States grows slower than the rest of the world.

Figure A.6 reports the results of this simulation. Panel (a) shows that F's tariff grows roughly linearly in its size, from 7.0% in our baseline calibration to 19.0% when F grows to fully subsume

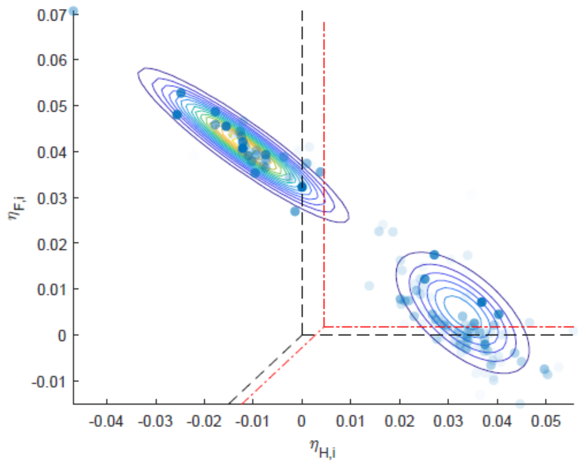
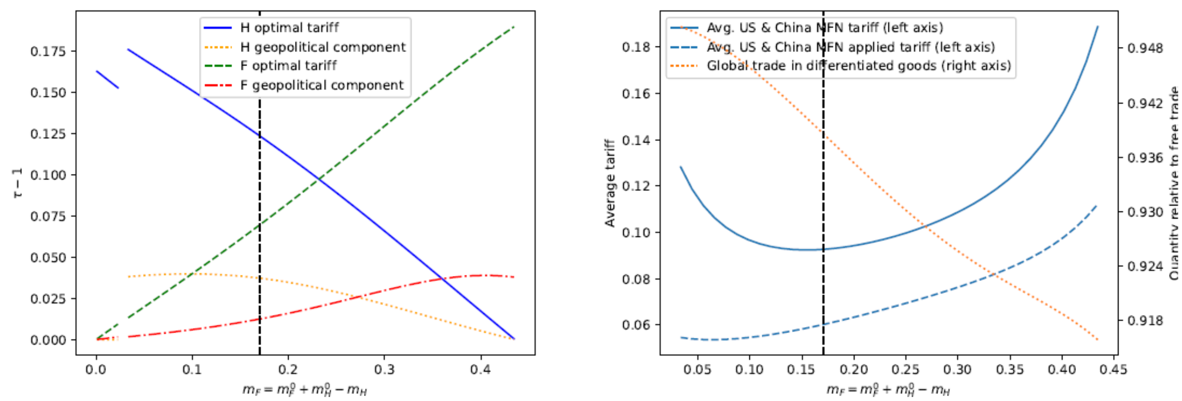


Figure A.5: Estimated Alignment Costs for Bipolar Case Under Free Trade and Nash Tariffs



(a) (b)

Figure A.6: Expansion of F at the Expense of H

H's mass. As in the comparative static we consider in the main text, most of this tariff hike reflects F's rising market power, while the geopolitical component for its tariff remains modest.

Somewhat less similarly to the main text, H's tariff now responds sharply to the growth of F. This is unsurprising because, as F's growth comes at H's expense, any growth in F reduces H's market power and so—except when it affects H's decision about whether to offer an FTA—decreases H's optimal tariff.[50]

Panel (b) shows that, as in the main text, average tariffs respond non-monotonically to changes in the size of country F while global trade overall falls. However, the decline in trade is substantially more gradual than in the comparative static considered in the main text. This reflects the simple fact that, in the comparative static considered here, the total share of world GDP controlled by tariff-imposing great powers does not rise as F grows, since H shrinks one-for-one in this growth. The fact that global trade falls, despite this fact, mainly reflects that—since a smaller mass of

[50] The jump in H's tariff when F reaches about 3% of world GDP reflects that H does not offer FTA's when it is sufficiently large compared to F.

countries align with F than H—F applies its tariff to more small countries.