Trade Negotiations under Fire: Social Identity and the Rising Opposition to Free Trade

– Online Appendix –

Susanne Keller* Sebastian Krautheim†

December 2025

Abstract

This online appendix provides a detailed analysis of the model with asymmetric population shares.

* University of Passau
† Corresponding author, University of Passau and CESifo, email: sebastian.krautheim@uni-passau.de

C Asymmetric population shares

This section analyzes the model with unequal population shares and shows that our main results can also be obtained in this context. As starting point, we consider the 'old world' equilibrium in which country X imposes the high standard and country F imposes the low standard. For clarity, the analysis below focuses on the derivation of the 'new world' equilibrium where the government takes the NGO prototype as given. The derivation of the 'old world' equilibrium is analogous except that the alternative of X to produce under the high standard at cost c_h is not the foreign cost of low-standard production c_l^F but the domestic cost c_l.

Overall, there are two major differences to our analysis with $\lambda_n = \lambda_e$ in the main paper: first, when maximizing total welfare the government puts different weights on the utilities of the two consumer groups. Second, asymmetric population shares affect the consumer decision about identification with society at large. The reason is that dissonance costs increase in the distance between the individual and the group average. When an individual is part of the majority group, it is also closer to the average and is therefore prone to identification, while the minority group will dissociate for relatively low levels of inequality (see equations 6 and 7). Therefore, in the analysis with unequal population shares lemma 2 does not apply and we have to consider four possible identification regimes under the low standard, $(r_n^X, r_e^X, r_e^N) = \{(1,1,1), (0,1,1), (1,0,1), (0,0,0)\}$. Figure C.1 illustrates the 'new world' equilibrium prior to NGO campaigns for $\lambda_n < \lambda_e$ and fig. C.2 for $\lambda_n > \lambda_e$. Different to our main analysis with equal population shares, there are four convex curves $U^{(1,1,1)}(c_l^F), U^{(0,1,1)}(c_l^F), U^{(1,0,1)}(c_l^F)$ and $U^{(0,0,1)}(c_l^F)$ representing aggregate utility under the low standard for the four different identification regimes (prior to NGO campaigns). In appendices section C.2 and section C.3, we derive all elements that describe fig. C.1 and fig. C.2.

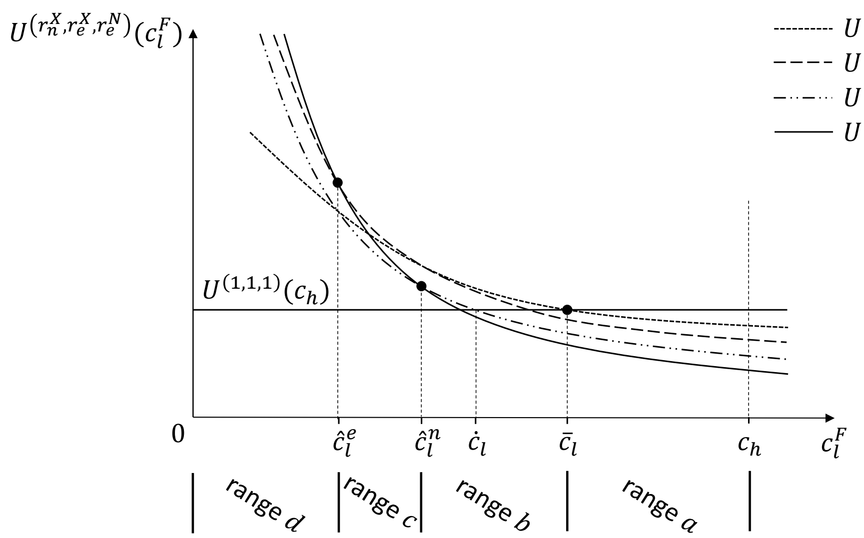

Figure C.1: Equilibrium when $\lambda_n < \lambda_e$

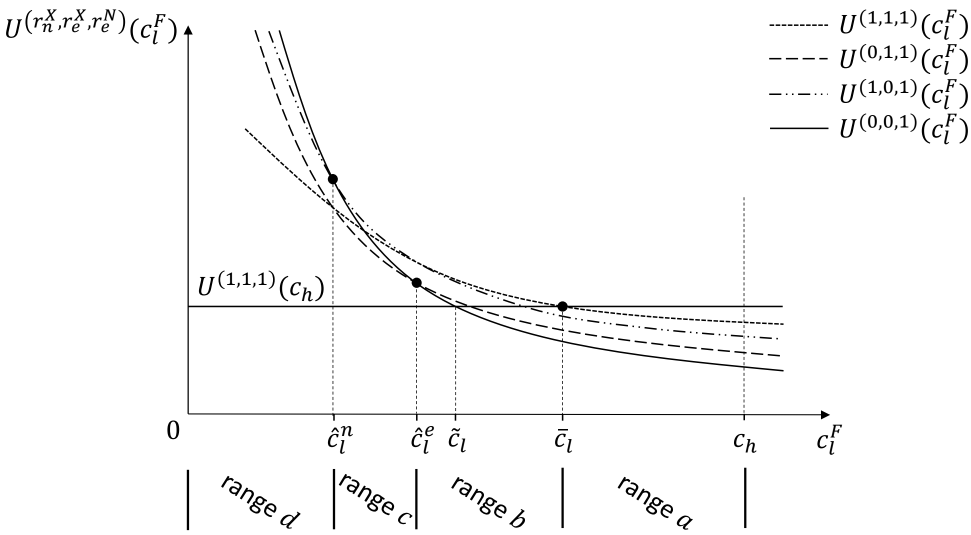

Figure C.2: Equilibrium when $\lambda_n > \lambda_e$

C.1 Adjusted parameter restrictions

The assumptions presented in appendix A.2 in the main paper slightly change for the analysis with asymmetric population shares. The adjusted assumptions are stated below and their formal derivations are provided in appendices C.3.1 and C.4.4 (assumption 5 and assumption 6), C.4.1 (assumption 7), and C.2 (assumption 8).

Assumption 5. $\frac{\alpha^X(\lambda_n+\delta\lambda_e)}{(1-\delta)(1-\lambda_i)} < \beta^X < \frac{(1+\alpha+\alpha^X)\lambda_n(1-\delta)+\lambda_e\eta\theta\alpha^X(\lambda_n+\delta\lambda_e)}{(2+\eta\theta)\lambda_e\lambda_n(1-\delta)}$ *for* $i = n, e$.

Note that the inequalities hold iff:

$$\delta < min\left\{\frac{(1+\alpha+\alpha^X)\lambda_n-\alpha^X 2\lambda_n\lambda_e}{(1+\alpha+\alpha^X)\lambda_n+\alpha^X 2\lambda_e^2}, \frac{(1+\alpha+\alpha^X)\lambda_n-\lambda_n\alpha^X[2\lambda_n+2\lambda_n\eta\theta-\eta\theta]}{(1+\alpha+\alpha^X)\lambda_n+\lambda_e\alpha^X[2\lambda_n+2\lambda_n\eta\theta-\eta\theta]}\right\},$$

that is, assumption 5 implies a sufficiently high perceived gap in ethical quality between the high and the low standard.

Assumption 6.

$$\left(\frac{(1+\alpha+\alpha^X)\Phi\big[\beta^X\lambda_n(1-\delta)-\alpha^X(\lambda_n+\delta\lambda_e)\big]}{A^X[(1+\alpha+\alpha^X)(\lambda_n+\delta\lambda_e)-2\lambda_e\lambda_n(1-\delta)\beta^X]-\lambda_e A^N\eta\theta[\beta^X\lambda_n(1-\delta)-\alpha^X(\lambda_n+\delta\lambda_e)]}\right)^{\frac{1}{\gamma}} < c_h$$

$$< \left(\frac{\Phi\big(\beta^X(1-\delta)(1-\lambda_e)-\alpha^X(\lambda_n+\lambda_e\delta)\big)}{\delta A^X}\right)^{\frac{1}{\gamma}} \textit{ with } \Phi = \gamma^\gamma(1-\gamma)^{(1-\gamma)}E.$$

In appendix C.5, we show that given assumption 5 there exists a non-empty set of values of c_h that satisfies this condition.

Assumption 7. $A^N < A^X$.

Assumption 8. $1+\alpha > \alpha^X$.

C.2 Properties of aggregate utility

Differentiating aggregate utility (eq. (A.2)) in the 'old world' with respect to c_s with $s = l, h$ delivers:

$$\begin{aligned}\frac{\partial U^{(r_n^X, r_e^X, r_e^N)}}{\partial c_s} = & -\gamma \Phi c_s^{-\gamma-1} \left\{ (1+\alpha)\left(\lambda_n + \left[\mathbb{I}_s + (1-\mathbb{I}_s)\delta\right]\lambda_e\right) \right. \\ & +\lambda_n \mathbb{I}_n^X \left[\alpha^X\left(\lambda_n + \left[\mathbb{I}_s + (1-\mathbb{I}_s)\delta\right]\lambda_e\right) - \beta^X(1-[\mathbb{I}_s + (1-\mathbb{I}_s)\delta])(1-\lambda_n)\right] \\ & \left. +\lambda_e \mathbb{I}_e^X \left[\alpha^X\left(\lambda_n + \left[\mathbb{I}_s + (1-\mathbb{I}_s)\delta\right]\lambda_e\right) - \beta^X(1-[\mathbb{I}_s + (1-\mathbb{I}_s)\delta])(1-\lambda_e)\right]\right\} .\end{aligned}$$

Aggregate utility decreases in cost c_s if the term in curly brackets is positive. This implies:

$$\begin{aligned} 0 \; < \; & (1+\alpha)\left(\lambda_n + \left[\mathbb{I}_s + (1-\mathbb{I}_s)\delta\right]\lambda_e\right) \\ & + \lambda_n \mathbb{I}_n^X \left[\alpha^X\left(\lambda_n + \left[\mathbb{I}_s + (1-\mathbb{I}_s)\delta\right]\lambda_e\right) - \beta^X(1-[\mathbb{I}_s + (1-\mathbb{I}_s)\delta])(1-\lambda_n)\right] \\ & + \lambda_e \mathbb{I}_e^X \left[\alpha^X\left(\lambda_n + \left[\mathbb{I}_s + (1-\mathbb{I}_s)\delta\right]\lambda_e\right) - \beta^X(1-[\mathbb{I}_s + (1-\mathbb{I}_s)\delta])(1-\lambda_e)\right] .\end{aligned}$$

According to lemma 1 when the regulated good is produced under the high standard $s = h$, we only have to consider the case where all consumers identify with society at large $(1, 1, 1)$. When the regulated good is produced under the low standard $s = l$, there are four possible cases, $(1, 1, 1)$, $(0, 1, 1)$, $(1, 0, 1)$ and $(0, 0, 1)$. The derivative is negative when the high standard is implemented and both groups identify with society at large, $s = h$ and $(1, 1, 1)$, but also in the case of low-standard production and no identification with society, $s = l$ and $(0, 0, 1)$. This is obvious to see since aggregate utility does not include dissonance costs in these two cases, i.e. in the inequality above line two and three does not include the last term in squared brackets, $\beta^X(1 - [\mathbb{I}_s + (1 - \mathbb{I}_s)\delta])(1 - \lambda_i)$. For low standard production combined with society, $s = l$ and $(1, 1, 1)$, the derivative is negative iff

$$\frac{(1+\alpha+\alpha^X)(\lambda_n + \delta\lambda_e)}{2\lambda_n\lambda_e(1-\delta)} > \beta^X . \tag{C.1}$$

This condition is less strict than the upper bound for β^X stated in assumption 5 and therefore always fulfilled given our parameter restrictions. For low standard production combined with only caring consumers identifying with society, $s = l$ and $(0, 1, 1)$, the condition is

$$\begin{aligned} & (1+\alpha)(\lambda_n + \delta\lambda_e) + \lambda_e\alpha^X(\lambda_n + \delta\lambda_e) - \lambda_n\lambda_e\beta^X(1-\delta) > 0 \\ \leftrightarrow \quad & \frac{(1+\alpha+\lambda_e\alpha^X)(\lambda_n + \delta\lambda_e)}{\lambda_n\lambda_e(1-\delta)} > \beta^X , \end{aligned} \tag{C.2}$$

and for low standard production combined with only non-caring consumers identifying with society, $s = l$ and $(1, 0, 1)$:

$$(1+\alpha)(\lambda_n + \delta\lambda_e) + \lambda_n\alpha^X(\lambda_n + \delta\lambda_e) - \lambda_n\lambda_e\beta^X(1-\delta) > 0$$

$$\leftrightarrow \quad \frac{(1+\alpha+\lambda_n\alpha^X)\,(\lambda_n+\delta\lambda_e)}{\lambda_n\lambda_e(1-\delta)} > \beta^X \; . \tag{C.3}$$

Throughout appendix C, we assume that $1+\alpha > \alpha^X$ (stated by assumption 8), which implies that eq. (C.2) and eq. (C.3) are always larger as – and therefore consistent with – the upper bound stated in assumption 5.It follows that under assumption 5, aggregate utility $U^{(r_n^X, r_e^X, r_e^N)}$ decreases strictly monotonically in the cost c_s and so do the four curves in fig. C.1 and fig. C.2.

C.3 'New world' cutoff cost levels

C.3.1 Dissociation cutoff

Consumers of type $i = n, e$ are indifferent to identification with society at large if the benefits from identification (i.e., the group status) are offset by disutility caused by inequality between non-caring and caring consumers. Note that this approach is equivalent to setting $U^{(1,1,1)}(c_l^F) = U^{(0,0,1)}(c_l^F)$ which reduces to:

$$A^X + \alpha^X \bar{\nu}^X(\hat{c}_l^i) = \beta^X(1-\delta)(1-\lambda_i)\nu_n(\hat{c}_l^i) \; .$$

The cutoff level is therefore given by:

$$\hat{c}_l^i \;=\; \left(\frac{\Phi\left[\beta^X(1-\delta)(1-\lambda_i) - \alpha^X(\lambda_n+\delta\lambda_e)\right]}{A^X}\right)^{\frac{1}{\gamma}} \; . \tag{C.4}$$

For $c_l^F < \hat{c}_l^i$ consumers of type i dissociate from society. By definition $c_l^F > 0$. Hence, for $\hat{c}_l^i$ being an interior cutoff we need $\frac{\alpha^X(\lambda_n+\delta\lambda_e)}{(1-\delta)(1-\lambda_i)} < \beta^X$, which constitutes the first element of assumption 5. For uniqueness, we derive utility from identification with society with respect to c_l^F, which strictly monotonically increases in cost c_l^F iff $\frac{\alpha^X(\lambda_n+\delta\lambda_e)}{(1-\delta)(1-\lambda_i)} < \beta^X$. Therefore, assumption 5 assures existence and uniqueness of the dissociation cutoff $\hat{c}_l^i$.

Note that the dissociation cutoff of the larger group is lower than for the smaller group. The reason is that dissonance costs increase in the distance between the individual and the group average. When an individual is part of the majority group, it is also closer to the average and is therefore prone to identification, while the minority group will dissociate for relatively low levels of inequality, i.e. for relatively high cost levels c_l^F (see equations 6 and 7). The two distinct cutoffs are illustrated by fig. C.1 for $\lambda_n < \lambda_e$ in which $\hat{c}_l^n > \hat{c}_l^e$ and by fig. C.2 for $\lambda_n > \lambda_e$ in which $\hat{c}_l^n < \hat{c}_l^e$. For our analysis, it is particularly relevant that caring consumers initially identify with society at large in order to be able to investigate NGO protests leading to a polarization of caring consumers. Therefore, the relevant dissociation cutoff for us is $\hat{c}_l^e$ and we focus on the cost range $c_l^F > \hat{c}_l^e$. In fig. C.1 this comprises range a, b and c; in fig. C.2 range a and b.

Caring consumers' preference for the high standard. Moreover, we derive a condition for $\hat{c}_l^e$ to be above c_l^e where the latter determines the range in which caring consumers prefer the high standard (for the derivation of c_l^e see appendix A.1). Like this, we assure that in the relevant parameter space for the TTIP- and CETA-cases all caring consumers prefer production under the high standard (which they may not, for example, when the low-standard productions costs go to zero). This is also conceptually important, as preference for the high standard is a prerequisite for NGO-identification. A priori identification with society at large by caring consumers as well as caring consumers' preference for the high standard are essential for the emergence of the TTIP- and CETA-cases (which we analyze in appendix C.4 and which can occur in range b and c in fig. C.1 and range b in fig. C.2). Initial identification with society by caring consumers is necessary to analyze the (effects of a) polarization of caring consumers which takes the form of (some) caring consumers dissociating from society at large. This implies that the cost level c_h has to be sufficiently low, so that in the relevant cost range $c_l^F > \hat{c}_l^e$ caring consumers prefer the high standard. Therefore, we take eq. (A.1) and derive under which condition caring consumers prefer the high standard at $c_l^F = \hat{c}_l^e$. This can be easily done by plugging eq. (C.4) for $i = e$ into eq. (A.1) which delivers the following condition. There exists a cost range $c_l^e \leq \hat{c}_l^i < c_l^F$, in which consumers identify with society at large and caring consumers always prefer the high standard iff:

$$c_h < \left(\frac{\Phi \left(\beta^X (1-\delta)(1-\lambda_e) - \alpha^X \left(\lambda_n + \lambda_e \delta \right) \right)}{\delta A^X} \right)^{\frac{1}{\gamma}} . \tag{C.5}$$

This condition constitutes the second element of assumption 6.

C.3.2 Cutoff for the low standard: case (1,1,1)

We now turn to the derivation of the ratification cutoffs for the four identification regimes. These are determined by the intersection of the horizontal line with the four convex curves. For the case $(1, 1, 1)$, where both groups identify with society, the government signs the trade agreement when $U^{(1,1,1)}(c_h) < U^{(1,1,1)}(c_l^F)$.

Derivation of the cutoff. Setting $U^{(1,1,1)}(c_h) = U^{(1,1,1)}(c_l^F)$ and solving for c_l^F delivers the cutoff $\bar{c}_l$:

$$\bar{c}_l = c_h \left(\frac{\left(1 + \alpha + \alpha^X\right) \left(\lambda_n + \delta \lambda_e\right) - 2\beta^X \lambda_n \lambda_e \left(1 - \delta\right)}{1 + \alpha + \alpha^X} \right)^{\frac{1}{\gamma}} . \tag{C.6}$$

For $(1, 1, 1)$ and $c_l^F > \bar{c}_l$ the government rejects the trade agreement.

Existence of an interior cutoff. By definition, $c_l^F \in (0, c_h)$. To prove that $\bar{c}_l \in (0, c_h)$, we compare $U^{(1,1,1)}(c_h)$ to $U^{(1,1,1)}(c_l^F)$ with $c_l^F = \{0, c_h\}$ and for a given value of c_h to find the optimal standard when c_l^F converges to its limits. Therefore, we solve eq. (C.6) for c_h

and consider it for the limits of c_l^F.

$$\lim_{c_l \to 0} c_l^F \left(\frac{1+\alpha+\alpha^X}{(1+\alpha+\alpha^X)(\lambda_n+\delta\lambda_e) - 2\beta^X\lambda_n\lambda_e(1-\delta)} \right)^{\frac{1}{\gamma}} = 0 \ .$$

As $c_h > 0$, the low standard is preferred when c_l approaches zero.

$$\begin{aligned} &\lim_{c_l^F \to c_h} c_l^F \left(\frac{1+\alpha+\alpha^X}{(1+\alpha+\alpha^X)(\lambda_n+\delta\lambda_e) - 2\beta^X\lambda_n\lambda_e(1-\delta)} \right)^{\frac{1}{\gamma}} \\ &= c_h \left(\frac{1+\alpha+\alpha^X}{(1+\alpha+\alpha^X)(\lambda_n+\delta\lambda_e) - 2\beta^X\lambda_n\lambda_e(1-\delta)} \right)^{\frac{1}{\gamma}} \ , \end{aligned}$$

which is smaller than c_h iff

$$\frac{1+\alpha+\alpha^X}{(1+\alpha+\alpha^X)(\lambda_n+\delta\lambda_e) - 2\beta^X\lambda_n\lambda_e(1-\delta)} \ < \ 1 \ .$$

Using assumption 5 this can be rewritten into:

$$1+\alpha+\alpha^X \ < \ -2\beta^X\lambda_n \ .$$

By definition, $\alpha, \alpha^X > 0$, therefore this inequality never holds. This implies that for sufficiently small distances between c_h and c_l^F (i.e. when $c_l^F \to c_h$) the domestic high standard production maximizes aggregate utility for the case $(1,1,1)$. As shown in appendix C.2, $U^{(1,1,1)}(c_l^F)$ decreases monotonically in c_l^F conditional on assumption 5, therefore $\bar{c}_l$ is a unique cutoff that lies between 0 and c_h. Overall, assumption 5 assures existence and uniqueness of the cutoff $\bar{c}_l$.

C.3.3 Cutoff for the low standard: case (0,1,1)

For the case $(0,1,1)$, where only caring consumers identify with society, the government signs the trade agreement when $U^{(1,1,1)}(c_h) < U^{(0,1,1)}(c_l^F)$. Setting $U^{(1,1,1)}(c_h) = U^{(0,1,1)}(c_l^F)$ and solving for c_l^F delivers the cutoff $\breve{c}_l$:

$$\breve{c}_l \ = \ \left(\frac{\Phi(1+\alpha+\lambda_e\alpha^X)(\lambda_n+\delta\lambda_e) - \beta^X\lambda_n\lambda_e(1-\delta)}{(1+\alpha+\alpha^X)\Phi c_h^{-\gamma} + \lambda_n A^X} \right)^{\frac{1}{\gamma}} \ . \tag{C.7}$$

For $(0,1,1)$ and $c_l^F < \breve{c}_l$ the government accepts the trade agreement

Existence of an interior cutoff. By definition, $c_l^F \in (0, c_h)$. Given assumption 5, this cutoff is larger than zero. To prove that $\breve{c}_l < c_h$, we compare $U^{(1,1,1)}(c_h)$ to $U^{(0,1,1)}(c_l^F)$ for a given value of c_h to find the optimal standard when c_l^F converges to c_h. Therefore, we solve eq. (C.7) for c_h and consider it for the upper limit of c_l^F.

$$\lim_{c_l^F \to c_h} \left(\frac{1+\alpha+\alpha^X}{(c_l^F)^{-\gamma}\left[(1+\alpha+\lambda_e\alpha^X)(\lambda_n+\delta\lambda_e) - \beta^X\lambda_n\lambda_e(1-\delta)\right] - \lambda_n A^X} \right)^{\frac{1}{\gamma}}$$

$$= \left(\frac{1+\alpha+\alpha^X}{c_h^{-\gamma}\left[\left(1+\alpha+\lambda_e\alpha^X\right)\left(\lambda_n+\delta\lambda_e\right)-\beta^X\lambda_n\lambda_e\left(1-\delta\right)\right]-\lambda_n A^X} \right)^{\frac{1}{\gamma}} ,$$

which is smaller than c_h iff

$$\frac{\left(1+\alpha+\alpha^X\right)}{c_h^{-\gamma}\left[\left(1+\alpha+\lambda_e\alpha^X\right)\left(\lambda_n+\delta\lambda_e\right)-\beta^X\lambda_n\lambda_e\left(1-\delta\right)\right]-\lambda_n A^X} \;<\; c_h^{\gamma} .$$

Under assumption 5 this can be rewritten to:

$$(1-\delta)\lambda_e(1+\alpha)+\underbrace{\left(1-\lambda_n\lambda_e-\delta\lambda_e^2\right)}_{>0}\alpha^X \;<\; -\beta^X\lambda_n\lambda_e\left(1-\delta\right)-\lambda_n A^X c_h^{\gamma}$$

This inequality never holds. Hence, the government rejects the agreement when c_l^F converges to c_h. As shown in appendix C.2, $U^{(0,1,1)}(c_l^F)$ decreases monotonically in c_l^F conditional on assumption 5, therefore $\breve{c}_l$ is a unique cutoff that lies between 0 and c_h. Assumption 5 therefore assures existence and uniqueness of the cutoff $\breve{c}_l$.

C.3.4 Cutoff for the low standard: case (1,0,1)

For the case $(1,0,1)$, where both groups identify with society, the government signs the trade agreement when $U^{(1,1,1)}(c_h) < U^{(1,0,1)}(c_l^F)$.

Derivation of the cutoff. Setting $U^{(1,1,1)}(c_h) = U^{(1,0,1)}(c_l^F)$ and solving for c_l^F delivers the cutoff $\dot{c}_l$:

$$\dot{c}_l \;=\; \left(\frac{\Phi(1+\alpha+\lambda_n\alpha^X)(\lambda_n+\delta\lambda_e)-\beta^X\lambda_n\lambda_e\left(1-\delta\right)}{(1+\alpha+\alpha^X)\Phi c_h^{-\gamma}+\lambda_e A^X} \right)^{\frac{1}{\gamma}} . \tag{C.8}$$

For $(1,0,1)$ and $c_l^F < \dot{c}_l$ the government accepts the trade agreement.

Existence of an interior cutoff. By definition, $c_l^F \in (0, c_h)$. Given assumption 5, this cutoff is larger than zero. To prove that $\dot{c}_l < c_h$, we compare $U^{(1,1,1)}(c_h)$ to $U^{(1,0,1)}(c_l^F)$ for a given value of c_h to find the optimal standard when c_l^F converges to c_h. Therefore, we solve eq. (C.8) for c_h and consider it for the upper limit of c_l^F.

$$\lim_{c_l^F\to c_h} \left(\frac{1+\alpha+\alpha^X}{(c_l^F)^{-\gamma}\left[\left(1+\alpha+\lambda_n\alpha^X\right)\left(\lambda_n+\delta\lambda_e\right)-\beta^X\lambda_n\lambda_e\left(1-\delta\right)\right]-\lambda_e A^X} \right)^{\frac{1}{\gamma}}$$
$$= \left(\frac{1+\alpha+\alpha^X}{c_h^{-\gamma}\left[\left(1+\alpha+\lambda_n\alpha^X\right)\left(\lambda_n+\delta\lambda_e\right)-\beta^X\lambda_n\lambda_e\left(1-\delta\right)\right]-\lambda_e A^X} \right)^{\frac{1}{\gamma}} ,$$

which is smaller than c_h iff

$$\frac{1+\alpha+\alpha^X}{c_h^{-\gamma}\left[\left(1+\alpha+\lambda_n\alpha^X\right)\left(\lambda_n+\delta\lambda_e\right)-\beta^X\lambda_n\lambda_e\left(1-\delta\right)\right]-\lambda_e A^X} \;<\; c_h^{\gamma} .$$

Using assumption 5 this can be rewritten to:

$$(1-\delta)\lambda_e(1+\alpha)+\underbrace{\left(1-\lambda_n^2-\delta\lambda_e\lambda_n\right)}_{>0}\alpha^X \;<\; -\beta^X\lambda_n\lambda_e\,(1-\delta)-\lambda_e A^X c_h^{\gamma}\,.$$

This inequality never holds. Hence, the government rejects the agreement when $c_l^F \to c_h$. As shown in appendix C.2, $U^{(1,0,1)}(c_l)$ decreases monotonically in c_l, therefore $\dot{c}_l$ is a unique cutoff that lies between 0 and c_h. Assumption 5 therefore ensures existence as well as uniqueness of the cutoff $\dot{c}_l$.

C.3.5 Cutoff for the low standard: case (0,0,1)

For the case $(0,0,1)$, where both groups dissociate from society, the government signs the trade agreement when $U^{(1,1,1)}(c_h) < U^{(0,0,1)}(c_l^F)$. Setting $U^{(1,1,1)}(c_h) = U^{(0,0,1)}(c_l^F)$ and solving for c_l^F delivers the cutoff:

$$\tilde{c}_l \;=\; \left(\frac{(1+\alpha)(\lambda_n+\delta\lambda_e)\Phi}{(1+\alpha+\alpha^X)\Phi c_h^{-\gamma}+A^X}\right)^{\frac{1}{\gamma}}\;. \tag{C.9}$$

For $(0,0,1)$ and $c_l^F < \tilde{c}_l$ the trade agreement maximizes aggregate welfare.

Existence of an interior cutoff. By definition, $c_l^F \in (0, c_h)$. Obviously, this cutoff is larger than zero. To prove that $\tilde{c}_l < c_h$, we compare $U^{(1,1,1)}(c_h)$ to $U^{(0,0,1)}(c_l^F)$ for a given value of c_h to find the optimal standard when c_l^F converges to c_h. Therefore, we solve eq. (C.9) for c_h and consider it for the upper limit of c_l^F.

$$\lim_{c_l^F\to c_h}\left(\frac{(1+\alpha+\alpha^X)\Phi}{(1+\alpha)(\lambda_n+\delta\lambda_e)\Phi(c_l^F)^{-\gamma}-A^X}\right)^{\frac{1}{\gamma}} = \left(\frac{(1+\alpha+\alpha^X)\Phi}{(1+\alpha)(\lambda_n+\delta\lambda_e)\Phi c_h^{-\gamma}-A^X}\right)^{\frac{1}{\gamma}}\;,$$

which is smaller than c_h iff

$$\frac{(1+\alpha+\alpha^X)\Phi}{(1+\alpha)(\lambda_n+\delta\lambda_e)\Phi c_h^{-\gamma}-A^X} \;<\; c_h^{\gamma}$$

$$c_h \;<\; \left(\frac{\Phi\left[(1+\alpha)(\lambda_n+\delta\lambda_e)-(1+\alpha+\alpha^X)\right]}{A^X}\right)^{\frac{1}{\gamma}}\;,$$

which never holds since the right-hand side is negative. This implies that domestic high-standard production becomes optimal when $c_l^F \to c_h$. As shown in appendix C.2, $U^{(0,0,1)}(c_l^F)$ decreases monotonically in c_l^F, therefore $\tilde{c}_l$ is a unique cutoff that lies between 0 and c_h.

C.3.6 Cutoff ordering

There exist several possible cutoff orderings which crucially depend on the level of c_h that determines the location of the horizontal line $U^{(1,1,1)}(c_h)$ which may shift up or down for different values of c_h. We now derive a parameter condition for cutoff orderings that include the range of c_l^F in which the trade agreement is welfare-maximizing and either social cohesion or identification of caring consumers with society at large is preserved. This range is defined by $c_l^F \in (\hat{c}_l^e, max\{\check{c}_l, \bar{c}_l\})$. A sufficient condition is that $U^{(1,1,1)}(c_h) < U^{(1,1,1)}(c_l^F)$ for $c_l^F = \hat{c}_l^e$, i.e., at $c_l^F = \hat{c}_l^e$ the horizontal line lies below the curve $U^{(1,1,1)}(c_l^F)$ implying $\hat{c}_l^e < \bar{c}_l$. Therefore, we derive a lower bound for c_h (pinning down the maximum height of the horizontal line that still assures the emergence of the TTIP- and CETA-cases). From appendix C.3.2, we already know that $U^{(1,1,1)}(c_h) < U^{(1,1,1)}(\hat{c}_l^e)$ can be reduced to:

$$(1+\alpha+\alpha^X)c_h^{-\gamma} \; < \; (\hat{c}_l^e)^{-\gamma}\left[\left(1+\alpha+\alpha^X\right)(\lambda_n+\delta\lambda_e) - 2\beta^X\lambda_n\lambda_e\,(1-\delta)\right] .$$

Plugging in eq. (C.4) and solving for c_h gives:

$$c_h \; > \; \left(\frac{(1+\alpha+\alpha^X)\Phi\left[\beta^X(1-\delta)\lambda_n - \alpha^X(\lambda_n+\delta\lambda_e)\right]}{\left[(1+\alpha+\alpha^X)\,(\lambda_n+\delta\lambda_e) - 2\beta^X\lambda_n\lambda_e\,(1-\delta)\right]A^X}\right)^{\frac{1}{\gamma}} . \qquad \text{(C.10)}$$

This condition is consistent with the lower bound on c_h stated in assumption 6. Under assumption 5, the right-hand side of this inequality is positive and a binding lower bound for c_h. Note that the right-hand side of the inequality above depends on the relation of λ_n and λ_e. The cutoff $\hat{c}_l^e$ is larger for $\lambda_n > \lambda_e$ than for $\lambda_n < \lambda_e$. Therefore, when $\lambda_n > \lambda_e$, the horizontal line $U^{(1,1,1,)(c_l^F)}$ has to be lower (so that $\hat{c}_l^e < \bar{c}_l$), which requires a higher value of c_h. This implies that for $\lambda_n < \lambda_e$ the cost range in which the trade agreement is ratified and social cohesion or identification of caring consumers with society is maintained (fig. C.1 range *b* and *c*) and in which the TTIP- and CETA-case can occur is larger than for $\lambda_n > \lambda_e$ (fig. C.2 range *b*).

Figures C.1 and C.2 each represent only one of the possible cutoff ordering that allow for the emergence of the TTIP- and CETA-cases as well. The cutoff ordering that applies is determined by the level of high-standard production cost c_h, i.e., the location of the horizontal line $U^{(1,1,1)}(c_h)$ in figures C.1 and C.2. For $\lambda_n < \lambda_e$, the condition for the emergence of the TTIP- and CETA-case is that the horizontal line lies below the intersection of the $(0,0,1)$ and the $(0,1,1)$ curve at $\hat{c}_l^e$. This means that the horizontal line in fig. C.1 could shift upwards to a point where all intersections of the horizontal with the four curves lie within the range $c_l^F \in (\hat{c}_l^e, \hat{c}_l^n)$, i.e. all four ratification cutoffs, $\tilde{c}_l, \dot{c}_l, \check{c}_l, \bar{c}_l$, lie within this range. For $\lambda_n > \lambda_e$, the condition for the emergence of the TTIP- and the CETA-case is that the horizontal line lies below the intersection of the $(1,0,1)$ and the $(1,1,1)$ curve at $\hat{c}_l^e$. This implies that the horizontal line could cross the $(0,0,1)$ and the $(0,1,1)$ curve already below $\hat{c}_l^e$.

C.4 Proof of the TTIP- and CETA-case

C.4.1 Fragmentation of caring consumers

As in the analysis with equal population shares, we are interested in the case where NGO protests against the negotiated trade agreement trigger fully congruent caring consumers (fraction $1-\eta$ with $\tilde{\theta} = 1$) to dissociate from society, while partially congruent caring consumers (fraction η with $\tilde{\theta} = \theta < 1$) dissociate from the NGO. Therefore, we derive conditions that allow for this interior solution – i.e. the fragmentation of caring consumers – in cost range $c_l^F \in (max\{\tilde{c}_l, \dot{c}_l\}, max\{\breve{c}_l, \bar{c}_l\})$. Caring consumers are indifferent between identification with society at large and identification with the NGO when the last two terms in eq. (7) are equalized:

$$\tilde{\theta} A^N = A^X + \alpha^X \bar{\nu}^X(\hat{c}_l^N) - \beta^X(1-\lambda_e)\left[\nu_n(\hat{c}_l^N) - \nu_e(\hat{c}_l^N)\right] .$$

Plugging in $\bar{\nu}^X(\hat{c}_l^N)$, $\nu_n(\hat{c}_l^N)$ and $\nu_e(\hat{c}_l^N)$, this can be rewritten into

$$\tilde{\theta} A^N = A^X - \Phi(\hat{c}_l^N)^{-\gamma}\left(\beta^X \lambda_n(1-\delta) - \alpha^X(\lambda_n + \delta\lambda_e)\right)$$

and solved for cutoff $\hat{c}_l^N$

$$\hat{c}_l^N = \left(\frac{\Phi\left(\beta^X \lambda_n(1-\delta) - \alpha^X(\lambda_n + \delta\lambda_e)\right)}{A^X - \tilde{\theta} A^N}\right)^{\frac{1}{\gamma}} . \tag{C.11}$$

Caring consumers identify with the NGO (society at large) for $c_l^F < \hat{c}_l^N$ ($c_l^F > \hat{c}_l^N$). Assumption 5 and assumption 7 ensure that the threshold $\hat{c}_l^N$ (eq. (C.11)) is positive. Moreover, assumption 7 implies that $\hat{c}_l^N$ lies above the dissociation cutoff $\hat{c}_l^e$ (eq. (A.5)), i.e. in the cost range where identification with society is beneficial for caring consumers. From eq. (C.11) follows that $\hat{c}_l^N|_{\tilde{\theta}=\theta} < \hat{c}_l^N|_{\tilde{\theta}=1}$. For $c_l^F \in (\hat{c}_l^N|_{\tilde{\theta}=\theta}, \hat{c}_l^N|_{\tilde{\theta}=1})$, a fragmentation of caring consumers occurs. To show that a polarization of caring consumers can generate the TTIP-case, it is sufficient to prove that the range $c_l^F \in (\hat{c}_l^N|_{\tilde{\theta}=\theta}, \hat{c}_l^N|_{\tilde{\theta}=1})$ can cover the range $c_l^F \in (\hat{c}_l^e, max\{\breve{c}_l, \bar{c}_l\})$ that includes cost levels for which the TTIP-case may occur.[1] Hence, we consider the limits of eq. (A.10):

$$\lim_{\tilde{\theta}\to 0}\left(\frac{\Phi\left(\beta^X \lambda_n(1-\delta) - \alpha^X(\lambda_n + \delta\lambda_e)\right)}{A^X - 0\cdot A^N}\right)^{\frac{1}{\gamma}} = \left(\frac{\Phi\left(\beta^X \lambda_n(1-\delta) - \alpha^X(\lambda_n + \delta\lambda_e)\right)}{A^X}\right)^{\frac{1}{\gamma}} = \hat{c}_l^e ,$$

$$\lim_{\tilde{\theta}\to 1, A^N\to A^X}\left(\frac{\Phi\left(\beta^X \lambda_n(1-\delta) - \alpha^X(\lambda_n + \delta\lambda_e)\right)}{A^X - 1\cdot A^X}\right)^{\frac{1}{\gamma}} = \left(\frac{\Phi\left(\beta^X \lambda_n(1-\delta) - \alpha^X(\lambda_n + \delta\lambda_e)\right)}{0}\right)^{\frac{1}{\gamma}} = \infty .$$

From this follows that for a sufficiently low θ the cutoff $\hat{c}_l^N|_{\tilde{\theta}=\theta}$ lies below the range where the TTIP-case can occur, while for a sufficiently large θ and A^N the cutoff $\hat{c}_l^N|_{\tilde{\theta}=1}$ lies above the range where the TTIP-case can occur.

[1] For $\lambda_n < \lambda_e$ the range where the TTIP-case is possible is given by $c_l^F \in (max\{\tilde{c}_l, \dot{c}_l\}, max\{\breve{c}_l, \bar{c}_l\})$ (fig. C.1, range b and c) depending on the cost c_h, while for $\lambda_n > \lambda_e$ this range simply is $c_l^F \in (\dot{c}_l, \bar{c}_l)$ (fig. C.2 range b).

C.4.2 New 'new world' cutoff: case $(1, \eta, 1-\eta)$

The altered social identification leads to a shift of the cutoffs $\bar{c}_l$ and $\breve{c}_l$. The new cutoff $\bar{c}_l'$ is determined by $U^{(1,1,1)}(c_h) = U^{(1,\eta,1-\eta)}(\bar{c}_l')$:

$$\leftrightarrow \quad \bar{c}_l'(\eta) \;=\; \left(\frac{\Phi\left[\left(1+\alpha+(\lambda_n+\eta\lambda_e)\alpha^X\right)(\lambda_n+\delta\lambda_e)-(1+\eta)\,\beta^X\lambda_n\lambda_e(1-\delta)\right]}{(1+\alpha+\alpha^X)\Phi c_h^{-\gamma}+(1-\eta)\lambda_e A^X+\lambda_e A^N\eta\theta} \right)^{\frac{1}{\gamma}} . \tag{C.12}$$

For $(1, \eta, 1-\eta)$ and $c_l^F > \bar{c}_l'$ the government rejects the trade agreement. To prove that $\bar{c}_l' \in (0, c_h)$, we compare $U^{(1,1,1)}(c_h)$ to $U^{(1,\eta,1-\eta)}(c_l^F)$ with $c_l^F = \{0, c_h\}$ for a given value of c_h to find the optimal standard when c_l^F converges to its limits. Therefore, we solve eq. (C.12) for c_h and consider it for the limits of c_l^F.

$$\lim_{c_l^F \to 0} \left(\frac{(1+\alpha+\alpha^X)\Phi}{\left[\left(1+\alpha+(\lambda_n+\eta\lambda_e)\alpha^X\right)(\lambda_n+\delta\lambda_e)-(1+\eta)\beta^X\lambda_n\lambda_e(1-\delta)\right]\Phi(c_l^F)^{-\gamma}-(1-\eta)\lambda_e A^X-\lambda_e A^N\eta\theta} \right)^{\frac{1}{\gamma}}$$
$$= 0 \,,$$

as the denominator goes to infinity. Since $c_h > 0$, the low standard is preferred when c_l^F approaches zero.

$$\lim_{c_l^F \to c_h} \left(\frac{(1+\alpha+\alpha^X)\Phi}{\left[\left(1+\alpha+(\lambda_n+\eta\lambda_e)\alpha^X\right)(\lambda_n+\delta\lambda_e)-(1+\eta)\beta^X\lambda_n\lambda_e(1-\delta)\right]\Phi c_h^{-\gamma}-(1-\eta)\lambda_e A^X-\lambda_e A^N\eta\theta} \right)^{\frac{1}{\gamma}} ,$$

which is smaller than c_h iff

$$\frac{(1+\alpha+\alpha^X)\Phi}{\left[\left(1+\alpha+(\lambda_n+\eta\lambda_e)\alpha^X\right)(\lambda_n+\delta\lambda_e)-(1+\eta)\beta^X\lambda_n\lambda_e(1-\delta)\right]\Phi c_h^{-\gamma}-(1-\eta)\lambda_e A^X-\lambda_e A^N\eta\theta} \;<\; c_h^{\gamma} .$$

Since assumption 5 implies that $\left(1+\alpha+(\lambda_n+\eta\lambda_e)\alpha^X\right)(\lambda_n+\delta\lambda_e)-(1+\eta)\beta^X\lambda_n\lambda_e(1-\delta) > 0$, this can be rewritten into:

$$c_h \;<\; \left(\frac{\Phi\left[\left(1+\alpha+(\lambda_n+\eta\lambda_e)\alpha^X\right)(\lambda_n+\delta\lambda_e)-(1+\eta)\beta^X\lambda_n\lambda_e(1-\delta)-(1+\alpha+\alpha^X)\right]}{(1-\eta)\lambda_e A^X+\lambda_e A^N\eta\theta} \right)^{\frac{1}{\gamma}} .$$

This inequality never holds since the right-hand side is negative. This implies that aggregate utility is maximized by the high standard when c_l^F converges to c_h. Overall, assumption 5 ensures existence as well as uniqueness of the cutoff $\bar{c}_l'$.

C.4.3 New 'new world' cutoff: case $(0, \eta, 1-\eta)$

The new cutoff $\breve{c}_l'$ is determined by $U^{(1,1,1)}(c_h) = U^{(0,\eta,1-\eta)}(\breve{c}_l')$:

$$\breve{c}_l'(\eta) \;=\; \left(\frac{\Phi\left[\left(1+\alpha+\eta\lambda_e\alpha^X\right)(\lambda_n+\delta\lambda_e)-\eta\lambda_e\beta^X\lambda_n(1-\delta)\right]}{(1+\alpha+\alpha^X)\Phi c_h^{-\gamma}+(1-\eta)\lambda_e A^X+\lambda_e A^N\eta\theta} \right)^{\frac{1}{\gamma}} \tag{C.13}$$

For $(0, \eta, 1-\eta)$ and $c_l^F > \breve{c}_l'$ the government rejects the trade agreement. To prove that $\breve{c}_l' \in (0, c_h)$, we compare $U^{(1,1,1)}(c_h)$ to $U^{(0,\eta,1-\eta)}(c_l^F)$ with $c_l^F = \{0, c_h\}$ for a given value of c_h to find the optimal standard when c_l^F converges to its limits. Therefore, we solve eq. (C.13) for c_h and consider it for the limits of c_l^F.

$$\lim_{c_l^F \to 0} \left(\frac{(1+\alpha+\alpha^X)\Phi}{\left[\left(1+\alpha+\eta\lambda_e\alpha^X\right)(\lambda_n+\delta\lambda_e) - \eta\beta^X\lambda_n\lambda_e(1-\delta)\right]\Phi(c_l^F)^{-\gamma} - (1-\eta)\lambda_e A^X - \lambda_e A^N\eta\theta} \right)^{\frac{1}{\gamma}}$$
$$= 0 \ ,$$

as the denominator goes to infinity. Since $c_h > 0$, the low standard is preferred when c_l^F approaches zero.

$$\lim_{c_l^F \to c_h} \left(\frac{(1+\alpha+\alpha^X)\Phi}{\left[\left(1+\alpha+\eta\lambda_e\alpha^X\right)(\lambda_n+\delta\lambda_e) - \eta\beta^X\lambda_n\lambda_e(1-\delta)\right]\Phi c_h^{-\gamma} - (1-\eta)\lambda_e A^X - \lambda_e A^N\eta\theta} \right)^{\frac{1}{\gamma}} ,$$

which is smaller than c_h iff

$$\frac{(1+\alpha+\alpha^X)\Phi}{\left[\left(1+\alpha+\eta\lambda_e\alpha^X\right)(\lambda_n+\delta\lambda_e) - \eta\beta^X\lambda_n\lambda_e(1-\delta)\right]\Phi c_h^{-\gamma} - (1-\eta)\lambda_e A^X - \lambda_e A^N\eta\theta} \ < \ c_h^{\gamma} \ .$$

Since assumption 5 implies that $\left(1+\alpha+(\lambda_n+\eta\lambda_e)\alpha^X\right)(\lambda_n+\delta\lambda_e) - (1+\eta)\beta^X\lambda_n\lambda_e(1-\delta) > 0$, this can be rewritten into:

$$c_h \ < \ \left(\frac{\Phi\left[\left(1+\alpha+\eta\lambda_e\alpha^X\right)(\lambda_n+\delta\lambda_e) - \eta\beta^X\lambda_n\lambda_e(1-\delta) - (1+\alpha+\alpha^X)\right]}{(1-\eta)\lambda_e A^X + \lambda_e A^N\eta\theta} \right)^{\frac{1}{\gamma}} \ .$$

This inequality never holds since the right-hand side is negative. This implies that aggregate utility is maximized by the high standard when c_l^F converges to c_h. Overall, assumption 5 ensures existence as well as uniqueness of the cutoff $\breve{c}_l'$.

C.4.4 Cutoff ordering

We now determine conditions for a cutoff ordering that allows for the TTIP- and CETA-case to appear. As with equal population shares, a sufficient condition is that $U^{(1,1,1)}(c_h) < U^{(1,\eta,(1-\eta))}(c_l^F)$ at $c_l^F = \hat{c}_l^e$. From eq. (C.12), we already know that $U^{(1,1,1)}(c_h) < U^{(1,\eta,(1-\eta))}(c_l^F)$ boils down to:

$$(1+\alpha+\alpha^X)\Phi c_h^{-\gamma} + (1-\eta)\lambda_e A^X + \lambda_e A^{NGO}\eta\theta$$
$$< \ \Phi(\hat{c}_l^e)^{-\gamma}\left[(1+\alpha+(\lambda_n+\eta\lambda_e)\alpha^X)(\lambda_n+\delta\lambda_e) - (1+\eta)\beta^X\lambda_n\lambda_e(1-\delta)\right] \ .$$

Plugging in $\hat{c}_l^e$ as given by eq. (C.4) gives:

$$(1+\alpha+\alpha^X)\Phi c_h^{-\gamma} + (1-\eta)\lambda_e A^X + \lambda_e A^N\eta\theta$$

$$< \ \Phi \left[\left(\frac{\Phi \left[\beta^X (1-\lambda_e)(1-\delta) - \alpha^X (\lambda_n + \delta\lambda_e)\right]}{A^X} \right)^{\frac{1}{\gamma}} \right]^{-\gamma}$$

$$\cdot \left[(1+\alpha+(\lambda_n+\eta\lambda_e)\alpha^X)(\lambda_n+\delta\lambda_e) - (1+\eta)\beta^X \lambda_n \lambda_e (1-\delta) \right] \ ,$$

which can be solved for c_h:

$$c_h \ > \ \left(\frac{(1+\alpha+\alpha^X)\Phi\left[\beta^X\lambda_n(1-\delta) - \alpha^X(\lambda_n+\delta\lambda_e)\right]}{A^X\left[(1+\alpha+\alpha^X)(\lambda_n+\delta\lambda_e) - 2\lambda_e\lambda_n(1-\delta)\beta^X\right] - \lambda_e A^N \eta\theta \left[\beta^X\lambda_n(1-\delta) - \alpha^X(\lambda_n+\delta\lambda_e)\right]} \right)^{\frac{1}{\gamma}} \ . \tag{C.14}$$

Given assumption 5 part one, the numerator in eq. (C.14) is positive. The threshold for the cost level c_h that allows for the TTIP- and the CETA-case to occur is interior if the denominator of eq. (C.14) is positive as well. This is the case iff:

$$\beta^X < \frac{\left(1+\alpha+(1+\lambda_e\eta\theta)\alpha^X\right)(\lambda_n+\delta\lambda_e)}{(2+\eta\theta)\lambda_e\lambda_n(1-\delta)} \ ,$$

This condition is less strict than the upper bound for β^X stated in assumption 5 and therefore always fulfilled given our parameter restrictions.

Condition (C.14) constitutes the first element of assumption 6 since it is more restrictive on c_h than eq. (C.10). Considering figures C.1 and C.2, this implies that the emergence of the TTIP- and CETA-case requires a relatively higher c_h (i.e., lower horizontal line) compared to the initial situation in appendix C.3.6. To show this, we compare the right-hand side of eq. (C.14) with the right-hand side of eq. (C.10). This boils down to the following condition:

$$0 > -\lambda_e A^N \eta\theta \left[\beta^X\lambda_n(1-\delta) - \alpha^X(\lambda_n+\delta\lambda_e)\right] \ . \tag{C.15}$$

Since the right-hand side is negative, this inequality always holds.

C.5 Consistency of assumption 6

To assure that assumption 6 defines a non-empty set of values of c_h, note that the inequalities in assumption 6 hold iff:

$$A^N \frac{\lambda_e\eta\theta\left[\beta^X\lambda_n(1-\delta) - \alpha^X(\lambda_n+\delta\lambda_e)\right]}{\lambda_n(1-\delta)\left[M - 2\lambda_e\beta^X\right]} < A^X \ . \tag{C.16}$$

With assumption 7, a sufficient condition for this condition to hold is that the scaling factor of A^N on the left-hand side of the inequality above is smaller than one. This is the case iff:

$$\beta^X < \frac{(1+\alpha+\alpha^X)\lambda_n(1-\delta) + \lambda_e\eta\theta\alpha^X(\lambda_n+\delta\lambda_e)}{(2+\eta\theta)\lambda_e\lambda_n(1-\delta)} \ . \tag{C.17}$$

This condition delivers the upper bound of assumption 5, which therefore assures that there exists a set of values of c_h (stated in assumption 6) that allows for the TTIP- and the CETA-case to occur.