CESifo WORKING PAPERS

9920
2022

August 2022

The Costs and Benefits of Rules of Origin in Modern Free Trade Agreements

Emanuel Ornelas, John L. Turner

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 9920*

The Costs and Benefits of Rules of Origin in Modern Free Trade Agreements

Abstract

We study the welfare impact of rules of origin in free trade agreements where final-good producers source customized inputs from suppliers within the trading bloc. We employ a property-rights framework that features hold-up problems in suppliers' decisions to invest, and where underinvestment is more severe for higher productivity firms. A rule of origin offers preferred market access for final goods if a sufficiently high fraction of inputs used in the production process is sourced within the trading bloc. Such a rule alters behavior for only a subset of suppliers, as some (very-high-productivity) suppliers comply with the rule in an unconstrained way and some (very-low-productivity) suppliers choose not to comply. For those suppliers it does affect, the rule increases investment, but it also induces excessive sourcing (for given investment) within the trading bloc. From a social standpoint, it is best to have a rule that affects high-productivity suppliers. The reason is that the marginal net welfare gain from tightening the rule increases with productivity. Therefore, when industry productivity is high, a *strict* rule of origin is socially desirable; in contrast, when industry productivity is low, no rule of origin is likely to help. Regardless of the case, a sufficiently strict rule can (weakly) ensure welfare gains.

JEL-Codes: F130, F150, L220, D230.

Keywords: hold-up problem, sourcing, incomplete contracts, regionalism.

Emanuel Ornelas
Sao Paulo School of Economics-FGV
Sao Paulo / Brazil
eaornelas@gmail.com

John L. Turner
Department of Economics, Terry College of Business, University of Georgia
Athens / GA / USA
jlturner@uga.edu

August 24, 2022
We thank Bruno Daré for competent research assistance and Bernardo Guimaraes, Giovanni Maggi, John McLaren, Kamal Saggi, Bob Staiger, Feodora Teti, Dan Trefler and seminar participants at the 2022 NBER conference "Trade and Trade Policy in the 21st Century," the 2022 CESifo Area Conference on Global Economy, the 2022 GEP "Globalisation and Economic Policy" workshop, the 2022 Bielefeld "Trade Policy and Trade Agreements" workshop, the 2022 ESAM meetings, the Sao Paulo School of Economics-FGV, and the University of Georgia for helpful comments. Ornelas acknowledges financial support from the Brazilian Council of Science and Technology (CNPq).

1 Introduction

Since the conclusion of the Uruguay Round of multilateral negotiations in 1994, the main mode of trade liberalization has been through the formation of free trade agreements (FTAs). In tandem with this phenomenon, rules of origin (ROOs) have multiplied. Under these rules, goods whose inputs come substantially from within-bloc countries are traded freely, but otherwise face tariffs. Their formal goal is to prevent the trans-shipment of imported goods from low-tariff to high-tariff countries within the same trading bloc. Interestingly, Felbermayr et al. (2019) find the scope for such "trade deflection" to be extremely limited, due to transportation costs and to similarity of external tariffs. Nevertheless, ROOs are widespread, with their stringency varying substantially across agreements and across products within agreements. Generally, economists take a dim view of ROOs, as they tend to lower welfare in competitive models (Grossman, 1981; Krishna, 2006). The conventional view is that ROOs are distortionary because they "prevent final good producers from choosing the most efficient input suppliers around the world" (Conconi et al., 2018, p. 2336). Accordingly, the empirical literature customarily interprets findings that ROOs have a negative effect on imports from third countries as evidence of welfare-reducing trade diversion.

While this is true in some contexts, it may not be when within-FTA sourcing is inefficiently low absent ROOs. That is the case when firms need to make relationship-specific, unrecoverable investments that cannot be fully contracted upon, a situation that typically arises when inputs are customized. In the modern era of global sourcing, customization is indeed prevalent in many industries. This is particularly common in the context of FTAs, which both promote and are promoted by global/regional value chains (Baldwin, 2011; Johnson and Noguera, 2017; Ruta, 2017). In such cases, terms of trade are determined via bargaining and input suppliers may face hold-up problems. But then, if a ROO induces changes in input mixes, there is not necessarily a welfare loss, and there may be welfare gains. We capture and study these effects in this paper, where we provide the first analysis of ROOs using a property-rights model.

Building on Ornelas, Turner and Bickwit (2021) – henceforth OTB – we consider an environment where specialized input suppliers with heterogeneous productivity form vertical chains with producers of final goods. Both firms are in countries that belong to the same FTA. Within each vertical chain, the supplier invests in technology (marginal cost curve reduction) prior to bargaining with the producer over inputs; those investments are non-contractible. The producer also has the option to purchase generic inputs from a competitive world market, and the model is set up so that the producer always sources a mix of specialized and generic inputs in equilibrium. Under free trade, investments are inefficiently low due to the possibility of hold up. Crucially, because investment affects more units of inputs in more productive firms, the underinvestment problem is more severe for higher productivity firms.

To this basic setup we add a rule such that the final goods produced by a vertical chain enjoy preferred market access as long as the fraction of inputs produced within the bloc exceeds a prespecified level.[1] We find that this rule is more likely to increase aggregate welfare when it is *stricter*. The reason is as follows. A rule of origin will typically affect choices made by just a subset of suppliers, and a stricter rule targets higher productivity firms (under a strict rule, those with low productivity choose to forgo the benefits from compliance). For those high-productivity suppliers, a relatively strict rule is desirable because the marginal net welfare gain from tightening the rule increases with productivity. This happens because the extra investment that the rule induces is more socially beneficial when it comes from the suppliers more affected by the original underinvestment problem – i.e., those with higher productivity.

The heterogeneous incidence echoes both early and new work on ROOs (e.g., Grossman, 1981; Celik et al., 2020; Head et al., 2022), which show that a ROO matters only for particular levels of the supply curve and there are three cases to consider. In our model, we can classify the cases according to supplier productivity. The highest-productivity suppliers have the lowest marginal cost curves and produce very high levels of intermediate inputs. Their vertical chains comply with the ROO without altering investments, so the rule has no effect on their behavior. At the opposite extreme, the lowest-productivity suppliers have the highest marginal cost curves and produce very low levels of intermediate inputs. Because compliance would be too costly for them, their vertical chains are unwilling to comply with

[1]There are different ways to determine "origin" (see the detailed discussion by Inama, 2009). Kniahin and Melo (2022), studying 370 trade agreements, find that the most common definition is a minimal regional value content.

the ROO, so the rule does not affect their sourcing behavior either. Finally, vertical chains with suppliers within an intermediate productivity range find it optimal to source extra specialized inputs within the FTA to gain preferential access for their final goods. Naturally, this split is mediated by the level of the final-good tariff in the FTA partner, which determines the extent of the gains from compliance with the ROO.[2]

For the suppliers that comply in a constrained way, we say the rule is *binding*. The extra inputs serve as a commitment device for the supplier to invest more. This additional investment tends to improve efficiency. However, for given investment, there will also be excessive sourcing within the FTA – the effect on which most of the literature has concentrated. A socially optimal rule should trade off the benefits from mitigating the hold-up problem against the costs due to excessive within-FTA sourcing for the population of suppliers. Because the marginal gain from tightening the rule is higher, and the corresponding marginal cost is lower, for high-productivity firms, a hypothetical optimal rule set at the firm level would be increasing in productivity. Thus, a relatively strict rule is both binding for more productive suppliers and is particularly beneficial (from a social standpoint) when it affects their behavior.

Considering the whole distribution of suppliers, then, a stricter ROO is more likely to generate a positive welfare effect for all affected suppliers because it binds for more productive suppliers, and the potential welfare gains are higher when those suppliers are affected. Now, whether aggregate welfare rises or falls with the stringency of a ROO does depend on the specific distribution of supplier productivity. Nevertheless, a very lenient rule is likely to be harmful, because it affects the behavior primarily of firms whose original underinvestment problem is mild and whose excessive FTA sourcing due to the rule would be severe. In contrast, a sufficiently strict rule *ensures* a welfare gain, as it affects only the behavior of firms whose original underinvestment problem is severe and whose excessive FTA sourcing due to the rule would be mild. We highlight this with an example where productivity follows a Pareto(k) distribution. For any shape parameter k, the welfare effect is negative for low r

[2]Most obviously, if that tariff is close to zero, any ROO will be innocuous because vertical chains will have nearly nothing to gain by altering their sourcing choices. In contrast, if the final-good tariff is very high, then every supplier complies regardless of the rule. The tripartite case arises when the final-good tariff is not extreme.

and positive for high r, although what "high" or "low" r means does hinge on k.

An external tariff on intermediate inputs alters the welfare consequences of a ROO. A positive input tariff increases investment by all suppliers, leading some to overinvest. An tighter ROO is then needed to boost welfare, as it is necessary to tailor the ROO to affect higher-productivity suppliers that are still underinvesting. However, when the external tariff on inputs is sufficiently high, all suppliers overinvest and any binding ROO would only worsen welfare, because it is ill-suited to address overinvestment problems.

If we consider limiting cases of our environment, it is possible to find other situations where a ROO always worsens welfare. First, if there is no investment decision by the supplier, then there is no underinvestment problem and the ROO leads just to trade diversion. This is the case usually considered in the literature, which justifies the interpretation of many empirical analyses. Second, if suppliers have full bargaining ability, then investments are efficient without ROOs, and any rule will lead to trade diversion and excessive investments. Third, if suppliers have no bargaining ability, then ROOs cannot affect investments and, again, yield only trade diversion. Put together, these cases make clear that ROOs can be beneficial only if (1) hold-up problems are present and (2) investment responds to policy.

The paper is organized as follows. After discussing the related literature in the next subsection, we set up the basic model in section 2. In section 3, we study the choice of input mix for given level of investment. We then move to understand firms' choice of investment in section 4. This allows us to analyze the welfare impacts of ROOs in section 5. In section 6, we extend the analysis to the case where there is a strictly positive external tariff on inputs, and in section 7 we discuss modeling alternatives, extensions and positive implications of our model. We conclude in section 8.

1.1 Related Literature

The theoretical literature on ROOs goes back to Grossman (1981), who studies the consequences of local content requirements rules, including the case where access to preferential treatment for exports requires a minimum level of domestic value added. For that situation, in a competitive setting, he identifies the three cases of non-compliance, unconstrained compliance, and constrained compliance, as we also find in our environment. Krishna (2006) offers a similar perspective. Also in a competitive model, Falvey and Reed (1998) show how ROOs can distort allocative efficiency and underline the many ways in which "origin" can be defined. In a recent paper, Chung and Perroni (2021) depart from the competitive benchmark to study the effects of ROOs when markets are oligopolistic and show that stricter ROOs tend to lead to higher prices for intermediate goods.

Recently, Head et al. (2022) evaluate the effect of NAFTA's ROOs for auto parts. They use a structural model that yields a "ROO Laffer curve" with respect to the share of intra-bloc inputs used. The curve reflects the tripartite separation of firms regarding compliance depending on their cost structures. Intuitively, at low levels, tightening the rule induces more within-bloc sourcing, but eventually the effect reverses, as more firms choose to not comply under a strict rule. A similar pattern emerges in our model as well.

Naturally, the motives behind ROOs are diverse, and raising aggregate welfare may be far from the main goal in some circumstances. For example, as Krueger (1999) forcefully notes, ROOs can be imposed for protectionist reasons and constitute a source of economic inefficiency in FTAs. In turn, Celik et al. (2020) study the optimal design of ROOs when countries use them to affect the distribution of gains from an FTA. More generally, Maggi et al. (2022) show that there is a rationale for using non-tariff barriers – such as ROOs – when trade agreements restrict the use of tariffs. We do not study how or why ROOs are actually chosen, but we show that a relatively strict rule can be welfare improving, regardless of its motivations.[3]

Empirically, several papers evaluate the impact of ROOs on trade flows from a reduced-form perspective (see, for example, the studies in Cadot et al., 2006). In contrast, Cherkashin et al. (2015) structurally estimate an heterogeneous-firm model with firm-market specific demand shocks to study the impact of changing the costs of meeting ROOs faced by Bangladeshi apparel exporters. They find that fewer requirements for meeting ROOs are associated with

[3]In Celik et al.'s (2020) model, because firm entry is inefficiently low due to backward and forward linkages, a strict ROO can be welfare-enhancing when those linkages are sufficiently strong. Tsirekidze (2021), building on the theoretical analysis of FTA formation by Saggi and Yildiz (2010), shows instead that ROOs could be useful to facilitate the achievement of global free trade. Although in a very distinct context, our analysis connects with those in that we also uncover a potentially beneficial role for strict ROOs.

greater exports and more entry in the long run. Conconi et al. (2018) provide an insightful product-level analysis of NAFTA, finding that ROOs induce a relocation of sourcing from outside to inside the bloc.[4] That type of result is often taken as evidence that ROOs are distortionary. Yet here we show that greater within-FTA can be welfare-enhancing. Thus, an implication of our analysis is that market structure and firm organization matter for the interpretation of empirical results about the effects of ROOs.

Other empirical papers document the incomplete use of preferences in FTAs, and often associate that to poorly designed ROOs. Recently, Crivelli et al. (2021) reveal that preferences in the European Union FTAs, although widely used, are still far from being used by all potential beneficiaries. They provide a detailed account of product-specific ROOs that are too strict to be useful by some firms. Our analysis makes clear that partial utilization is generally inevitable, as firms with different levels of productivity will have different incentives to comply. In particular, the vertical chains with low-productivity suppliers in the FTA will choose not to comply. From a social welfare perspective, this tends to be beneficial, because they are the least affected by hold-up problems. Related to this point, Krishna et al. (2021) show how documentation costs to satisfy ROOs prevent the full use of preferences in FTAs. Interestingly, they find that the fixed cost of documentation falls over time at the firm level, suggesting that firms learn how to satisfy ROOs as they accumulate experience exporting a product to an FTA partner. We do not consider fixed costs to use ROOs in our main analysis, but we discuss how they could be incorporated in section 7.

Finally, our modeling approach is related to that of other papers studying trade in intermediates under incomplete contracts (e.g., Antràs and Helpman, 2004; Ornelas and Turner, 2008, 2012), especially those in the context of trade agreements (e.g., Antràs and Staiger, 2012).[5] As indicated above, the paper is methodologically closest to OTB. The main difference is the focus. In OTB we study the welfare effects of an FTA due to within-bloc reduction of input tariffs. Considering the institutional design of FTAs around the world, here we take the next logical step. Given free trade in inputs within the FTA, we analyze

[4]Sytsma (2022) offers a related analysis for the impact of relaxing ROOs in the European Union's Generalized System of Preferences, finding similar results. Bombarda and Gamberoni (2013) show how diagonal cumulation slackens the restrictiveness of ROOs.

[5]For a recent review of models of international trade featuring incomplete contracts, bargaining and specialized components in the context of global value chains, see Antràs (2020).

how free trade in final goods within the bloc, when mediated by ROOs – as in all existing FTAs – affect the desirability of the trading bloc. Thus, this paper helps toward a fuller understanding of the welfare implications of free trade agreements.

2 The Model

We build on OTB's model, with the key departure being the introduction of a ROO requirement to make the trade of final goods tariff-free within an FTA. Except for that, we keep the final good's market as simple as possible, focusing instead on the inputs market. As we will see, the ROO alters sourcing decisions which, in turn, alter investment decisions. This has implications for the incidence of hold-up problems, for the efficiency of sourcing and, therefore, for the welfare consequences of FTAs.

There is a final good x whose production is carried out by final-good producers (F) – or simply *producers*, for brevity – located in the *Home* country. Those firms transform intermediate inputs into good x. Consumption of good x increases consumers' utility at a decreasing rate. Its world price is p_w^x. We consider that $Home$ is small in world markets, and therefore its producers take p_w^x as given. Under free trade, *Home* is an exporter of x. There is also an homogeneous numéraire good y that enters consumers' utility function linearly. Thus, if they purchase any y, extra income will be directed to the consumption of that good. We assume relative prices are such that consumers always purchase some y. Production of one unit of y requires one unit of labor, the market for y is perfectly competitive, and y is traded freely. This sets the wage rate in the economy to unity and effectively shuts down general equilibrium effects.

Home is in a free trade agreement with *Foreign*, which is an importer of x. The implication is that trade between them is free provided that rules of origin, when present, are respected. The FTA's ROO requires that fraction $r \in (0, 1)$ of the inputs used to produce x come from within the FTA. If final-good producers fail to comply with the ROO, they must pay *Foreign*'s MFN specific tariff $\tau > 0$ on final goods sold to *Foreign*. Hence, compliance yields savings of τ times $Home$'s exports of x to $Foreign$.

When sourcing, each producer may purchase generic inputs z available in the world market (ROW) and/or specialized inputs q from a *Home* supplier (S).[6] Generic inputs are priced in the world market at p_w^z, and decisions in the FTA do not affect p_w^z. In our baseline model, there are no tariffs assessed on z. We relax this assumption in section 6.

As in Grossman (1981), we assume that intermediate goods are perfect substitutes. We define units so that one unit of generic input and one of specialized input have the same revenue-generating value for a producer. Under this normalization, F's revenue only depends on the total number of intermediate inputs he purchases, Q, and not its composition. Note that the ROO requires $q \geq rQ$.

Now, to acquire customized inputs, F and S must first specialize their technologies toward each other, constituting a *vertical chain*. This implies that a producer purchases specialized inputs from only one supplier. All producers are identical, whereas each supplier is identified by ω, an heterogeneity parameter that indexes (the inverse of) her productivity. The distribution of suppliers follows a continuous and strictly increasing distribution $G(\omega)$, with an associated density $g(\omega)$, where ω lies on $[0, p_w^z]$.[7]

Once F and S are specialized toward each other, S makes a non-contractible relationship-specific investment that lowers her marginal cost. The investment is observed by both F and S, but is not verifiable in court. The analysis would remain analogous if the producer also made an ex-ante investment.

After investment is sunk, the firms bargain over how much to trade and at what price. The specialized inputs are not traded on an open market, and have no scrap value. Furthermore, the parties cannot use contracts to affect their trading decisions.[8] If bargaining breaks down, S produces the numéraire good and earns zero (ex post) profit, while F purchases only generic inputs. If bargaining is successful, F imports z from ROW and purchases q

[6]It would make no difference for the analysis if specialized suppliers were located in *Home*'s FTA partner, *Foreign*, provided that they could sell q to producers in *Home* without incurring tariffs.

[7]We take the FTA and the structure of matching as given. Our specification is consistent with the "large natural trading partner" case from OTB, where all specialized suppliers are in the same country and changes to FTA characteristics do not affect matches. In section 7.4, we discuss how endogenizing matching would affect our results. See Grossman and Helpman (2021) for a detailed analysis of how discriminatory tariffs affect the structure of global value chains in a setting where matching and bargaining take center stage.

[8]This is the same approach used by Antràs and Staiger (2012), among others analyzing related environments. It can be formally justified if, for example, quality were not verifiable in a court and the supplier could produce either high-quality or low-quality specialized inputs, with low-quality inputs entailing a negligible production cost for the supplier but being useless to the producer.

from S.

The timing of events within each vertical chain is therefore as follows. First, (i) S makes an irreversible relationship-specific investment. Once the investment is sunk, (ii) F and S bargain over price and quantity of q. If bargaining is successful, production and trade of q takes place; otherwise, $q = 0$ and S produces the numéraire good. Subsequently, (iii) F purchases z. Then, (iv) production occurs and final good x is sold, with payments dependent on whether the ROO is satisfied. We solve the game by backward induction, from the perspective of a single vertical chain. As in OTB, we use the term *Y-chain* when referring to the entire supply chain, as distinct from the F-S vertical chain.[9]

2.1 Cost and Production Functions

To produce q customized inputs, which requires only labor, S incurs cost $C(q, i, \omega)$, where i is the level of her relationship-specific investment. Investment reduces both cost and marginal cost of production, while ω has the opposite effects: $C_i < 0, C_{qi} < 0; C_\omega > 0, C_{q\omega} > 0$. The marginal cost curve is positively sloped ($C_{qq} > 0$) and the cost of investment, $I(i)$, is increasing and convex ($I' > 0, I'' > 0$). We assume a linear-quadratic specification, so that third derivatives of functions $C(.)$ and $I(.)$ are nil.

While this set of assumptions is sufficient for some results, to generate closed-form analytical solutions we adopt the following specific functional forms:

$$\begin{aligned} C(q, i, \omega) &= (\omega - bi)q + \tfrac{c}{2}q^2, \\ I(i) &= i^2. \end{aligned} \tag{1}$$

Here, ω is the intercept of the marginal cost curve; c is the slope of the marginal cost curve; and b represents the effectiveness of investment in reducing marginal costs. We assume that $2c > b^2$ to ensure that the choice of investment is always finite.

We adopt the following specification for the production function of x:

$$x = \begin{cases} Q & \text{if } Q \le \bar{Q} \\ \bar{Q} & \text{if } Q > \bar{Q}. \end{cases} \tag{2}$$

[9] We use *Y-chain* because visually the supply chain resembles a Y: it includes two sources of upstream supply and one downstream final-good producer.

That is, F can transform Q into x under constant returns to scale until capacity $\bar{Q}$ is reached; beyond that, additional units of Q are no longer useful. We further assume that $(p_w^x - p_w^z)$ is sufficiently large to ensure that $Q = \bar{Q}$ is always optimal.[10] As will become clear in section 5, this specification makes it possible to isolate the welfare generated in the inputs market. This setup is, obviously, artificial. We know, since Grossman (1981), that ROOs can have either a positive or a negative impact on final good production and exports. We ignore that possibility to focus on a pedagogically more useful setup, which allows us to concentrate on the first-order effects stemming from the market for inputs. Nevertheless, we indicate in section 7 how our analysis and results would change once one allows final-good production to vary as well.

3 The Choice of Inputs Conditional on Investment

After S chooses her investment, F and S bargain over the number and price of the specialized intermediate inputs. We assume the outcome follows Generalized Nash Bargaining and specify the supplier as having bargaining ability $\alpha \in (0, 1)$. The two firms jointly choose the number of specialized inputs q and their price p^s according to

$$\max_{\{q,p^s\}} \left[U_S^T - U_S^0\right]^{\alpha} \left[U_F^T - U_F^0\right]^{1-\alpha},$$

where U_j^m is the gross profit (i.e., profit absent transfers) that firm $j = F, S$ would receive under scenario m. The two possible scenarios are either bargaining and trading ($m = T$) or not reaching a bargain and thus not trading ($m = 0$).

Conditional on inverse productivity ω, investment i, specialized inputs q and total inputs $\bar{Q}$, producer utilities are:

$$U_F^T = \begin{cases} (p_w^x + \tau)\bar{Q} - p_w^z z - p^s q & \text{if } q \geq q_r \\ p_w^x \bar{Q} - p_w^z z - p^s q & \text{else;} \end{cases}$$

$$U_F^0 = \begin{cases} (p_w^x + \tau - p_w^z)\bar{Q} & \text{if } q \geq q_r \\ (p_w^x - p_w^z)\bar{Q} & \text{else,} \end{cases}$$

[10]See section 7 and Appendix B for discussion of the case where firms might choose total inputs $Q < \bar{Q}$.

where $q_r \equiv r\bar{Q}$. Note that the producer's utility depends on whether the ROO is satisfied. If it is, then he obtains an extra surplus of $\tau\bar{Q}$. If the bargain fails, $q = 0$, $z = \bar{Q}$, and the $q \geq q_r$ constraint is satisfied only if $r = 0$.

For supplier utilities, we have

$$\begin{aligned} U_S^T &= p^s q - C(q, i, \omega); \\ U_S^0 &= 0. \end{aligned}$$

Observe that, while the supplier's utility under a bargain, U_S^T, does not depend directly on whether the ROO is satisfied, the supplier shares part of the additional producer utility through the input price p^s.

For $r > 0$, the bargaining surplus $\Sigma \equiv \left[U_F^T - U_F^0\right] + \left[U_S^T - U_S^0\right]$ is

$$\Sigma = \begin{cases} \Sigma^{RC} \equiv \tau\bar{Q} + p_w^z q - C(q, i, \omega) & \text{if } q \geq q_r \\ \Sigma^{NC} \equiv p_w^z q - C(q, i, \omega) & \text{else.} \end{cases}$$

The RC superscript denotes rule-of-origin compliance, while the NC superscript denotes non-compliance.[11] The bargaining surplus corresponds to the savings from producing q units of inputs within the vertical chain, rather than paying p_w^z for each generic alternative. Furthermore, if the ROO is satisfied, it includes the additional payoff $\tau\bar{Q}$. In an efficient bargain, the vertical chain solves:

$$\max_q \ \Sigma.$$

We restrict the analysis to the case where F purchases both generic and specialized inputs in equilibrium.[12] A sufficient condition for that is $\frac{2p_w^z}{2c-b^2} < \bar{Q}$. In that case, the marginal cost of the most productive supplier ($\omega = 0$) is high enough so that F wants to purchase some generic inputs even when matched with that best supplier, and even when α is near one. Focusing on the dual sourcing case simplifies the analysis significantly, but the important requisite is that the producer must have the option of buying generic inputs

[11] In the special case of $r = 0$, the bargaining surplus always equals Σ^{RC} because compliance is assured regardless of bargaining.

[12] This is in line with the findings of Boehm and Oberfield (2020), who document that mixing customized and standardized inputs is a common practice for Indian manufacturing plants.

when negotiating with his specialized supplier, because that establishes the threat point in the bargaining process.

Now, ignoring the $q \geq q_r$ constraint on Σ^{RC} for a moment, observe that the same choice of inputs that maximizes Σ^{RC} also maximizes Σ^{NC}. This choice, denoted by q_0, equalizes the marginal cost of generic and specialized inputs:

$$p_w^z \equiv C_q(q_0, i, \omega).$$

Using the implicit function theorem, it follows from the properties of $C(.)$ that q_0 is increasing in investment.

Reimposing the constraint, if q_0 complies with the ROO ($q_0 \geq q_r$), then it optimizes Σ and yields bargaining surplus $\Sigma^{RC}(q_0, i, \omega)$. This holds for a sufficiently high investment,

$$i \geq i_{UC}(\omega), \tag{3}$$

where the *unconstrained-compliance investment threshold* $i_{UC}(\omega)$ solves $q_0(i_{UC}(\omega), \omega) \equiv q_r$. This threshold is increasing in ω.[13] With lower productivity, investment must be higher for the vertical chain to comply with the ROO in an unconstrained way.

If $i < i_{UC}(\omega)$, input level q_0 does not comply with the ROO and would yield bargaining surplus $\Sigma^{NC}(q_0, i, \omega)$. The vertical chain then has an additional consideration. By choosing $q \geq q_r$, it can earn the extra surplus $\tau\bar{Q}$ while sacrificing efficiency by producing some specialized inputs at a marginal cost higher than p_w^z. Since Σ^{RC} is strictly decreasing in q as q rises above q_0, the best choice satisfying the constraint is $q = q_r$, which yields bargaining surplus $\Sigma^{RC}(q_r, i, \omega)$. This is then compared to the optimal bargaining surplus under non-compliance, $\Sigma^{NC}(q_0, i, \omega)$. If $\Sigma^{NC}(q_0, i, \omega) > \Sigma^{RC}(q_r, i, \omega)$, non-compliance is optimal. This holds for a sufficiently low investment,

$$i < i_{NC}(\omega, \tau), \tag{4}$$

[13] Because it follows from the properties of $C(.)$ that q_0 is decreasing in ω, we have that $\frac{di_{UC}}{d\omega} = -\frac{\partial q_0/\partial\omega}{\partial q_0/\partial i} > 0$. Under functional forms (1), $i_{UC} = \frac{\omega - p_w^z + cq_r}{b}$.

where the *non-compliance investment threshold* $i_{NC}(\omega,\tau)$ solves $\Sigma^{RC}(q_r, i_{NC}, \omega) = \Sigma^{NC}(q_0, i_{NC}, \omega)$. Intuitively, chains choose not to comply when investment is so low that producing q_r would require pushing marginal cost to an excessively high level. For $i \in [i_{NC}(\omega,\tau), i_{UC}(\omega))$, however, that distortion is worth incurring and $q = q_r$ is optimal. The $i_{NC}(\omega,\tau)$ threshold is also increasing in ω.[14]

The following lemma summarizes optimal sourcing for given investment levels. All proofs are in the Appendix.

Lemma 1 *Conditional on ω, there exist investment thresholds $i_{NC}(\omega,\tau)$ and $i_{UC}(\omega)$, with $i_{NC}(\omega,\tau) \leq i_{UC}(\omega)$, such that the equilibrium level of inputs q_i satisfies the following:*

$$q_i = \begin{cases} q_0 = \frac{p_w^z - \omega + bi}{c} & \text{if } i < i_{NC}(\omega,\tau) \\ q_r = r\bar{Q} & \text{if } i \in [i_{NC}(\omega,\tau), i_{UC}(\omega)] \\ q_0 = \frac{p_w^z - \omega + bi}{c} & \text{if } i > i_{UC}(\omega). \end{cases}$$

In the subsequent analysis, the following definitions are useful:

Definitions (ROO Effectiveness)

1. *A rule of origin that yields equilibrium output choice $q_i = q_r \geq q_0$ is* **binding**.

2. *A rule of origin that yields equilibrium output choice $q_i = q_0$ is* **innocuous**.

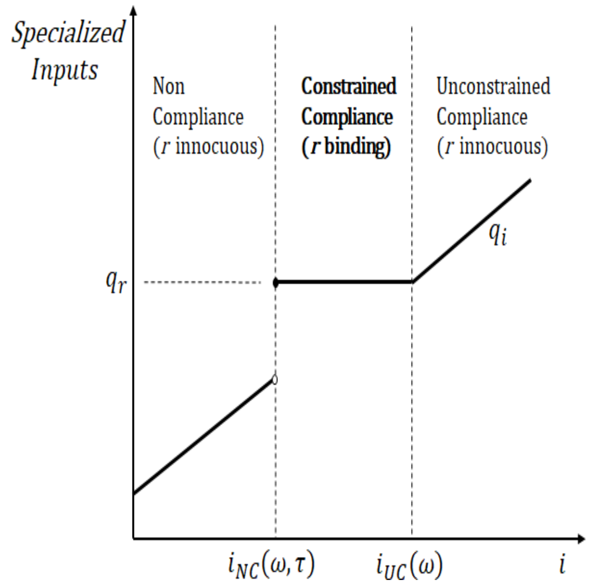
According to these definitions, a ROO is binding for $i \in [i_{NC}(\omega,\tau), i_{UC}(\omega)]$ and is innocuous for other levels of i. Figure 1 illustrates, for a fixed ω, these definitions and the relationship between investment and the vertical chain's choice of inputs. Note that an increase in the final-goods tariff τ slacks the non-compliance investment threshold $i_{NC}(\omega,\tau)$,[15] but has no effect on the unconstrained-compliance investment threshold $i_{UC}(\omega)$. With no final-goods tariff, the investment thresholds coincide, $i_{NC}(\omega, 0) = i_{UC}(\omega)$, since there would be no reason to deviate from q_0 to achieve compliance; any ROO would be innocuous.

The finding that the ROO typically binds for some but not all producers echoes early

[14] We have that $\frac{di_{NC}}{d\omega} = \frac{C_\omega(q_0) - C_\omega(q_r)}{C_i(q_r) - C_i(q_0)} > 0$, where the positive sign follows because $C_{qi} < 0, C_{q\omega} > 0$, and $q_r > q_0$ at this point. Under functional forms (1), $i_{NC} = i_{UC} - \frac{\sqrt{2c\tau\bar{Q}}}{b}$.

[15] This follows from $\frac{di_{NC}}{d\tau} = \frac{\bar{Q}}{C_i(q_r) - C_i(q_0)} < 0$.

Figure 1: *Optimal Specialized Inputs Conditional on Investment*



Note: Conditional on a given ω, this diagram illustrates the relationship between investment i and the choice of specialized inputs q_i that obtains under Generalized Nash Bargaining.

findings from the literature (Grossman, 1981; Krishna, 2006). However, as we will see, our model yields novel implications for welfare. Before that, we study the investment decision.

4 The Choice of Investment

Consider the supplier's choice of investment. She solves

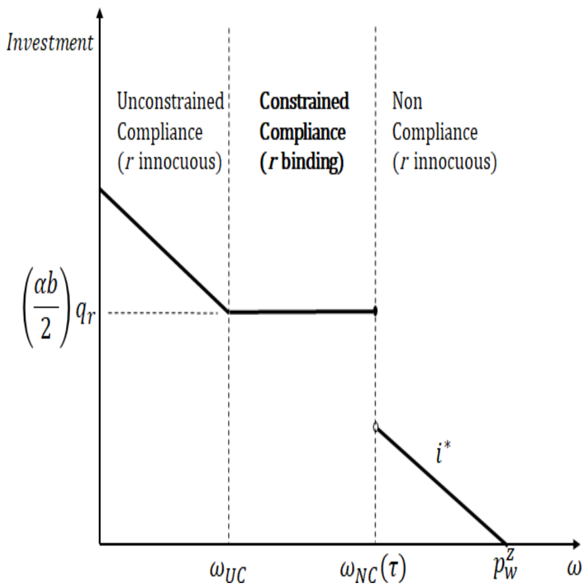
$$max_i \quad U_S(i) \equiv \alpha\Sigma(i,\omega) - I(i), \tag{5}$$

where

$$\Sigma(i,\omega) = \begin{cases} \Sigma^{NC}(q_0,i,\omega) & \text{if} \quad i < i_{NC}(\omega,\tau) \\ \Sigma^{RC}(q_r,i,\omega) & \text{if} \quad i \in [i_{NC}(\omega,\tau), i_{UC}(\omega)] \\ \Sigma^{RC}(q_0,i,\omega) & \text{if} \quad i > i_{UC}(\omega) \end{cases}$$

incorporates ROO-effectiveness. Just as the choice of inputs conditional on investment is tripartite, the investment decision as a function of inverse productivity ω is also tripartite.

Figure 2: *Equilibrium Investment Conditional on* ω



Note: This diagram illustrates the relationship between inverse productivity ω and equilibrium investment i^*.

Proposition 1 *Equilibrium investment* i^* *is defined implicitly by* $-\alpha C_i(q_{i^*}, i^*, \omega) = I'(i^*)$. *There exist thresholds* ω_{UC} *and* $\omega_{NC}(\tau)$, *with* $0 \leq \omega_{UC} \leq \omega_{NC}(\tau) \leq p_w^z$, *such that* i^* *satisfies:*

$$i^* = \begin{cases} i_0^* = \frac{\alpha b(p_w^z - \omega)}{2c - \alpha b^2} & \text{if } \omega < \omega_{UC} \quad \text{High Productivity;} \\ i_r^* = \frac{\alpha b q_r}{2} & \text{if } \omega \in [\omega_{UC}, \omega_{NC}(\tau)] \quad \text{Medium Productivity;} \\ i_0^* = \frac{\alpha b(p_w^z - \omega)}{2c - \alpha b^2} & \text{if } \omega > \omega_{NC}(\tau) \quad \text{Low Productivity.} \end{cases}$$

Figure 2 highlights the pattern of equilibrium investment. For a given r, suppliers can be grouped into high, low and medium productivity categories, partitioned by the cutoffs ω_{UC} and $\omega_{NC}(\tau)$. For high-productivity suppliers, the ROO is innocuous because they comply unconstrained. For low-productivity suppliers, the ROO is similarly innocuous, but because they fail to comply. The ROO is binding only for medium-productivity suppliers. By complying, they produce more inputs than they otherwise would; this compels them to also invest more than they otherwise would.

Intuitively, the supplier's investment always equalizes its marginal cost (I') with fraction α of its marginal benefit $-C_i$. If the supplier expects to produce extra inputs due to constrained compliance with the ROO, then the marginal benefit of investment is higher and she invests more. The $I'(i^*) = -\alpha C_i(q_{i^*}, i^*, \omega)$ condition also reveals a fundamental hold-up problem in our setting; whenever $\alpha < 1$, the supplier under-invests in the absence of policies.

Under our functional forms, the choice of investment satisfies

$$i^* = \frac{\alpha b}{2} q_{i^*}. \tag{6}$$

Substituting, we can then write equilibrium inputs as

$$q_{i^*} = \begin{cases} q_0^* = \frac{2(p_w^z - \omega)}{2c - \alpha b^2} & \text{if } \omega < \omega_{UC} \\ q_r^* = r\bar{Q} & \text{if } i \in [\omega_{UC}, \omega_{NC}(\tau)] \\ q_0^* = \frac{2(p_w^z - \omega)}{2c - \alpha b^2} & \text{if } \omega > \omega_{NC}(\tau). \end{cases}$$

When the thresholds ω_{UC} and $\omega_{NC}(\tau)$ are interior (as in Figure 2),[16] they satisfy

$$\begin{aligned} \omega_{UC} \quad &= p_w^z - q_r \left(\tfrac{2c - \alpha b^2}{2}\right) \text{ and} \\ \omega_{NC}(\tau) \quad &= p_w^z - q_r \left(\tfrac{2c - \alpha b^2}{2}\right) + \sqrt{(2c - \alpha b^2)\tau \bar{Q}}. \end{aligned}$$

An increase in r shifts the range of affected suppliers to the left, while an increase in τ shifts $\omega_{NC}(\tau)$ to the right and widens the range.

In the interior case,

$$\omega_{NC}(\tau) - \omega_{UC} = \sqrt{(2c - \alpha b^2)\tau \bar{Q}},$$

so that the width of the range of supplier productivity affected by the ROO is independent of the stringency of the rule, r. Unsurprisingly, the width is increasing in $\tau\bar{Q}$, the bargaining-surplus bonus from compliance. Moreover, it is decreasing in both supplier bargaining ability, α, and in the effectiveness of investment, b. When investment sharply reduces marginal cost (b is high) and the investing party is very responsive (α is high), the range narrows. Intuitively, in that case investment is a "key decision" for suppliers, and therefore most of them are unwilling to distort it to reap the gain from compliance. Conversely, for low α and low b, the

[16]The interior case holds for $r \in \left(2\sqrt{\frac{\tau}{\bar{Q}(2c - \alpha b^2)}}, \frac{2p_w^z}{\bar{Q}(2c - \alpha b^2)}\right)$.

role of investment is diminished for suppliers, hence most of them become willing to distort it to earn $\tau\bar{Q}$.

If r is sufficiently high, then $\omega_{UC} = 0$ and the ROO binds even for the highest-productivity suppliers. If r is sufficiently low, then $\omega_{NC}(\tau) = p_w^z$ and the ROO binds even for the lowest-productivity suppliers. A high τ has a similar effect, pushing $\omega_{NC}(\tau)$ up. Thus, we could have situations where r and τ are high enough so that the ROO binds for all suppliers because the gain from compliance is sizable (τ is very high) and compliance requires a great share of within-FTA inputs (r is very high). We can think of this as the limiting case when the tripartite equilibrium described in Proposition 1 collapses to one where the ROO is binding for every vertical chain.[17]

5 Welfare

We now consider the welfare effects of a rule of origin r. The welfare generated by a single Y-chain has potentially several components: (1) F's profit; (2) S's profit; (3) consumer surplus (CS) from the final good in *Home*; (4) CS from the final good in *Foreign*; (5) tariff revenue (TR) from imports of x in *Foreign*; (6) TR from imports of z in *Home* (in section 6). Thus, there is welfare due to actions both in the inputs and the final-good markets.

The literature on ROOs has concentrated on how they affect the inputs market. As the discussion in the introduction highlights, ROOs are often associated with welfare-reducing trade diversion in the inputs market, because they tend to induce excessive sourcing within the bloc. We follow the literature by concentrating the analysis on the inputs market.[18] To do so, we shut down welfare effects in the final-goods market.

Specifically, we consider a variation of Grossman and Helpman's (1995) "enhanced protection" case. In that scenario, final demand in *Foreign* is large enough relative to supply in *Home*, so that *Home* can sell all it produces in *Foreign* without affecting prices there. Thus, in *Foreign* the only welfare change is lost tariff revenue on everything it imports from

[17] The other limiting case is when $r = 0$, in which case $\omega_{UC} = p_w^z = \omega_{NC}(\tau)$ and every Y-chain (trivially) complies unconstrained to the rule.

[18] Clearly, this is not without loss of generality. For example, a ROO could generate trade diversion in inputs but reduce trade diversion in final goods.

Home under the agreement. Each producer in *Home* will sell all of his production in *Foreign*. *Home* will import everything it consumes, but since the price does not change (it is given by the world price p_w^x before and after the FTA), consumers are indifferent. The implication is that the welfare components (3) and (4) above are unaffected by the FTA and its ROOs.

In Grossman and Helpman's (1995) model, where the goal is to study the final-goods market, under enhanced protection the net welfare effect is negative because the gain for *Home* producers is less than the loss of tariff revenue in *Foreign*. This happens because, as *Home* producers expand production, they work up in their marginal cost curves. The difference, negative, is the cost of trade diversion for the countries. Here, this does not happen because of production function (2). Under that specification, in the final-good market we have *trade creation = trade diversion = 0*. That is, there is no net gain or loss from trade in the final good x due to the FTA, only a transfer between the two countries. Specifically, *Home* benefits from the free internal trade in final goods (under ROO compliance) because each of its final-good producers earns $\tau\bar{Q}$ with the higher price earned by exporting $\bar{Q}$ to *Foreign*. Meanwhile, *Foreign* loses exactly $\tau\bar{Q}$ in tariff revenue in each of these transactions under the FTA. The net effect for the bloc is therefore nil with respect to the trade of the final good.

Thus, for each Y-chain, the welfare component (5) above is fully offset by the gain in the final-good market of *Home*'s producers, which is part of the welfare component (1). That is, from the bloc's viewpoint, *the FTA is welfare-neutral with respect to the trade of good x*. It follows that the welfare changes due to an FTA between *Home* and *Foreign* stem *only* from changes in S's profit and in F's profit, net of the direct benefit from ROO compliance.

Such joint welfare due to a single Y-chain can be written as the sum of producer utility (net of the private gain from ROO compliance) and supplier utility, if we ignore terms that do not change with the FTA. Specifically:

$$\hat{\Psi}(q,i,\omega) = [p_w^x\bar{Q} - p_w^z(\bar{Q}-q) - p^s q] + [p^s q - C(q,i,\omega) - I(i)].$$

Subtracting the constant $(p_w^x - p_w^z)\bar{Q}$ and rearranging, we obtain

$$\Psi(q,i,\omega) = p_w^z q - C(q,i,\omega) - I(i). \tag{7}$$

Thus, ignoring constant terms, the contribution of a single Y-chain to aggregate welfare corresponds to the savings due to the production of q units of inputs by the specialized supplier, rather than importing those units of generic inputs, net of investment costs. Note that Ψ does not depend directly on the rule of origin r. However, equilibrium welfare does depend on r through its effects on equilibrium investment and input choice.

We denote by $\Delta\widetilde{\Psi}(r, \omega) \equiv \Psi\left(q_{i^*}, i^*, \omega\right) - \Psi\left(q_0^*, i_0^*, \omega\right)$ the welfare effect of moving from no rule of origin to rule of origin r for a single chain involving a supplier with parameter ω – the "Y-chain-level welfare effect." The aggregate welfare effect integrates this term over all levels of productivity:

$$\Delta W(r) = \int_0^{p_w^z} \Delta\widetilde{\Psi}(r, \omega) g(\omega) d\omega. \tag{8}$$

In subsection 5.1, we focus on the Y-chain-level welfare effect for a given ω. This analysis identifies complete welfare effects for the case of a degenerate $g(\omega)$ distribution – that is, absent firm heterogeneity. Comparative statics analysis yields insights that are useful for the aggregate welfare analysis. We turn to aggregate welfare effects with non-degenerate distributions of suppliers in subsection 5.2. Finally, to highlight the role of the hold-up problem in our results, in subsection 5.3 we consider the special cases where investment is useless ($b = 0$) and where the hold-up problem is either unsolvable ($\alpha = 0$) or nonexistent ($\alpha = 1$).

5.1 Homogeneous Suppliers

Fix $\omega \in [0, p_w^z)$ and focus on the investment choice. Conditional on q_i satisfying the privately optimal sourcing condition, $p_w^z = C_q(q_i, i, \omega)$, the *first-best* investment that maximizes (7) satisfies

$$-C_i(i^{fb}) = I'(i^{fb}).$$

That is, i^{fb} equalizes the marginal cost of investment to its marginal return. However, first-best welfare is not achievable generally due to the hold-up problem. As Proposition 1 shows, equilibrium investment is inefficiently low because the supplier captures only share α of the returns from the investment but pays all investment costs.[19]

Following OTB, we define the extent of the hold-up problem as

$$HUP_0 \equiv i^{fb} - i_0^*.$$

The following result holds here and is likely to apply more generally:

Remark 1 *The fundamental hold-up problem is more severe for higher-productivity suppliers.*

The inefficiency in the supplier's investment choice increases with the share $(1-\alpha)$ of the returns to investment.[20] In turn, those returns are increasing in supplier productivity, because higher-productivity suppliers produce more, and therefore investment lowers the cost of more units when productivity is higher.[21]

Now, because a ROO can affect investment decisions, it can also affect the severity of the hold-up problem. This is a potential source of gain. However, there are three potential difficulties. First, the ROO may not be binding. Second, when the ROO is binding, it will distort the choice of inputs. Third, the resulting investment may exceed the first best. For future use, it is useful to define the (potential) excess of investment under a binding ROO as

$$EXC_r \equiv i_r^* - i^{fb}.$$

The following lemma describes the Y-chain-level welfare impact of a rule of origin.

Lemma 2 *For any $\omega \in [0, p_w^z)$, there exist thresholds $r_{UC}(\omega)$ and $r_{NC}(\omega, \tau)$, with $0 < r_{UC}(\omega) \leq r_{NC}(\omega, \tau)$, such that the rule of origin is binding if and only if $r \in [r_{UC}(\omega), r_{NC}(\omega, \tau)]$.*

[19]For the above condition, we further assume that $\frac{2p_w^z}{2c-b^2} < \bar{Q}$. If $\frac{2p_w^z}{2c-\alpha b^2} < \bar{Q} < \frac{2p_w^z}{2c-b^2}$, then first-best investment would be a corner solution.

[20]Specifically, $HUP_0 = \frac{2(1-\alpha)bc(p_w^z-\omega)}{(2c-b^2)(2c-\alpha b^2)}$.

[21]The characteristic that the returns to the non-contractible choice increase with firm productivity is not specific to our model. It is also present, for example, in the model of Antràs and Helpman (2004).

The Y-chain-level welfare effect of the rule of origin is

$$\Delta\widetilde{\Psi}(r,\omega) = \begin{cases} 0 & \text{if } r < r_{UC}(\omega) \\ \Delta\Psi(r,\omega) \equiv (p_w^z - \omega)(q_r^* - q_0^*) - \frac{\left(2c - 2\alpha b^2 + \alpha^2 b^2\right)\left(q_r^{*2} - q_0^{*2}\right)}{4} & \text{if } r \in [r_{UC}(\omega), r_{NC}(\omega,\tau)] \\ 0 & \text{if } r > r_{NC}(\omega,\tau). \end{cases}$$

The welfare effect $\Delta\widetilde{\Psi}(r,\omega)$ incorporates two components: (1) The *potential welfare effect* $\Delta\Psi(r,\omega)$; and (2) threshold values of r that determine whether the ROO binds. The potential welfare effect is the difference in welfare for the case of a binding ROO versus the case of no ROO. If the ROO does bind, then the potential welfare effect obtains as the welfare effect, $\Delta\widetilde{\Psi}(r,\omega) = \Delta\Psi(r,\omega)$. If the ROO does not bind, then the welfare effect is nil.

The threshold values $r_{UC}(\omega)$ and $r_{NC}(\omega,\tau)$ are inversions of the ω_{UC} and $\omega_{NC}(\tau)$ thresholds, respectively. If $r \in [r_{UC}(\omega), r_{NC}(\omega,\tau)]$, then the ROO binds ($q_{i^*} = q_r^*$). If τ is small, the difference between $r_{UC}(\omega)$ and $r_{NC}(\omega,\tau)$ is also small: if the private gain from compliance is modest, most levels of r will be unable to affect investment and sourcing decisions. In the limit when $\tau = 0$, $r_{UC}(\omega) = r_{NC}(\omega,0)$ and no rule of origin alters investments or welfare.[22] Observe also that both $r_{UC}(\omega)$ and $r_{NC}(\omega,\tau)$ decrease with ω. This happens because higher-productivity suppliers always invest more and sell more inputs; thus, it takes a higher r to match the fraction of inputs they produce absent a ROO.

Relying on Lemma 2, the following proposition characterizes the welfare-optimizing r for the case of homogeneous suppliers.

Proposition 2 *Let the distribution of inverse productivity be degenerate and centered on any $\omega \in [0, p_w^z)$. There exist $\underline{\tau}(\omega) \geq 0$ and $\hat{r}(\omega) \in [0,1]$ such that:*

(i) If $\tau \geq \underline{\tau}(\omega)$, the welfare effect $\Delta\widetilde{\Psi}(r,\omega)$ is maximized with rule of origin $r^* = \hat{r}(\omega)$;

(ii) If $\tau < \underline{\tau}(\omega)$, the welfare effect $\Delta\widetilde{\Psi}(r,\omega)$ is maximized with rule of origin $r^* = r_{NC}(\omega,\tau)$;

(iii) $r^* = \min\{\hat{r}(\omega), r_{NC}(\omega,\tau)\}$.

Maximizing welfare requires maximizing the potential welfare effect subject to the constraint that $r \in [r_{UC}(\omega), r_{NC}(\omega,\tau)]$. The potential welfare effect, $\Delta\Psi(\omega, r)$, is an inverted-U

[22] Given our definitions, the ROO technically is binding ($q_r = q_0$) if $r = r_{UC}(\omega) = r_{NC}(\omega,0)$, but does not alter the input or investment choices.

function of r with unique maximizer

$$\hat{r}(\omega) = \frac{2(p_w^z - \omega)}{\bar{Q}(2c - 2\alpha b^2 + \alpha^2 b^2)}.$$

It is obvious that $\hat{r}(\omega) > r_{UC}(\omega)$.[23] Thus, if $\hat{r}(\omega) \leq r_{NC}(\omega, \tau)$, then $\hat{r}(\omega)$ binds and is optimal. The threshold $\underline{\tau}$ implicitly solves $\hat{r}(\omega) = r_{NC}(\omega, \underline{\tau})$; hence $\hat{r}(\omega) \leq r_{NC}(\omega, \tau)$ is optimal if and only if $\tau \geq \underline{\tau}$. If $\tau < \underline{\tau}$, then $r_{NC}(\omega, \tau) < \hat{r}(\omega)$ is the r closest to $\hat{r}$ such that the ROO binds, so it is optimal.

Regardless of τ, it follows from the proofs of Lemma 2 and Proposition 2 that both $r_{NC}(\omega, \tau)$ and $\hat{r}(\omega)$ decrease with ω. Therefore, the optimal ROO in the homogeneous case is increasing in productivity.

Corollary 1 *When suppliers are homogeneous, the optimal rule of origin is increasing in productivity.*

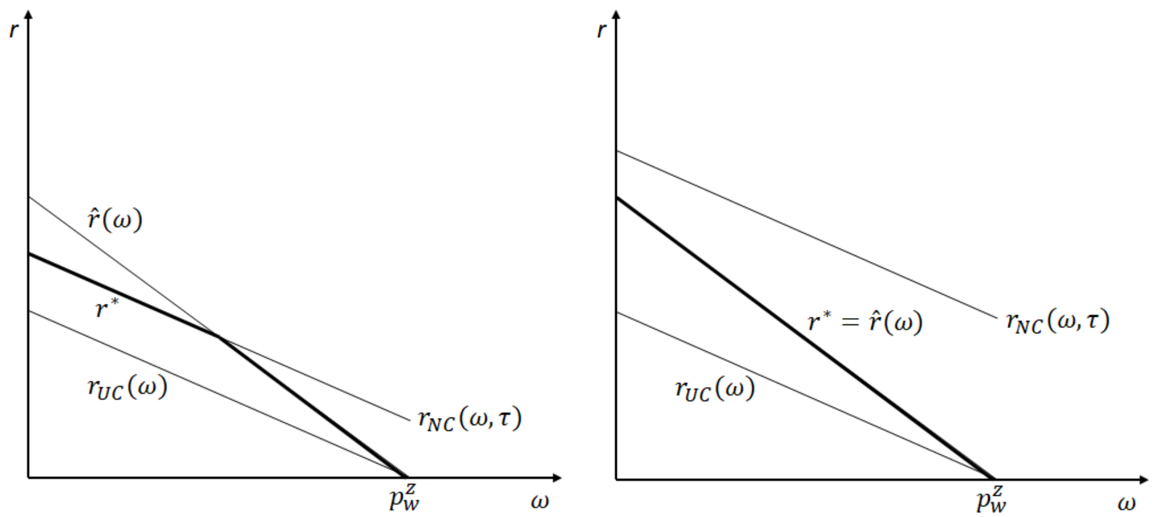
Figures 3a-b highlight key characteristics of the optimal $r^*(\omega)$. In Figure 3a, $r^*(\omega)$ equals $r_{NC}(\omega, \tau)$ if ω is low and $\hat{r}(\omega)$ if ω is high. In Figure 3b, τ is high enough so that $\hat{r}(\omega)$ is always binding, so $r^*(\omega) = \hat{r}(\omega)$ everywhere. In both figures, $r^*(\omega)$ is decreasing in ω.

Observe that the optimal ROO is complementary to τ. Generally, welfare under the optimal r^* increases with τ and strictly increases with τ on $[0, \underline{\tau}]$, i.e., whenever $\hat{r}(\omega) > r_{NC}(\omega, \tau)$. Intuitively, a higher τ provides greater scope of action for a ROO, which allows for higher welfare if r is chosen optimally. Thus, if the FTA-importing country has high tariffs on the final good, the ROO has the potential to be more useful to mitigate hold-up inefficiencies and increase welfare.

The potential welfare effect, $\Delta\Psi(r, \omega)$, is an inverted-U function of r because it reflects two opposing effects of the ROO on welfare. On the one hand, it induces more investment, helping to alleviate the hold-up problem (although it can also induce too much investment). On the other hand, it yields a socially excessive number of specialized inputs because, for any i, the supplier's marginal cost of the final unit exceeds the unit price of generic in-

[23]Intuitively, this is due to the hold-up problem; there is always some margin to use ROOs to overcome this. Mathematically, it holds because $2c - 2\alpha b^2 + \alpha^2 b^2 < 2c - \alpha b^2$.

Figure 3: *Optimal ROO, Homogeneous Suppliers, Varying* τ



(a) *Final Goods Tariff* τ (b) *Final Goods Tariff* $\tau' > \tau$

Note: These diagrams highlight the optimal ROO for the case where all suppliers have the same level of inverse productivity. The x-axis is the level of inverse productivity ω that all suppliers have, while the bold-face function is the optimal ROO $r^*(\omega)$ for that set of suppliers. In the left panel, τ is relatively small, which constrains $r^*(\omega)$ for low ω. In the right panel, τ is high enough so that $r^*(\omega)$ is unconstrained.

puts, $C_q(q_r, i) > C_q(q_0, i) = p_w^z$. Analogously to OTB, we call the former effect *relationship strengthening* and the latter effect *sourcing diversion.*[24]

More precisely, we define relationship strengthening as

$$\Delta\Psi_{RS} = p_w^z (q_1 - q_0^*) + [C(q_0^*, i_0^*, \omega) - C(q_1, i_r^*, \omega)] - [I(i_r^*) - I(i_0^*)],$$

where q_1 denotes the level of specialized inputs that equalizes the marginal costs of the two types of inputs when $i = i_r^*$. That is, q_1 solves $C_q(q_1, i_r^*) = p_w^z$. Using our functional forms and manipulating, this expression can be rewritten as

$$\Delta\Psi_{RS} = \left(\frac{2c - b^2}{2c}\right)(i_r^* - i_0^*)(HUP_0 - EXC_r). \tag{9}$$

[24] OTB define those forces to describe the effects of tariff preferences, without any ROO. Here they are defined to describe the effects of a ROO, without any tariff preference (until section 6, when we consider both). Indeed, those two terms can be used to describe the effects of any policy that affects firm choice of investment and its subsequent impact on production. However, the precise form of the two effects vary with the specifics of the policy in analysis.

This effect is positive if and only if the ROO moves investment *closer* to the first-best level, relative to the situation without ROOs. Differentiating equation (9) with respect to r, while noticing that r affects it only via i_r^* and EXC_r (through its effect on i_r^*), one finds that

$$\frac{\partial \Delta \Psi_{RS}}{\partial r} = \left(\frac{2c - b^2}{2c} \right) \alpha b \bar{Q} \left(i^{fb} - i_r^* \right),$$

which is positive if and only if $i_r^* < i^{fb}$. That is, increasing r improves welfare through the relationship-strengthening effect provided that it does not induce investment in excess of i^{fb}. Since $\partial \Delta \Psi_{RS} / \partial r^2 < 0$, it follows that $\Delta \Psi_{RS}$ is concave in r.

In turn, we define sourcing diversion as

$$\Delta \Psi_{SD} = C(q_1, i_r^*, \omega) - C(q_r, i_r^*, \omega) + p_w^z (q_r^* - q_1).$$

With our functional forms, this expression can be rewritten as

$$\Delta \Psi_{SD} = -\frac{c}{2} \left(q_r^* - q_1 \right)^2, \tag{10}$$

which is negative because $q_r^* > q_1$. This negative effect increases in magnitude as ω rises. The reason is that, with lower productivity, i_0^* is lower, so ROO compliance generates a bigger increase in investment above i_0^*, which yields a greater sourcing distortion, $q_r^* - q_1^*$. Differentiating equation (10) with respect to r, we have that

$$\frac{\partial \Delta \Psi_{SD}}{\partial r} = -c \bar{Q} \left(q_r^* - q_1 \right),$$

so increasing r makes sourcing diversion monotonically worse. Since $\partial^2 \Delta \Psi_{SD} / \partial r^2 < 0$, $\Delta \Psi_{SD}$ is also concave in r, and so is $\Delta \Psi$.

Note that, if $r = r_{UC}(\omega)$, then $i_r^* = i_0^*$ and the ROO does not alter choices. For slightly higher r, i_r^* rises and $\Delta \Psi_{SD}$ grows in magnitude, but the loss due to sourcing diversion is of second order. By contrast, the relationship-strengthening effect is of first order. However, as r increases further, the loss due to sourcing diversion grows larger, while the relationship-strengthening effect starts to decrease eventually. Hence, the net effect is an inverted-U curve, reaching its maximum at $\hat{r}(\omega)$.

Recalling Figures 3a-b, the function $\hat{r}$ decreases in ω. This reflects substitutability in $\Delta\Psi$ between r and ω (or equivalently, complementarity between r and productivity). To understand that, note first that the marginal social gain (due to higher investment) when r increases is greater for high productivity Y-chains, because the hold-up problem is more severe for them (Remark 1). Given a binding ROO,

$$\frac{\partial^2 \Delta\Psi_{RS}}{\partial r \partial \omega} = \left(\frac{2c - b^2}{2c}\right) \alpha b \bar{Q} \frac{\partial i^{fb}}{\partial \omega} < 0.$$

In addition, the marginal social loss (due to sourcing diversion) when r increases is smaller for high productivity Y-chains, because they already use more within-FTA specialized inputs anyway, so the distortion due to raising q up to q_r is smaller. Formally,

$$\frac{\partial^2 \Delta\Psi_{SD}}{\partial r \partial \omega} = \left(\frac{2c - \alpha b^2}{4c}\right) \bar{Q} \frac{\partial q_0}{\partial \omega} < 0.$$

Thus, considering the Y-chain-level welfare effect, it is best to raise r as productivity increases. Every supplier with $\omega < p_w^z$ underinvests. The smaller ω is (that is, the higher the supplier productivity), the more severe is the hold-up problem and the less severe is sourcing diversion. Therefore, the rule of origin should be tighter, inducing a greater increase in investment, the smaller ω is.

5.2 Heterogeneous Suppliers

Now consider the effect of a ROO in the more realistic case where there is a distribution of heterogeneous suppliers. Incorporating ROO-effectiveness, the aggregate welfare effect can be written as

$$\Delta W(r) = \int_{\omega_{UC}}^{\omega_{NC}(\tau)} \Delta\Psi(r, \omega) g(\omega) d\omega.$$

This integrates out over the density of Y-chains with suppliers $\omega \in [\omega_{UC}, \omega_{NC}(\tau)]$, for whom the ROO binds. The following proposition characterizes how the level of r affects welfare.

Proposition 3 *There exists an $r^+ \geq 0$ such that, if $r \geq r^+$, then the welfare effect of the rule of origin, $\Delta W(r)$, is positive for any distribution of suppliers. If in addition $\tau < 4\underline{\tau}(0)$, then* *when* $r < r^+$ *there exists an* $\omega^0 \in (\omega_{UC}, \omega_{NC}(\tau))$ *such that* $\Delta\Psi(r,\omega) > 0$ *for* $\omega \in (\omega_{UC}, \omega^0)$ *and* $\Delta\Psi(r,\omega) < 0$ *for* $\omega \in (\omega^0, \omega_{NC}(\tau))$.

The intuition for this result is easiest to see for the case where ω_{UC} and $\omega_{NC}(\tau)$ are interior. The potential welfare effect $\Delta\Psi(r,\omega)$ is an inverted-U function of ω, with $\Delta\Psi(r,\omega_{UC}) = 0$ and $\Delta\Psi(r,\omega)$ increasing in ω at $\omega = \omega_{UC}$. Thus, if $\Delta\Psi(r,\omega_{NC}(\tau)) \geq 0$, then $\Delta\Psi(r,\omega) \geq 0$ for all ω such that the ROO is binding. That condition is equivalent to $r \geq r^+$. Now, if r is lower, then $\Delta\Psi(r,\omega_{NC}(\tau))$ is negative. Provided that τ is sufficiently low, $\Delta\Psi(r,\omega)$ will be positive for ω relatively close to ω_{UC} but negative for ω relatively close to $\omega_{NC}(\tau)$. The cutoff ω^0 determines what "close" represents.[25]

For any rule of origin that yields interior ω_{UC} and $\omega_{NC}(\tau)$, some Y-chains provide a positive contribution for welfare, and a sufficiently tight rule of origin (weakly) improves welfare for all Y-chains. To see the reasons behind the first claim, note that welfare is unchanged for Y-chain ω_{UC}. But for any rule of origin, no matter how "light," welfare must increase for a Y-chain with ω that is higher but sufficiently close to ω_{UC}. The reason is that the ROO induces a first-order relationship-strengthening effect, which is higher than the second-order loss from the sourcing diversion induced by the ROO.[26] For Y-chains with ω well above ω_{UC}, however, the sourcing-diversion effect may dominate.

Now, for a sufficiently high r, the welfare effect is positive for all $\omega \in [\omega_{UC}, \omega_{NC}(\tau)]$.

[25] If τ is not sufficiently low, then welfare may be negative for all affected ω. The simplest example of this is the limiting case where $\alpha(1-\alpha)b^2 = 0$, in which case the cutoff for τ is zero and $\omega^0 = \omega_{UC}$ when the latter is interior. We analyze this case in detail in subsection 5.3. In the less extreme case where $\alpha(1-\alpha)b^2$ is close to zero, so that $4\underline{\tau}(0)$ is also positive but close to zero, if τ is high then it is possible to find levels of r such that ω_{UC} obtains at the boundary while $\omega_{NC}(\tau)$ is interior *and* $\omega^0 < 0$. In that case $\Delta\Psi(r,\omega_{UC}) < 0$, and indeed $\Delta\Psi(r,\omega) < 0$ for all $\omega \in [\omega_{UC}, \omega_{NC}(\tau)]$. See also footnote 27.

[26] The relationship-strengthening effect can be rewritten as

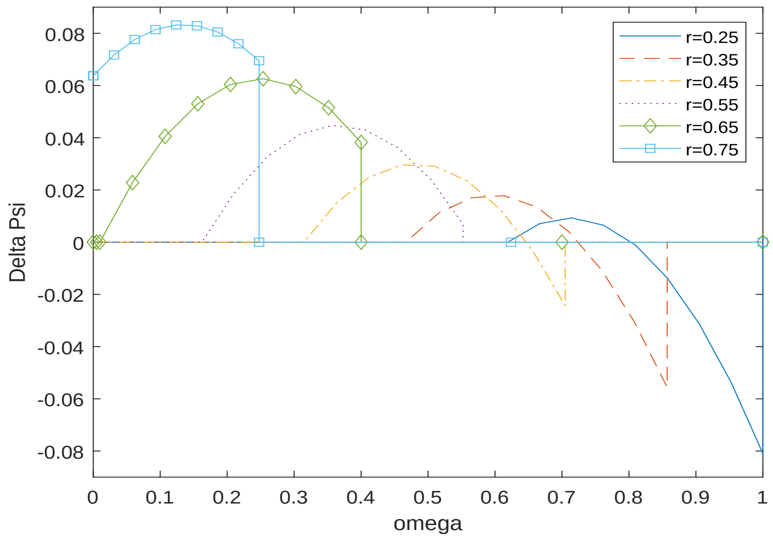
$$\Delta\Psi_{RS} = \frac{\alpha b(2c-b^2)}{2c(2c-\alpha b^2)}\left(\omega - \omega_{UC}\right)\left(HUP_0 - EXC_r\right),$$

while the sourcing diversion effect can be rewritten as

$$\Delta\Psi_{SD} = -\frac{1}{2c}\left(\omega - \omega_{UC}\right)^2.$$

The former is first-order and positive for ω just above ω_{UC}, while the latter is negative but second-order for such ω.

Figure 4: *The Y-Chain-Level Welfare Effect of a ROO*



Note: This diagram illustrates the Y-chain-level welfare effect $\Delta\Psi(r, \omega)$ using an example. We consider $r \in \{0.25, 0.35, 0.45, 0.55.0.65, 0.75\}$. For each r, $\Delta\Psi(r, \omega)$ is shown for all ω. Other parameters are $\alpha = 0.5$, $b = 1.25$, $c = p^z_w = 1$, $\tau = 0.05$ and $\bar{Q} = 2.5$.

Intuitively, a "tight" rule of origin only affects the behavior of high-productivity suppliers (because low-productivity ones will choose to not comply). Since those suppliers' underinvestment is more severe, the rule has a more crucial role in mitigating that problem, relative to the sourcing distortion (for given investment) that it also engenders.

This result implies that, with a continuous density of productivity, a stricter ROO is more likely to generate a positive welfare effect for all affected suppliers. Figure 4 highlights this using an example where we consider different levels of r. The ω_{UC} and $\omega_{NC}(\tau)$ cutoffs are both interior for $r = \{.35, ..., .65\}$, but not for $r = .25$ and $r = .75$, when, respectively, $\omega_{NC}(\tau)$ and ω_{UC} are at a corner. Following Proposition 3, $\Delta\Psi(r, \omega)$ is increasing at $\omega = \omega_{UC}$ for all r and is concave on $\omega \in [\omega_{UC}, \omega_{NC}(\tau)]$.

Using the results in Proposition 3, we can solve to find $r^+ \approx .53$ under the parametrization used in Figure 4. Hence, in the figure, $\Delta\Psi(r, \omega) > 0$ for all $\omega \in [\omega_{UC}, \omega_{NC}(\tau)]$ when $r \geq 0.55$. If r is in this upper region, then the aggregate welfare effect of the ROO is positive for *any* distribution of ω. We also see in the figure that $\Delta\Psi(r, \omega) < 0$ at $\omega = \omega_{NC}(\tau)$ for $r \leq 0.45$.

If r is in this lower region, then the aggregate welfare effect of the ROO may be positive or negative depending upon the distribution of ω.

The ultimate magnitude of $\Delta W(r)$, and how it changes with r, depends on the distribution of ω. Although generally ambiguous, it is useful to sort out the various forces in play. To do so, let us write the general expression of how r affects $\Delta W(r)$. Writing the ω cutoffs as functions of r and suppressing the τ argument in ω_{NC}, the Leibniz formula yields:

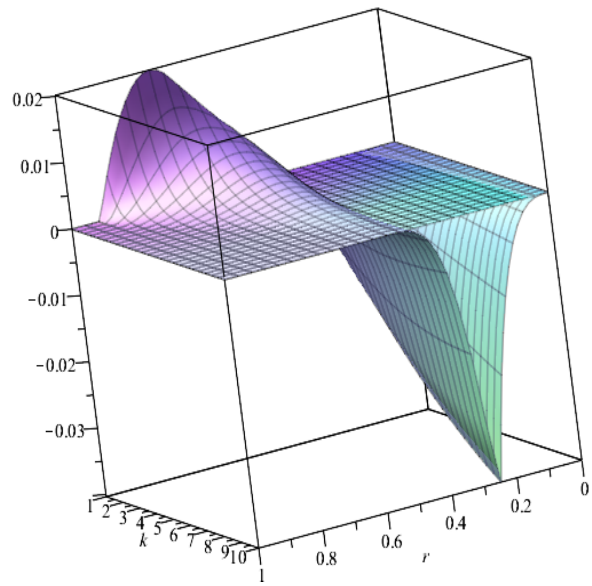
$$\frac{d\Delta W(r)}{dr} = \int_{\omega_{UC}(r)}^{\omega_{NC}(r)} \left(\frac{d\Delta\Psi(r,\omega)}{dr}\right) g(\omega)d\omega - \Delta\Psi(r,\omega_{UC})g(\omega_{UC})\omega'_{UC}(r) + \Delta\Psi(r,\omega_{NC}(r))g(\omega_{NC})\omega'_{NC}(r). \tag{11}$$

There are three sets of effects. The first term captures the way welfare changes for chains that comply in a constrained way both with ROO r and with ROO $r + dr$. The second term captures the way welfare changes for chains that enter into constrained compliance as r increases to $r + dr$. The third term captures the way welfare changes for chains that exit constrained compliance (and enter non-compliance) as r increases to $r + dr$.

Consider the first term. For each $\omega \in (\omega_{UC}, \omega_{NC})$, the chain complies in a constrained way with ROO r and for ROO $r + dr$. As r increases, welfare changes according to $\frac{d\Delta\Psi(r,\omega_{UC})}{dr}$ for a given chain. This change is then multiplied by the density at that point, $g(\omega)$. For each $\omega \in (\omega_{UC}, \omega_{NC})$, these changes are added up, yielding the first term in equation (11). Let us analyze each ω. For ω just above ω_{UC}, at the status-quo ROO, the chain chooses investment i_r^* and inputs q_r^*, but these levels are only slightly above i_0^* and q_0^*, respectively. Thus, investment is too low because of the hold-up problem. As r increases, these chains increase their investments and inputs further, and welfare rises. From Proposition 2, we know that for any ω there is an $\hat{r}(\omega)$ that maximizes $\Delta\Psi$. For a given r, we can invert this expression to find $\hat{\omega}(r)$. For all $\omega < \hat{\omega}(r)$, $\Delta\Psi(r,\omega)$ rises with a higher r. For all $\omega \geq \hat{\omega}(r)$ such that the chain complies both with status-quo ROO r and with the new ROO $r + dr$, the higher ROO exacerbates the over-investment problem. Hence, welfare falls for these chains. The entire welfare effect captured by the first term in equation (11) therefore depends on the distribution. If it has sufficient mass of high-productivity suppliers, welfare rises due to this force, but otherwise it falls.

Consider now the second term. At ω_{UC}, as r increases, the vertical chain moves from

Figure 5: *The Welfare Effect of a ROO with Pareto Density*



Note: This diagram illustrates the welfare effect where $\frac{1}{\omega}$ is distributed according to a Pareto(k) with scale parameter $\frac{1}{p_w^z}$ and shape parameter $k \geq 1$. Other parameters are as in Figure 4

unconstrained compliance into constrained compliance. The welfare effect changes from zero to $\Delta\Psi(r, \omega_{UC})$, but this is also zero because the unconstrained compliance input choice just meets the ROO constraint. Hence, there is no welfare change from this effect.

Finally, consider the third term of equation (11). At ω_{NC}, as r increases, the vertical chain opts out of constrained compliance and into non-compliance. The welfare effect changes to zero, so welfare falls by $\Delta\Psi(r, \omega_{NC})$. This loss/gain is weighted by the density of chains at that point, $g(\omega_{NC})$, and by the pace at which suppliers substitute out of constrained compliance into non-compliance, $\omega_{NC}^{'}(r)$. Naturally, this happens when ω_{NC} is interior. If it is not, then this term vanishes.

Putting all this together, the two non-zero welfare effects are: (1) the changes in welfare from the chains that comply in a constrained way for both r and $r + dr$; and (2) the change in welfare from the ω_{NC} chains opting out of compliance as r increases to $r + dr$, when ω_{NC} is interior.

To visualize the aggregate effects more clearly, let us consider the familiar case where inverse productivity $\frac{1}{\omega}$ is distributed Pareto with scale parameter $\frac{1}{p_w^z}$ and shape parameter $k \geq 1$, so $G(\omega) = (\omega/p_w^z)^k$. In this case, we can solve explicitly for $\Delta W(r)$ as a function of primitives. Using the same parametrization of Figure 4, Figure 5 shows how $\Delta W(r)$ changes with r and k.

When $r = 0$, $\omega_{UC} = \omega_{NC} = p_w^z$, so there is no effect (trivially) because every chain complies unconstrained. As r starts to increase, $0 < \omega_{UC} < p_w^z = w_{NC}$, so the chains with ω smaller but very near p_w^z comply in a constrained way. But these are the chains for which the initial hold-up problem is very mild, so the welfare effect coming from them is surely negative. As r keeps increasing while $0 < \omega_{UC} < \omega_{NC} = p_w^z$, the ROO-induced overinvestment problem of the high-ω chains gets worse and some additional chains with lower-but-still-high ω start to overinvest, although by not as much. Thus, ΔW becomes more negative. Observe that, because $\Delta\Psi(\omega)$ is an inverted-U function and is positive for ω near ω_{UC} (proof of Proposition 2), some chains do yield a positive welfare effect, but for low r this effect is too small to overturn the negative effect due to the least productive chains.

Once r increases enough so that ω_{NC} becomes slightly smaller than p_w^z, and $0 < \omega_{UC} < \omega_{NC} < p_w^z$, then the least productive chains stop complying. This eliminates the most negative contributions to ΔW, which then starts to increase from a negative level. As r keeps rising, the range of chains for which r generates a positive effect increases and the positive relationship-strengthening effect grows, while at the same time other low-productivity chains stop complying and overinvesting. This pushes the welfare effect higher and eventually turns it positive.

Now, as r rises further so that ω_{UC} obtains at the 0 boundary, the mass of high-productivity chains that are affected decreases, and this lowers ΔW. Finally, as r becomes high enough so that $\omega_{UC} = \omega_{NC} = 0$, then no chain complies and $\Delta W = 0$ again.

Importantly, all of these effects must be weighted by the corresponding densities. Specifically, the negative effect from the high-ω chains overinvesting, as well as the gain when they stop complying, is more relevant when the distribution of productivity is shifted toward low-productivity suppliers. Similarly, the positive effect from the low-ω chains investing more is less relevant when the distribution of productivity is shifted toward low-productivity suppliers. Under a Pareto distribution, these happen when k is high.

The broader lessons from the Pareto distribution are general. First, when the ROO is very lenient, there is ample room for negative welfare effects. A low r does not affect the high-productivity chains that should be affected, and does affect the low-productivity ones that should not (from a social standpoint). It is especially harmful when the distribution of productivity is skewed toward low-productivity suppliers. The opposite happens when the ROO is relatively strict, as it induces more investment precisely by the chains that are underinvesting more. It is especially beneficial when the distribution of productivity is not too skewed toward low-productivity suppliers. Put in a different way, when there are plenty of low-productivity suppliers, there is little hope for a welfare-improving ROO, but when there is a sufficiently large share of high-productivity suppliers, a relatively strict ROO is likely to generate the highest possible welfare gain.

5.3 Special Cases

Several prior papers have studied rules of origin in competitive environments. Generally, their view is that rules of origin reduce welfare. This is consistent with our analysis. Intuitively, when firms act as price takers, there are no hold-up problems. Accordingly, there is no scope for welfare-enhancing relationship strengthening and ROOs produce only distortions in the inputs market.

Indeed, welfare also falls for several special limiting cases in our model. Consider first the possibility that either investment is useless ($b = 0$) or the hold-up problem cannot be solved ($\alpha = 0$). Then there is effectively no investment decision, because equilibrium investments are zero and do not change with a rule of origin (recall equation (6)). Hence, the relationship-strengthening effect is zero. Meanwhile, the firms still have some incentive to comply with the ROO, inducing sourcing diversion. Welfare cannot improve, and it strictly decreases whenever the ROO is binding for any Y-chain.

Welfare also falls for sure when the hold-up problem is non-existent, $\alpha = 1$. In this case, investment equals the first-best with no ROO. Hence, $HUP_0 = 0$ and $EXC_r \geq 0$, so the relationship-strengthening effect cannot be positive and is strictly negative if the ROO is binding. Again, welfare cannot improve, and it strictly decreases whenever the ROO is binding.[27]

Proposition 4 *In the limiting cases of ineffective investment ($b = 0$) or extreme supplier bargaining ability ($\alpha = 0$ or $\alpha = 1$), any binding ROO decreases welfare.*

In sum, a ROO can be helpful only if it mitigates the original hold-up problem.

6 Positive Input Tariff

Now consider the case where the input external tariff is strictly positive, $t > 0$. Absent a binding ROO, equilibrium investment and output become

$$i^* = i_t^* = \tfrac{\alpha b(p_w^z + t - \omega)}{2c - \alpha b^2},$$
$$q_i^* = q_t^* = \tfrac{2(p_w^z + t - \omega)}{2c - \alpha b^2}.$$

Note that $i_t^* > i_0^*$ and $q_t^* > q_0^*$.[28] The input tariff increases the privately optimal number of inputs in the F-S bargain; in response, S invests more.

This can have either a positive or a negative effect on welfare. To see this, note first that, for $t = 0$, the level of investment that obtains when $r = \hat{r}(\omega)$ is binding is

$$i^{sb} = \frac{\alpha b(p_w^z - \omega)}{2c - 2\alpha b^2 + \alpha^2 b^2}.$$

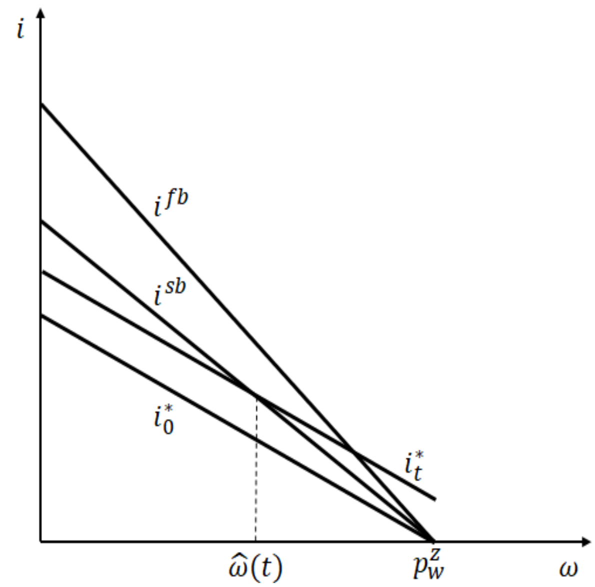
This level of investment is, essentially, a *second-best* level. It is easily seen from the formulas that $i^{fb} > i^{sb} > i_0^*$ and that the differences $(i^{fb} - i^{sb})$ and $(i^{sb} - i_0^*)$ decrease with ω.

We have that $i^{sb} > i_0^*$ because i^{sb} is the level of investment that optimally trades off the gains from relationship strengthening against the losses from sourcing diversion. In turn, $i^{fb} > i^{sb}$ because i^{fb} is the level of investment that solves the hold-up problem assuming efficient sourcing. Notice that the gap between i^{fb} and i_0^* is wider for lower ω. This reflects

[27] Note that this result ties in to the "high τ" case left out of Proposition 3 and discussed in footnote 25. The cutoff value of τ, $4\underline{\tau}(0)$, is proportional to $\alpha(1-\alpha)b^2$. This cutoff is zero for all of the cases considered in Proposition 4, in which case $\omega^0 \leq \omega_{UC}$, so $\Delta\Psi(r,\omega) \leq 0$ for all $\omega \in [\omega_{UC}, \omega_{NC}(\tau)]$. Note that if $r \geq r^+$, $\Delta\Psi(r,\omega) = 0$ for all $\omega \in [\omega_{UC}, \omega_{NC}(\tau)]$, consistent with Proposition 3.

[28] In this section we further assume $\frac{2(p_w^z + t)}{2c - b^2} < \bar{Q}$ to guarantee dual-sourcing in equilibrium.

Figure 6: *Benchmark Investment Levels*



Note: This diagram illustrates the relationship between inverse productivity ω and first-best, second-best and equilibrium investment, with and without an input external tariff.

the fact that the hold-up problem becomes more severe as productivity increases (Remark 1). The two differences decrease with ω because i^{sb} incorporates elements from both i^{fb} and i^*_0.

Figure 6 shows these benchmark investment levels, together with i^*_t. Clearly, investment absent a ROO may now be either below or above i^{sb}. Indeed, $i^*_t \geq i^{sb}$ if and only if $\omega \geq \hat{\omega}(t)$ in the diagram.

The input tariff alters both the effectiveness of a ROO and its ability to increase welfare. Regarding effectiveness, we return to the homogeneous supplier case of subsection 5.1 and redefine the Y-chain-level welfare effect as $\Delta\widetilde{\Psi}(r, t, \omega)$.[29] The cutoffs $r_{UC}(\omega)$ and $r_{NC}(\omega, \tau)$ become now functions of t :

$$r_{UC}(\omega, t) \equiv \tfrac{2(p_w^z + t - \omega)}{\bar{Q}(2c - \alpha b^2)} \text{ and}$$

$$r_{NC}(\omega, \tau, t) \equiv \tfrac{2(p_w^z + t - \omega)}{\bar{Q}(2c - \alpha b^2)} + 2\sqrt{\tfrac{\tau}{\bar{Q}(2c - \alpha b^2)}}.$$

[29] Observe that the aggregate welfare effect of a ROO is still given by the integral over the joint profit of all Y-chains, gross of tariff payments. The reason is that the tariff revenue due to imports of z is a transfer from the Y-chains to the *Home* government, and therefore neutral from society's perspective.

Because specialized sourcing is increasing in t, it becomes more likely that vertical chains will meet a given r unconstrained. Thus, to be binding, the ROO needs to be tighter. The upshot is that the input tariff shifts $[r_{UC}(\omega, t), r_{NC}(\omega, t, \tau)]$ up.

This shift may improve welfare, provided that τ is relatively low. In particular, if $r_{NC}(\omega, \tau, 0) < \hat{r}(\omega) < r_{NC}(\omega, \tau, t)$, then input tariff $t > 0$ enables the use of ROO $\hat{r}(\omega)$ to yield investment i^{sb}. On the other hand, if τ is high enough so that $\hat{r}(\omega) < r_{NC}(\omega, \tau, 0)$, then $\hat{r}(\omega)$ is binding with no input tariff but there is no additional welfare improvement to be had. But if t is sufficiently high, then $\hat{r}(\omega) \leq r_{UC}(\omega, t)$, so that the optimal ROO at the Y-chain level is infeasible. In this case, $i_t^* \geq i^{sb}$, as in Figure 6 for $\omega \geq \hat{\omega}(t)$.[30] In such cases, it is optimal to choose an innocuous ROO. We have the following proposition.

Proposition 5 *Let the distribution of inverse productivity be degenerate and centered on any $\omega \in (0, p_w^z)$. There exists a $\hat{t}(\omega)$ such that:*
(i) If $t < \hat{t}(\omega) - \sqrt{(2c - \alpha b^2)\tau\bar{Q}}$, then $\Delta\widetilde{\Psi}(r, t, \omega)$ is maximized with rule of origin $r^*(\omega, t) = r_{NC}(\omega, \tau, t)$;

(ii) If $t \in [\hat{t}(\omega) - \sqrt{(2c - \alpha b^2)\tau\bar{Q}}, \hat{t}(\omega)]$, then $\Delta\widetilde{\Psi}(r, t, \omega)$ is maximized with rule of origin $r^*(\omega, t) = \hat{r}(\omega)$;

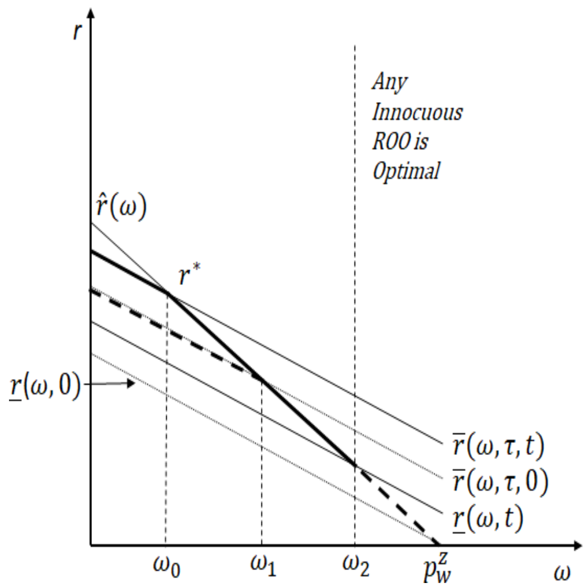
(iii) If $t > \hat{t}(\omega)$, then welfare cannot be improved with a rule of origin.

The lower cutoff, $\hat{t}(\omega) - \sqrt{(2c - \alpha b^2)\tau\bar{Q}}$, is found by setting $\hat{r}(\omega) = r_{NC}(\omega, \tau, t)$ and solving for t. The higher cutoff, $\hat{t}(\omega)$, is found by setting $\hat{r}(\omega) = r_{UC}(\omega, t)$ and solving for t. It maximizes the Y-chain-level welfare effect when there is a strictly positive input tariff but no ROO is used. If t is higher than that, then investment already exceeds i^{sb} with no ROO. Because the ROO can only increase investment, it is best not to have a binding ROO.

Figure 7 shows how $r^*(\omega)$ changes with a positive t. The input external tariff shifts the range $[r_{UC}(\omega, t), r_{NC}(\omega, \tau, t)]$ up for all ω. Generally, the input tariff is more likely to be helpful when productivity is higher; it makes $r^*(\omega)$ strictly higher for $\omega \leq \omega_1$. For $\omega \leq \omega_0$, $r^*(\omega) = r_{NC}(\omega, \tau, t)$. For $\omega \in (\omega_0, \omega_1]$, $r^*(\omega) = \hat{r}(\omega)$. Welfare rises in each of these cases. For $\omega \in (\omega_1, \omega_2)$, $r^*(\omega)$ is the same as in Figure 3a (and note the optimal ROO in the $t = 0$ case

[30] Note that $\hat{\omega}(t)$ is the level of inverse productivity such that the tariff is exactly $t = \hat{t}(\hat{\omega}(t))$, where $\hat{t}$ is derived in Proposition 5.

Figure 7: *Optimal ROO, Homogeneous Suppliers, $t > 0$*



Note: This diagram highlights the optimal ROO for the case where all suppliers have the same level of inverse productivity and there is a positive input tariff ($t > 0$). The x-axis is the level of inverse productivity ω that all suppliers have, while the bold-face function is the optimal ROO $r^*(\omega)$ for that set of suppliers. The optimal ROO r^* under $t = 0$ (Figure 3a) is shown by the dashed line for $\omega < \omega_1$ and $\omega > \omega_2$, and is the same as the optimal ROO with $t > 0$ for $\omega \in [\omega_1, \omega_2]$.

is otherwise represented by the bold dashed line). For $\omega > \omega_2$, the input tariff eliminates the usefulness of a ROO. To the right of the vertical dashed line, any effective ROO will worsen welfare. An innocuous ROO is therefore optimal.

Note that τ may be both a complement and a substitute for t. In Figure 7, τ and t are complements for some $\omega < \omega_0$. If τ rises, then an increase in t enables higher welfare for additional levels of ω. But if $\tau = \tau'$ is as large as in Figure 3b, then with no input tariff, $r^*(\omega) = \hat{r}(\omega)$ for all ω. With an input tariff, $r_{UC}(\omega, t)$ shifts up as in Figure 7, and for $\omega > \omega_2$ a ROO now cannot help. But because τ' is so large, there are no ω where the input tariff enables an increase in welfare. The final-good tariff only substitutes for the input tariff.

Now consider the effect of a ROO for the case of heterogeneous suppliers. Incorporating ROO-effectiveness, the aggregate welfare effect can now be written as

$$\Delta W(r) = \int_{\omega_{UC}(t)}^{\omega_{NC}(\tau,t)} \Delta\Psi(r, t, \omega) g(\omega) d\omega,$$

where the thresholds $\omega_{UC}(t)$ and $\omega_{NC}(\tau, t)$ are analogous to those from Proposition 1. When interior, they satisfy

$$\begin{aligned} \omega_{UC}(t) &= p_w^z + t - q_r\left(\tfrac{2c-\alpha b^2}{2}\right) \text{ and} \\ \omega_{NC}(\tau, t) &= p_w^z + t - q_r\left(\tfrac{2c-\alpha b^2}{2}\right) + \sqrt{(2c-\alpha b^2)\tau\bar{Q}}. \end{aligned}$$

Defining

$$\underline{\tau}(\omega, t) \equiv \left(\frac{1}{\bar{Q}(2c-\alpha b^2)}\right)\left(\frac{(p_w^z - \omega)\alpha(1-\alpha)b^2}{2c - 2\alpha b^2 + \alpha^2 b^2} - t\right)^2$$

as the analog to $\underline{\tau}(\omega)$, the following result characterizes how the level of r affects welfare.

Proposition 6 *There exist values r_t^- and r_t^+ such that, for any distribution of suppliers, the welfare effect of the rule of origin, $\Delta W(r)$, has the following properties:*
(i) If $r < r_t^-$, then $\Delta W(r) \leq 0$.
(ii) If $r > r_t^+$, then $\Delta W(r) \geq 0$.
If in addition $\tau < 4\underline{\tau}(0, t)$, then if $r \in [r_t^-, r_t^+]$, there exists an ω_t^0 such that $\Delta\Psi(r, t, \omega) \geq 0$ for $\omega \in [\omega_{UC}(t), \omega_t^0]$ and $\Delta\Psi(r, t, \omega) < 0$ for $\omega \in (\omega_t^0, \omega_{NC}(\tau, t)]$.

As with $t = 0$, in the interior case a ROO is more likely to be beneficial if it is tighter, and for the same reasons: a tight rule affects the behavior of high-productivity suppliers. The key difference when $t > 0$ is that the aggregate welfare effect can be strictly negative for any distribution of ω (and τ) because a ROO may aggravate overinvestment by *all* affected suppliers.[31]

Note that both r_t^- and r_t^+ increase with t. With a higher external tariff on intermediate inputs, more suppliers overinvest and fewer underinvest, so a higher r is needed to affect only those that underinvest. For a sufficiently high t, all suppliers overinvest and welfare improvements are impossible.

Corollary 2 *If $t > \frac{\alpha(1-\alpha)b^2\bar{Q}}{2}$, then the introduction of any rule of origin strictly decreases welfare.*

[31] As with Proposition 3, there are scenarios ($\tau > 4\underline{\tau}(0, t)$) where r increases to the point where $\omega_{UC}(t)$ obtains at the 0 boundary, $\omega_t^0 < 0$ and $\omega_{NC}(\tau, t)$, in which case the welfare effect becomes strictly negative for $\omega \in [\omega_{UC}(t), \omega_{NC}(\tau, t)]$. This occurs, as before, when $\alpha(1-\alpha)b^2$ is small relative to τ.

7 Extensions

In this section, we discuss how our results would change under some important alternative specifications to our benchmark model. We also describe some testable implications of the model.

7.1 Location of Inputs

Since a large fraction of global value chains are actually regional, and are often circumscribed to members of FTAs, we conduct our analysis assuming that the key economic relationship between firms takes place within a trading bloc. This, of course, is not always the case, and specialized and generic suppliers may operate in geographic regions other than those in our baseline model.

Consider first the situation where both types of suppliers are in ROW. In that case, the ROO would be immaterial because there would be no within-bloc sources of input supply. The ROO would be equally mute if both types of inputs are fully available within the trading bloc: regardless of the input mix, compliance would be assured.

A more interesting alternative obtains when specialized suppliers are in ROW while generic suppliers are in either $Home$ or $Foreign$. Essentially, this reverses the location considered in our analysis. In that case, constrained compliance with a ROO would induce more sourcing of generic within-bloc inputs, crowding out the supply of specialized inputs. Thus, the ROO would yield inefficient sourcing while also reducing relationship-specific investments, thereby worsening hold-up problems. This is clearly bad for welfare. Extending the model in this direction would therefore reinforce a central insight from our analysis, made clear in section 5.3: if ROOs do not stimulate relationship-specific investments and attenuate holdup problems, then they cannot improve welfare.

Finally, there could be a mixed situation, in which some specialized suppliers are within the FTA but others are in ROW, as is the supply of generic inputs. In that case, our analysis applies to the Y-chains with specialized suppliers in the FTA, but not to the others. Thus, our results are more relevant, the more important is customized sourcing within the FTA.

7.2 Adjustments in Q

In our benchmark model, we impose assumptions that make it optimal for vertical chains to always choose $Q = \bar{Q}$. This is very helpful to highlight the implications of a ROO for investment and the sourcing of inputs, but is of course an artificial assumption. More generally, vertical chains have two margins of adjustment when complying with a ROO: (1) the mix of inputs for given level of production; and (2) the level of production for a given mix of inputs. Our analysis shuts down the second margin, but in general firms may find it optimal to alter both margins.

In particular, firms may reduce the total number of inputs, Q, to comply with the ROO. Lowering Q reduces the need to increase q, and this could be a less costly way to comply with the ROO in some circumstances. When this occurs, the welfare effect of the ROO changes both because fewer final goods are produced within the bloc and because this reduces the need to increase investments. Importantly, whether lower production of the final good within the bloc increases or decreases welfare by itself is, in principle, ambiguous. In the case of enhanced protection considered here, it is welfare-improving because fewer final goods mitigate trade diversion, but the opposite would happen if there were trade creation in final goods.

Handling this possibility is possible in our setting, but cases where $Q < \bar{Q}$ obtains in equilibrium require a number of conditions to hold simultaneously.[32] To see the intuition, note that to preclude this outcome, it suffices to assume that $(p_w^x - p_w^z)$ is sufficiently large. By reducing Q, a vertical chain costs itself the margin $(p_w^x - p_w^z)$ on every unit not produced in the equilibrium bargain that would be produced absent that bargain. When this margin is large, the chain prefers to set $Q = \bar{Q}$, and instead just adjusts the mix of inputs to comply with the ROO, as in our benchmark analysis. A high final-good tariff also pushes the chain in this direction, because reducing Q sacrifices τ of surplus per unit of foregone final-good production. A lenient ROO and high supplier productivity similarly push the chain to keep Q high, because both changes make it easier to comply by increasing specialized inputs. Thus, the most likely situation where a chain would adjust Q downwards would be if r is

[32]See Appendix B for the basic analysis.

both large and binding for relatively unproductive suppliers, but where the final-good tariff is relatively low. Recalling that a high r tends to bind for high-productivity suppliers and a low τ decreases the range in which the ROO binds, it follows that the conditions that make $Q < \bar{Q}$ optimal are relatively difficult to satisfy and, when they are satisfied, the ROO is not particularly relevant (i.e., τ is low and the rule has little bite).

In the general case where the production function of the final good is smooth in the number of inputs, Q, constrained compliance will typically imply both lower Q and more specialized inputs, q. Thus, our current analysis of how a ROO induces more sourcing of within-bloc specialized inputs and its novel welfare implications would extend to that general case. The difference is that the conventional – and ambiguous – welfare changes due to the final-goods market would need to be taken into account as well.

7.3 Administrative Cost of Compliance

We assume that, if a vertical chain chooses quantities so that $q \geq rQ$, then it automatically satisfies the ROO. In reality, there are additional costs of compliance related to the necessity to prove to the customs authority that, indeed, the firm's choice of inputs satisfies $q \geq rQ$. These include, for example, the cost of keeping additional records of transactions and of filling out additional border documents. Such costs are often considered equivalent to an increase in marginal costs (or to a reduction of the preferential margin), although the magnitudes of the estimates vary with the study and with the trading bloc in analysis (e.g., in the analyses in Cadot et al., 2006, they vary from 2 to 7 percent). They could also include a fixed cost component. Regardless of how they are considered, numerous authors argue that such costs are significant enough to induce non-compliance by many firms.

It is relatively straightforward to incorporate administrative costs into our setup. Recall that the gain from compliance for a vertical chain is $\tau\bar{Q}$, the total tariff savings when it exports $\bar{Q}$ units of the final good to the FTA partner. Following the literature, the documentation costs of compliance could be represented as $\delta\bar{Q}$, thus reducing the gain from compliance to $(\tau - \delta)\bar{Q}$. Naturally, if the per-unit documentation cost were larger than the final-good tariff, $\delta \geq \tau$, then no vertical chain complies constrained and the ROO does not affect the inputs market. Otherwise, the analysis of the private decisions of the vertical chains carries over just as before, but for an "adjusted" final-good tariff of $\tau' \equiv \tau - \delta$.[33]

Now, the welfare analysis does change with the introduction of the fixed documentation/administrative costs, because such costs are a real burden to the economy, and not just a transfer (as in the case of the tariff proceeds, $\tau\bar{Q}$). Therefore, whenever there is rule compliance, we would need to subtract $\delta\bar{Q}$ from the welfare calculation. The upshot is that, the higher the documentation and administrative costs, the less attractive are ROOs for the society, all else equal.

7.4 Endogenous Matching

We consider that producers of final goods and suppliers of customized inputs have already matched. By doing so we simplify along two grounds. First, and most obviously, we do not study how a ROO would disturb those matches. While potentially important, many of the insights from the analysis of how the formation of an FTA affects matching in OTB would carry over for the introduction of a ROO. In particular, while the prospect of free trade in final goods increases the attractiveness of matching inside the bloc, the ROO requirement tends to decrease the extent of "matching diversion" – i.e., the matching with low-productivity suppliers inside the FTA at the expense of matching with high-productivity suppliers outside the bloc.

Second, even if we left aside how a ROO would affect the matching equilibrium, we know from OTB that, by endogenizing the matching process, an inevitable consequence is that low-productivity suppliers will not match, being relegated in favor of specialized suppliers outside the FTA. The implication is that the potential for a ROO to cause harm is reduced. Consider, for example, Figure 4. The only source of welfare loss stems from high-ω suppliers (when r is relatively low). If those suppliers are not matched into Y-chains and revert to producing the numéraire good, then it becomes more likely that a ROO will induce a welfare gain, even when r is relatively low.

[33] Observe that, in our setting, we can think of $\delta\bar{Q}$ also as a fixed cost.

7.5 Positive Implications

An empirical assessment of the welfare predictions of our model would be challenging, and would probably require a quantitative model. On the other hand, testing the observable implications of the model would be more straightforward. For that, one would need to rely on the introduction of ROOs in a new FTA, or changes in existing ROOs, in addition to firm-level information.

First, our model is precise about what type of firm is likely to comply with ROOs: high-productivity ones. Moreover, as the rules become stricter, eventually the range of firms complying shortens, and the ones that stop complying are those with intermediate levels of productivity, for whom compliance was barely profitable initially.

Second, our model is also clear about which firms, among the compliers, will change behavior because of ROOs. Our results show that, although high-productivity firms will generally satisfy the rules, they will not change their behavior – that is, for them the rule is innocuous. On the other hand, the compliers with lower productivity will increase within-FTA sourcing and investment. Accordingly, their **observed** productivity (which considers both their "fundamental" productivity parameter ω and their investment) should increase as a result of introducing (or tightening) ROOs.

This heterogeneous behavior, with greater investment reaction from mid-range productivity firms, is similar to what Lileeva and Trefler (2010) find when studying the U.S.-Canada FTA. As data on utilization rates becomes more available (see Kniahin and Melo, 2022), testing these implications will become more feasible, and falsifying the building blocs of the model will become easier.[34]

8 Conclusion

We study the welfare effects of rules of origin in free trade agreements with a property-rights model. Given the nature of modern global value chains and their prevalence within FTAs, this approach seems natural. We design the details of the model so that we can derive

[34]The model also has other, more straightforward, testable implications. For example, a higher tariff on final goods induces compliance by a wider range of firms. This has been documented in several FTAs.

clear-cut analytical solutions, but the two key assumptions on which the analysis rests are much more general. First, hold-up problems matter. Second, trade policies affect investment incentives. This makes clear when our conclusions do and do not hold. A prominent case when they do not is a competitive market.

In addition to being novel, the implications of ROOs from a property-rights perspective are vastly different from those under a competitive setting. We show that welfare may rise or fall with the imposition of such a rule, but that the effects tend to be more positive when the rule requires a higher fraction of within-bloc inputs. Moreover, a sufficiently strict rule can ensure a positive welfare impact, at least if input external tariffs are not exceedingly high. Thus, our analysis provides an efficiency rationale for ROOs even when, as Felbermayr et al. (2019) forcefully argue, they are not necessary to prevent trade deflection.

Some analysts (e.g., Crivelli et al., 2021) suggest that more lenient rules should be used to induce higher levels of preference utilization. This is also a common message of several papers that find that ROOs induce changes in sourcing patterns, which implicitly assume that sourcing and investment are efficient in the absence of ROOs. Our analysis recommends caution in such policy proposals. In our setting, making rules of origin more lenient would induce some low-productivity suppliers to comply and make the rule innocuous for some high-productivity suppliers. The former effect tends to decrease welfare, while the latter may imply forgoing gains from mitigating hold-up problems precisely when they really matter. More generally, our analysis shows that understanding the organization of the firms affected by the rules is critical for their normative assessment.

Clearly, the design of ROOs has several practical dimensions that we bypass in our analysis. The lack of transparency and clarity in actual ROOs, their multiplicity across products and agreements, and the distinct ways of defining origin are all important dimensions from a practical perspective. Attempts at defining "best practices" and at "multilateralizing" the rules under the auspices of the World Trade Organization have so far failed, but developments in those directions would surely be helpful.[35] Indeed, the insights from our analysis would be more useful following those developments.

[35]See, for example, the discussion and suggestions of Hoekman and Inama (2018) and Kniahin and Melo (2022).

References

Antràs, Pol. "Conceptual Aspects of Global Value Chains," *World Bank Economic Review* 34(3), 551-574, 2020.

Antràs, Pol and Elhanan Helpman. "Global sourcing," *Journal of Political Economy* 112, 552-580, 2004.

Antràs, Pol and Robert Staiger. "Offshoring and the Role of Trade Agreements," *American Economic Review* 102(7), 3140-3183, 2012.

Baldwin, Richard. "21st century regionalism: Filling the gap between 21st century trade and 20th century trade rules," World Trade Organization, Staff Working Paper ERSD-2011-08, 2011.

Boehm, Johannes and Ezra Oberfield. "Misallocation in the Market for Inputs: Enforcement and the Organization of Production," *Quarterly Journal of Economics* 135(4), 2007–2058, 2020.

Bombarda, Pamela and Elisa Gamberoni. "Firm Heterogeneity, Rules of Origin, and Rules of Cumulation," *International Economic Review* 54(1), 307-328, 2013.

Cadot, Olivier, Antoni Estevadeordal, Akiko Suwa-Eisenmann, and Thierry Verdier. The origin of goods: Rules of origin in regional trade agreements, Oxford: Oxford University Press, 2006.

Celik, Levent, Bilgehan Karabay, and John McLaren. "Fast-track authority: A hold-up interpretation," *Journal of International Economics* 127, article 103392, 2020.

Cherkashin, Ivan, Svetlana Demidova, Hiau L. Kee, and Kala Krishna. "Firm heterogeneity and costly trade: A new estimation strategy and policy experiments," *Journal of International Economics* 96(1), 18–36, 2015.

Chung, Wanyu and Carlo Perroni. "Rules of Origin and Market Power," CEPR Discussion Paper DP15897, 2021.

Conconi, Paola, Manuel Garcia-Santana, Laura Puccio, and Roberto Venturini. "From Final

Goods to Inputs: The Protectionist Effects of Rules of Origin," *American Economic Review* 108, 2335-65, 2018.

Crivelli, Pramila, Stefano Inama, and Jonas Kasteng. "Using Utilization Rates to Identify Rules of Origin Reforms: The Case of EU Free Trade Area Agreements," EUI Working Paper RSC 2021/21, 2021.

Falvey, Rod and Geoff Reed. "Economic effects of rules of origin," *Review of World Economics* 134, 209-229, 1998.

Felbermayr, Gabriel, Feodora Teti, and Erdal Yalcin. "Rules of origin and the profitability of trade deflection," *Journal of International Economics* 121, article 103248, 2019.

Grossman, Gene. "The Theory of Domestic Content Protection and Content Preference," *Quarterly Journal of Economics* 96, 583-603, 1981.

Grossman, Gene and Elhanan Helpman. "The Politics of Free-Trade Agreements," *American Economic Review* 85, 667-690, 1995.

Grossman, Gene and Elhanan Helpman. When Tariffs Disrupt Global Supply Chains, mimeo, 2021.

Head, Keith, Thierry Mayer, and Marc Melitz. "Moving parts: When more restrictive content rules backfire," mimeo, 2022.

Hoekman, Bernard and Stefano Inama. "Harmonization of Rules of Origin: An Agenda for Plurilateral Cooperation?" *East Asian Economic Review* 22(1), 3-28, 2018.

Inama, Stefano. Rules of Origin in International Trade, Cambridge University Press: New York, NY, 2009.

Johnson, Robert and Guillermo Noguera. "A Portrait of Trade in Value Added Over Four Decades," *The Review of Economics and Statistics* 99(5), 896-911, 2017.

Kniahin, Dzmitry and Jaime de Melo. "A Primer on Rules of Origin as Non-Tariff Barriers," CEPR Discussion Paper 17076, 2022.

Krishna, Kala. "Understanding Rules of Origin." In: Olivier Cadot et al. (eds.), The Origin of Goods: Rules of Origin in Regional Trade Agreements, 191–212. Oxford: Oxford University Press, 2006.

Krishna, Kala, Carlos Salamanca, Yuta Suzuki, and Christian Volpe-Martincus. "Learning to Use Trade Agreements," NBER Working Paper 29319, 2021.

Krueger, Anne. "Free trade agreements as protectionist devices: rules of origin." In J. Melvin, J. Moore and R. Riezmond (eds.), Trade, Theory and Econometrics: Essays in Honor of John C. Chipman, 91–102. London: Routledge, 1999.

Lileeva, Alla and Daniel Trefler. "Improved Access to Foreign Markets Raises Plant-Level Productivity ... for Some Plants," *Quarterly Journal of Economics* 125(3), 1051-1099, 2010.

Maggi, Giovanni, Monika Mrázová, and Peter Neary. "Choked by red tape? The political economy of wasteful trade barriers," *International Economic Review* 64 (1), 161–188, 2022.

Ornelas, Emanuel and John L. Turner. "Trade Liberalization, Outsourcing and the Hold-Up Problem," *Journal of International Economics* 74, 225-41, 2008.

Ornelas, Emanuel and John L. Turner. "Protection and international sourcing," *Economic Journal* 122, 26–63, 2012.

Ornelas, Emanuel, John L. Turner, and Grant Bickwit. "Preferential Trade Agreements and Global Sourcing," *Journal of International Economics* 128, Article 103395, 2021.

Ruta, Michele. "Preferential trade agreements and global value chains: Theory, evidence, and open questions." In: Global Value Chain Development Report: Measuring and Analyzing the Impact of GVCs on Economic Development, World Bank, Washington DC, 175-185, 2017.

Saggi, Kamal and Halis Yildiz. "Bilateralism, multilateralism, and the quest for global free trade," *Journal of International Economics* 86, 26-37, 2010.

Sytsma, Tobias. "Improving Preferential Market Access through Rules of Origin: Firm-Level Evidence from Bangladesh," *American Economic Journal: Economic Policy* 14(1), 440-472, 2022.

Tsirekidze, David. “Global supply chains, trade agreements and rules of origin,” *The World Economy* 44(11), 3111-3140, 2021.

Appendix A: Proofs

Proof of Lemma 1. Ignoring the ROO constraint, note that both Σ^{RC} and Σ^{NC} are strictly concave in q and are maximized by $q_0 = (p_w^z - \omega + bi)/c$. Then, because $\Sigma^{RC}(q_0, i, \omega) > \Sigma^{NC}(q_0, i, \omega)$, if q_0 is ROO-compliant, then q_0 is optimal. The expression for $i_{UC}(\omega) \equiv \frac{\omega-(p_w^z+t)+cq_r}{b}$ is found by setting $q_0 = q_r$ and solving for ω, and it follows that $q_0 > q_r$ precisely when $i > i_{UC}$. This shows that $q_i = q_0$ if $i > i_{UC}$.

To find the expression for i_{NC}, note first that Σ^{RC} and Σ^{NC} are both increasing functions of i, with $\frac{d\Sigma^{RC}(q_r,i,\omega)}{di} = bq_r$ and $\frac{d\Sigma^{NC}(q_0,i,\omega)}{di} = bq_0$. As i falls below i_{UC}, we know that $q_r > q_0$; hence $\Sigma^{RC}(q_r, i, \omega)$ falls faster than $\Sigma^{NC}(q_0, i, \omega)$, and $\Sigma^{RC}(q_r, i, \omega) = \Sigma^{NC}(q_0, i, \omega)$ for a unique value of i. The cutoff $i_{NC}(\omega)$ is the level of investment that solves this. It follows that $\Sigma^{RC}(q_r, i, \omega) < \Sigma^{NC}(q_0, i, \omega)$ for $i < i_{NC}$, so that $q_i = q_0$ if $i < i_{NC}$. It also follows that $\Sigma^{RC}(q_r, i, \omega) > \Sigma^{NC}(q_0, i, \omega)$ for $i \in (i_{NC}, i_{UC})$, so that $q_i = q_r$ if $i \in (i_{NC}, i_{UC})$. Note that, if $i = i_{UC}$, then $q_r = q_0$ and $\Sigma^{RC}(q_r, i_{UC}, \omega) = \Sigma^{RC}(q_0, i_{UC}, \omega)$. ■

Proof of Proposition 1. We first show that $U_S(i)$ is continuous on $\left[0, \frac{bp_w^z}{2c-b^2}\right]$. Because q_0 and q_r are continuous functions, we have that $\Sigma^{NC}(i, q_0, \omega)$ is continuous on $[0, i_{NC}]$, that $\Sigma^{RC}(q_r, i, \omega)$ is continuous on $[i_{NC}, i_{UC}]$, and that $\Sigma^{RC}(q_0, i, \omega)$ is continuous for $i \in [i_{UC}, \frac{bp_w^z}{2c-b^2}]$, where the upper bound of this interval corresponds to S's choice of investment when $\omega = 0$ and $\alpha = 1$.

It remains to establish that $U_S(i)$ is continuous at $i = i_{NC}$ and at $i = i_{UC}$. For the first condition, the proof of Lemma 1 shows that i_{NC} is defined as the unique i such that $\Sigma^{RC}(q_r, i, \omega) = \Sigma^{NC}(q_0, i, \omega)$. Hence $\lim_{i \to i_{NC}} U_S(i) = \Sigma^{RC}(q_r, i_{NC}, \omega) = \Sigma^{NC}(q_0, i_{NC}, \omega)$, and $U_S(i)$ is continuous at $i = i_{NC}(\omega)$. For the second condition, note that $i_{UC}(\omega)$ is defined so that $q_0 = q_r$. Hence, $\Sigma^{RC}(q_0, i_{UC}, \omega) = \Sigma^{RC}(q_r, i_{UC}, \omega)$ and $U_S(i)$ is continuous at $i = i_{UC}$.

Given that $U_S(i)$ is continuous on $\left[0, \frac{bp_w^z}{2c-b^2}\right]$, it attains a maximum. Using the envelope theorem, it is straightforward to see that maximizing U_S when either $\Sigma = \Sigma^{RC}(q_i, i, \omega)$ or $\Sigma = \Sigma^{NC}(q_i, i, \omega)$ yields $I'(i^*) = -\alpha C_i(q_{i^*}, i^*, \omega)$. If $q_{i^*} = q_0$, then with our functional forms we can solve to find

$$\begin{aligned} q_0^* &= \tfrac{2(p_w^z-\omega)}{2c-\alpha b^2}, \\ i_0^* &= \tfrac{\alpha b(p_w^z-\omega)}{2c-\alpha b^2}. \end{aligned} \tag{12}$$

If $q_{i^*} = q_r$, then $i^* = i_r^* = \frac{\alpha b q_r}{2}$. From Lemma 1, we know that q_0 maximizes Σ when i is such that $q_0 \geq q_r$. Hence, i_0^* maximizes $U_S(i)$ if q_0^* in (12) exceeds $q_r^* = r\bar{Q}$. The cutoff

$\omega_{UC} = p_w^z - \frac{r\bar{Q}(2c-\alpha b^2)}{2}$ is found by setting $q_0^* = q_r^*$ and solving for ω. Because q_0^* is decreasing in ω, we have shown that $i^* = i_0^*$ if $\omega < \omega_{UC}$.

To find the expression for $\omega_{NC}(\tau)$, note that $\Sigma^{RC}(q_0^*, i_0^*, \omega)$ equals $\Sigma^{NC}(q_0^*, i_0^*, \omega)$ plus a constant term that reflects the additional producer surplus coming from ROO compliance. Thus, the envelope theorem implies that the rate of change of supplier profit is

$$\frac{dU_S}{d\omega} = -\alpha C_\omega(q_0^*, i_0^*, \omega) = -\alpha q_0^*.$$

Because q_0^* is itself a decreasing function of ω, these profit functions are decreasing and strictly convex functions of ω. The envelope theorem also applies for the rate of change for profit under constrained compliance:

$$\frac{dU_S(q_r^*, i_r^*, \omega)}{d\omega} = -\alpha C_\omega(q_r^*, i_r^*, \omega) = -\alpha q_r^*.$$

If $\omega > \omega_{UC}$, then $q_r^* > q_0^*$. Hence, $U_S(q_r^*, i_r^*, \omega)$ has a steeper slope than $U_S(q_0^*.i_0^*, \omega)$. This implies that there is a unique value $\omega_{NC}(\tau)$ satisfying $U_S(q_r^*, i_r^*, \omega_{NC}(\tau)) = \Sigma^{NC}(q_0^*, i_0^*, \omega_{NC}(\tau))$. Solving this equation yields $\omega_{NC}(\tau) = p_w^z - \frac{r\bar{Q}(2c-\alpha b^2)}{2} + \sqrt{(2c - \alpha b^2)\tau\bar{Q}}$.

Moreover, it follows that $\Sigma^{RC}(q_r^*, i_r^*, \omega) < \Sigma^{NC}(q_0^*, i_0^*, \omega)$ for $\omega > \omega_{NC}(\tau)$, so that $i^* = i_0^*$ if $\omega > \omega_{NC}(\tau)$. It also follows that $\Sigma^{RC}(q_r^*, i_r^*, \omega) > \Sigma^{NC}(q_0^*, i_0^*, \omega)$ for $i \in (\omega_{UC}, \omega_{NC}(\tau))$, so that $i^* = i_r^*$ if $\omega \in [\omega_{UC}, \omega_{NC}(\tau)]$. Note that if $\omega = \omega_{UC}$, then $i_r^* = i_0^*$ and $\Sigma^{RC}(q_r^*, i_r^*, \omega_{UC}) = \Sigma^{RC}(q_0^*, i_0^*, \omega_{UC})$. ■

Proof of Lemma 2. By definition, $\Delta\tilde{\Psi}(\omega, r) = \Psi(\omega, q_{i^*}, i^*) - \Psi(\omega, q_0^*, i_0^*)$. The threshold values $r_{UC}(\omega)$ and $r_{NC}(\omega, \tau)$ are inversions of the ω_{UC} and $\omega_{NC}(\tau)$ thresholds, respectively. The lower threshold, found by setting $\omega_{UC} = \omega$ and solving for r, is

$$r_{UC}(\omega) \equiv \frac{2(p_w^z - \omega)}{\bar{Q}(2c - \alpha b^2)},$$

and is always between 0 and 1. Any $r < r_{UC}(\omega)$ yields unconstrained compliance ($\omega < \omega_{UC}$). The higher threshold, found by setting $\omega_{NC}(\tau) = \omega$, is

$$r_{NC}(\omega, \tau) \equiv \frac{2(p_w^z - \omega)}{\bar{Q}(2c - \alpha b^2)} + \frac{2\sqrt{(2c - \alpha b^2)\tau\bar{Q}}}{\bar{Q}(2c - \alpha b^2)}.$$

It increases in τ and exceeds 1 for sufficiently large τ. Any $r > r_{NC}(\omega, \tau)$ yields noncompliance ($\omega > \omega_{NC}(\tau)$).

For r outside $[r_{UC}(\omega), r_{NC}(\omega, \tau)]$, $i^* = i_0^*$ and $q_{i^*} = q_0^*$, so $\Delta\tilde{\Psi}(r, \omega) = 0$. For $r \in [r_{UC}(\omega), r_{NC}(\omega, \tau)]$, $i^* = i_r^*$ and $q_{i^*} = q_r^*$. We can then write

$$\begin{aligned}\Psi(q_{i^*}, i^*, \omega) &= p_w^z q_r^* - C(q_r^*, i_r^*, \omega) - I(i_r^*) \\ &= p_w^z q_r^* - (\omega - b i_r^*) q_r^* - \tfrac{c}{2} q_r^{*2} - i_r^{*2} \\ &= (p_w^z - \omega) q_r^* + b\left(\tfrac{\alpha b q_r^*}{2}\right) q_r^* - \tfrac{c}{2} q_r^{*2} - \left(\tfrac{\alpha b q_r^*}{2}\right)^2 \\ &= (p_w^z - \omega) q_r^* - \tfrac{(2c - 2\alpha b^2 + \alpha^2 b^2) q_r^{*2}}{4}.\end{aligned}$$

Similarly, we can derive

$$\Psi(\omega, q_0, i_0^*) = (p_w^z - \omega) q_0^* - \frac{(2c - 2\alpha b^2 + \alpha^2 b^2) q_0^{*2}}{4}.$$

Collecting terms, we then have

$$\Delta\tilde{\Psi}(\omega, r) = (p_w^z - \omega)(q_r^* - q_0^*) - \frac{(2c - 2\alpha b^2 + \alpha^2 b^2)(q_r^{*2} - q_0^{*2})}{4},$$

concluding the proof. ■

Proof of Proposition 2. Potential welfare,

$$\Delta\Psi(r, \omega) = (p_w^z - \omega)(q_r^* - q_0^*) - \frac{(2c - 2\alpha b^2 + \alpha^2 b^2)(q_r^{*2} - q_0^{*2})}{4},$$

is a strictly concave function of r that equals 0 at $r = r_{UC}(\omega)$ (where $q_0^* = q_r^*$). The first and second derivatives are

$$\begin{aligned}\tfrac{d\Delta\Psi(\omega, r)}{dr} &= (p_w^z - \omega)\bar{Q} - \tfrac{(2c - 2\alpha b^2 + \alpha^2 b^2)(2 q_r^* \bar{Q})}{4}, \\ \tfrac{d^2\Delta\Psi(\omega, r)}{dr^2} &= -\tfrac{(2c - 2\alpha b^2 + \alpha^2 b^2)\bar{Q}^2}{2}.\end{aligned}$$

The second-derivative is negative, so the function is strictly concave. Evaluating the first derivative at $r = r_{UC}(\omega)$ and rearranging, we have

$$\frac{d\Delta\Psi(\omega, r)}{dr}\Big|_{r = r_{UC}(\omega)} = \frac{(p_w^z - \omega)\alpha(1 - \alpha) b^2 \bar{Q}}{2c - \alpha b^2} > 0.$$

Hence, $\Delta\Psi(\omega, r)$ is increasing at $r = r_{UC}(\omega)$. Setting the first derivative equal to zero and solving, we find that $\Delta\Psi(\omega, r)$ is maximized at

$$\hat{r}(\omega) = \frac{2(p_w^z - \omega)}{\bar{Q}(2c - 2\alpha b^2 + \alpha^2 b^2)} > r_{UC}(\omega).$$

Hence, if $\hat{r}(\omega) \leq r_{NC}(\omega, \tau)$, then $\hat{r}(\omega)$ is binding and $r^* = \hat{r}(\omega)$. We find $\underline{\tau}(\omega)$ by setting $\hat{r}(\omega) = r_{NC}(\omega, \tau)$ and solving for τ. Because $r_{NC}(\omega, \tau)$ is increasing in τ, it follows that for any $\tau > \underline{\tau}(\omega)$, the unconstrained optimum $\hat{r}(\omega)$ is lower than $r_{NC}(\omega, \tau)$, so $\Delta\Psi(\omega, r)$ is an inverted-U function of r on $[r_{UC}(\omega), r_{NC}(\omega, \tau)]$ and is maximized at $r^* = \hat{r}(\omega)$ (part (i)). If $\tau \leq \underline{\tau}(\omega)$, then $\hat{r}(\omega) \geq r_{NC}(\omega, \tau)$ and $\Delta\Psi(\omega, r)$ is strictly increasing on $[r_{UC}(\omega), r_{NC}(\omega, \tau)]$ and is maximized at $r^* = r_{NC}(\omega, \tau)$ (part (ii)). It follows from this discussion that the optimal r^* is the minimum of $\hat{r}(\omega)$ and $r_{NC}(\omega, \tau)$ (part (iii)). ■

Proof of Corollary 1. From the derivation of $r_{NC}(\omega, \tau)$ in the proof of Lemma 2 and of $\hat{r}$ in the proof of Proposition 2, it is clear that both decrease with ω. Hence, $r^* = \min\{\hat{r}(\omega), r_{NC}(\omega, \tau)\}$ is likewise decreasing in ω, and therefore increasing in productivity. ■

Proof of Proposition 3. Potential welfare can be rewritten as

$$\Delta\Psi(r, \omega) = (q_r^* - q_0^*)\left[(p_w^z - \omega) - \frac{(2c - 2\alpha b^2 + \alpha^2 b^2)(q_r^* + q_0^*)}{4}\right]. \tag{13}$$

We start with the case where $0 < \omega_{UC} < \omega_{NC}(\tau) < p_w^z$ and handle the boundary cases later. For the interior case, if $\omega = \omega_{UC}$, $q_r^* = q_0^*$, so $\Delta\Psi(r, \omega_{UC}) = 0$. Recalling that q_0^* is a function of ω, the first derivative is

$$\begin{aligned} \tfrac{d\Delta\Psi(r,\omega)}{d\omega} &= (q_r^* - q_0^*)\left[-1 - \left(\tfrac{2c-2\alpha b^2+\alpha^2 b^2}{4}\right)\left(\tfrac{-2}{2c-\alpha b^2}\right)\right] \\ &+ \left(\tfrac{2}{2c-\alpha b^2}\right)\left[(p_w^z - \omega) - \tfrac{(2c-2\alpha b^2+\alpha^2 b^2)(q_r+q_0)}{4}\right], \end{aligned}$$

or equivalently,

$$\frac{d\Delta\Psi(r,\omega)}{d\omega} = -(q_r^* - q_0^*)\frac{(2c - \alpha^2 b^2)}{2(2c - \alpha b^2)} + \left(\frac{2}{2c - \alpha b^2}\right)\left[(p_w^z - \omega) - \frac{(2c - 2\alpha b^2 + \alpha^2 b^2)(q_r + q_0)}{4}\right]. \tag{14}$$

At $\omega = \omega_{UC}$, the first term vanishes. We can then solve to find

$$\frac{d\Delta\Psi(r,\omega)}{d\omega}|_{\omega=\omega_{UC}} = \frac{r\bar{Q}\alpha(1-\alpha)b^2}{2c - \alpha b^2} > 0.$$

Hence, $\Delta\Psi(r, \omega)$ is increasing in ω at $\omega = \omega_{UC}$. The second derivative is, after some rearranging,

$$\frac{d^2\Delta\Psi(r,\omega)}{d\omega^2} = \frac{-2(2c - \alpha^2 b^2)}{2c - \alpha b^2} < 0.$$

Hence, $\Delta\Psi(r, \omega)$ is concave.

With these characteristics, $\Delta\Psi(r,\omega)$ on $\omega \in (\omega_{UC}, \omega_{NC}(\tau)]$ can cross zero at most once and only from above. The cutoff r^+ is found by inserting $\omega = \omega_{NC}(\tau)$ into (13), setting it to zero, and solving for r. This yields

$$r^+ = \left(\frac{2c - \alpha^2 b^2}{\alpha(1-\alpha)b^2}\right)\sqrt{\frac{\tau}{\bar{Q}(2c - \alpha b^2)}}.$$

We now prove the first sentence of the proposition. Because $\omega_{NC}(\tau)$ is strictly decreasing in r, and because of the aforementioned concavity and single-crossing characteristics of $\Delta\Psi(r,\omega)$, it follows that in the interior case, for any $r > r^+$ and for any $\omega \in (\omega_{UC}, \omega_{NC}(\tau))$, $\Delta\Psi(r,\omega) > \Delta\Psi(r, \omega_{NC}(\tau)) > 0$. Now consider the possibility that r is such that ω_{UC} and $\omega_{NC}(\tau)$ are not both interior. Simple algebra shows that at $r = r^+$, $\omega_{NC}(\tau) < p^z_w$, i.e., the non-compliance threshold does not bind at the top. Hence, there are two boundary cases to consider: (1) $\omega_{UC} = 0 < \omega_{NC}(\tau) < p^z_w$; and (2) $\omega_{UC} = \omega_{NC}(\tau) = 0$. If $r \geq r^+$ and case (1) obtains, then by the preceding analysis $\Delta\Psi(r,\omega) > 0$ for all $\omega \in [0, \omega_{NC}(\tau)]$, where $\omega_{UC} = 0$. If $r \geq r^+$ and case (2) obtains, then the ROO is innocuous and potential welfare is zero. So continuing to assume $r \geq r^+$, let $g(\omega)$ be arbitrary. We know that $\Delta\Psi(r,\omega) \geq 0$ for all $\omega \in [\omega_{UC}, \omega_{NC}(\tau)]$. Hence $\Delta W(r) = \int_{\omega_{UC}}^{\omega_{NC}(\tau)} \Delta\Psi(r,\omega) g(\omega) d\omega \geq 0$.

The cutoff

$$\omega^0 \equiv p^z_w - q_r \left[\frac{(2c - \alpha b^2)(2c - 2\alpha b^2 + \alpha^2 b^2)}{2(2c - \alpha^2 b^2)}\right]$$

is found by setting the term in brackets in (13) equal to zero and solving for ω. Ignoring boundary conditions, it is easily shown that $\omega^0 \in [\omega_{UC}, \omega_{NC}(\tau)]$ if and only if $r \leq r^+$, and exceeds $\omega_{NC}(\tau)$ for higher r. Hence, for the interior case, assuming $r < r^+$, and noting again the concavity and single-crossing characteristics of $\Delta\Psi(r,\omega)$, it follows that $\Delta\Psi(r,\omega) > 0$ for all $\omega \in (\omega_{UC}, \omega^0)$ and $\Delta\Psi(r,\omega) < 0$ for all $\omega \in (\omega^0, \omega_{NC}(\tau)]$.

Now, let $\tau < 4\underline{\tau}(0)$. This assumption guarantees that, at $r = r^+$, $\omega^0 = \omega_{NC}(\tau) > 0$. In addition, if $r < r^+$, $0 < \omega^0 < \omega_{NC}(\tau)$, in which case ω_{UC} may equal 0 but it remains true that $\Delta\Psi(r,\omega) > 0$ for all $\omega \in (\omega_{UC}, \omega^0)$ and $\Delta\Psi(r,\omega) < 0$ for all $\omega \in (\omega^0, \omega_{NC}(\tau)]$. ■

Proof of Proposition 4. Suppose either $b = 0, \alpha = 0$ or $\alpha = 1$. Then, from the proof of Proposition 3, we know that in the interior case,

$$\frac{d\Delta\Psi(r,\omega)}{dr}\Big|_{r=r_{UC}(\omega)} = \frac{(p^z_w - \omega)\alpha(1-\alpha)b^2 r \bar{Q}^2}{2c - \alpha b^2} = 0,$$

and that $\frac{d^2\Delta\Psi(r,\omega)}{dr^2} < 0$. Summarizing, $\Delta\Psi(r_{UC}(\omega), \omega) = 0$ and $\Delta\Psi(r,\omega)$ is a decreasing and strictly concave function on $r \in [r_{UC}(\omega), r_{NC}(\omega,\tau)]$. Any binding ROO requires $r \in [r_{UC}(\omega), r_{NC}(\omega,\tau)]$. Hence, if the ROO binds, then $\Delta\Psi(r,\omega) \leq 0$ and welfare falls. ■

Proof of Proposition 5. Potential welfare with a positive tariff is

$$\Delta\Psi(r,t,\omega) = (p_w^z - \omega)(q_r^* - q_t^*) - \frac{(2c - 2\alpha b^2 + \alpha^2 b^2)\,(q_r^{*2} - q_t^{*2})}{4}.$$

This is a strictly concave function of r that equals 0 at $r = r_{UC}(\omega,t)$ (where $q_t^* = q_r^*$). The first and second derivatives are

$$\begin{aligned} \tfrac{d\Delta\Psi(\omega,t,r)}{dr} &= (p_w^z - \omega)\bar{Q} - \tfrac{(2c-2\alpha b^2+\alpha^2 b^2)\left(2q_r^*\bar{Q}\right)}{4}, \\ \tfrac{d^2\Delta\Psi(\omega,t,r)}{dr^2} &= -\tfrac{(2c-2\alpha b^2+\alpha^2 b^2)\bar{Q}^2}{2}. \end{aligned}$$

The second-derivative is negative, so the function is strictly concave. Evaluating the first derivative at $r = r_{UC}(\omega,t)$ and rearranging, we have

$$\frac{d\Delta\Psi(\omega,t,r)}{dr}|_{r=r_{UC}(\omega,t)} = \bar{Q}\left[\frac{(p_w^z - \omega)\alpha(1-\alpha)b^2}{2c - \alpha b^2} - \left(\frac{2c - 2\alpha b^2 + \alpha^2 b^2}{2c - \alpha b^2}\right)t\right].$$

This is zero for

$$t = \hat{t}(\omega) = \frac{\alpha(1-\alpha)b^2(p_w^z - \omega)}{(2c - 2\alpha b^2 + \alpha^2 b^2)}$$

and is strictly negative if $t > \hat{t}(\omega)$. Because $\Delta\Psi(\omega,t,r)$ is strictly concave, if $t > \hat{t}(\omega)$, then $\Delta\Psi(\omega,t,r) < 0$ for $r \in (r_{UC}(\omega,t), r_{NC}(\omega,\tau,t)]$. This proves part (iii).

Setting the first derivative equal to zero and solving, we find that (relaxing the effectiveness constraint) $\Delta\Psi(\omega,t,r)$ is maximized for $r = \hat{r}(\omega)$. If $\hat{r}(\omega) \in [r_{UC}(\omega,t), r_{NC}(\omega,\tau,t)]$, then $\hat{r}(\omega)$ is binding and $r^* = \hat{r}(\omega)$. If $\hat{r}(\omega) > r_{NC}(\omega,\tau,t)$, then $\hat{r}(\omega)$ is innocuous. Then by the concavity of $\Delta\Psi(\omega,t,r)$, $r^* = r_{NC}(\omega,\tau,t)$. We thus find the upper bound on t in part (ii) by setting $\hat{r}(\omega) = r_{UC}(\omega,t)$ and solving for t. We find the lower bound on t in part (ii) and the bound in part (i) by setting $\hat{r}(\omega) = r_{NC}(\omega,\tau,t)$ and solving for t. ■

Proof of Proposition 6. Rewrite potential welfare with a positive input tariff as

$$\Delta\Psi(r,t,\omega) = (q_r^* - q_t^*)\left[(p_w^z - \omega) - \frac{(2c - 2\alpha b^2 + \alpha^2 b^2)\,(q_r^* + q_t^*)}{4}\right]. \tag{15}$$

We start with the case where $0 < \omega_{UC}(t) < \omega_{NC}(\tau,t) < p_w^z$, handling the boundary cases later. For the interior case, if $\omega = \omega_{UC}(t)$, $q_r^* = q_t^*$, so $\Delta\Psi(r,t,\omega_{UC}(t)) = 0$. Recalling that q_t^* is a function of ω, the first derivative is

$$\begin{aligned} \tfrac{d\Delta\Psi(r,t,\omega)}{d\omega} \;&=\; (q_r^* - q_t^*)\left[-1 - \left(\tfrac{2c-2\alpha b^2+\alpha^2 b^2}{4}\right)\left(\tfrac{-2}{2c-\alpha b^2}\right)\right] \\ &+\; \left(\tfrac{2}{2c-\alpha b^2}\right)\left[(p_w^z - \omega) - \tfrac{(2c-2\alpha b^2+\alpha^2 b^2)\left(q_r^*+q_t^*\right)}{4}\right]. \end{aligned}$$

At $\omega = \omega_{UC}(t)$, the terms on the first line vanish. We can then solve to find

$$\frac{d\Delta\Psi(r,t,\omega)}{d\omega}|_{\omega=\omega_{UC}(t)} = \left(\frac{2}{2c-\alpha b^2}\right)\left(\frac{r\bar{Q}\alpha(1-\alpha)b^2}{2} - t\right).$$

Hence, $\Delta\Psi(r,t,\omega)$ is increasing in ω at $\omega = \omega_{UC}(t)$ if and only if $r \geq r_t^- \equiv t\left(\frac{2}{\bar{Q}\alpha(1-\alpha)b^2}\right)$.

The second derivative is, after some rearranging,

$$\frac{d^2\Delta\Psi(r,t,\omega)}{d\omega^2} = \frac{-2(2c-\alpha^2 b^2)}{2c-\alpha b^2} < 0,$$

so $\Delta\Psi(r,t,\omega)$ is concave. Because it is concave and $\Delta\Psi(r,\omega_{UC}(t)) = 0$, it follows that if $r < r_t^-$, then $\Delta\Psi(r,t,\omega_{UC}(t))$ is negative for all $\omega \in [\omega_{UC}(t), \omega_{NC}(\tau,t)]$. Note that it is obvious that this also holds if $\omega_{UC}(t)$ or $\omega_{NC}(\tau,t)$ occurs at a boundary. Let $g(\omega)$ be arbitrary. We know that $\Delta\Psi(r,\omega) \leq 0$ for all $\omega \in [\omega_{UC}(t), \omega_{NC}(\tau,t)]$. Hence $\Delta W(r) = \int_{\omega_{UC}(t)}^{\omega_{NC}(\tau,t)} \Delta\Psi(r,\omega)g(\omega)d\omega \leq 0$. This proves part (i).

If $r > r_t^-$, then (returning to the interior case) $\frac{d\Delta\Psi(r,t,\omega)}{d\omega}|_{\omega=\omega_{UC}} > 0$. Given the concavity of the function, $\Delta\Psi(r,t,\omega)$ on $\omega \in (\omega_{UC}(t), \omega_{NC}(\tau,t)]$ can cross zero at most once and only from above. The cutoff r_t^+ is found by inserting $\omega = \omega_{NC}(\tau,t)$ into (15), setting it to zero, and solving for r. This yields

$$r_t^+ \equiv t\left(\frac{2}{\bar{Q}\alpha(1-\alpha)b^2}\right)\left(\frac{(2c-\alpha^2b^2)(4c-3\alpha b^2+\alpha b^2)}{2(2c-\alpha b^2)^2}\right) + \left(\frac{(2c-\alpha^2b^2)\sqrt{\tau\bar{Q}}}{\sqrt{2c-\alpha b^2}\alpha(1-\alpha)b^2}\right).$$

We now prove part (ii). Because $\omega_{NC}(\tau,t)$ is strictly decreasing in r and because of the aforementioned concavity and single-crossing characteristics of $\Delta\Psi(r,t,\omega)$, it follows that in the interior case, for any $r > r_t^+$ and for any $\omega \in (\omega_{UC}(t), \omega_{NC}(\tau,t)]$, $\Delta\Psi(r,t,\omega) > \Delta\Psi(r,t,\omega_{NC}(\tau,t)) > 0$. Now consider the possibility that r is such that $\omega_{UC}(t)$ and $\omega_{NC}(\tau,t)$ are not both interior. Simple algebra shows that at $r = r_t^+$, $\omega_{NC}(\tau,t) < p_w^z$, i.e., the non-compliance threshold does not bind at the top. Hence, there are two boundary cases to consider: (1) $\omega_{UC}(t) = 0 < \omega_{NC}(\tau,t) < p_w^z$; and (2) $\omega_{UC}(t) = \omega_{NC}(\tau,t) = 0$. If $r > r_t^+$ and case (1) obtains, then by the preceding analysis $\Delta\Psi(r,\omega) > 0$ for all $\omega \in [0, \omega_{NC}(\tau,t)]$, where $\omega_{UC}(t) = 0$. If $r > r_t^+$ and case (2) obtains, then the ROO is innocuous and potential welfare is zero. So continuing to assume $r > r_t^+$, let $g(\omega)$ be arbitrary. We know that $\Delta\Psi(r,\omega) > 0$ for all $\omega \in [\omega_{UC}(t), \omega_{NC}(\tau,t)]$. Hence $\Delta W(r) = \int_{\omega_{UC}(t)}^{\omega_{NC}(\tau,t)} \Delta\Psi(r,\omega)g(\omega)d\omega > 0$.

The cutoff

$$\omega_t^0 \equiv p_w^z - t\left(\frac{2c-2\alpha b^2+\alpha^2b^2}{2c-\alpha^2b^2}\right) - q_r\left[\frac{(2c-\alpha b^2)(2c-2\alpha b^2+\alpha^2b^2)}{2(2c-\alpha^2b^2)}\right]$$

is found by setting the term in brackets in (15) equal to zero and solving for ω. Ignoring boundary conditions, it is easily shown that $\omega_t^0 \in [\omega_{UC}(t), \omega_{NC}(\tau, t)]$ if and only if $r \in [r_t^-, r_t^+]$. Hence, for the interior case, assuming $r \in [r_t^-, r_t^+]$ and noting again the concavity and single-crossing characteristics of $\Delta\Psi(r, t, \omega)$, it follows that $\Delta\Psi(r, t, \omega) \geq 0$ for all $\omega \in [\omega_{UC}(t), \omega_t^0]$ and $\Delta\Psi(r, t, \omega) < 0$ for all $\omega \in (\omega_t^0, \omega_{NC}(\tau, t)]$.

Finally, let $\tau < 4\underline{\tau}(0, t)$. This assumption guarantees that at $r = r_t^+$, $\omega^0 = \omega_{NC}(\tau, t) > 0$. In addition, if $r \in [r^-, r^+]$, then $0 < \omega^0 \leq \omega_{NC}(\tau, t)$, in which case $\omega_{UC}(t)$ may equal 0 but it remains true that $\Delta\Psi(r, \omega) \geq 0$ for all $\omega \in [\omega_{UC}(t), \omega^0]$ and $\Delta\Psi(r, \omega) < 0$ for all $\omega \in (\omega^0, \omega_{NC}(\tau, t)]$. This proves the final part. ■

Proof of Corollary 2. If $t > \frac{\alpha(1-\alpha)b^2\bar{Q}}{2}$, then $r_t^- > 1$. Therefore, any ROO satisfies $r < r_t^-$, so by Proposition 6 welfare is negative for any ω such that the ROO is binding. ■

Appendix B: Variable Q

Relax the assumption that $Q = \bar{Q}$. In the default of no bargain and no compliance, production of $\bar{Q}$ units is always optimal. Thus, we can write

$$U_F^T - U_F^0 = [(p_w^x + \tau)Q - p_w^z z - p^s q] - \left[(p_w^x - p_w^z)\,\bar{Q}\right].$$

With this expression, the bargaining surplus becomes

$$\Sigma^{RC} \equiv \tau Q + p_w^z q - C(q, i, \omega) - (p_w^x - p_w^z)\,(\bar{Q} - Q).$$

Obviously, if $Q = \bar{Q}$, this collapses back to the case considered in the main analysis. But if $Q < \bar{Q}$, the bargaining surplus is reduced because there is a margin of $(p_w^x - p_w^z)$ that is earned on $(\bar{Q} - Q)$ extra units in the default (versus the equilibrium).

Incorporating this, we can then write the maximization problem as

$$\max_{\{q,Q\}} \qquad \Sigma^{RC} \equiv \tau Q + p_w^z q - C(q, i, \omega) - (p_w^x - p_w^z)\,(\bar{Q} - Q)$$

such that

$$q \geq rQ,$$
$$Q \leq \bar{Q}.$$

Imposing the first constraint, relaxing the second and optimizing over Q, we find

$$Q^*(i) = \frac{\tau + (p_w^x - p_w^z) + (p_w^z - (\omega - bi))\,r}{r^2 c}.$$

Substituting for $i^* = \frac{\alpha b r \bar{Q}}{2}$, we obtain

$$Q^* = \frac{2\left(\tau + (p_w^x - p_w^z) + r(p_w^z - \omega)\right)}{r^2(2c - \alpha b^2)}.$$

It is easy to show that Q^* is decreasing in r and ω. Hence, the condition $Q^* \geq \bar{Q}$ holds more easily if r is low or if ω is low.

To guarantee that $Q^* \geq \bar{Q}$ for any r, we need it to hold for $r = 1$. This implies

$$\tau + p_w^x - \omega \geq \frac{(2c - \alpha b^2)\bar{Q}}{2}.$$

This condition is easiest to meet with low ω, i.e., high-productivity suppliers. To guarantee further that $Q^* \geq \bar{Q}$ for any ω, we need it to hold for $\omega = p_w^z$. This implies

$$\tau + (p_w^x - p_w^z) \geq \frac{(2c - \alpha b^2)\bar{Q}}{2}.$$

This condition holds more easily if either τ or p_w^x are higher, and can hold for negligible τ as long as p_w^x is sufficiently high.