RIETI Discussion Paper Series 21-E-036

Wake Not a Sleeping Lion: Free Trade Agreements and Decision Rights in Multinationals

MUKUNOKI, Hiroshi
Gakushuin University

OKOSHI, Hirofumi
Okayama University

RIETI
Research Institute of Economy, Trade & Industry, IAA

The Research Institute of Economy, Trade and Industry
https://www.rieti.go.jp/en/

RIETI Discussion Paper Series 21-E-036
April 2021

Wake Not a Sleeping Lion: Free Trade Agreements and Decision Rights in Multinationals*

Hiroshi Mukunoki
Gakushuin University

Hirofumi Okoshi
Okayama University

Abstract

Free trade agreements (FTAs) with rules of origin (ROO) affect the location of input production for vertically integrated, multinational enterprises (MNEs). FTA-induced relocation changes the allocation of decision rights within MNEs and the purpose of transfer pricing from avoiding high taxes to strengthening their product-market competitiveness. This study shows that an FTA with ROO may hurt both MNEs and local firms despite tariff elimination, when the relocation occurs and the decision rights change from centralized to decentralized. Moreover, such an FTA can hurt consumers. Nevertheless, ROO increase the feasibility of FTAs thanks to larger tax revenues.

Keywords: Rules of origin; Free trade agreements; Transfer pricing; Input relocation
JEL classification: F15; F23; H26; L13; L23

*This study is conducted as a part of the Project "Economic Policy Issues in the Global Economy" undertaken at the Research Institute of Economy, Trade and Industry (RIETI). The authors are grateful for helpful comments and suggestions by Jay Pil Choi (Michigan State University), Taiji Furusawa (University of Tokyo), Andreas Haufler (University of Munich), Keisaku Higashida (Kwansei Gakuin University), Jung Hur (Sogang University), Jota Ishikawa (Hitotsubashi University), Yoshimasa Komoriya (Chuo University), Ngo Van Long (McGill University), Kiyoshi Matsubara (Nihon University), Martin Richardson (Australian National University), Eiichi Tomiura (Hitotsubashi University), Akihiko Yanase (Nagoya University) and Discussion Paper seminar participants at RIETI.

1 Introduction

For the last few decades, a proliferation of regional trade agreements (RTAs) has played a key role in liberalizing trade among countries.[1] As RTAs eliminate or reduce trade barriers, they are supposed to lower consumer prices and raise export prices, benefiting consumers and exporters. However, the preferential nature of RTAs may make their effects more complicated than they seem because of specific rules in implementing RTAs. Among others, setting rules of origin (ROO) is indispensable to form a free trade agreement (FTA) to avoid exports from third countries to FTA regions through a member country with the lowest external tariff, which is known as "trade deflection." To prevent trade deflection, ROO require firms to prove that the exported products originated within the FTA.[2]

Although ROO are aimed at preventing trade deflection, Felbermayr et al. (2019) show that most trade deflection is not profitable even without ROO because of small differences in external tariffs and non-negligible transportation costs. Nevertheless, ROO affect exporting firms' strategies such as their input procurement or location choices. For example, Conconi et al. (2018) concluded that the ROO of the North American Free Trade Agreement (NAFTA) reduced imports from non-member countries, which indicates that ROO cause inefficiency in input procurement.

Such a change in global procurement patterns can create another negative impact on multinational enterprises (MNEs) from the viewpoint of tax avoidance. If an MNE procures inputs from its related company outside an FTA region, the MNE freely determines the price of intra-firm trade, which makes it possible that the related entity saves overall corporate tax payments by shifting profits via price manipulation on the traded products. This price manipulation is called transfer pricing and the manipulated price is called the transfer price.[3] As some MNEs relocate their input production in the same country inside the FTA region to satisfy ROO, they lose the opportunity of saving overall tax payments. Thus, FTA formation with ROO can prevent MNEs from tax avoidance.

The change in MNEs' input procurement can also negatively influence local firms inside the FTA. After vertically-integrated MNEs locate their upstream affiliate inside the FTA, they can del-

[1] As of March 2021, 343 RTAs are in force. See `http://rtais.wto.org/UI/PublicAllRTAList.aspx`. Many articles have investigated them both theoretically and empirically. See Freund and Ornelas (2010) for a review of the literature on RTAs.

[2] Unlike a custom union, member countries of an FTA independently set their external tariffs against non-member countries. If the external tariffs are different for the same product, firms producing outside the FTA can save tariff payments by exporting a product to the member country whose external tariff is lowest and then re-exporting it to other FTA member countries whose external tariffs are higher. For instance, Stoyanov (2012) empirically examined firms' incentive to transship goods through FTA members. To forestall firms from tariff avoidance, member countries of FTAs stipulate ROO.

[3] Some empirical research has provided evidence of transfer pricing to save tax payments. For instance, Cristea and Nguyen (2016), Davies et al. (2018), and Liu et al. (2020) provide empirical evidence of transfer price manipulation. Blouin et al. (2018) found conflicting motives of transfer price when MNEs use it for corporate tax saving and also for tariff saving.

egate decision rights to their downstream affiliates and strategically use their internal prices on inputs to make their downstream affiliates more aggressive in the product market. Existing studies have suggested the role of internal input pricing to make the behavior of managers of downstream affiliates aggressive and shift profits from rival firms.[4] For example, Fershtman and Judd (1987) and Sklivas (1987) showed that firm owners have incentives to give their managers decision-making rights in the product market. It has also been reported that some MNEs conduct transfer pricing for managerial use. For instance, Czechowicz et al. (1982) reported that 89% of U.S. MNEs manipulated a transfer price for both tax avoidance and strategic motives. Ernst & Young (2003) also reported that over 80% of their sample firms use a single transfer price for these two purposes. As MNEs can use their internal prices solely for managerial purposes after they locate both an upstream and a downstream entity in the same country, the output decision becomes more aggressive. These changes may hurt other local firms in the country.

Although the above argument seems realistic, no studies have considered all the above factors simultaneously. Exploring MNEs' location decisions, their delegation decisions, and FTA formations is an essential step for both researchers and policymakers to understand the impacts of trade liberalization through FTAs.

1.1 Preview of the model and the results

Based on the hypothesis argued above, this study builds an international duopoly model to investigate how an MNE responds to FTA formation with ROO when it manipulates transfer pricing and how the welfare effect of the FTA formation depends on the MNE's response. The MNE produces a final good within an FTA member country and exports the good to another FTA member country. The MNE's location for input production is either in the final-good-production country or in a low-tax country outside the FTA. The MNE can shift profits across countries by manipulating the transfer price if it locates its input production outside the FTA countries. The MNE competes with the local firm, which also produces the final good within the FTA and exports its product to the same member country.

In the absence of ROO, the MNE prefers to locate its upstream and downstream affiliates in different countries when the tax gap is large. This is because a larger tax differential increases the MNE's gains from tax savings. The decision-making of the two affiliates is centralized in this case because a high transfer price to save tax payments discourages the downstream affiliate if the MNE chooses decentralization. In this case, the MNE (regarded as a "Lion") is sleeping in the sense that

[4]See Gök and Schiller (2006) for a survey of transfer pricing manipulation for strategic motives.

it is less aggressive in the product market. When the tax gap is small, however, the MNE prefers to locate both affiliates in the same country, and the decision-making of the downstream affiliate is decentralized for managerial purposes. Then, the upstream affiliate sets a low input price to make the downstream affiliate more competitive in the product market.[5] In this case, the MNE as the Lion is waking in the sense that it is aggressive in the product market and uses its transfer price to "attack" the rival firm.

We show that the formation of an FTA can induce the relocation of the MNE's input production from a country outside the FTA to a country inside the FTA. A notable result is that FTA formation with the MNE's input relocation may hurt the local firm even though the local firm benefits from tariff elimination as part of the FTA. This is because the loss from the strategic effect of transfer pricing outweighs the gain from tariff elimination for the local firm. Thus, an FTA hurts the local firm despite tariff elimination because it awakens the Lion.

In the presence of ROO, the MNE is not always qualified for tariff-free exports and chooses between the following two options: (i) producing inputs inside an FTA to comply with the ROO and enjoying tariff elimination, or (ii) producing inputs outside an FTA and manipulating the transfer price to save tax payments. If the MNE produces inputs in the same country where its downstream affiliate is located, the MNE is able to comply with the ROO and exports its product without tariff, though it cannot save tax payments by transfer pricing. Instead, it delegates the output decision to its downstream affiliate and adjusts the transfer price to shift rents in the product market. Alternatively, if the MNE locates its upstream affiliate in a country outside the FTA, the MNE saves corporate tax payments by setting a high transfer price with tariff. Therefore, this model exhibits the MNE's choice of "tariff elimination versus tax avoidance" via its input procurement strategies.

Unlike the case without ROO, FTA formation can reduce the profits of both the MNE and the local firm, even though both of them comply with ROO and make tariff-free exports within the FTA. This happens when the MNE relocates its upstream production to a country within the FTA. The MNE's (post-tax) profit decreases because it no longer saves tax payments, whereas the local firm's profit decreases because the strategic effect of low intra-firm input price of the MNE intensifies market competition. This exporter-hurting FTA occurs when the tariff is low because the direct gains from elimination of the tariff are small. An FTA with ROO hurts the awakened Lion itself in addition to the local firm, but the MNE's profit is still larger than the profit when it is sleeping.

We also show that an FTA with ROO can decrease total exports and hurt consumers in the

[5]This is in line with Schjelderup and Sorgard (1997), who showed that, when the decisions of the headquarters and foreign affiliates are decentralized and those affiliates compete with rival firms in the product market, MNEs may use their transfer price as a strategic tool to shift rents from rival firms.

importing country, even if all firms comply with ROO and make tariff-free exports. When the tax gap is large, the MNE has an incentive to make a large amount of exports to increase the value of intra-firm transaction and save its tax payments. After the FTA formation, if the MNE changes the location of input production from offshoring to inshoring to comply with ROO, it no longer makes large exports to avoid a high tax. Although the inshoring provokes a strategic delegation of the decision rights and has an effect of increasing exports, it decreases exports and hurts consumers if the former effect dominates the latter. The negative trade effect of RTAs are found in some empirical studies. For instance, Cipollina and Salvatici (2010) conducted a meta-analysis on the coefficients of RTA dummy variables in gravity analyses, which reported that the estimated coefficients are even negative in 312 out of 1827 studies.

Furthermore, we also identify the condition under which ROO improve the total welfare of countries inside the FTA. In the absence of ROO, FTA formation can worsen the welfare of FTA countries. This is because the MNE shifts more of its tax base to an outside FTA country, and the host country collects a small or no tax revenue from the MNE. In this situation, ROO can transform a welfare-worsening FTA into a welfare-improving one because the host country can collect tax revenue from the MNE if it induces input relocation. Besides that, even if ROO do not induce input relocation, they can still improve the total welfare of the FTA countries when the MNE does not comply with the ROO. Because tariff is not eliminated for the MNE in this case, the profit of the local firm increases and the countries inside the FTA can collect tariff revenues from the MNE.

1.2 Relationship to the literature

The welfare effects of FTAs with ROO have been previously analyzed by some articles, but their focuses are mainly on intermediate goods markets. Krishna and Krueger (1995) showed that ROO may work as a hidden protection against the input suppliers outside the FTA. Ju and Krishna (2005) showed that ROO increase the price of FTA-made inputs and reduce the total output if the ROO are not overly stringent such that all firms comply with them. However, ROO have the opposite effects if they are sufficiently stringent such that some firms choose not to comply with them. In Ju and Krishna (2005), the price of the output is fixed, and they did not consider how ROO affect consumers. Demidova and Krishna (2008) extended Ju and Krishna (2005) to include heterogeneity in productivity among final-good producers and showed that productivity sorting ensures a negative relationship between the stringency of ROO and the demand for FTA-made inputs. Ishikawa et al. (2007) focused on final good markets and showed that ROO have a role

in segmenting markets within the FTA; both inside and outside firms producing final goods may benefit from ROO at the cost of consumers. Mukunoki (2017) showed that FTAs with ROO may be consumer-hurting if they change the outside firms' location decisions. Mukunoki and Okoshi (2021) investigated a firm's manipulation of the output price for complying with ROO. None of these articles, however, have considered the effects of FTA formation when an MNE manipulates its transfer price. The most closely related article is Mukunoki and Okoshi (2020), which explores the impact of a value-added criterion of ROO on a monopolistic MNE's transfer pricing. However, the focus of this article is different in the sense that it considers strategic interactions between firms and explores how FTA formation affects the MNE's delegation of decision rights to its downstream affiliate.

Another strand of literature related to this article is the analyses on the use of transfer pricing for managerial purposes. For instance, Elitzur and Mintz (1996) derived the optimal transfer price when it is used for both saving tax payments and increasing the effort level of the local manager of a foreign affiliate. Schjelderup and Sorgard (1997) considered a decentralized MNE and calculated the optimal transfer pricing when there is a trade-off between saving tax payments and shifting rents from rival firms. Hyde and Choe (2005) also considered two international tax schemes and focused on the use of two books for transfer pricing, namely an MNE setting two internal prices: one for saving tax payments and the other for providing appropriate incentives to local managers. Some articles considered an MNE's decision between centralization and decentralization. For instance, Nielsen et al. (2008) showed that an MNE chooses centralization when the tax gap between countries is large but chooses decentralization when it is small. Dürr and Gök (2011) showed that using one transfer price for both tax and managerial purposes, rather than two transfer prices for each purpose, may increase an MNE's profit. Our article also considers the endogenous choice of an MNE between centralization and decentralization, and the decision is linked to the formation of an FTA with ROO and the MNE's location choice.

This article also contributes to the literature on the link between tax avoidance and location choice. Peralta et al. (2006) and Stöwhase (2013) considered tax competition between two (potential) host countries. Kato and Okoshi (2019) also investigated the impact of transfer price regulation on the location of an upstream affiliate. Our article is different from these articles in that we consider the impact of FTA formation with ROO on the location choice of an MNE.

The rest of the article is organized as follows. Section 2 sets up a model. Section 3 derives the equilibrium without ROO and the equilibrium with ROO separately, and compares them. Section

4 investigates the welfare effects of FTA formation. Section 5 discusses the determination of the decision rights and the robustness of the main results by relaxing some key assumptions. Section 6 summarizes and concludes the article.

2 Model

We consider a three-country model with two firms, an MNE (firm M) and a local firm (firm L). The model is illustrated in Figure 1. Two of the three countries are potential FTA member countries, whereas the rest is the outside, non-member country (country O). The headquarters of the MNE is located in country O and is owned by residents in country O. The MNE has a downstream affiliate (firm M_D) in one of the member countries to supply final goods to consumers in the FTA region.[6] We assume the two downstream firms, firm M_D and firm L, are located in one of the member countries, which is referred to as the host country (country H). Country H is chosen because it has location advantages to attract firms, such as low factor prices, a large pool of skilled workers, and so on. The two downstream firms produce homogeneous goods and serve them to consumers in another member country (country F). Countries H and F are potential members of an FTA. To simplify the analysis, the baseline model does not consider the output market in country H. As is discussed in section 5.3, this assumption does not qualitatively change our main results.

The representative consumer's utility in country F is given by $U = a(x_L + x_M) - \frac{(x_L+x_M)^2}{2}$, where x_i is the consumption of the final good produced by firm i ($i \in \{L, M\}$). By utility maximization, the inverse demand function is given by $p = a - (x_L + x_M)$, where p is the price of the final good.

As policy instruments, we consider both a corporate tax and an import tariff. The governments in countries O and H respectively impose t and T as a corporate tax on reported profits.[7,8] Hereafter, we focus on the case of $T \geq t$, with which our main findings are obtained.[9]

In addition, country F imposes a specific tariff, τ, on imports of the final good. An FTA between countries H and F eliminates this tariff. In order to focus on the impact of FTA formation on competition in the final-goods market, tariffs on inputs are assumed away. We focus on the case where

[6]This type of foreign direct investment (FDI) is known as export-platform FDI, by which FDI firms export their products from the host country to other countries. For example, see Tekin-Koru and Waldkirch (2010) for evidence from Mexico of an increasing role of export-platform FDI. Tintelnot (2017) also shows that the share of U.S. MNEs' outputs exported to countries outside the host country increases. For instance, the share in Belgium was 63% in 2004, which was the third highest share.

[7]Note that we use the terms "tax rate" and "tax revenue" to represent the corporate tax rate and corporate tax revenue, respectively. The tax rate and tax revenue are distinguished from tariff rate and tariff revenue.

[8]In this model, we postulate that both governments in countries O and H adopt a territorial tax system rather than a worldwide one. After the U.S. moved from a worldwide tax system to a territorial tax system, most OECD countries have adopted a territorial tax system.

[9]This situation is consistent with real-world observations. For instance, Mexico and Belgium have higher corporate taxes than other countries, and these countries are major host countries of export-platform FDIs. See also footnote 6.

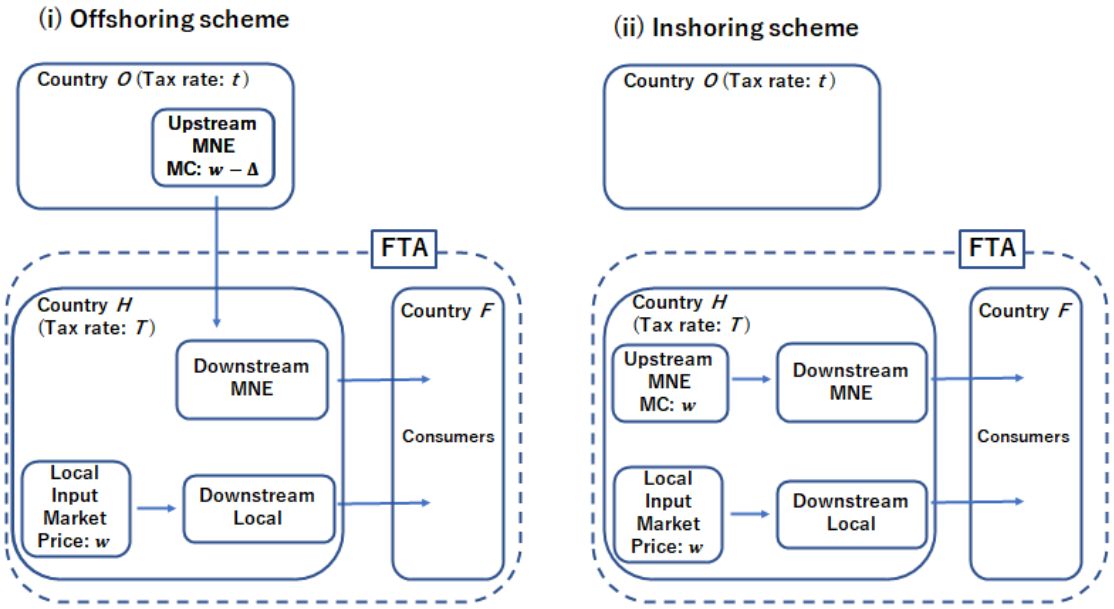

Figure 1: Model

both firms always supply their products in country F.[10] In the presence of ROO, the downstream firms need to meet the ROO to be eligible for the non-application of τ. We assume that the inputs imported from country O cannot satisfy any criteria of ROO.[11] Therefore, the only way to meet the ROO is to produce inputs in the FTA countries or procure inputs from the local markets in these countries.[12]

The downstream firms use the same production technology, where one unit of inputs is transformed into one unit of final products. Other production costs are constant and normalized to zero.[13] Firm L is always eligible for the FTA tariff because it always procures inputs with the input price w from the perfectly competitive input market in country H.

Meanwhile, firm M produces inputs by itself that are to be used for the production of firm M_D. It establishes an upstream affiliate (firm M_U), either in country H or O. Firm M_U produces inputs more efficiently in country O than in H because country O has location advantage of input production. Specifically, if firm M_U produces inputs in country H, its marginal cost is given by w. If inputs are produced in country O, firm M_U's marginal cost is given by $w - \Delta$. This implies that locating firm M_U in country O gives firm M not only a cost advantage over local input suppliers but

[10]The detailed conditions will be shown in Section 3.

[11]For instance, the ROO of the US-Mexico-Canada Agreement (USMCA) requires exporters of automobiles to produce some core inputs, such as engines and shafts, inside USMCA.

[12]If a value-added criterion of ROO is employed, the multinational firm has an option to adjust the transfer price to meet the ROO without changing the location of its input procurement. This possibility is analyzed in another article of ours, Mukunoki and Okoshi (2020).

[13]We assume that only the MNE has an option to procure inputs from country O, even though both firms share the same production technology. Empirical evidence, such as Tomiura (2007), suggests that some firms engage in global production such as outsourcing and FDI whereas the others do not, even if they have similar productivity.

	Scheme	λ_M^s	λ_L	Variable v
No FTA	Inshoring	1	1	"$*$" (v^{s*})
	Offshoring	1	1	
FTA w/o ROO	Inshoring	0	0	"$\widehat{\ }$" ($\widehat{v}^s$)
	Offshoring	0	0	
FTA w/ ROO	Inshoring	0	0	"$\widetilde{\ }$" ($\widetilde{v}^s$)
	Offshoring	1	0	

Table 1: Schemes considered in the model

also a tax saving opportunity. In this case, firm M_U exports the produced inputs to its downstream affiliate by charging an intra-firm transfer price denoted by r^O. However, firm M cannot meet the requirement of ROO, and it does not enjoy tariff-free access to the market even after the formation of an FTA.[14] We call this case scheme O (Offshoring).

To utilize the FTA tariff, firm M must comply with ROO by procuring inputs in the host country. We call this case scheme I (Inshoring). Let λ_M^s denote a state variable that takes zero if firm M is qualified for zero tariff and takes unity otherwise. The same rule is applied to λ_L for firm L. To distinguish the equilibrium variables in the three regimes, we use asterisk "$*$" for pre-FTA equilibrium, hat "$\widehat{\ }$" for the post-FTA equilibrium without ROO, and tilde "$\widetilde{\ }$" as a circumflex for the post-FTA equilibrium with ROO. These are summarized in table 1.

The MNE chooses its organization structure, and it depends on the location of the upstream affiliate. When the decision making is decentralized, firm M_D becomes more aggressive in the product market if the input cost, that is, the level of the transfer price, is lower. This implies that the MNE can shift rents from the local firm by lowering the transfer price, r^I. This strategic motive of transfer pricing is the reason why the MNE always chooses decentralization when the two affiliates are located in the same country. In this case, firm M_D sets the quantity to maximize its own profits, although the objective of the upstream affiliate is to maximize the total profits.

Alternatively, when the locations of the two affiliates are separate, the MNE may prefer centralized decisions to the decentralized decisions under $T \geq t$ because of the tax-avoidance motive of transfer pricing. To avoid the high tax of the host country, the MNE sets a transfer price that is higher than the marginal cost of input production. The MNE chooses centralization in most cases, so as not to discourage the downstream firm's decision in quantity setting.[15] Since our main results are obtained when an FTA formation changes the MNE's allocation of decision rights from

[14]Some empirical evidence shows that not all firms can use FTA tariffs because of the existence of ROO, which means the impacts of FTA formation are heterogeneous across firms. See, for example, Takahashi and Urata (2010) and Hayakawa et al. (2013).

[15]Nielsen et al. (2008) demonstrates that centralization is more profitable than decentralization when $T \geq t$ and the tax gap is large. When the tax gap is small, decentralization realizes higher profits for the MNE.

centralization to decentralization, hereafter we focus on the case where the MNE chooses centralization when it locates the upstream affiliate in country O. In Section 5.1, we discuss how the MNE decides the allocation of the decision rights. Thus, the MNE faces a trade-off in the location of its input production: the MNE is able to use the transfer price to save tax payments in producing inputs in country O, whereas it can use the transfer price to take advantage of a strategic effect of decentralization in producing them in country H.

We solve the following three-stage game. In the first stage, the headquarters of the MNE decides the location of firm M_U. In the second stage, the headquarters determines the optimal input price. In the third stage, the MNE and firm L compete à la Cournot in country F.

3 The equilibrium

This section derives the equilibrium outcomes in each scheme and investigates how the MNE chooses the location of its input production.

3.1 Market equilibrium

Let us first derive the market equilibrium determined in the last stage. The unit cost of the local firm, firm L, in producing a final good and exporting it to country F is given by $c_L = w + \lambda_L \tau$. Then, firm L maximizes the following (pre-tax) profit,

$$\pi_L = (p - c_L)x_L. \tag{1}$$

As described in the previous section, there are two schemes for the MNE's input sourcing: (i) an offshoring scheme in which the MNE produces inputs in country O and (ii) an inshoring scheme in which the MNE produces inputs in country H. Below, we subsequently derive the equilibrium in each scheme.

Offshoring scheme When the MNE locates its input production in country O, the MNE centralizes its decision-making and determines the amount of supply to maximize the following global post-tax profit:

$$\begin{aligned}\Pi_M^O &= (1-t)(r^O - (w-\Delta))x_M^O + (1-T)(p - r^O - \lambda_M \tau)x_M^O \\ &= (1-T)(p - c_M^O)x_M^O,\end{aligned} \tag{2}$$

where

$$c_M^O = \frac{(1-t)(w-\Delta) + (1-T)\lambda_M^O \tau - (T-t)r^O}{1-T} \tag{3}$$

is the *perceived marginal cost*. The centralized MNE's decision in the product market is based on the perceived marginal cost, which is different from the sum of the input production cost and trade cost, $w - \Delta + \lambda_M^O \tau$.[16] In this cross-border production, the MNE's unit cost is adjusted by the tax differential. As a marginal increase in the transfer price, r^O, saves per-unit tax payments as much as $(T-t) > 0$, it reduces the effective marginal cost of firm M_D in the production of the final good. This is because the per-unit, post-tax profit is larger with tax avoidance, and it gives the MNE an incentive to increase the quantity of sales. Therefore, the "perceived marginal cost" becomes lower, and the MNE supplies more as r^O becomes higher. Note that the perceived marginal cost under an offshoring scheme c_M^O is equivalent to $w - \Delta + \lambda_M^O \tau$ only if $T = t$ holds and is decreasing in T and increasing in t. This means that the perceived marginal cost is less than the true marginal cost, $c_M^O \leq w - \Delta + \lambda_M^O \tau$, when $T > t$ holds. In other words, transfer pricing makes the MNE more aggressive in the product market under an offshoring scheme.

Inshoring scheme When the MNE locates its input production in country H, the MNE's decision-making is decentralized. Thus, how much to produce of the final good is delegated to the manager of the downstream affiliate (i.e., firm M_D), who only takes into account the profit of firm D, which is given by

$$\pi_D = \{p - \underbrace{(r^I + \lambda_M^I \tau)}_{c_M^I}\} x_M. \tag{4}$$

For expositional convenience, we denote c_M^I as the perceived marginal cost under the inshoring scheme. Firm M_D's decision is based on c_M^I, whereas the true marginal cost of the MNE is $w + \lambda_M \tau$.

By maximizing (1), (2), and (4) with respect to each firm's quantity, we have the equilibrium outputs of the firms as:

$$x_M^s = \frac{a - 2c_M^s + c_L}{3}, \quad \text{and} \quad x_L^s = \frac{a - 2c_L + c_M^s}{3}, \qquad s \in \{O, I\}. \tag{5}$$

3.2 Manipulation of the transfer price

Next, we consider how the MNE sets the transfer price in the second stage. As described above, depending on the MNE's location choices in input production, there are two motives for which the

[16] The terminology "perceived marginal cost" is often used in the analysis of a vertically-related industry in the context of industrial organization. See Choi et al. (2020) for an application of this terminology in tax avoidance literature.

MNE manipulates the transfer price. We derive the optimal transfer prices separately in these cases.

Offshoring scheme Given (2) and (5), the first derivative of Π_M^O with respect to r^O is always positive. Therefore, the optimal transfer price is set as high as possible. This means that $p - r^O - \lambda_M^O \tau = 0$ holds.[17] We have

$$r^O = w - \Delta + \underbrace{\frac{(1-T)\{a - w + 2\Delta - (2\lambda_M^O - \lambda_L)\tau\}}{1 - t + 2(1-T)}}_{\text{Tax avoidance motive}}. \tag{6}$$

By substituting (6) into (3), the perceived marginal cost that reflects the equilibrium transfer price becomes

$$c_M^O = w - \Delta + \lambda_M \tau - \frac{(T-t)\{a - w + 2\Delta - (2\lambda_M^O - \lambda_L)\tau\}}{1 - t + 2(1-T)}, \tag{7}$$

which is decreasing in T. The corresponding equilibrium output and profits, respectively, become

$$x_M^O = \frac{(1-t)(a - w + 2\Delta - (2\lambda_M^O - \lambda_L)\tau)}{1 - t + 2(1-T)} \quad \text{and} \quad \Pi_M^O = (1-T)(x_M^O)^2. \tag{8}$$

x_M^O is the smallest when $\lambda_M^I = 1$ and $\lambda_L = 0$, and it is positive if and only if $\tau < \frac{a-w+2\Delta}{2}$. The equilibrium output and profits of the local firm, respectively, become

$$x_L^O = \frac{(1-T)(a - w - \lambda_L \tau) - (1-t)\{\Delta + (\lambda_L - \lambda_M^O)\}\tau}{1 - t + 2(1-T)} \quad \text{and} \quad \Pi_L^O = (1-T)(x_L^O)^2. \tag{9}$$

In the offshoring scheme, x_L^O is smaller in the pre-FTA equilibrium ($\lambda_M^I = \lambda_L = 1$). It is positive if and only if $\tau < a - w - \left(\frac{1-t}{1-T}\right)\Delta$.

The optimal transfer price is higher than the marginal cost of producing inputs. The transfer price is set to shift profits from a high-tax country H to a low-tax country O. The second term of (6) represents a *tax-avoidance motive*, whose sign is always positive. Moreover, (8) indicates that the MNE's output expands as the corporate tax in country H, T, is higher. This is because the perceived marginal cost is decreasing in T. As the induced increase in the output lowers the equilibrium price of the final good, p, the transfer price that realizes zero profits of firm M_D also becomes lower.

Inshoring scheme Given (5), the overall profit of the MNE is

$$\Pi_M^I = (1-T)\left[\{r^I - w + p - (r^I + \lambda_M^I \tau)\}\left(\frac{a + w - 2r^I - (2\lambda_M^I - \lambda_L)\tau}{3}\right)\right]. \tag{10}$$

[17] We assume away an additional cost for manipulating the transfer price. We relax this assumption by introducing a concealment cost in section 5.4.

By differentiating (10) with respect to r^I, the optimal transfer price becomes

$$r^I = w \underbrace{- \frac{a - w - (2\lambda_M^I - \lambda_L)\tau}{4}}_{\text{Strategic motive}}. \tag{11}$$

The corresponding equilibrium output and profits of the MNE are given by

$$x_M^I = \frac{a - w - (2\lambda_M^I - \lambda_L)\tau}{2}, \text{ and } \Pi_M^I = \frac{(1-T)}{2}(x_M^I)^2. \tag{12}$$

The equilibrium output and profits of the local firm are given by

$$x_L^I = \frac{a - w - (3\lambda_L - 2\lambda_M^I)\tau}{4}, \text{ and } \Pi_L^I = (1-T)(x_L^I)^2. \tag{13}$$

We have either $\lambda_M^I = \lambda_L = 1$ or $\lambda_M^I = \lambda_L = 0$ in the inshoring scheme, and the equilibrium outputs are smaller in the former case. Therefore, x_M^I and x_L^I are positive if and only if $\tau < a - w$ holds.

In the inshoring scheme, the production and export decisions in the final-goods market are delegated to firm M_D. Then, the MNE uses the transfer price to make firm M_D behave more aggressively in the product market by setting a low transfer price. In other words, lowering the transfer price works as a "strategic intra-firm subsidy" and shifts rents from firm L to firm M. This is captured by the second term in (11), which is always negative. Unlike the offshoring scheme, the optimal transfer price under inshoring is independent of T.

3.3 Location choice of the input production

In the first stage, the MNE chooses between country O or H for the location of firm M_U. We focus on the situation where $\tau < \min\left[\frac{a-w+2\Delta}{2}, a - w - \left(\frac{1-t}{1-T}\right)\Delta\right]$ holds, such that exports of both firms are positive. Rearranging this inequality, we have the maximum level of T, $T^{max} \equiv 1 - \frac{(1-t)\Delta}{a-w-\tau} (< 1)$, below which the equilibrium exports of both firms are positive given $\tau < \frac{a-w+2\Delta}{2}$.

By comparing the profits between the two schemes, we have

$$\Pi_M^O \geq \Pi_M^I \iff T \geq \frac{3-t}{2} - (1-t)\left(\frac{a - w + 2\Delta - (2\lambda_M^O - \lambda_L)\tau}{a - w - (2\lambda_M^I - \lambda_L)\tau}\right)\sqrt{2}, \tag{14}$$

The MNE chooses the offshoring scheme and shifts profits for the tax-avoidance motive when the corporate tax in country H is sufficiently high. Otherwise, it chooses the inshoring scheme and enjoys the strategic effect of transfer pricing. Specifically, the MNE prefers the offshoring scheme

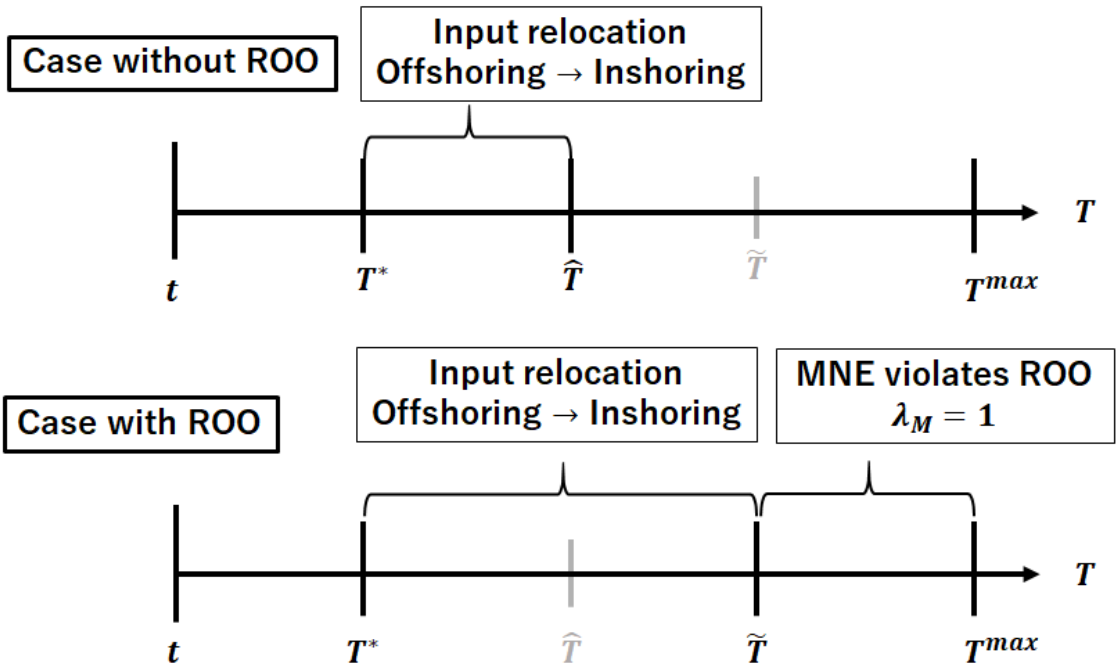

Figure 2: The MNE's input production

to the inshoring scheme if and only if $T > T^* \equiv \frac{3-t}{2} - \frac{(1-t)(a-w+2\Delta-\tau)\sqrt{2}}{a-w-\tau}$ holds before formation of an FTA, $T > \widehat{T} \equiv \frac{3-t}{2} - \frac{(1-t)(a-w+2\Delta)\sqrt{2}}{a-w}$ holds after formation of an FTA without ROO, and $\widetilde{T} \equiv \frac{3-t}{2} - \frac{(1-t)\{a-w+2(\Delta-\tau)\}\sqrt{2}}{a-w}$ holds after formation of an FTA with ROO. We can easily confirm that $T^* < \widehat{T} < \widetilde{T}$ hold.

The elimination of the tariff increases the equilibrium sales of the MNE, magnifying the strategic motive of transfer pricing. If $T > \widehat{T}$ holds, however, the tax-avoidance motive still dominates the strategic motive even after FTA formation. Thus, by formation of an FTA without ROO, the MNE that initially chooses the offshoring scheme ($T^* < T$) changes its scheme to the inshoring scheme if $T^* < T < \widehat{T}$ holds.

The result also indicates that FTA formation causes efficiency loss, similar to the conventional, trade-diversion effect. The trade-diversion effect is caused by the substitution of imports from less efficient member countries for those from more efficient non-member countries. In our model, input production is relocated from a more efficient country whose production cost of the input is $w - \Delta$ to a less inefficient country whose production cost is w.

In the absence of ROO, the location of input production is unrelated to tariff elimination. In the presence of ROO, however, the MNE complies with the ROO, and the tariff on the final good is eliminated only if the MNE locates input production within FTA countries. Thus, firm M faces a trade-off between tax avoidance and tariff elimination. Because the MNE needs to incur tariff under

scheme O, the threshold of T is larger than that of the case without ROO, $\widehat{T} < \widetilde{T}$. This implies that ROO expand the range of T such that input relocation occurs. Moreover, as a higher tariff discourages the MNE to choose scheme O, $\frac{\partial \widetilde{T}}{\partial \tau} > 0$ holds.

The equilibrium choice of the MNE is depicted in Figure 2. We summarize the equilibrium outcomes in the following proposition.

Proposition 1. *FTA formation induces the MNE to relocate its input production from an outside country to an inside FTA country if $T^* < T < \widehat{T}$ holds in the absence of ROO and if $T^* < T < \widetilde{T}$ holds in the presence of ROO, where $\widehat{T} < \widetilde{T}$ holds. In the presence of ROO, the MNE produces inputs in an outside country and does not comply with ROO if $\widetilde{T} \leq T$ holds.*

The following two points are notable. First, input relocation induced by FTA formation is consistent with empirical results. For instance, Hanson et al. (2005) empirically showed a substantial increase in inward investments to Mexico and Canada after the formation of NAFTA. The increase in FDI to these countries might be caused by input relocation of MNEs, which leads to the magnified strategic effect of transfer pricing. Furthermore, as Conconi et al. (2018) showed, ROO lower the likelihood of input procurement from outside FTA countries. This corresponds to the case with $\widehat{T} \leq T \leq \widetilde{T}$. As the MNE's input production remains in country O in this range of T if ROO are absent, they may have a role to prevent the MNE's tax avoidance.

Second, as Takahashi and Urata (2010) and Hayakawa et al. (2013) empirically pointed out, some firms may not utilize FTA tariffs because of the burden of ROO, which corresponds to the case with $\widetilde{T} \leq T$. Although traditional interpretation attributes non-compliance to the costs of meeting ROO such as documentation costs and adjustment costs in input procurement, our model indicates that the loss of tax avoidance opportunities can be another reason why firms choose non-compliance with ROO.

4 The welfare effects of FTA formation

In this section, we subsequently explore the effects of FTA formation on the MNE, the local firm, consumers, and the total welfare of the FTA countries.

4.1 The effect on the MNE

In the absence of ROO, FTA formation always benefits the MNE. It is obvious that an FTA favors the MNE when it does not change the MNE's procurement strategy because only the gains from

tariff elimination exist. When the MNE relocates its input production, it loses an opportunity to save tax payments. Nevertheless, the gains from the magnified strategic effect exceed the loss of the tax saving opportunity in the absence of ROO.

In the presence of ROO, FTA formation can hurt the MNE. There are three cases, depending on the level of T. First, when the MNE chooses the inshoring scheme both before and after the formation of an FTA (i.e., $T < T^*$), the FTA always benefits the MNE because the MNE does not engage in tax avoidance and only the gains from tariff elimination exist. Second, when the MNE chooses the offshoring scheme both before and after the formation of an FTA (i.e., $\widetilde{T} < T$ holds), the FTA formation necessarily reduces the profits of the MNE because the MNE incurs the tariff whereas firm L takes advantage of tariff elimination.

Third, when FTA formation induces the MNE to shift from the offshoring scheme to the inshoring scheme (i.e., $T^* < T < \widetilde{T}$), it can either benefit or hurt the MNE. FTA formation with ROO decreases the MNE's profits if the tax difference is large enough. Specifically, there exists a unique threshold of T, T_M, such that $\widetilde{\Pi}_M^I = \Pi_M^{O*}$ holds, which is given by

$$T_M \equiv \frac{3-t}{2} - \frac{(1-t)(a-w+2\Delta-\tau)\sqrt{2}}{(a-w)}. \tag{15}$$

Note that $\widehat{T} < T_M < \widetilde{T}$ always holds, implying that only FTA formation with ROO can hurt the MNE.[18] As the corporate tax rate in country H gets higher, the loss from the missed opportunity of tax avoidance gets larger. Therefore, FTA formation with ROO hurts the MNE if $T > T_M$ holds and benefits the MNE otherwise. The following lemma summarizes the effects of FTA formation on the MNE's profits (see Appendix A.1 for the proof).

Lemma 1. *FTA formation without ROO always benefits the MNE. FTA formation with ROO hurts the MNE if the post-FTA MNE's choice is (i) not complying with ROO by choosing the offshoring scheme or (ii) complying with ROO by choosing the inshoring scheme and $T > T_M$ holds. Otherwise, it benefits the MNE.*

It is noteworthy that the MNE becomes worse off by FTA formation even if it complies with ROO and the tariff imposed on the MNE's product is eliminated. Nevertheless, the MNE prefers to comply with ROO because the tariff is also eliminated for the local firm, which is its rival in the product market. If the MNE chooses the offshoring scheme and does not comply with ROO in this situation, the MNE's profits further reduce. Therefore, the elimination of tariff on the rival firm forces the MNE to give up using transfer pricing for tax purposes, and the negative effect from paying more corporate taxes outweighs the positive effect from tariff elimination.

[18] Because $\widehat{T} < T_M$ holds, FTA formation without ROO never hurts the MNE even if the FTA induces input relocation.

4.2 The effect on the local firm

When an FTA without ROO does not affect the MNE's procurement strategy (i.e., $T \leq T^*$ or $T \geq \widehat{T}$), it benefits the local firm. When it changes the MNE's location of input production (i.e., $T^* < T < \widehat{T}$), however, firm L can be negatively influenced by the FTA. This is because input relocation makes the MNE more aggressive in the product market.

The equilibrium profit of firm L is increasing in its equilibrium output. By comparing the post-FTA output ($\widehat{x}_L^I$) with the pre-FTA output (x_L^{O*}) of firm L, we have

$$\widehat{x}_L^I \leq x_L^{O*} \iff T \leq T_L \equiv 1 - \frac{(1-t)(a-w+4\Delta)}{2(a-w-2\tau)}. \tag{16}$$

We can easily confirm that $\frac{\partial T_L}{\partial \tau} < 0$ holds. As the tariff becomes lower, the positive effect from the tariff elimination becomes lower, and FTA formation that induces input relocation is more likely to hurt firm L. If the tariff is sufficiently small such that $\widehat{T} < T_L$ holds without ROO and $\widetilde{T} < T_L$ holds with ROO, an FTA that induces input relocation always hurts firm L because the negative effect from the intensified competition dominates the positive effect from the tariff elimination. On the other extreme, if the tariff is high enough such that $T_L \leq T^*$ holds, an FTA always benefits firm L irrespective of the MNE's input relocation.

In the intermediate tariff level, it depends on the level of T whether an FTA inducing input relocation benefits or hurts the local firm. If $T^* < T_L \leq \widehat{T}$ holds in the absence of ROO and $T^* < T_L \leq \widetilde{T}$ holds in the presence of ROO, an FTA hurts the local firm if $T^* < T < T_L$ holds.[19] Otherwise, it benefits the local firm. The following lemma summarizes the effects of FTA formation on the local firm (see Appendix A.2 for the proof).

Lemma 2. *FTA formation hurts the local firm if $T^* < T < \min[\widehat{T}, T_L]$ holds without ROO and if $T^* < T < \min[\widetilde{T}, T_L]$ holds with ROO. Otherwise, it benefits the local firm.*

It is counterintuitive that ROO may hurt the local firm because they only restrict the MNE's actions. It contrasts with previous arguments, such as those of Krishna and Krueger (1995), who state that ROO work as a "hidden protection" policy for both the domestic upstream and downstream industries. In our model, although ROO induce the MNE to procure inputs inside an FTA country where the production cost of inputs is higher, they also provoke the MNE to delegate its decision right of final good production to the downstream affiliate. Then, the MNE uses its transfer price as a commitment device to make the downstream MNE more aggressive in the product market. In other

[19] We can derive the two cutoff levels of the tariff, $\widehat{\tau}_L$ and τ_L^*, such that $T^* < T_L \leq \widehat{T}$ holds for $\widehat{\tau}_L \leq \tau < \tau^*$. Similarly, we can derive $\widetilde{\tau}_L$ such that $T^* < T_L \leq \widetilde{T}$ holds for $\widetilde{\tau}_L \leq \tau < \tau^*$.

words, our model indicates that ROO magnify the pro-competitive effect of an FTA by changing the MNE's allocation of the decision rights from centralization to decentralization. Note also that the tax difference is a key to this result. If the tax difference is small, the MNE chooses the inshoring in the pre-FTA equilibrium and the change from centralization to decentralization does not occur.

Another intriguing result is that FTA formation with ROO may hurt the MNE and firm L at the same time, even though both comply with ROO and qualify for zero-tariff exports. As explained, an FTA with ROO hurts the MNE if $T > T_M$ holds, whereas it hurts the local firm if $T^* < T < \min[T_L, \widehat{T}]$ holds. Thus, an FTA with ROO hurts both firms at the same time if $T_M < T < \min[T_L, \widehat{T}]$ holds. Because an increase in τ increases T_M and decreases T_L, we have $T_M < T_L$ for a sufficiently low level of the initial tariff. Specifically, by (15) and (16), we have

$$T_L > T_M \iff \tau < \frac{(2-\sqrt{2})(a-w)}{4}. \tag{17}$$

Figure 3 illustrates the effects of FTA formation with ROO on the firms' post-tax profits when $\tau < \frac{(2-\sqrt{2})(a-w)}{4}$ holds. The red curves represent the equilibrium profits of the MNE whereas the black curves represent those of the local firm. The solid curves depict the post-FTA profits whereas the dashed curves depict the pre-FTA profits. We can confirm that an FTA with ROO hurts both firms when $T_M < T < T_L$ holds.

Proposition 2. *FTA formation with ROO hurts both the MNE and the local firm if* $\tau < \frac{(a-w)\left(2-\sqrt{2}\right)}{4}$ *and* $T_M < T < T_L$ *hold.*

This result is novel because existing studies on FTAs suggest that FTA formation benefits at least some exporting firms producing within the FTA. If we take into account the MNE's location choice and its manipulation of the transfer price, then FTA formation with ROO decreases the post-tax profits of all exporting firms, even though they comply with ROO and all tariffs are eliminated.

4.3 The effect on consumers

Here, we investigate the effect on consumers in the country where the final good is imported and consumed (country F). Because input relocation increases the MNE's production cost of the final good, it is not obvious whether FTA formation increases the total amount of exports to country F. We can confirm, however, that FTA formation without ROO always benefits consumers because the positive effect from the tariff elimination dominates the negative effect from the inefficiency in input production.

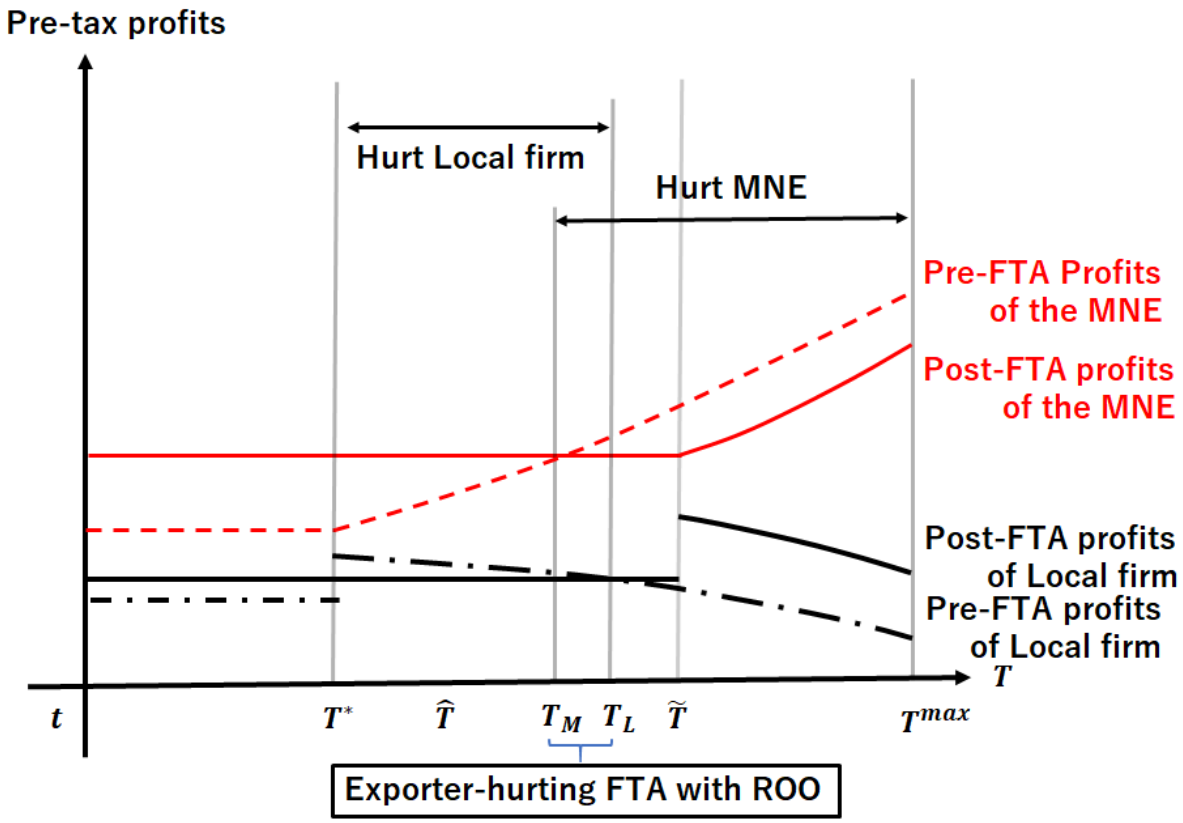

Figure 3: The effects of FTA formation with ROO on the firms' profits

Concerning FTA formation with ROO, the outcomes in the post-FTA equilibrium are the same as those without ROO when $T \leq \widehat{T}$ holds. Therefore, consumers gain from FTA formation in this range of T. For $\widehat{T} < T < \widetilde{T}$, it is ambiguous whether ROO increase or decrease the exports of the MNE. On the one hand, input relocation induced by ROO increases the MNE's marginal cost and has a negative impact on the volume of exports of the MNE. On the other hand, input relocation is accompanied by the MNE's decentralization decision, which makes the MNE more aggressive in the product market and has a positive impact on the volume of its exports. Note that the perceived marginal cost of the MNE is lower and the outputs of the MNE in the pre-FTA equilibrium are larger as the corporate tax in country H is higher. Larger total exports in the pre-FTA equilibrium implies that consumer surplus is higher, which increases the likelihood of a consumer-hurting FTA formation. Thus, we can derive a threshold denoted by $\widetilde{T}_{CS}$, such that FTA formation with ROO decreases total exports when $T > \widetilde{T}_{CS}$ holds. If the initial tariff is sufficiently low, we have $\widetilde{T} < \widetilde{T}_{CS}$ and FTA formation benefits consumers for all T in $T \in (\widehat{T}, \widetilde{T})$. We can derive the threshold of the initial tariff, $\widetilde{\tau}_{CS}$, such that $\widetilde{T} < \widetilde{T}_{CS}$ holds if $\tau < \widetilde{\tau}_{CS}$. Otherwise, $\widetilde{T} > \widetilde{T}_{CS}$ holds, and FTA formation with ROO reduces total exports and hurts consumers when $\widetilde{T}_{CS} < T < \widetilde{T}$ holds.

For $\widetilde{T} \leq T$, the MNE produces inputs in country O both before and after the formation of an

FTA with ROO. In this case, the MNE does not comply with ROO, and the tariff is applied only to the MNE. Although a part of the MNE's exports are replaced by exports of the local firm that is less efficient in output production, total exports always increase and the FTA benefits consumers. The following proposition summarizes the results (see Appendix A.3 for the proof).

Proposition 3. *FTA formation without ROO always benefits consumers in a foreign country. An FTA with ROO hurts consumers if* $\tau \geq \widetilde{\tau}_{CS}$ *and* $\widetilde{T}_{CS} < T < \widetilde{T}$ *hold. Otherwise, an FTA with ROO benefits consumers.*

Note that consumers may lose from FTA formation, even though all the exporters are eligible for tariff-free exports. The result is surprising because this happens when the initial tariff is large and the gains from tariff elimination are large. Owing to ROO, a larger tariff induces the MNE's input relocation. Input relocation increases the production cost of the MNE, because input production is less efficient in country H and the perceived marginal cost of the MNE in the pre-FTA situation is relatively large when the tax gap is in the middle range satisfying $\widetilde{T}_{CS} < T < \widetilde{T}$. Therefore, the negative effect of the increase in the marginal cost outweighs the positive effect from the tariff elimination in this case.[20]

4.4 The effect on the welfare of inside countries

We have shown that FTA formation may hurt exporters and consumers. FTA formation also changes both tariff revenues and tax revenues of member countries. Thus, we next explore whether these effects are consistent with countries' incentives to form an FTA.

We exclude the MNE's profits from welfare in country H because the MNE is owned by residents in country O. We denote the MNE's taxable profits in country H in scheme s ($s \in \{I, O\}$) as

$$\pi_M^s = \begin{cases} (p^O - r^O - \lambda_M^O \tau) x_M^O = 0, \\ (p^I - w - \lambda_M^I \tau) x_M^I = \frac{\{a - w - (2\lambda_M^I - \lambda_L)\tau\}^2}{8}. \end{cases} \tag{18}$$

Then, the equilibrium welfare of country H and that of country F in scheme s is respectively given by

$$W_H^s \equiv \pi_L^s + T\pi_M^s \quad \text{and} \quad W_F^s \equiv CS_F^s + TR_F^s, \tag{19}$$

where $CS_F^s \equiv \frac{\left(x_L^s + x_M^s\right)^2}{2}$ is the consumer surplus in country F and $TR_F^s \equiv \tau\left(\lambda_L x_L^s + \lambda_M^s x_M^s\right)$ is the tariff revenue in country F. We suppose that member countries are able to arrange transfers of

[20] We can confirm that an FTA with ROO never hurts consumers and all exporters at the same time. It hurts consumers when τ is sufficiently large but hurts exporters when τ is sufficiently small and these ranges of τ do not overlap. Thus, our model predicts that an FTA with ROO is more likely to hurt exporters as the initial tariff becomes lower and hurt consumers as it becomes higher.

welfare with the FTA formation, which can be done by making mutual concessions in other sectors, for instance. This means that formation of an FTA is feasible if it improves the total welfare of the member countries, which is given by $W^s_{FTA} \equiv W^s_H + W^s_F$.

We found that FTA formation without ROO can worsen the total welfare of the member countries. When $T < T^*$ holds, the MNE produces inputs in country H, regardless of FTA formation. In this case, although FTA formation increases the consumer surplus and the profit of the local firm, the loss of tariff revenues exceeds these benefits if $\tau < \frac{2(a-w)}{13}$ holds. This is because, under imperfect competition, a tariff has a strategic role to shift profits from the foreign firm to the welfare of the importing country as tariff revenue. The strategic effect tends to be relatively larger as the tariff becomes smaller. As the MNE pays a corporate tax in country H, the net negative effect on $\pi^I_L + CS^I_F + TR^I_F$ is covered by increases in tax payments, $T\pi^I_M$ when T is large. Thus, under $T < T^*$, FTA formation without ROO decreases the total welfare of the member countries if $\tau < \frac{2(a-w)}{13}$ and $T < T^I_W$ hold, where $T^I_W \equiv \frac{2(a-w)-13\tau}{4\{2(a-w)-\tau\}}$ is the threshold level of T. Note that T^I_W can be higher or lower than T^*. If $\tau < \frac{2(a-w)}{13}$ and $T^* \leq T^I_W$ hold, FTA formation without ROO always worsens the welfare of the FTA countries.

Another possible effect of a welfare-reducing FTA occurs when $\widehat{T} < T$ holds and the MNE produces inputs in country O, both before and after FTA formation. In this case, the MNE always shifts all the taxable profits in country H to country O, and a higher T increases the profits of the MNE due to a decrease in the perceived marginal cost. The resulting increase in the MNE's sales is more likely to reduce the total welfare inside the FTA as T approaches $\widetilde{T}$. With a sufficiently large T, however, the negative effect can be covered by a large increase in total exports because a part of the output supplies is shifted from the less productive local firm to the more productive MNE. Given that $\tau < (a-w) + 2\Delta - \sqrt{\frac{2(a-w)^2+10(a-w)\Delta+11\Delta^2}{3}}$ holds such that consumer gains from tariff elimination are not large, we can derive the two thresholds, $\underline{T}^O_W$ and $\overline{T}^O_W$, such that FTA formation without ROO decreases the total welfare of the member countries if $\underline{T}^O_W < T < \overline{T}^O_W$ holds. We can confirm that $\widehat{T} < \overline{T}^O_W$ always holds, but it is ambiguous whether $\underline{T}^O_W$ is higher or lower than $\widehat{T}$.

When $T^* < T < \widehat{T}$ holds, FTA formation induces the input relocation of the MNE. It seems that FTA formation can be welfare-reducing because of the loss of tariff revenue and possible loss of the local firm. In the absence of ROO, however, we confirm that an FTA with input relocation improves the total welfare because country H collects tax revenue from the MNE, and this positive revenue effect outweighs possible negative effects. The following proposition summarizes the welfare effect for FTA countries in the absence of ROO (see Appendix A.4 for the proof).

Proposition 4. *In the absence of ROO, FTA formation worsens the total welfare of member countries when (i)* $\tau < \frac{2(a-w)}{13}$ *and* $T < \min[T_W^I, T^*]$ *hold or (ii)* $\tau < (a-w) + 2\Delta - \sqrt{\frac{2(a-w)^2+10(a-w)\Delta+11\Delta^2}{3}}$ *and* $\max[\underline{T}_W^O, \widehat{T}] < T < \overline{T}_W^O$ *hold. Otherwise, an FTA improves the total welfare.*

This result points out a new negative aspect of transfer pricing from the viewpoint of trade liberalization. Due to tax avoidance, the MNE's output decision becomes more aggressive, and its market share of the export market is larger than the market equilibrium without transfer pricing. However, increased operating profits due to the increased market share are not taxed inside the FTA. Therefore, transfer pricing makes FTA formation more likely to be welfare-reducing, which implies that the FTA formation is infeasible.

ROO neither change the equilibrium outcome nor the welfare property of FTA formation when $T \leq \widehat{T}$ holds, but ROO change them when $\widehat{T} < T$ holds. There are two sub-cases with $\widehat{T} < T$. First, when $\widetilde{T} < T$ holds, the MNE's input production takes place in country O irrespective of FTA formation, but an initial tariff is still applied to the MNE because the MNE does not comply with ROO. In this case, country H cannot collect tax revenue from the MNE either before or after the FTA formation. However, country F can collect tariff revenue from the MNE. Furthermore, because only the MNE incurs the tariff after the FTA formation, the local firm earns more profits with the FTA, which contributes to increasing the welfare of the member countries. Therefore, the MNE is treated as if it produces the final good outside the FTA, and the rent-shifting effect makes FTA formation welfare-improving for the member countries.

Second, when $\widehat{T} < T < \widetilde{T}$ holds, an FTA induces the MNE to relocate its input production, and the members of the FTA can collect tax revenue from the MNE. The tax revenue can cover the loss of consumer surplus owing to the reduced output of the MNE. Although the welfare effect of FTA formation itself is ambiguous in this region of T, we can show that ROO make an infeasible FTA without ROO a feasible one, if the MNE's cost advantage of input production in country O is small enough to satisfy $\Delta < \frac{(7\sqrt{2}-8)(a-w)}{32}$. This result is summarized as the following proposition (see Appendix A.5 for the proof).

Proposition 5. *If* $\Delta < \frac{(7\sqrt{2}-8)(a-w)}{32}$ *and* $\underline{T}_W^O < T < \overline{T}_W^O$ *hold, ROO transform an infeasible FTA into a feasible one.*

When $\underline{T}_W^O < T < \overline{T}_W^O$ holds, an FTA without ROO is infeasible because it reduces the total welfare of the member countries. With ROO, however, FTA formation either prevents tariff elimination to the MNE owned by a non-member country or generates tax revenue by inducing input relocation of the MNE. These effects enhance the positive welfare effects of the FTA on the member countries,

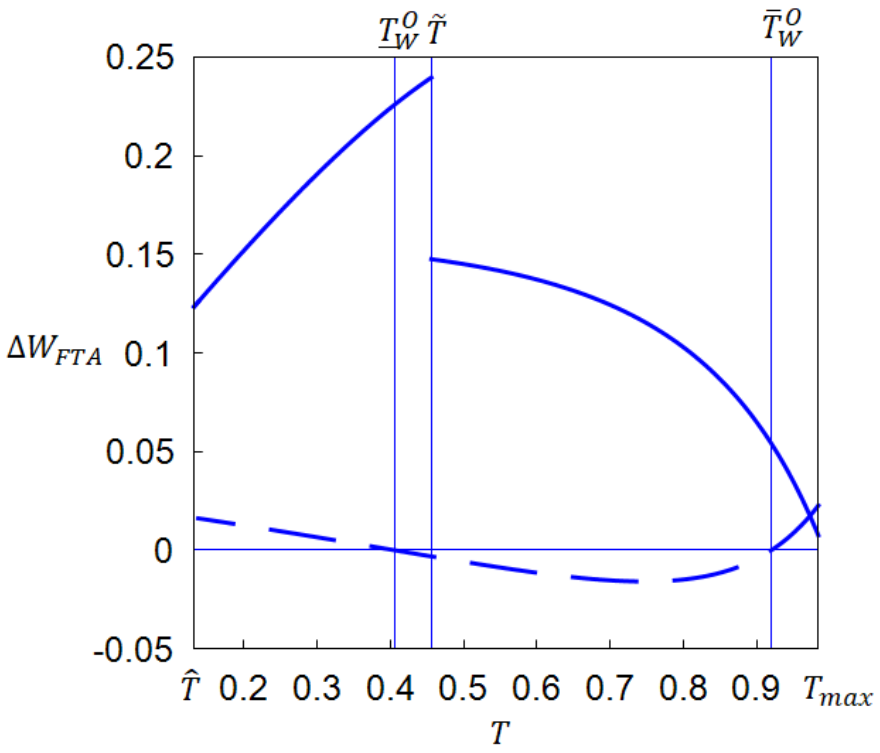

Figure 4: The effect of FTA formation on the welfare of FTA countries

and the FTA becomes feasible if the relative production efficiency of the MNE, Δ, is small enough. Figure 4 provides a numerical example that illustrates the welfare impacts of an FTA with and without ROO.[21] The dotted curve depicts the change in total welfare without ROO whereas the solid curves represent the ones with ROO. For $\hat{T} < T \leq \tilde{T}$, the MNE relocates input production with ROO whereas it does not without ROO. For $\tilde{T} < T$, the MNE produces inputs outside the FTA and it gives up utilizing an FTA tariff with ROO although the MNE is eligible for tariff elimination without ROO. We can see that, for $\underline{T}_W^O < T < \overline{T}_W^O$, the welfare of FTA formation is positive for the FTA countries with ROO whereas it is negative without ROO.

5 Discussion

Our analysis has provided new results that have not been explored in the extant literature. In this section, we explain how the MNE allocates its decision rights, and also argue the robustness of these results by relaxing some assumptions made in the benchmark model.

[21] In Figure 4, the parameters are set at $a = 3$, $w = 1$, $t = 0.1$, $\Delta = 1/32$ and $\tau = 1/4$.

5.1 The allocation of decision rights

In the baseline model, we suppose that the MNE chooses the centralization of its decision rights when it chooses the offshoring scheme, while the MNE chooses the decentralization when it chooses the inshoring scheme. Here, we explain the reason why the MNE makes these choices.

In the inshoring scheme, the MNE cannot use the transfer price to avoid a high tax in country H. The MNE's overall profit under the decentralized decision-makings is always larger than that under the centralized decision-making, because the MNE is able to manipulate the transfer price for the strategic motive in the former case. Therefore, the MNE always chooses the decentralization whenever it chooses the inshoring scheme.

In the offshoring scheme, if the MNE chooses the decentralization, it can use the transfer price for both the tax-avoidance motive and the strategic motive. Then, the MNE strikes the balance between the two motives and sets the transfer price as

$$r^{OD} = w - \Delta + \{a - w + 2\Delta - (2\lambda_M^I - \lambda_L)\tau\} \left[\frac{3(T-t)}{2\{3(1-t) - 2(1-T)\}} - \frac{1}{4} \right]. \tag{20}$$

The first-term of the last parenthesis reflects the tax-avoidance motive, while the second-term reflects the strategic motive. By using (20), we can calculate the overall profit of the MNE under the offshoring scheme with decentralization, which is denoted by Π_M^{OD}. If the MNE chooses the centralization under the offshoring scheme as in the benchmark model, it manipulates the transfer price solely for the tax-avoidance motive, and its overall profit is given by Π_M^O. By comparing Π_M^O with Π_M^{OD}, we have

$$\Pi_M^O \geq \Pi_M^{OD} \iff T \geq T^{OD} \equiv \frac{5(1+t) - 2(1-t)\sqrt{5}}{10}. \tag{21}$$

The MNE chooses the centralization in the offshoring scheme unless the tax gap is small enough. We can confirm that there are a range of parameter values under which T^{OD} is smaller than T^M, $\tilde{T}_{CS}$, and $\underline{T}_W^O$. Therefore, the main results of the baseline model, such as an exporter-hurting FTA and a consumer-hurting FTA , remain unchanged even if we consider a possibility that the MNE can choose the decentralization under the offshoring scheme.

5.2 The timing of setting transfer price

In the benchmark model, we assumed that the MNE's decision to set the input price is earlier than that of quantity setting. The pre-commitment of the input price along with the delegation of

decision rights to the downstream affiliate generates more profits for the MNE under the inshoring scheme because of the strategic effect. Under the offshoring scheme, however, the MNE may earn more profits by setting quantity first and then adjusting its transfer price to avoid a high tax.

If the timing of setting the transfer price and that of setting quantity are reversed, the post-tax profit is a standard duopoly profit taxed with a corporate tax rate in country O, $\Pi_M^{O,NC} = (1-t)\left(\frac{a+c_L-2(w-\Delta+\lambda_M^O\tau)}{3}\right)^2$. Thus, we can derive the condition under which price commitment is more profitable than quantity commitment for the MNE:

$$\Pi_M^O - \Pi_M^{O,NC} \geq 0 \iff T \leq \frac{3-t}{4}. \tag{22}$$

Hence, as long as the tax rate in the high-tax country, T, is sufficiently low, the MNE chooses the timing of the game in the benchmark model, even if the timing of moves is endogenously determined.

5.3 Home market

We assumed that the final product is consumed only in country F in the benchmark model. Here, we discuss how the results are affected by considering the market in country H. For simplicity, suppose that the market size of country H is the same as that of country F. We preserve the other assumptions that the MNE establishes a single plant in either country O or H and cannot report negative profits in each country.

We also assume that the production process in one plant cannot be separated depending on the destination of the products because it is highly costly for the MNE to adjust production processes depending on the destination of the goods. Thus, it is not possible that the MNE procures local inputs to comply with ROO and at the same time uses the inputs produced outside the FTA only for the domestic supply of the product.

The optimal transfer price under offshoring is the one such that $(p_H - r)x_{MH}^O + (p_F - r - \lambda_M^O\tau)x_{MF}^O = 0$ holds, whereas the first-order condition of profit maximization provides the optimal internal price under inshoring. We put the subscript HM in the presence of the home market, whereas no subscript corresponds to the benchmark case. We have the following rankings of the input prices.

$$\widetilde{r}_{HM}^I = \widetilde{r}^I < r_{HM}^{I*} < r^{I*}, \tag{23}$$

$$\widehat{r}^O < \widehat{r}_{HM}^O < r^{O*} < r_{HM}^{O*} < \widehat{r}_{HM}^O = \widehat{r}^O, \tag{24}$$

In the inshoring scheme, the pre-FTA level of the input price is lower with the home market ($r^{I*}_{HM} < r^{I*}$). Because a tariff is incurred in supplying to country H, the strategic effect of input-price adjustment is larger in the market of country H than that of country F. If the MNE could discriminate its input prices across the markets, the input price set for the market of country H would be lower with the home market. Thus, the uniform input price is lower with the home market. After FTA formation, the tariff is eliminated, and the market size of both countries becomes the same. Therefore, the input price with the home market is the same as that without it ($\widetilde{r}^I = \widetilde{r}^I_{HM}$).

In the offshoring scheme, the input price with the home market is the same as the benchmark case after FTA formation without ROO ($\widehat{r}^O_{HM} = \widehat{r}^O$). Because the effective market size is the same in the two countries, the transfer price that realizes zero profit in country H remains unchanged. After FTA formation with ROO, the MNE faces a tariff in country F, and the effective market size is larger in country H. Therefore, even if the MNE's profit from exporting to country F is zero at $r = \widetilde{r}^O$, the overall profits of the downstream affiliate of the MNE are still positive because it earns a positive profit from the home market in country H. Then, the MNE sets a higher transfer price ($\widetilde{r}^O < \widetilde{r}^O_{HM}$), such that a negative profit in country F is just covered by a positive profit in country H. By the same reason, the MNE sets a higher transfer price before FTA formation ($r^{O*} < r^{O*}_{HM}$).

Thus, in both the inshoring and offshoring schemes, the MNE's incentive to manipulate the input price gets stronger if we introduce home market competition. Nevertheless, the main results of the benchmark still hold. Because of the analytical complication, we present a numerical example to confirm the existence of the main results. Figure 5 shows that we have the same welfare property as the benchmark case shown in Figure 4, where the dotted curve depicts the change in total welfare without ROO whereas the solid curves represent the ones with ROO. As in the benchmark model, ROO transform an infeasible FTA into a feasible one for $T \in [\underline{T}^O_W, \overline{T}^O_W]$.[22] Furthermore, two additional curves are depicted in Figure 5. The downward curve represents how FTA formation with ROO changes the post-tax profits of the MNE, whereas the upward curve illustrates how it changes the pre-tax profits of the local firm. Thus, even if we consider the home market, FTA formation with ROO can hurt both the MNE and the local firm.

5.4 Concealment costs for transfer price manipulation

In the benchmark model, there is no cost of manipulating the transfer price. In practice, MNEs need to explain the plausibility of transfer pricing in order to shift profits across country. This concealment cost should increase as MNEs shift more profits because explaining the reasons for the

[22] In Figure 5, the parameters are set at $a = 3$, $w = 1$, $t = 0.05$, $\Delta = 1/32$, and $\tau = 1/4$.

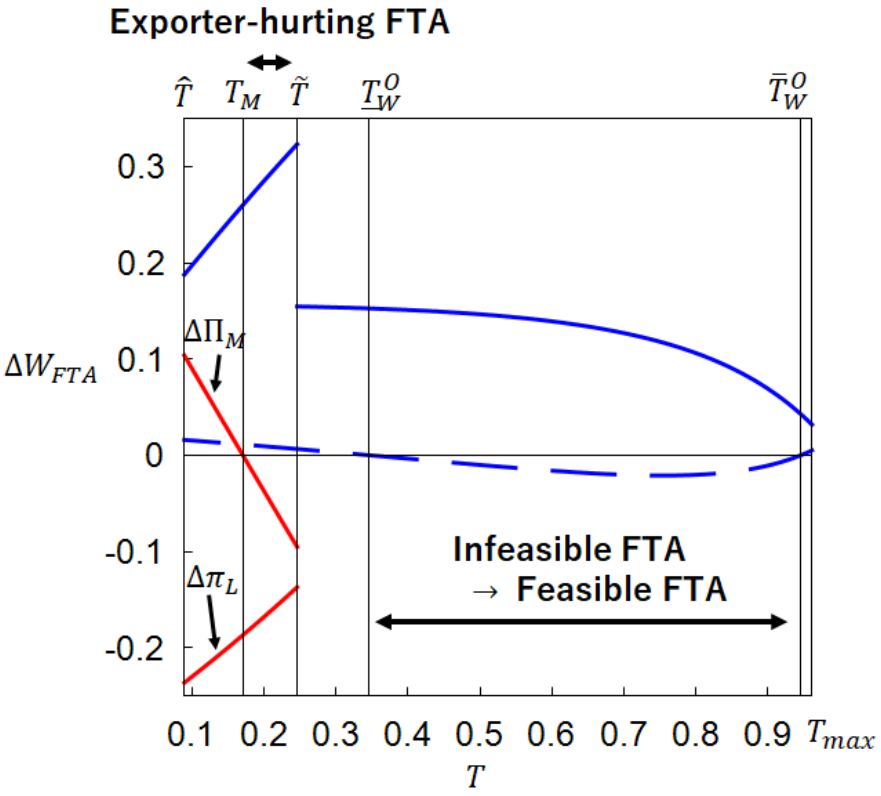

Figure 5: The effects of FTA formation with home market competition

greater deviation from the "appropriate price" (or arm's length price) becomes more difficult. Here, we show that introducing a concealment cost of transfer pricing does not change the main results.

Following the literature on transfer pricing, we introduce the following quadratic concealment cost in the case of offshoring:

$$C(r^O) = \frac{\delta\{r^O - (w - \Delta)\}^2}{2}, \tag{25}$$

where δ is a parameter that captures the difficulty of concealing tax avoidance. A higher δ corresponds to a more difficult environment of profit shifting owing to a well-enforced tax code by authorities, for example. In the case of inshoring, the MNE does not incur such a cost of profit shifting. The introduction of the concealment cost does not influence the MNE's actions under the inshoring scheme. Therefore, we only discuss the effect of introducing a concealment cost under the offshoring scheme. The modified post-tax profits are given by

$$\begin{aligned}\Pi_M^O &= (1-t)[\{r^O - (w-\Delta)\}x_M^O] + (1-T)[(p - r^O - \lambda_M\tau)x_M^O] - C(r^O) \\ &= (1-T)(p - c_M^O)x_M^O - C(r^O).\end{aligned} \tag{26}$$

Unlike the benchmark case, the cost of profit shifting may prevent the MNE from shifting all the

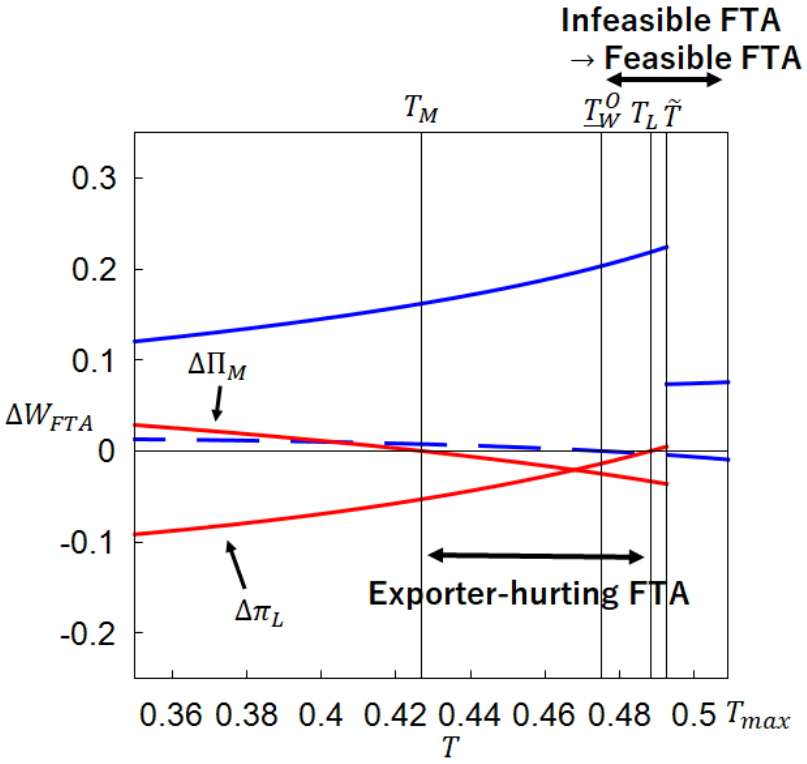

Figure 6: The effects of FTA formation with a concealment cost

profits.[23]

Figure 6 illustrates a numerical example of the case with the concealment cost, where the dotted curve again depicts the change in total welfare without ROO whereas the solid curves represent the ones with ROO.[24] We can again obtain qualitatively similar results as the benchmark analysis, though the upper bound of the thresholds, $\overline{T}_W^O$, disappears in this extension. Thus, the main findings in the benchmark analysis are robust even if we consider the cost of profit shifting.

6 Conclusion

A recent proliferation of FTAs has played a key role in advancing trade liberalization between countries. Trade liberalization usually increases the profits of firms via exporting and benefits consumers in importing countries. Because FTAs entail ROO, however, the profits and the welfare effects of FTAs are much more complicated than they seem. Specifically, ROO require exporting firms to use inputs produced in FTA countries to be eligible for tariff elimination and thereby affect

[23]As the amount of shifted profits depends on the parameter δ, a partial profit shifting occurs with the concealment cost if the tax in country H, T, is less than the level that either makes the local firm's supplies to be zero ($x_L^O = 0$) or shifts the entire tax base of the MNE to country O ($\pi_M^O = 0$).

[24]The parameters are set at $a = 3$, $w = 1$, $t = 0.1$, $\Delta = 1/32$, $\tau = 1/8$, and $\delta = 1$ in Figure 6.

location choices of exporting firms for input production. In particular, if an exporting firm is a vertically-integrated MNE that conducts both input and output production in different affiliates, it is able to manipulate its transfer price to avoid a high corporate tax or to increase the downstream affiliate's competitiveness in the product market. The decision rights of the MNE are centralized in the former case and decentralized in the latter case. Therefore, FTA formation that affects MNEs' location choices also affects MNEs' allocation of decision rights and strategies to manipulate their transfer prices. This study investigated the effects of an FTA with ROO when an MNE chooses its location of input production and manipulates its transfer price set for intra-firm transactions.

If the tax gap between potential host countries of input production is wide, the MNE prefers to locate its input production in a country outside an FTA in the pre-FTA equilibrium. In this case, the decision rights are centralized, and the MNE manipulates its transfer price to avoid a high corporate tax (i.e., the Lion is sleeping). If the tax gap is narrow, the MNE locates its input production in the same country as the output production in the pre-FTA equilibrium. In this case, the decision rights are decentralized, and the MNE manipulates its transfer price to shift profits from the local firm in the product market (i.e., the Lion is waking up). FTA formation can induce the MNE to relocate input production to inside an FTA country and delegate its output decisions to the downstream affiliate because the strategic motive of transfer pricing becomes more important than the tax-avoidance motive. Thus, an FTA may wake the Lion and, in that case, the induced change in the nature of the product-market competition substantially affects the welfare properties of FTA formation.

A remarkable result is that both the local firm and MNE can lose from FTA formation in the presence of ROO, even if the tariffs imposed on them are eliminated. The local firm suffers from the FTA formation because the MNE relocates its input production and becomes more aggressive in the product market. The MNE suffers from the FTA because it needs to pay a high tax with input relocation. Nevertheless, the MNE chooses input relocation because it cannot comply with ROO, and the tariff is imposed only on the MNE without the relocation.

FTA formation with ROO can also decrease total exports within the FTA and hurt consumers of the importing country, even though all firms comply with ROO and all tariffs are eliminated. This is because input relocation increases the production cost of the MNE and the negative effect from the increase in the marginal cost can outweigh the positive effect from the tariff elimination. Although the presence of ROO tends to induce the inefficient relocation of input production, which can cause unfavorable effects on the MNE and consumers, it can be a tool to improve the total welfare of the FTA countries. An FTA without ROO may worsen the total welfare of the FTA countries because the

member countries collect no tax revenue from the MNE. With ROO, input relocation is necessary for the MNE to be eligible for tariff elimination. With the induced input relocation, the member countries of the FTA collect tax revenue from the MNE. Even without the input relocation, the FTA countries are able to collect tariff revenue from the MNE because the MNE does not comply with ROO. Because of these effects, ROO can transform an infeasible FTA into a feasible one. These results point to a new role for FTA formation in the presence of the MNE's tax avoidance through transfer pricing: An FTA works as a tool to prevent tax avoidance by the MNE, whose effectiveness is reinforced by the ROO.

These results provide important policy implications amid the real-world prevalence of intra-firm trade and export-platform FDIs. Policy makers should note that, even if firms comply with ROO and make tariff-free exports, an FTA does not always benefit these firms or consumers when the MNEs manipulate their transfer prices. Besides that, although Felbermayr et al. (2019) conclude that there is no rationale for having ROO because tariff circumvention is usually not profitable, our model suggests that ROO work as a tool to prevent *tax* circumvention, if not *tariff* circumvention.

There remains scope for further research. We have assumed that tax rates and tariff rates are exogenously given. It is intriguing to investigate how the formation of an FTA affects the outcomes of tax competition among countries and how it affects the optimal tariffs that FTA members set. Examining the effects of regulations on transfer pricing, such as the arm's length principle, in this setting will also be a possible extension. Empirical investigations into the relationship between ROO and transfer pricing will be essential to strengthen the real-world relevance of our results.

Appendix

A.1 Proof of Lemma 1

The comparison of supplies is equivalent to that of post-tax profits if the MNE's input location is unaffected by FTA formation. The equilibrium supplies of firms are given by

$$x_M^s = \begin{cases} \frac{a-w-(2\lambda_M-\lambda_L)\tau}{2} & \text{when inshoring} \\ \frac{(1-t)\{a-w+2\Delta-(2\lambda_M-\lambda_L)\tau\}}{3-2T-t} & \text{when offshoring} \end{cases}, \tag{a1}$$

$$x_L^s = \begin{cases} \frac{a-w-(3\lambda_L-2\lambda_M)\tau}{4}, & \text{when inshoring} \\ \frac{(1-T)(a-w)-(1-t)\Delta-\{(1-T)\lambda_L-(1-t)(\lambda_L-\lambda_M)\}\tau}{3-2T-t} & \text{when offshoring} \end{cases}. \tag{a2}$$

When an FTA does not affect the MNE's location of input production, tariff elimination clearly increases these supplies, and equivalently, the post-tax profits of both firms. This means that the total supply also increases. When $T^* \leq T \leq \widehat{T}$ holds, the MNE changes the country of input production from country O to country H. The change in the MNE's post-tax profits is computed by using (a1), which is given by

$$\widehat{\Pi}_M^I - \Pi_M^{O*} \propto \widehat{x}_M^I - \sqrt{2}x_M^{O*} = \frac{a-w}{2} - \frac{(1-t)(a-w+2\Delta-\tau)\sqrt{2}}{3-2T-t} \gtreqless 0$$
$$\iff T \lesseqgtr \frac{3-t}{2} - \frac{(1-t)(a-w+2\Delta-\tau)\sqrt{2}}{a-w} \equiv T_M,$$

Although the MNE's post-tax profits would decrease if $T > T_M$ holds, because $\widehat{T} < T_M$ always holds, FTA formation always benefits the MNE without ROO.

A.2 Proof of Lemma 2

Without ROO, the amount of supplies by the local firm decreases if and only if

$$\widehat{x}_L^I = \frac{a-w}{4} > \frac{(1-T)(a-w-\tau)-(1-t)\Delta}{3-2T-t} = x_L^{O*} \iff T > 1 - \frac{(1-t)(a-w+4\Delta)}{2(a-w-2\tau)} \equiv T_L$$

holds. By comparing T_L with T^* and $\widehat{T}$, we have

$$\widehat{T} - T_L = (1-t)\left\{\frac{a-w-\tau+2\Delta}{a-w-2\tau} - \frac{(a-w+2\Delta)\sqrt{2}}{a-w}\right\} \gtrless 0$$
$$\iff \tau \gtrless \frac{(\sqrt{2}-1)(a-w)(a-w+2\Delta)}{(2\sqrt{2}-1)(a-w)-4\Delta} \equiv \widehat{\tau}_L, \quad \text{and}$$
$$T^* - T_L = (1-t)\left\{\frac{a-w-\tau+2\Delta}{a-w-2\tau} - \frac{(a-w+2\Delta-\tau)\sqrt{2}}{a-w-\tau}\right\} \gtrless 0$$
$$\iff \tau \gtrless \left(\frac{\sqrt{2}-1}{2\sqrt{2}-1}\right)(a-w) \equiv \tau_L^*.$$

We can confirm that $\widehat{\tau}_L < \tau_L^*$ holds. Therefore, FTA formation hurts firm L when $T^* \leq T \leq T_L$ holds with $\widehat{\tau}_L < \tau < \tau^*$, or when $T^* \leq T \leq \widehat{T}$ holds with $\tau \leq \widehat{\tau}_L$.

With ROO, (a2) suggests that FTA formation benefits the local firm. By subtracting T_L from $\widetilde{T}$,

we have

$$\widetilde{T} - T_L \propto \frac{a-w-\tau+2\Delta}{a-w-2\tau} - \frac{(a-w-2\tau+2\Delta)\sqrt{2}}{a-w} \geq 0 \iff \tau \geq \widetilde{\tau}_L,$$

$$\text{where} \quad \widetilde{\tau}_L \equiv \frac{\left(4\sqrt{2}-1\right)(a-w)+4\Delta\sqrt{2}-\sqrt{\left(8\sqrt{2}+1\right)(a-w)^2+8\left(4+3\sqrt{2}\right)(a-w)\Delta+32\Delta^2}}{8\sqrt{2}}.$$

Therefore, $\widetilde{T} \geq T_L$ holds if and only if $\tau \geq \widetilde{\tau}_L$ holds.

A.3 Proof of Proposition 3

First, we analyze the change in total exports in the absence of ROO. From (a1) and (a2), it is obvious that total exports increase when an FTA does not change the MNE's strategy. If the MNE changes the production location owing to an FTA, the change in total exports is positive if

$$\left(\widehat{x}_M^I + \widehat{x}_L^I\right) - \left(x_M^{O*} + x_L^{O*}\right) = \frac{3(a-w)}{4} - \frac{(2-T-t)(a-w-\tau)+(1-t)\Delta}{3-2T-t} > 0$$
$$\iff T < 1 - \frac{(1-t)\{a-w-4(\tau-\Delta)\}}{2(a-w+2\tau)} \equiv \widetilde{T}_{CS}$$

holds. By subtracting $\widetilde{T}_{CS}$ from $\widehat{T}$,

$$\widetilde{T}_{CS} - \widehat{T} = (1-t)\left(\frac{(a-w+2\Delta)\sqrt{2}}{a-w} - \frac{a-w-\tau+2\Delta}{a-w+2\tau}\right)$$

is obtained. Because the second term is decreasing in τ, $\widetilde{T}_{CS} - \widehat{T}$ is minimized at $\tau = 0$. At $\tau = 0$, $\widetilde{T}_{CS} - \widehat{T} = (1-t)(\sqrt{2}-1)\left(\frac{a-w+2\Delta}{a-w}\right) > 0$. Therefore, $\widehat{T} < T_{CS}$ always holds, which means that FTA formation always benefits consumers in country F in the absence of ROO.

As ROO are redundant when $T < \widehat{T}$ holds, we investigate the case of $\widehat{T} < T < \widetilde{T}$ and that of $\widetilde{T} < T$, respectively. When $\widehat{T} < T < \widetilde{T}$ holds, the FTA increases total exports under regime I if and only if $T < \widetilde{T}_{CS}$ holds. We have

$$\widetilde{T} - T_{CS} = \frac{a-w-\tau+2\Delta}{a-w+2\tau} - \frac{(a-w+2\Delta-2\tau)\sqrt{2}}{a-w}$$
$$\propto -(\sqrt{2}-1)(a-w)(a-w+2\Delta) - \left(a-w+4\Delta\sqrt{2}\right)\tau + 4\tau^2\sqrt{2}.$$

$\widetilde{T} - T_{CS}$ takes the minimum value at $\tau = \frac{a-w+4\Delta\sqrt{2}}{8\sqrt{2}}$, which is negative. Furthermore, $\widetilde{T} - T_{CS}$ is maximized at either $\tau = 0$ or $\tau = \min\left\{\frac{a-w+2\Delta}{2}, a-w-\left(\frac{1-t}{1-T}\right)\Delta\right\}$. It is negative at $\tau = 0$. Besides

that, we have

$$\frac{a-w-4\Delta\sqrt{2}}{8\sqrt{2}} < \min\left\{\frac{a-w+2\Delta}{2}, a-w-\left(\frac{1-t}{1-T}\right)\Delta\right\} \equiv \tau^{max}.$$

When $T \leq \frac{a-w-2(2-t)\Delta}{a-w-\Delta}$ holds, we have $\tau^{max} = \frac{a-w+2\Delta}{2}$, at which

$$\left(\widetilde{T} - \widetilde{T}_{CS}\right)\Big|_{\tau=\frac{a-w+2\Delta}{2}} = \frac{(a-w)(a-w+2\Delta)}{2} > 0.$$

This implies that there exists a unique threshold, $\widetilde{\tau}_{CS}$, such that $\widetilde{T} - \widetilde{T}_{CS} > 0$ holds for $\tau > \widetilde{\tau}_{CS}$. By definition, $\frac{a-w+2\Delta}{2} = a-w-\left(\frac{1-t}{1-T}\right)\Delta$ holds at $T = \frac{a-w-2(2-t)\Delta}{a-w-\Delta}$, and thus $\widetilde{T} - \widetilde{T}_{CS}\Big|_{\tau=a-w-\left(\frac{1-t}{1-T}\right)\Delta} > 0$ also holds when T is close to $T = \frac{a-w-2(2-t)\Delta}{a-w-\Delta}$. Specifically, the threshold is calculated as

$$\widetilde{T} > \widetilde{T}_{CS} > 0 \iff \tau > \widetilde{\tau}_{CS},$$
$$\text{where} \quad \widetilde{\tau}_{CS} = \frac{a-w+4\Delta\sqrt{2}+\sqrt{(33-8\sqrt{2})(a-w)^2+8(8-3\sqrt{2})(a-w)\Delta+32\Delta^2}}{8\sqrt{2}}.$$

Therefore, an FTA with ROO hurts consumers when $\widetilde{T}_{CS} < T < \widetilde{T}$ holds, which is the case when $\tau > \widetilde{\tau}_{CS}$ holds.

When $\widetilde{T} < T$ holds, tariff elimination is applied only to the local firm. Then, we have

$$\left(\widetilde{x}_M^O + \widetilde{x}_L^O\right) - \left(x_M^{O*} + x_L^{O*}\right) = \frac{(1-T)\tau}{3-2T-t} > 0.$$

In this case, FTA formation benefits consumers.

A.4 Proof of Proposition 4

Suppose that ROO are absent. When $T < T^*$ holds, the effect of FTA formation on the total welfare of member countries is given by

$$\widehat{W}_{FTA}^I - W_{FTA}^{I*} = \frac{\tau}{32}\left((11+4T)(2(a-w)-\tau) - 24(a-w-\tau)\right).$$

If $\tau \geq \frac{2(a-w)}{13}$ holds, we have $\widehat{W}_{FTA}^I \geq W_{FTA}^{I*}$ irrespective of the level of T. If $\tau < \frac{2(a-w)}{13}$ holds, however, the sign of $\widehat{W}_{FTA}^I - W_{FTA}^{I*}$ depends on T. We can calculate that

$$\widehat{W}_{FTA}^I \gtreqless W_{FTA}^{I*} \iff T \gtreqless T_W^I \equiv \frac{2(a-w)-13\tau}{4\{2(a-w)-\tau\}}$$

holds. Because T_W^I can be either higher or lower than T^*, FTA formation worsens the total welfare if $\tau < \frac{2(a-w)}{13}$ and $T < \min[T_W^I, T^*]$ hold.

When $\widehat{T} < T$ holds, the welfare effect of FTA formation becomes

$$\widehat{W}_{FTA}^O - W_{FTA}^{O*} = \frac{\tau\left\{-2(1-T)\{(T-t)(a-w)+3\Delta(1-t)\}+(t^2+T^2+4Tt-6t-6T+6)\tau\right\}}{2\{1-t+2(1-T)\}^2}.$$

We can obtain the threshold level of T, $\underline{T}_W^O$ and $\overline{T}_W^O$, by solving $\widehat{W}_{FTA}^O = W_{FTA}^{O*}$. We have $\widehat{W}_{FTA}^O < W_{FTA}^{O*}$ if

$$\underline{T}_W^O \equiv 1 - \frac{(1-t)\{a-w+3\Delta-2\tau+\sqrt{\theta}\}}{2(a-w)+\tau} < T < 1 - \frac{(1-t)\{a-w+3\Delta-2\tau-\sqrt{\theta}\}}{2(a-w)+\tau} \equiv \overline{T}_W^O$$

holds, where $\theta \equiv (a-w)^2 - 6(\tau-\Delta)(a-w) + 3(\tau-\Delta)(\tau-3\Delta)$. If $\tau < (a-w)+2\Delta - \sqrt{\frac{2(a-w)^2+10(a-w)\Delta+11\Delta^2}{3}}$ holds, there is a real solution of $\underline{T}_W^O$ and $\overline{T}_W^O$ and an FTA becomes a welfare-reducing one if $\underline{T}_W^O < T < \overline{T}_W^O$ holds. Otherwise, there is no real solution of $\underline{T}_W^O$ and $\overline{T}_W^O$ and $\widehat{W}_{FTA}^O \geq W_{FTA}^{O*}$ always holds. We can calculate that $\underline{T}_W^O$ can be higher or lower than $\widehat{T}$ whereas $\widehat{T} < \overline{T}_W^O$ always holds. Therefore, FTA formation worsens the total welfare if $\tau < (a-w)+2\Delta - \sqrt{\frac{2(a-w)^2+10(a-w)\Delta+11\Delta^2}{3}}$ and $\max[\underline{T}_W^O, \widehat{T}] < T < \overline{T}_W^O$ hold.

When $T^* < T < \widehat{T}$ holds, the second derivative of the welfare effect of FTA formation is given by

$$\frac{\partial^2(\widehat{W}_{FTA}^I - W_{FTA}^{O*})}{\partial T^2} = -\left[\frac{a-w+3\tau}{\{1-t+2(1-T)\}^3} + \frac{6\{(T-t)(a-w-\tau)+3(1-t)\Delta\}}{\{1-t+2(1-T)\}^4}\right] < 0,$$

which implies that $\widehat{W}_{FTA}^I - W_{FTA}^{O*}$ takes the minimum value at either $T = T^*$ or $T = \widehat{T}$. Furthermore, at $T = T^*$, we obtain

$$\left(\widehat{W}_{FTA}^I - W_{FTA}^{O*}\right)\Big|_{T=T^*} = \frac{(8-5\sqrt{2})(a-w)^2 - 2(16-11\sqrt{2})(a-w)\tau + (5\sqrt{2}+24)\tau^2}{64\sqrt{2}} + \frac{T^*(a-w)^2}{8},$$

whose first term takes the minimum value $\frac{(1187\sqrt{2}-1456)(a-w)^2}{16\sqrt{2}(5\sqrt{2}+24)^2} > 0$ at $\tau = \frac{(16-11\sqrt{2})(a-w)}{5\sqrt{2}+24}$. This means that $\left(\widehat{W}_{FTA}^I - W_{FTA}^{O*}\right)\Big|_{T=T^*} > 0$ holds.

At the other edge, we have

$$\left(\widehat{W}_{FTA}^I - W_{FTA}^{O*}\right)\Big|_{T=\widehat{T}} = \frac{(8-5\sqrt{2})(a-w)^2(a-w+2\Delta)^2 - 2(a-w)\Gamma_1\tau + \Gamma_2\tau^2}{64\sqrt{2}(a-w+2\Delta)^2} + \frac{\widehat{T}(a-w)^2}{8},$$

where $\Gamma_1 \equiv \{(16-11\sqrt{2})(a-w)^2 + (56-38\sqrt{2})(a-w)\Delta + (48-32\sqrt{2})\Delta^2 > 0$ and $\Gamma_2 \equiv \{(5\sqrt{2}+$

$24)(a-w)^2+(48+32\sqrt{2})(a-w)\Delta+32\sqrt{2}\Delta^2\}>0$. It takes the minimum value at $\tau=\frac{(a-w)\Gamma_1}{\Gamma_2}$, which is given by

$$\left(\widehat{W}^I_{FTA}-W^{O*}_{FTA}\right)\Big|_{T=\widehat{T},\tau=\frac{\Gamma_1}{\Gamma_2}}=\frac{\widehat{T}(a-w)^2}{8}+\frac{4(a-w)^2\Gamma_3}{64\sqrt{2}(a-w+2\Delta)^2\Gamma_2}>0.$$

where $\Gamma_3\equiv(68\sqrt{2}-89)(a-w)^4+(536\sqrt{2}-708)(a-w)^3\Delta+4(396\sqrt{2}-529)(a-w)^2\Delta^2+2^5(65-88)(a-w)\Delta^3+2^7(8\sqrt{2}-11)\Delta^4>0$. Thus, in the absence of ROO, FTA formation that induces input relocation always improves the total welfare of member countries.

A.5 Proof of Proposition 5

First, consider the case of $\widetilde{T}<T<T^{max}$, where the MNE procures inputs from country O both before and after FTA formation. Let $\widetilde{W}^O_{FTA}-W^{O*}_{FTA}$ be the welfare gains from an FTA with ROO. We obtain

$$\begin{aligned}\frac{\partial(\widetilde{W}^O_{FTA}-W^{O*}_{FTA})}{\partial T}&=\frac{(1-t)\tau\{2(a-w-\tau)+5\Delta)t-(2(a-w+\Delta)-5\tau)T-3(\tau+\Delta)\}}{\{1-t+2(1-T)\}^3}\\&<\frac{(1-t)\tau\{2(a-w-\tau)+5\Delta)T-(2(a-w+\Delta)-5\tau)T-3(\tau+\Delta)\}}{\{1-t+2(1-T)\}^3}\\&=\frac{-3(1-T)(\tau+\Delta)}{\{1-t+2(1-T)\}^3}<0.\end{aligned}$$

Therefore, $\widetilde{W}^O_{FTA}-W^{O*}_{FTA}$ takes the minimum value at $T=T^{max}$, which is given by

$$\widetilde{W}^O_{FTA}-W^{O*}_{FTA}\Big|_{T=T^{max}}=\frac{\Delta\tau\{2(a-w-\tau)^2+\Delta(4(a-w)-\tau)\}}{2(a-w+2\Delta-\tau)^2}>0.$$

Therefore, we have

$$\widetilde{W}^O_{FTA}-W^{O*}_{FTA}\Big|_{T=\widetilde{T}}>\widetilde{W}^O_{FTA}-W^{O*}_{FTA}\Big|_{T=T^{max}}>0.$$

Next, consider the case with $\widehat{T}<T<\widetilde{T}$, where input relocation happens. As indicated in Appendix A.4, $\widehat{W}^I-W^{O*}$ may take the minimum value at $T=\widetilde{T}$ and be negative. Considering this possibility, we compare it with $\widetilde{W}^O-W^{O*}$ at $T=\widetilde{T}$, which is given by

$$\begin{aligned}\left(\widehat{W}^I-W^{O*}\right)\Big|_{T=\widetilde{T}}-\left(\widetilde{W}^O-W^{O*}\right)\Big|_{T=\widetilde{T}}&=\left(\widehat{W}^I-\widetilde{W}^O\right)\Big|_{T=\widetilde{T}}\\&=\frac{(a-w)[(7\sqrt{2}-8)(a-w)-32\Delta+\{4(4-\sqrt{2})(a-w)+32(\Delta-\tau)\}t]}{64\sqrt{2}}.\end{aligned}$$

Note that FTA formation without ROO reduces the total welfare when $\tau<a-w+2\Delta-$

$\sqrt{\frac{2(a-w)^2+10(a-w)\Delta+11\Delta^2}{3}}$ holds. Then, it is computed as

$$\left(\widehat{W}^I - \widetilde{W}^O\right)\Big|_{T=\widetilde{T}} = \frac{(a-w)}{64\sqrt{2}}[(7\sqrt{2}-8)(a-w) - 32\Delta + \{4(4-\sqrt{2})(a-w) + 32(\Delta-\tau)\}t]$$
$$\propto (7\sqrt{2}-8)(a-w) - 32\Delta + 4\left\{8\sqrt{\frac{2(a-w)^2+10(a-w)\Delta+11\Delta^2}{3}} - ((4+\sqrt{2})(a-w)+8\Delta)\right\}t,$$

where the last bracketed expression is positive. Thus, if $\Delta < \frac{(7\sqrt{2}-8)(a-w)}{32}$ holds, we obtain $\widehat{W}^I_{FTA} - W^{O*}_{FTA} > 0$.

References

Blouin, J.L., Robinson, L.A., Seidman, J.K., 2018. Conflicting transfer pricing incentives and the role of coordination. Contemporary Accounting Research 35, 87–116.

Choi, J.P., Furusawa, T., Ishikawa, J., 2020. Transfer pricing regulation and tax competition. Journal of International Economics 127, 103367.

Cipollina, M., Salvatici, L., 2010. Reciprocal trade agreements in gravity models: A meta-analysis. Review of International Economics 18, 63–80.

Conconi, P., García-Santana, M., Puccio, L., Venturini, R., 2018. From final goods to inputs: the protectionist effect of rules of origin. American Economic Review 108, 2335–65.

Cristea, A.D., Nguyen, D.X., 2016. Transfer pricing by multinational firms: New evidence from foreign firm ownerships. American Economic Journal: Economic Policy 8, 170–202.

Czechowicz, I.J., Choi, F.D., Bavishi, V.B., 1982. Assessing Foreign Subsidiary Performance: systems & practices of leading multinational companies. Business International Corporation.

Davies, R.B., Martin, J., Parenti, M., Toubal, F., 2018. Knocking on tax haven's door: Multinational firms and transfer pricing. Review of Economics and Statistics 100, 120–134.

Demidova, S., Krishna, K., 2008. Firm heterogeneity and firm behavior with conditional policies. Economics Letters 98, 122–128.

Dürr, O.M., Göx, R.F., 2011. Strategic incentives for keeping one set of books in international transfer pricing. Journal of Economics & Management Strategy 20, 269–298.

Elitzur, R., Mintz, J., 1996. Transfer pricing rules and corporate tax competition. Journal of Public Economics 60, 401–422.

Ernst & Young, 2003. Transfer pricing 2003 global survey. International Tax Services .

Felbermayr, G., Teti, F., Yalcin, E., 2019. Rules of origin and the profitability of trade deflection. Journal of International Economics 121, 103248.

Fershtman, C., Judd, K.L., 1987. Equilibrium incentives in oligopoly. The American Economic Review , 927–940.

Freund, C., Ornelas, E., 2010. Regional trade agreements. Annual Review of Economics 2, 139–166.

Göx, R.F., Schiller, U., 2006. An economic perspective on transfer pricing. Handbooks of management accounting research 2, 673–695.

Hanson, G.H., Mataloni Jr, R.J., Slaughter, M.J., 2005. Vertical production networks in multinational firms. Review of Economics and statistics 87, 664–678.

Hayakawa, K., Hiratsuka, D., Shiino, K., Sukegawa, S., 2013. Who uses free trade agreements? Asian Economic Journal 27, 245–264.

Hyde, C.E., Choe, C., 2005. Keeping two sets of books: The relationship between tax and incentive transfer prices. Journal of Economics & Management Strategy 14, 165–186.

Ishikawa, J., Mukunoki, H., Mizoguchi, Y., 2007. Economic integration and rules of origin under international oligopoly. International Economic Review 48, 185–210.

Ju, J., Krishna, K., 2005. Firm behaviour and market access in a free trade area with rules of origin. Canadian Journal of Economics/Revue canadienne d'économique 38, 290–308.

Kato, H., Okoshi, H., 2019. Production location of multinational firms under transfer pricing: The impact of the arm's length principle. International tax and public finance 26, 835–871.

Krishna, K., Krueger, A., 1995. Implementing free trade areas: Rules of origin and hidden protection. Technical Report. National Bureau of Economic Research.

Liu, L., Schmidt-Eisenlohr, T., Guo, D., 2020. International transfer pricing and tax avoidance: Evidence from linked trade-tax statistics in the united kingdom. Review of Economics and Statistics 102, 766–778.

Mukunoki, H., 2017. The welfare effect of a free trade agreement in the presence of foreign direct investment and rules of origin. Review of International Economics 25, 733–759.

Mukunoki, H., Okoshi, H., 2020. Tariff elimination versus tax avoidance: Free trade agreements and transfer pricing. Munich Discussion Paper .

Mukunoki, H., Okoshi, H., 2021. Rules of origin and consumer-hurting free trade agreements. The World Economy Forthcoming.

Nielsen, S.B., Raimondos-Møller, P., Schjelderup, G., 2008. Taxes and decision rights in multinationals. Journal of Public Economic Theory 10, 245–258.

Peralta, S., Wauthy, X., Van Ypersele, T., 2006. Should countries control international profit shifting? Journal of International Economics 68, 24–37.

Schjelderup, G., Sorgard, L., 1997. Transfer pricing as a strategic device for decentralized multinationals. International Tax and Public Finance 4, 277–290.

Sklivas, S.D., 1987. The strategic choice of managerial incentives. The RAND Journal of Economics 18, 452–458.

Stöwhase, S., 2013. How profit shifting may increase the tax burden of multinationals: A simple model with discrete investment choices. Journal of Public Economic Theory 15, 185–207.

Stoyanov, A., 2012. Tariff evasion and rules of origin violations under the Canada-US free trade agreement. Canadian Journal of Economics 45, 879–902.

Takahashi, K., Urata, S., 2010. On the use of FTAs by japanese firms: Further evidence. Business and Politics 12, 1–15.

Tekin-Koru, A., Waldkirch, A., 2010. North–south integration and the location of foreign direct investment. Review of International Economics 18, 696–713.

Tintelnot, F., 2017. Global production with export platforms. The Quarterly Journal of Economics 132, 157–209.

Tomiura, E., 2007. Foreign outsourcing, exporting, and FDI: A productivity comparison at the firm level. Journal of International Economics 72, 113–127.