Global Production Decisions with Demand Uncertainty and Tax Cross-Crediting

The U.S. tax law taxes a multinational firm (MNF)’s global incomes at its home country tax rate. To avoid double taxation, it permits tax cross-crediting. Namely, the global firm can use excess foreign tax credits (FTCs) (the portion of foreign tax payments that exceed the home country tax liabilities) generated from a subsidiary located in a high-tax country to offset the tax liabilities of its low-tax divisions. This paper studies production decisions in a subsidiary of an MNF with tax cross-crediting. Casting the problem on a newsvendor model with the objective of maximizing the global firm’s worldwide after-tax profits, we show that the optimal production decision under the effects of tax cross-crediting can behave very differently from that of the traditional newsvendor model. In particular, we show that an improvement in the firm’s after-tax profitability (through tax cross-crediting, increased profit margin, or reduced tax rate) may reduce the optimal production quantity; and that the optimal production decision under certain circumstances can be made without the knowledge of the demand distribution. We also discuss the issue of how to motivate the division manager to use an after-tax performance measure with a managerial tax rate.

1. Introduction

The globalization of the world’s economy allows multinational firms (MNFs) to expand their global businesses rapidly beyond home countries. Many of these companies increasingly recognize the strategic value of aligning international tax planning with their global supply chain decisions (Murphy and Goodman 1998). One of the major challenges for an MNF to develop such a so-called tax-aligned or tax-effective global supply chain strategy is to effectively motivate and coordinate the supply chain decisions at its subsidiaries to achieve the global company’s overall operational and financial objectives.

Wilson (1993) conducted a study of nine U.S. MNFs on how they integrate tax planning with supply chain decisions such as capacity expansion and product production. Most of the firms surveyed used pretax measures to evaluate their division managers’ performances. However, some firms recognized that the pretax measures may create potential conflicts between a global firm’s goal of maximizing worldwide after-tax profits and its local operations decisions that are based on pretax performance. They therefore made various informal and formal adjustments on managerial books to mitigate the potential coordination frictions. Two firms in the study used certain forms of after-tax performance measures. However, these after-tax measures “did not provide individual managers with a direct incentive to coordinate activities such as production.” Informal communication of
tax planning was also deployed to make further adjustments. The Wilson study demonstrates the complexity of designing and implementing a decentralized performance measure in an MNF to “ensure the correct distribution of (production) quantities to maximize after-tax worldwide income” (Wilson 1993). Further, it suggests that many MNFs used pretax performance measures simply because “determining the correct tax rate for these after-tax performance measures is a complex, if not impossible, problem that requires considerable coordination between corporate and foreign tax planners”.

There are ample empirical evidences that MNFs with greater tax planning opportunities tend to use after-tax performance measures to provide incentives for their managers to explicitly consider the impacts of tax on their operation decisions. Specifically, large companies with a greater number of operating divisions and a higher level of multinational operations are found to be more likely to select after-tax rather than pre-tax measures. (See for example, Newman 1989, Atwood et al. 1998, Carnes and Guffey 2000, Phillips 2003, and Powers et al. 2013.) However, the literature has been relatively quiet on the impacts of international taxation on an MNF’s global operations decisions. For example, almost all analytical models in the Operation Management literature that characterize operational behaviors use pretax financial objectives (typically cost minimization or pretax profit maximization).

In this paper, we examine the impact of international taxation on the optimal production quantity decision at an U.S. MNF’s overseas subsidiary. U.S.-based MNFs must pay taxes for their worldwide incomes at the U.S. (home country) tax rate. To avoid double taxation, Section 901 of the Internal Revenue Code permits foreign tax credits (FTCs) for the taxes an MNF has paid to foreign countries. Thus, a subsidiary located in a high-tax country (with a tax rate higher than that of its home country), which will have paid more taxes when its incomes are repatriated back to its parent company, will generate excess FTCs or excess credits (ECs). These excess credits can be used to offset the tax obligations for the incomes remitted from the global firm’s subsidiaries located in low-tax countries. This practice is referred to as tax cross-crediting (Scholes et al. (2009)). A global firm is said to be in an excess credit position (or binding foreign tax credits) if its total foreign tax credits are more than enough to offset its total foreign tax liabilities. Because the amount of excess credits that a global firm possesses determines the effective tax rate that is applied to the profits earned by the low-tax division, it is natural to expect that the firm’s excess credit position influences its operational decisions such as production decision (see Billings et al. (2009), Devereux and Griffith (1998), and Kemsley (1998) for empirical evidences).
We study the production quantity decision problem by expanding the traditional newsvendor model to include the after-tax profit maximization performance measure and tax cross-crediting consideration. Specifically, we characterize the optimal production quantity decision at an MNF’s subsidiary located in a low-tax country. Our study indicates that when facing tax cross-crediting opportunities, the production decision at the low-tax division can exhibit characteristics that are surprisingly different from that of the traditional inventory models. For example, conventional wisdom suggests that a production manager’s decision should be more aggressive, i.e., to stock more, as his product becomes more profitable. This intuition no longer holds when the division production manager is given the incentive to cross-credit a certain amount of ECs. We show that as the ECs increase, which increases the division’s after-tax profits, the division’s optimal production quantity does not necessarily increase. Instead, as the ECs increase from zero to a sufficiently high level, the optimal production quantity first decreases, then increases, and finally stays flat. We also show that as the product’s profit margin increases, the optimal production quantity does not necessarily increase. Instead, as the profit margin increases from zero to a sufficiently high level, the optimal production quantity first increases, then decreases, and finally increases.

Another surprising effect of tax cross-crediting on the production decision is that there is a range of modest levels of ECs within which the corresponding optimal production decision can be determined simply by matching the maximum tax liabilities with the existing ECs. In other words, the concern of tax planning mutes the concern of matching supply with uncertain demand, and becomes the sole determinant for the optimal production decision. Further, this range of ECs widens when the difference between the subsidiary’s local tax rate and the home country tax rate becomes larger.

Our characterization of the optimal production quantity has implications for how the MNF should use an internal managerial tax rate to induce its subsidiary to make production decision that will maximize the company’s worldwide after-tax expected profits. We show that it is possible for the parent company to use an optimal managerial rate in an after-tax performance measure to incentivize the division manager to choose globally optimal production quantity, but such an optimal managerial tax rate can be even larger than the home country tax rate (which is already higher than the local tax rate). We also compare three easily implementable and intuitive performance measures: namely, measures based on pretax, after local tax, and after home country tax incomes (see Phillips (2003) for a survey of the pretax and after-tax performance measures used by 206 U.S.-based MNFs). We show that if the existing ECs are low (high), then the MNF’s best choice among the three performance measures is the after home country (local) tax measure.
To sharpen our study’s insights on the impacts of tax planning on the MNF’s production decisions, our analysis will initially focus on a newsvendor-style model with the objective of maximizing the expected after-tax profits within a planning year. The choice of a single-period stochastic demand model is appropriate for the business setting in our study as tax planning and reporting in an MNF are typically conducted on annual basis. The requirement that the production decision must be made before realization of uncertain demand reflects an important feature of products that have short selling season (or life cycle) and relatively long production lead-time, such as fashion, toys, and high-tech hardware. We will then discuss the robustness of our results by extending the base model to a number of more general settings, including general profit functions, profit loss and excess credits carryback and carryforward, and endogenously generated ECs.

The remainder of this paper is organized as follows: Section 2 discusses the related literature. Section 3 describes the model. Section 4 characterizes the optimal production decision in a benchmark case with the objective of after-tax profit maximization, but without tax cross-crediting. Section 5 characterizes the optimal production decision with tax cross-crediting and offers some insights on the impacts of excess credits, profit margin, and tax rates on the optimal production decision. Section 6 addresses the issue of how the global firm can use internal managerial tax rates to effectively motivate the division manager to make desirable production decisions. Section 7 discusses extensions of the newsvendor model to a number of more general settings. Section 8 offers concluding remarks with some thoughts on future research.

2. Literature

Our work falls into the area of research that integrates international tax planning with global operational decisions. Many of the papers in this area focus on income shifting, a strategy used by MNFs to shift their incomes from high-tax divisions to subsidiaries in low-tax jurisdictions through transfer pricing (Scholes et al. (2009)). Some of these papers consider the integration of transfer pricing and operations decisions to improve the performance of an MNF’s global supply chain operations (Cohen et al. (1989), Kouvelis and Gutierrez (1997), Vidal and Goetschalckx (2001), Goetschalckx et al. (2002) and Miller and de Matta (2008)). The interactions of transfer pricing and production/distribution decisions are typically formulated as non-linear mathematical models, which are then often solved by heuristics.

Shunko et al. (2009) study a global firm that sells a product in its domestic market, but produces the product either at a subsidiary or an external manufacturer located in a foreign country. They investigate various centralized and decentralized production and pricing (including transfer pricing) strategies which balance the tradeoffs between tax and production costs.
We do not address the issue of income shifting. We assume that the low-tax subsidiary produces a single product locally (without capacity restrictions) to satisfy its local demands and that there are no inter-company (or cross-division) transactions between the global firm’s divisions. Thus, our paper has a distinct focus on the impacts of tax cross-crediting on the optimal production decision.

Some papers in the research area of tax and operations interface, including Munson and Rosenblatt (1997), Wilhelm et al. (2005), Li et al. (2007) and Hsu and Zhu (2011), study the impacts of indirect taxes, such as tariff and value added taxes, on operational decisions. These papers do not consider the interaction of incomes from high-tax and low-tax divisions of an MNF.

More generally, our paper is also related to study in the global supply chain literature that include after-tax profit maximization in their objectives (see Meixell and Gargeya (2005) for a comprehensive review). Most of these papers consider deterministic profits and, therefore, the optimal decisions often guarantee positive profitability. In their cases, the effects of taxes can often be simply treated as parts of the variable costs of the products. For example, they would add a percentage of a tariff to the cost of a product imported into a country, or subtract a percentage of the corporate taxes from the product made in a country. Furthermore, these papers do not consider tax cross-crediting.

3. The Models

Suppose that an MNF with home country tax rate $\tau_h > 0$ has a wholly-owned subsidiary incorporated in a low-tax foreign country with tax rate $\tau_l < \tau_h$. Following the typical single-period production quantity decision setting, we assume that the subsidiary sells a single product to its uncertain local market in a single selling season at a per unit selling price (or retail price) $p$. Due to the long production lead time, the production quantity $q$ has to be decided in advance of the selling season, at a per unit procurement cost (or wholesale price) $c$, where $p > c$. The market demand, denoted by $D$, is a random variable with probability density function $g(\cdot)$ and distribution function $G(\cdot)$ with the support $[0, +\infty)$. After the random demand $D$ is realized, the subsidiary collects pretax profits $R(q, D)$, where $R(q, D) = p \min(q; D) - cq$.

Ignoring tax issues, the traditional operations literature chooses a production quantity $q^0(0)$ that maximizes the expected pretax profits (with a tax rate of zero), i.e., $q^0(0) = \arg \max_q \{E_D R(q, D)\}$. However, when tax is considered, $q^0(0)$ may not coincide with the quantity that maximizes the firm’s expected after-tax profits. Suppose that a tax rate $\tau \geq 0$ applies to the firm’s pretax profits $R(q, D)$ whether it is positive or negative; that is, the firm is levied at the tax rate $\tau$ when it is profitable, and it receives a subsidy at the same rate $\tau$ for its losses. Then, the quantity $q^\tau(0)$,
which is optimal for the traditional model, also maximizes the expected after-tax profits because taxation only scales the profits (positive or negative) by 

\( (1 - \tau) \) and thus does not affect the optimal quantity. In practice, however, the tax rate for such a firm is \( \tau \) when it is profitable, but becomes zero when it incurs losses. Under this phenomenon of tax asymmetry (Eldor and Zilcha (2002)), the quantity that maximizes the expected after-tax profits is

\[
q^o(\tau) = \arg \max_q E_D[R(q, D) - \tau R^+(q, D)], \tag{1}
\]

where \( R^+(q, D) = \max\{0, R(q, D)\} \). Note that when considering the tax asymmetry based on local tax, the low-tax subsidiary should produce quantity \( q^o(\tau_l) \) that maximizes the subsidiary’s expected after-local-tax profits.

However, since \( q^o(\tau_l) \) disregards the global firm’s worldwide FTC planning opportunities, it may still be suboptimal for the MNF. Suppose that with its international tax planning the MNF has decided to remit back to the home country a certain amount of deferred repatriation income from its high-tax subsidiaries. Such a repatriation plan will generate a certain amount of ECs, denoted by \( C \), in the planning year. On the other hand, for any given quantity \( q \) and realized market demand \( D \), the low-tax division’s pretax profits are \( R(q, D) \) which attract local taxes of \( \tau_l R^+(q, D) \) and thus generate \( (\tau_h - \tau_l) R^+(q, D) \) tax liabilities. As a result of tax cross-crediting, the global firm can offset part (or all) of the tax liabilities up to the available ECs, \( C \).

Consequently, the global firm’s net repatriation taxes owed to its home country are \( (\tau_h - \tau_l) R^+(q, D) - \min(C, (\tau_h - \tau_l) R^+(q, D)) \). Letting \( \Pi(q, C) \) be the global firm’s expected worldwide after-tax profits, the MNF’s problem is

\[
\max_q \Pi(q, C) = \max_q E_D\{R(q, D) - \tau R^+(q, D) + \min(C, (\tau_h - \tau_l) R^+(q, D))\}. \tag{2}
\]

Let \( q^*(C) = \arg \max_q \Pi(q, C) \) be the global firm’s optimal production decision that considers both tax asymmetry and tax cross-crediting in the presence of market uncertainty.

Note that the subsidiary would typically conduct its businesses in its local currency. However, when remitting its incomes back to the parent company at the end of the year, these incomes need to be translated into home country currency using the exchange rate at the time of remittance (Yang and Jeffers (2008)). We assume that the global firm is able to hedge currency exchange risk through some derivative instruments such as forward contracts. Therefore, all monetary parameters

\[1\] We assume any leftover ECs after cross-crediting has zero value. Therefore, it is optimal to redeem ECs as much as possible to offset the tax liabilities. We will relax the assumption of zero salvage value in \( \S 7.2 \), and it remains to be an optimal strategy for the global firm to redeem ECs as early as possible (which is intuitive because a unit of EC redeemed later will not generate more value than by redeeming it now).
in this paper are expressed in the home country currency using a fixed exchange rate at the end of the planning year. In the conclusion, we will further discuss possible future research that will incorporate currency risk.

4. Optimal Production Decision without Tax Cross-Crediting

As a benchmark for later discussions, we will characterize the impact of tax asymmetry on the optimal production quantity decision $q^o(\tau)$ defined by (1) with the objective to maximize the expected after-tax profits. For fixed $D$, it is easy to verify that $R(q, D) - \tau R^+(q, D)$ is concave in $q$. Thus, the objective function of problem (1) is also concave in $q$. The concavity property implies that the optimal solution $q^o(\tau)$ can be obtained through its first-order condition, which will be derived next through the marginal cost analysis.

Given any production quantity $q$, if the realized demand exceeds the supply, i.e., $D > q$, the excess demand is lost and the firm loses the opportunity to make more profits because of insufficient supply. Thus, the marginal underage cost, i.e., the gain in the after-tax profits by satisfying a unit of lost demand, is equal to $(1 - \tau)(p - c)$ ($p > c$ to ensure profitability). Because underage occurs when demand exceeds supply, the expected marginal underage cost is $(1 - G(q))(1 - \tau)(p - c)$.

When the demand is lower than the supply, i.e., $D < q$, the leftover inventory has zero value and the firm incurs overstock cost. However, in contrast to the above result that the marginal understock cost is a constant and does not depend on how much excess demand is lost, the marginal overage cost depends on the demand realization and on whether or not the division is profitable. Specifically, two distinct cases arise. If $D < cq/p$, the firm’s sales revenue $pD$ is not even enough to recoup the procurement cost $cq$, resulting in net losses and thus zero tax payment. The marginal overstock cost is simply the unit procurement cost $c$. On the other hand, if $D \in [cq/p, q)$, the firm makes profits and pays taxes at the rate $\tau$. The firm’s marginal overstock cost in this case is $(1 - \tau)c$ because it receives a tax deduction $\tau c$ from the unit production cost $c$. Because these two cases occur with probability $G(cq/p)$ and $G(q) - G(cq/p)$, respectively, the expected marginal overstock cost is $G(cq/p)c + (G(q) - G(cq/p))(1 - \tau)c$.

Note that as $q$ increases from zero to positive infinity, the expected marginal understock cost decreases from $(1 - \tau)(p - c)$ to zero and the expected marginal overstock cost increases from zero to $c$. It follows from the marginal cost analysis that the firm’s optimal production quantity $q^o(\tau)$ is the unique solution to the equation

$$G(cq/p)c - \frac{\tau}{1 - \tau} + G(q)c = (1 - G(q))(p - c),$$

where the first term captures the effect of the tax asymmetry while the second and third terms are the typical newsvendor marginal costs.
Proposition 1. \( q^*(\tau) \) increases as \( \tau \) decreases, or \( c \) decreases, or \( p \) increases.

The first part of Proposition 1 indicates that the optimal production quantity decision \( q^*(\tau) \) with tax consideration is smaller than \( q^*(0) \), the optimal newsvendor quantity without considering tax; and the difference becomes more significant as the tax rate \( \tau \) increases. This distinctive result is caused by the interplay between demand uncertainty and a salient feature of the tax system whereby taxes are charged only when the firm makes nonnegative profits. Specifically, because of demand uncertainty, it is impossible to perfectly match supply with demand, resulting in two types of costs of supply/demand mismatch: overstock and understock costs. If the tax simply scales down both types of costs by a constant factor, then the optimal production quantity would not change. However, when understock occurs, the firm always makes profits and thus the understock cost is scaled down by the tax rate \( \tau \); in contrast, when overstock occurs, there is the possibility that the firm incurs losses in which case the scale-down of the overstock cost does not occur. In other words, the tax asymmetry effect mitigates the concern of understocking more than it does to the concern of overstocking, thus resulting in more conservative production quantity decisions than what would otherwise occur in the traditional newsvendor model.

The results in Proposition 1 also support the intuition, which holds for the traditional newsvendor model, that a production manager’s decision tends to become more aggressive (i.e., the optimal production quantity increases) in response to a change in business conditions which make his product more profitable; e.g., when the unit production cost decreases, or the unit selling price increases, or the tax rate decreases. We will show in the next section that this intuition no longer holds for the firm’s optimal production decision with tax cross-crediting.

5. Optimal Production Decision with Tax Cross-Crediting

This section studies the problem (2) where tax cross-crediting is considered. Based on Theorem 5.5 (pp.35) in Rockafellar (1970) that the pointwise infimum of an arbitrary collection of concave functions is also concave, the objective function of problem (2) is concave in \( q \) and therefore its optimum \( q^*(C) \) can be determined by its first-order condition.

However, characterizing the optimal production decision \( q^*(C) \) is no longer as straightforward as in the benchmark discussed in the previous section. With the addition of the home country tax rate \( \tau_h \) and the interplay of tax asymmetry and tax cross-crediting, some of the insights from the marginal cost analysis for the benchmark are no longer applicable. To see this, suppose that the division makes a production decision \( q \) and the realized demand exceeds the supply, i.e., \( D > q \). Then, the division’s pretax profits are \((p - c)q\), which incur \((\tau_h - \tau)(p - c)q\) tax liabilities. In
contrast to the benchmark model in which a single tax rate applies to any marginal profit, now the tax rate applied to the marginal profit depends on whether the available ECs can be fully utilized to offset the tax liabilities. In particular, if the ECs are in shortage, i.e., if \( C < (\tau_h - \tau_l)(p - c)q \), the marginal profit will be taxed at the home country tax rate \( \tau_h \). If, however, there are sufficient ECs to offset the tax liabilities, i.e., if \( C > (\tau_h - \tau_l)(p - c)q \), then the tax rate applied to the marginal profit becomes the local tax rate \( \tau_l \).

Defining \( q_1(C) = \frac{C}{(\tau_h - \tau_l)(p - c)} \), three regions of production quantity \( q \) emerge from the observations above: \( q < q_1(C) \), \( q > q_1(C) \), and \( q = q_1(C) \), which are labeled as Regions I, II, and III, respectively. We will show in this section that there are threshold levels of ECs, within which the optimal production quantity \( q^*(C) \) will fall into one of these three regions. Furthermore, the optimal production decision in each of these regions takes very different forms and behaves quite differently in response to a change in some of the firm’s business environments. In the next subsection, we will first characterize the optimal production decision in each of the three regions in sequel. Then in the second subsection, we will discuss the impacts of a few key business factors such as ECs, tax rates, and prices on the optimal production decision.

5.1. Characterization of Optimal Production Decision

**Region I.** \( q < q_1(C) \).

Suppose that the optimal production quantity \( q^*(C) \) satisfies \( q^*(C) < q_1(C) \), which implies that the available ECs are sufficiently large so that they will always exceed the tax liabilities generated by the division even if the entire production quantity \( q^*(C) \) is sold out. Thus, the repatriation tax \( \tau_h - \tau_l \) generated by one dollar of local profit can be fully offset by the ECs, resulting in the local country tax rate \( \tau_l \) being applied to every dollar of profits generated by the division. As a result, \( \Pi(q, L) \) can be rewritten as

\[
\Pi(q, L) = ED\{R(q, D) - \tau_l R^+(q, D)\}.
\]

It is therefore intuitive to expect the global firm to choose \( q^*(L) = q^*(\tau_l) \), the production quantity which maximizes the expected after-local-tax profits. The remaining question is under what conditions will \( q^*(\tau_l) \) fall into Region I, i.e., \( q^*(\tau_l) < q_1(C) \). The answer follows directly from the definition of \( q_1(C) \): \( q^*(\tau_l) < q_1(C) \) if and only if \( C > \overline{C} \), where

\[
\overline{C} = (\tau_h - \tau_l)(p - c)q^*(\tau_l).
\]  

The above discussions lead to the following result.
**Proposition 2.** The firm’s optimal production quantity $q^*(C)$ lies in Region I if and only if $C > C$. When it does, $q^*(C) = q^*(\tau_1)$.

The necessary and sufficient condition in Proposition 2 further suggests that when $C$ is no larger than $C$ (i.e., $C \leq C$), any quantity in Region I, i.e., $q < q_1(C)$, cannot be optimal. The rationale for this claim can be explained as follows. At any quantity $q < q_1(C)$, the firm will never be able to fully use the ECs $C$ and thus always enjoys the lower tax rate $l$, in which case the critical quantity that balances the two types of supply/demand mismatch cost is $q^*(\tau_1)$. By the definition of $C$ in (4), the condition $C \leq C$ can be alternatively expressed as $q_1(C) < q^*(\tau_1)$, implying that producing $q_1(C)$ is already too conservative with respect to the critical quantity $q^*(\tau_1)$. Producing any quantity $q < q_1(C)$ will be even more conservative and therefore suboptimal.

**Region II.** $q > q_1(C)$.

For any given production quantity $q > q_1(C)$, if the realized demand exceeds the supply ($D > q$), the tax liabilities $(\tau_h - \tau_1)(p - c)q$ generated by the division are more than enough to redeem the entire ECs $C$. Thus, the marginal gain in pretax profits by selling an additional unit of the product is subject to the home country tax rate $h$, implying that the marginal understock cost is $(1 - h)(p - c)$. Consequently, the expected marginal understock cost is

$$c_o(q, C) = (1 - \tau_h)(p - c)(1 - G(q)).$$

When the supply $q$ exceeds the demand $D$, the pretax profits are $pD - cq$. Three distinct scenarios emerge depending on the level of $D$. First, if $pD - cq \leq 0$, or equivalently, $D \leq cq/p$, then the division earns no profit and pays no tax. The marginal overstock cost is $c$. Second, if $pD - cq > 0$ and the resulting tax liabilities $(\tau_h - \tau_1)(pD - cq) \leq C$, or equivalently, $D \in (cq/p, C/(p(\tau_h - \tau_1)) + cq/p]$, then the division earns positive profits and the tax liabilities can be fully offset by the available ECs $C$. This suggests that the local country tax rate $\tau_l$ applies to all profits. Similar to the marginal cost analysis for the benchmark model, in this case an overstocked unit will incur a cost $c$ but gain $\tau_l c$ through tax deduction, resulting in a marginal overstock cost $(1 - \tau_l)c$. Third, if $pD - cq > 0$ and $(\tau_h - \tau_1)(pD - cq) > C$, or equivalently, $D \in (C/(p(\tau_h - \tau_l)) + cq/p, q)$, then the division earns positive profits and the ECs are insufficient to fully offset the tax liabilities. This suggests that the home tax rate $\tau_h$ applies to the marginal profit and thus the marginal overstock cost is $(1 - \tau_h)c$. Combining these three scenarios, the expected marginal overstock cost is

$$c_o(q, C) = cG(cq/p) + c(1 - \tau_l)(G(C/(p(\tau_h - \tau_l)) + cq/p) - G(cq/p)) + c(1 - \tau_h)(G(q) - G(C/(p(\tau_h - \tau_l)) + cq/p)).$$
Let $q_2(C)$ be the critical quantity at which the expected marginal overstock costs equal understock costs. That is, $q_2(C)$ is the solution to equation $c_u(q, C) = c_o(q, C)$. We conclude that if the optimal production quantity falls into Region II, then $q^*(C) = q_2(C)$. The next question is under what conditions $q_2(C)$ exists in Region II, i.e., $q_2(C) > q_1(C)$. To this end, let $\hat{q}$ be the solution to the equation

$$cG(cq/p) + c(1 - \tau_i)(G(q) - G(cq/p)) = (p - c)(1 - \tau_h)(1 - G(q))$$

and define $\hat{C} = (p - c)(\tau_h - \tau_i)\hat{q}$. It can be verified that $q_1(\hat{C}) = \hat{q}$ and $q_2(\hat{C}) = \hat{q}$, implying that $q_2(\hat{C}) = q_2(C)$. With the additional observation from the definition that $q_1(C)$ increases in $C$ and $q_2(C)$ (if exists) decreases in $C$ (because $c_o(q, C)$ is increasing in $C$, but $c_u(q, C)$ is invariant to $C$), we conclude that for any $C < \hat{C}$

$$q_2(C) > q_2(\hat{C}) = q_1(\hat{C}) > q_1(C).$$

The above arguments lead to the following characterization of the condition under which the optimal production decision lies in Region II.

**Proposition 3.** The firm’s optimal production quantity $q^*(C)$ lies in Region II if and only if $C < \hat{C}$. When it does, $q^*(C) = q_2(C)$.

Proposition 3 implies that if the ECs are sufficiently low ($C < \hat{C}$), then the global tax consideration transforms the global firm’s production decision problem into one with three distinct tax rates $\tau_i$, $\tau_h$, and 0, depending on the realization of the demand. Specifically, the higher tax rate $\tau_h$ applies to any marginal gain when understock occurs because the tax liabilities exceed the available ECs. When overstock occurs, as the realized demand increases, the tax rate applied to the marginal profit changes from 0 to $\tau_i$ and then to $\tau_h$ as the division’s profits increase from negative to positive, or equivalently, as the tax liabilities increase from zero to an amount below $C$, and then above $\hat{C}$.

**Region III.** $q = q_1(C)$.

When the conditions in Propositions 2 and 3 do not hold, i.e., when $C \in [\hat{C}, \overline{C}]$, the optimal production quantity can only lie in Region III. This is summarized in the following proposition, which completes the characterization of $q^*(C)$.

**Proposition 4.** The firm’s optimal production quantity $q^*(C)$ lies in Region III if and only if $C \in [\hat{C}, \overline{C}]$. When it does, $q^*(C) = q_1(C)$. 
5.2. Properties of Optimal Production Decision

In this subsection we investigate how the firm’s optimal production decision responds to a change in the firm’s business environments. Specifically, we study how three groups of factors, namely, the available ECs $C$, the profit margin $p - c$, and the tax rates $\tau_h$ and $\tau_l$, which are the main drivers of the MNF’s global after-tax profitability, affect the optimal production decision. The following proposition, which describes the impacts of $C$ on the optimal production decision, follows directly from the analysis in the previous subsection.

**Proposition 5.** As $C$ increases over $[0, \hat{C}]$, $q^*(C) = q_2(C)$ decreases from $q^*(\tau_h)$ to $\hat{q}$; as $C$ increases over $[\hat{C}, \bar{C}]$, $q^*(C) = q_1(C)$ increases from $\hat{q}$ to $q^*(\tau_l)$ at a constant rate $\frac{1}{(\tau_h - \tau_l)(p - c)}$; as $C$ increases over $[\bar{C}, +\infty)$, $q^*(C)$ stays unchanged at $q^*(\tau_l)$.

When $C = 0$, no tax cross-crediting is possible and the firm only needs to be concerned with the tax asymmetry effect. Thus, the optimal quantity $q^*(C)$ is $q^*(\tau_h)$, the optimal quantity that maximizes the expected after-tax profits with tax rate $\tau_h$. Intuitively, as the ECs $C$ increase, the pretax profits generated from the division should become more valuable for the global firm because more of the resulting tax liabilities owed to the home country can be offset by the increased ECs. However, Proposition 5 indicates that this enhanced value of pretax profits does not necessarily mean that the firm should be more aggressive in its production decision. Instead, the optimal production quantity decreases in $C$ when $C$ is small ($C \in [0, \hat{C}]$).

This result can be explained as follows. For a small value of $C$, the optimal production quantity lies in Region II so that the tax liabilities in the event of stockout (i.e., $D > q$) are more than enough to fully redeem the ECs. Thus, an increase of $C$ does not affect the marginal understock cost. On the other hand, the earlier marginal cost analysis in Region II indicates that an increase of $C$ expands the regime of demand realization under which the ECs can not be fully utilized, but shrinks the regime of demand realization under which the ECs are fully utilized. Note that the marginal overstock cost in these two regimes are $(1 - \tau_l)c$ and $(1 - \tau_h)c$, respectively. Hence, an increase of $C$ drives up the overstock cost but does not affect the understock cost, thereby pushing the optimal production quantity downward.

When ECs are within a range of modest levels ($C \in [\hat{C}, \bar{C}]$), Proposition 5 suggests that the global firm’s optimal production decision is extremely simple: produce an amount so that the maximum tax liabilities that the division can possibly generate (in the event of soldout) match exactly the ECs $C$. Interestingly, this optimal production quantity is independent of the demand distribution. This is in a stark contrast with the well-known result from the classical newsvendor
model that knowledge about the demand distribution is crucial in making optimal production decisions. Furthermore, note that with fixed $\tau_h$, $\bar{C} - \hat{C}$ increases as $\tau_i$ decreases. Thus, the simple production decision above is more likely to be optimal when the two tax rates are further apart from each other.

When $C$ becomes sufficiently large; that is, when $C \geq \bar{C}$, more ECs have no impact on the optimal production quantity because the tax liabilities generated from any demand realization can be fully offset by the ECs. Thus, all incomes enjoy the lower tax rate $\tau_i$, so $q^*(C) = q^0(\tau_i)$.

For a given $C > 0$, we now turn to the sensitivity of the optimal production quantity $q^*(C)$ with respect to a change in the profit margin $p - c$.

**Proposition 6.** With fixed $c$, as the retail price $p$ increases from $c$ to $+\infty$, $q^*(C)$ first increases, then decreases, and finally increases.

Note that for a fixed $c$, as $p$ increases from $c$ to $+\infty$, the profit margin $p - c$ increases from 0 to $+\infty$. Conventional wisdom suggests that such steady increase of profit margin will lead to a monotonic increase of the optimal production quantity. Proposition 6 shows that this intuition, true for the traditional newsvendor model and even for the benchmark model studied in Section 4 (see Proposition 1), is no longer valid for the model with tax cross-crediting consideration. This counter-intuitive result can be explained as follows. As the profit margin increases, all else being equal, the division’s pretax profits improve and so do the resulting tax liabilities. Thus, the increase in profit margin has the opposite effect on the optimal production quantity as that of the increase in ECs. As a mirror image of the result that the optimal production quantity increases in $C$ for a modest range of $C$ values in $[\hat{C}, \bar{C}]$, Proposition 6 indicates that as the profit margin increases from 0 and upwards, the optimal production quantity $q^*(C)$ will at some point fall in Region III.

As suggested by Proposition 4, the global firm will source the quantity within this region with the goal of fully redeeming the given ECs $C$ in the event of stockout. As the profit margin continues to increase, a smaller production quantity is needed to accomplish this goal.

For a fixed $p$, the response of the optimal production quantity to a decrease in $c$ or an increase in $p - c$ is similar to that described in Proposition 6 except that the possible change of $c$ is between 0 and $p$. Thus, for a sufficiently large fixed $p$, a decrease in $c$ from $p$ to 0 will cause $q^*(C)$ to first increase, then decrease, and finally increase.

To conclude this subsection, we examine the impact of a change in the tax rates, $\tau_i$ or $\tau_h$, on the optimal production decision. We observe that for a large $C$, the optimal production quantity $q^*(C) = q^0(\tau_i)$, which is decreasing in $\tau_i$ but is independent of $\tau_h$. For a modest $C$, $q^*(C) = C/(p-$
\((\tau_h - \tau_l)\), which decreases as \(\tau_h\) increases or as \(\tau_l\) decreases. For a small \(C\), \(q^\ast(C)\) is not monotonic in \(\tau_l\) or \(\tau_h\); it may either increase or decrease as any of the tax rate increases.

In particular, it is possible that \(q^\ast(C)\) may increase as \(\tau_l\) increases, a result that is counter-intuitive and distinct from the benchmark model presented in Section 4. The intuition is as follows. Recall from our discussion of Region II in Section 5.1, that for a small \(C\), as \(\tau_l\) increases, it becomes more likely that the ECs will be fully used in the event of overstock, resulting in a greater chance of incurring the lower marginal overstock cost \((1 - \tau_h)c\). Therefore, the tax cross-crediting effect tends to move the production quantity upward as \(\tau_l\) increases. However, as in the benchmark model, there is the loss-of-profit-margin effect that moves the production quantity downward as \(\tau_l\) increases. When the tax cross-crediting effect dominates the loss-of-profit-margin effect, the optimal production quantity increases as \(\tau_l\) increases.

6. Managerial Tax Rates

It is well-known that most MNFs prefer to delegate local production decisions to local business units through certain decentralized control mechanisms with pretax or after-tax performance measures (see for example, Wilson (1993), Phillips (2003)). These MNFs would be interested in designing a performance measure to induce local production decisions that will maximize the global firms’ expected worldwide after-tax profits. However, according to Wilson (1993), one of the challenges is how to determine a “correct tax rate” for setting after-tax performance measures. In the context of our modeling framework, we assume that the global firm will evaluate the subsidiary’s performance using an expected after-tax profit maximization measure with a managerial tax rate \(\hat{\tau}\), which may or may not be the same as the division’s local tax rate. We will first characterize the optimal managerial tax rate and then discuss the effectiveness of a few simpler pretax and after-tax measures.

Recall from discussions of the benchmark model in Section 4 that for a managerial tax rate \(\hat{\tau}\), the division’s best production quantity is \(q^\circ(\hat{\tau})\). Because \(q^\circ(\hat{\tau})\) decreases in \(\hat{\tau}\), the inverse function \(q^{-1}(\cdot)\) is well defined. Thus, by internally communicating the optimal managerial tax rate \(\hat{\tau}^\ast(C) \equiv q^{-1}(q^\ast(C))\) to the division, the production quantity that the division manager chooses to maximize its own “after-tax incomes” based on the benchmark model coincides with the globally optimal production quantity \(q^\ast(C)\). The following results follow immediately from Proposition 5.

\textbf{Proposition 7.} For \(C = 0\), \(\hat{\tau}^\ast(C) = \tau_h\); for \(C \in [0,\bar{C}]\), \(\hat{\tau}^\ast(C)\) increases in \(C\); for \(C \in [\bar{C},\overline{C}]\), \(\hat{\tau}^\ast(C)\) decreases in \(C\); for \(C \geq \overline{C}\), \(\hat{\tau}^\ast(C) = \tau_l\).
For every dollar of pretax profits generated by the division, the division pays $\tau_l$ local taxes, which generate $\tau_h - \tau_l$ tax liabilities that can be partially or fully offset by ECs. So the effective tax rate (defined as the ratio of total tax expense to the pretax incomes) falls somewhere between $\tau_h$ and $\tau_l$. Thus, because of tax cross-crediting, it seems intuitive to expect that the optimal managerial tax rate should fall somewhere between $\tau_l$ and $\tau_h$. However, Proposition 7 implies that the optimal managerial tax rate $\tau^*(C)$ could be set even higher than the home country tax rate $\tau_h$. This result suggests that in designing the optimal managerial tax rate for the division, the firm ought not to look for the effective tax rate which reflects the after-tax contribution of the division’s pretax incomes. Instead, the global firm may even need to impose a managerial tax rate higher than the home country tax rate, which is already high compared with the local tax rate.

The difficulty of determining the “correct” managerial tax rates, the concerns over ease of communication, and the perceived fairness are perhaps some of the reasons that many global firms resort to adopting simpler and more intuitive managerial tax rates to facilitate production decisions. Of the 209 MNFs surveyed in a study by Phillips (2003), 143 used pretax measures and 66 used after-tax measures to evaluate their division managers’ performances. Phillips (2003) further identifies three types of after-tax performance measures used by MNFs in practice. The first type would extensively allocate the company’s total tax liabilities among business units. Our optimal managerial rate may serve well in such an active coordination scheme. The second type attributes only local taxes to a local division. Setting managerial tax rate to the local tax rate would be a good measure. The third type uses a company-wide percentage (e.g., effective tax rate) to allocate tax expenses among business units. In our problem, the effective tax rate for the low-tax division is between the two tax rates $\tau_h$ and $\tau_l$. In particular, when the tax liabilities generated by the division are more than enough to redeem the available ECs, the effective tax rate for the division is the home country tax rate $\tau_h$.

Based on the discussions above, we investigate in the remainder of this section the effectiveness of the three easily implementable performance measures, namely, the pretax measure (which sets the managerial tax rate to zero) and the two after-tax measures using the local tax rate $\tau_l$ and the home country tax rate $\tau_h$ respectively. Recall that the division’s optimal production quantities are $q^o(0)$, $q^o(\tau_l)$, and $q^o(\tau_h)$, respectively, under these three measures. Since the tax asymmetry effect becomes more salient as the tax rate increases, $q^o(0) \geq q^o(\tau_l) \geq q^o(\tau_h)$ (see Proposition 1). Let $\Pi_0$, $\Pi_l$, and $\Pi_h$ be the corresponding expected after-tax profits of the global firm, i.e., $\Pi_0 = \Pi(q^o(0), C)$, $\Pi_l = \Pi(q^o(\tau_l), C)$, and $\Pi_h = \Pi(q^o(\tau_h), C)$, where $\Pi(q, C)$ is defined by (2) in Section 3.
Proposition 8. There exists a threshold $\tilde{C} \in [\tilde{C}, C]$ such that $\Pi_h \geq \Pi_l \geq \Pi_0$ for $C \in [0, \tilde{C}]$ and $\Pi_l \geq \max\{\Pi_h, \Pi_0\}$ for $C \geq \tilde{C}$.

Several observations are noteworthy. First, the pretax measure is dominated by the other two after-tax measures. Second, if the ECs are less than a threshold value, then the after-home–country-tax measure is the best; otherwise, the after-local-tax measure prevails. By Proposition 5, when $C$ is small ($C \in [0, \tilde{C}]$), the firm’s optimal production quantity is even lower than $q^o(\tau_h)$, the smallest quantity among those induced by the three measures above. Thus, all of the three measures lead to overstocking in the production decision, but the extent of overstocking is ascending in the order of after-home–country-tax, after-local-tax, and pretax measure. Consequently, the after-home-country-tax measure is the best, whereas the pretax measure is the worst. However, once $C$ exceeds $\tilde{C}$, the firm’s optimal production quantity increases as $C$ increases. That is, it will exceed $q^o(\tau_h)$ and approach $q^o(\tau_l)$, where the after-home–country-tax measure leads to understocking and the other two measures result in overstocking in the division’s production decision. As $C$ continues to increase, the extent of the understocking under the after-home–country-tax measure becomes more pronounced. Since the after-local-tax measure ends up stocking more than the after-home-country-tax measure, the former performs better than the latter.

Based on results from Propositions 7 and 8, we can make the following suggestions to an MNF that uses a managerial tax rate to coordinate the production decisions at its high-tax division. When the available ECs are low and therefore the global firm anticipates its tax liabilities will exceed ECs, allowing the division to maximize its own after-home–country-tax profits is an effective measure. However, because the home-country tax rate $\tau_h$ prompts the division’s manager to stock more than the optimal quantity, the global firm may consider making some additional downward adjustments of the division’s inventory. Such adjustments should become more aggressive as $C$ increases.

On the other extreme, when the given ECs are sufficiently high and are expected to be more than the tax liabilities generated at the low-tax division, the parent company should use its local tax rate to evaluate the division’s after-tax profitability. Similarly, because such a policy tends to induce the division manager to stock more than the optimal quantity, additional downward adjustment should be made to compensate for the suboptimality; and such adjustments should be increasingly less aggressive as $C$ increases.

When $C$ approaches the threshold value $\tilde{C}$, neither policy (based on $\tau_l$ or $\tau_h$) will work very well without proper adjustments, suggesting that heavier coordination efforts are needed or even that the optimal managerial tax rate should be adopted.
7. Robustness Results

In this section, we discuss a number of extensions of our base model. In §7.1, we generalize the newsvendor profit function in the base model to a general profit function satisfying several mild and reasonable assumptions. In §7.2, we relax the assumption that the leftover ECs have zero salvage value, and consider the single-period setting with exogenous (positive) value of leftover ECs and then the multi-period setting with endogenous value of leftover ECs allowing both ECs carryback and carryforward. Similar conclusions hold for the loss carryback and carryforward. In §7.3, we extend our single-division problem with exogenous ECs to a two-division problem where the ECs are endogenously determined by the profits generated from the division in the high-tax country.

7.1. General Profit Function

This subsection generalizes the newsvendor profit function to a general profit function. We show that almost all results and insights established earlier for the base model still hold under this general model, demonstrating that these results and their implied managerial insights are not driven by the newsvendor profit function. However, as will be seen in our analysis in this section, the marginal cost analysis in §5 is no longer effective for the generalized problem with a general profit function. A new set of technically non-trivial analysis along with some new insights will be developed.

Consider a general function \( R(Q;S) \), where \( Q \) is the production quantity and \( S \) represents the uncertain market condition. Let \( s \) be the realized value of \( S \). Let \( R_1(Q;s) = \frac{\partial R(Q;s)}{\partial Q} \); \( R_{11}(Q;s) = \frac{\partial^2 R(Q;s)}{\partial Q^2} \); \( R_{12}(Q;s) = \frac{\partial^2 R(Q;s)}{\partial Q \partial s} \); \( R_2(Q;s) = \frac{\partial R(Q;s)}{\partial s} \). We make the following mild assumptions.

(A1): \( R(Q;s) \) is concave in \( Q \) for any \( s \in [\underline{s}, \overline{s}] \), i.e., \( R_{11}(Q;s) \leq 0 \);
(A2): \( R(Q;s) \) has increasing difference in \( Q \) and \( s \), i.e., \( R_{12}(Q;s) \geq 0 \);
(A3): \( R(Q;s) \) increases in \( s \), i.e., \( R_2(Q;s) \geq 0 \);
(A4): \( R(0,s) = 0 \) for any \( s \in [\underline{s}, \overline{s}] \);
(A5): \( R(Q,s) \leq 0 \) for any \( Q \geq 0 \).

We can re-write the objective function of our base model (2) in Section 3 as
\[
\Pi(Q) = E_S\{\min(R(Q,S) - \tau_1 R^+(Q_S), R(Q,S) - \tau_0 R^+(Q,S) + C)\}. \tag{5}
\]
The global firm’s optimal production decision is \( Q^* = \arg \max_Q \Pi(Q) \).

As implied by (5), \( \Pi(Q) \) is bounded above by \( E_S\{R(Q,S) - \tau_1 R^+(Q,S)\} \) for any \( Q \). Therefore, assumption (A1) implies that \( E_S\{R(Q^*(\tau_1),S) - \tau_1 R^+(Q^*(\tau_1),S)\} \) is an upper bound on \( \Pi(Q) \) for any \( Q \), where \( Q^*(\tau) = \arg \max_Q E_S[R(Q,S) - \tau R^+(Q,S)] \). Define \( C = (\tau_0 - \tau_1) R^+(Q^*(\tau_1),\overline{s}) \). If \( C \geq \overline{C} \), then assumption (A3) suggests that the tax liabilities, \( (\tau_0 - \tau_1) R^+(Q^*(\tau_1),s) \), generated by
the division under the production quantity $Q^s(\tau_i)$ never exceed $C$ for any realization $s$. Hence, 
$\Pi(Q^s(\tau_i)) = E_S\{R(Q^s(\tau_i), S) - \tau_1 R^+(Q^s(\tau_h), S)\}$, attaining the upper bound of $\Pi(Q)$. This leads to the following result.

**Lemma 1.** If $C \geq \bar{C}$, then the optimal production quantity that maximizes the global firm’s expected after-tax profits is $Q^* = Q^s(\tau_i)$.

Now we turn to the case in which $C < \bar{C}$. We first summarize a property of the optimal production quantity $Q^*$.

**Lemma 2.** If $C < \bar{C}$, $(\tau_h - \tau_i) R^+(Q^*, \bar{s}) \geq C$.

By Lemma 2, without loss of generality we can focus on the set of $Q$ that satisfies $(\tau_h - \tau_i) R^+(Q, \bar{s}) \geq C$. Thus, $R^+(Q, \bar{s}) = 0$ (see (A5)) and the monotonically increasing property of $R(Q, s)$ in $s$ (see (A3)) yield the result that for any $Q$ satisfying $(\tau_h - \tau_i) R^+(Q, \bar{s}) \geq C$, there exists a cutoff value of the realized market condition $s \in [\underline{s}, \bar{s}]$ under which the tax liabilities generated at $S = s$ equal the ECs $C$, i.e., $(\tau_h - \tau_i) R^+(Q, s) = C$. Further, for $S < s$, the tax liabilities are not enough to fully redeem the ECs, so the firm’s after-tax profits are $R(Q, s) - \tau_1 R^+(Q, s)$. For $S > s$, the tax liabilities generated exceed the ECs, so the firm’s after-tax profits are $R(Q, s) - \tau_h R^+(Q, s) + c$. Hence, the global firm’s production problem can be rewritten as $\max_Q J(Q, s(Q))$, where for $s \in [\underline{s}, \bar{s}]$,

$$J(Q, s) = \int_{\underline{s}}^{\bar{s}} [R(Q, x) - \tau_i R^+(Q, x) + C] g(x) dx + \int_{\underline{s}}^{\bar{s}} [R(Q, x) - \tau_1 R^+(Q, x)] g(x) dx,$$

and $s(Q)$ satisfies

$$(\tau_h - \tau_i) R^+(Q, s(Q)) = C.$$

Let $\tilde{Q}(s) = \arg \max_Q J(Q, s)$ for $s \in [\underline{s}, \bar{s}]$, where the minimum is chosen in the case of multiple maximizers. Finally, for a given $C \geq 0$, let $s^*(C) = \min \{ s \mid (\tau_h - \tau_i) R(\tilde{Q}(s), s) = C \}$. We now characterize the optimal production quantity decision $Q^*$ in terms of $\tilde{Q}(\cdot)$.

**Proposition 9.** For $C = 0$, $Q^* = Q^s(\tau_h)$; for $C \in [0, \bar{C}]$, $Q^* = \tilde{Q}(s^*(C))$; for $C \geq \bar{C}$, $Q^* = Q^s(\tau_i)$.

From Proposition 9, we see that the optimal production decision $Q^*$ under tax cross-crediting with given ECs $C$ maximizes the weighted average of two types of incomes, which, if positive, are taxed at two distinct tax rates of $\tau_h$ and $\tau_i$, respectively. Specifically, when $C = 0$, no tax cross-crediting is possible; the firm only needs to concern itself with the tax asymmetry effect. Thus, the optimal quantity $Q^*$ is $Q^s(\tau_h)$, the optimal quantity that maximizes the expected after-tax profits with tax rate $\tau_h$. As $C$ increases, those incomes generated under market conditions $s \in (0, s^*(C)]$, 


which are taxed at the home country tax rate \( \tau_h \), carry increasing weight in the firm’s overall expected after-tax profits.

Define \( \hat{Q}(s) = \arg\max Q R(Q, s) \) for \( s \in [\underline{s}, \bar{s}] \). Although the optimal production decision needs to be made ex-ante (i.e., before observing the market condition), the ex-post optimal production quantity \( \hat{Q}(s) \) can be seen as the ideal production quantity under a certain market condition \( s \).

**Lemma 3.** (a) \( \hat{Q}(s) \) increases in \( s \);
(b) There exists \( s_1 \in [\underline{s}, \bar{s}] \) such that \( \hat{Q}(s_1) = \bar{Q}(s_1) \);
(c) \( \hat{Q}(s) < \bar{Q}(s) \) for \( s < s_1 \), and \( \hat{Q}(s) \geq \bar{Q}(s) \) for \( s \geq s_1 \).

Let \( s_1 \) be the minimum solution that satisfies \( \bar{Q}(s_1) = \hat{Q}(s_1) \) and define \( \bar{C} = (\tau_h - \tau_l)R(\hat{Q}(s_1), s_1) \). We have the following results:

**Proposition 10.** \( Q^* = \bar{Q}(s^*(C)) \) decreases from \( Q^0(\tau_h) \) in \( C \) for \( C \in [0, \bar{C}] \) and then increases to \( Q^0(\tau_l) \) in \( C \) for \( L \in [\bar{C}, \bar{C}] \).

Recall in our earlier discussion that the incomes generated under a group of market conditions \( s \in (0, s^*(C)] \) are taxed at the local tax rate \( \tau_l \). As the firm’s ECs increase incrementally from \( C \) to \( C + \Delta \), this group of lower-taxed market conditions, denoted as \( \Omega(C) \equiv (0, s^*(C)] \), expands to \( \Omega(C + \Delta) \). We first observe from Lemma 3 that if \( s^*(C) < s_1 \), then \( \hat{Q}(s^*(C)) < \bar{Q}(s^*(C)) \). Thus, as \( C \) increases from zero incrementally, we expect \( \hat{Q}(s) < \bar{Q}(s^*(C)) \) for any market condition \( s \in (C, C + \Delta] \) which emerges in the expanded group. In other words, the ideal production quantities under these newly emerged market conditions are lower than the original optimal production quantity \( \bar{Q}(s^*(C)) \).

We observe further that the marginal after-tax profit is higher under the newly emerged market conditions \( s \in (C, C + \Delta] \) when compared with those outside the group \( \Omega(C + \Delta) \). Therefore, when deciding the optimal production quantity that maximizes its expected after-tax profits, the firm should respond more favorably to these newly emerged market conditions. That is, the optimal production quantity should tilt toward the ideal production quantities \( \hat{Q}(s), s \in (C, C + \Delta] \), which, according to an earlier observation, are lower than \( \bar{Q}(s^*(C)) \). As a result, \( Q^* \) decreases in \( C \) for \( C \in [\underline{C}, \bar{C}] \).

However, once the ECs \( C \) exceed the threshold \( \bar{C} \), any further increase of \( C \) expands the lower-taxed group of market conditions beyond \( s_1 \). In other words, we have \( s > s_1 \) and, based on earlier observation, \( \hat{Q}(s) > \bar{Q}(s^*(C)) \) for all \( s \in (C, C + \Delta] \). Thus, the incremental increase of \( C \) will push the optimal production quantity upward towards the ideal production quantities for newly
emerged market conditions in the expanded lower-taxed group. Consequently, $Q^*$ increases in $C$ for $C \in [\hat{C}, \overline{C}]$.

When $C$ reaches $\overline{C}$, the lower-taxed group $\Omega(C)$ expands to the entire range of market conditions, i.e., $[s, \overline{s}]$. As a result, the division’s pretax profits are subject to the local tax rate $\tau_l$. Consequently, the optimal quantity $Q^*$ stays at $Q^*(\tau_l)$, the optimal quantity that maximizes the expected after-tax profits with tax rate $\tau_l$.

Finally, all of the results in §6 about the effectiveness of the managerial tax rates still hold in the generalized model with the general profit function studied in this subsection.

7.2. Loss and Excess Credits Carryback and Carryforward

Many countries’ tax laws permit businesses to carry their losses and ECs backward and/or forward for a number of years. For example, the U.S. permits its MNFs to carry ECs backward for up to 1 year and forward for up to 10 years (ECs carryback and carryforward). To make our analysis more concise, our base model assumes that the global firm is interested in maximizing its worldwide after-tax earnings within a tax year. Such a performance measure is consistent with the current U.S. General Accepted Accounting Principals (GAAP)’s earning report requirements and is used by many companies, (e.g., Powers et al. 2013 identify 905 firms that pay annual cash bonuses to their CEOs based on after-tax earnings), but it ignores the effects of possible loss or EC carryback and carryforward. This restriction will be relaxed in this subsection.

Loss and EC carryforwards are considered part of a company’s deferred tax assets (DTAs). If these DTAs are recognized at their full values in our base model, then the optimal production decision is identical to that of the traditional newsvendor model. However, this is usually not the case in practice due to uncertainties of these DTAs’ realization and their potential loss of values in the future. Since DTAs reflect potential tax savings arising from future tax deductions, GAAP rule requires that a firm only includes DTAs on its balance sheets when it has available evidence to show that this portion of DTAs are more likely than not (a likelihood of more than 50%) to be realized (Petree, Gregory and Vitray 1995). There are strong evidences in practice that many companies will not be able to use up all DTAs and therefore tend not to recognize part of their DTAs in their balance sheets on the grounds of managerial prudence. For example, a survey by Cooper and Knittel (2006) finds that over 10 year window, about 40-50% net operating loss, a form of DTAs, are used as loss carryforward deduction, but 25-30% are lost (the firm no longer exists) and the other 10-20 remain unused.²

² There are also anecdotal evidences that firms apply valuation allowances to reduce their DTAs. For example, Booz...
In what follows, we show the robustness of our main results by studying two distinct approaches to recognizing partial values of DTAs. The first approach, which is aligned with accounting literature, imposes an exogenous value function to recognize values of DTAs still within a single-period model. The second approach extends the single-period model into a multi-period model where the value of DTAs in each tax year is determined endogenously by solving the MNF’s optimal operational decisions in the remaining planning horizon.

**Exogenous Value**

We discuss two methods suggested in accounting literature to recognize partial values of DTAs in a single period. The first method follows the U.S. Financial Accounting Standards (FAS) which requires a company to reduce its DTAs by a certain amount of *valuation allowance* if there is a more than 50% chance that all or part of the DTAs will not be realized in the future. Following De Waegenaere, Sansing and Wielhouwer (2003), we can add to our base model two upper limits, denoted by $A$ and $B$, for loss and EC carryforwards respectively. The MNF’s objective in this case is to maximize its current year’s expected after-tax profits plus its carryforward DTAs reported on its balance sheet as per FAS requirements. The objective of our base model can be modified accordingly as follows:

$$
\Pi(Q, C) = E_S \{ R(Q, S) - \tau_h R^+(Q, S) + \tau_1 \min(A, R^-(Q, S)) + \min(C, (\tau_h - \tau_i) R^+(Q, S)) + \min(0, [C - (\tau_h - \tau_i) R^+(Q, S)]) \},
$$

where $R^-(Q, S) = -\min(R(Q, S), 0)$.

The second method, which is adopted by Eldor and Eilcha (2002) in their study of tax asymmetry, recognizes the partial value of a DAT through a discount rate. In the context of our study, we will use two discount rates, $\sigma_1$ and $\sigma_2$ ($0 \leq \sigma_1, \sigma_2 < 1$), for loss and EC carryforwards, respectively. The modified model, which includes the base model as a special case with zero discount rate, is described as follows. The MNF’s objective function now becomes

$$
\Pi(Q, C) = E_S \{ R(Q, S) - \tau_h R^+(Q, S) + \sigma_1 \tau_i R^-(q, S) + \min(C, (\tau_h - \tau_i) R^+(Q, S)) + \sigma_2 [C - (\tau_h - \tau_i) R^+(Q, S)]^+ \}.
$$

The same analysis developed in the previous subsection can be used to show that Proposition 9 and 10 continue to hold for the optimal production quantity under both methods.

Allen Hamilton Corp. determined that the $42,379,000 out of the $302,953,000 deferred tax assets fall below the fifty percent threshold and thus deducted the $42,379,000 of valuation allowance from its deferred tax assets in the fiscal year of 2011. (Booz Allen Hamilton Corp., Form 10-K, Annual Report to the SEC, March 31, 2011, p. F-23.) In 2010, American International Group, Inc. reports a gross deferred tax asset of $37.4 billion, but reduces that asset by a valuation allowance of $25.8 billion. The company explains that the valuation allowance is necessary because of “recent negative evidence of cumulative operating losses and a lack of predictable profits.” (American International Group, Inc. Form 10-K Annual Report to the SEC, December 31, 2010, p. 357.)
Endogenous Value

We extend our single-period model to a dynamic multi-period model where the value of leftover ECs is determined endogenously by solving the MNF’s operational problem in the remaining time periods. For expositional simplicity, we focus only on the ECs carryback and carryforward, by noting that loss carryback and carryforward can be incorporated into the model without loss of generality.

Consider a planning horizon consisting of $T$ time periods. Let $C_n$ be the ECs newly arising in period $n$ for $n = 1, 2, ..., T$. The sequence of events in period $n$ is as follows. First, the leftover ECs in the previous period are carried over to period $n$. We assume that all of the leftover ECs are still eligible to offset the tax liabilities generated in period $n$. This assumption holds if the entire planning horizon is short relative to the maximum amount of ECs carry-forward time (which is ten years). Naturally, the ECs $C_n$ newly arising in period $n$ are also eligible. Because any ECs can be carried back to at most one immediately preceding tax year, the ECs $C_{n+1}$ generated in period $n + 1$, but not those ECs generated after period $n + 1$, should also be counted into the total ECs that are eligible to offset the tax liability generated in period $n$. We denote the total ECs (i.e., the sum of the three types of aforementioned ECs that are eligible for offsetting the tax liabilities in the current period $n$) by $C$. Second, after observing $C$, the MNF decides the production quantity $Q$. Third, the market condition $S_n$ is realized, resulting in pre-tax profits $R(Q, S_n)$ and tax liabilities $(\tau_h - \tau_l)R^+(Q, S_n)$ which are offset up to $C$. Consequently, the total ECs that are eligible to offset the tax liability in period $n + 1$ become $(C - (\tau_h - \tau_l)R^+(Q, S_n)) + C_{n+2}$, where the ECs $C_{n+1}$ newly arising in period $n + 1$ are already counted into $C$ by first completely carrying $C_{n+1}$ back to period $n$ and then carrying any unused ECs forward to period $n + 1$. This allows us to formulate the MNF’s production quantity problem as the following dynamic programming problem:

$$V_n(C) = \max_Q E_{S_n, C_{n+2}} \{ R(Q, S_n) - \tau_l R^+(Q, S_n) - ((\tau_h - \tau_l)R^+(Q, S_n)) + V_{n+1}((C - (\tau_h - \tau_l)R^+(Q, S_n)) + C_{n+2}) \},$$

for $n = 1, 2, ..., T$, where $V_n(C)$ is the MNF’s optimal total expected after-tax profits from period $n$ onward given that the total ECs that are eligible to offset the tax liabilities in period $n$ are equal to $C$. Let $Q^*_n(C)$ be the maximizer. Instead of assuming an exogenous value for leftover ECs as in

---

3 Because there is a possibility that any unit of EC, if carried over into the future period, may not always be redeemed due to insufficient tax liabilities, redeeming it in the earliest time always produces higher value than carrying it forward for the future use. Therefore, it is optimal for the global firm to redeem the ECs as early as possible, justifying the assumption that the firm will use as much ECs from $C$ as possible to offset the tax liabilities arising in the current period.
our single-period model, the multi-period model described above allows the value of leftover ECs in every period to be endogenously determined by solving the MNF’s optimal production quantity problem in the remaining periods.

We make two mild and reasonable assumptions. First the per-period profit function $R(\cdot, \cdot)$ satisfies (A1)-(A5). Second, the end-of-horizon salvage value function, i.e., $V_{T+1}(C)$, is a concave increasing function and $V_{T+1}’(C) \in [0, 1]$ for any $C \geq 0$. The concavity property implies the diminishing return of ECs and the assumption $V_{T+1}(C) \in [0, 1]$ implies that any unit of excess credit is always valuable but its value is no more than 1 because of the possibility of not having sufficient tax liability to offset in the future.

**Lemma 4.** $V_n(C)$ is a concave increasing function and $V_n’(C) \in [0, 1]$ for any $C \geq 0$ and $n = 1, 2, ..., T$.

The above lemma implies that the two properties, i.e., the diminishing return property and the marginal value of ECs being bounded above by 1, which we have imposed to the end-of-horizon salvage value function as an intuitively appealing assumption, are preserved for the endogenously-determined optimal value function of any given ECs in every period $n$.

**Proposition 11.** There exist thresholds $\hat{C}_n$ such that the optimal production quantity $Q_n^*(C)$ decreases in $C$ for $C \in [0, \hat{C}_n]$. Further, $\hat{C}_n \geq \hat{C}$.

The above proposition implies that the non-intuitive result, i.e., the optimal production quantity decreases in ECs when it is small, is robust in the multi-period model with carryback and carryforward of ECs. Further, the result that $\hat{C}_n \geq \hat{C}$ indicates that such a regime in interest is expanded by considering the endogenous value of leftover ECs due to the ECs carryback and carryforward.

### 7.3. Endogenously Generated ECs

In this subsection, we consider a situation in which the ECs to be cross-credited by the low-tax division are endogenously generated by another subsidiary of the same MNF from a high-tax foreign country. To differentiate the global firm’s two subsidiaries, we use the subscripts $l$ and $f$ to indicate the respective low-tax division ($S_l$) and high-tax division ($S_f$). Thus, $\tau_l < \tau_h < \tau_f$. We will use bold font to indicate a vector variable. For an arbitrary production decision $q$, the global firm’s expected worldwide after-tax profits are given by

$$\Pi^C(q) = E_D \left\{ \sum_{i=1}^{l} [R_l(q_i, D_i) - \tau_l R_l(q_i, D_i)^+] - (\tau_h - \tau_l) R_l(q_l, D_l)^+ + \min \{ (\tau_f - \tau_h) R_f(q_f, D_f)^+, (\tau_h - \tau_l) R_l(q_l, D_l)^+ \} \right\},$$

(6)
where the first term within the expectation is the profit repatriated to the home country after paying local taxes, the second term represents the excess limitations from the low-tax subsidiary, and the last term is the amount of excess credits being used to offset the tax liabilities. The objective of the global firm is to maximize its expected worldwide after-tax profits from both subsidiaries.

Note that (6) can be rewritten as

$$\Pi^C(q) = E_D \{ \min \{ \Pi_f(\tau_f, q_f) + \Pi_l(\tau_l, q_l), \Pi_f(\tau_f, q_l) + \Pi_l(\tau_l, q_f) \} \},$$

where for a tax rate $\tau$, $\Pi_i(\tau, q_i) \equiv R_i(q_i, D_i) - \tau R_i(q_i, D_i)^+$, $i = f, l$. Based on Theorem 5.5 (pp.35) in Rockafellar 1970, the pointwise infimum of an arbitrary collection of concave functions is also concave. Thus $\Pi^C(\cdot)$ is a concave function on its domain. Because of the concavity, the optimal quantity can be obtained by solving the first-order conditions. The characterization of these first-order conditions is quite involved and is omitted here. However, the concavity property allows us to demonstrate via a numerical study that the key insights from our base model continue to hold qualitatively.

Consider a global firm with two subsidiaries operating in a business environment in which $\tau_f = 0.45$, $\tau_l = 0.20$ and $\tau_h = 0.35$. The two divisions have the same cost parameters $r_i = 40$, $c_i = 35$, where $i = f, l$, and both face demands with Normal distribution. We fix the mean $\mu_l = 50$ and the standard deviation $\sigma_l = 40$ at the low-tax division $S_l$ but vary the mean demand $\mu_f$ at the high-tax division $S_f$ from 10 to 135, while keeping its coefficient variation fixed at $\sigma_f/\mu_f = 0.8$.

Note that as $\mu_f$ varies from 10 to 135, the global firm’s expected available ECs increase accordingly. We see in Figure 1 that the optimal production decision at the low-tax division $q^*_f$ behaves consistently with the analytical results obtained earlier in Proposition 5: it first decreases, then increases, and finally plateaus as $\mu_f$ increases.

Turning now to the managerial tax rates, we will follow the discussions in §6 by focusing on the three decentralized policies: the pretax measure (setting the managerial tax rate for each subsidiary to zero), the after-local-tax measure (setting the managerial tax rate to the local tax rate $\tau_f$ and $\tau_l$ for $S_f$ and $S_l$ respectively), and the after-home-country-tax measure (setting the managerial tax rates for both subsidiaries at the home country’s tax rate $\tau_h$). We will denote these three policies as $D_0$, $D_l$, and $D_h$ respectively. The corresponding production decisions under these policies are $q(D_0)$, $q(D_l)$, and $q(D_h)$. To measure the effectiveness of a production decision $q$, we define $P(q) \equiv [\Pi^C(q^C) - \Pi^C(q)]/\Pi^C(q)$, which represents the relative gap between the expected worldwide after-tax profits corresponding to the production decision $q$ and that of the optimal centralized decision $q^C$ which maximizes $\Pi^C(q)$ in (6).
Figure 1  Optimal Quantity at Division $S_i$

Figure 2  Performance of Suboptimal Policies
Figure 2 shows the effectiveness of the three proposed policies as the mean demand $\mu_f$ varies from 10 to 135. We first observe that policy $D_0$ performs poorly as it ignores both tax asymmetry and cross-crediting effects. Policy $D_h$ performs quite well for small $\mu_f$ since the firm has small ECs available for cross-crediting. As $\mu_f$ increases beyond 85, $D_l$ performs much better than $D_h$. For $\mu_f$ between 85 and 90, the ECs and tax liabilities generated by the two subsidiaries are very close to each other, that is, tax liabilities will be nearly cross-credited by all of the available ECs. This is the range in which both policies $D_l$ and $D_h$ do not perform well relative to the centralized decision. These observations are again consistent with the analytical results obtained earlier in §6.

8. Concluding Remarks

This paper studies global production decisions at a subsidiary of a US-based MNF located in a low-tax country. Casting the problem in a newsvendor model that captures the effects of tax asymmetry and cross-crediting in a global business environment, our research demonstrates the importance of integrating international tax planning and global supply chain management decisions.

In particular, our study offers a number of managerial insights on managing global production decisions with FTC planning. Some of these insights are counter-intuitive when compared with the conventional understanding of the production decisions without tax consideration. For example, we show that an improvement in a firm’s after-tax profitability (through increased ECs, or increased profit margin, or reduced tax rate) may induce a production manager to produce less, not more; and that the optimal production decision under some quite probable circumstances can be made without knowledge of the demand distribution.

Our study also provides a number of surprising insights into the choice of managerial tax rate in a decentralized decision structure. For example, it suggests that in order to align a low-tax division’s production decision with that of the global firm, the parent company may need to use a managerial tax rate that is even higher than the home-country tax rate, and that the two easily implementable after-tax measures with home country tax rate and local tax rate should be used in place of the pretax measure which is still popular among many MNFs.

We remark that though our paper focuses on the study of a low-tax division with given ECs, all of its results and insights still carry over, with some straightforward modifications, to the mirror case where there is a high-tax division making production decisions under any given excess limits resulted from profit repatriation from low-tax divisions. We therefore omit the repetitive analysis of such a case.

Our model assumes that the global firm counts the US tax liabilities (associated with the low-tax divisions profits) into its after-tax profits at the time when the profits are generated. This
assumption is consistent with the requirement that for a low-tax division, its US tax liabilities must be recorded in consolidated financial statements for eventual home country taxation regardless when the divisions profits are repatriated back to the home country. Even though US GAAP permits a US-based MNF to delay repatriating its overseas profits and forgo recording a deferred tax liability, it requires the company to have proper documentations (which are subject to audits from tax authorities) to support the intent and ability to indefinitely reinvest the profits (so called permanently reinvested earnings or PRE). This suggests that in delaying profit repatriation, global firms face not only internal constraints such as the needs for dividend distribution but also external constraints such as meeting the high hurdles for designating PRE. Therefore, to relax our exogenous profit repatriation assumption to include endogenous profit repatriation decisions, constraints ought to be imposed to reflect the firm’s partial flexibility in delaying profit repatriation. We leave it to future research.

Appendix

Proof of Proposition 1. Dividing $1 - \tau$ on both sides of the equation prior to Proposition 1 yields

$$G(eq/p)c/(1-\tau) + (G(q) - G(eq/p))c = (1-G(q))(p-c). \quad (7)$$

Because the right (left) hand of (7) side decreases (increases) when $q$ increases, and the left hand side increases as $\tau$ increases, $q^c(\tau)$ decreases as $\tau$ increases. For any given production quantity $q$, as $p$ increases or $c$ decreases, the expected marginal overstock cost $G(eq/p)c + (G(q) - G(eq/p))(1-\tau)c$ decreases and the expected marginal understock cost $(1-G(q))(1-\tau)(p-c)$ increases and therefore, the optimal production quantity increases.

Proof of Proposition 6. As $p$ increases from $c$ to $+\infty$, both $\bar{C}$ and $\hat{C}$ increase from 0 to infinity. Hence, by Proposition 5, as $p$ increases from $c$ to $+\infty$, $q^*(C)$ changes from $q^c(\tau_l)$ to $q_1(C) = \frac{c}{(\tau_h-\tau_l)(p-c)}$ and then to $q_2(C)$. The proposition follows by noting that both $q^c(\tau_l)$ and $q_2(C)$ increase in $p$ and that $q_1(C)$ decreases in $p$.

Proof of Proposition 8. Note that for any $C$, $q^*(C) \leq q^c(\tau_l) \leq q^c(0)$. Because $\Pi(q,C)$ is a concave function of $q$, we have $\Pi(q^*(0),C) \leq \Pi(q^c(\tau_l),C)$ for any $C$. It remains to compare $\Pi(q^c(\tau_l),C)$ and $\Pi(q^c(\tau_h),C)$.

By (2),

$$\Pi(q,C) = E_D\{R(q,D) - \tau_h R^+(q,D) + \min(C,(\tau_h-\tau_l)R^+(q,D))\} = \int_0^{s(q,C)} [R(q,x) - \tau_l R^+(q,x)]g(x)dx + \int_{s(q,C)}^{+\infty} [R(q,x) - \tau_h R^+(q,x) + C]g(x)dx$$
where \( s(q, C) = \{ s | (\tau_h - \tau) R^+ (q, s) = C \} \). Therefore,

\[
\frac{\partial \Pi(q^o(\tau_i), C) - \Pi(q^o(\tau_i), C)}{\partial C} = G(s(q^o(\tau_i), C)) - G(s(q^o(\tau_i), C)) \geq 0,
\]

because \( q^o(\tau_i) \leq q^o(\tau_i) \) and \( s(q, C) \) decreases as \( q \) increases. This, together with the fact that \( \Pi(q^o(\tau_i), 0) \leq \Pi(q^o(\tau_i), 0) \) and \( \Pi(q^o(\tau_i), C) \geq \Pi(q^o(\tau_i), C) \), suggests that there exists a threshold \( \tilde{C} \leq C \), such that \( \Pi(q^o(\tau_i), C) \leq \Pi(q^o(\tau_i), C) \) for \( C \leq \tilde{C} \) and \( \Pi(q^o(\tau_i), C) \geq \Pi(q^o(\tau_i), C) \) for \( C \geq \tilde{C} \). The result that \( \tilde{C} \geq \tilde{C} \) follows from \( q^o(\tilde{C}) \leq q^o(\tau_i) \leq q^o(0) \) and thus \( \Pi(q^o(0), \tilde{C}) \leq \Pi(q^o(\tau_i), \tilde{C}) \leq \Pi(q^o(\tau_i), \tilde{C}) \).

**Proof of Lemma 2.** We prove by contradiction. If \( (\tau_h - \tau) R(Q^*, s) < C \), then \( \Pi(Q^*) = E_s[R(Q^*, S) - \tau R^+ (Q^*, s)] \). If \( Q^* < Q^o(\tau_i) \), there exists a sufficiently small positive value \( \varepsilon \) such that \( (\tau_h - \tau) R(Q^* + \varepsilon, s) < C \) and hence \( \Pi(Q^*) \leq \Pi(Q^* + \varepsilon) \). By the definition of \( Q^* \), this inequality must be binding, i.e., \( \Pi(Q^*) = \Pi(Q^* + \varepsilon) \) which is equivalent to \( E_s[R(Q^*, S) - \tau R^+ (Q^*, S)] = E_s[R(Q^* + \varepsilon, S) - \tau R^+ (Q^* + \varepsilon, S)] \). However, \( Q^* < Q^* + \varepsilon < Q^o(\tau_i) \) contradicts the definition of \( Q^o(\tau_i) \). If \( Q^* > Q^o(\tau_i) \), there exists a sufficiently small positive value \( \varepsilon \) such that \( (\tau_h - \tau) R(Q^* - \varepsilon, s) < C \) and hence \( \Pi(Q^*) \geq \Pi(Q^* - \varepsilon) \). Hence, \( Q^* - \varepsilon \) is also a maximizer of \( \Pi(Q) \), which is in contradiction with the definition of \( Q^* \).

**Proofs of Proposition 9, Lemma 3, and Proposition 10.** We prove these results in four steps: First, we present and prove several technical results, Lemma a1 through a5. Second, we prove Lemma 3. Third, we present and prove two more middle results, Lemma a6 and a7. Finally, we prove Propositions 9 and 10 by using the lemmas in the previous steps.

**Step 1. Lemma a1 to a5.**

**Lemma a1.** \( \bar{Q}(s) \) increases in \( s \).

**Proof of Lemma a1.** Recall \( \bar{Q}(s) = \arg\max_Q R(Q, s) \). By assumption (A2), \( R(Q, s) \) has increasing difference in \( Q \) and \( s \). Thus, the parametric monotonicity property implies that \( \bar{Q}(s) \) increases in \( s \).

**Lemma a2.** \( \bar{Q}(s) \leq Q^o(\tau_h) \leq Q^o(\tau_i) \leq \bar{Q}(s) \).

**Proof of Lemma a2.** We first prove \( Q^o(\tau_h) \leq Q^o(\tau_i) \). Define \( H(Q, \tau) = E_s[R(Q, S) - \tau R^+ (Q, S)] \) and \( Q^o(\tau) = \arg\max_Q H(Q, \tau) \). Because \( \tau_i < \tau_h \), it suffices to show that \( Q^o(\tau) \) decreases in \( \tau \). Let \( s^o(\tau) = \sup \{ s | R(Q, s) < 0 \} \) and rewrite \( H(Q, \tau) \) as follows:

\[
H(Q, \tau) = \int_s^{s^o(\tau)} R(Q, x) g(x) dx + \int_{s^o(\tau)}^T (1 - \tau) R(Q, x) g(x) dx.
\]

Taking the partial derivative with respect to \( Q \) yields

\[
\frac{\partial H(Q, \tau)}{\partial Q} = \int_s^{s^o(\tau)} R_1(Q, x) g(x) dx + \int_{s^o(\tau)}^T (1 - \tau) R_1(Q, x) g(x) dx.
\]

(8)
Note that for \( x \in [s, s^o(Q)] \), \( R(Q, x) < 0 \). Thus, assumptions (A1) and (A4) lead to \( \hat{Q}(x) \leq Q \) for any \( x \in [s, s^o(Q)] \). Hence, \( R_1(Q, x) \leq 0 \) for any \( x \in [s, s^o(Q)] \). Setting (8) to zero yields that the maximizer \( Q^o(\tau) \) satisfies

\[
\int_{s^o(Q)}^{\tau} (1 - \tau) R_1(Q^o(\tau), x) g(x) dx = - \int_{s^o(Q)}^{\tau} R_1(Q^o(\tau), x) g(x) dx \geq 0.
\]

Thus, without loss of generality we can impose the constraint \( \int_{s^o(Q)}^{\tau} (1 - \tau) R_1(Q, x) g(x) dx \geq 0 \). Consequently,

\[
Q^o(\tau) = \arg \max_{Q} H(Q, \tau), \text{ subject to } \int_{s^o(Q)}^{\tau} (1 - \tau) R_1(Q, x) g(x) dx \geq 0.
\]

Note that

\[
\frac{\partial^2 H(Q, \tau)}{\partial Q^2 \partial \tau} = - \int_{s^o(Q)}^{\tau} R_1(Q, x) g(x) dx \leq 0,
\]

where the inequality holds because of the imposed constraint. Thus, the parametric monotonicity property yields that \( Q^o(\tau) \) decreases in \( \tau \).

It remains to show \( \hat{Q}(\hat{s}) \leq Q^o(\tau) \leq \hat{Q}(\hat{s}) \) for any \( \tau \in (0, 1) \). Because \( H(Q, \tau) \) is concave in \( Q \) and \( Q^o(\tau) \) is the maximizer of \( H(Q, \tau) \), it suffices to prove that \( \frac{\partial H(Q, \tau)}{\partial Q} \bigg|_{Q = \hat{Q}(\hat{s})} \geq 0 \) and \( \frac{\partial H(Q, \tau)}{\partial Q} \bigg|_{Q = \hat{Q}(s)} \leq 0 \). By assumption (A1) and Lemma a1 that \( \hat{Q}(\hat{s}) \leq \hat{Q}(s) \), \( R_1(\hat{Q}(\hat{s}), s) \geq 0 \) for any \( s \in [\hat{s}, \hat{s}] \) and therefore \( \frac{\partial H(Q, \tau)}{\partial Q} \bigg|_{Q = \hat{Q}(\hat{s})} \geq 0 \). Likewise, because of assumption A1 and Lemma a1 that \( \hat{Q}(\hat{s}) \geq \hat{Q}(s) \), \( R_1(\hat{Q}(\hat{s}), s) \leq 0 \) for any \( s \in [\hat{s}, \hat{s}] \) and therefore \( \frac{\partial H(Q, \tau)}{\partial Q} \bigg|_{Q = \hat{Q}(\hat{s})} \leq 0 \).

**Lemma a3.** Let \( q^*(s) = \arg \max_{N_2(s) \geq Q \geq N_1(s)} M(Q, s) \) for \( s \in [s', s''] \), where the minimum is chosen if there are multiple maximizers. If both \( N_1(s) \) and \( N_2(s) \) increase in \( s \), \( M(Q, s) \) is concave in \( Q \) and has decreasing difference in \( Q \) and \( s \) for \( Q \geq N_1(s) \), then one of the following statements must be true: 1) \( q^*(s) \) decreases in \( s \), 2) \( q^*(s) = N_1(s) \), 3) \( q^*(s) = N_2(s) \), for \( s \in [s', s''] \).

**Proof of Lemma a3.** We prove by contradiction. Suppose the statements are false. Equivalently, there exists \( s_0 \in [s', s''] \) such that \( dq^*(s_0)/ds > 0 \) and \( N_2(s_0) > q^*(s_0) > N_1(s_0) \), then the assumption that \( N_1(s) \) and \( N_2(s) \) increase in \( s \) implies that there exists a sufficiently small value \( \varepsilon > 0 \) such that \( N_2(s_0 + \varepsilon) \geq N_2(s_0) \geq q^*(s_0 + \varepsilon) > q^*(s_0) \geq N_1(s_0 + \varepsilon) \geq N_1(s_0) \). That is, both \( q^*(s_0 + \varepsilon) \) and \( q^*(s_0) \) satisfy the constraints \( N_2(s) \geq Q \geq N_1(s) \) for \( s = s_0 \) and \( s = s_0 + \varepsilon \), respectively, and thus are feasible solutions to the respective problems \((P-s_0)\) and \((P-s_0 + \varepsilon)\), where \((P-s)\) denotes the optimization problem \( \max_{N_2(s) \geq Q \geq N_1(s)} M(Q, s) \) for \( s \in [s', s''] \). Because \( M(Q, s) \) has decreasing difference in \( Q \) and \( s \) for \( Q \geq N_1(s) \), \( M(q^*(s_0 + \varepsilon), s_0 + \varepsilon) - M(q^*(s_0), s_0 + \varepsilon) \leq M(q^*(s_0 + \varepsilon), s_0) - M(q^*(s_0), s_0) \leq 0 \), where the last inequality holds because of the definition of \( q^*(s_0) \) and the fact that \( q^*(s_0 + \varepsilon) \) is
Therefore, the maximizer of \( Q(q^*(s_0 + \varepsilon), s_0 + \varepsilon) - M(q^*(s_0), s_0 + \varepsilon) \leq 0 \). This, together with the fact that \( q^*(s_0) \) is a feasible solution to (P-\( s_0 + \varepsilon \)), suggests that \( q^*(s_0) \) is a maximizer of (P-\( s_0 + \varepsilon \)). However, \( q^*(s_0) < q^*(s_0 + \varepsilon) \). This is in contradiction with the definition of \( q^*(s_0 + \varepsilon) \).

**Lemma a4.** Let \( q^*(s) = \arg \max_{Q \leq N(s)} M(Q, s) \) for \( s \in [s', s''] \), where the minimum is chosen if there are multiple maximizers. If \( N(s) \) increases in \( s \), \( M(Q, s) \) is concave in \( Q \) and has increasing difference in \( Q \) and \( s \) for \( Q \leq N(s) \), then \( q^*(s) \) increases in \( s \) for \( s \in [s', s''] \).

**Proof of Lemma a4.** We prove by contradiction. If the statements are false, then there exists \( s_0 \in [s', s''] \) such that \( dq^*(s_0)/ds < 0 \). Thus, there exists a sufficiently small positive value \( \varepsilon \) such that \( q^*(s_0) > q^*(s_0 + \varepsilon) \).

On one hand, because \( M(Q, s) \) has increasing difference in \( Q \) and \( s \) for \( Q \leq N(s) \), \( M(q^*(s_0), s_0 + \varepsilon) - M(q^*(s_0), s_0 + \varepsilon) \geq M(q^*(s_0), s_0) - M(q^*(s_0 + \varepsilon), s_0) \geq 0 \), where the last inequality holds because of the definition of \( q^*(s_0) \) and \( q^*(s_0 + \varepsilon) \leq N(s_0) \). On the other hand, because \( q^*(s_0 + \varepsilon) \) is the maximizer of \( M(Q, s_0 + \varepsilon) \) and \( q^*(s_0) \leq N(s_0 + \varepsilon) \), \( M(q^*(s_0), s_0 + \varepsilon) - M(q^*(s_0 + \varepsilon), s_0 + \varepsilon) < 0 \).

Therefore, \( M(q^*(s_0), s_0) - M(q^*(s_0 + \varepsilon), s_0) = 0 \), implying that \( q^*(s_0 + \varepsilon) \) is also a maximizer of \( M(Q, s_0) \). However, \( q^*(s_0) > q^*(s_0 + \varepsilon) \). This is in contradiction with the fact that \( q^*(s_0) \) is the minimum maximizer of \( \max_{Q \leq N(s_0)} M(Q, s_0) \).

**Lemma a5.** For any \( s \in [\bar{s}, \bar{s}] \), \( \bar{Q}(s) = Q^*(\tau_h) \) or \( \bar{Q}(s) < Q^*(s) \), where \( Q^*(s) = \sup \{ Q : R(Q, s) \geq 0 \} \).

**Proof of Lemma a5.** It suffices to show that \( J(Q, s) \leq J(Q^*(\tau_h), s) \) for \( Q \geq Q^*(s) \), which follows from the definition of \( Q^*(\tau_h) \) and \( J(Q, s) = \int_{\bar{s}}^{s(Q)} R(Q, x)g(x)dx + \int_{s(Q)}^{s}\tau hR(Q, x)dx \) for \( Q \geq Q^*(s) \).

**Step 2. Proof of Lemma 3.**

We have proved part (a) in Lemma a1. Note that \( \bar{Q}(s) = Q^*(\tau_h) \) and \( \bar{Q}(s) = Q^*(\tau_h) \). This, together with Lemma a2 that \( \bar{Q}(s) \leq Q^*(\tau_h) \leq Q^*(s) \), suggests that \( \bar{Q}(s) \leq Q^*(s) \). Hence, there exists \( s_1 \in [\bar{s}, \bar{s}] \) such that \( \bar{Q}(s_1) = Q^*(s_1) \) and \( \bar{Q}(s) < Q^*(s) \) for \( s < s_1 \). It remains to show that \( \bar{Q}(s) \geq Q^*(s) \) for \( s \geq s_1 \).

We prove this result by contradiction. Suppose there exists \( s_2 \geq s_1 \) such that \( \bar{Q}(s_2) < \bar{Q}(s_2) \). Then there exists an interval \( [s_3, s_4] \) within the interval \( [s_1, s_2] \) such that \( \bar{Q}(s) \neq Q^*(\tau_h) \), \( \bar{Q}(s) > \bar{Q}(s) \) and \( d\bar{Q}(s)/ds > d\bar{Q}(s)/ds \) for \( s \in [s_3, s_4] \). By Lemma a5 and \( \bar{Q}(s) \neq Q^*(\tau_h) \), we have \( \bar{Q}(s) < Q^*(s) \). This, together with \( \bar{Q}(s) > \bar{Q}(s) \), suggests the maximization problem \( \max_{Q \geq \bar{Q}(s)} J(Q, s) \), where \( s \in [s_3, s_4] \), is equivalent to \( \max_{Q \geq \bar{Q}(s)} J(Q, s) \), where \( s \in [s_3, s_4] \). Note that for \( Q^*(s) \geq Q \),

\[
J(Q, s) = \int_{\bar{s}}^{s(Q)} R(Q, x)g(x)dx + \int_{s_3}^{s} [R(Q, x) - \tau hR(Q, x)]g(x)dx
\]
\[ + \int_s^\pi [R(Q,x) - \tau_h R(Q,x)]g(x)dx. \]

Hence, \( \partial^2 J(Q,s)/\partial Q \partial s = (\tau_h - \tau_l)R_1(Q,s) \leq 0 \) for \( Q \geq \tilde{Q}(s) \). It then follows from Lemma a3 that one of the following three statements must be true: 1) \( \tilde{Q}(s) \) decreases in \( s \), 2) \( \tilde{Q}(s) = Q^o(s) \), 3) \( \tilde{Q}(s) = \hat{Q}(s) \), for \( s \in [s_3, s_4] \). Note that 1) is in contradiction with \( d\tilde{Q}(s)/ds > d\hat{Q}(s)/ds \geq 0 \), 2) is in contradiction with the earlier result \( \tilde{Q}(s) < Q^o(s) \), and 3) is in contradiction with \( \hat{Q}(s) > \tilde{Q}(s) \).

**Step 3. Lemma a6 and a7.**

**Lemma a6.** \( \tilde{Q}(s) \) decreases in \( s \) for \( s \in [\underline{s}, s_1] \) and increases in \( s \) for \( s \in [s_1, \overline{s}] \).

**Proof of Lemma a6.** By Lemma a5, for \( s \in [\underline{s}, s_1] \), \( \tilde{Q}(s) = Q^o(\tau_h) \) or \( \tilde{Q}(s) < Q^o(s) \). Now suppose \( \tilde{Q}(s) < Q^o(s) \). By Lemma 3, for \( s \in [\underline{s}, s_1] \), \( \tilde{Q}(s) < \tilde{Q}(\pi) \). Thus, the maximization problem \( \max_Q J(Q,s) \), where \( s \in [\underline{s}, s_1] \) is equivalent to \( \max_{Q^o(s) \geq Q \geq \tilde{Q}(s)} J(Q,s) \), where \( s \in [\underline{s}, s_1] \). Note that \( \partial^2 J(Q,s)/\partial Q \partial s = (\tau_h - \tau_l)R_1(Q,s) \leq 0 \) for \( Q \geq \tilde{Q}(s) \). Hence, it follows from Lemma a3 that \( \tilde{Q}(s) \) either decreases in \( s \), \( \tilde{Q}(s) = Q^o(s) \), or \( \tilde{Q}(s) = \tilde{Q}(\pi) \) for \( s \in [\underline{s}, s_1] \). Because \( \tilde{Q}(s) < \tilde{Q}(\pi) < Q^o(s) \), \( \tilde{Q}(s) \) must decrease in \( s \) for \( s \in [\underline{s}, s_1] \). Therefore, we have proved that \( \tilde{Q}(s) = Q^o(\tau_h) \) or \( \tilde{Q}(s) \) decreases in \( s \) for any \( s \in [\underline{s}, s_1] \). This, together with the fact that \( \tilde{Q}(\pi) = Q^o(\tau_h) \), proves that \( \tilde{Q}(s) \) decreases in \( s \) for any \( s \in [\underline{s}, s_1] \).

By Lemma 3, for \( s \in [s_1, \overline{s}] \), \( \tilde{Q}(s) \geq \tilde{Q}(\pi) \). Thus, the maximization problem \( \max_Q J(Q,s) \), where \( s \in [s_1, \overline{s}] \) is equivalent to \( \max_{Q^o(s) \geq Q \geq \tilde{Q}(\pi)} J(Q,s) \), where \( s \in [s_1, \overline{s}] \). Note that the constraint \( Q \leq \tilde{Q}(s) \) implies that \( Q \leq Q^o(s) \). Hence, \( \partial^2 J(Q,s)/\partial Q \partial s = (\tau_h - \tau_l)R_1(Q,s) \geq 0 \) for \( Q \leq \tilde{Q}(s) \). It then follows from Lemma a4 that \( \tilde{Q}(s) \) increase in \( s \) for \( s \in [s_1, \overline{s}] \).

**Lemma a7.** \( R(\tilde{Q}(s),s) \) increases in \( s \).

**Proof of Lemma a7.** For \( s \in [\underline{s}, s_1] \), \( d\tilde{Q}(s)/ds \leq 0 \) (Lemma a6) and \( R_1(\tilde{Q}(s),s) \leq 0 \) (because of \( \tilde{Q}(s) < \tilde{Q}(\pi) \) in Lemma 3 and assumption A1). This, together with assumption (A3) that \( R_2(\tilde{Q}(s),s) \geq 0 \), implies that \( dR(\tilde{Q}(s),s)/ds = R_1(\tilde{Q}(s),s)d\tilde{Q}(s)/ds + R_2(\tilde{Q}(s),s) \geq 0 \).

Similarly, for \( s \in [s_1, \overline{s}] \), \( d\tilde{Q}(s)/ds \geq 0 \) (Lemma a6) and \( R_1(\tilde{Q}(s),s) \geq 0 \) (because of \( \tilde{Q}(s) \geq \tilde{Q}(\pi) \) in Lemma 3 and assumption A1). This, together with assumption (A3) that \( R_2(\tilde{Q}(s),s) \geq 0 \), implies that \( dR(\tilde{Q}(s),s)/ds = R_1(\tilde{Q}(s),s)d\tilde{Q}(s)/ds + R_2(\tilde{Q}(s),s) \geq 0 \).

**Step 4. Proof of Proposition 9 and 10.**

**Proof of Proposition 9.** Lemma 2 allows us to impose the constraint \( C \leq (\tau_h - \tau_l)R(Q,\overline{s}) \) to the global firm’s profit maximization problem without loss of generality. Depending on the realization of demand \( s \), the effective tax rate changes from 0 to \( \tau_l \), and then to \( \tau_h \), thus (5) can be rewritten as follows:

\[
 \max_Q \Pi(Q) = \int_{s^o(Q)}^{s^o(Q)} [R(Q,x)]g(x)dx + \int_{s^o(Q)}^{s^o(Q)} (1 - \tau_l)R(Q,x)g(x)dx
\]
Because $\Pi(C)$ implying that its first order derivative with respect to $s$ for $s > s^*(Q)$, $R(Q, s) \geq 0$ and otherwise, $R(Q, s) < 0$. Because $\Pi(Q)$ is concave, it suffices to find a solution to the following first order condition,

$$
\Pi'(Q) = \int_{s(Q,C)}^\infty [ (1 - \tau_h) R(Q, x) + C ] g(x) dx, 
$$

where $s(Q, C) = \{ s | (\tau_h - \tau_l) R(Q, s) = C \}$ and for $s > s^*(Q)$, $R(Q, s) \geq 0$ and otherwise, $R(Q, s) < 0$.

Define $s^*(C) = \{ s | (\tau_h - \tau_l) R(\tilde{Q}(s), s) = C \}$. Such $s^*(C)$ is well defined because $R(\tilde{Q}(s), s)$ increases in $s$ (by Lemma a7) from $R(\tilde{Q}(\bar{s}), s) \leq 0$ (by A5) to $R(\tilde{Q}(\bar{s}), \bar{s}) = \bar{C}$. Note that $s(\tilde{Q}(s^*(C)), C) = s^*(C)$. Hence,

$$
\Pi'(\tilde{Q}(s^*(C))) = \int_{s^*(C)}^{\infty} R_1(Q, x) g(x) dx + \int_{s^*(C)}^{\infty} (1 - \tau_l) R_1(Q, x) g(x) dx 
$$

$$
+ \int_{s^*(C)}^{\infty} [ (1 - \tau_h) R(Q, x) + C ] g(x) dx = 0
$$

where the last equality follows from the definition of $\tilde{Q}(-)$. Therefore, $\tilde{Q}(s^*(C))$ satisfies the first order condition, i.e., $Q^* = \tilde{Q}(s^*(C))$.

**Proof of Proposition 10.** Note that $s^*(C)$ increases in $C$ for $C \in [0, \bar{C}]$ and $s^*(\tilde{C}) = s_1$. This, together with Lemma a6, suggests that $\tilde{Q}(s^*(C))$ decreases in $C$ for $C \in [0, \tilde{C}]$ and increases in $C$ for $C \in [\tilde{C}, \bar{C}]$.

**Proof of Lemma 4.** We prove by induction. Note that the statement in the lemma is true for $T + 1$. Suppose it holds for $n + 1$. It suffices to show that it also holds for $n$. Let $s^*(Q) = \sup \{ s | R(Q, s) < 0 \}$ and $s(Q, C) = \{ s | (\tau_h - \tau_l) R(Q, s) = C \}$. We can rewrite $V_n(C)$ as follows.

$$
V_n(C) = \max_Q E_{C_{n+2}} \left\{ \int_{s^*(Q)}^{\infty} [ R(Q, x) + V_{n+1}(C + C_{n+2}) ] g_n(x) dx + \int_{s(Q,C)}^{\infty} [ (1 - \tau_l) R(Q, x) + V_{n+1}(C - (\tau_h - \tau_l) R(Q, x) + C_{n+2}) ] g_n(x) dx 
\right\},
$$

implying that its first order derivative with respect to $C$ is

$$
V_n'(C) = \max_Q E_{C_{n+2}} \left\{ \int_{s^*(Q)}^{\infty} V_{n+1}(C + C_{n+2}) g_n(x) dx + \int_{s(Q,C)}^{\infty} (1 - \tau_l) R(Q, x) + C + V_{n+1}(C_{n+2}) ] g_n(x) dx 
\right\},
$$

and its second order derivative is

$$
V_n''(C) = \max_Q E_{C_{n+2}} \left\{ \int_{s^*(Q)}^{\infty} V_{n+1}'(C + C_{n+2}) g_n(x) dx + \int_{s(Q,C)}^{\infty} (1 - \tau_l) R(Q, x) + C_{n+2}) ] g_n(x) dx 
\right\}.
$$
Therefore, the result that \( V'_n(C) \in [0, 1] \) follows directly from \( V'_{n+1}(C) \in [0, 1] \); and the result that \( V''_n(C) \leq 0 \) follows because \( V'_{n+1}(C) \leq 0 \), \( V'_{n+1}(C) - 1 \leq 0 \), and \( \partial s(Q, C)/\partial C \geq 0 \).

**Proof of Proposition 11.** We introduce the following notation that have been used in §7.1:

\[
J(Q, s) = \int_{s}^{s'(Q)} R(Q, x) g_n(x) dx + \int_{s'}^{\tilde{Q}(Q, s)} (1 - \tau_1) R(Q, x) g_n(x) dx + \int_{\tilde{Q}(Q, s)}^{V_n'(C) + C} (1 - \tau_1) R(Q, x) g_n(x) dx,
\]

where

\[
\tilde{Q}(s) = \arg\max_Q J(Q, s), \quad \tilde{Q}(s) = \arg\max_Q R(Q, s), \quad \text{and} \quad s^*(C) = \min\{s | (\tau_h - \tau_l) R(\tilde{Q}(s), s) = C\}.
\]

It follows from Lemma 3a that there exists a threshold \( s_1 \) such that \( \tilde{Q}(s) < \tilde{Q}(s) \) for \( s < s_1 \) and \( \tilde{Q}(s) \geq \tilde{Q}(s) \) for \( s \geq s_1 \). Let \( \tilde{C} = (\tau_h - \tau_l) R(\tilde{Q}(s_1), s_1) \). It follows from Lemma 3a (i.e., \( R(\tilde{Q}(s), s) \) increases in \( s \)) that \( s^*(C) < s_1 \) for \( C < \tilde{C} \), implying that \( \tilde{Q}(s^*(C)) < \tilde{Q}(s^*(C)) \) for \( C < \tilde{C} \). Therefore, \( R_1(\tilde{Q}(s^*(C)), x) \leq 0 \) for all \( x \leq s^*(C) \).

Note that

\[
V_n(C) = \max_Q \frac{1}{E_{C_{n+2}}} \left\{ \begin{array}{l} \int_{s}^{s'(Q)} R(Q, x) + V_{n+1}(C + C_{n+2})] g_n(x) dx \\ + \int_{s'}^{\tilde{s}'(Q)} [(1 - \tau_1) R(Q, x) + V_{n+1}(C - (\tau_h - \tau_l) R(Q, x) + C_{n+2})] g_n(x) dx \\ + \int_{\tilde{s}'(Q)}^{V_n'(C)} [(1 - \tau_1) R(Q, x) + C + V_{n+1}(C_{n+2})] g_n(x) dx \end{array} \right\}.
\]

Let \( \Pi(Q, C) \) be the maximand in the above equation. We have

\[
\partial \Pi(Q, C)/\partial Q = E_{C_{n+2}} \left\{ \begin{array}{l} \int_{s}^{s'(Q)} R_1(Q, x) g_n(x) dx \\ + \int_{s'}^{s'(Q)} [(1 - \tau_1) R_1(Q, x) - V_{n+1}'(C - (\tau_h - \tau_l) R(Q, x) + C_{n+2})] g_n(x) dx \\ + \int_{\tilde{s}'(Q)}^{V_n'(C)} (1 - \tau_1) R_1(Q, x) g_n(x) dx \end{array} \right\},
\]

implying that \( \partial \Pi(Q, C)/\partial Q \big|_{Q = \tilde{Q}(s^*(C))} = - \int_{s'}^{s'(Q)} V_{n+1}'(C - (\tau_h - \tau_l) R(\tilde{Q}(s^*(C)), x) + C_{n+2}) (\tau_h - \tau_l) R_1(\tilde{Q}(s^*(C)), x) g_n(x) dx \geq 0 \), where the inequality is due to \( V_{n+1}'(\cdot) \geq 0 \) and the earlier established result that \( R_1(\tilde{Q}(s^*(C)), x) \leq 0 \) for all \( x \leq s^*(C) \). Consequently, by the concavity of \( \Pi(Q, C) \) in \( Q \), we have that the maximizer \( Q^*(C) \geq \tilde{Q}(s^*(C)) \) for \( C \leq \tilde{C} \). Therefore, we can impose the constraint \( Q \geq \tilde{Q}(s^*(C)) \) to the optimization problem without loss of generality.

Note that

\[
\partial^2 \Pi(Q, C)/\partial Q \partial C = E_{C_{n+2}} \left\{ \begin{array}{l} - \int_{s'}^{s'(Q)} V_{n+1}''(C - (\tau_h - \tau_l) R(Q, x) + C_{n+2}) (\tau_h - \tau_l) R_1(Q, x) g_n(x) dx \\ (\tau_h - \tau_l) R_1(Q, s(Q, C)) g_n(s(Q, C)) \partial s(Q, C)/\partial C \end{array} \right\}.
\]

For \( Q \geq \tilde{Q}(s^*(C)) \), \( s(Q, C) \leq s^*(C) \). Hence, it follows from the earlier result (i.e., \( R_1(\tilde{Q}(s^*(C)), x) \leq 0 \) for all \( x \leq s^*(C) \)) that \( R_1(Q, x) \leq 0 \) for all \( x \leq s(Q, C) \) and \( Q \geq \tilde{Q}(s^*(C)) \), implying that \( \partial^2 \Pi(Q, C)/\partial Q \partial C \leq 0 \) for \( Q \geq \tilde{Q}(s^*(C)) \). Hence, \( Q^*(C) \) decreases in \( C \) over \( C \leq \tilde{C} \).
References


