The Economics of Administering Import Quotas with Licenses-on-Demand

Jana Hranaiova, James Falk

and Harry de Gorter*

Prepared for the World Bank’s Agricultural Trade Group

January 2003
Abstract

This paper examines the effects of rationing import quotas by licenses-on-demand. We analyze the Nash equilibrium for a game where prizes are allocated on a prorated basis and bidding is costless. Factors affecting the degree of inefficiency and rent shares among firms are assessed. With licenses-on-demand, a reduction of in-quota tariffs may result in decreased quota fill, and an increase in the quota may result in a 100 percent quota fill rate. This result is opposite to that when licenses are allocated on an historical share basis, for example. Other policy implications for the implementation of licenses-on-demand include the efficacy of non-use penalties and the inefficiency of (a) license fees (equivalent to an increase in in-quota tariffs); and (b) limits on the number of licenses a firm is allowed to receive.

Keywords: import quotas, rent dissipation, licenses-on-demand, Nash equilibrium, rent distribution

JEL Classification: F13 ; D45

*Authors are Assistant Professor at the University of New Mexico, Professor of Operations Research at The George Washington University and Associate Professor at Cornell University, respectively. Corresponding author: Jana Hranaiova, Anderson Schools of Management, University of New Mexico, 1924 Las Lomas NE, Albuquerque, NM 87131. Ph: (505) 277-5230, email hranai@unm.edu

The findings, interpretations, and conclusions of this paper are the authors’ own and should not be attributed to the World Bank, its management, its Board of Executive Directors, or the countries they represent.
The Economics of Administering Import Quotas with Licenses-on-Demand

1. Introduction

Import quotas remain an important policy instrument whether in the form of a simple quota or in combination with tariffs as a tariff rate quota. For example, the Agreement on Agriculture in the Final Act of the Uruguay Round 1994 established 1,371 tariff-quotas alone. Rents generated by import quotas provide an opportunity for firms to spend resources in competing for these rents, with the degree of dissipation depending critically on the method by which import licenses are allocated.

Of the seven principal methods of quota administration reported by the World Trade Organization (WTO), a licenses-on-demand is the most common method representing 47 percent of the total cases where the quota is binding (WTO, 2000). First-come, first-served, historical import shares and auction are the next most common methods of allocating import quotas, representing 21, 9 and 8 percent of the total cases where the quota is binding, respectively. The first-come, first-served method does not entail allocation of import or export licenses. While auctions have been widely discussed in the literature and historical shares method has been given some attention, literature is scarce on the most common method of licenses-on-demand. This paper aims to fill this gap.

The purpose of this paper is to analyze the economics of licenses-on-demand and assess the effects on rent shares and economic efficiency. The licenses-on-demand method assigns licenses based on the level of licenses requested by individual firms. If the sum of all requests is greater than the quota, then each firm receives a prorated number of licenses. Hence, licenses-on-demand is another way in which the government rations quota rents. The way in which a quota is administered can have a direct influence on

---

1 This excludes applied tariffs as a ‘method’ of tariff quota administration where the quota is non-binding.
both trade flows and the distribution of rents, and is, therefore, a highly political issue. There is an urgent need to provide more information on the economics of tariff quota administration methods. By comparing the economic implications of licenses-on-demand to those of auctions and first-come, first-served methods, our paper aims to contribute to the discussion on tariff quota reform and trade liberalization, and better rules for tariff quota administration. Auctions allocate import licenses to the highest bidders and the first-come, first-served method allocates imports without licenses to the firms that arrive the first to the port of imports’ entry.

We present a game-theoretic analysis of the competition for import quota licenses under different economic conditions when licenses-on-demand are implemented as the administration method. We show that there exists a unique equilibrium for an extended game where prizes are allocated on the prorated basis and bidding is costless. Inefficiency occurs because licenses-on-demand method of quota administration enables high cost firms to operate. In addition, higher cost firms receive the number of licenses closer to their desired allocation, while lower cost firms receive licenses progressively further from their desired allocation.

We confirm some intuitive results like greater inefficiency of the licenses on demand system of allocation compared to the ideal of a well functioning auction or tradable license market. But we also show that 1) the level of waste depends more on the variance in the cost structure among firms than on the level of costs, 2) presence of fixed costs has an ambiguous effect on efficiency 3) rent dissipation may be lower than with the first-come, first-served method but dead-weight cost can be larger due to inefficient allocation of resources, 4) a reduction of in-quota tariffs may result in decreased quota fill, and an increase in the quota may result in a 100 percent quota fill rate (a result opposite to that when licenses are allocated on an historical share basis). Other policy implications for the implementation of licenses-on-demand include the efficacy of non-use penalties and the inefficiency of (a) license fees (equivalent to an increase in in-quota tariffs); and (b) limits on the number of licenses a firm is allowed to receive. Effects of trade liberalization are examined through an increase in the quota amount and a change in the per unit
quota rent due to a change in tariffs or market prices. In addition, conditions affecting quota under-fill are identified.

The paper is organized as follows. The next section discusses the literature on quota administration methods and highlights the contribution of this paper to the literature. Section 3 introduces the model of a game when multiple prizes (licenses) are allocated on a prorated basis, followed by some numerical examples that illustrate the Nash equilibrium for three firms. Section 4 highlights the economic implications of the results and discusses some economic and policy factors that can affect the outcome. Section 5 explores the implication of relaxing several of the assumptions of the base model by allowing for constant marginal costs and no penalty for leaving licenses unused. The concluding section summarizes the policy implications and offers suggestions for future research.

2. Literature on quota administration methods

From the four most common methods of quota administration, auction has been most widely analyzed in the literature as a common method of allocating rents (Bergsten et al.; Krishna 1990; 1993a,b) or prizes (Holt; McAfee and McMillan). Yet, only 8 percent of the total agricultural quotas binding are allocated using this method. The economic analysis of first-come, first-served can be traced to the waiting-in-line literature (Holt and Sherman; Suen; Deacon and Sonstelie), while that of historical shares has been analyzed in the case of agriculture by Gervais and Surprenant. However, the most common method of licenses-on-demand has not been given any formal attention. Yet it is the most common method used in 47% of cases in the agricultural sector when the quota is binding. It also is used as a quota administration method in sectors other than agriculture such as for textile import quotas in the multi-fiber agreement. Our paper aims to fill this gap in the literature.

Like Krishna and Tan, and McCorriston and Sheldon, we focus on the implementation of import quotas in terms of the rules governing allocation and the effects on the distribution of rents and economic efficiency. Our paper makes two major contributions, one rather technical and the other involving policy
implications for implementation of the licenses on demand as a system of license allocation. First, we analyze the Nash equilibrium of a game where prizes are allocated on a prorated basis and bidding is costless and show that a unique equilibrium exists for any given number of bidding firms. Second, we determine the conditions affecting the degree of inefficiency and rent shares among firms and compare the type and degree of inefficiency to quota administration methods.

Although licenses-on-demand are important but not analyzed in the literature, there is a significant literature on the effects of auctions and first-come, first-served methods of allocation. If licenses are auctioned, then no rents are dissipated and the revenue to the government is the tariff equivalent of the quota (Bergsten et al.; Krishna 1990; 1993a,b; Holt). This assumes perfect competition, perfect information, certainty and the like. Krishna (1990; 1993a,b) for example, shows the effects of imperfect competition in reducing auction revenues, resulting in rent transfers away from taxpayers. Licenses are awarded to the lowest cost firms when auction is used as a method of quota administration, resulting in an efficient allocation of resources. The waiting-in-line literature implies rent dissipation through rent seeking (Barzel; Suen). Barzel shows that with perfect information rents are completely dissipated through waiting-in-line if agents are identical. With asymmetric valuations of the good, rent is dissipated only for the marginal participant. The infra-marginal participants are still able to earn a part of the rent, resulting in under-dissipation of rents. Suen argues that the degree of rent dissipation depends on the degree of heterogeneity in individual valuations.

An efficient tariff quota administration method is the one that allows for full utilization of the import quota and its allocation to the lowest cost firms. Rules such as tradability of quota licenses and license fees will affect the incentives for utilizing tariff quotas. Furthermore, underlying economic conditions like constant marginal costs affect efficiency of resource allocation. The method of allocating quotas can have important implications for the impact of trade liberalization. For example, it is shown in the literature that if licenses are allocated inefficiently by the historical share method to higher cost importers, a reduction of in-quota tariffs may result in increased quota fill, whereas an increase in the quota may
result in quota under-fill (IATRC). However, we will show that the opposite occurs with a licenses-on-demand method of allocating import quotas.

3. Model of Licenses-on-Demand

Consider a small country imposing an import quota for a single commodity. The government allocates the rights to this quota by a licenses-on-demand allocation system. Licenses-on-demand allocate licenses based on the level of licenses requested by individual firms. If the sum of all requests is greater than the quota, then each firm receives a prorated number of licenses. That is, if firm \( i \) asks for \( Q_i \) licenses, then it’s share of total licenses available will be \( Q_i / \sum_i Q_j \), where \( n \) is a number of importing firms.\(^2\) No individual firm is allowed to request more licenses than the quota itself.

The licenses are non-tradable. Only \textit{bona fide} importing firms qualify for licenses in the current period, which means that each firm had to be an importer in the previous period. Thus, the number of importing firms is determined from the previous period and is known to everyone. Each importing firm derives per unit revenues determined by the difference between the world price \( P_w \) and the domestic price \( P_d \), net of the in-quota tariff in case of tariff quotas.\(^3\) This “quota rent” per unit imported, \( (P_d - P_w) \), is the same for all firms. Each firm has an increasing cost structure of the form \( C_i(q_i) \), where \( q_i \) is the actual import quantity for firm \( i \). We consider both the cases of identical and asymmetric cost structures. Perfect information is assumed, such that firms know each other’s cost functions and desired quantities.

The game proceeds in two stages. In the first stage, firms determine their desired quantities that they want to import \( q_i^\ast, i=1,2,\ldots,n \). Based on their desired quantities, in the second stage firms determine the quantities they ask for (bids), \( Q_i^\ast, i=1,2,\ldots,n \). Working backwards, we first determine the optimal bids for given target quantities and then determine the optimal quantities desired.

\(^2\) The firms are assumed to be importers only, with no domestic production. Although our analysis is carried out for an importing country, the results hold for allocating export licenses as well.
Stage 2:

For given target quantities, firm $i$ asks for $Q_i$ licenses such that they receive $q_i$ as close as possible to the desired (target) import quantity $q_i^*$, subject to the constraint that no individual firm can ask for more licenses than the quota itself, $Q_i \leq \overline{Q}$ for all $i = 1, 2, \ldots, n$. Initially, we assume that there is a prohibitive penalty imposed for leaving licenses unused, such that each firm will use all the licenses it receives. For example, the European Union requires a down payment of 20 percent of the value of imports and the firm loses this if no imports occur that year. For a linear marginal cost function, the per unit profits foregone for importing less than the desired quantity is the same as that of importing more than the optimal amount.

The optimal quantities asked for (bids) $Q_i^*$ are as follows:

If $\sum_j^n q_j^* \leq \overline{Q}$, then $Q_i^* = q_i^*$ for all $i = 1, 2, \ldots, n$. This is because the government allocates the exact number of licenses requested when the number of requested licenses is less than the quota. Thus, every firm bids its desired quantity. Note that quota may be under-filled in this case.

If $\sum_j^n q_j^* > \overline{Q}$, then the government allocates the prorated number of licenses such that

$$q_i = \frac{Q_i}{\sum_j^n Q_j} \cdot \overline{Q}.$$

Thus, the optimal bid $Q_i^*$ is such that $q_i^* = \frac{Q_i^*}{Q_i^* + Q_j^*} \cdot \overline{Q}$ if $Q_i^* < \overline{Q}$ and $Q_i^* = \overline{Q}$ otherwise. The optimal quantity of each firm is a function of the other firms’ optimal bids and the solution is thus a Nash equilibrium. Because the firms will bid their desired quantity otherwise, the case $\sum_j^n q_j^* > \overline{Q}$ is the only interesting one to analyze.

**Symmetric firms:**

---

3 For the analysis to follow, details such as in-quota and out-of-quota tariffs and out-of-quota and over-quota imports are ignored.
When all firms have identical cost structures, $C_i(\cdot) = C(\cdot)$, their desired import quantities are also identical, $q_i^* = q^*$. In equilibrium, all firms ask for $Q_i^* = \overline{Q}$ licenses and receive $q_i = q = \overline{Q} / n$. This is because since $\sum_j q_j^* > \overline{Q}$, the allocated quantity is less than the desired quantity for a representative firm, $q^* > \overline{Q} / n$. Thus, no firm has an incentive to unilaterally lower their bid, as this would bring them even further from their desired quantity. Also, no bid below $\overline{Q}$ is a Nash equilibrium because firm i then has an incentive to unilaterally raise their bid and thus get closer to their desired quantity $q^*$.

**Asymmetric firms:**

Now consider asymmetric cost structures between individual importing firms. The ensuing desired quantities are rank ordered on the interval $(0, A)$, where $A \in R^+$, $0 < q_1^* < q_2^* < \ldots < q_n^*$. If $A > \overline{Q}$, all the firms whose desired quantities exceed or equal the quota automatically bid the quota amount. The optimal bid issue is interesting only for firms who desire to import a quantity lower than the quota. We therefore assume that no single firm has sufficient capacity to import the entire quota; i.e., $A < \overline{Q}$. Note that the lower indices indicate firms with high importing costs, as high cost firms desire lower quantities.

To illustrate the Nash equilibrium, first consider two importing firms, $n = 2$, with different cost structures whose desired import quantities are $q_1^* < q_2^*$ and $q_1^* + q_2^* > \overline{Q}$.

**Proposition 1:** In equilibrium, the low cost firm always bids the maximum amount. The high cost firm either also bids the quota or bids such that it receives exactly the amount desired. The high cost firm bid depends on the desired level of licenses. That is, firm 1 and 2’s ‘best response functions’ are:

\[
Q_2^* = \overline{Q} \quad \text{and} \quad Q_1^* = \frac{\overline{Q} \cdot q_1^*}{\overline{Q} - q_1^*} \quad \text{if} \quad q_1^* < \overline{Q} / 2 \quad \text{and} \quad Q_1^* = \overline{Q} \quad \text{otherwise.}
\]
The lower cost firm will always bid the maximum; i.e., the quota amount and the higher cost bids such that it receives the desired amount if this is less than a half of the quota amount. Otherwise the higher cost firm also bids the quota. Thus, the higher cost firm gets exactly or closer to what it wants. This is because in competing with the other firm the lower cost firm is the first one to have the constraint binding and thus is always constrained by the maximum.

Proof: Consider $Q_2^* = q_2^*$ and $Q_1^* = q_1^*$. This is obviously not a Nash equilibrium since $q_2^* > q_2^* \frac{Q_2^*}{q_2^* + q_1^*} \overline{Q}$ and $q_1^* > q_1^* \frac{Q_1^*}{q_2^* + q_1^*} \overline{Q}$ and the firms have an incentive to deviate by raising their bids. Consider firm 2 raising its bid to $Q_2^* = q_2^* + k$. Firm 1 can be at least as well off as before by raising its bid by the same amount $k$. Thus, firm 1 can match the increase until $Q_2^* = \overline{Q}$. Firm 1 now optimizes its bid knowing what firm 2 bids. Firm 1 would like to set $Q_1^*$ to satisfy the equation

$$q_1^* = \left( \frac{Q_1^*}{Q_1^* + Q} \right) \overline{Q} \text{ subject to } Q_1^* \leq \overline{Q}$$

This will be possible if

$$\left( \frac{q_1^*}{Q - q_1^*} \right) \leq 1 \text{ or } q_1^* \leq \frac{\overline{Q}}{2}.$$

For $n > 2$, we again assume rank ordering of desired quantities $0 < q_1^* < q_2^* < \ldots < q_n^*$ and $\sum_j q_j^* > \overline{Q}$.

**Proposition 2:** There exists a unique equilibrium, where the lowest cost firm always bids the quota amount and additional $n-t-1$ low cost firms may also bid the maximum quota. The remaining $t$ firms bid such that they receive exactly their desired levels of licenses. The equilibrium profile for the first $t$ firms, $t$
< n, can be found by solving the system of t linear equations \( \left( \frac{Q}{q_i} - 1 \right) Q_i^* - \sum_{j=1, j \neq i}^t Q_j^* = (n-t)\bar{Q} \) for 

\( i = 1, \ldots, t \) subject to the constraint that \( Q_i^* \leq \bar{Q} \).

The number of firms that will bid \( \bar{Q} \) is \( n-t \). It will be shown that a unique equilibrium exists for the highest \( t \) that satisfies the above equation. Firms \( 1, \ldots, t \) actually get the allocation that they desire, while firms \( t+1, \ldots, n \) all get the same allocation – an allocation less than what they desire, where the difference is progressively larger for firms with lower importing costs.

For all firms bidding \( \bar{Q} \), higher cost firms get the same amount of licenses as low cost firms even though the desired amount of the latter firms is higher. Therefore, high cost firms get closer to their desired levels than low cost firms. High cost firms are effectively able to ‘overbid’ more than the low cost ones and thereby receive licenses closer to their desired levels. For firms who bid less than \( \bar{Q} \), the desired level of licenses is obtained.

A closed form solution for \( Q_i \) (firm \( i \)'s `best response`) is presented in a later proposition.

**Proof:** Each firm \( i \) would like to set \( Q_i \) to match the quantity \( q_i = \left( \frac{Q_i}{\sum_{j=1}^n Q_j} \right) \cdot \bar{Q} \) to \( Q_i^* \) subject to the constraint that no individual bid can be greater than the quota amount. Noting that 

\[
\frac{\partial q_i}{\partial Q_i} = \sum_{j=1, j \neq i}^n Q_j \left( \sum_{j=1}^n Q_j \right)^2 > 0,
\]

we see that a firm can increase (decrease) its allocation by increasing (decreasing) \( Q_i \). Suppose a particular \( n \)-tuple \( Q_1^*, \ldots, Q_n^* \) is in equilibrium. Then there can be no incentive for a firm \( i \) to increase it's bid \( Q_i^* \); i.e., the number \( \left( \frac{Q_i}{\sum_{j=1}^n Q_j} \right) \cdot \bar{Q} \) cannot be smaller than
$q_i^*$. Likewise there can be no incentive for firm $i$ to decrease its bid; i.e., the number

$$\left(\frac{Q_i^*}{\sum_{j=1}^{n} Q_j^*}\right) \cdot \overline{Q}$$

cannot be larger than $q_i^*$. In equilibrium, there are two possibilities:

$$Q_i^* < \overline{Q} \text{ in which case } \left(\frac{Q_i^*}{\sum_{j=1}^{n} Q_j^*}\right) \cdot \overline{Q} = q_i^*$$

or

$$Q_i^* = \overline{Q} \text{ in which case } \left(\frac{Q_i^*}{\sum_{j=1}^{n} Q_j^*}\right) \cdot \overline{Q} \leq q_i^*.$$ 

Note our assumption that $0 < q_1^* < q_2^* < ... < q_n^*$. Let $q_i^* < q_j^*$ and let $(Q_1^*, ..., Q_n^*)$ be an equilibrium n-tuple. It can be shown that the bids are rank ordered in the same way as the desired quantities, $Q_i^* < Q_j^*$ (see Appendix).

By an argument similar to that used for the case $n = 2$, we can show that $Q_n^* = \overline{Q}$. From the previous result, we know that there is some maximal integer $t < n$ such that

$$Q_1^* < Q_2^* < ... < Q_t^* < \overline{Q} = Q_{t+1}^* = ... = Q_n^*.$$ 

Suppose for the moment that we know $t$. To find the equilibrium profile, we need to solve the system of equations:

$$\left(\frac{Q_i^*}{\sum_{j=1}^{n} Q_j^*}\right) \cdot \overline{Q} = q_i^* \text{ for } i = 1,...,t.$$ 

or, recalling that, $Q_{t+1}^* = ... = Q_n^* = \overline{Q}$.
This is a system of \( t \) linear equations in \( t \) unknowns. As we do not know the integer \( t \) beforehand, we must start solving the system for \( t = n - 1, n - 2,... \) and find the highest \( t \), for which the system of equations yields a feasible equilibrium. It can be shown that the defining equations are also solvable for integers less than \( t \); i.e., these solutions also satisfy the necessary conditions, but they do not constitute a Nash equilibrium. Consider a solution where \( n-s \) firms, \( s \in \{0,1,...,t-1\} \), bid \( Q \), i.e., \( Q_{s+1} = ... = Q_{t}^* = Q \). Then the first \( s \) firms receive their desired allocation and \( n-s \) firms receive less than what they desire. However, any firm \( j \), where \( j = s+1, s+2,..., t \), can get closer to their allocation by lowering their bid below \( Q \). A unique Nash equilibrium occurs at the maximal integer \( t \), for which the system of defining solutions solves and that satisfies the necessary conditions.

Proposition 2a shows how to determine the largest value of \( t \) for which the equations hold. Although we can show that the equations hold for all smaller (positive) values of \( t \), these solutions do not constitute a Nash equilibria and are therefore of no interest to us.

**Proposition 2a:** The closed form solution for an equilibrium bid for the largest \( t \)-th firm is

\[
Q_i^* = \left( \frac{(n-t)Q \cdot q_i^*}{Q - \sum_{j=1}^{t} q_j^*} \right),
\]

where \( Q_t^* \) satisfies the conditions \( Q - \sum_{j=1}^{t} q_j^* > 0 \) and

\[
(n-t) \cdot q_i^* \leq Q - \sum_{j=1}^{t} q_j^*.
\]

**Proof:** see Appendix.
The following examples for \( n = 3 \) illustrate the procedure for finding the Nash equilibrium. In example 1 both higher cost firms are perfectly satisfied, while in example 2 only the highest cost firm is perfectly satisfied. Example 3 illustrates the case when none of the firms obtained their desired allocation - all firms bid at the maximum and none obtain what they asked for exactly. But even then, the higher cost firm allocation is closer to their desired quantity.

**Example 1:**

Let \( n = 3 \), \( \overline{Q} = 1.0 \), and \( q_1^* = 0.2, q_2^* = 0.3, \) and \( q_3^* = 0.8 \). We know that \( Q_3^* = 1 \). At \( t = 2 \), the appropriate equations are

\[
(1/0.2 - 1)Q_1 - Q_2 = 1
\]

\[
- Q_1 + (1/0.3 - 1)Q_2 = 1
\]

and the solution is \( Q_1^* = 0.4, Q_2^* = 0.6 \). The equilibrium 3-tuple is \( (0.4,0.6,1.0) \) and the corresponding allocations are \( (0.2,0.3,0.5) \). Firms 1 and 2 obtain their desired allocation, and Firm 3 gets the best allocation that it can with a bid of 1.

Now suppose that \( t = 1 \). Then setting also \( Q_2^* = 1 \), the appropriate equation is

\[
(1/0.2 - 1)Q_1 = 2 \text{ or } Q_1^* = 0.5
\]

But the triplet of bids \( (0.5,1,1) \) and the corresponding allocation \( (0.2, 0.4, 0.4) \) is not a Nash equilibrium because firm 2 would want to lower its bid in order to get an allocation closer to 0.3.
Example 2:

Now let $\overline{Q} = 1.0$, and $q_1^* = 0.3, q_2^* = 0.5$, and $q_3^* = 0.8$. Again we know that $Q_3^* = 1$. If we assume $t = 2$, we expect $Q_1^*, Q_2^* < 1$. The appropriate equations are

$$(1/0.3 - 1)Q_1 - Q_2 = 1$$

$$- Q_1 + (1/0.5 - 1)Q_2 = 1$$

and the solution is $Q_1^* = 3/2, Q_2^* = 5/2$. which violates the condition that $Q_i^* \in [0, \overline{Q}]$. The solution does not satisfy necessary conditions. Now assume that $t = 1$, so that $Q_2^* = 1$ as well. The appropriate equation is $(1/0.3 - 1)Q_1 = 2$, resulting in $Q_1^* = 6/7$. The corresponding allocations are $(0.3, 0.35, 0.35)$. Firm 1 gets its desired allocation, and Firms 2 and 3 do not. This solution satisfies both necessary and sufficient conditions.

Example 3:

Now let $\overline{Q} = 1.0$, and $q_1^* = 0.5, q_2^* = 0.6$, and $q_3^* = 0.8$. Assume $t = 2$. The appropriate equations are

$$(1/0.5 - 1)Q_1 - Q_2 = 1$$

$$- Q_1 + (1/0.6 - 1)Q_2 = 1$$

and the solution is $Q_1^* = 6, Q_2^* = -7$. This is not feasible. Now assume that $t = 1$. The appropriate equation is $(1/0.5 - 1)Q_1 = 2$ and the solution is $Q_1^* = 2$, again infeasible. The only equilibrium 3-tuple for this example has $Q_i^* = \overline{Q}$ for all $i$. 
Stage 1:

Knowing how firms will behave in stage 1, each firm strategically decides about its desired quantity.

Proposition 3: Behaving strategically, firms’ desired quantities are Cournot quantities if the total quantity desired in Cournot equilibrium is less than the quota but competitive quantities if the total quantity desired in Cournot equilibrium exceeds the quota.

Without the interaction of the second stage, each firm behaves strategically by maximizing its profit given the quantities imported by the other firms. The equilibrium yields the Cournot quantities $q_i^C$ for all $i$ as the optimal quantities desired. If the total unconstrained quantity desired in Cournot equilibrium is less than the quota, $\sum_{i=1}^{n} q_i^C \leq Q$, the quota is undersubscribed. Without the quota constraint, all firms bid and also receive their Cournot quantities, $Q_i^* = q_i^C = q_i$ and the quota is underfilled.

If the total unconstrained quantity desired in Cournot equilibrium exceeds the quota

$$\sum_{i=1}^{n} q_i^C > Q,$$

the quota is oversubscribed and will be filled. This is because each firm bids at least their desired quantity, $Q_i^* = q_i^* = q_i^C$, thus the sum of the bids exceeds the quota and the actual quantities awarded will be determined on a prorated basis.

But if the quota is filled, the price is given and the Cournot quantities are no longer the optimal desired quantities. In fact, each firm takes the price (at the quota filled level) as given and thus re-solves for its optimum quantity desired, namely each firm would like to import their competitive quantity $q_i^{ce}$. Thus, the competitive quantities become the relevant desired quantities in this case, $q_i^* = q_i^{ce}$, for $i=1,2,\ldots,n$.

4. Economics of licenses-on-demand

This section discusses the economic implications of the theoretical results for the game that characterizes licenses-on-demand method and compares the type and degree of inefficiency to auctions and first-come, first-served methods. Factors affecting the degree of inefficiency are identified as well as the effects of
trade liberalization through a decrease in the in-quota tariff (for tariff-rate quotas) and an increase in quota.

One of the main results of the above analysis is that the higher cost the firm, the closer their license allocation to their desired quantity and the higher the probability that they receive exactly their desired quantity. This implies inefficiency in resource allocation. Allocating licenses-by-demand enables high cost firms to remain in business and achieve license allocation that under an efficient outcome would be assigned to low cost firms. The inefficient distribution of licenses across firms generates economic waste of quota rents with the extra costs associated with importation. Thus, relative to both the auction and first come first served methods, a license on demand administration method generates waste through its inability to eliminate the inefficient firms.

No rent seeking occurs under licenses on demand. This is a result of the one-period setting that we assume for our analysis, costless bidding, and no entry allowed. Being a bona fide importing firm ensures a firm the right to compete for licenses, thus no firm has an incentive to spend resources in the effort to obtain this right. Also, no resources are expended when competing for licenses, as bidding is costless. This result is the same for a one-time auction; since only the participants who can bid the highest amount receive and pay for the licenses and there is no incentive to rent seek. Under first-come, first-served, however, resources are spent by waiting-in-line and rents are dissipated either completely (if identical valuations) or incompletely (asymmetric valuations).

As long as the total quantity demanded exceeds the quota amount, the quota will be fully filled. This is the result of our assumption of a high penalty for failure to use awarded licenses. This penalty may take a form of a non-refundable down payment when firms receive their awarded licenses at the beginning of the year. In case of agriculture, European Union importers are often
required to pay 20% of the value of the product in advance. We relax this assumption in the next section. If perfect information is assumed, both auction and waiting-in-line result in a 100 percent fill rate.

The presence of fixed costs has an indeterminate effect on deadweight cost. It may decrease the inefficiency if the fixed costs prevent the highest cost importers from bidding. Relative to the case with zero fixed costs and the same variable cost function, fixed costs add to the deadweight cost because they constitute an additional cost for the firms that should be out of business. The same reasoning applies to a decrease in the revenue margin due to an increase in the in-quota tariff (for tariff quotas) or the world price, or a decrease in the domestic price due to some economic shock. The absolute level of inefficiency may decrease due to an elimination of the highest cost importers but relative distortion for the remaining firms still exists. An increase in the quota, on the other hand, decreases inefficiency through reducing the relative distortion. The high cost firms that originally received their optimal amount will remain at their desired levels and the quota expansion is transferred to the lower cost firms. In addition, per unit rent declines, increasing the probability that the highest cost firm will not participate. Hence, trade liberalization by reducing in-quota tariffs or expanding quotas will reduce the inefficiency of the licenses-on-demand method. When a limit on the licenses received by each firm is imposed, inefficiency increases because limit will be more binding on low cost firms. This result is in sharp contrast to that of non-transferable rents with historical allocation (as assumed in Lott, for example).

If firms are allowed to trade the licenses, Pareto optimality of the equilibrium outcome is restored. All firms bid the maximum allowable amount \( Q^* = \bar{Q} \) for all \( i = 1, 2, \ldots, n \) and each firm is allocated \( \bar{Q} / n \) licenses. Subsequently, the high cost firms sell their licenses to the low cost firms. Thus, efficiency is maximized but rent is transferred to the high cost firms.
The following section explores the implications of relaxing (one by one) several assumptions of the base model. We consider the case of constant marginal costs and zero non-use penalty.

5. **Constant marginal cost and zero non-use penalty**

*Constant marginal cost*

Assume now that each importing firm has different variable costs that are invariant to the level of imports. Each firm bids the maximum \( \overline{Q} \) and so receives \( \overline{Q}/n \) licenses. The deadweight costs for the example of \( n = 4 \) in Figure 1 are given by the shaded areas. If fixed costs are also present, then deadweight costs increase by the fixed costs of the two high cost firms remaining in the bidding and decrease by the deadweight cost of the highest cost firm that exits. An increase in the per unit quota rent brought about by either a decrease in the in-quota tariff or change in world or domestic prices will result in a decline in deadweight costs if the highest cost firm is forced to exit the bidding process. The per firm deadweight costs of the remaining higher cost firms increase but total deadweight costs decline. The identical effect on deadweight costs occurs if a license fee is imposed that forces the highest cost firm to exit.

**Figure 1: Deadweight costs for four firms with different levels of constant marginal cost**

An increase in the quota through trade negotiations will have an indeterminate effect on deadweight costs because domestic prices fall (and world prices increase if a large importer) and
so high cost importers may exit the bidding process (the outcome depends on the elasticity of excess demand (and excess supply if a large country importer)). If there is no exit, then deadweight costs increase. A limit on the import licenses received by a firm can increase (but not decrease) deadweight costs but quota under fill never occurs under the assumptions of our model unless the total quantity demanded is less than the quota. Note that the variance of costs is important in addition to the level of costs among importing firms in determining the degree of economic inefficiency.

**Zero non-use penalty**

In many cases, the government does not impose a penalty for leaving licenses unused. Since bidding for licenses and consequently leaving them unused is costless under this scenario, each firm will leave unused the licenses that exceed its desired amount. All firms ask for the maximum number of licenses equal to the quota amount, \( Q^*_i = Q \) for all \( i = 1, 2, \ldots, n \) and each firm is allocated \( Q/n \) licenses.

Inefficiency increases because low cost firms on average get fewer licenses than with a penalty for unused quotas and quota under fill occurs that creates an additional dead-weight cost.

Transfer of licenses to the high cost participants increases the gap between the desired and actual number of licenses for the low cost participants. In addition, these transferred licenses are left unused by the high cost firms as these only retain their desired levels.

**Proposition 4:** When no penalty is imposed on non-use of licenses, the degree of rent dissipation increases with greater heterogeneity of firms.

An increase in variance of importing cost structures under no penalty causes an increase in the number of licenses left unused and thus an increase in inefficiency. This is because the higher the cost variation, the more very high cost firms there are. The difference between the desired quantity and the allocated
quantity is larger for these, generating more unused licenses that under penalty were allocated to lower cost firms.

**Proof:** The number of licenses unused, $L_U$, is the difference between the allocated quantity under no penalty and the desired quantity of the $t$ firms that would receive these quantities under penalty.

$$L_U = \sum_{i=1}^{t} \left( \frac{Q}{n} - q_i^* \right).$$

Let $q_i^* = \bar{q} + \sigma_q \varepsilon_i$, where $\varepsilon_i$ has a mean 0 and variance 1 for $i = 1, 2, ...n$. Then

$$L_U = \frac{1}{n} Q - t\bar{q} - \sigma_q \sum_{i=1}^{t} \varepsilon_i, \quad \text{and}$$

$$\frac{\partial L_U}{\partial \sigma_q} = -\sum_{i=1}^{t} \varepsilon_i > 0.$$

This is because $\sum_{i=1}^{n} \varepsilon_i = 0$, $t < n$ and $\varepsilon_i < \varepsilon_{i+1} < ... < \varepsilon_n$, resulting in $\sum_{i=1}^{t} \varepsilon_i < 0$.

6. **Concluding remarks**

In this paper, we discuss the implications of rationing import quotas by licenses-on-demand where licenses are based on the level of licenses requested by individual firms. If the sum of all requests is greater than the quota, then each firm receives a prorated number of licenses-on-demand. We assess the effects of such a scheme on rent shares and economic efficiency, and compare these to auctions and first-come, first-served methods. This paper analyzes the Nash equilibrium for a game when multiple licenses are allocated on a prorated basis and shows that a unique equilibrium exists for any number of firms. Numerical examples provide insights into the properties of the model and economic implications of alternative situations. We show that the interaction between determining desired quantities and subsequently determining the quantities asked for (bids) causes the firms to choose Cournot quantities in
the unconstrained case (if the sum of Cournot quantities desired is less than the quota) and competitive quantities if the total Cournot quantity desired exceeds the quota. We then analyze the competition of import quota licenses with licenses-on-demand under different economic conditions. We identify several conditions affecting the degree of inefficiency and rent shares among firms.

Under conditions of perfect competition and perfect information, we analyze the implications of various factors actually observed like fixed costs, penalty for non-use of licenses, and a limit on the number of licenses a firm is allowed to receive. Effects of trade liberalization are examined through an increase in the quota amount and a change in the per unit quota rent due to a change in tariffs or market prices. In addition, conditions affecting quota under-fill are identified. As expected, this system of allocation is more inefficient than a well functioning auction or tradable license market. We show that the level of waste depends more on the variance in the cost structure among firms than on the level of costs.

Interestingly, the licenses-on-demand method has very different implications for the impact of trade liberalization where a reduction of in-quota tariffs may result in decreased quota fill, and an increase in the quota may result in a 100 percent quota fill rate. This result is opposite to that of when licenses are allocated on an historical share basis, for example, where opportunity costs may change overtime, preventing the least cost-importing firms from operating (Lott). Other policy implications for the implementation of licenses-on-demand include the efficacy of non-use penalties and the inefficiency of (a) license fees (equivalent to an increase in in-quota tariffs); and (b) limits on the number of licenses a firm is allowed to receive. Some countries and commodity sectors implement each of these rules for the licenses-on-demand allocation method.

Future research should focus on relaxing the assumption of a fixed number of importing firms. Rents may be such that it pays for firms to enter the business by importing at out-of-quota tariffs in the short run and so gain status as an importer to bid for licenses. This will have the effect of dissipating rents further, not only because of the rent seeking costs of importing at out-of-quota tariffs but also in adding to the
number of higher cost importing firms. The analysis can also be extended to include imperfect information and risk aversion on the part of firms when they make strategic decisions in their bidding levels.
References


International Agricultural Trade Research Consortium (IATRC) 2001 Issues in Reforming Tariff-rate Import Quotas in the Agreement on Agriculture in the WTO Commissioned Paper #13 (May), University of Minnesota, St Paul, Minnesota.


World Trade Organization (WTO), 2000. *Tariff Quota and Other Quotas* Background paper G/AG/NG/S/7 May 23
Appendix

Proof that bids are rank ordered in the same way as the desired quantities:

There are four possible cases:

\[ Q_i^* = Q_j^* = \bar{Q}, \]

\[ Q_i^* < \bar{Q}, Q_j^* = \bar{Q} \]

\[ Q_i^* = \bar{Q}, Q_j^* < \bar{Q} \text{ and } \]

\[ Q_i^*, Q_j^* < \bar{Q}. \]

Cases 1 and 2 are trivial. In case 3, the inequality \( Q_j^* < \bar{Q} \) implies that

\[ q_j^* = \left( Q_j^* / \sum_{k=1}^{n} Q_k^* \right) \cdot \bar{Q} \]

so that

\[ \left( Q_j^* / \sum_{k=1}^{n} Q_k^* \right) \cdot Q_i^* = q_j^* \cdot. \]

Then \( Q_j^* < \bar{Q} \) implies that

\[ \left( Q_i^* / \sum_{k=1}^{n} Q_k^* \right) \cdot \bar{Q} > q_j \left( Q_j^* / \sum_{k=1}^{n} Q_k^* \right) \cdot \bar{Q} > q_j^* \cdot. \]
Since the left-hand side is smaller than or equal to $q_i^*$, we have the contradiction $q_i^* > q_j^*$ and we conclude that this case is infeasible.

The last case implies that

$$
\left( \frac{Q_i^*}{\sum_{k=1}^{n} Q_k^*} \right) \bar{Q} = q_i^* < q_j^* = \left( \frac{Q_j^*}{\sum_{k=1}^{n} Q_k^*} \right) \bar{Q}
$$

so that

$$
Q_i^* / \sum_{k=1}^{n} Q_k^* < Q_j^* / \sum_{k=1}^{n} Q_k^*
$$

and the result is immediate.

**Proof of proposition 2a:**

The general form of the equations is:

$$
\left( \frac{Q_i^*}{q_i^* - 1} \right) Q_i - \sum_{j \neq i}^{t} Q_j = (n - t) \bar{Q} \quad \text{for } i = 1, \ldots, t
$$

For each $i = 1, \ldots, t$, subtract equation $t$ from equation $i$ to get $Q_i - (q_i^* / q_t^*) Q_i = 0$. Note that the ratio $(q_i^* / q_t^*)$ is smaller than 1 under our assumption. This means that if $Q_i^* \in (0, \bar{Q}]$ we are done. To determine $Q_t$, substitute the above expressions into the equation

$$
\left( \frac{\bar{Q}}{q_t^* - 1} \right) Q_t - \sum_{j=1}^{t-1} Q_j = (n - t) \bar{Q} \quad \text{for } i = 1, \ldots, t
$$

and solve for it. You get the closed form solution.
\[ Q_t^* = \left( \frac{(n-t) \cdot \overline{Q} \cdot q_t^*}{\overline{Q} - \sum_{j=1}^{t} q_j^*} \right) \]

Recalling that we are done if \( Q_t^* \in (0, \overline{Q}] \) we see that the only permissible \( t \)'s are those such that

a) \( \overline{Q} - \sum_{j=1}^{t} q_j^* > 0 \) so that \( Q_t^* > 0 \) and

b) \( (n-t)q_t^* \leq \overline{Q} - \sum_{j=1}^{t} q_j^* \) so that \( Q_t^* \leq \overline{Q} \).

Any \( t \) such that these two conditions are true will lead to an equilibrium. One might note that condition a) is actually implied by condition b), as the right hand side of b) becomes negative for \( t \)'s larger than the largest \( t \) that satisfies a). The opposite is not true as the following example shows. However since a) is so easy to check, it might be prudent to find the largest \( t \) which satisfies it, and work down from there.

For example, with: \( \overline{Q} = 1 \), and \( q^* = (0.1, 0.2, 0.3, 0.5, 0.6) \) we see that the largest \( t \) for which condition a) holds is \( t = 3 \). Testing condition b), we see that the largest \( t \) for which it holds is \( t = 2 \).

In order to actually compute the equilibrium, you do need to solve the appropriate equations. We have

\[ Q_t^* = \left( \frac{(n-t) \cdot \overline{Q} \cdot q_t^*}{\overline{Q} - \sum_{j=1}^{t} q_j^*} \right) \]

and

\[ Q_t^* = \left( \frac{(n-t) \cdot \overline{Q} \cdot q_t^*}{\overline{Q} - \sum_{j=1}^{t} q_j^*} \right) \]
Finally we show that the equations are satisfied for all \( t \) smaller than the largest \( t \) that satisfies the above equations. Suppose \( t > 1 \) satisfies b). Then we know that

\[
(n-t) \cdot q_i^* \leq \overline{Q} - \sum_{j=1}^{i} q_j^*
\]

or

\[
(n-t+1) \cdot q_i^* \leq \overline{Q} - \sum_{j=1}^{i-1} q_j^*
\]

We want to show that the first of these expressions is true when \( t \) is replaced by \( t-1 \); i.e.,

\[
(n-t+1) \cdot q_{i-1}^* \leq \overline{Q} - \sum_{j=1}^{i-1} q_j^*.
\]

But this follows immediately from the previous expression since we know that \( q_{t-1}^* < q_t^* \).