Economic Analysis of TRQ Administrative Methods

Rationing and Markets

Tariff quota administration amounts to rationing—which may be either efficient or inefficient. This section examines in depth the assets and liabilities of the seven basic methods to administer tariff quotas, particularly by the GATT criteria of quota fill and distribution of trade.

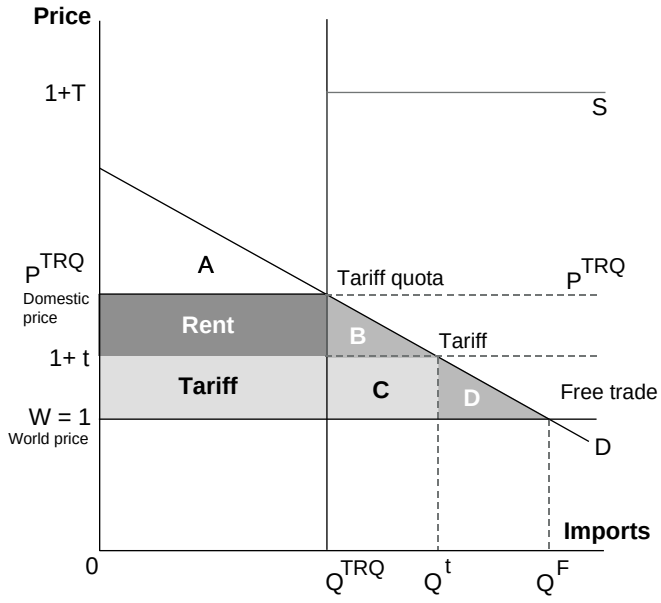
A market generates maximum surplus if all buyers and sellers to the left of the intersection of supply and demand curves can find each other and exchange. Such buyers are inframarginal buyers. Their willingness to pay exceeds the market-clearing price. Extramarginal buyers, represented on the demand curve right of equilibrium, have a willingness to pay less than the market-clearing price. Similarly, inframarginal suppliers have a willingness to accept less than the market-clearing price; extramarginal suppliers, represented on the supply curve right of equilibrium, have a willingness to accept greater than the market-clearing price.

The displacement of inframarginal traders by extramarginal traders is the primary source of inefficiency in TRQ administration. The availability of quota rents provides an incentive for extramarginal traders to enter the market. Market-based administrative methods, auctions, for example, remove the incentive posed by quota rents and thus remove the risk of displacement. The further TRQ administration deviates from market-based administrative methods, the greater the risk of displacement and the greater the risk of biased trade distribution.

Welfare Analysis of the Rationing Problem

The welfare analysis of various methods of TRQ administration is illustrated in figure 4. The international supply curve is a horizontal line at one. The in-quota tariff (t) is applied to the first Q^{TRQ} units of imports, which shifts the effective import supply curve upward to 1 + t until the volume Q^{TRQ} is attained. At the quota volume (Q^{TRQ}), there is a vertical jump in the supply curve. Imports in excess of Q^{TRQ} are charged the over-quota tariff (T > t), so the effective supply curve continues at 1 + T.

Figure 4
Welfare analysis of TRQ rationing



For the case in which import demand is sufficient to fill the quota, but not so great as to import at the over-quota tariff (fig. 4), consider the differences between a tariff quota, a simple tariff, and free trade. With free trade, an unlimited quantity may be imported at the world price. The domestic market clears with imports of Q^F and the domestic price equal to the world price, P = W = 1, and all demand inframarginal to P = 1 is satisfied. The large triangle below the demand curve and above the supply curve is the economic surplus gained from international exchange, that is, the sum of areas labeled: A + Rent + Tariff + B + C + D.

Under a simple tariff, unlimited imports are allowed at the in-quota tariff (t). The domestic market clears with imports of Q^t and the domestic price is 1 + t. Demand inframarginal to P = 1 + t is satisfied. Domestic consumers' surplus equals the area: A + Rent + B, and the domestic government collects the area Tariff + C in tariff revenue. Triangle D is the deadweight welfare loss from imposing the tariff.

The tariff quota limits imports at the in-quota tariff to Q^{TRQ} and the domestic market clears at P^{TRQ}, which results in a further reduction in welfare. Domestic consumers' surplus is reduced to the triangle labeled A and tariff revenue declines to the rectangle labeled "Tariff." The area "Rent" represents the arbitrage profits from the opportunity to import Q^{TRQ} units at the cost of 1 + t, while the domestic market value is P^{TRQ}. These tariff quota rents are the result of rationing Q^{TRQ} units of supply over Q^t units of demand. Rents

are neither good nor bad, they simply exist as the result of rationing the opportunity to bridge the gap between domestic and world prices. Because there is no clear title to these rents, they are a common property resource, and, as such, the fact that they exist can stimulate wasteful rent-seeking behavior. How rights to the rent are distributed largely determines the pattern and volume of exchange.

If imports are limited to Q^{TRQ}, the greatest welfare possible is A + Rent. Drawing the figure in this textbook manner assumes that a market mechanism is employed to ration Q^{TRQ} units of supply over Q^t units of demand. That is, it assumes the rent is perfectly and automatically allocated in the best possible manner, a strong assumption. If it does not hold, the outcome can be far inferior to the auction allocation.[3]

Auction Allocation

To minimize biased trade, given a tariff-quota constraint on imports, extramarginal suppliers must be excluded. This is the beneficial, discriminatory role that prices play in free markets. The quota rent and the incentive it transmits induces inefficiencies. Demand inframarginal to W but extramarginal to P^{TRQ} will enter the market as will supply inframarginal to P^{TRQ} but extramarginal to W. If there were no quota rents, only inframarginal traders would enter the market and welfare would be maximized. An auction neutralizes quota rents.[4] The opportunity to buy something for W and sell it for $P^{TRQ} – (1 + t)$ is worth $R = P^{TRQ} – (1 + t)$. Inframarginal traders will bid R for the opportunity; extramarginal firms bid less than R; and the required discrimination is realized.

In a TRQ auction, consumers would bid, at the margin, $R = P^{TRQ} - (1 + t)$—the difference between the domestic price P (given imports of Q) and the world price plus the in-quota tariff, 1 + t. If all winning bids are charged the marginal winning bid (uniform price auction) then auction revenue is the shaded rectangle "rent." The consumers who obtain the quota rights are those with a willingness to pay of at least P^{TRQ}. These consumers realize a consumer surplus equal to the area A. The domestic economy realizes gains from trade equal to the auction revenue plus A. This allocation is identical to the allocation that would result from the tariff-equivalent of a tariff quota (given market conditions): $t^* = t + R$.

Suppose we employ an alternative allocation method, for example, first-come, first-served (FCFS). Under FCFS, any consumer with a willingness to pay more than W will attempt to import the instant the tariff quota season opens. The quota rent attracts all importers inframarginal to W, many of whom are extramarginal to W + R. To perform as well as an auction, alternative allocation methods must discriminate perfectly between agents inframarginal to W + R from those extramarginal to W + R.

Only 4 percent of all TRQs notified to the WTO in 1999 used auctions. If auctions are so wonderful, the question arises as to why they are so seldom used. Economically, auctions are most likely to outperform other rationing methods only if the market for the controlled product is sufficiently liquid—if the market has a large volume of trade and several competing traders. As a market becomes less liquid, its capacity to function as a price-discovery mechanism deteriorates. Illiquidity has resulted in the demise of many commodity futures contracts.[5] Those commodities for which active futures or cash markets exist are excellent candidates for quota auctions. If illiquidity diminishes the relative efficiency of an auction, then other methods, license on demand, for example, might be preferable. Related to the liquidity of the market is the number of active traders. Research shows that too few traders can diminish market efficiency, but it does not require more than a few entrants to realize close to 100-percent efficiency. Several TRQ allocation methods specify a maximum market share that can be obtained by any single trader. Such rules can limit the risks of having too few traders.

There is also a political explanation for the low number of auctions. Auctions are markets, and markets can be hard to control. If the government administering the TRQ has strong preferences about the countries or firms that receive quota rights, then it will choose not to ration by auction. Similarly, if a government prefers

[3]This point is generally overlooked in the literature. Vousden, 1990, pp. 60-83, for example, devotes a full chapter to quotas without considering how quota rights are allocated.

[4]Vickrey (1961) initiated the economic literature on auctions. Milgrom (1989) provides a cogent introduction to the auction literature. Campbell (1995) is a lucid guide to the more important proofs and places auctions and bidding in a demand-revelation context. Bergsten et al. (1987) advocate auctioning (absolute) quotas.

[5]Silber (1981), Black (1986), and Miller (1986) provide comparative analyses. Working (1953) and Sandor (1973) offer case studies. For the number of traders and liquidity, see Workshop on Double Auctions (1991) and Tomek (1980).

to transfer quota rents to a certain group rather than collecting the rents as auction revenue, it will not auction. As shown below, discretionary allocation—State trading and producer group—provides the importing government or industry the greatest control over the distribution of rent and of trade.

Quasi-Market Allocation

Several allocation methods are a mix of market and random processes. This section examines license on demand (LD) and first-come, first-served (FCFS). It considers, in turn, the risk these methods can pose of quota underfill and of biased distribution of trade.

License on Demand

License-on-demand allocation generally operates in the following manner. Before the quota period begins, potential importers are invited to apply for import licenses. Applicants specify the quantity of imports they want. Call the i^{th} applicant's demand q_i^*, and call the sum of all import application requests $Q^* = \Sigma q_i^*$. If domestic demand is sufficient, the quota binds: $Q^* > Q^{TRQ} = \lambda Q^*$. To ration license supply among license demand, application quantities are reduced proportionally by the factor $\lambda < 1$. If one applies for q^* units and the quota is binding, then a license is granted for λq^* units of imports at the in-quota tariff. Many countries also specify a minimum license amount, so that the allocation rule reads: λq^* units but no less than γ units. This minimum quantity rule can prevent the minimum shipload problem discussed below and in the appendix.

Quota Fill. The proportional reduction of requests complicates importing. First, if a trading firm accurately states its desired import volume and the quota is binding, it gets less quota than desired, which creates an incentive to overstate license requests. If all applicants overstate by the same proportion, the final allocation of shares is not affected. If the degree of overstatement is not uniform, however, shares will be reallocated. In any event, uncertainty makes planning unnecessarily difficult.

Second, shipments of goods are generally conducted in units such as full truck or full container loads. Unit transport costs are less, as a rule, if loads are full, because shipping has substantial fixed costs. Typically, a license request will be for a multiple of full loads, say, for L^* containers. With proportional reduction, only λL^* containers will be licensed for import at the in-quota tariff. Except in the rare case that λL^* results in an integer, proportional reduction will cause a remainder of one less-than-truckload shipment for each license application. Assuming for the moment that licenses cannot be divided and traded following initial allocation, shippers holding rights to a less-than-truckload shipment face what amounts to a rounding decision. Above some fraction of a full load, say, 80 percent, it is still profitable to ship, so the products are shipped. Below the critical load of 80 percent, the shipment is not made and the less-than-truckload portion of the license is not used. These unused remainders can result in underfill.

Distribution of Trade. The risk of a biased distribution of trade from TRQ administration requires an examination of the supply side of the rationing problem (fig. 5). To reduce clutter, assume that the in-quota tariff is zero; that the over-quota tariff is prohibitive, given market conditions; and that the domestic market clears at the price P. The unit quota rent is $R = P - W$. The figure plots an upward sloping supply curve "S." The supply elicited at price P is normalized to one.

The area under the supply curve represents payments to factors employed to produce the traded product. Suppliers extramarginal to W must spend more than W per unit to employ labor, capital, and other productive resources to produce a unit of output with a market value of only W. Rather than adding value, extramarginal production destroys value. This economically wasteful misallocation of resources is represented by the shaded triangle: extramarginal factor use.

License-on-demand allocation may be thought of as a form of lottery. All firms inframarginal to P have an incentive to enter the quota lottery. The sum of applications, as shown above in the analysis of underfill, exceeds the supply of quota: $Q^* = 1 > Q^{TRQ} = \lambda$. Each applicant wins a pro rata share of the global quota, that is, λq^* units. The effective supply curve for this uniform allocation of quota rights is S_0. The proportion of quota rights granted to suppliers inframarginal to W is β. Given the assumptions made, LD allocation causes the displacement of $\lambda(1 - \beta)$ inframarginal suppliers by $\lambda(1 - \beta)$ extramarginal suppliers. This expected distribution of trade differs from the tariff equivalent counterfactual distribution, which consists exclusively of inframarginal suppliers. The welfare loss from LD allocation is shown by the shaded triangle below S_0.

This welfare loss can be interpreted as an indicator of how the expected distribution of trade differs from the tariff equivalent counterfactual distribution of trade. The

maximum surplus possible is realized by market (auction) allocation. Reductions in welfare from this maximum occur because of the displacement of inframarginal suppliers by extramarginal suppliers. The greater the degree of displacement, the greater is the reduction in welfare, and the greater the difference in the distribution of trade from the counterfactual standard.

First Come, First Served

The standard FCFS allocation allows importation at the in-quota tariff until the quota has been filled. If domestic demand is sufficient to fill the quota, then there is one individual shipment that fills the quota; that is, there is one last shipment that will be fully or partially over-quota. Call this last in-quota shipment the n^{th} shipment and represent its volume as q_n. The tariff charged on the n^{th} shipment is variable: the first n-1 shipments enter at the in-quota tariff; only a portion (α) of the nth shipment enters at the in-quota rate with the over-quota balance ($1-\alpha$) being charged the over-quota tariff; and for shipments after n, the over-quota tariff is applied.

Quota Fill. Being the claimant of the n^{th} shipment can be rather costly. Consider some of the alternatives. If it is costly to break the shipment—suppose it is a container or truckload—then the importer must choose whether to cease importing and route the whole shipment elsewhere or to pay the over-quota tariff on the proportion over-quota. If the shipment can be broken, then the over-quota portion can be shipped elsewhere, or forfeited at the border, leaving only the in-quota portion to be imported.

The system just described assumes that all customs agents have timely and accurate information on the level of quota utilization. Several countries have unified electronic monitoring systems and it is possible to identify and inform the n^{th} shipper prior to clearing customs. Many countries lack such systems, however. Information that the quota has been filled is disseminated to customs offices days, weeks, or sometimes months later. It is possible to clear a shipment at the in-quota tariff and be notified later that payment of the tariff difference (T - t) is due on each unit previously cleared. This additional charge will be levied only on shipments after the n^{th} shipment, with the n^{th} shipment liable for payment on $(1 - \alpha)q_n$. If a shipment is fully over-quota only after the fact, it is generally impossible to "unimport" the shipment and avoid an extra tariff. If the over-quota tariff is significantly higher than the in-quota tariff, the cost of being n^{th} or later is very high. In the absence of timely information on quota fill, potential importers may avoid shipping if they believe that the quota is close to being filled and the risk of being caught over-quota is high. This rational risk-aversion can result in a lack of quota-fill, even when domestic prices exceed the in-quota landed value.

From the point of view of the government's choosing to administer via FCFS, there are incentives to delay notification. If collecting tariff revenue is a priority, then sending the signal that the quota is not filled when in fact it is can entrap importers at the over-quota tariff. If protecting the domestic import-competing industry is a priority, sending the signal that the quota is filled when in fact it is not will inhibit imports within the quota and yield the desired extra protection.

Another disadvantage of first-come, first-served allocation is that it can disrupt markets. For example, the U.S. peanut TRQ is administered on a first-come, first-served basis. The quota year starts April 1 and ends March 31 of the following year. Figure 6 plots the monthly distribution of quota fill for the U.S. peanut TRQ, which surges in April.[6] Three costs can result from these induced import surges: 1) an unnecessary dip in domestic prices; 2) unnecessary domestic storage costs; and 3) other, unnecessary rent-seeking costs induced by the existence of a common resource.

Distribution of Trade. To analyze FCFS allocation, some assumptions must be made about the relationship or correlation between a supplier's willingness to supply and its place in the FCFS queue. If a lower cost supplier always places ahead of every higher cost supplier, then, in figure 5, the effective quota supply curve is identical to the original supply curve (S) in the interval $(0, \lambda)$: S*. However, because all suppliers can sell on the international market, W becomes the lower bound on the willingness to supply the TRQ market for suppliers inframarginal to W. If each inframarginal supplier is equally likely to supply the TRQ market, it amounts to a random drawing from among the set of inframarginal suppliers (an operational test of nondiscrimination). Such a random drawing yields the effective quota supply curve: S^+. Such a selection process also achieves efficient discrimination between inframarginal and extramarginal suppliers and produces an expected distribution of trade equivalent to an auction allocation. As a practical matter, however, it is difficult

[6]Skully (1999) analyzes in depth the U.S. TRQs for peanuts, peanut paste, and peanut butter.

Figure 5
Quasi-market allocation

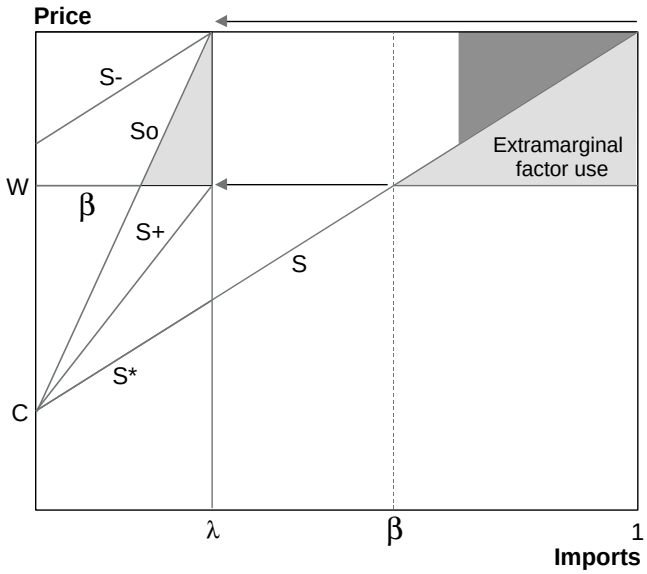
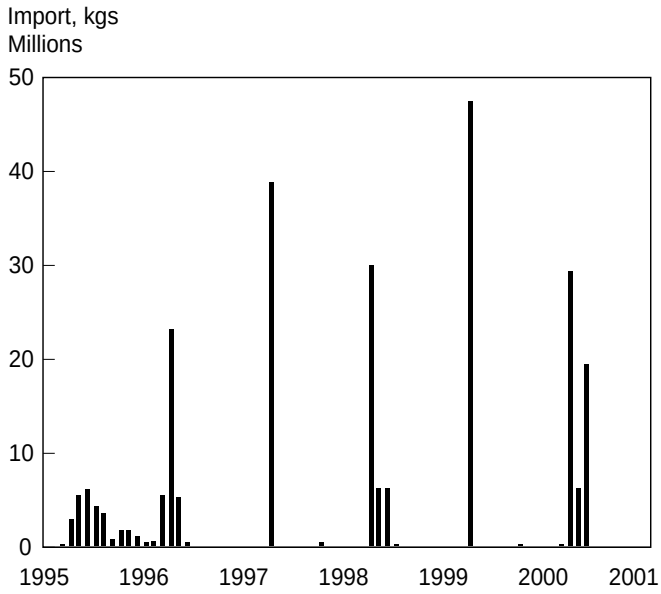


Figure 6
U.S. peanut imports under TRQ, volume by month



to conceive how such an efficient sorting of applicants could be achieved outside of using an auction.[7]

These two FCFS allocations (S* and S^+) assume that there is a perfect correlation between lower cost and place in the queue. This is a strong assumption, but it is not implausible to assume that there might be some imperfect correlation. License-on-demand allocation, discussed previously, is equivalent to FCFS allocation when there is zero correlation between cost and rank: a random selection of β units from a population uniformly distributed over the interval (C, P) has the expected distribution represented by the curve S^0. An FCFS process with a positive but imperfect correlation will generate an expected allocation of quota rights among suppliers which may be represented by a supply curve drawn within the area bounded by S^+ and S^0 over the range (0, λ).[8] In sum, if an FCFS process generates a positive yet imperfect correlation between low cost and place in the queue, then there is some expected inframarginal displacement. The lower the correlation, the greater the expected displacement.

Finally, consider the perverse case in which higher cost suppliers tend to queue before lower cost suppliers. This is not as implausible as it might seem. One could argue that rent-laden markets are the only markets where extramarginal producers can hope to cover their factor payments, so they may specialize in getting to the front of tariff quota queues. The necessary assumption is that loss aversion provides a greater incentive than profit maximization. When low cost is perfectly and inversely correlated with place in queue (i.e., ρ = -1), the expected outcome may be represented by the supply curve marked S-. This is the worst possible outcome, from a welfare and an Article XIII perspective. It results in the lowest realization of producers' surplus possible, given that the quota fills—all inframarginal suppliers are displaced by the "most" extramarginal suppliers. The triangle above S- is the producers' surplus realized by the most extramarginal suppliers. It is equivalent to the shaded triangle in the upper right corner of the graph. The area below S- in the range (0, λ) represents wasted resources and constitutes a welfare loss. Moreover, as established above, it indicates a biased distribution of trade relative to the counterfactual norm.

To summarize, this section surveys the quota fill and distribution of trade effects possible under auction, license-on-demand, and FCFS allocation. It shows that auction and license on demand, given the assumptions, are special cases of a general FCFS process. An auction is equivalent to FCFS with perfect correlation

[7]Vickrey (1961) proves that the only allocative mechanism that will guarantee this result is a uniform (second) price auction, now commonly known as a "Vickrey auction."

[8]The supply curve is a cumulative distribution of willingness to accept or to supply; it is a piecewise continuous, monotonically non-decreasing function. As the value of Pearson's ρ or Kendall's τ ranges from zero to one, the corresponding supply curve rotates clockwise from S^0 to S^+.

between low cost and place in queue. Such perfect sorting is not possible without recourse to an auction, however. License on demand is equivalent to a random FCFS process, with a zero correlation between cost and place in queue. The next section shows how historical allocation can also be characterized as a variant of FCFS allocation.

Historical Allocation

Historical allocation can be viewed as an extreme form of an FCFS process. FCFS and LD are annual lotteries. Each quota season there is a new drawing of applicants inframarginal to P. After several draws, the average realization will converge to its expected value. Historical allocation, in contrast, is essentially a one-time-only drawing. The historical base is infrequently revised: one particular realization is sustained for many years and remains invariant to changing market conditions. For example, exporter shares of the quota for U.S. sugar imports were first allocated in 1934 on the basis of trade volumes in 1931-33. Save for wartime controls, the allocation was essentially unchanged until 1948. Legislation in 1948 and 1956 made minor adjustments to the shares of the two major suppliers, Cuba and the Philippines. The trade embargo imposed on Cuba after the Cuban Revolution forced a re-assignment of the large Cuban share in 1961. It was formally reallocated in 1965 to countries other than the Philippines in proportion to their shares of the trade in 1963 and 1964. This allocation remained until 1974 when the 1948 quota was not renewed: imports were no longer limited by quota. A binding quota was re-imposed in 1982 on the basis of trade shares during 1975-81. This allocation was transferred unaltered into a tariff quota in 1995 and remains in effect. Each major change was prompted by an economic or political shock that, in each case, altered the structure of the sugar market. Despite this, the allocation of shares was based on the pattern of trade prevailing before the change. On average, there have been about 15 years between major reallocations. The U.S. sugar TRQ is considered below in a case study in which the extent of misallocation is calculated.

Article XIII and Historical Allocation

The discussion of Article XIII (beginning on p. 4) emphasized its advocacy of nondiscrimination. The analysis above has established that auctions, if practicable, are the best way to assure a non-discriminatory distribution of trade in a quota-constrained market. FCFS and license on demand are inferior to auctions, and will generally result in a biased distribution, and historical allocation amplifies the bias. Despite these predictable biases, all four methods are consistent with Article XIII. The sub-paragraphs on supplier quotas, XIII: 2(c) and 2(d), where the contradiction between advocacy of nondiscrimination and tolerance—if not advocacy—of discrimination is most clearly displayed.

GATT Article XIII: 2(c) states: "Except in the case of supplier tariff quotas import licenses shall not require that the imported product originate from a particular country or source."

Supplier tariff quotas, also known as "allocated tariff quotas," are tariff quotas that are allocated to supplying countries, rather than to domestic importers or traders. The particular country or exporting firm and country is specified by the assignment of the tariff quota rights. Article XIII: 2(c) essentially states that importing countries can employ TRQ rights as a GATT-consistent means of discrimination.

GATT Article XIII: 2(d) states:

> In cases in which a quota is allocated among supplying countries, the contracting party applying the restrictions may seek agreement with respect to the allocation of shares in the quota with all other contracting parties having a substantial interest in supplying the product concerned. In cases in which this method is not reasonably practicable, the contracting party concerned shall allot to contracting parties having a substantial interest in supplying the product shares based upon the proportions, supplied by such contracting parties during a previous *representative period*, of the total quantity or value of imports of the product, due account being taken of any *special factors* which may have affected or may be affecting the trade in the product. *(emphasis added)*

The two phrases italicized here have been the subject of further definition by the GATT in a series of interpretive notes to Article XIII. The convention has been to use an average of the 3 years prior to the imposition of a restriction as the representative period. Several disputes have arisen over base periods during which there were other restrictions on trade. The GATT recommends that shares be allotted according to the trade shares "which would correspond to what could reasonably have been expected in the absence of restrictions." Once again, this is the free trade counterfactual distribution of trade, the operational equivalent of nondiscrimination.

With regard to the meaning of special factors, Article XIII, GATT states:

> The term special factors as used in Article 22 [of the Havana Charter] includes among other factors the following changes, as between the various foreign producers, which may have occurred since the representative period:
> 1. Changes in relative productive efficiency;
> 2. The existence of new or additional ability to export; and
> 3. Reduced ability to export.

Thus XIII: 2(c) and 2(d) instruct member governments that they are allowed to transfer TRQ rights to incumbent exporters, but they should do so in such a way as to approximate the free trade counterfactual distribution of trade. This is not a simple task. The passage above elucidating the term "special factors" gives the impression that exporter shares can be reallocated in line with changing economic conditions, presumably without compensation. However, once exporters are vested with quota rights they tend to become upset when there is the least suggestion of taking them back or transferring them to a competitor. The author is unaware of a case where this kind of reallocation has occurred in accordance with Article XIII. The lack of such reallocations is hardly surprising. The primary reason an importing government chooses to allocate a "supplier quota" is to appease suppliers harmed by the imposition of a quota. The U.S. tobacco and sugar TRQs are examples. The quota rights are non-transferable; and the product delivered in-quota must be the domestic product of the exporter. Such restrictions are the cause of biased trade shares and often of quota underfill, as the U.S. tobacco TRQ example illustrates. The corollary to the last statement is that removing such restrictions can remedy, or at least substantially reduce, the risk of bias and underfill.[9]

Resale Markets in TRQ Rights

When TRQ rights are not transferable, one must export to realize the (compensatory) rent. Thus the distribution of rent and the distribution of trade are correlated. Allowing the transfer of TRQ rights liberates the distribution of trade from the distribution of rent. From the perspective of the WTO, the only relevant consideration is the distribution of trade, not the distribution of rent. This means, for the purpose of evaluating alternative methods of TRQ administration on the basis of nondiscrimination, one should ignore the redistribution of rent. Of course, it is the redistribution of rent that drives much of the politics of quota administration. As Gardner (1983) has demonstrated, the transfer efficiency of commodity programs (including schemes analogous to supplier quotas) is relatively low. It is far more efficient to compensate legitimate rent claimants by direct monetary transfers than through discrimination and manipulating markets.

The result above holds only if there is no resale of quota rights. Resale can reclaim much of the deadweight loss caused by extramarginal suppliers' displacing inframarginal suppliers. If there are no transactions or information costs, and all agents are rational profit-maximizers, then extramarginal quota holders will sell their quota rights to inframarginal suppliers discriminated against by quasi-market allocation processes. These trades, in a perfect market, occur at the price R, the marginal auction bid defined earlier. Extramarginal quota holders value quota at R or less, while inframarginal suppliers value the quota at R or more.

When resale is allowed, the final distribution of rent will differ from auction allocation only in that the TRQ sales revenue is captured by private traders rather than by the government. Suppliers inframarginal to W who receive quota rights keep them; suppliers extramarginal to W who gain quota rights in the primary allocation sell them. Essentially the auction revenue/rent is redistributed from the government or auction authority to private agents. Some countries, states, and provinces allow the resale of quotas and quota rights. In fact the government often organizes and supports the market exchange institutions.[10] Quota or rights trade has been allowed or devised to reduce air pollution and to prevent over-fishing.

U.S. Sugar TRQ: An Example of Historical Allocation

The U.S. sugar tariff quota is allocated to exporting countries on the basis of their "olympic average" market shares of U.S. sugar imports, 1975-81. This was a period of exceptionally high world sugar prices. So high, in fact, that in 1975, the United States removed the quantitative import restriction that had been in place since

[9]Skully (2000) evaluates the historical allocation of U.S. TRQs for peanuts, tobacco, and sugar.

[10]Alston (1986) examines egg quota trading in Australia. Dairy quotas are analyzed by Barichello (1996) and Chen and Meilke (1998) for Canada, and by Oskam (1991) and Pennings and Meulenberg (1998) for the European Union. Rucker, Thurman, and Sumner (1995) study the transfer of flue-cured tobacco quota in North Carolina.

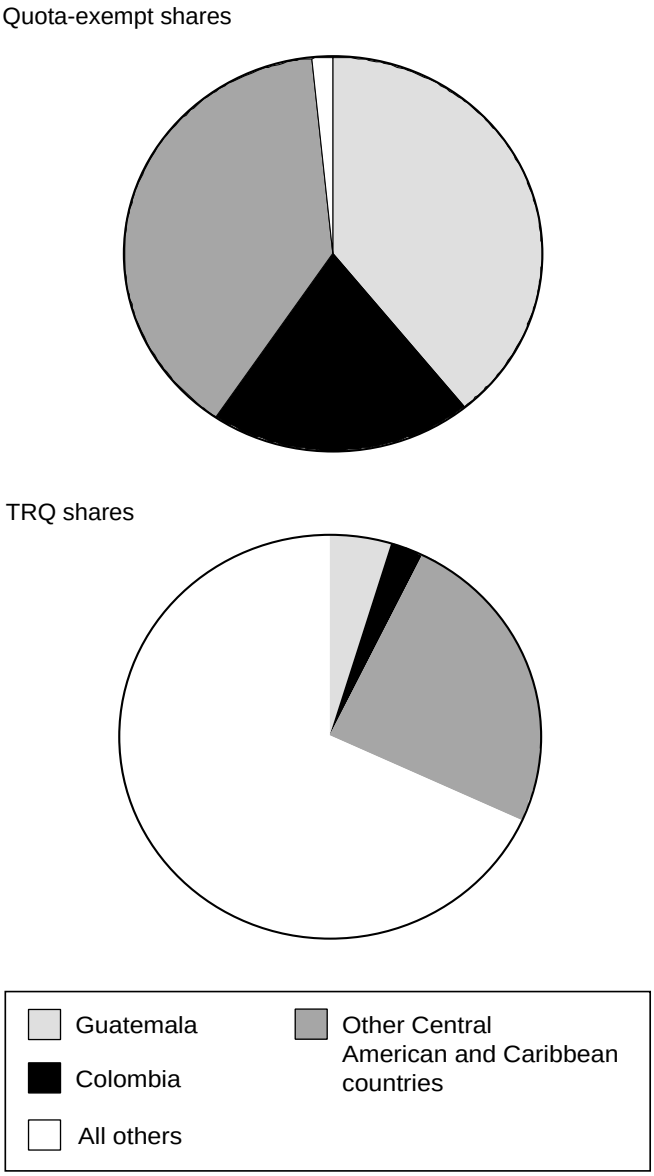
1934. During several months of the base period, the world price of sugar exceeded 30 cents per pound. At 30 cents, virtually everyone was an inframarginal sugar supplier. The market shares of U.S. imports during the base period 1975-81 included suppliers who were inframarginal for a few months, but were extramarginal under ordinary market conditions.[11]

Skully (1998) examines the pattern of imports for "quota-exempt re-export sugar." Raw sugar may be imported to the United States outside the quota if it is refined and re-exported within 90 days. This trade is not distorted by tariffs or quotas—save for the embargo on Cuba—and can be used as an estimate of the free trade counterfactual distribution of trade. This distribution is contrasted with the allocation of TRQ shares (fig. 7). Low-cost sugar producers located relatively close to U.S. refining centers in Gulf and Atlantic port cities dominate the quota-exempt distribution of trade.

The requirement that sugar imported under the TRQ must be produced in the country allocated the quota rights amounts to an anti-scalping law. This restriction induces costly transactions. Taiwan, for example, has tariff quota rights for the export of about 24,000 short tons of sugar to the United States.[12] Taiwan always fills its quota; however, this is the only sugar it exports. Taiwan's domestic production does not satisfy its domestic demand. It imports sugar (usually from Australia or Thailand) to cover the difference, including an additional 24,000 tons to cover the domestic production exported to the United States. It would be more efficient for Taiwanese quota holders to charter a shipment of 24,000 tons of sugar from Queensland or Guatemala to the United States and simply pocket the arbitrage rents. Similarly, the Philippines, the third largest quota holder (13 percent), has recently been unable to cover its domestic needs from domestic production. In fact, it has a TRQ to limit sugar imports. To procure domestic sugar to fill its U.S. tariff quota, the Philippine sugar authorities offered domestic mills 1.2 tons of imported raw sugar for every ton of domestic raw sugar delivered for export to the United States.

In both these examples, the ability to resell tariff quota rights would improve international factor allocation.

Figure 7
Distribution of trade for U.S. sugar imports



Because the revenue from resale or arbitrage could easily fund compensation (in the case of Taiwan) or fund investment (in the case of the Philippines), the resale would not have major domestic political repercussions. This is not the case for all current sugar TRQ holders. Several nations in the Caribbean region have both U.S. and EU preferences for sugar exports (e.g., Guyana, Barbados, St. Kitts-Nevis). Even with the quota rent income, sugar production is, at best, barely a viable economic activity. Because sugar production accounts for a large share of domestic employment, and sugar workers are well organized, the sale of quota rights would likely precipitate mass layoffs and cause political problems. In the absence of better alternatives, these governments

[11]Converting the absolute quota into a TRQ resolved a GATT dispute between Australia and the United States. Australia contended that the quota violated Article XI.

[12]A short ton is 2,000 pounds; a metric ton is 1,000 kilos or 2,204.62 pounds.

History of TRQ Governance in the GATT/WTO

GATT Article XIII governs the administration of quantitative restrictions (QRs). QRs were effectively prohibited in the Uruguay Round. Article XIII survives to govern TRQs which, while legally and technically not QRs, generally function as if they were QRs. Moreover, TRQs pose administrative problems identical to QRs. A brief history of the treatment of TRQs in international trade agreements follows. The narrative focuses on an inherent conflict between the GATT principle of nondiscrimination and the rationing problem posed by TRQs.

Imports were rarely restricted with absolute quotas or tariff quotas before World War I. The League of Nations sponsored a series of World Economic Conferences during the inter-war period, which attempted to reconcile how quantitative restrictions could be administered without discrimination, that is, consistent with Most Favored Nation (MFN) principles. By 1930, four positions were evident:[1]

- QRs are inconsistent per se with MFN.
- MFN requires that each country receive an equal share of the global quota.
- MFN can be approximated by allotting the global quota in proportion to the trade shares of current suppliers.
- Allow the global quota to be filled on a first-come, first-served basis.

The first position claims there is no just way to solve the quota-allocation problem. The second position argues for strict parity. If there are N parties to a trade agreement, then a fair allocation gives each party exactly 1/N of the global quota. The third position advocates proportionality, and the just basis for proportionality is the observed volume of trade in some recent representative period. Finally, the fourth position asserts (literal) priority in the form of first come, first served. As the brief history below shows, neither the League of Nations, nor the Havana Charter, nor the GATT resolved this issue. Instead of advocating one principle of distributive justice and proscribing all others, Article XIII allows for a conflicting set of distributive principles. Predictably, this conflict leads to disagreement about TRQ administration.

Starting in the early 1920's, political demands for agricultural protection and intervention emerged in Europe and North America. The United Kingdom and the Netherlands remained resolute free traders until the United States imposed the Smoot-Hawley Tariff Act in 1930 and protectionism cascaded. In 1932, the United Kingdom concluded the Ottawa Agreement that established a system of "imperial preference." The United Kingdom increased its MFN tariffs but granted a margin of preference to imports from its imperial dominions; the dominions, in turn, increased MFN tariffs but, also, they granted reciprocal preferences to the United Kingdom. The Smoot-Hawley Act raised tariffs on imports from all sources. By 1934, with the passage of the Reciprocal Trade Agreements Act (RTAA), the United States shifted to a trade strategy of "discriminatory liberalization." The RTAA generated a network of bilateral preferences between the United States and selected trading partners. By the outbreak of the World War II, Belgium, Britain, France, Germany, Italy, Japan, the Netherlands, the United States, among others, had developed systems of discriminatory trade preferences. The Allies attempted to maintain these systems after the war by incorporating them into the Post-War order.[2] The GATT, one of the three pillars of the Post-War International Economic order, required devices, for example, Article XIII to preserve these obvious violations of the principle of nondiscrimination.[3] Hudec (p. 178) notes:

> When governments decided, after World War II, that QRs would be permitted in many situations . . . [it] became necessary to say, whether it was true or not, that QRs could be applied in a manner consistent with the MFN concept. And so GATT Article XIII was written. Given its less-than-robust conceptual heritage, it is a small wonder that Article XIII proved to be a rather sickly child.

Article XIII is a "sickly child" because of a congenital deformity. It advocates both nondiscrimination and discrimination. The interpretation of Article XIII in this report emphasizes advocacy of the principle of nondiscrimination, the principle of distributive justice upon which GATT is constructed. The text takes an in-depth look at the subparagraphs of Article XIII that allow for discrimination.

[1] Hudec, 1997.

[2] See Skully, 1993 on competing visions of the governance of agricultural trade in the 1940's.

[3] The three pillars were to be the International Monetary Fund, IBRD, and International Trade Organization. The United States did not ratify the Havana Charter to establish the ITO. The GATT is the remnant of the ITO. Article 22 of the Havana Charter addresses the administration of QRs. Almost all of Article 22 was incorporated into GATT Article XIII.

might retain the quota rights to preserve the domestic status quo.[13]

Discretionary Methods

Discretionary methods of TRQ administration delegate the allocation process to a select group or organization. In the case of state trading, import rights within the TRQ (and sometimes outside the TRQ), are granted to a specialized government bureau. In the case of Producer Group administration, the import rights are granted to an organization that represents producer interests. How these organizations choose to exercise TRQ rights is limited only by the discretion granted them by their governments, or, in the case of producer groups, by their membership.

All TRQs administered by discretionary methods, 21 state trading TRQs and 9 producer group TRQs notified to the WTO for 1999, appear in table 2. A review of the list reveals that Thailand and South Korea account for the majority of state trading TRQs and all producer-group TRQs. Most are for products that are trivial from the perspective of global agricultural trade, e.g., pine nuts, raw ginseng, onion seed, garlic, capsicum, sesame, and potatoes. Eight state trading TRQs are potentially important—four are for rice: Indonesia, Japan, South Korea, and the Philippines. The Japanese TRQs for wheat, barley, and dairy are important simply because of the size of the Japanese market. Finally, there is the Thai tobacco TRQ. No producer-group TRQ is particularly important. While few TRQs are important from a global perspective, they may cause conflicts among interested suppliers.

Both state traders and producer groups tend to fill their quotas if there is sufficient domestic demand. So, quota-fill risk is generally not a problem under discretionary administration; the problem lies in the distribution of trade. Because public or quasi-public officials, not private traders, make import-sourcing decisions, many factors divorced from commercial considerations may determine market shares. That is, political considerations play a major role. It would not be fruitful to generalize further, because, unlike quasi-market methods, which can be reduced plausibly to algorithms, each discretionary institution has it own methods, or, perhaps more accurately, non-methods. The best one can do, at least at this initial stage of TRQ analysis, is to examine each individually. Thus, the balance of this section consists of a case study of how the Japanese Food Agency, a state trading organization, administers the Japanese TRQ for wheat. This case study, which is not a representative sample, was selected because of available data and secondary information, as well as its importance in U.S. agricultural trade diplomacy.

Wheat Imports by the Japanese Food Agency: Discretionary Allocation

Japanese grain and oilseed imports are not uniformly regulated. Corn, soybeans, and most other oilseeds are imported with relatively few restrictions; rice, until the Uruguay Round, was simply not imported; and wheat imports were, and continue to be, controlled by the Japanese Food Agency (JFA). The JFA has been notified to the WTO as a State Trading Enterprise (STE).[14]

The JFA is one of the world's largest importers of wheat. Consequently, it has been the subject of considerable study. Unfortunately, much of the economic analysis of the JFA's import procurement decisions has been predicated on erroneous "as if" assumptions about the JFA's objective function. This section draws on one study that presents a persuasive argument about the JFA's objectives. Alston, Carter, and Jarvis (1990) argue that the JFA operates a discriminatory quota scheme for wheat, and that the Japanese Livestock Industry Promotion Corporation (LIPC) operates a similar regime for beef imports. Each year the JFA, in consultation with the domestic milling industry and the Ministry of International Trade and Industry (MITI), determines domestic consumption targets for wheat and wheat products, estimates expected domestic wheat production, and determines aggregate import needs. The aggregate quota is procured by subcontracting to a pool of about 30 trading firms. The contracts specify the variety of wheat required as well as a designation: food wheat or feed wheat. As Alston et al. show, the food-feed distinction is not particularly meaningful. Rather, it allows an additional degree of freedom in determining supplier market shares. At the time Alston et al. drafted their article, only Australia and the United States were permitted to supply feed wheat. Canada was excluded as a feed-wheat source.

The essence of the Alston analysis is that the JFA's allocation of supplier shares for wheat is analogous to a set of Voluntary Export Restraints (VER). A VER

[13]Trela and Whalley (1995) calculate the welfare loss caused by the historical allocation of quotas by exporting countries under the Multi-fiber Agreement.

[14]Ackerman and Dixit (1999) analyze state trading in agriculture in a WTO context.

Table 2—TRQs allocated by discretionary methods

Country/commodity	1995	1996	1997	1998
		Percent		
TRQs allocated by state trading				
Canada				
Butter and dairy spreads	100	100	100	100
Indonesia				
Rice	100	100	100	100
Japan				
Designated dairy products for general use	100	100	100	100
Wheat, meslin, triticale & processed products	100	100	100	100
Barley & processedproducts	100	100	100	100
Rice & worked/prepared products	100	100	100	100
South Korea				
Potatoes	0	6	37	39
Onions	100	100	97	100
Garlic	77	82	100	100
Fruits of the genus capsicum	99	97	76	99
Beans	100	100	100	100
Ginger	96	36	100	90
Rice in the husk	100	100	100	100
Buckwheat	100	99	100	100
Groundnuts	100	99	95	98
Sesamum seeds	100	100	100	100
Philippines				
Rice	100	100	100	100
Thailand				
Garlic	0	0	0	0
Unmanufactured tobacco; tobacco leaves	100	100	100	100
Palm oil	100	100	100	100
Cane or beet sugar and sucrose	0	0	0	0
TRQs allocated by producer group				
South Korea				
Raw ginseng	100	1	100	43
Pine-nuts	100	100	100	100
Korean citrus	99	98	100	100
Thailand				
Onions	0	99	98	100
Onion seeds	100	100	100	100
Soya bean oil	92	27	0	0
Maize	100	100	100	100
Soya beans	100	100	100	100
Soya bean cake	100	100	100	100

shares quota rents with the exporting country that could be fully captured by the importing country. By the standard of domestic welfare maximization (the assumption employed in many studies of the JFA), a VER is clearly inferior to a global quota or its tariff equivalent. If an importing country chooses a VER, it reveals that the welfare of the preferred supplier(s) is important to the importing country's government.

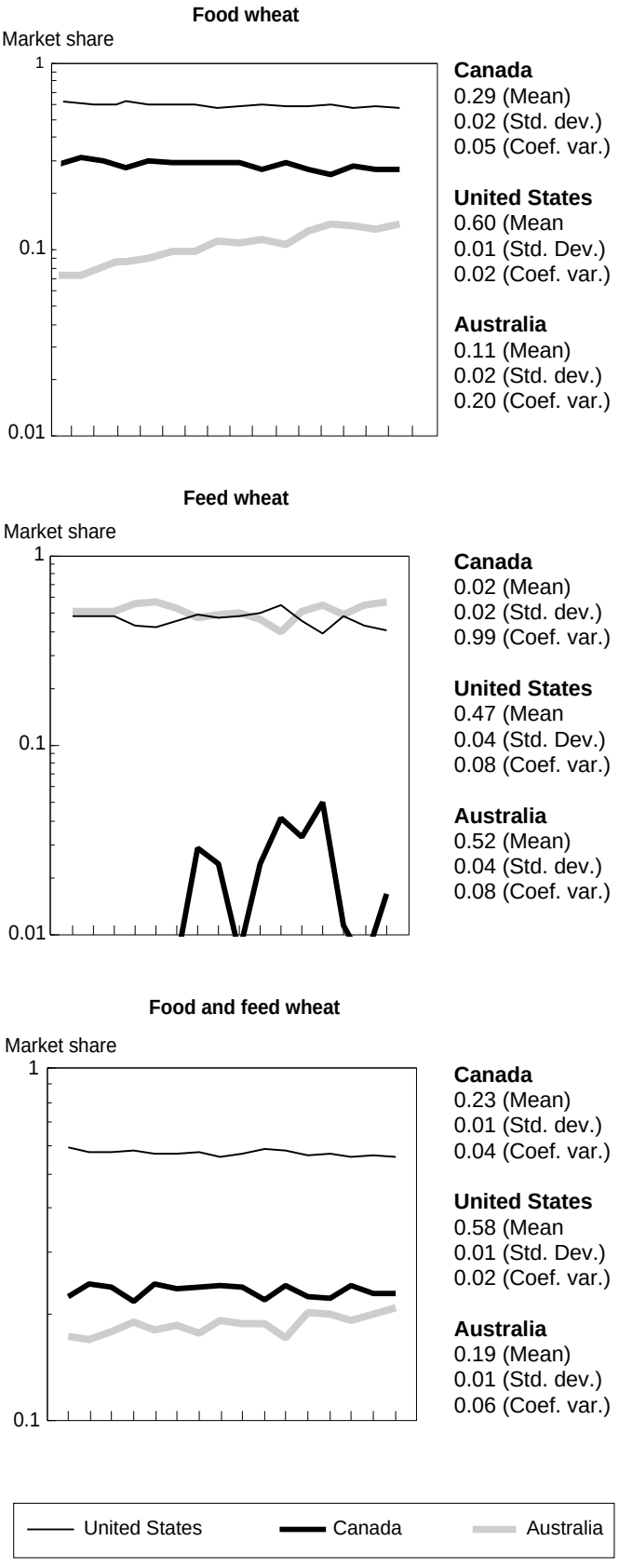
> [T]he government in the importing country attaches greater value to the appeasement of the foreign interests than it does to the efficiency costs of not using an otherwise economically superior instrument. Thus, we can explain the Japanese government's use of managed import quotas for beef and wheat, but to do so we must allow for the political influence of import suppliers. (Alston et al., p. 200)

Alston et al. based their conclusion on the observed imports of the JFA and the LIPC through 1989 (when they drafted the article), as well as supplementary evidence from interviews with principals in the trade. Supplier import shares are too stable, a telling piece of evidence. A market-driven allocation would show more variance than what is evident. As evidence of a JFA bias for U.S. wheat and against Canadian wheat, they note that Japanese milling firms consider Number One Canadian Western Red Spring wheat (CWRS) the most preferred variety. Mills always request more Number One CWRS in their annual consultation with the JFA, but the share procured remains too low in their view. On the basis of this evidence, Alston et al. concluded that a liberalization of the Japanese wheat market, for example, converting the JFA's discretion over imports into a tariff equivalent or auctioned TRQ, would result in a transfer of market share from the United States to Canada. Moreover, they argue that per capita Japanese wheat consumption is high and tariff reductions are unlikely to increase the volume of imports significantly. Consequently, liberalization might lead to an absolute decline in the volume of U.S. wheat exports to Japan. That is, the substitution effect would dominate the expansion effect (Alston et al., p. 210).

After a decade of additional trade data, we see if the Alston, Carter, and Jarvis thesis holds. Figure 8 shows the market shares of the three suppliers of wheat to the JFA, the only suppliers. Shares are shown for food, feed, and total (combined) wheat. Basic descriptive statistics are included in figure 8. The shares are plotted as values between zero and one using a logarithmic

Figure 8

Japanese Food Agency, wheat imports, 1982-97



scale. The annual shares are stable. The only change is a minor increase in the Australian share of the food-wheat market. The increase is gradual and steady, indicated by the constant slope of the log-share line. The Australian expansion in the food-wheat market is mirrored by the allowance of Canadian "feed-wheat" imports—a minor innovation since Alston et al. (1990).[15]

Arnade and Gehlhar (2001) analyzed monthly wheat import data of 43 countries, which account for over 90 percent of world wheat trade, for evidence of importers' market power. In their discussion of importers' sourcing behavior during 1962-1995, they noted:

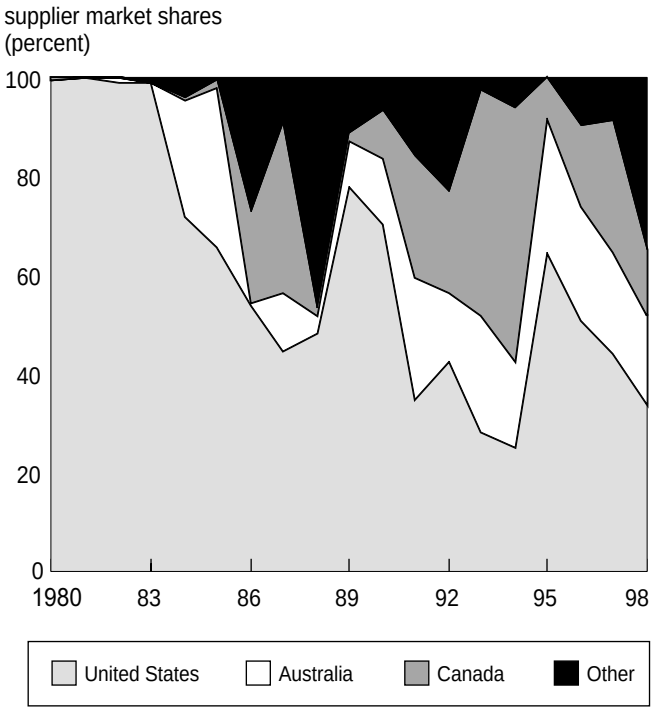
> Most importers have shifted suppliers somewhat. Japan however is a clear exception; its shares have been stable. It sources nearly all of its wheat from the United States, Canada, and Australia in the fixed proportions of 55%, 27% and 18% respectively.

Arnade and Gehlhar also examined the monthly trade data for the frequency of multiple suppliers. They argue that a high frequency of single-source observations may represent a series of corner solutions by very price-sensitive importers. These countries buy only from the least cost supplier. Conversely a high frequency of multiple suppliers may reflect an inelastic demand for a particular supplier's product characteristics, or, possibly a concern for a diversified supply. A high frequency of multiple suppliers is also consistent with the VER hypothesis of Alston et al. Arnade and Gehlhar found that Japan has imported from its three regular suppliers 95 percent of the time; in only 4 percent of the months observed did it procure from only two of the three suppliers, and in less than 1 percent of the time from only one supplier. The persistence and stability of its pattern of imports is consistent with the VER hypothesis.

---

[15]JFA prefers several Australian wheats for noodle production, such as Australian Standard White (ASW). For example, the Eradu and Gamenya varieties from western Australia are considered ideal for udon noodle production. The formation of the Western Australian Noodle Wheat Growers' Association in 1992, and the Australian Wheat Board's policy of testing and segregating noodle varieties, initiated in the 1992-93 season, have further enhanced Australian noodle wheat's appeal in Asian markets. This may account for the JFA's administrative increase in the Australian share of the food-wheat market. For development of varietal noodle wheats, see Crosbie (1994) and Lin and Vocke (1998).

Figure 9
South Korean wheat imports, supplier market shares



Source: Korean Statistical Yearbook of Foreign Trade.

South Korea is the only country with a pattern of wheat imports similar to Japan's. The reason for this similarity is that, until 1983, the Korean Flour Mills Industrial Association (KOFMIA), a state-coordinated umbrella organization, held monopoly control over wheat imports.

> The U.S. share of the Korean milling-wheat import market was virtually 100 percent until 1983, when private importers were permitted to enter the market. . . . U.S. milling wheat has been displaced in the Korean market primarily by Australian wheat, including Australian standard white (ASW), Australian Hard (AH) and Australian Soft (AS). Canadian Western Red Spring (CWRS) 13.5-percent protein wheat has begun to make inroads and will probably capture a larger share of the market. (Raney and Morgan, 1994, p. 8)

Figure 9 plots the market shares of Korea's major suppliers since 1980. The U.S. share since 1983, when KOFMIA was liberalized, has averaged 50 percent. The share is volatile, as expected in a competitive market, and shows a secular decline. The Korean case cannot be taken as a perfect counterfactual for the liberalization of the JFA, but a decline in the U.S. share is likely, and an increase in its volatility is certain.

The JFA fills its import quota, but this is not surprising. In practice, state trading poses a low risk of underfill. Competing suppliers carefully scrutinize TRQs that implement Minimum Access commitments. If there is market demand, quotas fill.[16] The greatest risk is that the discretion over sourcing, in the hands of public employees, is particularly vulnerable to political pressure. If such pressure alters the distribution of trade from its tariff-equivalent counterfactual distribution, that is, based on supplier efficiency alone, it violates the principle of nondiscrimination.

[16]If not, political pressure is applied, bilaterally or through the WTO. The recently resolved United States-Philippine dispute over the Philippines' administration of its TRQ for pork follows this pattern. The quota eventually fills.