

Quantitative restrictions on trade (QRs) have been the main means of increasing protection in the world economy in the last 25 years. Textiles, steel, and autos are well-known examples. Taken with the concurrent reduction in tariffs, it is only slightly inaccurate to treat protection among developed countries as primarily by QRs. Any protection is usually inefficient.¹ A major and unappreciated fact is that the QR method of protection is itself a source of further inefficiency.

The reason is intuitively pleasing to an economist. QRs limit the operation of markets more than tariffs, reducing the efficiency of the price system. There are two aspects: the reduction in *arbitrage efficiency* and the reduction in *competitive efficiency*. Under product heterogeneity, the allocation of quota licenses ordinarily does not achieve a market solution (equal rent in all uses). Thus it destroys the *arbitrage efficiency* of the tariff. The second inefficiency is that under imperfect competition the response of the foreign firms is more limited than under a tariff, which reduces the demand elasticity facing domestic firms and enhances their monopoly power. This destroys the *competitive efficiency* of the tariff. The principal message of this book is that these aspects are ubiquitous, especially the first, and that they are quantitatively important.

The efficiency of the price system can trade off against other values, on the other hand, so there are motives for protection that can reverse the ranking under some assumptions. A second message of this book is that their combined structure is rather special. For practical purposes, protection in developed countries would be more cheaply done with tariffs. Even in special cases where a pure quota might dominate a pure tariff, feasible tariff schedules that vary the tax rate with some variable such as import volume or foreign price will generally dominate quotas.

The conventional wisdom is that tariffs and quotas as a practical matter can usually be treated as equivalent, and that where this rule fails, nothing

very general can be said to rank the two. The starting place for the usual analysis is a simple model in which tariffs and quantitative restrictions are equivalent. All undergraduate international trade texts start with this case. While pedagogically convenient, in that it shows that the same incidence analysis developed for tariffs can immediately be used for QRs, the equivalence analysis is fundamentally misleading in suggesting the *practical* equivalence of the two. All the texts with which I am familiar follow the equivalence analysis with only slight attention to giving circumstances in which the two are not equivalent. Sometimes a presumption for tariffs is vaguely sketched (a tariff is more flexible). The undergraduate reader is inevitably left with the impression that nonequivalence is a minor and esoteric matter. This book shows that the nonequivalence is ubiquitous, and empirically important. The equivalence model also dominates the evaluation of protection in practice. For example, the equivalence structure is built into the standard template for case studies of the effects of protection done for the U.S. government. Alternative and operational methods are provided below that demonstrate the dangerous downward bias of standard estimates of the cost of QR protection. On the other hand, for the technically trained, there is a vast technical literature on the nonequivalence of tariffs and quotas. The point of these exercises is often to display sufficient conditions for the ranking to go either way. The impression created is that the ranking depends on special features of models, with no general principles available. This book shows that two important themes dominate the analysis, working always in favor of tariffs.

This chapter develops the common structure of essays written by the author alone and with Leslie Young over the last eight years that establish a presumption against quotas. Section 1.1 reviews the equivalence analysis. Section 1.2 considers the pure case of quota inefficiency. It shows that QRs are substantially inferior to tariffs in the presence of product heterogeneity or imperfect competition. The simplest comparison occurs when the motive is to hit a target level of import aggregate quantity or value. Such constraints most naturally arise due to external factors, such as a desire to reduce dependence on external oil. They also arise due to crude responses to domestic factors, such as a desire to protect employment or to raise a given amount of revenue. In an important special case, the crude response is optimal. Section 1.3 goes behind import control to review the most common motives for protection in more general models. In principle, these could qualify the ranking of section 1.2. Instead, save for special cases, quotas are an inefficient means of achieving protectionist targets in models of realistic complexity.

Section 1.4 concludes with some observations on QRs and political economy. The appendix develops the methodology used in the book in terms of a template model, and indicates how the special cases of the various chapters adapt it. The methods used are likely to have other applications in the analysis of policy, so the reader is encouraged to digest it.

1.1 Tariff and Quota Equivalence

In a competitive model with a single homogeneous product, any tariff has an equivalent quota and vice versa. The only difference between the two (absent administrative costs) lies in who receives revenue from the trade control instrument. In the tariff case, it is the government. In the quota case, it is the holders of licenses to sell in the controlled market. If the government holds a competitive auction to sell the licenses, then the two instruments are identical, (again, absent administrative cost differences).

To see these points, consider the market for clothing imports. P^* is the foreign relative price of clothing, and P is the domestic relative price. Consider a specific tariff t (a tax of t dollars per unit of imports) and a quota Q^0 that yields the same amount of imports. In figure 1.1, S is the foreign export supply curve and D is the home import demand curve. A tariff t shifts up the supply curve facing consumers to S' , so that S' and S have the same slope at each quantity. The home country will have imports selling at P , the foreign country will receive P^* , and quantity consumed is Q . A quota in the amount $Q^0 = Q$ (i.e., Q^0 worth of import licenses are issued) will

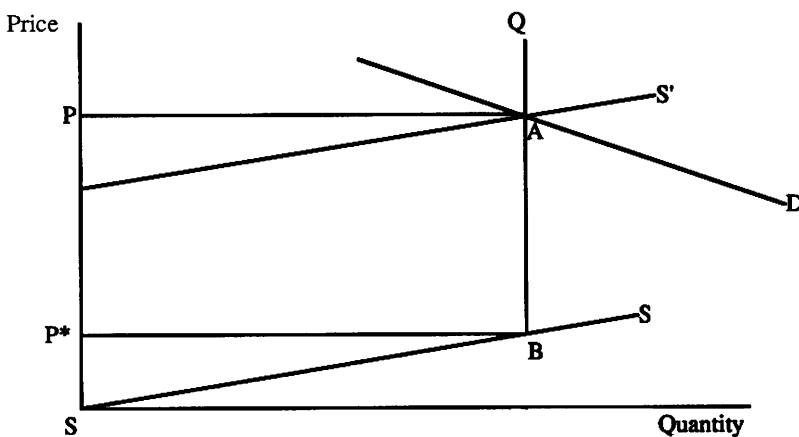


Figure 1.1

make the supply curve facing consumers be the foreign supply curve up to Q and then be vertical beyond it: SBQ . Evidently the price and quantity consequences for importing and exporting countries are identical for the tariff and quota. The revenue associated with either instrument is $PABP^*$, but in the case of the tariff it goes automatically to the government. In the case of the quota, it goes to the holders of the licenses, unless they have paid for the licenses.

One significant complication with QRs is that they are frequently forced on the exporting country. The importer wants to limit trade but requires the exporter to control it. In this case the quota is a voluntary export restraint (VER). The consequences are as in figure 1.1, save that foreign owners of export licenses get the revenue. VERs are especially bad from the viewpoint of domestic national net welfare. In a world welfare analysis, of course, it makes no difference who gets the revenue.

1.2 Tariffs Are Preferred to Quotas: The Pure Case

The simple equivalence model of section 1.1 implies that economists should be indifferent to the choice of an instrument. In fact, in more complex and realistic models there are compelling reasons for believing that tariffs are considerably more efficient instruments for attaining the protectionists' goals. Two important complications are imperfect competition and product heterogeneity. In either case the vague intuition that tariffs are better because more flexible is confirmed. QRs further confine the responsiveness of economic agents in these cases, and this results in reductions in the efficiency of the price system via losses in *arbitrage efficiency* and *competitive efficiency*.

1.2.1 Imperfect Competition

Imperfect competition has long been known to imply tariff dominance (Bhagwati, 1965). Quotas seal off domestic monopolists from the price discipline of the international market, permitting monopolization of the home market segment that cannot be filled from imports. Tariffs, in contrast, leave the potential monopolist as a price taker. The monopolist's well-known profit-maximizing markup formula is $(P - C_Y)/P = YP_Y/P$, where P is the price of the monopolist's output, Y is the quantity of output, C_Y is its marginal cost, and YP_Y/P is the inverse elasticity of demand facing the monopolist. Under a tariff, where the foreign good is a perfect substitute for the home good, the inverse elasticity is zero. Under a quota it is

based on the elasticity of the residual, net of imports, demand curve. Output restriction becomes profitable since domestic consumers cannot escape to imports at the margin, and the left-hand side of the equation must rise in response.

Chapter 2 considers the more realistic case where the home and foreign firms are duopolists. Essentially, the Bhagwati insight is shown to carry over to the home duopolist's rational "perceived" demand curve. As compared to tariffs, quotas raise the relevant "perceived" inverse demand elasticities under reasonably general conditions. Equivalence arises if the type of trade instrument has no effect on the nature of competition between firms. If instead, more reasonably, the firms' perceptions of their rivals' reactions are formed rationally,² quotas imply less competitive home firm behavior than tariffs that achieve the same quantity or import price. Interestingly, this conclusion holds for either price or quantity-setting duopolists. The quota is then shown to be inferior. This is the *competitive inefficiency* of quotas.

Chapter 2 also considers the case where the home firm is able to see through the government's trade policy to its strategic dependence on the firm's output decision. For example, the government may be using trade policy to hit an employment target. The output game then has a collusive solution, and tariffs are equivalent to quotas. More complex models of this type with asymmetric information are a fruitful area for future research.

A related form of nonequivalence arises when *nations* exercise monopoly power with strategic trade policy, as in the optimal tariff and retaliation literature (Rodriguez, 1974). Again tariffs lead to superior outcomes. Recent game-theoretical literature has been deepening insights in this area, but this chapter will not review it. Save for the singular case of OPEC, such behavior is rare, so it has little to do with the pragmatic analysis of protection.

1.2.2 Heterogeneity

The most basic and ubiquitous reason for nonequivalence is product heterogeneity. In practice, protection is granted for a product class: autos, cheese, oil, etc. An aggregate number of units, e.g., 1.85 million Japanese cars, is authorized. The distribution of the licenses across members of the controlled group (e.g., Toyotas, Subarus, Isuzus, Mitsubishis, Nissans) is left to an administrator. Instead of using the market system to price the licenses, the administrator allocates according to some simple rule, like base year market shares (e.g., if Subaru had 5% of the market in 1979 before the quota system, they receive 5% of the allocation of 1.85 million, or 92,500

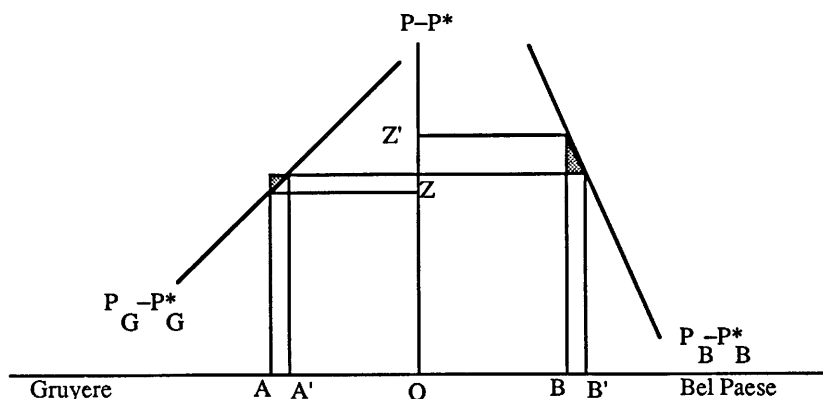


Figure 1.2

units). In general, this system will not imply rent earned on a unit license being the same across products, so long as resale is frustrated (which it usually is) or is inefficient (apparently the case for Hong Kong textile export licenses under the VER system).

Elementary economic theory teaches that exchange is beneficial, and this insight applies to quota licenses too. What would the outcome of a competitive market process for licenses be? Evidently, the rent earned on the use of a license would be the same for all members of a quota constrained group. If not, profitable arbitrage opportunities exist impelling reallocation. Figure 1.2 illustrates this. The aggregate amount of cheese (measured in pounds) across the two categories Gruyere and Bel Paese (G and B), is AB . The allocation of licenses to the two categories depends on a bureaucrat's discretion. In the case shown, $1/2$ of the licenses are given to each. The value of a license to the holder is $P_i - P_i^*$, where i is either G or B . Evidently the allocation leaves a higher premium on B licenses. Transfer of a license from G to B , given the initial allocation, earns additional rent $= (P_B - P_B^*) - (P_G - P_G^*) = ZZ'$. This is also a social benefit under the representative consumer model, since P_i is the marginal social benefit of another unit of imports of category i (measured by consumers' willingness to pay), and P_i^* is the marginal social cost (if a world welfare viewpoint is adopted, or if the country is small; S is horizontal in figure 1.1). A competitive equilibrium in the allocation of licenses will equate the rent, solving for the quantities A' and B' that yield a uniform $P - P^*$ in figure 1.2. The standard gains from trade analysis reveals that the competitive equilibrium yields higher social surplus by the amount of the shaded triangles.

Note that equal rent across uses is equivalent to a specific tax uniform across the product class, save for distribution of the revenue. The shaded triangles are the *arbitrage efficiency* gain due to the tariff. Relative to any allocation, the tariff secures such triangles: this is the arbitrage efficiency property of the tariff. This implies a basic result: given an aggregate constraint on import quantity, a uniform specific tax satisfying the constraint is more efficient than any quota allocation system (save the unique allocation equivalent to the specific tax allocation of imports). Quota inefficiency arises because it prohibits beneficial exchange.

Chapter 3 is an empirical study of the U.S. cheese quota system (from Anderson, 1985); with essentially this structure for 9 categories of imported cheese, I showed that the added inefficiency of the quota allocation amounted to over 15% of the base expenditure on the controlled categories. That is, switching to a specific tax designed to yield the same aggregate cheese imports (announced as a goal by the chief administrator) saved 15% of base expenditure. In an alternative scale, the inefficient quota allocation increased by 30% the unavoidable loss due to the constraint on aggregate imports. Moreover, aggregation bias causes these to be substantial underestimates of the true relative inefficiency. This is a very substantial loss relative to the usual magnitudes found in studies of the cost of protection.

The potential arbitrage efficiency loss of QRs is also ubiquitous, since this is typically the situation of quota-ridden products. No matter how finely the product specification is divided, for manufactured products there are usually remaining elements of heterogeneity. An administrator is then left with the discretion to solve an allocation problem. The prominent industries receiving quota protection are highly heterogeneous (steel, textiles, autos). Even where the product is homogeneous, the allocation issue arises across countries of origin. U.S. sugar import quotas are notoriously inefficiently allocated, for example. VERs are inherently subject to the same allocative inefficiency, since the quota is negotiated with the original market incumbents, with new entrants being sealed off after a small incursion, and initial allocations receiving property right status.

A closely related form of product heterogeneity occurs when the future is not certain. Let a possible outcome of that which is not known (like the annual rainfall next year) be a state of nature. Then in essence, one pound of imported sugar is a different commodity in each state of nature. Let G be a good year for domestic sugar production and B be a bad year.

Note that a fixed quota Q is a given allocation of imports for a fixed time interval (usually a year) that is the same across states of nature

(assuming for simplicity that the quota always binds). Is this efficient? Suppose that the constraint on policy is to achieve an average import volume. Since imports are not in principle required to be Q in each state of nature, one can conceive of arbitrage of state-contingent licenses occurring across the product group, subject to the import average constraint.

This can be analyzed as in figure 1.2. Let $Q = OA = OB$. Imports are worth more in the bad year B . Suppose that good and bad years occur with probability $1/2$. The permissible average level of imports is then $AB/2$, and any allocation of state-contingent licenses that preserves AB is feasible. Competitive arbitrage equilibrium occurs where rent on use of a license is equal in all states; this is equivalent to a uniform-over-states specific tariff. It is found by sliding the line segment AB along the horizontal axis until the perpendiculars at the end points A and B intersect the inverse demand curves at the same height. The resulting allocation is OA' and OB' . This is the most efficient allocation due to the same gains from exchange argument as above for cheese. A more formal treatment is in chapter 4, based on Young and Anderson (1980).

Such an auction of contingent licenses is infeasible in practice, due to the cost of enumerating a very large number of states and monitoring the volume of exchange in them to impose the average constraint. *Therefore, no quota system can ever achieve in practice the arbitrage efficiency of a tariff.* This is a significant point, since the government is now considering auctions of licenses in future protection cases (e.g., this was in fact the ITC recommendation to President Reagan in the shoe case in 1985). An auction *would* resolve the inefficiency created in the previous type of heterogeneity. It would also provide information on the restrictiveness of the quota, which might reduce the level of protection under the analysis of the political economy model. While desirable, auctions can nevertheless not restore the efficiency of the tariff.

The shifts in the allocation of imports shown in figure 1.2 may of course have negative impact on some deeper government policy goal than aggregate or average imports AB . For example, aggregate or average employment, output, wages, or profits in an import-competing sector may be targeted. In less-developed countries, revenue from trade distortions may be a target. Generally these cases are treated in section 1.3. In the important special case of weak separability, analyzed in chapter 5, the deeper target is affected by the aggregate *value* of imports alone, not by the detailed allocation within-group. Allocations of import value quotas across the heterogeneous types (foreign exchange licenses) can be analyzed as in figure 1.2, dividing the vertical variables by P^* and multiplying the hori-

zontal variables by P^* . The new constraint is in terms of aggregate foreign exchange value of imports, a constant length on the horizontal axis for the transformed variables $P_i^* Q_i$. Under this constraint, a uniform *ad valorem* tariff achieves the competitive equilibrium in arbitrage of foreign exchange licenses. Once again, the appropriate type of tariff has the *arbitrage efficiency* property relative to any quota allocation system. Chapter 4 applies the same analysis where the element of heterogeneity is the state of nature and the constraint is on expected foreign exchange value of imports.

To gain a simple understanding of the argument of chapter 5, suppose for simplicity that an auction of licenses under certainty *did* make a quantity quota equivalent to a specific tariff. The *ad valorem* rate of duty on high priced imports is less than on low priced imports. The substitution effect will lead to *quality upgrading* (Baldwin, 1982; Falvey, 1979). This could imply that in order to meet, for example, an employment constraint, the quota must be tightened, possibly implying that the quantity quota would be inefficient relative to a system that did not induce quality upgrading. Chapter 5 shows that for a protectionist target rate of unemployment (which stands in for a wide variety of domestic economy targets), the optimal trade instrument is a uniform *ad valorem* tariff under production or preference structures weakly separable with respect to the partition between imports and domestic goods. The reason is that under this structure, the choice of goods to demand for input use or consumption proceeds according to a decision tree in which the choice between aggregate expenditure on home and foreign goods is made at the upper level, with allocation among imported goods given the aggregate import expenditure decided at a lower level. Then what matters for the link of imports with domestic goods or factors demand is only the aggregate expenditure on imports. The *arbitrage efficiency* of the uniform *ad valorem* tariff can be secured. Under more general production or preference structures, the circumstances that lead to dominance of the specific tax over the *ad valorem* are shown to be rather special.

Finally, product heterogeneity is significant along the time dimension. Tariffs and quotas generally have different implications for the temporal structure of domestic prices of imports. First, most simply, note that "non-economic" constraints on the (discounted) sum of the quantity or value of imports over time act like the average constraints in the case of uncertainty and all the same analysis applies. An *ad valorem* tariff leaves the intertemporal import domestic price ratio equal to the foreign; a specific tariff leaves the intertemporal marginal net benefit of imports constant over time. Both are efficient relative to a quota that freezes the value or quantity

imported at each point in time, causing intertemporal fluctuations in the license premium.

But more deeply, a quota is a permit to import over an interval of time. Quota licenses (foreign exchange licenses) cannot generally be priced so that they convert into specific (*ad valorem*) tariffs in the presence of uncertainty, since they are effectively *options*. The quota will be restrictive; licenses have positive value, even for years when the limit on imports is not attained. Presumptively, the added distortion of intertemporal relative prices is inefficient if the “noneconomic” constraint is not an absolute annual limit. Chapter 10 sets out a model of option pricing of quota licenses. The use of quota licenses is usually encumbered with restrictions, and the chapter shows that the usual administrative procedures for control of license use are inefficient.

Both the *competitive inefficiency* and the *arbitrage inefficiency* are effectively overcome by the virtues of the invisible hand of the market mechanism. The relative inefficiency of quotas thus rests on the same base as other invisible hand propositions, and is subject to the same qualifications. It could be that quotas outperform tariffs in ameliorating some other distortion, for example. Such claims must be checked on a case-by-case basis. It turns out that typically even when the claim is true, the arbitrage efficiency of the tariff is the dominant consideration.

The next section reviews the principal forms of other distortions, some of which could lead to quota dominance in principle. In practice, tariffs dominate.

1.3 Impure Cases: The Practical Dominance of Tariffs

The principal reasons for protection in developed countries are (1) to raise sectoral-specific employment or wages, or sometimes the associated output or profits, and (2) to limit imports of agricultural price-supported commodities. In less-developed countries, (1) and (2) may be operative, and a host of other second-best claims are made, in which a trade instrument is devoted to fixing a nontrade problem such as a capital market distortion. What lies ultimately behind the protectionist target may be either efficiency or equity motives. Where these might matter to instrument choice, a further consideration is developed.

This section reviews some recent work done on comparing tariffs and quotas under the main forms of motives for protection in the presence of heterogeneity. Employment and agricultural price protection are considered as the most practically relevant cases for developed countries.

There is almost no end to the number of possible second-best cases relevant to less-developed countries, but two interesting ones are revenue constraints and capital market failure. These are reviewed below.

1.3.1 Employment Protection

The most important distortion for explaining protection in industrial countries is in the labor market. Sometimes this is expressed in a concern for employment, sometimes for wages, and sometimes for both. Employment concerns can arise naturally in a representative consumer model in the presence of labor market distortions. Somewhat more awkwardly, distributive justice concerns can explain a desire to protect sectoral employment, wages, or both. Below it is shown that the exact form of the constraint is actually unimportant for the purpose of ranking tariffs versus quotas; a variety of sectoral targets or embedded distortions have the same implications for the ranking of tariffs versus quotas.

In chapter 6, a study of employment protection in the U.S. textile industry (SIC 22), I show that the arbitrage efficiency of tariffs is the dominant consideration, even in a model that does not reduce to achieving the goal by a simple restriction of imports, such as average value or quantity. The optimal (but infeasible) allocation of imports over states is revealing. It requires equality of the marginal net benefit of another unit of imports, $P - P^*$, with the marginal cost to employment, which is proportional to l_Q , the employment displacement effect of another unit of imports. For constant l_Q over states, this implies the optimality of a fixed specific tariff. The closer to linearity the employment function is, the better the tariff will do relative to the quota. Even for nonlinear models the structure that permits quota dominance is very special.

A simulation of a loglinear model of the textile industry (typical of the type of cost-of-protection model used in numerous case studies) reveals that tariffs are 30% less costly (quotas increase by 30% the unavoidable loss due to the constraint) in achieving the same 10% increase in average employment. The aggregation bias due to treating textiles as a homogeneous product implies that this is a substantial underestimate of the true relative inefficiency. Interestingly, tariffs in this study came within a few percent of the optimal (but infeasible) state-contingent tariff or quota system that shifts the protection in each state of nature.

In contrast to this, many union and management lobbyists press for quotas because they argue that their effects are more certain. It is unclear what they mean is more certain: quantity is, but price is less so. Other

magnitudes, like employment, may or may not be. In the simulation the difference in the variation in employment under the QR and tariff was trivial. A "noneconomic" constraint on average employment appears consistent with social norms in the United States, but constraints that are risk averse with respect to variation in employment are easily implemented. It is possible with such constraints and the "right" structure that a fixed quota could dominate a fixed tariff.³ On the other hand, tariff quotas, in which the rate of tariff steps upward with the volume of imports, are feasible and in fact fairly common. Chapter 7, from Anderson and Young (1982), sets out circumstances in which they are optimal. Essentially, high import volume is more damaging. For each range of imports subject to a given tax rate, tariffs dominate quotas due to their arbitrage efficiency. It seems clear that such mixed instruments can almost always dominate a fixed quota.

The design of the textile study seems representative of methods that would be used in most sectors, so I believe the results are representative also. In sensitivity analysis, the ranking of tariffs over quotas was preserved for a wide range of key elasticities, so the results appear to be robust. I conclude that tariffs are significantly more efficient in employment protection.

Output and profit protection motives should lead to the same results, since in this type of model all three move together. Furthermore, it can be shown (see chapter 6) that any fixed distortion that prevents equality of supply price and demand price produces the same qualitative structure of optimal instruments and thus should have the same ranking of tariffs versus quotas. This covers a wide variety of possible qualifications of the invisible hand theorem, including several versions of labor market distortion, regulated prices, fixed markup pricing by producers, fixed excise taxation, etc. For example, if the fixed distortion is g and it affects activity Y , the optimal policy involves $P - P^*$ proportional to gY_Q , where Y_Q is the marginal response of Y to imports.

1.3.2 Agricultural Protection

Agricultural products that are price supported have a variety of restrictions on competitive imports (and occasional export subsidies). In the EEC (European Economic Community) the system of variable levies taxes imports or subsidizes exports to maintain the price target. In the United States, the primary policy instrument is the domestic support purchase program, but import quotas and occasional export subsidies arise in order to limit domestic support budgets.⁴ What lies behind the support programs

are apparently two motives: to raise farm income on average and to reduce its variation. If the former motive dominates (or if the market were reasonably certain, which is counterfactual), the analysis of optimal policy may be entirely subsumed under that of chapter 5, with the arbitrage efficiency of *ad valorem* tariffs presumptively dominant.

The new element is the reduction of variance motive. A deep consideration of the capital market failure that makes this desirable has yet to be done and is beyond the scope of this book. So a “noneconomic” lower limit price is exogenously given. Trade policy alone could be used, leading to variable levies. Or as in the U.S. case, buffer stock policy can also be used.

Chapter 8 considers commercial and buffer policy in the presence of price supports and deficit limits. In the absence of storage cost, it is always welfare improving to add buffering. But the buffer agency is clearly subject to some restriction in its budget. Suppose this takes the form of an average constraint on the net deficit to be covered by a subsidy from general revenue. If the randomness is primarily domestic, the arbitrage efficiency of the tariff is desirable (buy more imports when domestic willingness to pay is high), and the tariff dominates. On the other hand, the arbitrage efficiency of tariffs (buy more imports when external cost is low) turns out to be a disadvantage if the source of randomness is external price shifts. In this case, the quota does a better job of protecting the domestic buffer agency’s budget. Greater imports when price is low increase the budget pressure precisely in those states where it is more acute. The optimal policy turns out to be a feasible combination of a specific tariff and an *ad valorem* subsidy on imports, so the appropriate tax system is still superior. The rather bizarre nature of optimal policy is a reminder that price supports are an inefficient method of increasing and stabilizing farmers’ incomes. Capital market methods are presumptively preferable, though a satisfactory theoretical treatment is lacking. Empirical work on the ranking of simple tariffs and quotas in practice for agriculture remains to be done.

1.3.3 Revenue Constraints

When trade distortion revenue is important to the government budget (as in many LDCs—less-developed countries), it is *possible* that a fixed quota is superior to a fixed tariff in raising revenue under uncertainty (Young, 1980). Most empirical work is based on demand and supply curves with either (1) additive or (2) multiplicative random terms. For these cases, a fixed specific (1) or fixed *ad valorem* (2) tariff dominates a fixed quota. Thus quota dominance is unlikely. For more nonlinear types of randomness, the

arbitrage efficiency of the tariff can be outweighed by the quota's superior ability to confine trade so that the elasticity of import demand or export supply is closer to its optimal value in each state, given by the inverse elasticity formula. This is advantageous.⁵

For the certainty case, chapter 5 shows that if imports enter preferences or the technology weakly separably and in addition within-group expenditure shares are invariant to aggregate import expenditure, the optimal revenue tariff is uniform *ad valorem*. More generally, the arbitrage efficiency of the tariff creates a presumption in favor of the *ad valorem* tariff, since "nonneutrality" must enter in the right way and strongly to overcome it.

For developed economies, trade revenue is a trivial and disregarded fraction of government receipts, so even aside from the esoteric nature of possible quota dominance under a revenue motive, this seems unlikely to produce a motive for quota preference.

1.3.4 Income Smoothing

Finally, when capital markets are insufficient to ensure adequate risk reduction, there may be reason to intervene in traded goods markets to make trade policy do some of the work of insurance markets at the aggregate level.⁶ Aggregate risks such as external price or domestic productivity shocks cannot be diversified, and without complete markets in Arrow's sense (due to transactions cost) individuals bear risk. The social surplus magnitudes of figure 1.2 are now weighted by the marginal utility of income. Chapter 9, from Young and Anderson (1982), is an example of this type of analysis.⁷ With no other distortion, the optimal small country policy remains free trade, but a "noneconomic" constraint on average imports introduces trade policy—which then has an income-smoothing component. For domestic shocks, the arbitrage efficiency of tariffs also achieves real income smoothing over states, since the marginal utility of income will be positively correlated with the marginal net benefit of imports $P - P^*$. With external price disturbances, it *can* turn out that QRs dominate tariffs because they do a better job of buffering real income. A rise in the foreign price of imports reduces real income. On arbitrage grounds, it implies purchasing less imports. An increase in permitted imports in the given high price state, on the other hand, would raise income in that state, and hence buffer the real income from the price shock. This type of possibility is most likely when other forms of real income smoothing are few, when the imports controlled account for a large share

of expenditure, and when external randomness dominates. For developed economies, at least two and more likely all three of these conditions are not met, but they are relevant for some less-developed countries and other small open economies.

This section has shown that tariffs ordinarily dominate quotas. Even in cases where the opposite is possible, a stepped tariff function, which is feasible, is superior since it can preserve part of the exchange efficiency of the tariff.

What structure could lead to the converse? Evidently, this occurs when the rate of social damage from imports is steeply rising at level Q .⁸ It is difficult to think of normal circumstances that imply this property and yet allow positive trade. For example, a zero quota on heroin is no doubt optimal, since even a small amount causes great damage to health. The examples of side effects considered here, such as employment displacement, do not have the same leap in the rate of social damage at a critical value with plausible supply and demand configurations.

The most reasonable exception is in wartime. For example, physical shipping constraints severely limited the import volume capacity of the United Kingdom in 1942, essentially requiring an aggregate fixed quota constraint. Consider applying the analysis of arbitrage efficiency to the problem. Effectively, the marginal transport cost was infinitely responsive to variations in quantity at the limit. Any *arbitrage* efficiencies of reallocation over states would be overwhelmed by the response of transport costs. Also, for reallocation across goods, in the absence of a market for war material, and with the price control system for consumer goods, arbitrage over quota allocations within the aggregate constraint would be likely to impact other “distortions” of large magnitude. It could well be inefficient.⁹

1.4 Political Economy and Quotas: Conclusion

This chapter argues that quotas are an inefficient means of achieving protectionist goals. For a given distortion of trade, national income will be lower; hence an enlightened planner will presume the tariff is better.

Recent developments in political economy models of protection suggest that the preceding discussion of quota inefficiency is incomplete, since it holds the intervention exogenous. Endogenous choice of protection also has interesting implications for the relative inefficiency of quotas. This section will develop a political economy model of relative quota inefficiency and contrast it with a bureaucratic model that attempts to establish the converse.

I assume it is reasonable to characterize the postwar era as one where protection occurs by means of quotas: a quota *regime*. Prior to World War II protection was by means of tariffs: a tariff *regime*. Regime choice is still held to be exogenous to the actions of agents in any particular market for protection. Protection can in fact be either in QR or tariff form. For the United States, it is the outcome of either a “political” or an “administrative” process (see Baldwin, 1987, for a description). Quotas in the form of VERs are generally the outcome of the “political” route along which the president diverts intense lobbying pressure by ordering the Special Trade Representative to negotiate such agreements. The “administrative” route, in which the International Trade Commission hears evidence and makes recommendations to the president upon complaint by interested parties under various legislative mandates, can result in either sort of instrument, at much lower cost to the parties. Protection changes via this route have been trivial in the last 30 years. Thus the United States may fairly be characterized as being in a QR regime in recent times. Prior to the New Deal, protection was obtained through Congress exclusively in the form of tariffs. Thus it was in a tariff regime.

Subsection 1.4.1 develops the political economy model of relative quota inefficiency, and subsection 1.4.2 presents the bureaucrats’ alternative.

1.4.1 The Political Economy Model

Relaxing the assumption that the height of the tariff or size of the quota is exogenous, I now shall argue that for a given configuration of political economic forces the market for protection will yield a worse distortion under a quota regime by (a) raising the level of protection achieved and (b) generating greater resource loss in “rent-seeking” behavior for any level of protection achieved.

In political-economy models protection is regarded as a commodity subject to supply and demand analysis like any other (see Baldwin, 1987, for a review). In implausible circumstances, the amount of protection exchanged on political markets would be socially optimal; ordinarily this is a zero level. The main reason this outcome is never observed is the free rider problem in the organization of consumers, who lose from protection. Smaller producer and labor groups have lesser free rider problems and sharper perceptions of their interest in protection. Thus they are able to obtain “excess” protection in the political market.

There are four factors that shift the supply curve to the right under a QR regime relative to a tariff regime. First, quotas have an impact on price that

by its nature is more difficult to establish than tax rates, which are published. In practice, even for a sophisticated investigator, it takes a great deal of work to find the gap between P^* and P . Frequently it cannot be done at all with any degree of accuracy. Thus the implied subsidy to domestic production and employment is hidden more effectively. For the naive consumer, QRs do not immediately suggest any increase in price, particularly when lobbying campaigns emphasize irrelevant but plausible catch phrases like “fair” market share and “level playing fields.” By lowering consumer group resistance the pursuit of protection via QRs is made cheaper for interest groups: the cost curve is lowered.

A second factor is also operative in some circumstances. General government revenue is often less valuable to officials and politicians than a smaller amount of rent more directly under their control. The lucky recipients of annual quota licenses can easily be identified and shaken down for campaign contributions and other political favors in return for renewal of their annual licenses. This is particularly so if resale is forbidden (which simplifies finding the rent) and if government does not sell them in the first place (which transfers the rent from the general revenue to the license holders). Both are nearly universal practice. Even in the case of VERs, where the recipients are foreign nationals, the identifiability of the beneficiaries is useful to governments and bureaucrats. The sugar quota is infamous for the degree to which political/national “security” motives dominate economic considerations in the allocation of licenses. A trade tax, in contrast, simply dumps into general government revenue the entire sum $PABP^*$ in figure 1.1. If this is less valuable to key officials and politicians than a smaller sum that can be recaptured from license holders, a QR regime lowers the supply curve for protection.

Third, bureaucrats enjoy extra prerogatives under a QR regime. The quota limit is a number with much less apparent significance than a tax. It must be interpreted relative to a technical model that relates it to the employment or other target of protection. Further, the quota must be allocated to individuals by some process that subdivides the original single number into hundreds of further numbers. All this requires extra staff and allows the agency to display competence in producing “scientific” numbers. If resale is prohibited, as in the U.S. cheese import quota system, further staff members are required to monitor the use of licenses. These features serve to lower resistance to protection among bureaucrats, which shifts the supply curve for protection facing the lobbyists to the right.

Fourth, quota licenses that are not auctioned transfer rent to potential opponents of protection (importers and foreign exporters). They may thus

not oppose the lobbying efforts of proponents, or offer only token resistance. This shifts the supply curve for protection to the right.

Finally, two factors shift the demand curve for protection to the right under a QR regime relative to a tariff regime. First, the identifiability of beneficiaries may help ease the free rider problem. Pursuit of protection costs money, which must be raised by voluntary subscription. The rent transfer implied by the gift of a quota license may convert an opponent into a proponent who will contribute to the lobbying for protection. There was some suggestion in the press that Toyota and Nissan were in this position in the last stages of the Japanese auto VER debate.

Second, under imperfect competition (see below), the rents earned by protected firms will be higher under a QR regime than with a tariff that permits the same volume of imports. Thus they should be willing to spend more in pursuit of those rents, shifting the demand curve to the right.

The last observation points to the other source of relative inefficiency of a QR regime in the political economy model. Rent-seeking behavior involves the expenditure of real resources on capturing or defending rents. Krueger (1974) has argued that this is greater under a QR regime. Bhagwati and Srinivasan (1983, for example) have disagreed without a detailed consideration. Most analysts have sided with Krueger, also without a detailed consideration. The above arguments imply higher protection under a QR regime, hence higher absolute amounts of waste on rent seeking. But the last two factors imply higher relative rent seeking. This is compounded by factors having to do with heterogeneity. QRs confront bureaucrats with an allocation problem that tariffs solve automatically. Considerable rents accrue due to the allocation problem, and considerable resources are spent to capture or defend such rents.

The political economy model implies that when "constitutional" arrangements establish a QR regime, the level of protection will be higher than if a tariff regime obtained, and that expenditures on rent seeking will be higher. In principle these are testable propositions, but in practice it would be hard to control for all other factors to find episodes of protection where the only difference is in the permitted mode of protection. Casual empiricism supports the conclusion in two ways. First, the experience of LDC trade liberalizations under QRs is usually that the extent of trade increase is a surprise; protection was greater than informed analysts expected. Second, many QRs in developed countries have very high tariff equivalents (where they can be measured). They are not only high with respect to current tariff levels (which is scarcely surprising) but high with respect to historic high tariffs, such as Smoot-Hawley. Their distortionary

effect is exacerbated by the generally low tariffs on non-QR imports, so that they have a great effect on relative price.

The political economy model suggests that a simple method of reducing the attractiveness of QRs is routinely to establish government monitoring of foreign and domestic prices of quota-constrained categories. No other category of government subsidy has so little accountability, so this should not be objectionable in principle. The cost of the necessary staff additions at the Bureau of Labor Statistics should be part of the cost-benefit analysis of any QR proposal. Alternatively, government auction of quota licenses would establish a similar information base.

Finally, the choice of regime can be encompassed within the political economy model. Above, this is treated as exogenous, but it is of course ultimately endogenous. Cassing and Hillman (1985) have initiated discussion of this point, asking whether politicians should prefer a tariff or a quota regime. Their answer is that it is ambiguous, an answer also suggested by the discussion above. The outcome for interest groups is also ambiguous. Consumers, of course, should prefer a tariff regime.

The nature of the arguments above suggests a further relative defect of quotas. Milton Friedman connects the efficiency virtues of the market to its virtues in preserving individual freedom. The connection holds in comparing tariffs and QRs: the arbitrage efficiency property of tariffs also protects liberty. In contrast to quotas, tariffs are *anonymous* in their effects; beneficiaries need not be related to officials or judged "worthy" in order to receive benefits. Obnoxious and inevitable features of QRs are their secretiveness (because it is so difficult to monitor them), absence from public review, dispensation by arbitrary officialdom, and most important, creation of personality.

1.4.2 The Bureaucrats' Weak Case for Quotas

For completeness consider three reasons often proposed by bureaucrats in favor of quotas. They appear weak on reflection.

1. One reason given is that quotas are more certain in their impact on the domestic market than tariffs, sometimes expressed as concern about "import surges." Protectionist lobbyists have succeeded in diverting policy analysis from essentials with this classic bit of propaganda, suggesting storm surges falling on an unprotected shore.

In an environment with random shifts in the relevant demand and supply curves, an increase in imports can be caused by either an outcome

of the random process or a change in the underlying parameters of supply and demand. If it is the latter, any form of intervention should in general be altered following a resimulation of the model on which the original intervention was based.

If it is the former, a little reflection reveals that more certainty about quantity means less certainty about price. The price of the import is what controls the position of the demand curve for the domestic product. Thus it is not clear that this argument has any merit even considering only the interests of the domestic import-competing firms or workers.

The preoccupation of this view with quantity usually comes from the fundamental error of ignoring substitution effects. Outside the economics profession this error is nearly universal, despite such examples as the recent humiliation of the "limits-to-growth" modelers (e.g., Forrester, 1971).

A more sophisticated error leading to a focus on quantity is based on the mercantilist view of the world as heavily cartelized in all markets. Imperfectly competitive environments lead, however, to quotas being relatively inefficient (see chapter 2).

2. In international political terms, bureaucrats argue that the VER is superior to a tax because (a) it avoids multilateral negotiating over the tariff retaliations automatically produced under GATT and (b) it provides compensation to the foreigners losing markets in the form of quota rents. This argument is not really compelling outside the short run. First, VERs are not much simpler to negotiate than taxes. In fact, such trade arrangements typically end up in increasingly cumbersome multilateral negotiations. Over time the arrangements become more and more complex, as new entrants are subjected to VERs while "fairness" compels continued compensation of the original incumbents. Second, it has always seemed to me somewhat implausible to rationalize a selfish national policy under the guise of its ability to subsidize foreigners. Even if this is accepted, the advantage of VERs over tariffs in compensating the original incumbents becomes a disadvantage as new entrants must be dealt with. If distributive justice is to be served, it is generally the newest entrants who are most worthy of special consideration. In any event, compensation could equally well be served by voluntary export taxes. Another layer of complexity is added by "noneconomic" criteria, as in the recent (1986) debate over inclusion of South Africa in the textile agreement. Thus the "bureaucratic cost" of "orderly marketing arrangements" (OMAs, multilateral VERs) is not in practice lower than for tariff changes under GATT. Indeed, it seems higher in the long run.

3. Government officials claim that the above argument implies that VERs are very flexible, and can thus easily be removed when the need has passed. In practice this claim rings hollow. The textile industry has been protected for almost 30 years and has long since become competitive in the eyes of analysts. The only example of removal extant is that for Japanese autos, and the degree to which trade in autos has actually liberalized is debatable. Furthermore, it seems clear that had Chrysler been joined in its campaign to keep the VERs by GM and Ford, the campaign would have succeeded. The latter wanted an end to the VERs in order to bring in more imports from their Japanese affiliates.

If protection cannot be avoided, it should be done with taxes in normal times. If the rent transfer property of VERs is a significant contributor to international harmony, it should be accomplished by voluntary export taxes (VETs). Foreign governments would secure the revenue, which they could use to dispense trade adjustment assistance or any other worthy cause.

Appendix 1.A: The Reduced Form Primal Methodology

The methods used in succeeding chapters are specializations of a general *reduced form primal* model. This appendix offers a formal treatment of the general case. There are two reasons for doing so. First, the underlying unity of the work will be more clearly revealed, and the reader's passage across chapters eased.

Second, the usefulness of the *reduced form primal* method seems to be underappreciated in the profession. Dual methods offer added simplicity over primal methods by embedding optimizing behavior and obtaining the dependence on price that is usually the focus of econometrics and partial analysis of the behavior of agents. *Reduced form primal* methods embed *market equilibrium* requirements into dual methods. On the other hand, by inverting to make prices depend on quantities, they appear to be a step backward. Their main general advantage over dual methods for welfare economics is that the derivatives with respect to quantities are *equilibrium* prices, hence have an immediate interpretation in terms of taxation (see Deaton, 1979, for a similar argument). Primal methods also have first derivatives in terms of prices, but these equal equilibrium prices only at an optimum. Thus they are not directly useful in evaluating away from an optimal point. The power of this feature of the *reduced form primal* method is revealed in the numerous applications below to the ranking of second-best instruments.

The models of this book have in common a representative consumer whose interests are advanced by a benevolent government planner, subject to unavoidable distortions. These are usually in the form of "noneconomic" objectives that constrain the planner's actions. Some concern is paid to relating the "noneconomic" to deeper "economic" targets, but they are always exogenous to the planner (thus suppressing political economy for simplicity). Imports are usually

infinitely elastically supplied by the world economy for simplicity, in order to suppress the routine “optimal tariff” effect.

Some aspects of general equilibrium linkages are used throughout. The revenue accruing to trade interventions is usually assumed to be distributed back to consumers without altering their decisions at the margin. Profits in domestic import-competing activities are also distributed to consumers without altering decisions at the margin. Prices in domestic import-competing activities are jointly determined with domestic prices of imports; further interaction with the domestic economy is usually suppressed by the assumption that relative prices in the rest of the economy are not altered by a change in output in the import-competing sector.

Subsection 1.A.1 sets out a general canonical model. Subsection 1.A.2 notes its specializations in succeeding chapters.

1.A.1 The General Model

The indirect utility function of the representative consumer is $V(P, H, I)$, where P is the domestic price of imports, H is the domestic price of import competing goods, and I is consumer expenditure. Other prices r are suppressed in the numeraire. When dealing with randomness, $V(P, H, I)$ is for a given realization of the random process (state of nature, indexed by s) and welfare is the mean of V , $E[V(\cdot)]$; i.e., additive separability over states of nature is imposed on preferences. When dealing with a set of imports with no randomness, P is a vector. When domestic and foreign goods are perfect substitutes, P and H are identical.

Expenditure is constrained by income (in the absence of intertemporal shifting via assets). Various aspects of general equilibrium linkage are developed, depending on the purpose of the exercise. In partial production equilibrium, the sources of consumer income are exogenous income I^0 , government revenue, G , and profits in the domestic import-competing sector, π : $I = I^0 + G + \pi$. When a general equilibrium of production is developed, $I^0 + \pi$ becomes the revenue function $R(H, r)$. When the other domestic good is developed explicitly, its production is X and consumption is Z . Consumers always take I as exogenous. G will be the rent on quota licenses or the tariff revenue: $G = [P - P^*]Q$, where Q is the quantity imported and $[P - P^*]$ is the margin between P and the foreign price P^* . The home industry profits π are a function of the import-competing good's price H , $\pi(H)$. When randomness is studied, the producer decisions are assumed for simplicity to all occur ex post, after uncertainty is resolved. Then output is π_H by Hotelling's lemma. Finally

$$I = I^0 + [P - P^*]Q + \pi(H), \quad (\text{A.1})$$

or

$$I = R(H, r) + [P - P^*]Q. \quad (\text{A.1}')$$

Market equilibrium constrains prices in terms of quantities. For a given set of binding (for simplicity, with later relaxation where it is useful) import quotas Q^0 , the market clearing prices satisfy

$$-V_P/V_I = Q^0 \quad (\text{using Roy's equality}), \quad (\text{A.2})$$

$$-V_H/V_I = \pi H \quad (\text{using Roy's equality and Hotelling's lemma}). \quad (\text{A.3})$$

Substituting (A.1) into (A.2)–(A.3) yields a system of equations sufficient to determine P, H in terms of Q^0 .

Denote the reduced form price functions based on the implicit solution to (A.1)–(A.3) as $P(Q^0), H(Q^0)$. Substitution back into the indirect utility function yields the *reduced form primal utility function*:

$$v(Q^0) = V(P(Q^0), H(Q^0), [P(Q^0) - P^*]Q^0 + \pi(H(Q^0)) + I^0), \quad (\text{A.4})$$

or

$$v(Q^0) = V(P(Q^0), H(Q^0), [P(Q^0) - P^*]Q^0 + G(H(Q^0))). \quad (\text{A.4}')$$

Due to the envelope properties of efficient production and consumption, and equilibrium as expressed in (A.2)–(A.3), the first derivative of (A.4) or (A.4') is remarkably simple:

$$\begin{aligned} v_Q &= -V_I Q^0 P_Q + V_I Q^0 P_Q + V_I [P - P^*] \\ &\quad - V_I D H_Q + V_I Y H_Q \\ &= V_I [P - P^*], \end{aligned} \quad (\text{A.5})$$

where D is the consumption of the domestic good and Y is its production. By market clearance $Y = D$; hence the second equality follows. (A.5) is the marginal net benefit of an additional unit of imports. With no constraints the optimum is attained where $v_Q = 0$, which implies free trade.

LEMMA Under risk aversion, $V_{II} \leq 0$, v is concave in Q .

Proof of Lemma The lemma holds if the second derivative matrix is negative semidefinite. Let $\Phi = (P, H)$ and $\Psi = (Q, D - Y)$, the excess demands. The prices are implicit functions of Q through

$$Q = Q(P, H, G(H) + [P - P^*]Q),$$

$$0 = D(P, H, G(H) + [P - P^*]Q) - R_H(H).$$

To obtain the second derivative matrix of $v(\cdot)$, I first obtain the price derivatives Φ_Q . Using the Slutsky decomposition and the envelope theorem, market clearance requires

$$(i) \quad \Phi_Q = [\partial \Psi^{\text{comp}} / \partial \Phi]^{-1} \{K - \Psi_I (P - P^*)\},$$

so that Φ_Q equals

$$\Phi_Q^{\text{comp}} \{K - \Psi_I (P - P^*)\}.$$

Differentiating (A.5),

$$(ii) \quad v_{QQ} = V_I P_Q + [P - P^*][(\partial V_I / \partial \Phi)' + V_{II} \Psi'] \Phi_Q + V_{II} [P - P^*][P - P^*].$$

P_Q is the matrix formed by deleting the last row of Φ_Q : $P_Q = I^{\Psi'} \Phi_Q$. Now use the Slutsky decomposition for the marginal utility of income:

$$(iii) \quad \partial V_I / \partial \Phi = V_{II}(-\Psi) + V_{I\Phi}^{\text{comp}} = V_{II}(-\Psi) - V_I \Psi_I.$$

Substituting (iii) into (ii),

$$(iv) \quad v_{QQ} = V_I I^{\Psi'} \Phi_Q + V_{II}[P - P^*][P - P^*]' + V_I[P - P^*](-\Psi_I') \Phi_Q.$$

Collecting terms and using (i),

$$(v) \quad v_{QQ} = V_I \{ I^{\Psi'} - [P - P^*] \Psi_I' \} \Phi_Q^{\text{comp}} \{ I^{\Psi} - \Psi_I(P - P^*) \} \\ + V_{II}[P - P^*][P - P^*]'. \quad (v)$$

Under the hypothesis of risk aversion $V_{II} \leq 0$ so the second term is negative semidefinite. The matrices to the right and left of Φ_Q^{comp} are transposes and Φ_Q^{comp} is negative semidefinite; hence the first term is negative semidefinite. QED

The canonical problem facing the planner is to maximize (A.4) (or its expectation) over the heterogeneous elements of Q^0 (or Q^0 for each state of nature). The constraint is assumed to be a concave function of Q^0 (or the expectation of a concave function of Q^0). These functions are well-behaved representations of "noneconomic" objectives.

1.A.2 Special Cases

A number of applications (chapters 4 and 6–10) deal with randomness. The objective function is the expectation of (A.4) or (A.4'). Save for chapter 9, it is assumed that variation in the marginal utility of income is trivial; hence the objective function is expected surplus. This can be justified on practical grounds in partial equilibrium (the variation in the income focused on is trivial relative to national income). It can also be justified as risk diversification opportunities exist such that consumers can achieve complete smoothing of marginal utility of income over states. Chapter 9 relaxes this assumption to consider a general equilibrium economy with no risk smoothing possibilities for consumers.

Chapters 2, 3, and 5 suppress randomness. In chapters 3 and 5, imports are heterogeneous, so that Q is a vector. In chapter 2, there is one import and one domestic substitute, but a tariff or quota can differ due to shifting the structure of competition between firms; hence the Y choice associated with a given Q depends on the nature of the trade control. In chapter 3, the direct primal method is used to develop the optimal import quota vector; the interested reader can check that the method of this book gives the same optimality conditions (inessentially generalizing the model to include variable production).

The "noneconomic" constraint facing the planner under heterogeneity is often taken to be average imports (chapters 4, 7, and 9), or aggregate imports (chapters 2 and 3). A deeper approach uses employment or average employment (chapters 2, 5, and 6) and considers constraints on the variability of imports (chapter 7) or government price support budgets (chapter 8). In chapter 10, the constraint on variability is assumed to forbid all imports above a target level.

Symbol Table

The following symbols are used throughout the book (other symbols are locally defined as needed):

- Q = import demand quantity,
- Y = domestic import-competing production quantity,
- D = domestic import-competing demand quantity,
- Z = exportable good demand quantity,
- X = exportable good production quantity,
- L = labor demand quantity in import-competing production,
- I = consumer income,
- R = production revenue,
- G = government revenue,
- P = domestic price of imports,
- P^* = foreign price of imports,
- H = price of import-competing good,
- W = wage rate,
- r = price of exportable good,
- $e(\cdot)$ = expenditure function,
- E = expectations operator,
- $V(\cdot)$ = indirect utility function,
- $v(\cdot)$ = reduced form indirect utility function,
- $l(\cdot)$ = reduced form labor demand function,
- $w(\cdot)$ = reduced form expected surplus function,
- C = cost function in import-competing production,
- C^* = foreign import cost function,
- s = index of the state of nature,
- i, j = indices of import good type,
- t = specific import tax,
- τ = *ad valorem* import tax,
- ρ = coefficient of relative risk aversion.

Conventions

Subscripts denote differentiation unless they equal i or j . In the latter case they denote an import category index.

Prime denotes a particular value of a variable, save when it is used to denote a vector transpose.

A vector such as Q denotes either a row or a column vector, with context determining which. Where necessary for clarity, the row or column identity is made explicit.

\log denotes natural logarithm (to the base e).

Notes

1. Strictly, this is true in the broad class of models in which distortions are not too significant or are not perverse in just the right way. Models of the alternative class have structure that seems implausible to most economists. Even when distortions can justify protection, there is generally a superior domestic instrument available. The profession has had near unanimity of opinion for a liberal trade policy for over a century.

2. Conjectural variations are assumptions firms make about their rivals. Firm i assumes that when it raises output by one unit, its rivals will shift output by the "conjectural variation." Cournot behavior is when all firms assume a zero conjectural variation. Market share preservation is implied by an assumed equiproportional conjectural variation. This is cooperative behavior and is equivalent to joint monopoly. Rational, or "consistent," conjectural variations are when the assumed response is correct.

3. One candidate for this type of structure involves "dumping" of the type analyzed by Ethier (1982). Industries with fixed cost facing random shocks to demand or cost may occasionally find it optimal to sell below long-run marginal cost. When this occurs across international frontiers, foreign firms can in essence "export the adjustment cost" to domestic firms. Ethier disavows commercial policy based on his model because it is too rudimentary a basis for a welfare analysis. In a further development it is conceivable that a fixed quota would dominate a fixed tariff, because it is costly for home firms to deviate from their planned output, which implies rising costs of imports.

4. The empirical example of chapter 3 assumes the size of the support budget is of no concern for dairy products (save as reflected in the aggregate import constraint). That is a natural consequence of the assumed structure of a bureaucrat operating under limited information (not knowing cross-price elasticities), and was reflected in the publicly stated rationale for import quotas (see chapter 8). Nevertheless, for price-supported products, a concern for the support budget is appropriate. Under randomness for a single product, the administrator has better information for judging the relative performance of tariffs and quotas.

5. The same type of reasoning can lead to quota dominance in the "optimal tariff" case where a fixed tariff or quota is implemented to exploit monopoly power in trade.

6. The situation can be either a complete absence of insurance markets or a variety of intermediate cases. In the complete absence of markets for risk sharing, with capital committed in advance of knowing the state of nature, laissez-faire is not the

optimal policy. Second-best policies involve intervention in capital markets, but trade intervention can dominate laissez-faire. With the limited form of stock market analyzed by Helpman and Razin (1978), free trade remains optimal.

7. Their model takes as given a "noneconomic" constraint on average imports. They assume all decisions are taken *ex post*. This model can readily be extended to cover *ex ante* capital commitment with a Helpman-Razin stock market. In either case the optimality of free trade holds in the absence of the noneconomic constraint. Given the necessity of intervention, the insurance aspects of trade policy come into play, so long as capital markets are incomplete—do not fully smooth out real income fluctuations.

8. It is useful to relate this to the prices versus quantities literature inspired by Weitzman (1974). In the trade case the social marginal benefit function is given by $P - P^*$, and social cost by a variety of suggested numbers or functions. Picking a Q or a tax instrument always involves hitting a point on the marginal benefit function.

In Weitzman's case, simplifying to make production choice be *ex post*, the price or quantity instrument selects a point on the social cost function (which gives the producer's supply response). This somewhat alters the analysis, but the result is the same: the preference depends on the relative slope of marginal benefit and marginal cost functions.

In the trade case the point is simply that the steep rise in marginal cost needed for quota preference is implausible.

9. Furthermore, in such a shortage situation, distributive justice considerations among consumers can imply that a quota system without sale of ration coupons is superior to a tax (see Baumol, 1987, chapter 4). In normal times the imposition of quotas or tariffs is not accompanied by such a concern for distributive justice, the import premium being charged to all consumers equally.

VERs *may* be rationalized on distributive-justice-among-suppliers grounds, although this seems farfetched to me. An analysis of VERs along Baumol's "fairness" lines may be an interesting exercise.

References

- Anderson, James E. (1985), "The Relative Inefficiency of Quotas: The Cheese Case," *American Economic Review*, 75, 178–190.
- Anderson, James E., and Leslie Young (1982), "The Optimality of Tariff-Quotas," *Journal of International Economics*, 13, 337–351.
- Baldwin, Robert E. (1982), "The Inefficacy of Trade Policy," Frank Graham Memorial Lecture, Princeton University.
- Baldwin, Robert E. (1987), *The Political Economy of U.S. Trade Policy*, Cambridge, MA: MIT Press.
- Baumol, William J. (1987), *Superfairness*, Cambridge MA: MIT Press.

Bhagwati, Jagdish (1965), "On the Equivalence of Tariffs and Quotas," in *Trade, Growth, and the Balance of Payments*, ed. R. Caves et al., Chicago: Rand-McNally.

Bhagwati, Jagdish, and T. N. Srinivasan (1983), *Lectures in International Trade*, Cambridge, MA: MIT Press.

Buffie, Edward, and Pablo Spiller (1986), "Trade Liberalization in Oligopolistic Industries: The Quota Case," *Journal of International Economics*, 20, 65–82.

Cassing, James H., and Arye L. Hillman (1985), "Political Influence Motives and the Choice between Tariffs and Quotas," *Journal of International Economics*, 19, 279–290.

Deaton, Angus (1979), "The Distance Function and Consumer Behavior with Applications to Index Numbers and Optimal Taxation," *Review of Economic Studies*, 46, 391–405.

Ethier, Wilfred J. (1982), "Dumping," *Journal of Political Economy*, 90, 487–506.

Falvey, Rodney E. (1979), "The Composition of Trade within Import-Restricted Categories," *Journal of Political Economy*, 87, 1105–1114.

Forrester, Jay (1971), *World Dynamics*, Cambridge, MA: Wright, Allen.

Helpman, Elhanan, and Assaf Razin (1978), *A Theory of International Trade under Uncertainty*, New York: Academic Press.

Krueger, Anne O. (1974), "The Political Economy of Rent-Seeking," *American Economic Review*, 69, 291–303.

Rodriguez, Carlos A. (1974), "The Non-Equivalence of Tariffs and Quotas under Retaliation," *Journal of International Economics*, 4, 295–298.

Weitzman, Martin (1974), "Prices vs. Quantities," *Review of Economic Studies*, 41, 50–65.

Young, Leslie (1980), "Optimal Revenue-Raising Trade Restrictions under Uncertainty," *Journal of International Economics*, 10, 425–440.

Young, Leslie, and James E. Anderson (1980), "The Optimal Instruments for Restricting Trade under Uncertainty," *Review of Economic Studies*, 46, 927–932.

Young, Leslie, and James E. Anderson (1982), "Risk Aversion and Optimal Trade Restrictions under Uncertainty," *Review of Economic Studies*, 49, 291–305.