

**BANCO INTERAMERICANO DE DESARROLLO
DEPARTAMENTO DE INTEGRACIÓN Y PROGRAMAS REGIONALES
DIVISIÓN DE INTEGRACIÓN, COMERCIO Y ASUNTOS HEMISFÉRICOS
INSTITUTO PARA LA INTEGRACIÓN DE AMÉRICA LATINA Y EL CARIBE**



EXPLORING THE POSSIBLE BENEFITS AND COSTS OF FTAA

ARNOLD C. HARBERGER
UNIVERSITY OF CALIFORNIA

SEPTEMBER 2002

Given the unclear results of Latin America's past experience with free trade areas and customs unions, one wonders whether the present degree of enthusiasm about the prospects of a much more grandiose effort in the same direction (the Free Trade Area of the Americas) is really justified. In this presentation I will try to take a coldly analytical look at the pluses and minuses surrounding a prospective FTAA. I do not want to be a spoilsport, taking the wind out of the sails of what looks like a beautiful "ship of the future", but I do want to focus on the economist's tried and true criteria of costs and benefits. A full awareness of these will enable readers to reach their own judgments about what is in each country's best interest, and may also help produce a design for this new ship which makes it more beneficial to the participants, and more seaworthy, helping it survive the storms and turmoils that will most certainly appear from time to time.

The Standard Analysis of Customs Unions

Those familiar with the customs union literature are probably painfully aware that it is very easy for the analysis, even when it follows what seems to be a simple path, to end up being very complicated. I have tried in this paper to keep things simple all the way, while preserving the essence of customs union theory. The trick is to treat the world prices of the relevant tradable goods as constant, but to consider imports from different sources as imperfect substitutes on the demand side. This "assumption" allocates the full benefits of the customs union to demanders of imports from within the union. In this case the implicit transfer of tax money takes place within each country. What were formerly the tariff receipts of a country's government on imports from future member countries, now are lost as tariff receipts, but are received in the form of benefits by demanders in that country. Exports, under this simplification, are assumed to be expandable, within each country, at relatively constant cost. This is

certainly a reasonable assumption for manufactured exports, and for many though not all agricultural products (considering a middle- to-long-term supply response).

As was just indicated, the principal “first-order” beneficiaries of free trade among union members are each country’s own demanders who now pay less for the same bundle of imports than before. But this benefit has a counterpart in lost revenue to the country’s own treasury. Standard applied welfare economics treats this package as a simple net transfer, with the treasury’s loss being matched by the gain to demanders. (We will proceed on this assumption for the moment, but will later return to this point in order to present and discuss some important modifications.)

The standard analysis goes on to point out that there is an additional gain to demanders. This gain applies to the extra units of imports ΔM_U that they now import from other members of the union. If τ_U was the average tariff previously paid on M_U^0 of imports, the gain on the increment will be (approximately) $1/2 \tau_U \Delta M_U$. This values the first dollar’s worth of extra imports at $\$(1+\tau_U)$ and the last unit at simply \$1 and takes a linear approximation to the demand function for M_U between these “beginning” and “ending” points. This gain is shown by the triangle in the upper panel of Figures 1a, 1b, and 1c.

Now we have what looks like a net gain for the country in question. But this is not the end of the story. The next step is to recognize that as they take more of M_U , demanders will reduce their demands for other things. Standard welfare economics says that, so long as there are no distortions in the markets for those other things, we can forget about them for this particular purpose. But if there are

distortions in the other markets that contract as M_U expands, we must take those distortions into account.

But when a free trade area is being formed, it is almost inevitable that such other distortions will exist, even that they will be important. The reason is that the tariff concessions that are made to fellow members do not apply to the rest of the world. Thus Latin American countries would give concessions to imports from the U.S. and Canada, but not from Europe, Asia, Australia and New Zealand. To the extent that imports from these sources are reduced by ΔM_j , with the applicable tariffs being τ_j , there is an induced welfare loss equal to $\sum_j \tau_j \Delta M_j$. This loss is felt directly by the treasury of the importing country. The losses associated with this set of reactions are attributed in the customs unions literature to trade diversion.

If the country in question had a single uniform tariff τ^* to begin with, we would have $\tau_U = \tau_j = \tau^*$. The original gain would then be $1/2 \tau^* \Delta M_U$, and the induced loss would be $-\tau^* \Delta M_D$, where ΔM_D is the sum of all the ΔM_j 's -- i.e., the aggregate amount of trade diversion. It is easy to see from the above that the country comes out even on these counts when $1/2 \tau^* \Delta M_U = -\tau^* \Delta M_D$, that is, when $\Delta M_D = -1/2 \Delta M_U$. In words, the country comes out even (under an initial uniform tariff) when trade diversion is responsible for half the country's increase in imports (brought about by the union). Now if half the increase in imports is due to trade diversion, the other half must represent trade creation. This case is shown in the bottom panel of Figure 1a.

Thus we have the rather familiar refrain -- that customs unions are good when trade creation exceeds trade diversion (see Figure 1b), and are bad when the reverse is true (see Figure 1c). It is easy

to see that this is not a general proposition. The union can yield a net benefit with trade diversion equal to trade creation, simply by τ_u being greater than τ_d (the average tariff rate applicable to the ΔM_j), and it can yield a net loss when the reverse is true ($\tau_d > \tau_u$). But it is clear that trade diversion is bad, and that its costs increase with the tariffs applicable to the imports from which demand was diverted. Moreover, the gross cost of trade diversion is simply the loss in revenue due to the diversion of imports from other sources -- i.e., $-\sum \tau_j \Delta M_j$.

On The “Shadow Price of Government Funds”

Traditional applied welfare economics has a lot to be said for it, but I believe that it has a major flaw (if not an Achilles’ heel) in its treatment of government revenues and outlays. Valuing government receipts and expenditures at one peso for one peso, or one dollar for one dollar, makes for easy analysis by economists, but is it the relevant way for economists and policymakers to be thinking about their problems? Let me take you down the road I myself traveled in becoming convinced of the need to incorporate a “shadow price of government funds” into our exercises in applied welfare economics.

It was over ten years ago when I was approached by a group at the World Bank to prepare a summary of what was then the “state of the art” in economic cost-benefit analysis. This request was, however, not quite open-ended, for me to do whatever I thought best on the topic. Instead, I was asked specifically to treat, among all the other matters, the issue of whether or not a shadow price of government funds “belonged” as an integral part of modern cost-benefit analysis. This, it turned out, had been a matter of dispute between two groups of World Bank professionals, which I was supposed, in a sense, to help mediate.

At that time I had been actively involved in cost-benefit analysis, and in applied welfare economics in general, for well over 30 years. And I had not up to then incorporated a shadow price of government funds into my thinking or my writing on the subject. Thus if anything I might have been expected to come down on the side of those who considered the shadow price of public funds to be an unnecessary or at best an unimportant concept. And, indeed, when the subject was first brought up, that is where I expected my analysis to come out.

But it did not come out that way at all. One of the great things about almost any field of science is that one is not free to choose one's conclusions. One is the prisoner of the evidence, on the one hand and of the internal logic of the analysis on the other. One is swept, as by powerful waves in a stormy ocean -- even to places where one may not at first want to go. But in the end one is content, even pleased, to go there, because that is where the evidence and scientific analysis lead. One really doesn't have any choice.

That is more or less how I was led, or driven, to the conclusion that the shadow price of government funds had to occupy an important place in modern applied welfare economics. Now let me take you down the precise path that I followed in reaching this conclusion.

In that part of cost-benefit analysis that deals with public investment projects, nearly all practitioners had for many years considered that the "natural" source of the government funds involved is the capital market. Not only does this make sense where investment projects are concerned, it is also true that month by month, quarter by quarter and year by year, the government's marginal funds nearly always come from the capital market. Budgets made in advance never turn out as expected, and

shortfalls are almost always made up by borrowing more than planned, while unexpected revenues lead to borrowing less than planned, or to paying down the debt.

Our cost-benefit analyses, then, were conducted under the assumption that the funds in question came from the government's natural margin in its budgetary operations -- namely borrowing. We didn't feel troubled or uneasy about this -- after all, borrowing is also the natural margin for most business firms -- at least for those that account for the bulk of most countries' business activity.

But when I had to struggle with the World Bank's question, I ended up looking at two extreme examples. The first was a public sector electricity project. Here I assumed that the electricity tariffs were well chosen, and that over the life of the project sufficient funds were generated to cover its full cost. This means that the funds that were borrowed to finance the project were paid back at the market rate of interest, and that enough surplus was left over to cover the estimated loss of government revenue from investments that were displaced, as the government entered the capital market to borrow for this project.

The electricity project outlined above is a perfect example in which our standard procedure (up to that time) worked perfectly. A project yielding precisely the rate of return that we called the economic opportunity cost of capital would turn out to be just on the margin of acceptability. If it yielded a higher rate of return it would be more than just acceptable. If it yielded a lower rate of return it should be rejected.

The key to the electricity case is that all the benefits of the project come in the form of cash. But what about the other extreme -- say a road project (that is not a toll road). Here the benefits take the form of reduced costs of getting from here to there. Much of this cost saving turns out to be

reductions in the cost of running and maintaining vehicles, plus savings of time for their drivers and passengers. These savings are genuine, but they do not involve cash receipts to the treasury. This is the real difference between the road case and the electricity case.

Anyway, the road case played a key role as I started to try to answer the problem posed by my sponsors at the World Bank. Suppose we have a road project that passes the standard cost-benefit test. That is, it yields a rate of return equal to or greater than a properly calculated economic opportunity cost of capital, but the benefits do not come in the form of cash receipts of the government. What I did was make the mental exercise of following such a project through to the end of its economic life. The capital cost would have been financed by government borrowing, and so too would its current costs, year by year. The debt generated in this way would simply accumulate over the life of the project -- properly calculated, this accumulation would take place at an interest rate equal to the economic opportunity cost of capital. And then, even after the project itself was dead -- its economic life finished -- that debt would keep on accumulating indefinitely.

I think you can appreciate how this simple example of a road project gave me a major wakeup call. The idea of a project's life being over, but the corresponding debt not just going on forever, but rolling over forever, accumulating new interest charges every year, seemed too much to take. Even if all this was just implicit in our standard procedures, it was a signal that something was wrong with these procedures. Something had to be done.

I came to the solution in two steps. The first step was to set an arbitrary cutoff date -- a day of reckoning -- at the end of a project's economic life. Up to that point everything would be "standard", except that there would be a separate accounting of the implicit debt burden associated with the project.

This would represent the accumulated value of those cash outlays that had not been offset by cash inflows. Then, at the day of reckoning we would require that this accumulated debt to be paid off with tax money, instead of being allowed to accumulate forever. This “convention” would not change anything if tax pesos or dollars had an economic cost of one for one. But we know that tax finance carries a significant excess burden of efficiency cost -- call it λ . Thus, taking this into account we want to charge the road project with $(1+\lambda)$ times the debt that would accumulate up to the day of reckoning. We could do the same with the electricity project, but if it accumulated no debt, there would be no need for any adjustment. And if it had cash benefits yielding a positive net present value, the adjustment would be made using the factor $(1+\lambda)$ but it would appear as a $\lambda\%$ benefit rather than a cost.

The second step in the solution was to get rid of the “day of reckoning”. This was a way of stopping a process of accumulation from going on forever, but it represented an arbitrary (even though sensible) point in time. Let me take you through a little math. The present value of a project, evaluated at its final period N , is

$$(1) \quad NPV_N = \sum_{t=0}^N F_t (1+r)^{N-t},$$

where F_t represents the net flow of benefits minus costs in period t . Now let us break this down into two parts -- “cash” (C_t) and “goods”, (G_t). That is, $F_t = C_t + G_t$. Then

$$(2) \quad NPV_N = NPV_{CN} + NPV_{GN} = \sum_{t=0}^N C_t (1+r)^{N-t} + \sum_{t=0}^N G_t (1+r)^{N-t}.$$

Our exercise to this point entails adding a cost (or benefit) equal to $\lambda \sum_{t=0}^N C_t (1+r)^{N-t}$ to equation

(2). But it is perfectly clear that we get the same result by multiplying each and every cash flow (to and from the public treasury) by $(1+\lambda)$. And the way that is done is to apply a shadow price of government funds (SPGF) to each and every such cash flow (C_t) .

Thus, by a somewhat roundabout route we arrive at the conclusion that it is in fact appropriate to apply a shadow price of government funds, attributing an extra cost of the fraction λ per peso or dollar of public outlays and an extra benefit of the same amount per peso or dollar of cash receipts. I see no way in which we could sensibly modify the logical steps by which we arrived at this conclusion that would permit us to justify setting $\lambda = 0$, which is what traditional applied welfare economics has done.

Estimating the Shadow Price of Government Funds

Figure 2 illustrates how the shadow price of government funds is estimated for a particular source of funds (a tax τ_i). One raises the rate of tax by an amount $\Delta\tau_i$. The increment of tax revenue is $Q_i^0 P_i^0 \Delta\tau_i + P_i^0 \tau_i^0 \Delta Q_i^0$. The first of these items is positive, the second negative (because the increase in the tax rate will induce a reduction in quantity demanded). In the case shown, the supply is infinitely elastic, so the full effect of the increase in tax is felt by demanders. In such a case the increment in tax revenue (ΔR_i) will be

$$(3) \quad \Delta R_i = Q_i^0 P_i^0 \Delta\tau_i + Q_i^0 P_i^0 \tau_i^0 (\Delta Q_i^0 / Q_i^0)$$

$$= Q_i^O P_i^O \left[1 + \frac{\eta_i \tau_i^O}{(1 + \tau_i^O)} \right] \Delta \tau_i.$$

This is equal to area B in Figure 2 ($= Q_i^O P_i^O \Delta \tau_i$) minus area A ($= \frac{\tau_i^O}{1 + \tau_i^O} Q_i^O P_i^O \eta_i \Delta \tau_i$).

Now the efficiency cost of a tax like τ_i is simply the triangle labeled E. The increment to this efficiency cost, generated by an increase in the tax, is area A. So the increase in efficiency cost per peso or dollar of extra revenue is $A/(B-A)$. Algebraically it is equal to:

$$(4) \quad \lambda_i = -\eta_i \tau_i^O / [1 + \tau_i^O + \eta_i \tau_i^O].$$

Table 1 shows how this marginal efficiency cost ratio varies with the elasticity of demand η_i and with the tax rate τ_i from which one started. One can see that it is of relatively small magnitude (.023) for a tax rate of 10% and an elasticity of demand of -0.25, and that it is quite dramatically high -- .400 for tax rate of 40% and an elasticity of -1.0. The relevant elasticity is the elasticity of demand for the taxed commodity in the case where the supply elasticity can be taken to be infinite (a flat supply curve). Where the supply curve is upward sloping the relevant elasticity is what I call the elasticity of “response”. It measures the percentage reaction of the equilibrium quantity to a one percentage point rise in the tax rate.

Now before going any further, let us take stock as to what might be “plausible” magnitudes of λ for real-world cases. Tax revenues in Latin America range from around 10% of GDP all the way to some 30% of GDP (in Brazil). Now since large chunks of GDP are free of tax, the effective average tax rate on the taxed portions of GDP will clearly be a lot higher than these figures. The saving feature is

that if one considers simultaneously raising taxes over the full range of taxed activities, the relevant elasticity of response is likely to be quite low. For the moment, let us be conservative and consider the most likely range of λ to be between .10 and .20.¹

Discussion of the Results When a SPGF Is Used

The best place to start this discussion is Figure 3. Figures 3a, 3b, and 3c exactly replicate Figures 1a, 1b, and 1c, but with a shadow price of government funds (λ) equal to 20%. It is easy to see how much worse this makes the outcome of a standard applied welfare economic analysis.

Readers will recall that losses cancel gains in Figure 1a. That means the net loss in Figure 3a is precisely the two additional rectangles representing the excess burden of raising the revenues to replace the tax losses involved in the union. Figure 1b was created to illustrate the case where the standard analysis produced an estimated net gain. But, as that figure is drawn, that net gain turns into a net loss when the SPGF is taken into account, as shown in Figure 3b. Quite obviously, Figure 3c shows a very large net loss, adding the two SPGF rectangles to what was already shown as a net loss in Figure 1c.

These pictures give the main messages of this paper. First, one has to take very seriously the likely efficiency costs involved in replacing the tariff revenue that would be lost under a prospective FTAA. And second, when one does so it is pretty hard to be very optimistic about the likely net benefits of such a free trade area, if all one counts are the benefits and costs usually considered in

¹In one of the pioneering articles on the shadow price of government funds, Edgar K. Browning (1976) estimated it to lie between .09 and .16 for the United States. Empirical work in this area has been sporadic, and different methodologies have been followed. The profession is in real need of a systematic effort at empirical estimation, based on a very carefully elaborated methodology.

applied welfare economic analysis (including, of course, the relative latecomer to that analysis, the shadow price of government funds).

Table 2 is intended to go beyond the pictures (which are purely didactic and illustrative) to an exploration of a range of alternative assumptions which is likely to be reasonably representative of reality for many Latin American countries. Perhaps it is better to say, since the range of parameters explored in Table 2 is quite wide (though still quite plausible) that such a range is likely to encompass the actual situations of most of the likely FTAA countries.

The key assumptions concern the own-price elasticity of demand for imports from the “union” countries, on the assumption that the prices of imports from these countries falls, relative to other goods and services including, of course, imports from “other” (i.e., non-union) countries. For this own-price elasticity we explore in Table 2 the values -1, -2, and -3. My own belief is that -1 is a conservative estimate and that -3 is a reasonably generous one. That elasticity, η_{uu} , is enough to allow us to estimate the direct benefit per peso of direct revenue loss as $-1/2 \eta_{uu}\tau$, where τ is the assumed average initial tariff rate, and is taken as applying (as an average) both to initial imports from member countries and to imports from nonmember countries that would end up being affected by trade diversion.

The only other variable with an assumed value (apart from λ , for which we set a range of 0.1 to 0.2 in the previous section) is f_d . This is the fraction of the primary increase in imports from union sources (as a result of the free trade area) that comes as a result of diverting imports from non-union sources. With $\tau_u = \tau_d$, we know that the standard benefits and costs (not counting SPGF) add up to zero when the fraction in f_d is 0.5. Consulting the table one finds that in each of the 6 places where f_d

is 0.5, the direct benefit shown in column (3) is exactly cancelled by the indirect loss due to trade diversion [column (4)].

The efficiency cost of replacing lost revenue represents the two rectangles that were added in the passage from Figure 1 to Figure 3. Since all the gains and losses shown in Table 2 are expressed as fractions of the direct revenue loss ($M_u^O \tau_u^O$), the rectangle of extra cost that results from applying λ to this figure is represented in column (5), per peso of $M_u^O \tau_u^O$, as simply λ , which in the case of Table 2 is represented either as -0.1 or -0.2, depending on whether λ is 0.1 or 0.2. The remaining term in column (5) represents the smaller of the new rectangles shown in Figure 3 -- the application of the factor λ to the revenue loss due to trade diversion.

The final column of Table 2 simply presents the sum of the benefits (+) and costs (-) shown in columns (3), (4) and (5). I have consciously allowed the tariff rate to be represented by τ (rather than introducing an explicit numerical assumption) because tariff rates do vary substantially across countries and commodities. Readers can apply the specific value of τ they consider relevant for a particular case. Alternatively, one can use the formulas in column (6) to calculate the initial tariff rate that would render net benefits equal to zero (if it applied to both direct benefits and to diverted trade). Thus the first entry in column (6) -- net benefit equal to $(-0.1 - .05\tau)$ -- is negative for all positive values of τ . A net benefit that is always negative similarly appears for all the cases where $f_d = 0.5$, because we know for such cases that the direct benefits are cancelled by the revenue losses due to trade diversion, leaving only the efficiency losses involved in the application of the SPGF.

Taking now the last entry in each panel we find that in panel one the net benefit is $-0.2 + .104\tau$, which equals zero when $\tau = .104/.200 = .502$. For the second panel, last row we have zero net benefit when $-0.2 + 0.2\tau = 0$, or when $\tau = 100\%$. In the third panel, last row, this happens when $\tau = 67\%$. Readers will find, performing this exercise all through the table; that rates of tariff needed to yield a positive net benefit (when that is even possible) are much higher than prevailing average tariff levels in the countries of the western hemisphere.

In What Way Might FTAA Make Sense?

As far as I can see, looking for positive net benefits of the standard type as a consequence of a potential Free Trade Area of the Americas is pretty close to a lost cause. Yet I am not a pessimist about it, nor would I consider myself an opponent of it.

How can I be optimistic when a traditional-type analysis points so gloomy a picture? The answer lies in a different type of benefits -- a type not contemplated in the traditional applied welfare economic analysis. I am not here going to refer to vague or ill-defined concepts such as “dynamic comparative advantage”, though what I have to say may overlap somewhat the things that people may have in mind when they appeal to such concepts.

I want instead to focus on the role of openness, and of competition in general, in the growth process. We know that improvements in total factor productivity (TFP), which I put under the more appealing label of “real cost reduction”, account for an important part of just about every episode of notable economic growth. Real cost reduction played a major role in the major spurts of growth in the main industrial countries, and also in the well known episodes of truly successful growth (Brazil 1968-79; Chile, 1985-2001; Panama, 1960-78; etc., etc.) in Latin America.

But while it is easy to show the importance of real cost reduction (RCR) in actual past episodes, it is not easy for policymakers to push buttons so as to create it at will. The lesson here is that all RCR really takes place at the level of the productive entity (the firm). It takes literally thousands of different forms -- different across activities and firms at any one time, and different across time for any firm, industry or sector. Governments can help a lot by providing an environment that is favorable to the search for real cost reductions. -- control of inflation, securing of property rights, fighting against market distortions, etc.

Reducing tariffs obviously fits into this picture as a way of reducing distortions, but we have already shaken that tree and found it rather barren in the particular case of a free trade area. But there is another way in which reducing tariffs can benefit a country -- namely, by exposing firms to a more rigorous discipline of competition. Studies of the growth process have repeatedly shown that important trade liberalization efforts seem to stimulate real cost reduction and economic growth. There can be little doubt that the way this works is to scare or shock lazy firms into a serious search for ways of reducing real costs. When they are protected by ample tariff and nontariff barriers, firms tend to pursue a comfortable way of life -- happy, working away, even maybe working pretty hard, but not subject to shocks that can threaten their profitability and even their very existence. Competition from the world marketplace has the capacity to provide a big adrenaline rush to owners and managers, to make them stay up nights and weekends, and to stretch the limits of their ingenuity and creativity. As a consequence, lower trade barriers means more real cost reduction, more investment as new methods are introduced, higher rates of return to capital, and more growth.

I wish we had a way of putting all this into simple diagrams and tables which would predict the magnitude of benefits of this type that come from a general trade liberalization. But we do not have such a way. Much less do we have it for the creation of a free trade area, with its twin faces of trade creation and trade diversion. But one thing, I believe, can be said, which is that in its role of opening an economy to competition, a free trade area or customs union does part of the job of a general program of liberalization. It brings some new competition, maybe quite a lot, even if it does not bring all the new competition that would be stimulated by free trade.

Implications for Design and Strategy

To me, the above considerations lead quite directly to the view of a regional free trade area as a halfway house -- a way station along the road to generalized freer trade. If the benefits of enhanced competition from within the union can be great, those that would come from enhanced competition on a worldwide basis would be even greater.

The alternative view would be that it is a good idea to free up trade within the region, but a bad idea to do the same on a global and nondiscriminatory basis. This conjures up the vision of a new world order in which protectionism is abjured within regional blocs but welcomed as between and among such blocs. Such a mindset could easily lead to thoughts of increasing tariffs and other trade barriers on an interregional basis. But this would entail greater efficiency costs and more trade diversion, and would also mean reduced challenges and stimuli for a continuing process of real cost reduction by producers within the region.

At bottom the arguments that favor freer trade among the countries of the region also favor freer trade among the regions of the world. A FTAA that is built on a recognition of this truth will surely be a

step in the right direction. One that is built on a denial of this truth would lack serious analytical underpinnings, and would be very questionable in terms of its long-run contribution to regional welfare.

APPENDIX

In this Appendix I address two questions of a somewhat technical nature, which supplement the analysis of the main text of this paper, but are not essential to it.

A. Customs Union Analysis With Upward Sloping Supply Curves

Figure 4 sketches the case of eliminating a tariff on a product whose supply within the region is less than infinitely elastic. One can easily see how this makes the analysis more complicated. There is the same rectangle of loss of revenue to the governments of country 1, but now only a part of this revenue loss is offset by a gain to that country's own demanders (rectangle marked G_1- , D_1+). The rest of the G_1 revenue loss is reflected in a rise in the prices received by suppliers in other member countries (rectangle marked G_1- , S_2+). With many different products affected under a free trade area, there could be many different patterns of distribution of these transfers. In some countries the gains to their demanders from their own tariff reductions, plus the gains to their suppliers from tariff reductions by other union members, would exceed the direct revenue losses of their governments, while in others they would fall short. Overall, however, across the entire free trade area, the direct revenue losses would be matched by the gains to demanders of the "own country" and to suppliers of the other member countries. On top of this would be the indirect revenue losses due to trade diversion, shown in the lower panel of Figure 4, plus, of course, the efficiency costs of replacing lost revenue, represented by λ and shown in Figure 3.

The point I would like to stress, however, can be seen in the upper panel of Figure 4 alone. It is that the direct benefits that are attributed under the assumption of infinitely elastic supplies are greater than what one would get under upward rising supply curves. This difference is shown by the triangle

labeled K. Thus, by making the simplifying assumption of infinitely elastic supplies, I have tended to overstate the benefits of a customs union, if in fact this assumption is not met in reality. In short, my main line of argument in this paper would be strengthened if we abandoned that assumption.

B. Discussion of Assumed Elasticities of Import Demand

The elasticity of demand for imports is one of the most difficult parameters one encounters in economics. In the first place, the demand for imports is itself an excess demand, being the difference between a nation's total demand for a good (or a set of goods) and its own national supply of that good (or set). This fact alone leads to a presumption of a relatively high elasticity of import demand. If $M \equiv D_N - S_N$, where D_N and S_N stand for national demand and supply, then

$$\eta_m = (D_N/M)\eta_d - (S_N/M)\epsilon_s.$$

That is, the elasticity of import demand is a multiple (D_N/M) of the elasticity of domestic demand for a good (or set), η_d plus another multiple (S_N/M) of the elasticity of domestic supply ϵ_s . The two are effectively additive because η_d is itself negative, while ϵ_s , being positive, is preceded by a minus sign.

The above excess demand formula is strictly true for fully homogeneous products, where the same change in price faces all demanders and suppliers. Where imperfect substitutes are concerned, the prices of close substitutes will change less than the prices of the imported goods themselves. The easiest way to incorporate this into our thinking is to allow for fractional responses g_d and g_s -- i.e.,

$$\eta_m = g_d(D_N/M)\eta_d + g_s(S_N/M)\epsilon_s.$$

Without elaborating on the details underlying such an adjustment, let me simply point out that these fractional responses work to mitigate the magnifying effects pointed out above.

Now we turn to the underlying elasticities η_d and ϵ_s . These can be tricky concepts in quite ordinary supply and demand situations, but perhaps they are even more tricky in the case of import demand. The reason is that one encounters the need to measure the response of imports of a given commodity to price changes of quite different types. Consider these four:

- i) The world price of wheat falls by 10%, other prices remain constant;
- ii) A country's real exchange rate appreciates by 10%, leading all tradable goods prices to fall by the same percentage;
- iii) A country liberalizes trade, starting from a uniform import tariff of 10%;
- iv) A country liberalizes trade, starting from a uniform tariff of 10%, but only for imports from within its region.

Obviously, even though we are talking about a 10% price change for wheat in each case, the effects on demand for imports of wheat will be different in all four of the listed cases. Case i) and ii) are at the extremes. In case i) only the price of wheat falls; in case ii) those of all importable and exportable goods fall by the same percentage. Case ii) is the premiere real-world example of a composite commodity (tradables in this case). Here it is well known that except for rare cases of complementarity the elasticity of response will be much lower for any given good, because the only substitution involved is with goods outside the composite (in this case the nontradables, with respect to which one can probably count on very low underlying elasticities η_d). Case iii) is like case ii), but the composite good in question is smaller, consisting of importables rather than all tradables. Because more goods are left out in this case, one would expect the elasticity of response to be greater (in absolute magnitude) than in case ii).

Finally, we come to case iv), which is the one that concerns us in this paper. In a sense, one can look at case iv) as a subspecies of case iii), operating on a subset of importables rather than on all importables. However, if so it must be regarded as a very special sort of subset, in that it does not make a distinction according to a commodity classification, but rather according to region of origin. There is a strong presumption that the degree of substitutability will be greater when a regional criterion is followed, than when the division is made by commodity classifications. The reason, of course, is that many commodities from a given region will have extremely close substitutes (up to and including the very same products) produced in and imported from the rest of the world.

What this means is: a) that clearly, the elasticity of demand for imports relevant to case iv) will be significantly higher than that associated with case iii) (full import liberalization) and a fortiori higher than that appropriate for case ii) (real exchange rate movement); and b) that the main reason why this is so, is the presumption of greater substitutability between imports from sources within the region and those of similar characteristics from outside the region. This in turn means that a program of regional liberalization will be virtually certain to have a higher elasticity than a program of liberalization by commodity class, but at the same time this extra elasticity will carry with it a significantly higher degree of trade diversion. Thus a reasonable case might be made for exploring elasticities of -2, -4, and -6, in place of the triad -1, -2, -3 pursued in Table 2. However, if I were to make that exploration, I would want to raise significantly the fraction f_d representing the relative share of trade diversion in the substitution process. My instinct here would be to consider a range like 0.5, 0.6, 0.7 for f_d . But this would automatically put us in the region where the costs of trade diversion always either equaled ($f_d = 0.5$) or actually outweighed ($f_d > 0.5$) the direct benefits from the liberalization process. My main

point -- that the net benefits of regional liberalization when measured by the criteria of standard applied welfare economics, are likely to be small, if indeed they are positive -- would have to be changed to say that such benefits are very unlikely to be positive. Hence in the end my choice of the elasticity range -1, -2, -3 and its companion range of 0.5, 0.4, 0.33 for f_d can be viewed as an effort on my part to “lean over backwards” to give as fair as possible a chance for the traditional analysis to show a positive net benefit from a regional free trade area.

TABLE 1**How [DEfficiency Costs/DRevenue] Varies****Under Different Tax Rates and Elasticities**

	Tax Rate (τ_i^0)			
Elasticity of Demand (η_i) Or “Response” (σ_i) ^a	0.1	0.2	0.3	0.4
-0.25	.023	.043	.061	.077
-0.50	.048	.091	.130	.167
-0.75	.073	.143	.209	.273
-1.00	.100	.200	.300	.400

^aThe elasticity of demand (η_i) is the relevant measure when the elasticity of supply (ε_i) is infinite. When this is not the case the relevant elasticity of “response” is σ_i . This measures the percentage change in equilibrium quantity that is generated by a rise in the tax rate equal to one percent of the pre-existing market price. This elasticity of response is given by $\sigma_i = \eta_i \varepsilon_i / (\varepsilon_i - \eta_i)$. It is easily verified that when $\varepsilon_i = \infty$, $\sigma_i = \eta_i$.

TABLE 2

FREE TRADE AREA BENEFITS AND COSTS(as a fraction of direct revenue loss $M_u^0 \tau_u^0$)Elasticity of Response of Imports From Partners = -1

λ	f_d	Direct Benefit (+) ($-1/2 \eta_{uu}\tau$)	Indirect Loss (-) Of Revenues = Cost of Trade Diversion	Efficiency Cost (-) Of Replacing Lost Revenue	Net Benefits (+) Or Costs (-)
(1)	(2)	(3)	(4)	(5)	(6)
0.1	0.5	0.5τ	-0.5τ	$-0.1 - .05\tau$	$-0.1 - .05\tau$
0.1	0.4	0.5τ	-0.4τ	$-0.1 - .04\tau$	$-0.1 + .06\tau$
0.1	0.33	0.5τ	-0.33τ	$-0.1 - .033\tau$	$-0.1 + .133\tau$
0.2	0.5	0.5τ	-0.5τ	$-0.2 - .10\tau$	$-0.2 - .10\tau$
0.2	0.4	0.5τ	0.4τ	$-0.2 - .08\tau$	$-0.2 + .02\tau$
0.2	0.33	0.5τ	-0.33τ	$-0.2 - .066\tau$	$-0.2 + .10\tau$

Elasticity of Response of Imports From Partners = -2

0.1	0.5	τ	$-\tau$	$-0.1 - .10\tau$	$-0.1 - .10\tau$
0.1	0.4	τ	-0.8τ	$-0.1 - .08\tau$	$-0.1 + .12\tau$
0.1	0.33	τ	-0.67τ	$-0.1 - .067\tau$	$-0.1 + .263\tau$
0.2	0.5	τ	$-\tau$	$-0.2 - .20\tau$	$-0.2 - .20\tau$
0.2	0.4	τ	-0.8τ	$-0.2 - .16\tau$	$-0.2 + .04\tau$
0.2	0.33	τ	-0.67τ	$-0.2 - .133\tau$	$-0.2 + 0.2\tau$

Elasticity of Response of Imports From Partners = -3

0.1	0.5	1.5τ	-1.5τ	$-0.1 - .15\tau$	$-0.1 - .15\tau$
0.1	0.4	1.5τ	-1.2τ	$-0.1 - .12\tau$	$-0.1 + .18\tau$
0.1	0.33	1.5τ	-1.0τ	$-0.1 - .10\tau$	$-0.1 + .40\tau$
0.2	0.5	1.5τ	-1.5τ	$-0.2 - .30\tau$	$-0.2 - .30\tau$
0.2	0.4	1.5τ	-1.2τ	$-0.2 - .24\tau$	$-0.2 + .06\tau$
0.2	0.33	1.5τ	-1.0τ	$-0.2 - .20\tau$	$-0.2 + .30\tau$

Table 2 (cont.)

Methodology Followed in Table 2

- i. We explore three alternative own-price elasticities (η_{uu}) of demand for imports from partner countries -- -1, -2, -3. Each of these corresponds to a panel in Table 2.
- ii. Within each panel we explore two alternative levels of the shadow price of government funds ($1+\lambda$). We alternatively set $\lambda = 0.1$ and $\lambda = 0.2$. Readers should recognize that these represent a “conservative” range for λ .
- iii. For each combination of η_{uu} and λ , we explore three values of f_d -- 0.5, 0.4 and 0.33. These correspond to trade diversion equals trade creation, trade creation equal to 1 1/2 times trade diversion, and trade creation equal to twice trade diversion.
- iv. The direct benefit of tariff reduction to partner countries is represented in column (3). This is the triangle of benefits shown in the top panel of Figures 1a, 1b, and 1c, and 3a, 3b, and 3c.
- v. The cost generated by trade diversion is shown in column (4). It is equal to twice the product of column (2) \times column (3). It is based on assumption, that the average tariff τ_u that is being eliminated and the average tariff τ_d applying to trade diversion are equal.
- vi. The efficiency cost of replacing lost revenue is equal to column (1) plus [column (1) times column (4)], all with a negative sign. It is shown in column (5).
- vii. Column (6) shows the overall net benefit (+) or cost (-). It is the sum of columns (3), (4) and (5).

Assumptions

- a. Average initial tariff τ_u^0 in imports from partners is equal to the average tariff τ_d affecting trade distortions from non-partners. Both are represented by τ in the table.
- b. f_d is the fraction of the total change ΔM_u in imports from partner countries that is represented by trade diversion ($-\Delta M_d$). In elasticity terms it is equal to $-\eta_{ou}/\eta_{uu}$ -- i.e., the cross-elasticity of demand for imports from other countries w.r.t the price of imports from partner countries, divided by the own-price elasticity of demand for imports from partner countries.