



INTER-AMERICAN DEVELOPMENT BANK
Integration and Regional Programs Department
Regional Operations Department 1

Deepening Integration of MERCOSUR: Dealing with Disparities

Regional disparities in regional blocs: theory and policy

Anthony J. Venables

CONTENTS

I. INTRODUCTION	1
II. FACTOR PRICES AND FACTOR MOBILITY	2
III. WHOSE TRADE IS DIVERTED?.....	6
IV. MARKET ACCESS AND INDUSTRIAL LOCATION.....	8
V. CLUSTERING AND AGGLOMERATION	10
A. Clustering and dispersion forces.....	11
B. Outcomes	12
C. Agglomeration and regional integration:.....	13
VI. CONCLUSIONS.....	16
FIGURES.....	17
BIBLIOGRAPHY.....	22

The opinions expressed herein are those of the authors and do not necessarily reflect the official position of the IDB or its member countries.

REGIONAL DISPARITIES IN REGIONAL BLOCS: THEORY AND POLICY

Anthony J. Venables*

I. INTRODUCTION

A longstanding concern of policy makers in regional blocs is that regional integration might be associated with widening disparities between countries or regions in the bloc. In Europe, Italian concerns about the threat integration posed to the Mezzogiorno led to the establishment of the European Investment Bank in the 1950s with an obligation to contribute to regional development. A regional directorate was added to the European Commission in 1974, and regional policy now amounts to over one third of European Union spending. The historical record in the EU is one of considerable convergence of per capita income levels between countries, although the record is patchy (Greek performance compared to Irish). In many countries within-country regional disparities have widened at the same time as between-country disparities have narrowed. Elsewhere in the world, a number of developing and middle-income regional blocs have seen strains develop because of the perception that gains are accruing to one region rather than – or even at the expense of – others. Mercosur is now alert to these possibilities.

The objectives of this paper are to explore the forces that may create disparities within a regional bloc and to assess their implications for policy. The thesis we will argue is that it is quite possible for regional integration to create disparities. Indeed, we expect it to encourage differences in the economic structures of countries, and should not be surprised if these are sometimes also associated with difference in factor prices and income levels. However, the policy response to such integration induced disparities is, loosely stated, more integration. This may sound contradictory, but is based on the idea that disparities are most likely to develop when some, but not all barriers to trade or factor mobility are reduced.

There is no existing unified theory, or evidence base, concerning the effect of regional integration on disparities between member countries and regions within these countries. Our method in this paper is to develop a number of arguments in a series of related models, and we focus on four main mechanisms. The first is that, even when all countries gain from trade, trade liberalisation will change factor prices, and may cause divergence of some of these prices. This in turn can induce factor mobility, such that a country or region may experience a decline in per capita income. The second mechanism arises from the logic of trade diversion. Preferential trade liberalisation brings with it the costs of trade diversion, and if these are unevenly distributed between member countries then it is possible for inequalities to increase, and even for some regions to suffer a real income decline. The third mechanism is to do with the location of firms in imperfectly competitive industries. Locations with good market access will tend to

* The author is from the London School of Economics and CEPR Paper prepared for the Inter-American Development Bank project on 'Deeper integration of Mercosur; dealing with disparities'. Thanks to participants in the project for helpful comments.

attract firms, and this can be a cause of disparity. Finally, we turn to cumulative causation mechanisms. From the early work of Perroux (1955) on ‘growth poles’ through to recent work on geographical economics, there has been the idea that regional integration might lead to concentration of activity that favours established centres at the expense of the periphery.

We organise the paper by reviewing these mechanisms in turn. Each mechanism may be more or less relevant to different contexts, and country experts will be able to judge their applicability to cases under study. The purpose of this paper is to set out the arguments, and thereby provide a toolkit that will facilitate this analysis. All of the mechanisms that we discuss suggest, in broad terms, a similar policy response; more trade liberalisation might be the cure for integration induced disparities. However, formulation of detailed country or region specific policies requires detailed analysis of the local situation and goes well beyond the scope of this paper.

II. FACTOR PRICES AND FACTOR MOBILITY

Deeply ingrained in the thinking of many economists is the idea that integration moves economies towards factor price equalisation, so the effect of trade is to reduce any initial differences in factor prices. This result holds under well known, but extremely restrictive conditions.¹ If these conditions are absent it is quite possible that trade liberalization disequalises factor prices. Indeed, there is a powerful argument suggesting that factor price disequalisation is quite likely. This is that, in a closed economy, the costs of an inefficiency in one sector are shared across the entire economy, because they are reflected in goods prices. However, once goods prices are determined in world markets, the inefficiency falls entirely on factors specific to the inefficient sector, so trade depresses returns to these factors. If trade triggers outflows of these factors, then it may reduce the size of the economy concerned.

The relationship between trade and factor prices is sometimes stated in terms of the substitutability or complementarity of goods trade and factor mobility. If trade moves countries towards factor price equalisation then the relationship is one of substitutability. In this case trade liberalisation reduces the incentives for factors to move internationally (and symmetrically, factor mobility would reduce the volume of trade²). Complementarity arises when trade increases international difference in prices of mobile factors, thereby increasing incentives for factor movements.

Under what circumstances might there be a complementary relationship between trade and factor mobility such that economic integration causes flight of the mobile factor? We will think of this as capital flight, covering both physical and human capital, and will see that it can occur for a country that has comparative disadvantage in the capital-intensive sector. In this case trade

¹ These include the requirement that countries have identical technologies, and that there are at least as many traded activities (goods or mobile factors) as there are immobile factors.

² Markusen (1983) poses the question of complementary/ substitutability in terms of the effects of factor movement on trade flows. There is an exact mathematical symmetry between the effect of goods prices on factor prices and the effect of factor movements on trade.

depresses the return of capital so (in the absence of factor price equalisation) may make it lower than in trading partners. We explore two possible reasons for this comparative disadvantage, the first based on endowment differences between trading countries, and the other based on technical efficiency differences.

We develop both these cases using the simplest possible specific factors model. There are two goods, manufacturing (M) and agriculture (A). Manufacturing uses capital (K) and labour (L), while agriculture uses land (G) and labour; this is the usual specific factors structure of one factor (labour) mobile between sectors and the other factors sector specific. For simplicity we will look at integration between a large and a small economy. The size difference is simply a matter of convenience allowing us to fix prices (there are no scale effects present) and we label the small economy, on which we focus, country 1. Furthermore, we will assume that this economy has Cobb-Douglas preferences and technologies and is symmetric, so the share of capital in manufacturing is the same as the share of land in agriculture. These assumptions are not necessary to the argument, but facilitate making the point.

The first case rests on country 1 being relatively labour abundant. That is, we assume that the relative endowments of the three factors are the same in both countries, except that country 1 has more land. What is the impact of integration on this land abundant economy? In an initial situation in which trade barriers are large enough for there to be no trade economy 1 (small and land abundant) has a somewhat lower price of agriculture relative to manufactures, and a lower land price than does the other economy. Real wages and returns to capital are higher, as the benefits of the additional land (relative to other factors) are passed on to the other factors.

As trade barriers fall there is an increase in the relative price of agriculture. Country 1 becomes an exporter of agriculture and there is labour reallocation from manufacturing to agriculture. As this happens so the wage increases and the return to capital therefore falls. Furthermore, it must be the case that the return to capital now falls below that in the other country.³ Factor price equalisation does not take place, because there are more non-traded factors of production than there are traded goods, and the economy's land abundance translates into a relatively high wage, this disadvantaging owners of capital.

³ The price of land remains relatively low in the land abundant economy; since the price of agricultural output is the same in both countries and equal to average costs, the wage in the land abundant country must be relatively high. Manufacturing is competitive at this wage only if the return to capital is low.

TABLE 1: FACTOR PRICE DISEQUALISATION AND CAPITAL FLIGHT

	Real wage	Real return on land	Real return on capital	Capital stock	Real domestic income
<i>10% more land endowment</i>					
Autarky	1.0160	0.9236	1.0160	1	1.0160
Trade	1.0164	0.9680	0.9680	1	1.0164
Trade and capital mobility	1	1	1	0.9	1
<i>10% lower efficiency in manufacturing</i>					
Autarky	0.9487	0.9487	0.9487	1	0.9487
Trade	0.9526	1.1021	0.8032	1	0.9526
Trade and capital mobility	0.8539	1.3722	1	0.336	0.8539

Symmetric Cobb-Douglas economy, labour share each sector = 2/3. All values expressed relative to a situation in which country 1 has the same technology and relative factor endowments as the partner country.

What if we now allow capital mobility? Capital flows out of the country, this reducing the wage and raising the returns to the specific factors. The tradability of capital gives factor price equalisation, but only by means of capital outflow. Illustrative numbers are given in the upper half of table 1, for the symmetric Cobb-Douglas economy in which the share of labour in each sector is 2/3rds. All numbers are expressed relative to the case where economy 1 differs from the large economy only by a scale factor (and therefore has the same prices). We see that under autarky this land abundant country has a lower price of land but higher prices of other factors and higher real domestic income. Free international trade (without capital mobility) raises the return on land but reduces the return on capital. Allowing capital mobility achieves factor price equalisation, but this is achieved by capital outflow. The economy becomes somewhat smaller, and the wage somewhat lower than in the initial autarkic situation. This is reminiscent of the Dutch disease literature. Land is a source of initial advantage and there are gains from trade, as there have to be in this perfectly competitive economy. However, the combination of trade and factor mobility gives a ‘deindustrialisation’, reducing domestic income below its initial level. Of course, if national residents are owners of the capital that is now employed abroad, then national income is greater than domestic. Alternatively, the mobile factor may be human capital, so the loss of capital is associated with a loss of skilled labour.

The potential effect of trade in disequalising factor prices can be seen even more clearly if

country 1's comparative advantage in agriculture derives not from a large land endowment, but from a low level of technical efficiency in manufacturing. Under autarky this inefficiency is divided between all factors. The price of manufacturing is raised and, in our symmetric example, all factors experience an identical reduction in real income, as illustrated in the lower part of table 1. Things are dramatically different once trade occurs. As before, the relative price of manufacturing falls and labour moves out of the sector, this reducing the real return to capital while raising the returns to land and labour. In the example a 10% efficiency difference causes the return on capital to fall 20% below that in the partner country while, prior to trade, it was only 5% below. Adding international capital mobility now causes a large outflow – in the example $2/3^{\text{rd}}$ of country 1's capital stock has to leave to give rate of return equal to that in the partner country. Both the wage rate and total income go from being just 5% below those in the partner country to being 15% below.

This second example demonstrates vividly how trade concentrates the cost of a technical inefficiency that was previously dispersed throughout the economy. Once the cost is concentrated in a mobile factor then the factor may relocate, its outflow reducing wages and income. Other mechanisms could produce the same effect. For example, a subsidy to the agricultural sector would be similar, tending to raise land rents and wages and reduce the return to capital. The inefficiency in manufacturing could arise from a geographical disadvantage if, for example, manufacturing requires inputs imported from the rest of the world, which are more expensive in country 1 than in the partner country.

The point of these examples is to show how it is quite possible, even in a 'perfect' economy with no distortions of any sort, that trade liberalization can change factor prices in a way that causes capital outflows and income reductions. What about policy responses to this sort of trade induced disparity? There are no market failures in this model, so from the point of view of the regional bloc as a whole – country 1 and its partner – any attempt to restrict trade or factor mobility would be a source of inefficiency and real income loss. A better policy comes from thinking about the logic of factor price equalisation, which requires that there are at least as many traded goods as there are immobile factors. In the examples outlined above trade by itself does not secure factor price equalisation because not enough things are traded – there are three non-traded factors and only two traded goods. Trade causes sectors with comparative disadvantage to contract, and the factors released encounter diminishing returns as they are redeployed in other sectors. Re-employment of factors is less likely to run into diminishing marginal returns the more alternative tradable activities there are. The policy response is therefore to make more goods tradable prior to liberalising mobility of factors.

We can see how this works by supposing that manufacturing production could be fragmented into two elements, such as components and assembly, each of them separately tradable and with different factor intensities. A low price of capital (human or physical) would then attract the capital-intensive production fragment, bidding up the price of capital. Making more things tradable would therefore have both a direct efficiency enhancing effect (the usual gains from trade) and would also reduce factor price disparities and the incentives for factors to move.

In practical terms, the idea points to the importance of diversifying export sectors prior to liberalising factor mobility. The experience of some European countries is instructive. For example, Ireland suffered significant outflows of skilled labour that could not be employed in a manufacturing sector that was weak compared to its customs union trading partners (first just the UK, then the EU as a whole). Only after strengthening its manufacturing sector in the 1990s (mainly through FDI inflows) did the brain drain go into reverse.

In summary then, this section shows that even without there being any market failures, it is possible that trade can disequalise some factor prices. If regional integration also increases the mobility of factors then countries or regions may experience factor outflow with negative consequences for incomes of some of the remaining factors. Whether or not this occurs depends on the patterns of factor endowments, technical differences, and goods tradability. Should it occur, the policy response is to widen the range of products in which trade is liberalised and trade flows occur, in order to try and drive the region towards factor price equalisation.

III. WHOSE TRADE IS DIVERTED?

Our second argument hinges on an imperfection that is at the heart of preferential trade liberalisation – trade diversion. The point of departure for analysis of preferential trading arrangements is Viner's (1950) study of trade creation and trade diversion. His analysis shows that while a reduction in tariffs with all trading partners raises welfare, a preferential reduction affecting only partners in the regional integration agreement (RIA) may reduce welfare. The reason is that in addition to creating trade, such a tariff change will also tend to divert it, possibly causing import supply to switch from the lowest cost source to a higher cost partner country whose exports benefit from preferential market access.

In general, it will be the case that the costs of trade diversion are uneven across members of an RIA. In particular, in an RIA between developing or middle-income countries it is the poorer countries that bear the costs of diversion, so that initial income disparities are magnified. The argument is based on Venables (2003) and outlined below.

As a framework for thinking about this, suppose that there are three countries. One is large and represents the world average, and the other two (economies 1 and 2) are the countries that are considering RIA formation. We shall concentrate on country 1, and refer to country 2 as the partner. We assign country 1 a particular comparative advantage and look at a range of different potential partners. The horizontal axis of figure 1 gives all possible partner types, ranked according to their comparative advantage. At the right-hand end of this axis are high income >northern= partner countries, whose endowment (abundant in human and physical capital) gives them a comparative advantage in a composite of goods we shall call >good M=. At the left-hand end are low income >southern= countries with comparative advantage in another good, >good A=. Country 1 has comparative advantage fixed at point I, and the world average is at point R. Thus, if the partner is at point I the two countries in the RIA have the same comparative advantage. If the partner is in range N, then country 1 forms a RIA with a high income Northern

partner. If the partner is in range HS then country 1 forms a RIA with a higher income >southern= country; in the range LS the partner is lower income than country 1.

The vertical axis of figure 1a has on it (as a function of the partner country's comparative advantage) the change in country 1's net imports of good M from the rest of the world, from the partner, and in total. If country 1 forms a RIA with a partner in set N, it will experience a large increase in imports of good M from the partner, as indicated by the dashed line, and a fall in imports from the rest of the world, indicated by the lower solid line. Total imports increase (and so, correspondingly, do exports of A). There are unambiguous welfare gains to country 1 from this change in trade, since it is an increase in imports of good M from a country which has a comparative advantage in good M, relative to country 1 and relative to the rest of the world.

If the partner's comparative advantage is in set HS, then the qualitative change in the direction of trade is similar, as it is determined by the partner's comparative advantage relative to that of country 1. We see an increase in imports from the partner, a fall in imports from the rest of the world, and an increase in overall imports of M (and exports of A). However, the welfare economics is quite different. A partner in HS has a comparative advantage in good M relative to country 1, but a comparative *disadvantage* in this good relative to the world. Since it displaces imports from the rest of the world there is Vinerian trade diversion which reduces real income. Preferential treatment is causing country 1 to divert the sourcing of its imports from the rest of the world (the lowest cost source of supply) to country 2.

What if the partner is a country with comparative advantage in set LS? In this case country 1 has comparative advantage in good M, relative to partner country 2, and will increase its *exports* of good M to country 2, this showing up as a reduction in net imports. Since country 1's increased exports of good M to country 2 raises the price of good M in country 1, there is also an increased flow of imports of good M into country 1 from the rest of the world. This is welfare improving because country 1 still has an import tariff on these imports.

The welfare effects of these changes is illustrated on figure 1b, (giving full general equilibrium welfare effects including changes in trade in good A). The U-shaped solid line is the change in country 1's welfare and reflects the discussion above. There are gains from forming a RIA with a Northern country as country 1 expands its imports from a country with comparative advantage relative to country 1 *and* relative to the rest of the world. Turning to South-South agreements, country 1 is likely to gain if it is the >*intermediate*= country, but lose if it is the >*extreme*= one. Thus, if the partner is in range LS we see that country 1 gains as it increases imports of good A from the partner, which has world comparative advantage in this good. However, if country 1 is extreme so its partner (in HS) lies between it and the world average, then trade diversion occurs, bringing country 1 welfare loss. Intuitively, country 1 is vulnerable to trade diversion when its partner has comparative advantage between its own and that of the rest of the world.⁴

⁴ The other solid line on figure 2b gives the welfare change of the RIA as a whole. For South-South agreements one of the countries has to be extreme, experiencing diversion, and meaning that the RIA as a whole gains little. For North-South agreements (partner to the right of R) the changes in trade are in line with global comparative advantage and the aggregate gains are considerably larger.

The main point from this analysis is that it is always the extreme country that is prone to diversion. Thus, a South-South integration scheme tends to increase any existing income differentials between the two countries, as it is the poorer of the member countries that has trade diverted to a partner intermediate between it and the world average. The opposite is true for a North-North integration scheme; in a North-North scheme the extreme country is the one with the higher income, whose imports are diverted to come from the intermediate country.

It is also noteworthy that the country that does better in a South-South agreement (the intermediate country) has its production structure moving in the *opposite direction* from what would occur with external trade liberalization. This occurs as it exploits its comparative advantage with the poorer country not with the rest of the world. In terms of the static model outlined here this is not, of itself, damaging to real income. However, in a dynamic setting in which regional integration is a way of developing the capacity or the efficiency of export industries, perhaps prior to broader trade liberalization, this is clearly problematic. The wrong sectors are being developed.

The policy response to these problems is evident. External trade liberalisation will reduce the potential for trade diversion, the root of the problem. The dashed line in figure 1b gives the effect on country 1's income of forming a RIA *and* implementing unilateral free trade with the rest of the world. We see that this is beneficial, as it must be for small countries in a perfectly competitive setting. Once again, the problem is with partial trade liberalisation, in this context preferential, rather than MFN reductions in trade barriers. Wider liberalisation removes trade diversion and the disparities that it creates.

IV. MARKET ACCESS AND INDUSTRIAL LOCATION

We now turn from models that are grounded in traditional trade theory, to those based on firms and increasing returns to scale. The location decisions of these firms are guided both by factor costs and by their proximity to consumers – ‘market access’. It turns out that trade liberalisation can have the effect of inducing relocation of some sectors and this in turn can create disparities in wages between countries. As in preceding sections, we outline the forces at work and discuss possible policy responses.

The basic structure is a model with two sectors, one perfectly competitive (often referred to as ‘agriculture’) and the other monopolistically competitive containing firms that produce with increasing returns to scale and set price in excess of marginal cost (‘manufacturing’). Firms engage in intra-industry trade, with each firm supplying all countries, although the presence of transport costs means that firms’ sales are skewed towards their home market. The standard workhorse model for analysing this is the Dixit-Stiglitz (1977) model of product differentiation and monopolistic competition, although other forms of oligopolistic interaction are possible. It

is well known that in such models firms have a bias towards locating in a region that has good market access. Thus, if two regions or countries are identical except that one is $k > 1$ times larger than the other, then (given transport costs between the regions) industrial production in the larger region will exceed that in the smaller by a factor greater than k . Furthermore, this fraction will vary with the level of trade costs.

To understand the basic logic, let us call the regions 1 and 2, with 2 being k times larger than 1. Could there be an equilibrium in which firms are located in proportion to the size of the regions, so 2 has k times more manufacturing firms than 1? If transport costs are prohibitively high the answer is yes; each market is supplied only by local firms, and the number of firms is proportional to the size of the market.⁵ As trade costs are reduced, two things happen. First, the country 1 market comes to be supplied by a large number of importers, while the country 2 market is only supplied by $1/k$ as many importers, this reducing the profitability of producers in 1. Second, each firm in 2 will pay transport costs on only a small part of their output (sales to the small country 1 market) while firms in 1 will pay transport costs on a larger part of their output (sales to the larger country 2 market). Both arguments suggest that firms in 2 become relatively more profitable, implying that in equilibrium with free entry the number of firms in 2 must exceed the number in 1 by a factor greater than k . The large region therefore has a disproportionately large share of manufacturing production, and is a net exporter of manufactures and importer of agriculture.

Notice several further points about this argument. First, it holds only if transport costs lie strictly between zero and a prohibitive level. If transport costs are prohibitive no firms ship any exports; autarky production has to equal local consumption and the location of industry is in proportion to the size of the regions. Conversely, if transport costs (and all other frictions) are zero, then obviously the argument collapses, as firms in all locations have equally good access to all markets. The argument shows that it is at intermediate levels of transport costs that manufacturing is pulled disproportionately into the large region.

What are the implications for factor prices? Unless factors are in perfectly elastic supply the changes in demand for factors for use in manufacturing tends to raise factor prices in the larger country and reduce them in country 1. This is illustrated in figure 2. The horizontal axis is the transport cost factor (a value of 1 corresponding to free trade, and 1.5 corresponding to transport costs equal to 50% of the value of output). The left-hand vertical axis of the figure is the share of manufacturing in the large region, and the right-hand vertical axis is the real wage in the large region relative to that in the small. For the example, the large region is assumed to be three times larger than the small, $k = 3$. Labour is the only factor used in manufacturing, while the other sector of the economy ('agriculture') uses labour and a specific factor.

Figure 2 shows very clearly both the pull of the larger market, and the non-monotonicity of this effect. The larger location gets share of manufacturing equal to its share of income, $(k/(1+k)) = 0.75$) at free trade and at autarky, and its share peaks at transport costs of around 15%. A large

⁵ Notice that this argument uses the Dixit-Stiglitz property that all firms are the same size in equilibrium.

manufacturing presence bids up wages in the larger region, and several effects underlie the relative real wage curve in the figure (the dashed curve). Real wages in country 2 are higher than those in country 1 under autarky, because of a variety effect – the large region has more varieties on offer. At intermediate transport costs the relative real wage in country 2 is increased further because of high labour demand created by the relocation of manufacturing. The magnitude of this effect depends on the elasticity of labour supply, being greater the less elastic is supply. Furthermore, if several primary factors are used in manufacturing this may well amplify some factor price disparities. For example, suppose that a fraction of manufacturing costs are met by a factor in perfectly elastic supply (capital) with the remainder being labour. In this case a similar divergence of firms' units costs would be associated with a much larger divergence of wages, now only a fraction of unit costs.

As transport costs fall below the turning point, the wage gap narrows for two reasons. First, the strength of the market access effect on labour demand declines. And second, as transport costs fall so the international differences in the consumer price index due to transport costs diminish. In the limit, of perfectly free trade, there is factor price equalisation.

Application of the insights from this basic model requires a much richer geography than is presented in the two location example. Consider, for example, the implications of one of the regions containing a port trading with the rest of the world. This region has good market access, but it also faces import competition. It will therefore tend to attract industry if comparative costs are such that the region is a net exporter of manufactures to the rest of the world. By contrast, if it is a net importer then local manufacturing will seek the natural protection of the region without the port. But the general point remains. The geography of market access is a force shaping industrial location and inter-industry wage differentials.

Finally, we turn to policy implications. As regional integration opens up the income gap between regions, what might policy do? One possibility is to freeze manufacturing, for example by the use of subsidies and taxes. Any self-financing scheme is inefficient, and will certainly reduce aggregate real income compared to the outcomes described above. This follows from the fact that equilibrium in the Dixit-Stiglitz model is efficient, given the constraint that firms in aggregate earn zero profits. However, the non-monotonicity that this model exhibits points to the obvious policy conclusion. Inequalities are greatest when there is *partial* integration of goods markets, and are reduced by full integration. Of course, the policy-maker does not know the exact level of trade costs at which the turning point in the wage schedule is encountered. Furthermore, trade costs themselves are a complex mixture of barriers, some of them artificial – tariffs and border controls – and other real barriers, requiring costly investments to bring down. Nevertheless, the fundamental logic of these effects is that, while integration can create regional disparities, further integration can reduce these disparities.

V. CLUSTERING AND AGGLOMERATION

The most striking feature of the economic geography of Mercosur is the presence of massive

concentrations of economic activity, most prominently in the Sao Paulo region. To understand such centres requires analysis of the clustering and agglomeration forces that support them. In section 5.1 we outline these forces, and in section 5.2 return to the effects of regional integration. Reducing trade costs facilitates clustering, in so far as it makes it cheaper to supply dispersed consumers from a single cluster of activity. Yet at the same time, falls in trade costs might make it easier to move some activities out of clusters. Net effects are ambiguous, and we discuss possible outcomes and policy implications.

A. Clustering and dispersion forces

Proximity facilitates many sorts of economic interactions and creates benefits from clustering of economic activities. We have already seen one of the implications of proximity. The analysis of the preceding section showed how market access effects mean that firms will locate disproportionately towards large centres of demand. This market access mechanism is amplified by the fact that demand for manufactured goods comes not just from final consumers but also from intermediate demand. A location with a lot of firms has a high demand for intermediates, making it an attractive location for intermediate producers. And the presence of intermediate producers makes the location profitable for firms that use these intermediate goods, as they economise on transport costs on inputs. There is thus a positive feedback between location decisions of upstream and downstream firms, tending to draw both types of firms together in the same location – agglomeration. These forces are just the backwards (demand) and forward (cost) linkages that figured so prominently in an earlier generation of development economics (in particular the writings of Hirschman 1958 and Myrdal 1957). Notice however that they constitute a force for clustering only if they are combined with increasing returns to scale, without which upstream and downstream firms could be broken into many small plants to meet local demand.

In addition to linkages between firms, a number of other arguments have been put forward to suggest the value of locating in a dense cluster of activity (see Duranton and Puga 2004 for a recent survey). One important set of arguments is to do with the efficiency advantages of thick labour markets. Pools of specialist workers and of the firms that use these skills benefit both from better matching of aptitudes with requirements, and from risk sharing if there are firm or worker specific fluctuations in demand or supply. Furthermore, incentives to acquire skills are greater if the skills are sought by several firms, so the worker is less likely to be subject to the monopsony power of a single employer. Labour turnover is one – but not the only mechanism – through which firms in a dense cluster of activity can benefit from the skills and knowledge of other firms. There is considerable evidence of productivity spillovers between firms, as they are able to learn about and imitate the practises of other firms in the industry. Silicon valley provides an example where knowledge exchange – formal and informal – is quite widespread. The knowledge may be about production methods, marketing skills, or simply knowledge about the location itself. Thus, multinational firms tend to cluster in particular locations, partly because one firm, observing the success (or failure) of another, learns about the quality of the business environment in the location. Hausman and Rodrik (2002) argue that very narrow patterns of specialisation in developing countries (for example, specialization in soccer ball

production) arise as producers learn about the efficiency of a particular location for producing a particular good.

Agglomeration forces can operate across more or less broad ranges of activity. For example, the key externalities and linkages might occur between firms in a particular industry or between firms that engage in a narrow field of R&D. Sectorally narrow effects of this type are sometimes referred to as ‘localisation’ economies. Alternatively they might operate at a much broader level – through aggregate demand as a whole, the development of general labor skills, or the provision of basic business infrastructure and inputs used by wide sectors of the economy. It is also argued that they may stem not from specialisation but from diversity in the activities of a location. Sectorally broad effects of this type are referred to ‘urbanisation’ economies, although we note that they can also occur over a wider spatial area than a single city – for example within a ‘city-region’ or a country.

Pulling in the opposite direction are forces for dispersion. These are of essentially three types. One is the supply of immobile factors, the prices of which will be bid up in centres of activity, encouraging firms to move to lower factor cost locations. The second is the extent of the market; geographically dispersed demand and costs involved in shipping goods create a force for dispersion. Thus, if labor is dispersed it encourages a dispersed location of firms for both supply and demand reasons. Finally, there may be other costs associated with concentration of activities, for example costs of commuting or of congestion.

B. Outcomes

Outcomes are determined by the balance between agglomeration and dispersion forces. The key point is that if agglomeration forces are strong enough, then locations that are identical in underlying characteristics may nevertheless have quite different equilibrium outcomes. Dispersion of economic activity equally between locations is not a stable equilibrium, as cumulative causation forces will concentrate activity in a subset of locations. There are typically multiple equilibria – the cluster could occur in any one of a number of locations, and its actual location is determined by history.

Table 2 below lists some of the possible forms that agglomeration might take. If dispersion forces are quite weak, in particular if labour is mobile, then cities or city-regions will develop. This is the sort of outcome predicted in the urban economics literature, and also by Krugman’s (1991) core-periphery model. While the spatial pattern of activity is extremely concentrated, real factor price differences are likely to be small. Labour migration narrows real wage differences, although prices of other immobile factors (eg land prices) may vary widely.

In the international context it is generally more appropriate to think of labour as being relatively immobile. While this is a force for dispersion of activity, it may nevertheless be the case that particular sectors (narrow or broad) cluster together. If the agglomeration economies are within fairly narrowly defined sectors, then this is consistent with small differences in factor prices; one region has engineering, another financial services, and so on. But if linkages are strong between

as well as within sectors then agglomeration will be associated with large spatial differences in labour demand and in equilibrium wages. Thus, agglomeration forces lead not just to spatial disparities in the concentration of activity, but also spatial disparities in real incomes, as in Krugman and Venables (1995).

TABLE 2: AGGLOMERATION: OUTCOMES

		Dispersion forces	
		Weak (eg factors mobile)	Strong (eg factors immobile)
Agglomeration forces	Narrow, 'localisation'	Specialized cities (Henderson 1974)	Industrial clusters vs. comparative advantage (Krugman and Venables 1996)
	Broad, 'urbanisation'	City formation (Fujita 1988) Core-periphery (Krugman 1991)	International income inequalities (Krugman and Venables 1995)

C. Agglomeration and regional integration:

How does regional integration change the balance of forces involved in clustering, and how does it change outcomes? Reductions in trade barriers unambiguously weaken one of the dispersion forces; lower trade costs mean that it is easier to supply dispersed consumers from a single location. However, it may also weaken some of the agglomeration forces, so its effect on industrial location and spatial disparities is ambiguous. To explore this we will outline some of the ideas developed in Fujita et al (1999). Throughout, the focus will be on the international dimension, maintaining the assumption of labour immobility.

Linkages; concentration and deconcentration

One of the few models of agglomeration that explicitly addresses the effects of changes in trade barriers between areas is that of Krugman and Venables (1995), see also Fujita et al (1999). The model has similar structure to the model of industrial location in section 4, containing a perfectly competitive sector (agriculture, using labour and a specific factor) and an imperfectly competitive sector, manufacturing (using labour alone). However, it adds linkages between firms in manufacturing, by assuming that each variety of product is used both as a final good and as an intermediate, and that each firm uses all varieties of intermediates. As we suggested above, such linkages can create agglomeration forces, as firms seek to locate close to other firms that use their output (demand linkage) and that produce their inputs (cost linkage).

The cleanest analysis of clustering is undertaken by assuming that there are two countries, both ex ante identical. However, clustering may cause manufacturing to locate asymmetrically, with a cluster in one of the countries and a smaller (possibly zero) level of activity in the other country. Figures 3a and 3b illustrate possible outcomes. The horizontal axis is the level of trade

costs, and the vertical in 3a is the share of manufacturing in each country. Since the two countries are symmetric, there is always an equilibrium in which manufacturing is divided 50 – 50 between the two locations. However, for an intermediate range of trade costs this symmetric equilibrium is unstable (dashed line). There is a new stable equilibrium in which shares of manufacturing employment in each country are traced out by the solid lines. Thus, as illustrated, most of manufacturing clusters in country 2. It is not profitable for a firm to relocate from country 2 to country 1, as any savings in labour costs would be offset by the costs of being outside the cluster – the firm would be distant from its main market and from its input suppliers.

The corresponding real wages are given in figure 3b. Reducing trade costs has a real benefit, hence the slope of these lines. But when clustering occurs, the country with the cluster has higher real wages, and the other country experiences real income decline. As before, there are two forces underlying this. Most important is the higher demand for labour in country 2 than in country 1, supplemented by the fact that the cost of living is lower in 1, since few manufactures have to be imported and bear trade costs.

For present purposes, the main point to note is that integration can create disparities – and further integration can eliminate them. What we observe is: when trade barriers are high each location has manufacturing to supply local consumers. When trade barriers are very low there is ‘death of distance’; a symmetric location of industry is driven by the equal factor endowments of the two countries. When trade barriers take ‘intermediate’ values the clustering forces are most powerful.⁶

Localisation and sectoral specialisation

In the previous example the only force supporting the agglomeration was linkages within manufacturing as a whole, and reducing trade barriers to zero removed the agglomeration force. As an alternative case, suppose that there are also localisation economies within industrial sectors arising, for example, because of thick labour market effects. We can easily speculate about the location of manufacturing as economies with this structure reduce trade barriers. At high trade costs all industries are present in all countries. At intermediate trade costs there is a tendency for all industries to cluster together. However, as this raises wages so it may become profitable for firms in some sectors to leave the agglomeration and form sectorally specialised clusters. Thus, if there are many locations and many sectors, we might expect to see deconcentration from a single centre to a number of sectorally specialised centres of activity. As this occurs and manufacturing employment becomes more widely dispersed, so wage disparities are reduced.

To the extent that inequalities are driven by agglomeration forces, what policy implications follow? The first point is that further integration, reducing trade barriers or transport costs, will reduce the cost of being outside an existing agglomeration. Wages typically decline as a

⁶ If the supply of labour to manufacturing was perfectly elastic then any degree of intermediate goods linkage within manufacturing, however small, would create agglomeration. With less than perfectly elastic labour supply, agglomeration forces must be sufficiently large relative to the response of wages to industrial relocation if agglomeration is to occur.

function of the distance away from centres of economic activity.⁷ The slope of these ‘wage gradients’ can be flattened by better integrating the hinterland with the centre of agglomeration. The second point follows directly from the analysis above. Clustering is most likely to occur at intermediate levels of trade costs (eg figure 3). The U-shaped nature of these relationships means that, once again, it is possible that further integration reduces or eliminates the disparities created by integration.

An important difference between this case and that of section 4 is the fact that agglomeration mechanisms are all sources of positive externality – either technological or pecuniary. This means that the real income gains from policy changes that facilitate clustering are typically quite large – there are real efficiency gains to letting an agglomeration form, even if the gains fall unequally across regions. But it also means that there is a potential for using policy to try to correct the agglomeration externalities.

Urbanisation provides a good example of these market failures. While there are efficiency gains to large cities, it may well be the case that developing country cities become even larger than is socially efficient. The reason is that the incentives for a firm to move to a new location are too low, essentially for two reasons. One is that firms fail to internalise the externalities that they create. And the other is that, in moving out of an established centre to a smaller city, they fail to anticipate (or they discount too heavily) the benefits of externalities that they will receive as the new city grows (Henderson and Venables 2004). Does this mean that active policy action should be followed? Design of policy to address this coordination failure is difficult – Pigovian subsidies to every source of externality are neither feasible nor desirable, and attempts to manage the creation of new cities have not generally been successful. However, the policy message is that institutions need to be designed that reduce some of the distortions that hold firms in existing centres (eg regulatory regimes that put a premium on access to government). Policy also needs to facilitate decentralisation by infrastructure investments and by decentralised institutions that can initiate and fund policies to overcome the coordination failure.⁸

⁷ See Redding and Venables (2004) for estimates of these wage gradients from international data.

⁸ For further discussion see Overman and Venables (2005).

VI. CONCLUSIONS

We have argued that regional integration can create factor price disparities between member countries of a trading bloc. The gains from economic integration come from allowing activities to relocate in line with comparative advantage or to achieve economies of scale and agglomeration. However, relocation changes factor demands in each country and can thereby open up wage gaps.

These disparities may well be worst at partial levels of integration. To yield any economic benefits integration requires reallocation of factors between sectors. However, if many sectors are non-traded (or subject to high trade costs) then expanding employment in these sectors will depress the price of output and give diminishing value marginal returns to factors. The lower are trade costs and the more activities are traded, the less likely is the economy to run into these diminishing marginal returns. This is the logic behind factor price equalisation, but equalisation is attained only if there are very low trade barriers across a wide range of activities.

This suggests that a way to mitigate regional disparities is to broaden both the width and the depth of economic integration. This does not necessarily mean that there is no scope for measures designed specifically to address regional disparities, although such measures must always be assessed using the sort of general equilibrium context that we have developed in this paper. For example, transport improvements may well narrow regional disparities – but we should be aware that, like tariff reductions, it is possible that they widen disparities. Developing human capital is valuable – but if skilled labour is mobile, it may just be training workers for emigration. Facilitating the growth new clusters of activity and urban centres is also important, particularly as we know that a range of market failures suggest that urban agglomerations in developing countries may be larger than socially efficient.

FIGURES

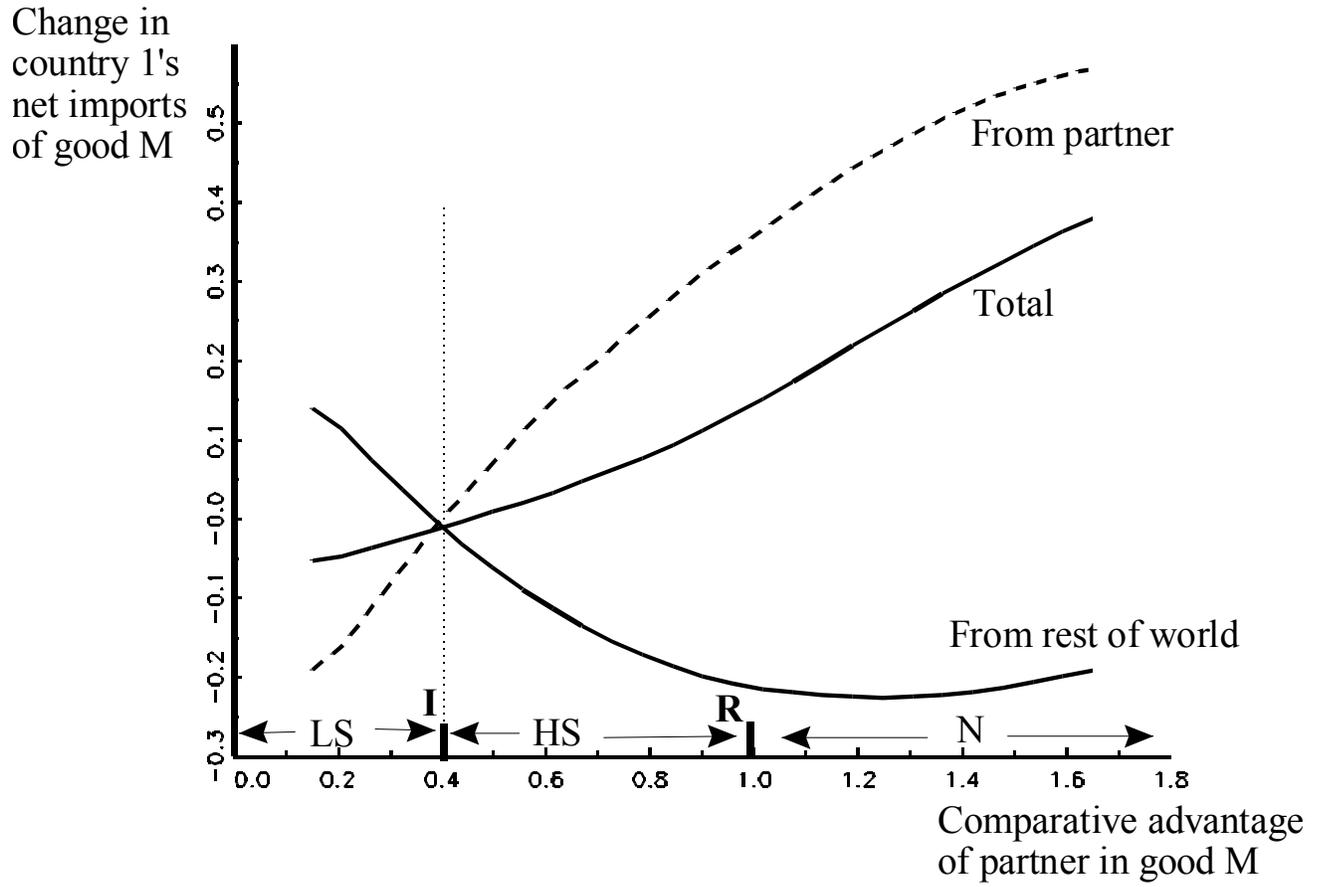


Figure 1a: Trade creation and diversion with different partners

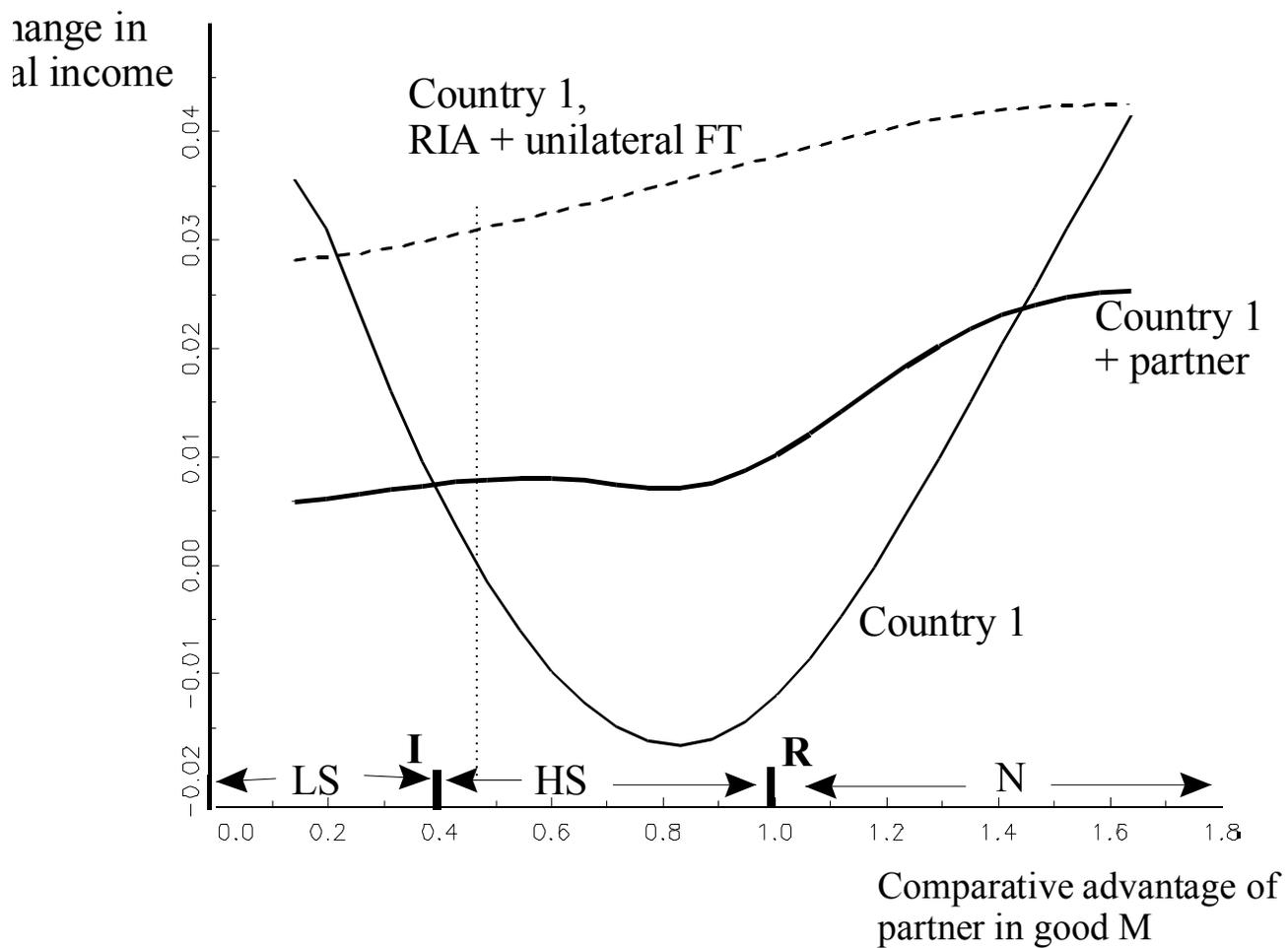


Figure 1b: Real income change

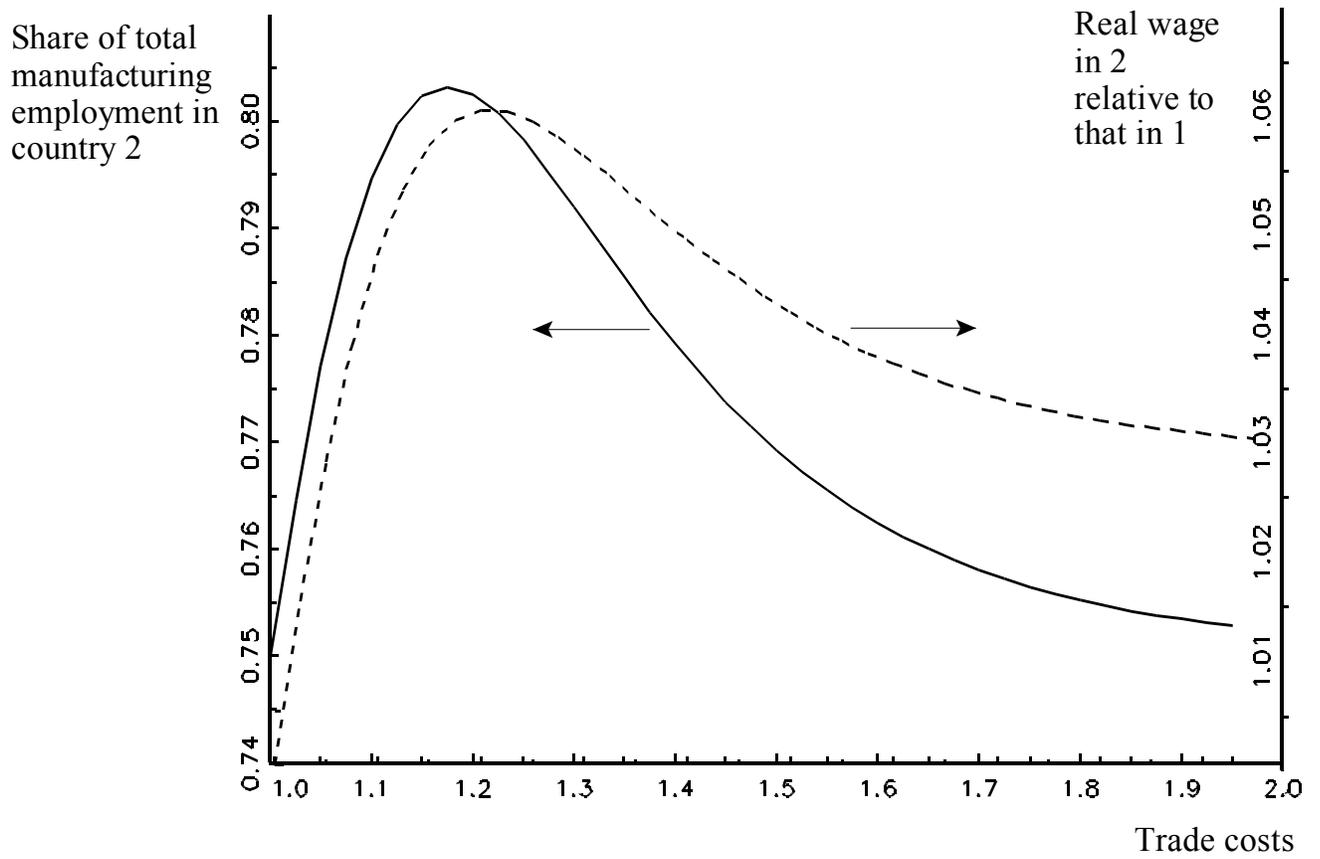


Figure 2: Relative manufacturing location and real wages

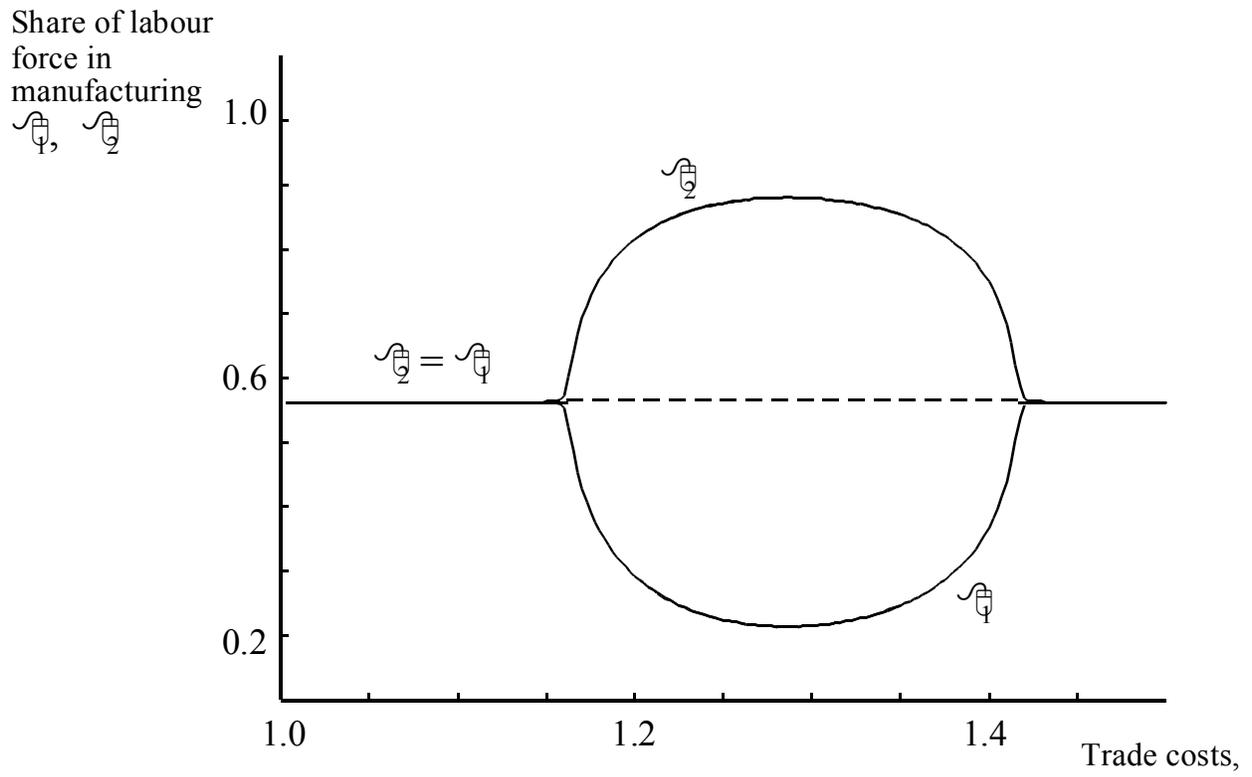


Figure 3a; Trade costs and a manufacturing cluster

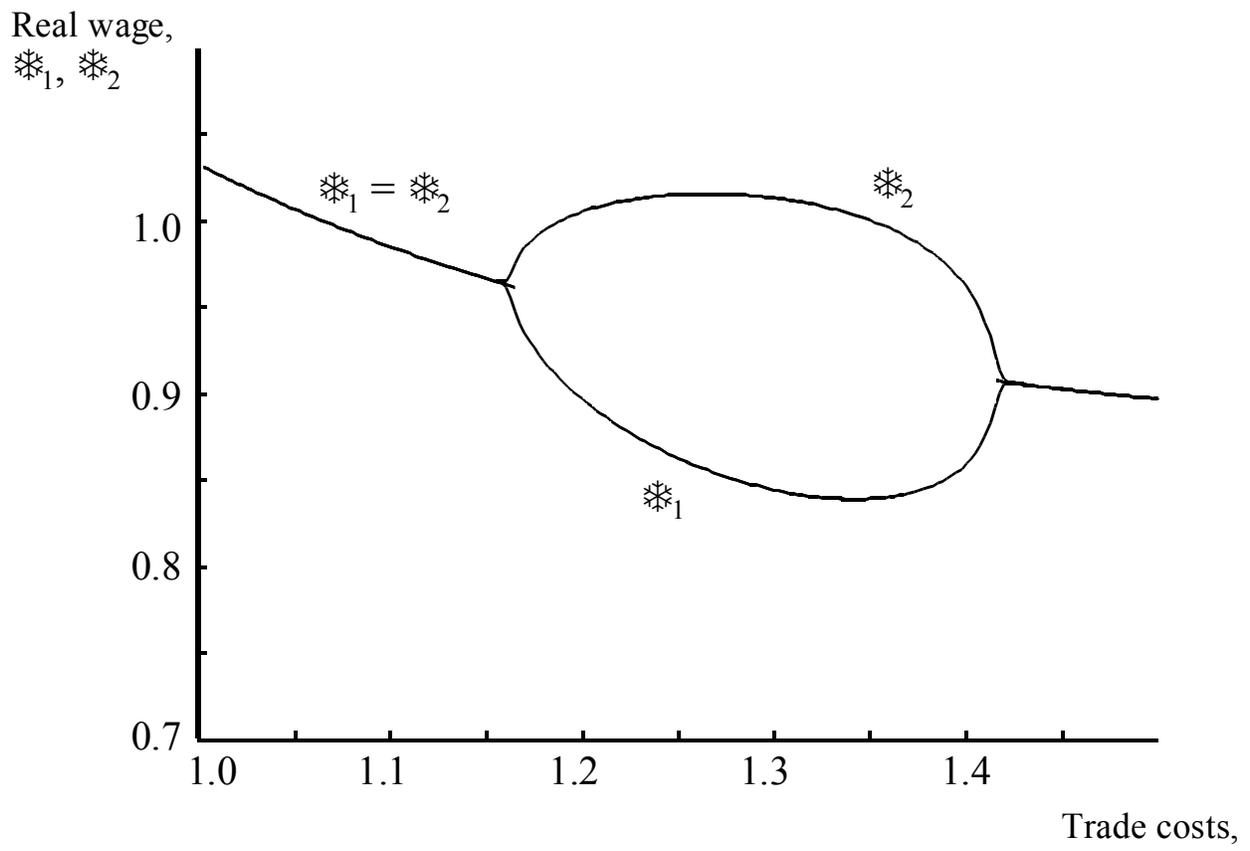


Figure 3b: Real wages

BIBLIOGRAPHY

- Baldwin, R. E. and A.J. Venables (1997), 'International Economic Integration', in G. Grossman and K. Rogoff (eds.) *Handbook of International Economics*, vol. 3. Amsterdam: North Holland.
- Dixit, A. and J. Stiglitz (1977), "Monopolistic Competition, and Optimum Product Diversity", *American Economic Review*, 67, 297-308.
- Duranton, G. and D. Puga (2004), "Micro-foundations of urban agglomeration economies" in J.F. Thisse and J. V. Henderson (eds) *Handbook of Urban and Regional Economics* (vol 4), North Holland.
- Fernandez, R. and Portes, (1998), 'Returns to Regionalism: An Analysis on Nontraditional Gains from Regional Trade Agreements'. *World Bank Economic Review* 12(2):197-220.
- Flores, R. (1997), 'The Gains from MERCOSUL: A General Equilibrium, Imperfect Competition Evaluation', *Journal of Policy Modeling*, vol. 19, pp. 1-18.
- Fujita, M., P. Krugman and A.J. Venables (1999), *The Spatial Economy: Cities, Regions, and International Trade*, MIT press, Cambridge MA.
- Hanson, G (1996b) 'Economic Integration, Intra-industry Trade, and Frontier Regions', *European Economic Review*, 40(3-5), 941-49.
- Hanson, G (1998) 'North American Economic Integration and Industry Location', *Oxford Review of Economic Policy*, 14(2), 30-44.
- Hausmann, R., and D. Rodrik. 2002. "Economic Development as Self-Discovery." Working Paper no. 8952. National Bureau of Economic Research, Cambridge, MA.
- Henderson, J. V. (1974), "The sizes and types of cities". *American Economic Review* 64(4): 640-656.
- Henderson, J.V., Z. Shalizi and A.J. Venables, (2001), 'Geography and development', *Journal of Economic Geography*, 1, 81-106.
- Henderson, J. V. and A. J. Venables (2004) "The dynamics of city formation: Finance and governance" Processed London School of Economics
- Krugman, P. R. (1991), "Increasing returns and economic geography". *Journal of Political Economy* 99(3):484-499
- Krugman, P. and A.J. Venables, (1995) "Globalization and the inequality of nations", with P. Krugman, *Quarterly Journal of Economics*, CX, 857-880.
- Krugman, P. and A.J. Venables, (1996) 'Integration, specialization and adjustment', *European Economic Review*, papers and proceedings, 40, 959-967.
- Lindert, P. and J. Williamson (2001) 'Does globalization make the world more unequal', mimeo Harvard University.
- Markusen, J.A. (1983) 'Factor movements and commodity trade as complements', *Journal of International Economics*, 14, 341-356.
- Overman, H.G and A.J. Venables (2005) 'Cities in the developing world' processed, LSE
- Perroux, F. (1955) 'Note sur la notion do pole de croissance', *Economie Applique*, 8.
- Puga, D and Venables, A.J. (1998) 'Trading arrangements and industrial development', *World Bank Economic Review*, 12, 221-249.
- Redding S.J. and A.J. Venables (2004), "Economic geography and international inequality", *Journal of International Economics*. 62, 53-82.

- Roberts, M. and J. Tybout, (eds.), (1996), *Industrial Evolution in Developing Countries: Micro Patterns of Turnover, Productivity, and Market Structure*, Oxford University Press
- Viner, J. (1950), *The Customs Union Issue*, Carnegie Endowment for International Peace.
- World Bank (2000), '*Trade Blocs*', World Bank Policy Research Report, Oxford University Press, Oxford
- Venables, A. J. (1996), "Equilibrium locations of vertically linked industries". *International Economic Review* 37(2):341–359
- Venables, A.J, (2003) "Winners and losers from regional integration agreements", *Economic Journal*, 113, 747-761.