Trade, Location, and Development: 
an overview of theory.*

Anthony J. Venables, London School of Economics and CEPR

1. Introduction
2. Comparative advantage, trade and income
   2.1 Endowments
   2.2 Location
   2.3 Production networks and vertical fragmentation.
   2.4 Increasing returns, market size, and supplier access
3. Changing the basis of comparative advantage
   3.1 Capital and foreign direct investment (FDI)
      3.1.1 Determinants of FDI
      3.1.2 Consequences of FDI
   3.2 Labour
      3.2.1 Skill acquisition
      3.2.2 Migration
   3.3 Spatial externalities and cumulative causation
      3.3.1 Agglomeration and North-South inequalities
      3.3.2 The spread of industry.
      3.3.3 External trade and internal geography
4. Policy issues
   4.1 Sectoral policies
   4.2 Infrastructure investment and the costs of distance
   4.3 Regional integration
5. Conclusions

1. Introduction

The theory of international trade has been transformed in recent decades, moving study away from the stylised world of perfect markets to a much richer recognition of imperfections of all sorts. Frictions to international trade and investment flows do not arise just because of tariff barriers, but are also due to geography, institutions, and information barriers. Markets are not all perfectly competitive but instead contain firms with market power, as well as imperfections in labour and capital markets. Production often involves increasing returns to scale, this requiring that focus be put on firms, rather than just the sectors of activity studied in competitive trade theory. Attention on firms has allowed the theory of foreign direct investment to be brought into the mainstream of trade theory. Dynamics and the processes of technology development and transfer have been analysed.

Importantly, these new elements have been studied in combination with each other and with the previous general equilibrium models of trade and specialization, and the combination has brought results that are much greater than the sum of the parts. For example, understanding international differentials in wage rates is an inherently general equilibrium subject. We have to know about labour endowments, and about how these change because of education or migration decisions. And we have to know about labour demands, which depend on decisions of firms. Firms’ profitability will be influenced not just by wage rates, but by considerations of access to markets and to supplies of other inputs. Combining these elements can give outcomes where small initial differences between countries translate into large differences in outcomes. With increasing returns to scale cumulative causation processes can operate, so both the economic structure and income levels of countries can follow divergent paths.

While these developments in international economics have been taking place, there has been relatively little work focussing on trade and development. This is in sharp contrast to the development economics literature of the 1960s and 70s where so much attention was directed to investigating the effects of alternative trade regimes. As will become clear in the course of this paper, more work is needed applying developments in trade theory (and in industrial organisation) to issues of economic development, as well as to drawing out policy conclusions. And above all, empirical work on trade, industry, and development is needed, establishing
exactly what shape industrial development patterns take in a world of new technologies, regional integration, and generally liberal trade regimes.

The remainder of this paper is organised into three main sections. The next section looks at the bases of comparative advantage, and the implications of these bases for production structure and the international distribution of income. This starts with traditional endowment based trade theory, but widens it out to show how geography and market size can themselves be a basis for comparative advantage.

Section 3 demonstrates how the bases of comparative advantage are themselves endogenous. There is brief review of the effects of the implications of capital mobility (through foreign direct investment), and of education and migration. More attention is devoted to ‘spatial externalities’. We use this term to capture the idea that the productivity of firms in a location is influenced by the activities of other firms in the same location, through knowledge spillovers or linkages of some kind. Pursuing this approach gives new ways of thinking about some trade and development issues, while also connecting modern economic analysis to older ideas of development economics such as those developed by Hirschman and Myrdal, as well as the ideas of business economists such as Porter.

Section 4 draws out some policy applications of the approaches discussed in previous sections, focussing on sectoral policies, infrastructure investments, and regional integration. Section 5 concludes. Throughout, the primary objective of the paper is to provide an exposition of developments in theory. Connections are made with empirical work at various points, although much more empirical work is needed on the issues addressed in this paper. Empirical research on the relationships between trade, industry and development had its heyday in the 1960s (eg Chenery et al), and new work is now badly needed.

2. Comparative advantage, trade and income
What determines the economic structure of an open economy, and how do openness and the ensuing structure influence factor prices and income levels? To answer these questions we proceed in two stages, looking first, in this section, at production structure and income given a
country’s comparative advantage (although we will define comparative advantage more broadly than would a standard approach). We then, in section 3, look at some of the ways in which the basis of countries’ comparative advantages can change, and at the implications of these changes.

The analytical approach of comparative advantage theory is to put a particular structure on the more general framework of competitive equilibrium theory. Competitive equilibrium theory is based on consumers and firms respectively maximising their utility and profits, and interacting only through perfectly competitive markets. The framework can be set up to contain many locations (countries) as well as many dates (as in intertemporal competitive equilibrium). Equilibrium determines the market clearing prices and quantities of all goods, and at equilibrium it will generally be the case that consumption and production are not equal for all goods in all locations, generating trade between locations. Trade theory puts a structure on this general framework by assuming that countries differ in certain well defined ways, and goods (or industries) also differ systematically. It then asks, what is the relationship between the characteristics of countries and the characteristics of the goods that it exports? We answer this question first for the traditional case, where the characteristics are factor endowments and intensities, and then show how the insights of this approach can be generalised.

2.1. Endowments:

Much analysis is based on the assumption that the only difference between countries is in their endowments of primary factors. This is based in turn on an assumption that some (or all) primary factors are internationally immobile, so countries’ endowments of these factors are, in the short run at least, exogenous. Correspondingly, commodities differ in their techniques of production, especially the intensities in which they use different primary factors (as well as differing in their physical and demand characteristics). The Heckscher-Ohlin-Samuelson theorem tells us that economies that are relatively labour abundant will export relatively labour intensive goods, and so on.

This interaction between country characteristics and industry characteristics is illustrated in figure 1. Suppose that there are many countries that differ only in the ratio in which they are endowed with labour and capital, \(L_P\), and one of the horizontal axes ranks countries according to
this ratio (over the set of countries, with individual countries indexed by subscript $i$). There are many industries, (indexed by superscript $k$), and these differ only in the labour intensity of production, which we denote $\gamma^k$. The industries are ranked along the other axis in the horizontal plane. The surface of the figure plots equilibrium output levels of each industry in each country (measured relative to the total size of each industry and country), so the height of the surface, at a particular value of $L_i$ and $\gamma^k$, gives the share of production of good $k$ in country $i$.

As expected, $L$-abundant countries have high production in industries in which the share of this factor is large (high $\gamma^k$) and low production in industries where it is low, giving a saddle shaped surface. The arrow marked R on the surface indicates how, in a particular industry, production varies with factor endowments; moving to more $L$-abundant countries increases output for products with high $\gamma$, and decreases it for products with low $\gamma$. The arrow marked H shows how, for a particular country, the structure of production depends on its factor endowment; an $L$-scarce economy has relatively high production in low $\gamma$ industries, while an $L$-abundant country has relatively low production in these industries. The effects illustrated by the R and H arrows can be thought of as generalizations of Rybczynski and Heckscher-Ohlin-Samuelson effects, showing how output of each industry depends on factor endowments, and how the structure of production of each country depends on factor intensities.

Several remarks are in order. First, we have constructed this example with countries and industries varying in just one dimension, $L$-abundance and $L$-intensity. In practise, they differ in many dimensions, each of which supports a surface like that in figure 1, although with countries and industries ordered differently (because the ranking of countries by land abundance may be quite different from the ranking by skilled labour abundance). However, the general principles outlined here remain applicable. Comparative advantage is determined by the interaction between an industry characteristic and a country characteristics. Furthermore, econometric analysis allows these surfaces to be recovered, (see Ellison and Glaeser 1999 and Midelfart-Knarvik, et al. 2001).

Second, discussion of ‘goods’, while standard, is perhaps misleading. In many activities the vertical production chain is now split, or ‘fragmented’, with different stages of production taking place in different locations according to their input intensities. Heckscher-Ohlin trade
theory applies at this level, so its predictions are that labour intensive stages of production should take place in labour abundant economies, etc. We return to this issue in discussion of production networks in section 3.1.

Finally, the model predicts that each country gains from trade (as must any trade model set in perfect market conditions). There is a tendency for international equalisation of prices of a particular factor, this implying that each country’s scarce factors suffer a real income loss (as they effectively come to compete with foreign factors) while abundant factors experiences real income gain (as they come to be exported, embodied in goods that are intensive users of the factors).

2.2. Location:
The insight that the pattern of production is determined by the interaction of country characteristics and industry characteristics, according to a surface qualitatively like that of figure 1, is far more general than sometimes appreciated. For example, country characteristics include a measure of institutional quality, and industry characteristics might include the effect of institutional quality on unit costs. Countries with good institutions would then have production skewed towards sectors where institutional quality was important, and so on. Where countries and industries differ in many dimensions the outcome is determined by combining these interactions.

We can think about the effects of geography on industrial structure using the same approach. Suppose that countries differ only in their distance from the world market, and commodities differ only in transport costs. Then the insight of von Thunen (1826) was that countries (or land area) close to markets would specialise in transport intensive commodities, and locations further out would produce goods with lower transport costs. This approach is developed in a trade theory framework in Venables and Limao (2001), and illustrated in figure 2a. The horizontal axis is the set of different countries, and point 0 is the ‘centre’, representing developed countries that export a good or composite of goods (good 0), while importing goods 1 and 2. Points \( z > 0 \) are countries lying at increasing distances from this centre, and potentially trading with it. The vertical axis gives the value of production of good 0, good 1, and good 2 in
each country. The structure of production is, as before, determined by the interaction between
country characteristics and commodity characteristics, these now being countries’ distances from
the centre, and goods’ transport intensity (transport intensity being high if transport costs are high
or the commodity is dependent on imported intermediates).

The figure illustrates an example in which endowments are the same in all locations, and
goods and factors are labelled such that good 1 has a higher transport intensity and higher labour
intensity than good 2. The following structure emerges. At locations close to the centre (low
values of \( z \)) there is a region, (zone I), in which countries specialise in the good with high
transport intensity, good 1. Since they export this good they import the other two goods, 0 and 2.
Moving further out (zone II) production of the less transport intensive good (good 2) commences,
and beyond some point countries become exporters of good 2. Further out again is zone III, in
which it is not profitable to export the high transport intensity good (good 1) and countries
become self sufficient in good 1, while continuing to export good 2 and import good 0. Thus,
between zones I and III the pattern of production and trade has reversed, even though factor
endowments are the same everywhere. Zone IV is one of import substitution – good 0 has
become so expensive that it is profitable to produce it locally. Within this zone each location
imports good 0 and exports good 2, but in smaller quantities as we move further out. Eventually,
in zone V, there is autarky.

Factor prices and real incomes corresponding to this example are given in figure 2b.
More remote locations have progressively lower real incomes as they forego the gains from trade.
The magnitude of this decline is determined by transport costs on final and intermediate goods,
and the share of intermediates in output. If imported intermediates account for 50% of gross
output and there are transport costs of just 10% on these and on exports of the final product, then
value added is reduced by a full 30%. Prices of each of the factors can vary in a more complex
way, depending on the factor intensities of the sectors. Thus, if the transport intensive good
(good 1) is also labour intensive (as assumed in this example), the wage falls rapidly.\(^3\)

While this is just an example, it illustrates an important general point. In this example all
activities have constant returns to scale and perfect competition, and all countries have identical
technologies, yet considerations other than factor endowments determine (in a systematic way)
production structure, factor prices, and income levels. In this example, geography matters for what countries do and how well off they are. However, the basic insight is simply that of comparative advantage trade theory; production structure is determined by the interaction of country characteristics and product characteristics, applied in this example to the geography of countries and transport intensities of products.

2.3. Production networks and vertical fragmentation.
The factor intensities of different stages of the production process vary at least as widely as do the factor intensities of finished products as a whole. The scope for trade therefore depends on the extent to which it is worth ‘fragmenting’ the vertical production process, and undertaking different stages of production in different locations. While the benefits of fragmentation are input costs saved, the costs are those incurred in shipping (including the costs of time in transit) and the package of costs associated with managing remote operations or supply chains. Recent decades have seen reductions in shipping costs and times (Hummels 1999, 2000) and perhaps more importantly the development of information and communications technologies (ICT) that have done much to reduce the costs of remote management. The consequent growth of ‘vertical specialisation is documented by a number of authors. Yeats (1998) estimates that 30% of world trade in manufactures is in components rather than final products. For Asian countries, Ng and Yeats (1999) report an overall pattern of assembly in lower wage economies, with components production taking place in Japan, Singapore and Taiwan. Hummels, Ishii and Yi (2001) chart trade flows that cross borders multiple times, as when a country imports a component and then re-exports it embodied in some downstream product. They find that (for 10 OECD countries), the share of imported value added in exports rose by one third between 1970 and 1990, reaching 21% of export value.

Analytically, the main basis of this trade is the interaction of factor abundance and factor intensity, as in Heckscher-Ohlin trade theory, but several comments are worth making. First, because products are likely to cross borders multiple times (as components are re-exported embodied in downstream stages of production) transport costs are likely to be a particularly important obstacle to this type of trade. Security of supply will also be important, as delay will
disrupt the entire production chain. Patterns of production and trade therefore depend on interactions of transport intensities and location, as well as factor intensities and endowments. Thus, proximity to other stages of the production process and final markets is likely to matter, particularly for bulky commodities. The fragmentation can affect middle-income countries if, for example, processing of primary products is relocated from them to low-wage countries.

Second, although this trade is based on factor price differences, an increase in the trade does not necessarily narrow international differences in these prices, as might be expected. The reason is that an activity relocating from a Northern economy to a Southern one may be unskilled labour intensive relative to other activities in the Northern economy, but skilled labour intensive relative to the endowment of the Southern economy. Feenstra and Hanson (1996) develop this idea, and argue that it may apply to many of the activities that have relocated from the US to Mexico; these activities are unskilled labour intensive compared to the US economy, but skilled labour intensive compared to the Mexican economy. The effect is to increase wage inequality in North (as unskilled labour intensive jobs leave) and in South, as the new pattern of labour demand is more skilled labour intensive than the previous Southern employment structure. 4

2.4. Increasing returns, market access and supplier access:
The ‘new trade theory’ that was developed from the late 1970s added to trade theory a systematic treatment of increasing returns to scale (at the level of the firm) and imperfectly competitive market structures. A number of insights come from the analysis. On the welfare economics side, there is improved understanding of the sources of gains from trade; in addition to comparative advantage, trade increases product variety, and is also likely to have a pro-competitive effect, reducing monopoly power while at the same time allowing firms to become larger and better exploit economies of scale. On the policy side, there was extensive study of ‘strategic’ industrial policy, although also a realisation that effects of policy were likely to be unpredictable, at least given the information available to policy makers (see Brander 1995 for a survey).

As regards the pattern of trade, the literature provides a theoretical explanation of the large volumes of intra-industry trade that are observed, and also shows how market size can provide a basis for comparative advantage. The main result (sometimes known as the home
market effect) is that economies with large markets will get a disproportionately large share of increasing returns industries. The intuition is derived from thinking about the location decisions of potentially footloose firms, operating in the presence of some transport costs on their output. Because of increasing returns to scale in production firms are faced with real location choices (if they had constant or diminishing returns it would be costless to split production and locate everywhere to save transport costs). These firms will seek to locate in the large market, in order to save transport costs and have lower marginal costs of supplying consumers, and this will create outcomes in which an economy with, say, twice the market size, will have more than twice as many firms as a smaller country. Obviously, in general equilibrium this will translate into higher factor prices in the larger economy, giving an outcome in which firms trade off benefits of market access against the higher wage costs of central locations. As in figure 1, general equilibrium will mean that economies with good market access will tend to specialise in industries where the home market effect is strong, and vice versa.

The ideas here and in section 2.2 are based on access to markets and to supplies of intermediate goods, and to make them operational they have to be generalised to a multi-country setting with a real geography of trade costs between locations, rather than just the linear world of figure 2. This can be done both in theory (eg Fujita, Krugman and Venables 1999) and empirically. Essentially, a measure of ‘market access’ can be computed for each location, and is a theoretically well founded version of the old idea of market potential, measuring demand at each location, weighted by a decreasing function of distance. While market access measures demand in the final market, an analogous measure of ‘supplier access’ can be constructed to measure proximity to suppliers of intermediate goods. Redding and Venables (2001) show how both the market access and supplier access of each country can be computed from trade data, and confirm empirically the importance of these variables in determining cross-country variations in income, as suggested by theory. Access to foreign markets and suppliers can explain one-third of the cross-country variation in per capita income, and the full measures of market access and supplier access explain up to two-thirds of the variation.
3. Changing the basis of comparative advantage

Analysis outlined above took as given the endowments, technologies, and market and supplier access of countries, and showed how this generates cross-country patterns of production, trade, and income distribution. However, in a developing country or region all these elements are subject to change, and analysis must be extended to handle this. We proceed in three stages. First, we look at foreign direct investment (FDI). Inflows of FDI may bring capital into the country, although are probably more important for bringing in technology and firm specific brands. We then look briefly at labour endowment issues, reviewing ideas on the interactions between trade and education, and between trade and migration. The third stage is to look at the implications of externalities between firms in an economy. These may be either technological or pecuniary externalities. Either way, they make comparative advantage endogenous, so their presence requires a major rethink of the basis of trade and of the effects of trade on incomes.

3.1 Capital and foreign direct investment

Multinational corporations provide an important vehicle for providing countries with capital and, more importantly, technologies for production and for gaining access to export markets. Their importance in the world economy has been extensively documented, with the overseas production of affiliates of US firms now three times larger than total US exports, and nearly half of US exports going directly to US affiliate companies.

3.1.1 Determinants of FDI

The literature focuses on three broad motives for undertaking FDI. The first is to gain access to natural resources, and we devote no further attention to this. The second is to gain better access to host country markets; this is known as horizontal FDI since it typically involves duplicating a part of the production process that is already operating elsewhere (eg, undertaking vehicle assembly in several regions). The third motive is to benefit from factor price differences by, for example, moving unskilled labour intensive activities to low wage economies; this is known as vertical FDI, as it involves breaking up the vertical production structure. It is the basis for the formation of production networks or production chains.
The decision to undertake horizontal FDI is usually posed as a trade-off between supplying a market by exports or by local production.\textsuperscript{5} If supply is through exports then marginal costs of supply might be high, because of transport costs, delivery lags, or failure to tailor the product to local circumstances. The alternative, supplying the market by local production, typically yields lower marginal costs but higher fixed costs, incurred in operating a further production plant. This suggests several circumstances where FDI is more likely to occur. One is where transport costs or other trade barriers are high, one example of this being ‘tariff-jumping’ FDI occurring simply to avoid the barriers created by import substituting regimes. Another is where local markets are large, this allowing the fixed costs of setting up an additional plant to be spread across a large volume of output. Thus, as with the ‘home market effect’ of section 2.4, increasing returns to scale activities are drawn into large markets. Additionally, there is considerable evidence that FDI projects cluster together in particular countries. This may just be because of the underlying characteristics of the country, but is likely to also be because of ‘herding’, as one FDI project has a demonstration effect, encouraging others.

For vertical investments, the trade-off is quite different. The advantage is access to cheap labour to undertake labour-intensive parts of the production process. The main costs are the additional costs incurred in managing an operation at a distance, and the transport costs (shipping costs, but also possible delay and increased uncertainty) involved. Transport costs are potentially extremely important, because such trade typically involves multiple border crossings as different stages of production take place in different countries.\textsuperscript{6} Thus, we expect production networks to form where lower wage economies are close to, or have very good transport links with, higher wage countries, and in commodities that are transport unintensive. Here too, there is evidence of the geographic clustering of investment projects.

### 3.1.2 Consequences of FDI

Horizontal and vertical FDI are likely to have substantially different effects, and here we briefly outline their impacts on trade and income.

Looking first at trade, horizontal FDI is likely to be a substitute for trade, as firms use FDI instead of imports to supply the market. Of course, this statement needs some qualification as
components are likely to be imported, total sales may increase, and a particular project may be used as a base from which to supply a larger regional market. Nevertheless, its basic function is to replace imports. By contrast, vertical FDI is a complement to trade, and may even create trade flows that are much larger than the value of the final good produced as component parts cross borders repeatedly, embodied in the product being shipped.

Turning to employment and income generation effects, horizontal inwards FDI will generally increase labour demand, as local production replaces imports. However, if local firms are operating in the industry then there are also likely to be competition effects, as the FDI project takes sales from these firms. In contrast, vertical FDI is typically export oriented, so will not have competition effects in the local market. It can be viewed as a mechanism for achieving the effects predicted by factor endowment trade theory, enabling countries to export their lower skilled labour embodied in the firms’ output.

Finally, it is often suggested that one of the main effects of FDI is to generate beneficial spillovers to the rest of the economy. The spillovers can arise through demonstration effects, direct linkages between FDI and local firms (eg collaboration to improved input product and process quality), and labour market turnover of trained workers. There is a good deal of case study evidence of the importance of all these mechanisms, although the picture from econometric studies is more mixed (see Blomstrom, Kokko and Zejan 2000 for a survey).

3.2 Labour

The quantity and quality of labour are obviously the most important elements of a country’s factor endowments, and these too are subject to change. Here we review interactions between trade and endogenous change in these endowments.

3.2.1 Skill acquisition

The previous two subsections took as given the quality of primary factors, and we now turn instead to determinants of the quality of the factors themselves, focussing on skill acquisition. The overall importance of skills to economic development goes beyond the scope of this paper, and here we concentrate just on interactions between trade and skills. There are two issues.
First, what is the impact of trade on the acquisition of skills? And second, what is the impact of the skill composition of the labour force on trade patterns and the structure of the economy?

Analytical work on the first of these questions takes the Heckscher-Ohlin model as point of departure. In this model a long established – although perhaps not so well known – result is that trade liberalization reduces the incentive for workers in unskilled labour abundant economies to acquire skills. For example, if the factors of production are skilled and unskilled workers then, under autarky (or restricted trade) an economy with few skilled workers will tend to have a high wage for these workers, compared to the more abundant unskilled workers. This creates an incentive to become skilled. Trade liberalization, however, allows skill intensive goods to be imported while expanding production of unskilled intensive. This reduces the demand for skilled labour and increases demand for unskilled labour, narrowing the wage gap between them and reducing the return to education.

This result can be challenged on several grounds. The first is that the effect depends on comparison of the skill composition of labour demands pre- and post trade liberalization. Once we move outside a simple two sector model it need not be the case that post-liberalization labour demand is less skill intensive. We have already seen how trade in new activities -- as when a stage of production can be detached and relocated -- can be unskilled intensive relative to Northern endowments, but skilled intensive relative to Southern endowments (section 2.3). Similarly, if non-traded activities are unskilled labour intensive, then opening up the economy and expanding the share of activity in tradeable sectors might increase demand for skilled labour. Thus, there may be factor endowment based trade that increases the incentive for Southern workers to acquire education. The dichotomy between skilled and unskilled labour is also too sharp, as there is evidently a continuum of skill levels. There is plenty of evidence (eg from the Asian experience) that at least primary level skills are at a premium in modern sector manufacturing export activities.

An alternative mechanism arises if trade liberalization allows movement of some factor (or knowledge) that is complementary to skilled labour. Tang and Wood (2000) suppose that globalization allows ‘knowledge workers’ to gain better access to cheap Southern unskilled labour. If knowledge workers combine with relatively skilled labour in South, then the effect
will be to raise wages of skilled workers in South and increase incentives for education. The same argument can be made for the movement of capital, for the package of activities embodied in FDI, and for transfer of new technologies. In addition, it may be the case that openness to trade changes the incentives for government to support education. A political economy analysis of this is undertaken by Bourguignon and Verdier (2000).

### 3.2.2 Migration

The basic variant of traditional trade theory takes factor endowments as given and immobile. When labour mobility is possible there is a strong presumption from theory that trade liberalisation decreases the incentive to migrate, by reducing international factor price differences. There is no completely general result here, as in full general equilibrium a narrowing of goods price differences (brought about by trade liberalization) can have complex effects on factor prices. For example, trade liberalisation that raises wages of unskilled labour could reduce returns to some skill types, encouraging their emigration. But we have also observed trade liberalization being the mechanism through which abundant types of skilled labour have become fully employed (the Bangalore experience).

What about the effects of emigration on economic performance and welfare? If the perspective is the real income of remaining citizens, then emigration is damaging if emigrants were being paid less than their full social marginal product, a concern more often voiced for skilled workers than unskilled. We take up some of these issues in the next section, where we look at the implications of externalities between sectors.

### 3.3. Spatial externalities and cumulative causation.

Study of cumulative causation processes have a long history in development economics, and were emphasised particularly in the writings of Hirschman (1958) and Myrdal (1957). In recent years they have come to attract renewed interest, both in policy circles (eg Stern 2001) and in formal economic analysis (eg Murphy, Shleifer and Vishny 1989). The trade and geography literature has shown how the location of production can be subject to these processes, and has drawn out their implications for the spatial structure of activity and of incomes. We now develop
some of the ideas in this literature.

Previous sections were based on the conventional assumption that firms could internalise the effects of their actions. We now turn to a case where externalities are – in some sectors at least – pervasive, and the actions of one firm have a direct impact on the performance of others. If an externality were to be transmitted equally to all industries, or to a particular industry in all countries, then comparative advantage would be unaffected. Usually however, externalities are limited in both sectoral and geographic scope. A new product or technique will only be used in some sectors, and the price of the product or availability of the technique will vary across countries. If externalities are sector and location specific in this way then they form a basis for comparative advantage, which now becomes endogenous. Externalities between firms may be either technological or pecuniary, the former arising when actions of one firm affect another without going through a market, and the latter where the interaction is through a market but firms are not able to capture the full benefit (or costs) of their actions. This section will explore the mechanisms behind these externalities and develop their implications. It will draw heavily on recent ideas on trade and geography, and (to a lesser extent) on trade and growth.

The main sort of technological externalities are knowledge spillovers of various types. These could take the form of demonstration effects, or the spillover of R&D, technical knowledge, or accumulated learning by doing. Early work modelling the international implications of these effects includes Krugman (1981, 1987). For example, Krugman (1981) assumes that labor input coefficients in manufacturing are decreasing functions of a region’s capital stock. These increasing returns imply that any initial differences between regions will be amplified as one gets ahead of the other, creating one developed and one underdeveloped region. The problem with this approach is that the key mechanism – the source of the technical spillover – is left as a black box, and most subsequent research has sought to look inside this box to identify the sources of the market failure that generates the externalities.

In contrast to technological externalities, pecuniary externalities are transmitted through markets. The key feature is that firms are unable to fully capture the benefits of their actions. For example, introduction of an improved intermediate good (or business service) may carry with it a net benefit to purchasers, who become recipients of a forwards linkage. If markets are less
than perfectly competitive then changes in quantities can be of positive value to the firms concerned, this creating demand (backwards) linkages. For example, a small increase in one firm’s purchases is of no value to firms selling, if they are selling at price equal to marginal cost. But if they sell at price greater than marginal cost (as under imperfect competition) then the extra sales increase profits. The geographical dimension to this arises if there are transport costs or other trade frictions, so the extra purchases are spatially concentrated. Then increased activity by downstream firms will raise demand for local upstream firms, raising their profits and perhaps attracting further entry. Entry of firms may in turn create forward linkages; as new upstream firms enter so new varieties will be developed or the market for existing ones will become more competitive. Thus, the benefits of entry are not all appropriated by the entrants, instead being transmitted to the downstream firms – a pecuniary externality. The combination of forward and backward linkages creates a potential process of cumulative causation: expansion of downstream activity increases demand for upstream output which attracts entry, improving the supply (price or varieties) of intermediates, attracting further downstream entry, and so on. Of course, this interaction of forward and backward linkages has been studied before in development economics. However, the mechanism is dependent on market imperfections, and it is only quite recently – following the development of models of trade with increasing returns to scale and imperfectly competitive market structures – that the mechanisms have been analysed formally, and their implications for trade and production drawn out in a rigorous manner.

Pecuniary externalities can arise in factor markets as well as product markets. For example, if one firm trains labour it may not be able to appropriate all the benefits as workers can quit and work for other firms. There may also be labour market ‘pooling’ effects, arising as firms and workers are better able to share random shocks in larger labour markets. Notice again, that the range of the externality will typically be limited across industries (depending on the specificity of labour skills) and across space (depending on the mobility of workers).

To see the implications of these externalities it is worth developing a simple example. Suppose that there are just two locations (countries) and an industry in which the total number of firms is fixed (at \( n \)). How is the distribution of firms between countries determined? Figures 3a and 3b give two alternative cases. The fixed number of firms is measured by the length of the
horizontal axis, while the numbers in country 1 and country 2, \( n_1 \) and \( n_2 \), are measured by
distance from the right and left hand ends of this axis respectively; thus, a point on the horizontal
axis measures a division of the industry between the two countries. The vertical axis measure the
profitability of a single firm in each location, assumed to depend just on the number of firms
active in the location, \( \pi_1(n_1) \) and \( \pi_2(n_2) \).

The first of these figures, 3a, indicates a standard ‘neoclassical’ case. The functions \( \pi_i(n_i) \)
are downward sloping, indicating diminishing returns to the activity in each location (perhaps
because expansion bids up the price of some scarce factor, or because it reduces the price of
output). The equilibrium is at point E, where returns are the same in both countries, so it is not
profitable for any firm to relocate. The figure is constructed under the assumption that country 1
has a comparative advantage in the sector so the \( \pi_1(n_1) \) line is higher than \( \pi_2(n_2) \). Consequently
country 1 has more of the firms in the equilibrium.

The second figure, 3b, gives the case where positive externalities are strong enough to
overturn any forces for diminishing returns, making the \( \pi_1(n_1) \) and \( \pi_2(n_2) \) schedules upwards
sloping. The intersection of these curves is now an equilibrium, but it is unstable (if a firm
relocated from country 2 to country 1 then it would raise \( \pi_1 \) relative to \( \pi_2 \), so other firms would
follow). There are two stable equilibria at the points on the axes labelled \( E_1 \) (where all firms are
in country 1 and \( \pi_1(n) > \pi_2(0) \)), and \( E_2 \) (where the converse applies). These points are equilibria,
as there is no incentive for any firm to relocate.

Several lessons can be drawn from this simple example. First, there are multiple
equilibria. Activity agglomerates, but there is nothing in the theory to tell us in which country.
We need to go outside the theory – to history for example – to resolve this indeterminacy.
Second, the equilibria are robust to parameter changes. Suppose that we reduce country 1’s
comparative advantage, pushing the \( \pi_1(n_1) \) schedule downwards. In figure 3a the equilibrium
moves continuously to the left, but in figure 3b the equilibria might be unaffected. For example,
if the equilibrium were at \( E_1 \) no firms relocate until \( \pi_1(n_1) \) reaches the dashed line illustrated, at
which point there is ‘catastrophic’ change. It becomes worthwhile for a single firm to relocate
(as \( \pi_2(0) > \pi_1(n) \)), and once one firm moves others follow. \( E_1 \) ceases to be an equilibrium, and
all activity moves to country 2 – now a unique equilibrium. These points demonstrate how there
is a possible first mover advantage. If a location becomes established in an activity then it will be difficult for another centre to become established. The cluster may well have a ‘deep’ comparative advantage, and be able to pay high wages or survive adverse shocks. However, survival of an established centre is only robust up to a point, beyond which collapse and relocation of activity might be rapid – like the ‘punctuated equilibria’ of evolutionary biology.

The example shows how activity may cluster in one location, but what implications does this have for real income in each location? The answer depends on the alternative uses of the factors employed in the industry. If the cluster is small then factor supply curves may be horizontal (over the relevant range), so having the cluster does nothing to raise wages or other factor prices. Benefits are instead all passed on to consumers (worldwide and local, depending on the magnitude of any trade costs). But if the cluster is large enough for its presence to change factor prices then the country with the cluster will, other things being equal, be better off: having the cluster raises labour demand and wages.

3.3.1 Agglomeration and North-South inequalities:
Models of clustering have been applied to a number of different trade and development issues, and we now review some of these applications. We base discussion on the Krugman- Venables (1995) model (see also Fujita, Krugman and Venables 1999), in which there are two different types of economic activity. One is termed ‘agriculture’, and represents a composite of all the perfectly competitive (and non-increasing returns) sectors of the economy. The other is ‘manufacturing’ and has two key features. One is that firms operate under increasing returns to scale and are potentially footloose, choosing where to locate according to market access and production costs. The other is that these manufacturing firms use primary factors and manufactures, and sell their output both to consumers and to other manufacturing firms. This input-output structure means that there are forward and backward linkages, and firms will tend to want to locate close to other firms who supply some of their inputs and provide some of their market. The labelling ‘agriculture’ and ‘manufacturing’ is intended to capture these differences, although is clearly very stylised. In the following section we start to disaggregate these sectors, and ask where the linkages are likely to be largest.
Krugman and Venables (1995) showed how this model could give rise to large wage and income inequalities between countries that have the same underlying characteristics of endowments, technology, preferences and location. Their results are summarised in figure 4, in which \( w_i \) measures the real wage in country \( i \), and there are just two countries in the world. The horizontal axis is a measure of trade barriers, and we see that when these are very high the two countries are identical \( (w_1 = w_2) \). This is because firms have to locate near final consumers which prevents clustering from occurring. At lower levels of trade costs the need to be close to final consumers is reduced so the clustering forces generated by linkages become relatively more powerful. Manufacturing then agglomerates in one country; theory does not say which, but in figure 4 it is assumed to be country 1, which as a consequence has higher labour demand and higher wages. This change is ‘catastrophic’, as the model passes through a bifurcation point at which there is a qualitative change in the structure of equilibria changes. The economic story is then, that as transport costs are reduced one region of the world ‘deindustrialises’ the other, and the world necessarily develops a dichotomous ‘North-South’ structure. Wages are much higher in North, but no firm wants to move to the South as by so doing it would forego the advantages of proximity to a large market and large number of supplier firms.

In this model further reductions in trade costs narrow the wage gap as the world enters the ‘globalization’ phase. The reason is that at low trade costs the pecuniary externality created by firms becomes less spatially concentrated – intermediate goods can be shipped more cheaply between countries. In the limit of perfectly free trade the model reduces to textbook international economics, with factor price equalisation. Factor price equalisation occurs as the limit of this model because the geographical spread of the pecuniary externalities is determined by trade costs on intermediate goods; at perfectly free trade we reach the ‘death of distance’. However, other sorts of pecuniary or technological externalities might be less sensitive to trade costs (for example externalities in the labour market) in which case factor price equalisation is not reached.

The general message of figure 4 is then, that at very high trade costs location of manufacturing is determined by the need to be close to final consumers: at low trade costs factor supply becomes important, and relocation of industry will narrow factor price differences: at intermediate trade costs the potential for clustering and consequent income inequalities is
3.3.2 The spread of industry:
The story outlined above is one of reductions in transport costs that, beyond some point, start to narrow international income differences by facilitating the spread of industry out from established clusters of activity. We now look at this spread of activity in more detail.

The first issue is the extent to which globalization and new technologies are leading to the death of distance, and allowing deconcentration of activities. Where are we on figure 4? It seems clear that, for most activities, transport costs are now low enough to allow some deconcentration to occur. Some commentators on new technologies are protagonists of the death of distance view (see for example Cairncross 2001), although this position can be exaggerated.\textsuperscript{11} Falling transport costs and new technologies have allowed some activities to relocate, but proximity remains important for the vast majority of economic transactions, as demonstrated by econometric estimates of gravity models for trade, investment, equity, and technology flows.\textsuperscript{12} Some activities can be fully digitized so have transport costs that are essentially zero. However, these are activities that are also likely to become fully automated, are rapidly falling in price, and probably account for no more than a few percent of world GDP.\textsuperscript{13} Many other information based activities require face-to-face contact, either because a high degree of interaction between participants is needed (R&D) or because of problems of contractual incompleteness and monitoring. In some activities, new technologies actually seem to increase the value of proximity; new stock control, lean production process and just-in-time technologies increase the costs of delay, and thereby necessitate closer location of related activities. Furthermore, an extremely important source of clustering is the labour market – the development of local pools of skilled labour. This clustering mechanism is (unlike the intermediate goods story of Krugman and Venables) essentially untouched by new technologies and falling transport costs. What all this suggests is the relocation of some activities from established centres, while other activities (in particular those where knowledge flows are too complex to be readily codified and digitized) are likely to remain entrenched.

The second issue is the geographical form that the spread of activities is likely to take.
How is the analysis of figure 4 modified if there are many countries? A conventional view of growth and convergence would suggest that countries will all converge at a more or less similar rate to common steady state values of capital and income per worker. The new economic geography view suggests, by contrast, that the world may consist of a rich club and a poor club, and development might take the form of selected countries in turn making a rather rapid transition from one group to the other. This is illustrated in figure 5, drawn from Puga and Venables (1996), who apply the preceding model in a multi-country framework. In the initial situation, illustrated at the left hand edge of the diagram, manufacturing is concentrated in country 1, and other countries have only agriculture. Exogenous technical progress increases demand for manufactures in the world economy (moving to the right on the figure), and this increases the wage gap between country 1 and other countries. At some point the wage gap becomes too large to be sustainable, and industry starts to relocate to other countries. However, it does not go to all other countries, because to do so would be an unstable equilibrium. If one country were to get just slightly ahead the linkages generated would cause this country to become the preferred location for further manufacturing expansion. In the case illustrated country 2 industrialises while 3 and 4 are left behind.

The process then repeats itself. Countries 1 and 2 have industry and a wage gap relative to the rest of the world. Continuing growth raises the wage gap until manufacturing once again spills over and spreads to a new location. Development is therefore the rapid transition of countries in sequence, from the poor club to an expanding rich club.

Puga and Venables argued that the model is suggestive of Asian development experience, although as theoretically modelled the structure is obviously too stark, particularly in its assumption that all candidate countries for industry are identical. Country differences that make a location an attractive host are the obvious ones of good institutions and low unit labour costs. Also important is proximity to established regions, so that intermediates can be imported and exports shipped at relatively low cost. Large market size is beneficial, providing a local market for developing industry. Combining some of these factors implies that a strong natural resource base has opposing effects: it increases local expenditures and market size, but by raising wages and unit costs is likely to deter industrialisation.14
The theoretical model also suggests that – as long as there are some remaining benefits to clustering – there is no guarantee that all countries will eventually experience industrialization. World demand for manufactures may perfectly well be met by efficient size clusters in a subset of countries. What might then determine the size of the income gap between countries with manufacturing and those without? First, of course, countries may be without manufacturing because they have other highly productive activities (the New Zealand model). But if not, the question becomes, how low must wages be to attract a firm to relocate from an established centre? The answer provides the upper bound on the sustainable wage gap.

Finally, we have so far couched the discussion largely in terms of an aggregate manufacturing sector. In practice, this must be disaggregated, so that the linkages between sectors depend on the structure of the input-output matrix as well as the tradeability of the product. What can then be said about the way in which industrial structure evolves during the development process?

Activities that are most easily detached from existing centres are, first, those with a high labour share as these benefit from the variation of wage costs across locations. Second, they are activities that receive few linkages, either forwards or backwards, from activities in the existing centres. Receipt of backwards linkages will be unimportant if the commodity is easily traded, or if it can be sold to a local market rather than being dependent on demand from existing centres. Receipt of forward linkages will be unimportant if intermediate manufactured inputs are a small share in production, or if intermediates are very easily traded. (Of course, in the case of primary processing industries linkages may be available from local supplies of natural resources or primary products). Finally, transferability of labour skills and technologies may be important (see Sutton 2000 for investigation of the determinants of transferability of firm capabilities).

While linkages received by the sector are important for assessing which activities can be detached from existing centres, linkages given off by the sector are important for determining both the speed and the shape of the development path. Mechanisms that attract further firms and sectors are: market enlargement (both through income and final demand, and through derived demands for intermediate goods); improved supply of intermediate goods and services; the development of an increased pool of (non-specific) labour skills. Against this, development of
initial sectors may have increased wages and prices of other factors, as well as creating negative externalities through congestion and pollution. Some preliminary investigations of these effects were undertaken in Puga and Venables (1999) and Fujita, Krugman and Venables (1998), who found that development was typically most rapid when upstream industries could be easily detached from existing centres, since these industries then created strong forward linkages facilitating the movement of downstream industries.

These ideas are suggestive, but empirical work has not yet been undertaken to confirm them. We return to the policy implications of this approach in section 4.

3.3.3 External trade and internal economic geography
Discussion so far has been based at the national level, and hence assumes that labour is not mobile between locations. What difference does it make if labour is mobile on a large scale, as is likely to be the case within a country and, in some circumstances, internationally?

Immobility of labour is a major force for dispersion of activity because as a cluster develops it raises wages choking off further growth of the cluster. Removing this labour supply constraint makes agglomeration more likely, and Krugman (1991) shows how falling transport costs could lead to agglomeration of manufacturing activity and population in one region. Essentially, any productivity advantage that one region gains will not be counteracted by rising wages if there is a sufficiently elastic supply of labour from other regions. Expansion of the region will continue until choked off by rising prices of immobile factors (land) or external diseconomies of scale, such as congestion. The approach has been developed to explore the development of cities, and the growth of mega-cities in developing countries (for example Puga 1998). Although this is not the place for a review of the internal economic geography of developing countries, several points are worth making.¹⁵

First, whether agglomeration forces will give rise to mega-cities and consequent emptying of regions depends on the breadth of activities over which clustering forces operate. Within developed countries clustering often takes quite a narrow sectoral form, arising because of externalities associated with particular sector specific skills or sector specific input requirements. This results in multiple sector- (or function-) specific clusters, each of which may be quite small
relative to the economy as a whole – vehicles in one region, electronics in another, financial services in a third, and so on. In contrast, it is possible that clustering in developing countries has a wider sectoral range, as would arise if the externalities occur at the level of provision of basic business services, public sector services or general (rather than sector specific) labour skills. In this case instead of multiple clusters, a single cluster -- a mega-city -- is likely to develop, until further expansion is choked off by other diseconomies of size.

Second, relatively closed and inward looking economies are likely to be more prone to development of these mega-clusters than are open ones (Krugman and Livas 1996). To the extent that clustering is driven by forward and backwards linkages, the strength of clustering will be lower the more outward orientated are firms, purchasing inputs and supplying outputs from and to world rather than local markets. Thus, it is suggested that an additional benefit of trade liberalisation is that it promotes the restructuring of a country’s internal economic geography, facilitating a deconcentration of activity from the prime city region.

4 Policy issues

The benchmark for assessing policy is, as always, the first theorem of welfare economics which establishes the efficiency of competitive equilibrium. Policy analysis is then based on the identification of market failures, and the design of policies targeted at these failures. This methodology remains applicable in the presence of increasing returns, imperfect competition and consequent pecuniary externalities, although policy design becomes more complex for two reasons. First, market failures may be pervasive throughout large sectors of the economy, in which case the usual second best arguments apply with force. Second, we have seen how these failures can lead to multiple equilibria and corner solutions. The fact that there are multiple equilibria means that policy instruments do not map uniquely into outcomes. Policy might seek not just to make a marginal adjustment to an interior equilibrium, but to cause a non-marginal change, shifting the economy to another equilibrium. Awareness of these difficulties, combined with awareness of the problems of policy capture and government failure, have made researchers reluctant to develop a theory of economic policy in this sort of environment. Nevertheless, some
remarks are possible, and in this section we discuss a number of policy issues, looking first at sectoral policies, then at general infrastructure and trade promotion policies, and finally at preferential trade promotion policies – regional integration.

4.1 Sectoral policies

**Differential rates of demand growth:**
Countries that are specialised in commodities with fast growing demand will tend to do better through time than countries whose terms of trade are declining because of slow growth of export demand. The differential rates of demand growth could be due to a number of reasons – income effects or technical change – and slow growth will lead to a larger terms of trade deterioration the less price elastic are demand and supply.\(^{16}\)

These observations have led to the suggestion that countries should actively seek to specialise in faster growing sectors. However, for this to constitute a valid reason for policy intervention the case has to be made that the private sector is either failing to perceive demand conditions, or failing to respond to them. The former is a hard case to make. There is no reason to believe that government has better information than the private sector in forecasting demand growth, and the record of governments in picking winners is not good.

The latter case – poor private sector response – provides a sounder basis for policy, but requires diagnosis of why the private sector response is lacking. The classic answer is to link poor response to capital market failure; the long run investments required to profit from growing demand are not made because of credit market constraints or a price of capital exceeding its social marginal cost. This is not an industry specific market failure, and the response should be in the capital market, not in the support of a particular sector.

**Externalities:**
The presence of externalities provides a well established case for the use of policy to expand activities that create positive externalities, and contract activities that create negative ones. Since most of the discussion of section 3 turned on the presence of externalities, what are the policy implications?
The first issue is, can sectors or projects that create relatively large externalities be identified. Theoretically, the answer is yes, although it depends on a complex combination of factors. How will derived demands for inputs be split between local firms and imports, determining the extent of backwards linkages? Is there imperfect competition in the upstream industry? What is the likely quantity response in the upstream industry? Will this quantity response create positive forward linkages? How firm specific are labour skills created by the project? Empirically, we know little about how these effects vary across industries. There is an inherent unpredictability, in so far as the externalities create value only in so far as they elicit positive quantity responses from other firms.

Second, the type of industrial linkages and externalities discussed above are generally reciprocal externalities — each firm creates them, and also receives them from other firms. The market failure is therefore often described as a coordination failure, rather than simple externality, as the problem could be solved by collective action of a group of firms. This is the logic behind policies such as business or science parks, and infrastructure, research and education policy to facilitate the development of clusters. Such policies need not provide sectorally targeted assistance, but do create the environment in which coordination failures are minimised.

Finally, the record of government action to provide sectoral support is, with a few notable exceptions, generally poor. There is also the catch-22 of this sort of policy. If we knew exactly what sort of sectoral policy worked, then it would be used by many governments, the price of output from these sectors would decline, and (at a policy equilibrium) there would be no return to another government employing the policy.17

4.2 Infrastructure investment and the costs of distance
We have seen at various points the possible costs of being remote from economic centres or of having high trade and transport costs. The real income loss associated with distance from an economic centre was discussed in section 2.2. The extent to which trade costs inhibit participation in production networks and receipt of vertical FDI was discussed in section 3.1. And section 3.3 referred to the fact that high transport cost locations are unlikely to be candidates for spreading industrial development. These theoretical arguments are confirmed in the work of
Sachs and his coauthors (Gallup, Sachs, and Mellinger, 1998, Radelet, and Sachs, 1998) who find that activity is concentrated in regions with geographical advantages, such as proximity to the coast.

These arguments suggest the benefits of open trade policy and point to the potentially high value of infrastructure investments or other measures that reduce the costs of international transactions. The appraisal of such investments requires, of necessity, full cost-benefit analysis. In addition to direct benefits that can be captured by investors, a major infrastructure project changes prices, creating economic surplus that accrues elsewhere in the economy, but should nevertheless be included in the calculation. The total surplus created by such a project will consist of the cost reduction times the existing quantity, plus an amount that is proportional to the elasticity of the quantity response with respect to the cost reduction. The message from theory is that this quantity response is potentially large, as the economy is drawn into fuller participation in the world trade. If production networks develop then quantities of trade will be large and, as we have argued, cumulative causation processes might also cut in. Of course, infrastructure investments alone are not sufficient to ensure effective participation, but they are a necessary part of reducing the costs of peripherality.

4.3 Regional integration.

In the light of the overview of modern trade theory presented here, what are the likely effects of regional integration? Economic effects can be grouped into three main types, corresponding to some of the theory arguments.

First, costs and benefits depend on the comparative advantage of member countries relative to each other and relative to the rest of the world. This is the basis of the traditional trade creation and trade diversion argument, and provides a strong argument for North-South rather than South-South agreements. South-South agreements are prone to trade diversion, as sectors develop in the member country that has comparative advantage in the sector relative to the partner country, but not relative to the world as a whole. Trade diversion occurs as members’ imports come to be sourced according to this regional comparative advantage rather than global comparative advantage.
Second, benefits are potentially derived from fuller exploitation of economies of scale, combining with pro-competitive effects as firms in member countries are brought into more direct competition with each other. To the extent that a regional integration agreement achieves a larger integrated market, it may be possible to have both more competition and larger firms. The home market effects discussed in section 2.3 might operate to strengthen manufacturing sectors as a whole.

Third, propensity to cluster brings both benefits and costs. Regional integration may create a larger integrated market which will increase the scale of activities, allowing critical mass to develop. However, the development might be in just one of the member countries rather than in all. We have argued above that in developing countries clustering is likely to involve a relatively wide range of activities rather than just occur in particular sectors. This suggests a potential for divergence of economic structure and income between member states of developing country regional integration agreements. The unequal distribution of costs and benefits implied by trade creation and diversion can be amplified by these mechanisms.

5. Conclusions

There is now compelling evidence that full participation in the world economy is an inherent part of modern economic growth. In the words of Lindert and Williamson (2001);

...’the empty set contains those countries that chose to be less open to trade and factor flows in the 1990s than in the 1960s and rose in the global living-standard ranks at the same time. As far as we can tell there are no anti-global victories to report for the postwar Third World.’

This conclusion is supported by the work of Dollar and Kray (2000) who compare the economic performance of a set of developing countries they term the ‘globalizers’ with the performance of all other developing countries. The globalisers are identified on the basis of the decline in their tariff rates between the 1980 and the late 1990s and the increase in their trade to GDP ratio\textsuperscript{20}. Their performance compared to other developing countries is given in table 1, the first two
columns of which indicate how very much more open these countries became relative to the non-globalizers. The rest of the table gives per capita growth rates for the last four decades. The striking point is that while these countries fared worse than others in the 1960s and 70s their performance was dramatically better during the 1980s and 1990s, with per capita growth of 5.3% pa compared to -0.8% pa for the non-globalizers.

Table 1: Growth and trade performance of the globalizers.

<table>
<thead>
<tr>
<th></th>
<th>% fall in tariffs, 1980 - late 90s</th>
<th>% increase in trade / GDP, 1980 - late 90s</th>
<th>Annual growth pc income 1960s</th>
<th>Annual growth pc income 1970s</th>
<th>Annual growth pc income 1980s</th>
<th>Annual growth pc income 1990s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Globalizers</td>
<td>64%</td>
<td>92%</td>
<td>1.0</td>
<td>1.7</td>
<td>2.6</td>
<td>5.3</td>
</tr>
<tr>
<td>Non-globalizers</td>
<td>29%</td>
<td>1%</td>
<td>2.2</td>
<td>2.8</td>
<td>0.2</td>
<td>-0.8</td>
</tr>
<tr>
<td>High income</td>
<td>50%</td>
<td>4.5</td>
<td>3.4</td>
<td>2.5</td>
<td>1.9</td>
<td></td>
</tr>
</tbody>
</table>


While these findings establish an association between trade performance and growth, they do not establish a causal relationship, and still less do they identify particular trade policy instruments as determinants for causing growth. Cross-country regression studies studies endeavouring to establish the effect of trade policy on growth have not, in general, been successful. This suggests that trade policy reforms, while perhaps necessary, are not sufficient for good economic performance. Institutions and geography matter, and theory tells us that, given cumulative causation, it is possible that very small differences in these initial conditions can translate into large differences in outcomes. Theory also tells us that where successful growth performance is achieved so too is strong export performance, as countries grow by developing and exploiting a comparative advantage in a range of activities.

Now that attention has been turned back to these issues of cumulative causation much more work – empirical work in particular – is needed on the determinants of manufacturing success in developing countries, on the path that industrialisation takes, and the policy levers that are most conducive to it. In Stern’s words, how countries can ‘design credible investment
reforms’ to ‘hoist themselves onto a virtuous circle characterised by increasing returns’ (Stern 2001).

Endnotes:

1. We choose not to get drawn into discussion of ‘dimensionality’ issues in trade theory, but note that if there are few distinct types of primary factors then there will be a greater degree of country specialization than suggested by the surface of figure 1. For implications of this see Leamer’s work on multiple cones (Leamer 1987) and the more recent work of Schott (2000).

2. The model outlined here gives each country a slightly differentiated variety of each product. The limiting case of this more general model is one in which products are perfect substitutes, in which case dimensionality issues (numbers of goods versus numbers of factors) become important.

3. Although in this example good 0 is also assumed labour intensive, so the wage starts to rise in the import substitution region, zone IV.

4. The Feenstra and Hanson model is one of capital mobility, but the same possibility arises in a pure trade model, see Venables (1999).


7. Only if the seller were a perfectly discriminating monopolist could it extract all the surplus from the new good.

8. Notably by Hirschman (1958) and Myrdal (1957).

9. Much of this literature constructed economies to be ex ante identical, in order to demonstrate how, even with this extreme assumption, in equilibrium the economies would have different economic structures and income levels.

10. Formally at high trade costs the symmetric equilibrium is unique. There is then a range of value of trade costs at which there are five equilibria. The symmetric equilibria is stable as are equilibria with agglomeration in either country, and between these stable equilibria are two unstable equilibria. At lower trade costs the symmetric equilibrium becomes unstable (the point on figure 4 where the $\omega_1 = \omega_2$ line ends), and there are three equilibria, the two agglomerated equilibria being stable.

11. This section draws on Venables (2001); see also Leamer and Storper (2001).

13. Transcription of medical records is a rapidly growing business for India’s IT enabled export sector, but in all likelihood the transcription will be completely computerised within a few years.

14. Crafts and Venables (2001) argue that the sheer size of the nineteenth century US economy, assisted by an import tariff, was sufficient to cause industrialisation reversing its endowment based comparative advantage in primary products. For the ambiguous role of tariffs in smaller economies see Puga and Venables.

15. See Henderson, Shalizi and Venables (2001) for a review of some of this material.

16. For the classic statement of the case see Prebisch (1950) and a more modern statement Matsuyama (2000).


20. They rank developing countries according to the decline in their tariff rates between the 1980 and the late 1990s, and the increase in their trade to GDP ratio, and select countries that are in the top 40 of both lists. There were 16 such countries, and the two African countries that came closest to the criterion were added, giving: Argentina, Bangladesh, Bolivia, Brazil, China, Costa Rica, Ghana, India, Malaysia, Mexico, Nepal, Philippines, Poland, El Salvador, Thailand, Uganda, Uruguay, Vietnam. The early trade liberalizers – Chile, Turkey, Hong Kong, Singapore, South Korea and Taiwan are in the rich country group, not in this list of globalizers.

21. See Rodriguez and Rodrik 1999 for a critique of these studies.

References


Ng, F. and A. Yeats (1999), ‘Production sharing in East Asia; who does what for whom and why’ World Bank policy research working paper no 2197.


mimeo, Yale (www.som.yale.edu/Faculty/pks4/files/research/papers/cones_1100.pdf)
(econ.worldbank.org/files/1732_stern1.pdf)
discussion paper no EI28, LSE.
processed, Institute of Development Studies, Brighton
(www.ids.ac.uk/ids/global/ttint.html)
Review, 43, 935-945.
2568, (econ.lse.ac.uk/staff/ajv/jpap4.pdf)
Venables, A.J. (2001) “Geography and international inequalities; the impact of new
    technologies”, paper prepared for ABCDE, World Bank, Washington DC
(econ.lse.ac.uk/staff/ajv).
    Thünen model of international specialisation’, mimeo, LSE, (econ.lse.ac.uk/staff/ajv)
Von Thünen, J.H. (1826), ‘Der Isolierte Staat in Beziehung auf Landtschaft und
    Nationalökonomie, Hamburg. (English translation by C.M. Wartenburg von Thünen’s
    Isolated State, Oxford, Pergamon Press.
Wood, A. (2000), ‘Globalisation and wage inequalities; a synthesis of three theories’, processed,
    Institute of Development Studies, Brighton (www.ids.ac.uk/ids/global/ttint.html)
Figure 1: Cost share and endowment
Figure 2a: Zones of specialization

Figure 2b: Factor prices and incomes
Figure 3a: Industry location: diminishing returns

Figure 3b: Industry location: increasing returns
Figure 4: Real wages and transport costs

Figure 5: The spread of industry