Chapter 3

Trade theories: a survey of the determinants of exports

The causes which determine the economic progress of nations belong to the study of international trade (Marshall, 1959: 255).

3.1   Introduction

International economics, the oldest branch of economics, has developed many theories explaining the causes or determinants of trade, with comparative advantage being the canon. In a world without borders, analysis of international trade would be unnecessary. There would be free movement of goods, services and factors, and global and domestic markets would react similarly. This chapter provides the historical and environmental context in which trade theory developed and its relevance to South Africa.

Chapter 2 showed that trade is a complex outcome of many factors operating simultaneously. International trade theory seeks to answer four main questions:

- Why do we trade? (Benefits of trade);
- What do we trade? (Pattern of trade);
- With whom do we trade? (Flow of trade); and
- What are the implications of trade? (Welfare implications and income redistribution through changes in commodity and factor prices).

Since the welfare benefits of trade are taken as given (as discussed in Chapter 1), the focus in this chapter is on the positive determinants of trade and will focus on the first three points. Therefore, this chapter will address the following issues:

1. Review and evaluate trade theory and empirical research (section 3.2 to section 3.6);
2. Categorise causes of trade (section 3.7); and
3. Identify theories applicable to South Africa and policy implications (section 3.8).

Following Adam Smith, economists have tried to explain why countries trade and what prevents them from trading. Surveys of the history of economic thought focusing on international trade theory abound. Maneschi (1998: ix) laments, "given the length of the period of the survey and the richness of literature on comparative advantage, I do not claim to have written a comprehensive account of the history of this important concept in the theory of international trade". Due to the scope of the topic, this chapter will only scratch the surface and present theories relevant to South Africa.

3.2   Historic development of theories of trade

In some cases, it is easy to explain the composition of trade ancient writers identified as "natural trade." South Africa, for example, with its abundant gold reserves, became the major supplier of gold. Besides natural resources, other factors determine the composition of a nation's trade and yet others, the direction of trade. Certain factors either inhibit or even prevent trade.

Economists differ in their understanding of how economies function. Economic theory, essentially a rationalisation of particular historical experiences, tries to explain and interpret observable data in terms of causal mechanisms. The different policies they propose can be traced back to the philosophical underpinnings that have developed since the Enlightenment. Each theory focuses on a different aspect of why countries and firms export.

Three canons – the liberal (orthodox) tradition, the nationalist (mercantilist) tradition, and new trade theories – have emerged. The liberal approach is traced from Adam Smith and David Ricardo to the Heckscher-Ohlin-Samuelson model and other neoclassical formulations. New trade theories complement these and identify alternative causes. The mercantilist approach is identified with, among others, the German historical school of the late 19th century. Besides these three canons, other philosophical considerations, such as the Providentialists, can be traced further back (Brue, 2000).

(i)   Providentialists

Providentialists provide the earliest, and certainly the most metaphysical view, on why nations trade. Religion and philosophy have greatly influenced each other since ancient times and have shaped the concepts of value, law, social interaction and generally how business is conducted. Providentialists hold that "a deity" endows them with certain attributes, allowing them to develop certain skills and products that can be traded. Providentialist determinants are associated with Heckscher and Ohlin, and even Ricardo (1951: 13) holds certain providentialist views:

> *If the weather is injurious to one soil, or to one situation, it is beneficial to a different soil and different situation; and, by this compensating power, Providence has bountifully secured us from the frequent recurrence of dearths.*

While these reasons were used as a basis for trade, French polemicists contend that Providence has made them self-sufficient, and argue for severe import restraints. Providentialist thinking can provide the basis for international economic mutualism by suggesting potential gains from trade, but does it not provide an analytical framework.

(ii)   Mercantilism

Mercantilism, the first reasonably systematic body of thought devoted to international trade, surfaced during the mid-16th century from pamphlets, studies and treatises written by the rising European bourgeoisie and courtiers. They equate wealth with gold and silver bullion possessions and believe these precious metals are the only form of wealth worth accumulating.

They advocate increasing the nation's wealth by concurrently encouraging production, increasing exports and holding down domestic consumption. International trade that provides a favourable balance of trade thus increases a nation's wealth and power. Mercantilists therefore advocate government intervention to achieve a surplus on the balance of trade. Their policies can be summarised:

> *Although a kingdom may be enriched by gifts received, or by purchase from some other nations, yet these are things uncertain and of small consideration when they happen. The ordinary means therefore to increase our wealth and treasure is by foreign trade, wherein we must ever observe this rule; to sell more to strangers yearly than we consume of theirs in value (Mun, 1664, text modernised).*

To the mercantilists, the government controls the determinants of trade that can be manipulated. Using tariff and non-tariff barriers, the government can limit imports. Export promotion, although a priority is not actively pursued but limited to indirect measures such as controlling the major trading groups, granting trading rights and prohibiting trade between colonies and foreign countries.

Brue (2000: 17) argues that although the publication of Adam Smith's book in 1776 ended this school, neo-mercantilists still abound today. Many modern and post-modern ideas and policies resemble those of 200-300 years ago. Neo-mercantilists argue that by restricting imports, unemployed domestic labour and un-utilised capital can be employed to satisfy the demand for previously imported goods to expand domestic output and income.

(iii)   Physiocrats

Even though the providentialists were the first to identify the determinants of trade, the physiocrats were the first formal school to analyse trade, though not the determinants (other than agriculture). They maintain that natural laws govern the operation of the economy, and although these laws are independent of human will, humans could objectively discover them, as they could the laws of natural science. The physiocrats focus on agriculture, viewing it as the determinant of wealth, as opposed to the mercantilists' concept that commerce produces wealth. They are not

interested in foreign trade, except for the export of grain. They view manufacturing as useful but sterile, especially the production of luxury items for the nobility. Nevertheless, by promoting laissez-faire the physiocrats removed the obstacles to capitalistic economic development. They hold that government should not interfere in economic affairs beyond the absolute minimum. Government's role is essentially to protect life and property, and to maintain freedom of contract.

3.2.1 Classicists and neo-classicists

The liberal (orthodox) tradition can be traced back to the classical school, born from the scientific revolution and later influenced by the industrial revolution, it generally had a decided impact on economic thinking. Although trade issues feature prominently in early classical economic thinking and they argue that trade is beneficial, they do not identify what gives rise to trade other than surplus production. The theory of absolute advantage was developed by Smith to show the benefit of trade.

(i)   Absolute advantage

Foreign trade is a genus of all trade, whether between individuals, town and country, regions of a country, or between countries. Smith demonstrates that a family would not attempt to manufacture a product that can be bought cheaper (because they have some advantage in its production). He holds that what is true for the individual or household is also true for the country and it does not matter if the advantages occur naturally or are acquired. Countries should therefore import commodities that cannot be produced at a lower cost than their trading partner can. Apart from singling out a country's climate and soil as determinants of manufacture, Smith does not explain the sources of absolute advantage, although he identified other determinants of trade that have been incorporated and developed in other models (Maneschi, 1998: 41).

(ii)   Comparative advantage – labour productivity

Subsequently, the classical economists showed that beneficial trade was possible even without absolute advantage and developed the theory of comparative advantage. It is the most important theory to emerge from the classical period. Torrens "baptized the theorem" and Ricardo "elaborated it and fought for it victoriously" (Schumpeter, 1954: 133). It was "developed in greater detail by John Stuart Mill, and taken up again with some modifications by Cairnes and Bastable" (Gide & Rist, 1948: 644). This theory shows how it is possible for all countries to benefit from trade without having an absolute advantage in any product.

The Ricardian model, showing how technologically inferior countries benefit from free trade, is constructed such that production technologies are the only difference between countries.

Differences in labour productivity achieved by different technology are therefore one of the main reasons why countries trade.

(iii)   Comparative advantage – endowments

In reality, labour productivity only partly explains trade and ignores other inputs and differences in countries' resources. Ohlin (1991: 31-32) states:

> *The fact that the productive factors enter into the production of different commodities in very different proportions, and that therefore relative prices of the factors being different in different countries and international specialisation of production is profitable, it is so obvious that it can hardly have escaped notice. Yet this fact was long ignored in international trade theory.*

The Heckscher-Ohlin paradigm, probably the foundation of prevailing theory, is economic theory's "most complex and impressive theoretical structure" ever developed (Chipman, 1965: 479). Differences in the abundance of factor endowments and differences in factor intensities of commodities give rise to differences in autarky commodity prices. Samuelson helped popularise the theory by casting it in the familiar 2×2×2 form to derive the mathematical conditions under which prediction was strictly correct (Jones, 1979).

The specific-factors model, a short-run version of the Heckscher-Ohlin model, has become an important addition to trade literature. Haberler (1937) makes the theory of comparative costs compatible with the neoclassical theory of value, freeing it from the Ricardian labour theory of value. Trade in the specific-factors model is determined by the stock of specific capital, assuming an identical endowment of labour (the mobile factor). Where there are differences in labour endowments, trade will depend on the nature of the production functions and on the allocation of capital (the stock of specific factors) (Markusen *et al.*, 1995: 139).

3.2.2 Empirical evaluation of classical theory

Although theoretically sound, classical trade theories did not always stand up to empirical verification. Before analysing the new theories that emerged, this section gives a brief overview of the empirical analysis of classical theories both internationally and in South Africa. After looking at methodological issues, the first part of this section reviews Ricardian comparative advantage, while second reviews Heckscher-Ohlin.

Because of the difficulty of fully expressing trade theories and suitable data, it took a long time before they could be submitted to rigorous testing with observable data. The general law of comparative advantage involves a comparison of autarky and free trade prices. Trading reduces

differences in comparative costs autarky and free trade prices are therefore unknown. Both natural impediments to trade (such as transport costs and imperfect knowledge) and tariff and non-tariff barriers ensure that these differences are not eliminated. Therefore, even though countries trade, there are observed differences, although not of the same magnitude as the theoretical autarky differences[1].

(i)   Empirical tests of the Ricardian model

Although the Ricardian theory was first tested in 1951, 134 years after its inception, there have been a number of subsequent empirical tests. The Ricardian theory besides emphasising that trade is beneficial also explains the crucial variables that determine the patterns of trade. It is the second proposition that will be discussed, which Leamer and Levensohn (1994: 1344) describe as "a country export[ing] the commodity in which it has a comparative labor cost advantage and import[ing] the commodity in which it has a comparative disadvantage".

Bhagwati (1964) analyses the early research undertaken and points out the problems when applying more than two commodities and two countries. The models used propose a chain in which all commodities are ranked in terms of their comparative factor-to-productivity ratios. This weaker proposition ensures that it will always be true that each country's exports will have a higher factor-to-productivity ratio than each of its imports.

MacDougall (1951), Balassa (1950) and Stern (1950) assumed labour was the only factor of production and that the relative export performance of the United States and the United Kingdom in third markets depended on differences in output per man by industry. This reflected successively in differences in unit wage costs, unit value added and unit prices. MacDougall also found a strong inverse relationship existed in 1937 between relative US, and UK wage costs per unit of output and relative exports.

Stryker (1968) used relative third market exports (row 10 in Table 14), reciprocal trade (row 11 in Table 14) and total trade (row 12 in Table 14), as dependent variable and regressed these against the relative productivity rations. These early results are given in the Table 14:

---

[1]   Any application of comparative advantage theory should therefore relate autarky prices to observable features of economies, thereby establishing a relationship between trade, price and comparative advantage. To test comparative advantage, autarky prices need to be known. Unfortunately, there is no evidence on autarky prices besides a study undertaken by Bernhofen and Brown (2000) on Japanese data from the mid-19th century. Their research provides actual historical evidence of the relationship between autarky prices and trade through an examination of Japan's opening to trade in the 1860s.

Table 1   Early research of the Ricardian model

Author	Row	Sample	Year	Regression result
MacDougall	(1)	25 industries	1937	Regression slope = -4
	(2)	39 industries	1950	x = -2,19 + 1,89y
Stern	(3)	24 (= MacDougall 1937 sample)	1950	x = 0,98 + 1,65y (0,57)
	(4)	39 industries	1950	x = -0,68 + 1,27y (0,43)
	(5)	25 industries	1950	x = 0,91 + 1,49y (0,59)
Balassa	(6)	28 industries	1950	x = -1,761 + 1,59y (0,181)
	(7)	28 industries	1951	x = -53,32 + 0,721y (0,103)
Krenin	(8)	14 industries	1947	x= 0,37 + 5,5y (1,76)
	(9)	14 industries	1947	x= 1,37- 6,68y (2,24)
Stryker	(10)	24 industries	1949-62	x= -0,02+1,98y (0,09)
	(11)	24 industries	1949-62	x= -0,02+1,72y (0,82)
	(12)	24 industries	1949-62	x= -0,02+1,93y (0,73)

Source: Bhagwati (1964: 12) and Solacha (1991) (as cited in Kohler and Bruce-Brand, 2002).

By contrast, Bhagwati (1964: 9) holds that relative export prices could not necessarily be approximated by labour productivities and the other measures employed and tests the relationship between comparative unit labour costs and labour productivity ratios on the one hand and export price ratios on the other. He concludes that "these results, limited as they are, cast sufficient doubt on the usefulness of the Ricardian approach." Nevertheless most economists (Deardorff 1984), accept the hypothesis and the findings of MacDougall, Balassa and Stern.

While the simple Ricardian model had only one input (that was mobile across the two sectors of the economy, but immobile internationally), the Ricardo-Viner model introduced two additional factors that are sector specific. This lent curvature to the production possibility curves and showed how international trade could affect income distribution. The sector-specific factors could be interpreted as technological inputs (Leamer and Levinson, 1994).

Harrigan (2001: 26) describes the empirical research on the static Ricardian model "was quiescent for nearly forty years." Recently Eaton and Kortum (2002) and Golub and Hsieh (2000) tested these. Golub (1994) tested the Ricardo model by applying time series regressions, cross-section regressions and panel data regressions to export patterns of the G7-countries. Sectoral trade balances are used as the dependent variable and unit labour costs as the independent variable. Golub finds that over time the changes of unit labour costs explain changes of sectoral trade

balances, although the levels of these balances are sometimes difficult to reconcile with the levels of unit labour costs.

(ii)   Empirical tests of the Ricardian theory in South Africa

Various empirical tests have been undertaken in South Africa. Kohler and Holden (1992) were unable to come up with a systematic relationship between South African output per worker and relative South African export quantities. They ascribe this to insufficient data at the industry level (on productivity differences, cost differences, price differences, and international trade for different countries) to fully test the Ricardian model. They find a relationship between labour productivity and labour compensation, the exchange rate and the structure of South African trade flows, although not at industry level, and conclude that this "constituted evidence in support of comparative cost theory".

Golub (2000) assessed South Africa's international price and cost competitiveness, particularly with regard to labour costs, to determine the relationship between South African cost competitiveness and trade performance. He defines competitiveness narrowly, in Ricardian terms, as international cost and price competitiveness, paying particular attention to labour costs in manufacturing. Labour cost per unit of output or unit labour costs in manufacturing are presumed to be the underlying determinant of competitiveness. Unit labour cost is equal to the ratio of wages to labour productivity. Productivity was calculated as real value added per employee, using the manufacturing value-added deflator to deflate nominal value added, which was then converted to rand using the mean real exchange rate. Wages were defined here as total remuneration of labour, inclusive of non-cash fringe benefits, divided by number of employees. Wages were converted to rand using the market exchange rate. Golub (2000) evaluated both South Africa's relative and absolute competitiveness, i.e., changes over time and levels of South Africa's prices and costs compared with those of other countries. He found South African labour to be competitive vis-à-vis most industrial countries, but uncompetitive relative to most developing countries. The improvement in international competitiveness using the Ricardian model, has been an important determinant of manufactured exports in the 1990s.

Golub and Edwards (2002) put forward the following cross-section equations to estimate relationships between relative unit labour costs (RULC) and export performance:

$$\text{Log} (X_{isa}/X_{ij}) = \alpha_{saj1} + \beta_{saj1}\text{rulc}_{isaj} + \varepsilon_{isaj1} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \quad (1)$$

$$\text{Log} (X_{isaj}/M_{isaj}) = \alpha_{saj2} + \beta_{saj2}\text{rulc}_{isaj} + \varepsilon_{isaj2} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \quad (2)$$

$$\text{Log } (X_{isaj}/M_{isaj}) = \alpha_{saj3} + \beta_{saj3}\text{rprod}_{isaj} + \beta_{saj3}\text{rwage}_{isaj} + \varepsilon_{isaj3} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{ (3)}$$

$$\text{Log } (X_{isaj}) = \alpha_{saj4} + \beta_{saj4l}\text{rulc}_{isaj} + \varepsilon_{isaj4}\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{ (4)}$$

$$\text{Log } (X_{isaj}) = \alpha_{saj5} + \beta_{saj5}\text{rprod}_{isaj} + \beta_{saj5}\text{rwage}_{isaj} + \varepsilon_{isaj5}\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{ (5)}$$

where:

- X_{isa} is South African exports of commodity i;

- X_{ij} is total exports of country j's exports of commodity $i;$

- M_{isaj} is South Africa's imports of commodity i from country $j;$

- rulc $_{isaj}$ is the relative unit labour cost (in logs) in South Africa relative to those in j for commodity $i;$

- rwage $_{isaj}$ is the relative unit wage rate (in logs) in South Africa relative to those in j for commodity $i;$ and

- rprod $_{isaj}$ is the relative productivity (in logs) in South Africa relative to those in j for commodity $i.$

Equation 1 is similar to the tests of MacDougall (1951) and Stern (1962) and uses the ratio of total trade as the dependent variable. Equations 2 and 3 use bilateral trade balances as the dependent variable, while equations 4 and 5 stress the structure of exports and not net trade. They find that although the ratio of wages relative to productivity has improved since the 1970s, it is much larger in developed than developing countries. No clear pattern emerges as regards the nature of South Africa's comparative advantage, although exports do respond strongly to relative unit labour costs.

Kohler and Bruce-Brand (2002) too, find robust support for the Ricardian model. They explain changes in the competitiveness of South African manufactures in terms of cross-country differences in labour requirements. They also recognise the shortcomings of focusing on a single factor of production, (i.e. labour), such as overlooking the important contributions to overall productivity of capital, technology and factors such as scale economies. The results confirm that labour costs per unit of output were a highly significant determinant of trade competitiveness in South African manufactures from 1970 to 2000. Strategically, their findings indicate that organised labour, business and government policy-makers need to restrain wages unless accompanied by higher productivity. High wages jeopardise South Africa's ability to compete internationally in manufactured goods, especially in labour intensive sectors where the wage bill is a significant proportion of overall production costs.

(iii) Comments on the Ricardian tests

Leamer and Levinson (1994) contend that while the theory is simple, there are great difficulties in translating it into predictions. They argue that the Ricardian propositions stating that "the observed terms of trade are bounded between the comparative labor cost ratios of two countries," are theoretically robust. However, they are either not testable or "quite fragile." The proposition that a country exports the commodity in which it has a comparative labour cost advantage and imports the commodity in which it has a comparative disadvantage provides a link between trade patterns and relative labour costs that is "too sharp to be found in any real data set." Although the Ricardian model serves as a reminder that technological differences can be an important source of comparative advantage, the one-factor model is too simple for empirical verification.

(iv) Empirical tests of the Heckscher-Ohlin model

The goal of this model is to predict the pattern of trade in goods between the two countries, based on their differences in factor endowments. The original Heckscher-Ohlin theorem did not specify the functional form. The formulation of the Heckscher-Ohlin model with two countries, two goods, and two factors (or the 2×2×2 model) is often called the Heckscher-Ohlin-Samuelson (HOS) model, based on the work of Paul Samuelson who developed a mathematical model. The formal model can be specified either in terms of trade in factor services or in terms of trade in goods (commodities) (Leamer, 1984).

To test the commodity composition of trade, researchers regress some degree of trade performance on measures, either of sector intensity, or factor abundance. The importance of these variables in determining trade is then assessed. In both versions it is important to distinguish how to deal with non-tradables. The factor content of non-traded inputs could be an important determinant of comparative advantage. Deardorff (1984) argues that gross factor intensities are more important as they determine autarky prices.

Leontief (1953) was the first to test the Heckscher-Ohlin model empirically, followed by Leamer (1980), Bowen, Leamer and Sveikauskas (1987), Trefler (1993; 1995), and Davis and Weinstein (2001). Using US input-output tables to calculate labour-output and capital-output ratios for various sectors, Leontief calculated factor inputs for US exports and imports on final or last-stage products. Although the Heckscher-Ohlin model predicts that the USA, a capital rich-country, would export capital-intensive products and import labour-intensive products, tests proved the opposite.

Leontief postulated that US labour was more productive by a factor of three. Helpman (1999) poses the question: "Why would US workers be so much more productive, especially after controlling for capital employment?" Other critical comments made were that Leontief should have counted workers' skills as human capital, that the USA imported natural resource-intensive products, and that the study was conducted too soon after the war. Vanek (1968) later formalised this model (see Appendix 8), popularly known as the Heckscher-Ohlin-Vanek (HOV) model, establishing the linear relationship between trade and end-factor endowments and extending it to include multi-goods and multi-factors. Trefler (1993) ratified Leontief's claim that the USA was actually a labour-abundant country when factors are measured in productivity-equivalent, efficiency units.

After various empirical tests, it can be concluded that, unless the assumption of identical technologies across countries is dispensed with, the HOV model performs poorly. The earlier Ricardian model of allowing for technological differences, which also implies differences in factor prices across countries, seems more plausible. Recent research has extended the empirical investigation of the HOV model in new directions with more probable explanations for trade patterns. An example is the work by Davis and Weinstein (2001), who relax the assumptions of the HOV model and find that the theory holds allowing for:

- Different technologies across countries;
- Different factor prices across countries; and
- A more general structure of demand.

(v)   Heckscher-Ohlin tests with historic data

The Heckscher-Ohlin model has been tested using historic data. O'Rourke and Williamson (1999), using historic price data from 1565, found that the Heckscher-Ohlin model was not particularly useful in explaining trade during the mercantilist period. It did explain the global era that followed, particularly from the 1820s. These authors argue that economic models may not hold over time, and attribute this to the smaller role played by agriculture today. They find that commodity prices explain land price movements but have a smaller impact on wages. The strong correlation between relative commodity prices and the wage-rental ratio was "driven more by movements of the latter's denominator than by movements in its numerator."

(vi)   Empirical tests of the Heckscher-Ohlin theory in South Africa

Holden and Holden (1981) use Leontief-type tests, while Ariovich (1979) uses simple correlations between export shares and capital intensity to test the applicability of the Heckscher-Ohlin model

in South Africa. Tests using export shares were also carried out by: Standish and Galloway (1991), and Edwards and Golub (2002). These are useful in that the commodity pattern of trade reveals differences in non-price factors and the differences in relative costs.

More recently, Tsikata (1999), the International Labour Organisation (ILO, 1999) and the International Monetary Fund (IMF, 2000) examined South Africa's trade pattern. Tsikata examines only the pattern of exports rather than net trade, raising questions of valid interpretation. She also uses international classifications of skilled and unskilled labour, resource and capital intensity, which tend to distort the results because the classifications are done *a priori* and do not necessarily conform to the South African industrial structure. Aggregated exports also mask variations across trading partners. The ILO (1999) suffers from similar generalisations, although it incorporates South Africa's endowment of natural resources which explains the inherent capital intensity of South African exports. Using the cross-commodity approach and focusing on net exports, using both ordinary least squares and logit analysis, the IMF (2000) finds that although the coefficient on capital-labour ratio is positive to high- and middle-income partners but not low-income partners, it indicates that the higher the capital-labour ratio the greater the probability of South Africa being a net exporter. After controlling for capital used in resource-intensive sectors, the results remained the same, contesting the ILO's (1999) findings. The IMF (2000) uses both the factor content and commodity composition approaches to explain South Africa's trade patterns. The factor content of South Africa's trade in manufactured goods for 1989 and 1997 was found to be overwhelmingly more capital-intensive than both developed and developing countries. Although a net importer of skilled and unskilled labour, comparing net trade to consumption, the IMF holds that South Africa is endowed with unskilled labour. As would be expected, exports to high- and medium-income countries are more unskilled labour intensive, while to low-income countries they are more skilled labour intensive.

(vii) Explaining the paradox in South Africa

As already described in Chapter 3, there are a number of plausible explanations for South Africa's capital-intensive exports. South Africa has inherited a bias toward capital-intensive production techniques. Mining operations are inherently capital-intensive. Therefore any exports of these products or even exports with these products as inputs are going to be reflect the capital-intensive nature of the industry even though South Africa is a capital poor country. This natural tendency was aggravated by policy decisions. South Africa adopted an import substitution policy in 1925. This distorted production along comparative advantage lines. This was aggravated by apartheid policies that limited access to the natural labour pool. As described above, black education was also inadequate. Consequently labour productivity was lower than trading partners. Relatively

low interest rates and a strong and volatile currency led to investments in capital-intensive production techniques.

(viii) Conclusion

Maskus (1991) describes the empirical analysis of international trade models as a "crude art". Leamer and Levinson (1994: 1338) offer researchers the following advice:

> *The first message is: "Don't take trade theory too seriously." In practice, this means "Estimate, don't test." … Understand that theorems are neither true nor false. They are sometimes useful and sometimes misleading. If we approach a database with the contrary attitude hoping to determine the verity or falsity of a theorem, we may statistically "reject" the theory, but leave it completely unharmed nonetheless. After all, we already knew it wasn't literally true.*

> *"Estimate, don't test" is important advice, but it can be taken too far when empirical analysis is done, without benefit of a clear theoretical framework. Our second piece of advice points in the opposite direction: "Don't treat the theory too casually." In practice this means: "Work hard to make a clear and close link between the theory and the data." We are convinced from several notable failures that it is important to have clear linkages between the theory and the data if empirical results are to have any hope of having a lasting impact. High, partial correlations by themselves are not enough. We need a good story.*

Empirical analysis invariably means making uncomfortable choices among various functional forms as well as definitions of variables. A further complication is the availability, quality and comparability of international trade data. Even though various academics and international bodies have tried to simplify the process, researchers are frequently faced with a multitude of incompatible data.

According to Wood and Mayer (1998), the Heckscher-Ohlin theory cannot provide a complete explanation of trade patterns because it neglects to consider economies of scale, sectoral differences in efficiency, imperfect competition, government policies and transport costs. While the theory is useful for explaining some major features of trade patterns, a half-century of empirical work has failed to find simple amendments that allow it to provide a unified description of international trade data. Generally the empirical tests of the Heckscher-Ohlin theory produced poor results (Bowen *et al.,*1987) and caused the following three reactions:

- Rejection of the Heckscher-Ohlin model;
- Re-specification of the model and rationalisation of the data used in the tests; and
- The development of new trade models (Helpman & Krugman, 1985).

3.3   Toward new explanations of trade

The standard trade model is the outcome of classical and neoclassical economic thinking and is widely accepted as a canon of international economics.   However, the classical economist identified other determinants of trade. Their ideas were often not formalised and therefore not analysed, although many of these have emerged later and become important in both theoretical and applied economics. These ideas influence policy and are incorporated into trade models. Smith was responsible for identifying or expanding the productivity theory, vent-for-surplus and the role of government, and even alluded to the gravity model. After Ricardo and Torrens, economists considered other potential determinants such as the demonstration effect. These early economists also looked at ways in which comparative advantage could be created, and developed the infant industry argument. Clearly, trade theory has developed over time.

Classical and neoclassical theories predict that nations will specialise in the production of products in which they have a comparative advantage, and that trade will therefore be inter-industry. Although intra-industry trade (IIT) was recognised by Marshall in the 1920s, this phenomenon has increased recently and defies the neoclassic models' predictions. This, together with empirical tests, raised doubts concerning the validity of the Heckscher-Ohlin model. New theories were needed. Constant returns to scale and perfect competition and other assumptions were challenged. Irwin (2000: 1) comments:

> *Anyone who has dipped into "Interregional and International Trade" (1933) knows that Bertil Ohlin's theoretical vision was much broader than the standard two-factor, two-good model that has been handed down to us as the textbook Heckscher-Ohlin-Samuelson model of international trade. Ohlin not only set out a theory of trade based on factor proportions, but also investigated trade in relation to increasing returns to scale, economic geography, international factor movements, and a host of other topics that continue to be active areas of economic research.*

Since the introduction of the constant elasticity of substitution production function by Arrow *et al.,* (1961), the similarity of production conditions in different countries has also been questioned. Arrow *et al.,* (1961: 246) conclude:

> *Although we began our empirical work on the naive hypothesis that observations within a given industry but for different countries at about the same time can be taken as coming from a common production function, we find subsequently that this hypothesis cannot be maintained. But we get reasonably good results when we replace it by the weaker, but still meaningful, assumption that international differences in efficiency are approximately neutral in their incidence on capital and labour. A closer analysis of international differences in efficiency leads us*

to suggest that this factor may have much to do with the pattern of comparative advantage in international trade.

Faced with the problems of empirical verification, economists started looking at the demand side of trade rather than simply focusing on endowments and static technology. The dynamic aspects of technology, especially technology transfer, have received attention. The product life cycle hypothesis and technology theories of international trade were developed.

3.3.1 Demand-side explanations

Focusing on the supply-side and using limiting assumptions, neutralises demand as a determinant of trade. However, assumptions can also neutralise technology and resource endowments, in which case trade flows are accounted for by differences in tastes. To get a complete picture of trade composition and flow, demand (and supply) must be included. Demand can be analysed from two angles:

- Firstly, domestic demand and tastes contribute to forcing firms to upgrade production or develop new products, thereby gaining a competitive position in global markets; and

- Secondly, firms respond to global or international demand. These firms identify and exploit global market niches.[2]

(i)   Linder hypothesises

Linder (1961) argues that the principles determining the trade in commodities are not the same as those determining the trade in manufactured goods. Endowments affect trade in primary goods, but not necessarily trade in manufactured goods. Additionally, much trade takes place between the developed countries with similar factor endowments, which highlights the role of demand. Linder hypothesises that entrepreneurs respond to perceived demand and produce the manufactured goods demanded. The range of products manufactured in the home market therefore depends on home market demand, because the entrepreneur is more familiar with this than with foreign demand. Once successful in the domestic market the entrepreneur expands to countries with similar demand patterns. Linder predicts that production and trade depend on local entrepreneurial activity, international income patterns and demand (Markusen *et al.*, 1995).

South Africa has produced a few products that follow the Linder hypothesis. Foremost, is the development of mining equipment, particularly equipment used in deep mines. According to Campbell (2004) South Africa accounts for only some 0,5 per cent of the world's GDP, but

accounts for about 10 per cent of the world's mining GDP (excluding hydrocarbons and industrial minerals). Consequently, South Africa is an exporter of locally designed mining capital goods and other related complex machinery.

Similarly, many South Africans have swimming pools that need to be kept clean. This resulted in the development of the Kreepy Krauly automatic pool cleaner. This was followed by other designs and exports to countries such as Australia and the USA that have similar conditions.

(ii)   Production sharing

Although the product life cycle theory of international trade provides a substantial description of the evolution of multinational operations, it has been criticised because most products do not follow this pattern (Gilligan & Hird, 1986: 19). In the late 1970s, Drucker (1980: 63-64) proposed the production-sharing hypothesis as an alternative. In terms of this theory, as levels of education increase, people have higher expectations. This causes the number of unskilled workers in the developed countries to decrease. Manufacturers in developed countries turn to the less-developed countries with surplus and cheaper labour to partially produce the labour-intensive component of the product. The final product is manufactured partially in both countries, which reduces labour costs.

Firms merge and forge strategic alliances nationally and internationally because of high research and development costs, the speed of technological change and the merging tastes. This trend has proved popular amongst manufacturers of cars, biotechnology, computers, aero-engines and in the space industry.

3.3.2 Externalities

Economies of scale occur internally at firm level, or externally at industry or regional level. Similar firms in the same region learn techniques and management practices from each other. Externalities are not reflected in the market and arise from production processes (production externalities) or consumption activities (consumption externalities).

Venables and Limão (2001: 8) note that "externalities generated in one sector will typically only affect firms in a set of industries, perhaps the same industry or others that are in some way linked to it". Technological externalities are knowledge spillovers, such as the spillover of research and

---

[2]   For instance, China, with a very small Christian population, supplies over 80% of world demand for Christmas ornaments.

development activities, technical or managerial know-how, or the knowledge accumulated with experience (i.e. learning by doing).

Geographic clusters not only allow for the creation of externalities, but also facilitate face-to-face communication which stimulates new and innovative activities that depend on the successful transfer of complex and uncodifiable messages. This requires "closeness" between the receiver and the sender, including customers, suppliers and competitors. Technology such as the internet may facilitate communication and may help to maintain relationships, but it cannot establish deep and complex contacts required to coordinate new and innovative activities.

The geographical concentration of the industries is not always explained by the presence of natural resources. Venables and Limão (2001) argue that firms that cluster, operate more efficiently than firms that operate in isolation. The clusters support the establishment of specialist suppliers, labour market pooling and fostered knowledge spillovers (Krugman & Obstfeld, 2001). Porter (1990) has led research on geographic clustering and how it affects the competitiveness of nations and trade.

Externalities can be achieved when industries in a geographic area create a pooled labour market. It is advantageous to both employers and employees, and unemployment problems and labour shortages are less likely to occur. Specialist labour can move more freely from one firm to another if both firms are situated in close proximity. Firms benefit as specialist labour will gravitate to the area and enhance all aspects of their operations. Productivity is likely to increase and product development will be quicker and cheaper. Historically, manufacturing clusters have not played an important role in South Africa. The role of sectoral export councils in this regard is discussed in Chapter 6.

3.3.3 Increasing returns to scale

Since Smith's (1776, Book 1, Chapter I) famous example of the pin factory, economists have recognised the importance of increasing returns to scale (albeit through the division of labour) that occur when the average cost decreases as the quantity produced increases because of:

- Economies of specialisation;
- Indivisibilities in production; and
- Technological reasons (large machines are more efficient than small machines).

Although focusing on factor endowments, Ohlin points out the relevance of returns to scale to international trade (Elmslie, 2000). Heckscher and Ohlin (1991: 47) allude to other factors that could cause differences in production techniques: "The only possible exception would be

differences in the absolute size of markets, which might be due to differences in the size of units of production and thus to differences in technique", and propose that they are due to economies of scale. With indivisibility of certain factors, activities concentrate geographically. Ohlin (1933), however, plays down the relative importance of economies of scale.

Overall, the differences in factor supply determine the course of interregional trade – unless regions are small. The advantages of large-scale production are more in the nature of a subsidiary cause, carrying the division of labour and trade a little further than they would otherwise go, but without changing their main characteristics.

Scale economies are closely related to technology and innovation, although they are only efficiently attained when there are longer production runs per item. The role of technology and innovation is discussed below.

Hufbauer (1970) postulates that large countries produce and export products that require relatively large plants, while small countries would produce and export products that required smaller plants for optimum production. He finds that even though the correlation between the size of an economy and the importance of scale economy is high, it is not statistically significant (Magee, 1980).

South Africa has some way to go before most of its sectors can attain scale economies in production. Because of trade policy (import substitution), Dutch disease, sanctions and a politically driven need for a self-sufficient economy, firms could never specialise. Anecdotal evidence shows that in certain sectors, export demands are so overwhelming for firms with small production runs that they experience difficulties accommodating a large one-off order. Trade promotion organisations, particularly the South African Foreign Trade Organisation (SAFTO),[3] encouraged firms to exploit niche markets and concentrate on customers requiring low volumes – in other words, to supply differentiated products.

Rankin (2001) analyses the production technology applying the standard Cobb-Douglas production to firm-level data. He finds all coefficients for capital, labour, material inputs and indirect inputs significant at the 1 per cent level, with the sum of coefficients equal to 1,01 suggesting constant returns to scale. The University of Pretoria's South African Macro Economic Model (SAMEM) confirms this and even suggests that overall South Africa has decreasing returns to scale (Du Toit, 1999: 62).

---

[3]   This was confirmed during a discussion with Mrs Anne Moore, (General Manager of SAFTO from 1989 to 1998) on 3 May 2004.

3.3.4 Imperfect markets

Imperfect markets tend to arise in a small, protected economy. Monopolists and oligopolists, with no threat of foreign competition, have very little incentive to behave competitively. Pro-competitive gains from liberalisation may occur because distortion created by imperfect competition is endogenous. Import parity pricing often occurs when protected industries adjust their prices according to the level of protection and tend to "dump" their products on foreign markets. Trade exposes the monopolist to foreign competition and an elastic world supply at a lower price. So-called non-comparative advantage gains from trade are gains which are achievable irrespective of the existence or not of technological or endowment differences between countries. Similar gains could have been achieved by eliminating the monopoly distortion even in an autarky situation, and are referred to as pro-competitive gains or the import discipline hypothesis. Pro-competitive gains from trade will typically be accompanied by gains from trade because of exploitation of existing or latent comparative advantages.

(i)   Oligopoly

The oligopoly is a prevalent form of market organisation despite most of the classical models of trade assuming perfect competition. Oligopolies sell either homogeneous products (steel and aluminium) or differentiated ones (branded products such as motor cars, household products, etc.). These markets are usually characterised by internal economies of scale. While possible, entry into the industry is not easy. Each firm is large enough to affect prices, though none has an unassailable monopoly.

As there are only a few firms in an oligopolistic market, the action of each firm will affect the other firms. Analysis of oligopolies is complex and controversial because the firms' pricing policies are interdependent. Each oligopolist will assume the reaction of the consumer to a change in price while also anticipating the reaction of its competitors, thus second-guessing each other's strategies.

(ii)   Monopolistic competition and product differentiation

The monopolistic competition model assumes many firms, each producing a variety of differentiated products. Firms have freedom to enter and exit (Hummels & Levinsohn, 1993). The model has a distinguished record dating back to Chamberlin (1936) and Robinson (1933). Dixit and Stiglitz (1977) developed the benchmark monopolistic competition model, with further notable contributions by Krugman (1979; 1981), Lancaster (1975; 1979; 1980), Dixit and Norman (1980), Helpman (1981) and summarised in Helpman and Krugman (1985). Monopolistic

competition is a relatively simple way of modelling how economies of scale give rise to trade. Under monopolistic competition, two key assumptions are made:

- Firms are able to differentiate their products; and

- Firms take the prices charged by their rivals as given.

Under these conditions, it can be expected that a firm will sell more the larger the total demand for its industry's product and the higher the prices charged by its rivals. It will sell less the greater the number of rival firms and the higher its own price.

Even though products with similar characteristics, appearances, functions and applications can be grouped in the same statistical category, consumers often view them differently. Enterprises segment markets and emphasise the product features that a particular market finds most appealing. In extreme cases, the same product is sold to different markets with only labels and branding varying. Product differentiation occurs because:

- Marketing executives have persuaded consumers that different brands have certain characteristics that differentiate them from other similar products;

- Consumers may demand higher-quality products;

- Taste differences and cultural factors lead to different demand factors and firms that produce to meet the demand; and

- Consumers tend to have a love for variety, with a propensity to consume imported goods.

Product differentiation typically, but not necessarily, involves economies of scale that limit the range of products that can profitably be supported by a small domestic market. The larger the market or the smaller the economies of scale, the more brands become available. International trade increases the size of the market and allows countries to specialise in different brands. With consumers demanding a wide variety of goods, international trade leads to an overlap of products traded, with brand-specific economies of scale resulting in IIT.

The theory of monopolistic competition is more applicable to manufacturers of final products, although intermediary inputs and capital equipment can be differentiated. South Africa, a small country, seldom enjoys more than 1 per cent share of the world market. From discussions with various South African manufacturers and export consultants[4], especially small and medium exporters, price is important. However, aspects such as product features, packaging, design and

---

[4]   This was confirmed during a discussion with Mrs Anne Moore, (General Manager of SAFTO from 1989 to 1998) on 3 May 2004 and Mrs Linda Holtes on 10 May 2004.

service are integral to the success of export penetration, emphasising the role of product differentiation.

Since the mid-1970s, both private sector and trade promotion organisations in South Africa have targeted the development of niche markets. Manufacturers and exporters were encouraged to differentiate their products and develop unique characteristics that would enhance their chances of gaining a larger share of the foreign market. Until the end of the 1980s, this was undertaken by moral suasion rather than a planned, integrated approach. Export incentives were generic and aimed at reducing the anti-export bias that was created by import substitution policies. In the 1990s, GEIS encouraged the export of manufactured goods for final consumption, as the government believed that the beneficiation would contribute to employment and lead to increased export of differentiated products.

3.3.5 Technology

Although classical and early neoclassical interpretations of why international trade takes place are static, these schools understand the importance of technology but do not explain how it is acquired or how it changes over time.

(i)   Technology transfer

Grossman and Helpman (1995) point out, "ever since David Ricardo published his *Principles of Political Economy*, cross-country differences in technology have featured prominently in economists' explanations of international patterns of specialisation and trade. Yet, until quite recently, the formal trade-theory literature has focused almost exclusively on the effects of technological disparities without delving much into their causes." They lament the absence of formal literature on the determinants of national productivity levels and the relationship between trade and technology, which is due to a lack of tools rather than lack of interest.

The production of one type of final good often leads to the development of new manufacturing processes because of learning spillovers. Firms can learn from the experience of other firms and thus develop new products. The emergence of this type of horizontal cluster has been observed in geographical areas. The rapid decline in the costs of information dissemination since the late 20th century suggests that the geographical dimension of industrial clusters might become less important. The process nonetheless demonstrates that a vibrant manufacturing sector can emerge in formerly agricultural or mining economies because of foreign investment, input trade and learning spillovers across firms. Trade liberalisation, by reducing the costs of imported inputs, can lead to agglomeration effects that result in significant welfare and productivity gains for

developing countries. In contrast, protectionism might lead to much less dynamic industrialisation (Puga & Venables, 1999).

Eaton and Kortum (1996) point out that despite the fact that benefits of innovation are experienced broadly, measures of innovative activity show it to be concentrated in a small number of countries. International trade is one conduit through which the benefits of innovation in a country can flow. These authors developed a bilateral gravity model of technology and trade, and found that the gravity specifications derived from models of monopolistic competition were quite successful at capturing the volume and pattern of trade in manufactured goods among OECD countries. Patterns of trade are influenced by stocks of knowledge as predicted by the theory.

Posner (1961) describes the need for imitation and holds that a great deal of the trade between developed countries is due to the emergence of new products and new processes. The innovating firm has a temporary monopoly in world markets, protected by patents and copyrights. Trade occurs while the imitation gap (the time needed for the innovation to be produced elsewhere) exceeds the demand gap (the time taken for an invention to be demanded in other countries). Typically, the USA or another OECD country will develop and export high-technology products until foreign producers acquire the technology and capture these markets. The developer then has to develop newer technology. Posner does not explain the size of the technology gap nor why it occurs.

(ii)   Technology gap

Hufbauer (1966) proposes the "technological gap" trade theory following observations that more advanced countries typically produce and export technological products. Winters (1985: 52) gives three reasons why technology develops in certain countries:

- Institutions such as strong patent laws encourage research and development. This is because inventors will be prepared to invest their resources as long as they can see potential rewards. Favourable tax laws will also encourage research.

- A country is well endowed with factors required for research and development and the appropriate mix of scientists, engineers, equipment and finance.

- Inventors need a market for their output. They understand what the market requires and have the necessary institutional arrangements in place to sell products.

The technology gap theory is therefore related to the Heckscher-Ohlin model.

In the late 1960s, Vernon (1966: 190-207) proposed a new explanation of world trade and investment patterns based on the product life cycle concept. He contends that on an international level, products move through four distinct phases:

- Initially the product is developed and manufactured in a highly industrialised country for local consumption. The product is subsequently exported to other countries with similar markets.

- The second phase begins to emerge as the technology develops and is transferable. Companies in other countries begin manufacturing the product and are able to produce it cheaper than the original manufacturer, because of lower input costs.

- The third phase is characterised by foreign companies competing against exports of the original manufacturer and further eroding its markets. They begin to withdraw from selected markets or invest in foreign manufacturing capacity to regain sales.

- Finally, having established a strong presence in their home and export markets, foreign companies export to the country of origin. Vernon (1966) shows how firms in industrialised countries eventually find themselves being squeezed out of their domestic markets after having enjoyed a monopoly.

The life cycle hypothesis holds that the pattern of trade will differ between developed or industrialised and developing countries, predicting that developed countries will export non-standard products, or products that require advanced high technology. These countries will import products using older technologies.

Gruber ,Mehta and Vernon (1967) find a strong correlation between exports and research and development, and reason that research and development create the temporary comparative advantage. Using research and development spending as a proxy, these authors confirm the validity of the product cycle and technology gap theories.

The question remains, why do developed countries undertake research and development? Maneschi (1998: x) bemoans the fact that the research is "somewhat ad hoc in nature" and does not "offer an overarching paradigm that could be regarded as a valid alternative to the traditional one." Porter (1990), on the other hand, points out that although the notion of a product life cycle left many questions unanswered, it represents the beginnings of a dynamic market. He suggests that demanding customers in the home market influence innovation.

Product cycle theory emphasises standardisation. Comparative advantage moves from the nations richly endowed with research and development facilities to nations with a comparative advantage in standardised products. Nations without research and development facilities should exploit and

develop comparative advantages in producing standardised products. In most cases, this will include an efficient and effective workforce that can adapt to new technology.

The product cycle theory could explain a component of South Africa's trade, as it has both Third World and First World components. There is some evidence to support this view. For instance, in the automotive sector, Nissan continues to build the 1400 half-ton utility vehicle, despite aging technology. This vehicle is not included in Nissan exports to the rest of Africa. Further evidence is that various component manufacturers continue to make and export parts for vehicles that have been discontinued. Dies and tools are imported after the original equipment manufacturer (OEM) is no longer able to sustain the low volumes. South Africa's comparative advantage in short-run production merely explains this phenomenon, rather than supporting the hypothesis.

(iii)   Innovation

Innovating firms spend a great deal of resources developing new ideas. Magee (1980) points out that the ideas are public goods and anyone who can figure out how to use them, may do so without reducing the use to others. However, the unauthorised use of new ideas reduces the profitability to the innovator. Magee defines the appropriability problem as the conflict between the social use of the good and the private returns to the innovator. This suggests that it is more efficient to transfer high technology globally inside a single firm rather than through the market. This reduces the risk of the technology being copied or stolen, and explains why many multinational corporations produce and market high-technology products and services.

Magee (1980) points out that even though the innovator may have a monopoly in the early stages of the product's life cycle, it should sell below the monopoly price. The monopoly price encourages emulation of production and reduces the present value of future production. As appropriability erodes the market potential, the innovator should steadily cut the price until the competitive level is reached.

Increasingly the production of goods and services, and especially the development of new products, requires specialist inputs. These inputs include machinery and specialised services. An individual company does not justify the creation of a support sector. An industry cluster solves this problem. Collectively, the firms provide a market large enough to sustain the range of specialist inputs required. Inputs become cheaper and more easily available as firms compete to provide them. Firms can outsource and focus on what they are best able to do.

Various studies (Clerides, Lach & Tybout, 1995; Bernard & Jensen, 1999) show that industrious firms often become exporters, and although this has a positive influence on their output, it does

not improve their productivity. Improvements in exporter productivity do not seem to be the result of any learning-by-exporting.

Empirical evidence seems to back up the technological explanations for trade. Most of the early work was informal, relying more on qualitative than quantitative analysis. Formal econometric models on technology theories require proxies for technological leadership and usually include either research and development (input) or patents (output). These are then regressed with a number of other variables to explain the country's composition of particular trade goods.

Technology gap theory seems to place emphasis on imitation while the product cycle emphasises standardisation. Despite the work done in this field, Winters (1981: 45) laments: "We are not yet in a position to develop very formal modes of technology akin to the Heckscher-Ohlin theory."

Porter (1990) poses the following questions:

- What happens when a demand originates simultaneously in different nations?
- Why do nations with a more slowly developing or small home market for a certain product often emerge as world leaders?
- Why is innovation continuous and not a once-and-for-all event followed by standardisation of technology as the product cycle theory implies?
- Why does the inevitable loss of advantage in Vernon's theory not take place in many industries? and
- Why do some nations' firms sustain advantage in an industry and others not?

Industries can be grouped to reflect varying determinants of success in export markets: traditional, knowledge-intensive, resource-intensive, scale mass production-sensitive, and high-technology or science-based industries. Unfortunately, these categories are blurred. Traditional industries now have access to use technological advances in both manufacture and distribution.

3.4   Geography

Geography plays an important role in trade. Historically South Africa has traded with Europe, North America and Asia, which are not physically close. Although distance between the countries and bilateral trade arrangement determine trade resistance, other geographic factors also play a role. Natural barriers such as mountains will inhibit trade even though countries are physically close. Countries without access to ports face additional man-made barriers such as customs. A gravity model using the three groups of variables identified above will show areas with trade potential. Some of these are likely to remain relatively unchanged, such as cultural difference,

language, distance and other geographical variables. Leamer (1974) identifies three groups of variables that determine trade patterns:

i.   State of development variables – GDP and population;
ii.   Resistance variables – tariffs, distance from market; and
iii.   Resource variables – capital intensity, research and development, education levels and electricity use (a proxy for industrial sophistication).

The gravity model will be dealt with in detail in Chapter 6.

3.5   Investment

Generally, cross-border supply can take three forms: arms-length trade, sale or licensing of technology, and direct production. The former two have been discussed above. In the case of direct production, the firm establishes (or acquires) an overseas affiliate to produce the commodity locally and then sell it. Foreign direct investment (FDI) occurs when an investor in one country acquires control of a business entity in another country through equity participation rather than portfolio investment in foreign stocks and other financial instruments.

Typically, foreign direct investment is a triangular process where firms:

- Own assets that can be profitably exploited on a comparatively large scale;

- Have some advantage by dividing production across countries, rather than simply exporting from the home country; and

- Acquire greater profits from exploiting in-house assets than from licensing the assets to foreign firms.

Firms with assets such as intellectual property like technology and brand names, organisational and managerial skills, and marketing networks with very small or no marginal costs can invest these in foreign countries. These foreign firms will have advantages over the domestic competition and will enjoy lower average costs.

According to research done by UNCTAD (2003), FDI is undertaken to generate profit and is motivated by:

- Resource seeking – the presence of natural resources. These firms need to locate near raw materials. Investment occurs when the cost of shipping the raw materials exceeds the cost of shipping finished goods (e.g. steel, basic chemicals);

- Market seeking – access to individual or regional markets. Investment occurs when industries produce heavier goods that are difficult to export (e.g. Coca-Cola);

- Efficiency seeking – more favourable cost base for operations; and

- Strategic asset seeking – protecting or advancing the global competitive advantage.

South Africa's diversified manufacturing base offers limited opportunities for market capture. Income disparities between the SADC countries further limit market seeking investment opportunities. Free trade agreements and other market-access measures can attract investment. Efficiency-seeking FDI tends to be located in countries with skilled, disciplined workforces and good technological and physical infrastructure. This, as well as a rationalisation in worldwide production, has driven investments in the automotive sector in the past few years. BMW, for instance, claims that its workforce is the reason for expanding output. Although incentives can make the difference at the margin between otherwise similar locations for this kind of investment, there is a possibility of a "race to the bottom" as countries find themselves subject to "incentive inflation."

FDI is still low in South Africa, with foreign firms tending to supply the local South African market and, to a lesser extent, the African market. In certain sectors, particularly the automotive sector, foreign firms are investing to exploit foreign markets. Distances from the major foreign markets will continue to hamper trade-induced FDI in South Africa, with foreign firms opting to invest closer to major markets.

With limited data on FDI in South Africa, particularly on a sectoral level, it is difficult to estimate the impact that such investment has on trade. Anecdotal evidence, particularly in the automotive sector, suggests that foreign ownership is a determinant of trade in South Africa. In his estimation of the South African production function, Rankin (2001) found that the only other significant variable, apart from capital, labour, material inputs and indirect inputs, was foreign ownership. In the South African context, this suggests that firms with some foreign ownership produce more output than those with only domestic owners.

3.6   Infant industry and factor creation

Hamilton (1791), List (1827; 1856) and Rae (1834) provide intellectual arguments for the protection of industry, for a limited time, to exploit potential comparative advantage that would otherwise not occur. This argument opened the door for governments (including ostensibly liberal governments) to impose restrictions to develop indigenous industries. Regrettably, the infants seldom grow up and protection continues without consumers realising the impact it has on their welfare. Other industries that are affected by the increased price of inputs are seldom as well organised or as influential, and are unable to address the issue. Although the nominal protection may be low, the effective rate of protection and the anti-export bias it creates limit the export potential of other sections that may enjoy natural comparative advantages.

Although South Africa used import substitution policies based on infant industry arguments, the impact on the volume of trade is ambiguous. It is true that exports of steel are higher than they would have been without these policies, but the impact on both forward and backward linkages is negative. The higher price of intermediary inputs depresses competitiveness.

Creating and upgrading factors is a traditional and important role of government. These factors include skilled human resources, basic scientific knowledge, economic information and infrastructure. Government's role in factor creation is justified by externalities. Failure to meet these obligations results in uncompetitive industries.

Human resource development is a vital contributor to factor creation in the modern international market. It is no longer sufficient simply to provide basic literacy. The government has a role in ensuring that the needs of modern industry are met. Education and training must have some practical orientation. Even government's policy towards emigration and immigration can contribute to a pool of specialised skills.

Technology is critical to improving efficiency, commanding higher prices through data quality, penetrating new industries and segments, and stimulating productivity growth. Because of externalities achieved through technology improvement, improvements in science and technology through research and development are acknowledged as a role of government. Virtually every developed nation has policies that encourage research. Porter (1990: 631) stresses the importance of an innovation policy rather than just a science and technology policy.

Trade is dependent on a nation's infrastructure. Transportation, logistics, information technology and telecommunications are necessary (although not sufficient) conditions for trade to occur. Although governments have played the leading role in providing infrastructure in the past, privatisation is increasingly occurring.

The supply and cost of capital are affected by government policy and actions. The supply of capital is influenced by, among other things, the size of government surpluses or deficits, personal savings and foreign capital flows. Mechanisms for allocating capital are also influenced by policy and regulation.

In a world dominated by information and communications technology, the role of information is undisputed. Not only is the availability of information important, but also the quality. Too much information – especially when it is contradictory – can have negative repercussions.   The government plays a prominent role in expanding the stock of information to firms. It also plays an important part in creating channels to distribute or disseminate the information. By supplying information, government can provide signals to industry.

Government has an important task in developing exports by encouraging an international outlook. Most governments do this through the provision and dissemination of information on foreign markets and technical information.

Government procurement can either enhance or weaken domestic industries. Domestic firms may be given preference over foreign firms. This will limit imports, but as discussed above, firms that exploit economies of scale from government business are able to penetrate foreign markets. Government affects the demand conditions of a nation and government spending can be used to improve the quality of domestic demand. Porter (1990: 645) highlights the following measures government can take in improving long-term competitive advantage: stimulating early demand, encouraging demanding and sophisticated buyers, and implementing procurement that reflects international needs and facilitates innovation, and encouraging competition.

Governments are responsible for trade negotiations and securing market access. This includes multilateral negotiations that level the playing field but also increasingly includes economic integration, which Balassa (1987: 35) describes as being "a process and a state of affairs". Matthews (1987) states that there are many forms of economic integration, ranging from an agreement by two or more nations to decrease bilateral impediments, to the complete integration of two or more independent economies. New market access into Europe, Africa and America is increasing the country's volume of trade.

3.7   Strategic trade theory

Strategic trade theory investigates the implications of market imperfections for government policy. Strategic interactions among firms occur when a change in the behaviour of one firm leads to a change in the optimal behaviour of another firm (a strategic response). By affecting the behaviour of firms, "strategic trade policy" can influence domestic and international markets by altering the strategic relationship between firms. Although frowned upon, governments do provide export incentives. Using these arguments and choosing an optimal import tariff or subsidy, the government can affect the strategic game played by firms in international markets to the advantage of domestic firms. The focus of strategic trade policy literature is normative. It shows how government intervention, in the presence of imperfect markets, can influence trading patterns. The literature contains contradictory results, depending on the type of market structure and the form of competitive rivalry. An optimal tariff can become an optimal subsidy or tax if one assumes that firms, instead of being Cournot players in the international markets (Brander & Spencer, 1985), would compete on prices (see Eaton & Grossman, 1986). Strategic trade policy provides intellectual arguments justifying a government subsidy on exports, allowing the government to change the reaction function of the firm.

The government's intervention is influenced by interest group pressures that are very likely to capture any benefits and may cause excessive intervention that benefits a small group of producers at the expense of a large number of non-represented agents. Political economy concerns and the difficulty of formulating useful interventions create a new case for free trade in a world of strategic trade policy.
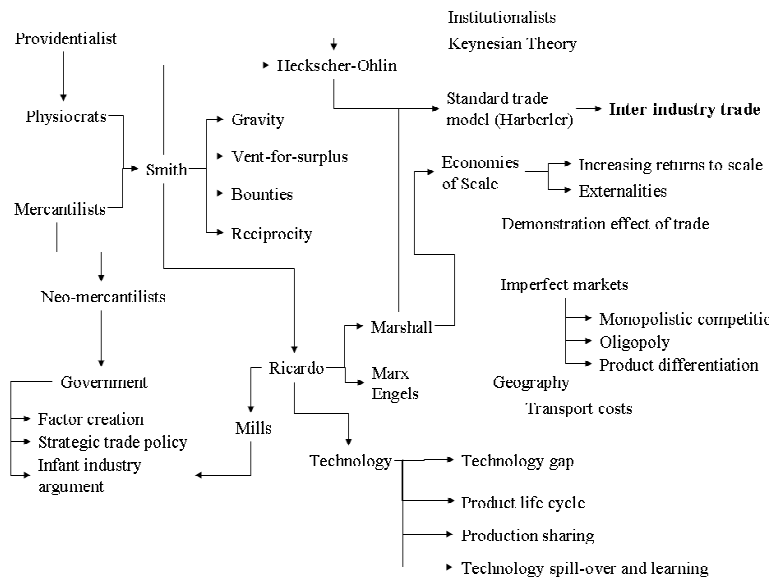
Winters (2000) points out that strategic trade policy has been abandoned by serious trade policy academics even though it holds elegant academic credentials. Its results are highly sensitive to specific assumptions about firm behaviour and its applicability, especially to developing countries, is questionable.

3.8   Summary and categorisation of theories

As the various trade theories developed over time, each emphasised certain determinants. Nevertheless they are interlinked. Figure 1 shows the origin and interaction between the main theories. Trade is a complex outcome of many factors operating simultaneously. Collectively, international trade theories show why we trade, highlight the benefits of trade; predict what we trade; and with whom we trade.

Early theorists focussed on the supply-side, or border-in factors. Figure 2 shows a composite of the various supply-side theories of trade. By simplifying the real world, the models do contribute to our understanding of trade. However, too many assumptions are made and therefore too many factors are ignored to make theories relevant in today's policy making environment. Ricardo, by focusing only on labour, ignored many factors that contribute to its productivity. In their analysis of trade, Heckscher-Ohlin miss many relevant factors. Nevertheless, although Figure 2 ignores the role of government and the impact tax, subsidies and import duties have on the cost of production, government does have a role in providing labour with skills. These skills can be generic, but more attention should be given to sector specific skills that will contribute to industries gaining or entrenching their comparative advantage.
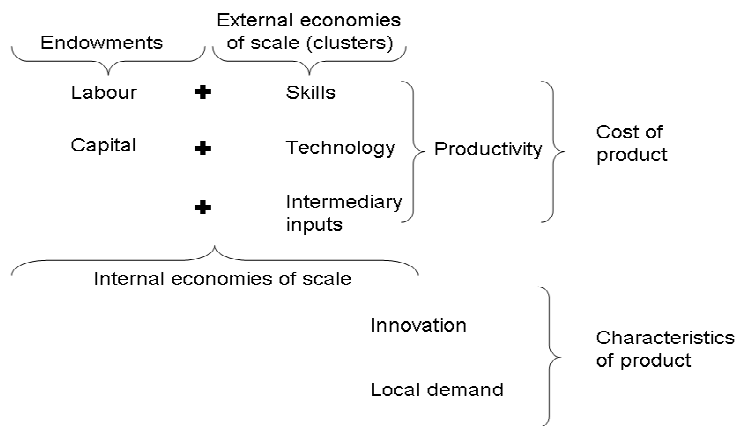
Figure 1 Origin and interaction between the various theories

Sources: Own compilation

The cost and characteristics of the product jointly determine the competitiveness of the firm in the international market, the supply-side determinants and the ability to export.

Figure 2   Various supply-side factors determining trade

Source: Own compilation.

The cost of producing the product has a direct influence on the products price and the firms' profit margin. Many factors that contribute to the cost can be influenced by policy. Similarly, policy can influence innovation, local demand and the characteristics of the product.

3.9   Policy implications

The government's role in influencing the composition and direction of trade has been recognised since ancient and medieval traders paid tolls for peaceful entry into foreign lands. Local, provincial and central government actions provide the general conditions under which entrepreneurs operate, and official steps to stimulate the economic development are taken on various fronts.

Table 2   Emphases of the major theories

Theory	Description of natural trade			Prescription of trade relations			
	How much is traded?	What products are trade?	With whom does trade take place?	Should government control trade?	How much should be traded?	What products should be traded?	With whom should trade take place?
Supply-side							
Providentialists							
Physiocrats							
Mercantilists				yes	x	x	x
Neo-mercantilists				yes	x		
Absolute advantage		x		no		x	
Ricardo		x		no		x	
Heckscher-Ohlin		x		no		x	
Economies of scale							
Strategic trade policy		x		yes		x	
Factor proportions		x	x				
Demand-side							
Country size	x	x					
Product life cycle		x	x				
Country similarity		x	x				
Dependence				yes		x	x
Composite determinants							
Porters' diamond		x					

Source: Adapted from Daniels and Radebaugh (2000: 161).

Government also regulates various industrial matters such as labour, working conditions, labour disputes, safety, and wage determinations. The provision of essential services (road, rail, air and sea transport, water supply, etc.) serves as a stimulant (South Africa, 1971). Table 2 shows the particular points that each theory highlights.

From a policy perspective, there are a few obvious lessons. Firstly, a country that wants to increase its welfare should trade. Some industries will be competitive and will therefore export, while products that cannot be produced competitively should be imported. Although there are a

few exceptions, free trade is the best policy option to maximise a country's welfare. Countries should adopt a neutral trade policy; neither supporting import substitution nor creating a pro-export bias.

Although the role of government may not be as prominent or important as technology or endowment differences, domestic distortions caused by government policy are a determinant of trade.   Porter (1990: 619) claims: "Government plays a prominent role in international competition, but it is a different one than is commonly supposed." He continues: "Governments do not control national competitive advantage; they can only influence it." Government achieves its policies through taxes and subsidies. Tariffs directly limit trade, with a wide range of government policies, or non-tariff barriers, influencing trade. These include health, safety and environmental policies and regulations. Government also has an influence, especially in the short term, on the value of a nation's currency through its monetary policy. This has a direct impact on the short-term determinants of trade. Devaluation increases the price of imported goods and increases the price of exported goods in terms of domestic currency. Porter (1990: 620) argues that the government should only act in areas where firms are unable to act. These include trade policy and where externalities cause the firms to under invest. Examples include general education, environmental quality, and some types of research and development that can boost productivity in many industries. In democratic countries, elections are held regularly and politicians are preoccupied with short-term economic fluctuations. Often policies tend to focus on short-term solutions, which include improving the trade balance through intervention, while ignoring desirable policies, such as factor creation, competition policy and upgrading demand quality. Industrial clusters have been highlighted as a determinant of trade. They are frequently concentrated in a city or region. The national government does play a role in upgrading industry; however, the role of the provincial (state) or local government is potentially as great or greater (Porter, 1990: 622).

3.10   Conclusion

This chapter has focussed on various trade theories and highlighted the major determinants of trade. Most of the theories predict that countries will specialise in products in which they have a comparative advantage, whether this has been acquired or not. Countries need to exploit their comparative advantages in order to maximise their welfare. Trade occurs only when at least one of the parties improves their position. How they perceive the improved position will depend on how they see their value improving. This, in turn, will be determined by how the value is created.

This chapter also studied the early determinants of international trade, beginning with the providentialists and the mercantilists. The providentialists believed that certain endowments were

valuable and that they were given by a deity. The physiocrats believed that wealth was created from the soil, or earth. Mercantilists believed that value was created by international exchange, which increased the local amount of precious metals. Ricardo showed how labour, combined with technology, could be used to create value. In comparison, the Heckscher-Ohlin model instead showed how factor endowments form the basis for trade. While this may be fine in theory, it has performed very poorly empirically and "is hopelessly inadequate as an explanation for historical or modern trade patterns unless we allow for technological differences across countries" (Feenstra, 2002: 11). This led to the development of the new trade theories that looked at various demand factors, market structures, externalities, and economies of scale.

Each of these theories implies different policy interventions. Despite the fact that these theories (particularly the neoclassical theories) predict specialisation, there has been an increase in intra-industry trade. Chapter 4 will adapt techniques used to measure this phenomenon and attempt to categorise products according to the input that affects them most. This will assist policy-makers to maximise trade and welfare benefits by using available scarce resources optimally.