AN EXTENSIVE EXPLORATION OF THEORIES OF FOREIGN DIRECT INVESTMENT

Patricia Lindelwa Makoni*

Abstract

The purpose of this study was to identify and examine the key foreign direct investment theories. The history and origins of FDI theories were considered, prior to dwelling in-depth on the theories themselves. FDI theories were classified under macroeconomic and microeconomic perspectives. Macroeconomic FDI theories emphasize country-specific factors, and are more aligned to trade and international economics, whereas microeconomic FDI theories are firm-specific, relate to ownership and internalisation benefits and lean towards an industrial economics, market imperfections bias. FDI theories are fairly complex to explain and apply. This paper is purely qualitative in nature, and attempted to explain the different FDI theories by providing an analysis of the key theories used in many scholarly works**.

Keywords: Foreign Direct Investment (FDI), Product Life Cycle Theory (PLC), OLI, Investment Development Path Theory (IDP), Multinational Corporations (MNCs), Imperfect Markets, Eclectic Paradigm

*Department of Finance, Banking and Risk Management at the University of South Africa (UNISA), P.O. Box 392, UNISA, 0003, South Africa
Tel: +27767538234

**This is article is derived from the author’s PhD work-in-progress

1 Introduction

According to the UNCTAD (2012), private capital flows consist of foreign direct investment (FDI), foreign portfolio investment (FPI), and other investment such as international banking flows and loans. An increase in international capital flows has resulted in faster financial globalisation than trade globalisation. As such, it has become more imperative to understand the underlying theories which help to explain this growth and movement in capital flows, mainly from the investor’s perspective. Our focus will be on FDI as it has been the dominant capital flow, especially amongst developing countries.

This paper therefore presents a theoretical perspective of FDI. The first section gives an overview of FDI definitions. The second section discusses the historical background and the origins of FDI theories, while the third section gives a classification of FDI theories. The fourth section presents the macroeconomic FDI theories, followed by the microeconomic ones. The final section of this article gives a concluding summary to the study.

2 Definitions of Foreign Direct Investment

Foreign Direct Investment is defined as international investment made by one economy’s resident entity, in the business operations of an entity resident in a different economy, with the intention of establishing a lasting interest (International Monetary Fund (IMF), 1993). According to the World Trade Organisation (1996), foreign direct investment (FDI) occurs when an investor based in one country (the home country) acquires an asset in another country (the host country) with the intent to manage that asset. The management dimension is what distinguishes FDI from portfolio investment in foreign stocks, bonds and other financial instruments. Alternatively, FDI can be considered as the ownership of 10 percent or more of the ordinary shares or voting stock of an enterprise which is usually considered to indicate ‘significant influence’ by an investor (IMF, 2000). This however differs from country to country and can even be determined by their policies, some of which restrict the levels of shareholdings of foreigners in local firms.

According to the World Bank (2004), Foreign Direct Investment is that foreign investment that establishes a lasting interest in or effective (active) management control over an enterprise. In its publication on The Benchmark Definition of FDI, the OECD (2008), defined FDI as the net inflows of investment undertaken to acquire a lasting management interest (10% or more of the voting stock) in a firm conducting business in any other economy but the investor’s home country. Emphasis is also placed on the fact that the 10% threshold commonly referred to is recommended to ensure statistical consistency across countries. For
investment to qualify as FDI, emphasis is placed on the fact that the investor must meet the 10% voting share threshold commonly referred to, which as the recommended mainly to ensure statistical consistency across countries (UNCTAD, 2009). Lipsey, Feenstra, Hahn and Hatsopoulos (1999) had earlier commented that this “lasting interest” implies the existence of a long-term relationship between the direct investor and the firm, as well as a significant degree of influence on the management of the firm.

3 The history and origins of FDI theories

The origins of FDI are not fully understood. Although there are many schools of thoughts which have been used to explain this phenomenon, there is still no consensus on any superior or general theory of FDI.

FDI theory dates as far back as the early work of Smith (1776) [as cited in Smith, 1937] and Ricardo (1817), and was related to international specialization of production. In Smith’s theory of absolute advantage, he explained that trade between two nations will occur if one country is able to produce and export goods using a given amount of capital and labour, more than its closest competitor (absolute advantage). However, Smith’s theory did not explain how trade arose between countries where one country was not in the business of production. It is then that the work of Ricardo (1817) emerged, to explain FDI using the theory of comparative advantage. Ricardo was more interested in international factor movements as he was of the opinion that labour and capital were mobile domestically but not across borders. His theory was however flawed because it was based on the assumptions of two countries, two products and perfect factor mobility, but still did not justify international capital movements. This is therefore in direct contrast to the notion that, in a world typified by perfect competition, FDI would not exist anyway (Kindleberger, 1969). According to Denisia (2010), if markets were efficient, with no barriers to trade or competition; international trade would be the only mode of participation in the global markets. It is against this background that when Hymer (1966) published his 1960 thesis, he laid the foundation for other authors to come up with more plausible theories of FDI. In his arguments, he found that FDI was motivated by the need to reduce or eliminate international competition among firms, as well as Multi-National Corporations’ (MNCs) wishes to increase their returns gained from using special advantages.

Mundell (1957) came up with a 2-sector model of international capital flows whereby capital flows were considered to be a substitute to international trade, resulting in factor price equalisation between countries. Mundell (1957) extended Ricardo’s theory of comparative advantage by developing a model encompassing two countries, two products, two factors of productions and two identical production functions in both countries (Denisia, 2010). However, Mundell’s model considered more short term, international portfolio type of investments rather than FDI, and therefore could not explain international production through FDI. Many of the earlier theories were based mainly on the U.S and Europe. To remedy the shortcomings of Mundell’s model, Kojima and Ozawa (1984) contextualised their model in Japan, and advanced an argument that FDI occurs if a country has comparative disadvantage in producing one product, while international trade depends on comparative advantage.

The emergence and trend of post-Second World War investments (a shift from exporting to FDI) made by US firms to Western European countries between 1950 and 1970 can be explained using Vernon’s (1966) product life cycle (PLC) theory. According to his theory, firms go through four production cycles: innovation, growth, maturity and decline. The underlying principles of this theory were technological innovation and market expansion; hence, while technology ensures the conceptualisation and development of a new product, the market size influenced the extent and type of international trade. In the initial stage, new products are invented, produced and sold in the internal markets. If the product is successful, production increases, new markets are penetrated and export develops. This is the transition from growth to maturity. It is also during this maturity phase that competitors emerge, and the product originator then sets up a production facility in the foreign market country to meet growing demand. Product standardisation occurs and incremental investment is then directed to any global site which offers the lowest input costs. After that, the product is exported back to the initial innovation country (exporter becomes importer as per the PLC) where it is eventually phased out, and the PLC starts all over again with the innovation of yet another product, since to emerge from the decline phase, the firm must be innovative again (Nayak & Choudhury, 2014). This is precisely what transpired when European firms began imitating the American products being exported to them; US firms had to set up production infrastructure in the local markets in order to maintain their market shares (Denisia, 2010).

Like other FDI theories, the PLC theory has its limitations. Primarily as pointed out by Boddeuwyn (1985), the product life cycle is but just a theory because it was not tested empirically. The PLC theory also does not take into account all FDI determinants, in that it, for example, only explains the location aspects of manufacturing infrastructure but not their ownership (e.g. manufacturing under licence or set up subsidiaries). The theory is a simplified decision-making process, which assumes a smooth-sailing, sequential journey with no obstacles, and is more applicable to industries that use technology for its innovation (Buckley & Casson, 1976). The PLC...
theory was further criticised for its failure to explain why it is profitable for a firm to pursue FDI rather than maintain its exporting strategy, nor the timing of the move to invest internationally (Nayak & Choudhury, 2014).

According to Bodewyn (1983), in the early 1980s, a cohort of researchers such as Casson (1979), Calvet (1981), Grosse (1985) and Rugman (1980) put forth their own versions of FDI theories. Although some of these researchers made a concerted effort to incorporate capital, location, industrial organization, growth of the firm, market failure, foreign exchange parity, investment portfolio and product lifecycle theories into one whole theory to attempt to explain the motives and patterns of FDI, most credit is given to Dunning’s eclectic paradigm (theory) of international production (Boddewyn, 1983). The best-known theory of FDI is Dunning’s 1977 Eclectic Paradigm in which he states that FDI occurs under different scenarios of ownership, locational and internalization advantages (OLI). This theory will be discussed in detail later, as it will be compared to more recent theories of FDI. It is for the above-discussed reasons that today, Popovici and Calin (2014) concluded that FDI theory is based on three integrative theories – the theory of international capital market, the firm theory and the theory of international trade. As such, it further necessitates the examining of FDI theories from two economic perspectives: the macroeconomic and the microeconomic views on FDI.

4 Classifying FDI theories

According to Denisia (2010), the macroeconomic perspective on FDI is that FDI itself is a type of cross-border capital flow, between home and host countries, and is captured in the balance of payments statement of countries, with the variable of interest being capital flows and stocks, revenues obtained from such investments. The microeconomic perspective on the other hand relates to the motives for investments across national boundaries, as seen from the investor’s point of view. This follows on from Shin (1998) who critically reviewed existing theories of FDI and cited various scholars who classified FDI theories in a similar manner. Petrochilos (1983) classified macroeconomic FDI decisions based on variables which determine the investment decision (as cited in Shin, 1998, p.186), and mimic corporate investment behaviour, under the importance of the market size of the host country as measured by the GDP, growth of the market size, factor prices, interest rates, profitability and investor protection against tariffs and other such elements. According to him, the microeconomic determinants, drawn from the theory of industrial organisation (theory of the firm), are more concerned with firm and industry features which would give MNCs certain advantages over domestic firms. Caves (1971) gives examples of these features as including product differentiation, technology, the product life cycle and the size of the firm as measured by its sales or the value of its assets. Another scholar who classified FDI theories along the macro and micro economic views was Gray (1981). According to him, macroeconomic FDI theories emphasize country-specific factors, and are more aligned to trade and international economics, whereas microeconomic FDI theories are firm-specific, relate to ownership and internalisation benefits and lean towards an industrial economics, market imperfections bias.

5 Macroeconomic FDI theories

Lipsy (2004) describes the macroeconomic view as seeing FDI as a particular form of the flow of capital across national borders, from home countries to host countries, measured in balance-of-payments statistics. These flows give rise to a particular form of stocks of capital in host countries, namely the value of home-country investment in entities, typically corporations, controlled by a home-country owner, or in which a home-country owner holds a certain share of voting rights. Lipsy (2004) further explains that the variables of interest are the flow of financial capital, the value of the stock of capital that is accumulated by the investing firms, and the flows of income from the investments. Macro-level determinants that impact on a host country’s ability to attract FDI include market size, economic growth rate, GDP, infrastructure, natural resources, institutional factors such as the political stability of the country, amongst others. The various theories are discussed below.

5.1 Capital Market Theory

This theory, also sometimes referred to as the “currency area theory”, is considered one of the earliest theories which explained FDI. Based on the work of Aliber (1970; 1971), it postulated that foreign investment in general arose as a result of capital market imperfections. FDI specifically was the result of differences between source and host country currencies (Nayak & Choudhury, 2014). According to Aliber (1970; 1971), weaker currencies have a higher FDI-attraction ability and are better able to take advantage of differences in the market capitalisation rate, compared to stronger country currencies. Aliber (1970; 1971) further adds that source country MNCs based in hard currency areas can borrow at a lower interest rate than host country firms because portfolio investors overlook the foreign aspect of source country MNCs. This gives source country firms the borrowing advantage because they can access cheaper sources of capital for their overseas affiliates and subsidiaries than what local firms would access the same funds for.

While this capital market theory holds true in the case of developed countries such as the United States, United Kingdom and Canada, it was challenged by
later scholars on the basis of ignoring basic currency risk management fundamentals. A major criticism of Aliber’s theory was made by Lall (1979) when he highlighted that the theory does not apply in the case of less developed countries with highly imperfect or non-existent capital markets, and those with heavily regulated foreign exchange rates. Also, Nayak and Choudhury (2014) allude to the fact that Aliber’s theory does not explain investment between two developed countries with similar strength currencies, nor how developing country MNCs with weaker currencies are able to invest in developed countries with much stronger currencies. This they exemplified using the case of Chinese firms with sizeable investments in USA and the UK.

5.2 Location-based approach to FDI theories

Although FDI location is influenced by firm behaviour (a macroeconomic element) insofar as the motives of its location, that is whether it is resource-seeking, market-seeking, efficiency-seeking or strategic asset seeking; the overarching decision is in fact taken on the basis of economic geography, which is a macroeconomic decision as it takes cognisance of country-level characteristics (Popovici & Calin, 2014). According to them, the theory explained the success of FDI among countries based on the national wealth of a country, such as its natural resources endowment, availability of labour, local market size, infrastructure and Government policy regarding these national resources. An off-shoot of this location-based theory is the gravity approach to FDI wherein it was assumed that FDI flows between two countries is highest, if those two countries are similar geographically, economically and culturally. Gravity variables such as size, level of development, distance, common language and additional institutional aspects such as shareholder protection and trade openness were regarded as important determinants of FDI flows (Popovici & Calin, 2014). This is however a very basic approach to the economics of FDI, because FDI flows are more complicated than just being about commonalities between nations. Being close together geographically may reduce transportation costs, but not necessarily the cost of labour, for example. Also, sharing the same culture may not necessarily result in increased profitability or trade between the two countries.

5.3 Institutional FDI Fitness theory

Developed by Wilhelms and Witter (1998), the term FDI fitness focuses on a country’s ability to attract, absorb and retain FDI. It is this country ability to adapt, or to fit to the internal and external expectations of its investors, which gives countries the upper-hand in harnessing FDI inflows. The theory itself attempts to explain the uneven distribution of FDI flows between countries. Wilhem’s institutional FDI fitness theory rests on four fundamental pillars – Government, market, educational and socio-cultural fitness. At the base of the pyramid are socio-cultural factors which according to Wilhelms and Witter (1998), are the oldest and most complex of all institutions. Above that is education, which the authors affirm to being necessary in ensuring an attractive environment for FDI as educated human capital enhances R&D creativity and information processing ability. The actual level of education does not seem to matter much for FDI as the requirements are dependent on the various skills needs of projects to be undertaken. However what is certain is that basic education may impact on the productivity and efficiency of FDI operations, making formative education such as the ability to speak, hear, understand, interpret and implement instructions key for attracting FDI.

The third pillar, that of markets, accounts for the economic and financial aspects of institutional FDI fitness, in the form of machinery (physical capital) and credit (financial capital). Developed and well-functioning financial markets are hence a prominent feature in the MNC’s investment decision-making process. The fourth and final pillar as put forth by Wilhelms is the Government. The role of a country’s political strength plays the biggest role in the FDI game. Government fitness requires the adoption of protective regulation to manage market fitness. Popovici and Calin (2014) add that Government fitness is considered to include economic openness, a low degree of trade and exchange rate intervention, low corruption and greater transparency. If policies are hostile and unfavourable towards investors, MNCs will shy away from such countries as the political instability increases the risk burden on their investments. (Wilhelms & Witter, 1998). The authors concluded that although the pyramid is represented in a specific order, the four institutional pillars in fact are inter-related and interact in unison in different forms. For example, Government policies shape markets, education and sociocultural activities; market forces impact on the Government, education and socio-culture; education affects human capital and hence Government, markets and sociocultural norms and practices; and finally, sociocultural systems are the origin of Government, markets and education, respectively (Wilhelms & Witter, 1998).

Interestingly, the theory of institutional FDI fitness has been empirically tested mainly in the African context. Muthoga (2003) (as cited in Popovici & Calin, 2014), investigated FDI determinants in Kenya for the period 1967-1999, in their PhD thesis. The author found that economic openness, GDP growth rate, level of domestic investment, internal rate of return and availability of credit – all proponents of Government economic policies – enhance a country’s attractiveness to foreign investors. Along the same ideologies, Musonera,
Nyamulinda and Karuranga (2010) evaluated the institutional FDI fitness model in the East African Community bloc, using Kenya, Tanzania and Uganda as their sample, and data drawn from 1995 to 2007. They found that for Tanzania and Uganda, FDI inflows were predetermined by more than a single country risk factor, such as population size, size of economy, financial market development, trade openness, infrastructure and other economic, financial and political risks. Their research also further refuted the perception that FDI inflows to Africa are attracted by natural resources. This was evidenced by that Tanzania and Uganda, both resource-poor countries, were also able to attract FDI on condition that their Governments fulfill two conditions: establish macroeconomic and political stability, and introduce an efficient regulatory framework, as well as eliminate corruption.

6 Microeconomic FDI theories

Lipsey (2004) also states that the microeconomic view examines FDI motivations from the investor’s perspective, which would be similar to take a firm-level or industry-level perspective in making a decision. This micro-view thus examines the consequences to the investor, and to home and host countries, of the operations of the multinationals or of the affiliates created by these investments, rather than the size of the flows or the value of the investment stocks or investment position. These consequences arise from their trade, employment, production, and their flows and stocks of intellectual capital, measured by the capital flows and stocks in the balance of payments, although some proxies for the flow of intellectual capital are part of the current account (Lipsey, 2004). According to Das (n.d.), microeconomic FDI theories attempt to shed light on why MNCs choose to locate their subsidiaries where they do, and why they specifically seek to penetrate those locations. Many of these microeconomic FDI theories are all based on the existence of imperfect markets.

According to the firm-specific advantage theory, developed by Hymer (1976), the decision of an MNC to invest abroad rests on certain advantages at its disposal, such as access to raw material, economies of scale, access to labour, low transaction costs, intangible assets in the form of brands and patents, amongst others. It is in fact a firm-level (firm-specific) decision, rather than a capital market one (Das, n.d.). Hymer’s theory which laid the foundation in explaining international production was also supported by scholars such as Kindleberger (1969) in his imperfect markets model; Knickerbocker’s (1973) oligopolistic reaction theory of following the market leader; the internalisation theory of Buckley and Casson (1976) in an international context, as well as Dunning’s (1974) eclectic paradigm. These theories are based on the same fundamental principle – the existence of imperfect markets, which then has a bearing on firm behaviour. As a result, other than Dunning’s eclectic theory, no further attention will be given to them, as they are accounted for in Dunning’s OLI paradigm.

6.1 The Eclectic Paradigm

This is probably the most well-known theory of FDI. On his way to winning the world acclaimed Nobel Prize, Dunning (1980) integrated various theories discussed above – being the international trade, imperfect markets (monopoly) and internalisation theories, and complemented these with the location theory, also briefly discussed earlier. According to Dunning (2001), in order for a firm to engage in foreign direct investment, it must simultaneously fulfill three conditions.

The firm should possess net ownership advantages over other firms serving particular markets. These ownership advantages are firm-specific and exclusive to that firm, in the form of both tangible and intangible assets such as trademarks, patents, information and technology, which would result in production cost reductions for the firm, enabling it to therefore compete with firms in a foreign country. These advantages were also emphasised by Hymer (1976) and Kindleberger (1969) in their market imperfections’ theories on firm-specific and monopolistic advantages, respectively.

Secondly, it must be more profitable for the firm possessing these ownership advantages to use them for itself (internalisation), rather than to sell or lease them to foreign firms through licensing or management contracts (externalisation). Boddevyn (1985) refers to this as the internalisation condition. Finally, assuming that the preceding conditions are both met, it must be profitable for the firm to exploit these advantages through production, in collaboration with additional input factors such as natural resources and human capital, outside its home country; failing which, the foreign markets would then be served through exports, and local markets by domestic production. Location-specific factors have to be taken into consideration by the investing firms, as per the economic geography and institutional FDI fitness theories discussed under the macroeconomic FDI theories.

Boddevyn (1985) emphasises that the more a country’s firms enjoy ownership advantages, the greater the incentive they have to internalise them, and the more profitable to exploit them outside their home country, then the higher the probability of engaging in FDI and international production. Because of the interrelatedness of the three conditions, it is important that they occur simultaneously, otherwise FDI cannot occur. The context and application of the Ownership, Location and Internalisation (OLI) paradigm differs from firm to firm, and hence the theory cannot be considered in isolation of theories which affirm the importance of the host country characteristics.
Although the Eclectic Theory was empirically tested by Dunning himself, it still has some limitations which critics have highlighted over the years. Boddewyn (1985) praised Dunning’s theory for explaining the initial FDI decision by MNCs, but however laments the lack of explanation with regard to subsequent FDI increases, which may only require changes only in some but not necessarily all the OLI factors. In addition to this, Shin (1998) questions the applicability of the theory to LDCs which generally do not monopolistic firm-specific advantages such as high knowledge content. Another criticism of the eclectic theory is that it incorporates so many variables that it ceases to be operationally practical as it does not explain FDI at the firm, industry and country levels. This is on the basis that Dunning attempted to combine several complementary theories of market imperfection, which even on their own are already fairly complex (Nayak & Choudhury, 2014).

To address these shortcomings, Dunning (1981) then came up with the Investment Development Cycle or Path (IDP) theory, in which he proposed a link between a country’s level of economic development and its investment positions. The IDP had four stages which followed a pattern similar to the product life cycle theory (introduction, growth, maturity and decline): no FDI; location-specific advantages arise due to Government intervention, hence attracting FDI inflows; domestic firms enjoy ownership advantages as wages rise, resulting in FDI outflows; countries finally become net outward investors in the fourth stage. The underlying hypothesis here is that due to the dynamic interaction between a country’s GDP and its economic policies, these have the potential to affect both domestic and foreign firms’ ownership advantages (Nayak & Choudhury, 2014). Despite these challenges, Dunning’s eclectic theory however still remains the most recognised FDI theory.

Another criticism of Dunning’s OLI paradigm was raised by Forssbaeck and Oxelheim (2008) when they questioned the monopolial role assigned to financial aspects in the FDI decision. In his defence, Dunning (1993) acknowledged the existence of a “financial asset advantage” which is a firm’s knowledge of and access to foreign sources of capital, but points out that this merely a by-product of the size, efficiency and knowledge of MNCs, and not necessarily a standalone advantage. Forssbaeck and Oxelheim (2008) argue that a strong financial strategy enables a firm to minimise its cost and maximise availability of capital; thus by lowering the discount factor of any investment, that firm’s likelihood to engage in FDI increases as a result of the financial advantage. To this end, they hypothesized that a firm will engage in FDI when, amongst other things, it has access to competitively priced equity, when it cross-lists its shares on a larger, more liquid stock market, when it enjoys strong investment credit ratings, and when it is able to negotiate reduced taxation and/ or attract subsidies. Forssbaeck and Oxelheim (2008) empirically tested their hypotheses using a sample of 1379 European non-financial firms’ international acquisitions. In their series of tests, they evaluated what effect including finance-specific variables has on Dunning’s OLI model, and found that there was a strong explanatory power of the financial variables, thereby concluding that financial factors are equally important in explaining FDI using the OLI model.

7 Conclusion

Having examined the available major FDI theories, it is clear that there is no single superior theory which comprehensively explains FDI. However, as it is necessary to conduct research from a specific theoretical background, it is hoped that the above classification and analysis of FDI theories provides an adequate grounding towards selecting the most appropriate theoretical framework for future scholarly work.

References

43. Smith, A. 1937, The wealth of nations [1776], na.