International Trade & Domestic Growth: Determinants, Linkages and Challenges

September 27, 2007
Firm Heterogeneity, Exporting and Foreign Direct Investment

David Greenaway
Richard Kneller
University of Nottingham

A rapidly expanding literature on firm heterogeneity and firm level globalisation strategies has developed over the last decade. There are new insights on why some firms export and others do not, why some firms fail to survive in export markets and some choose to produce overseas rather than export. This article provides a synthesis and evaluation of this literature. It reviews both new theories of firms in an open economy context and the extensive microeconometric evidence base, which has now developed. It highlights the implications of this evidence base for policy and includes an assessment of how the research agenda may evolve.

Interest in a range of aspects of firm and plant level adjustment to trade liberalisation and falling trade costs has exploded in recent years, and a new literature is leading to significant re-thinking of key drivers of the globalisation process: cross-border trade and cross-border investment. Like the last revolution in thinking in international trade (sometimes called new trade theory) which incorporated imperfect competition as a response to empirical observation of intra-industry trade, this new literature was also triggered by empirical observation, particularly the work of Bernard and Jensen (1995). That paper drew attention to the fact that exporting and non-exporting firms co-existed in the same industry but were marked by clear defining characteristics. The development of the literature since then into a progressive research programme has been fuelled by two

---

1 The authors acknowledge helpful comments on an earlier draft from three anonymous referees, Roberto Alvarez, Daniel Bernhofen, Ricardo López, Jim Markusen, Horst Raff, participants at the Singapore Economic Review Annual Conference 2005, the Otago Trade Workshop 2006 and at a SUFE-Orebro Conference in Shanghai in 2005. Financial support for The Leverhulme Trust under Programme Grant F114/BF is also gratefully acknowledged.

2 In so doing this paper fits into a broader literature on the within-industry heterogeneity of firms such as Olley and Pakes (1996), Roberts and Tybout (1996) and Aw et al. (1997).
complementary developments. First, major theoretical break-throughs associated with Melitz (2003), Helpman et al. (2004) and Bernard, Eaton et al. (2003) among others have resulted in new ways of thinking about firm heterogeneity and participation in international markets. Second, the growing availability of micro level datasets has facilitated detailed analysis of firm level adjustment in a large number of countries.

One dimension which has received particularly close attention is the relationship between firm level productivity, entry to and survival in export markets. Following Bernard and Jensen (1995) there is now an extensive body of empirical analyses on a large number of industrialized, transitional and developing countries. This addresses not only the characteristics of firms which enter export markets, but also those markers likely to be associated with survival. In addition, recent analysts have turned their attention to the issue of why firms choose to export rather than engage in direct production overseas. For both, the interaction of sunk costs and productivity heterogeneity is key.

At the most basic level what this literature adds to our understanding of export behaviour is clear: a combination of sunk costs and heterogeneity in the underlying characteristics of firms explains why not all firms export.\(^3\) We have moved from the new trade theory world of representative firms, where all firms export, to one in which firms are heterogeneous and some export, some do not. But the literature goes beyond this, for example to the recognition of potential complementarity between exporting and foreign direct investment (FDI), which challenges the traditional view of multinationals as different from other firms, with exporting and FDI being substitute strategies. Helpman et al. (2004) and others build on the Brainard (1987, 1993) model, which stresses trade-offs between proximity and concentration, but differ in that the export or FDI choice is predetermined by firm productivity. This provides a basis for understanding globalisation in a broader context and therefore in understanding how changes to the costs of exporting or foreign direct investment change production patterns within industries and across countries.

Within this literature, the direction of causation between productivity and internationalisation has been controversial. It has become something of a stylized fact that ex-ante productivity determines the choice of whether or not to export. In other words, firms have to become more productive before they export and causality runs from productivity to exports. Causality in the opposite direction is less clear. One can think of plausible reasons why a presence in export markets might raise productivity after entry, for instance exposure to best practice technology and learning,

---

\(^3\) Earlier and related insights into the role of sunk costs in sluggish adjustment of trade responses to exchange rate fluctuations are attributable to Baldwin (1988) and Baldwin and Krugman (1989).
but the empirical evidence is mixed. More generally, when studying the determinants of entry and exit from markets, most researchers include measures of international trade in the industry and at the firm level, with the notion that firm death is less likely when the firm is an exporter or in an industry in which exposure to imports is low. Entry and exit then lead to aggregate productivity changes as market shares change.

These are important issues from a policy perspective. Export promotion policies of one form or another are pervasive the world over, as a glance at a random sample of World Trade Organisation (WTO) Trade Policy Reviews would confirm. These can take many (transparent and opaque) forms and are often general rather than targeted. The point to note at this stage however is that if not all firms have the appropriate attributes to export, some may simply self select into export subsidies. So the literature is sharpening this policy debate.

In this article we provide a critical review of this new literature. Because it is growing so fast, we limit ourselves to firm heterogeneity, exporting and FDI. We begin our appraisal with a review of new theories of the firm and international trade. In section 2 we then focus on productivity, entry and survival, taking in evidence on exchange rates, agglomeration and changes in the policy environment. Section 3 moves on to exporting and FDI. In addition to evaluating these as alternative strategies we also examine links between the decision to establish production facilities overseas and exporting. In section 4 we discuss the emerging research agenda including for example new thinking on the boundaries of the firm, outsourcing and offshoring, associated with Antras (2003) and Antras and Helpman (2004). We also look more closely at the policy context in this section. Section 5 concludes.

1. New Theories of the Firm and International Trade

Although the standard workhorse Heckscher-Ohlin model of international trade has profit maximising firms in the background, operating under constant returns to scale, their boundaries are not well defined and they have no deterministic role in determining the pattern or commodity composition of trade. Economic activity takes place in sectors and international competitiveness is fashioned by relative factor endowments between potential trading partners. New trade theory associated with Krugman (1979) and others builds on Dixit-Stiglitz monopolistic competition and explicitly has firms. However in that framework all firms export, because each produces a unique variety that consumers, who have love of variety preference functions, want. In this setting any trade costs just absorb a proportion of a firm’s foreign revenue but do not stop it from exporting. Although new trade theory gave us new insights into the determinants of trade, a world where all firms export is manifestly at odds with what we observe in the real world, where some export and
others in the same industry do not. The reason why this happens in the models of Krugman (1979) and others is that firms do not face fixed costs of exporting.

The business community would take it as axiomatic that entering export markets incurs sunk costs: market research has to be done; option appraisals completed; existing products have to be modified; new distribution networks set up and so on. Clerides et al. (1998) were one of the first to model this explicitly in a discrete choice framework. In their model, more productive firms with lower marginal costs earn higher gross profits from producing, but not all firms export. Only those with sufficiently high profits to cover the sunk costs do so. This intuitively appealing result leads to the conclusion that self-selection is fundamental – sunk costs and firm heterogeneity interact and the most productive firms self-select into export markets. Its corollary is that firms have to raise productivity before they enter. So it follows that there is a direct connection between productivity and exporting (but if policymakers want to exploit that, they should target support at potential rather than actual exporters).

But this may not be the end of the story. Clerides et al. (1998) also raise the possibility of learning by exporting. In other words, once a firm has entered export markets, productivity growth may receive a further boost. They model this as an upward shift in the (stochastic) process that determines firms productivity and it can be rationalized in various ways. For example, actual involvement in export markets could sharpen incentives to innovate by raising returns to innovation, a possibility modelled by Holmes and Schmitz (2001). A second possibility is that export markets are more competitive than domestic markets, forcing firms to reduce X-inefficiency. Here, learning results in business process re-engineering for example. The point is that if learning by exporting occurs, firm productivity may grow after entry as well as before. If this were the case, it provides a plausible mechanism underpinning export-led growth, though it also complicates the calculation that faces policy makers. Ultimately it is an empirical issue to which we turn in section 2.

4 In a multi-country setting, between firm productivity differences can generate intra-industry trade in these models.
Everything we have said so far refers to intra-firm productivity. At the macro-level we often associate productivity growth with inter-sectoral reallocation, classically the shift of resources from agriculture to manufacturing. Can we say anything in the current context about inter-firm reallocation and industry productivity growth? The pioneering paper here is Melitz (2003), which is set out schematically in chart 1 from Falvey et al. (2005). He builds a dynamic industry model with heterogeneous firms operating in (Dixit-Stiglitz) monopolistically competitive industries. Firms incur a fixed cost to export. However, each has to make a productivity draw from an exogenous distribution which determines whether they produce and export, and an endogenously determined productivity threshold determines who does and does not export.\(^5\) The interaction of these raises industry productivity. First, there is a rationalisation effect. Exporting increases expected profit, which induces entry, pushes up the productivity threshold for survival and drives out the least efficient firms in a Schumpeterian wave of creative destruction. Clearly this raises average industry productivity. Second, exporting allows the most productive firms to expand and causes less productive firms to contract. The productivity distribution that results is set out in chart 2. This reallocation effect again acts to raise average industry productivity. This model, despite its microeconomic structure, helps us understand the correlation between exports and growth widely observed at the macro level.

---

5 Ederington and McCalman (2004) develop a model of firm heterogeneity with the opposite outcome. Heterogeneity is a consequence of the decision of some firms to start to export.
Melitz (2003) is an important model linking heterogeneous firms and industry productivity, with exporting being a key factor. It is not the only model to point to causal links between exporting and industry productivity. This is also a key output of Bernard, Eaton, et al. (2003). Their industrial organisation structure is different but they still derive rationalisation and reallocation effects, however, the former is driven by import competition and the latter from exporters penetrating more markets. Jean (2002) also identifies import driven and export driven contributors to industry productivity growth, in a two-country setting with differences in relative efficiencies across countries.

The core Melitz (2003) model is now being developed in various ways. Helpman et al. (2004) extend it to consider the decision to set up an overseas affiliate. As in Melitz (2003) increased globalisation is likely to lead to firm exit, where the probability is decreasing in whether the firm is an exporter or multinational firm. We return to this in section 3.

A number of recent papers extend Melitz to consider asymmetries between countries. Melitz and Ottaviano (2003) examine differences in the extent of competition between countries (proxied by differences in size) on equilibrium outcomes following trade liberalisation. They find that because competition is tougher in the large country, product choice is greater, average productivity higher, but firm survival lower, because new entrants have a higher probability of failure.
Trade liberalisation increases competition in both countries thereby raising aggregate productivity but these effects are felt disproportionately in the big country (because it attracts a disproportionate number of firms).

In Falvey et al. (2004) countries differ in the efficiency with which they use frontier technology. One interesting finding is that self-selection is stronger for industries in which the degree of substitution across products is higher. Therefore the probability of firm closure may be negatively correlated with the level of intra-industry trade. They also find the higher the average efficiency of the country the more likely firms are to survive in the export market, but the less likely they are to survive in the more efficient country, which leads us to expect that trade structure is important. The pattern of trade is determined by the physical size of countries and size of the efficiency gap. For a given efficiency difference, as the size falls, domestic production of the differentiated product falls. By contrast, for a given size difference, as the efficiency gap rises, domestic production of the differentiated product rises. The effect of falling trade costs is to raise the minimum productivity needed to survive—it raises the self-selection cut-off point. This effect is strongest in the more efficient country.

The approach of Bernard et al. (2007) is to combine heterogeneous firms with Helpman and Krugman (1985) assumptions of imperfect competition and scale economies, and Heckscher-Ohlin differences in factor endowments. The model generates predictions about reallocations of resources across industries by firms. Finally, Bernard, Redding and Schott (2003) develop a model to explain an alternative form of exit to death-industry switching. Productivity levels are again shown to be important, albeit in the context of a closed economy. Here product switching depends on the fixed costs associated with production of different products and heterogeneity in productivity. More productive firms endogenously choose to produce products with higher sunk costs. Although that paper does not identify a role for international competition in firm choices, an effect from increased openness to trade is possible to envisage. Firms alter their output mix towards industries in which they have a comparative advantage and therefore avoid competition from countries in industries where they do not. For OECD countries this is more likely towards the use of technologies with higher costs, where this decision is dependent on firm productivity.

As we can see from this brief review of this theoretical literature, modelling exporting activity at the firm level throws up a range of possible channels through which exporting might be causally linked to firm and industry productivity. We now turn to the econometric analysis of these issues.

---

6 A more comprehensive review of the theoretical literature can be found in Helpman (2005).
2. Evidence on Productivity, Export Market Entry and Survival

As we have seen, theory points to differing performance characteristics of exporters and non-exporters. But do these differences result from the decision to export or do only good firms become exporters? This question of causality between exports and productivity, sparked in part by the ongoing debate over the relationship between openness and growth at the aggregate level\(^7\) has, by some margin, received most attention within the micro literature on exports. Thus, we first consider determinants of export market entry and exit as well as evidence on potential feedback from export market participation into firm performance. To provide some structure we begin with evidence relating to participation in export markets more generally.

According to Melitz (2003) and others, participation decisions are determined completely by a combination of sunk-costs and firm productivity. Although in empirical counterparts to this, the set of firm characteristics has been extended to include factors such as size, age, human capital, capital-intensity, ownership and so on, these predictions are supported by the evidence. While there are differences in the exact methodology employed (the choice over logit or probit models and attempts to correct for bias from inclusion of lagged export status of the firm) results are for the most part robust, a point made forcefully in Wagner (2007). Some if not all firm level variables are strongly correlated with export market entry. It follows that episodes of entry and exit should be predicted by periods of change in these characteristics (which we discuss below).

Of the explanatory variables, that relating to persistence (proxied by lagged export status) almost always explains most of the variation in the data. Exporting next period is strongly correlated with exporting this period, even when other determinants of persistence have been controlled for. Its coefficient is usually interpreted as evidence of sunk-costs. While the exact magnitude varies across studies, past participation increases the probability that a firm will continue to export by between 36% in the US (Bernard and Jensen, 2004a) and 90% in Italy (Bugamelli and Infante, 2002). Entry is therefore likely to be determined by changes in sunk-costs. As Das et al. (2001) show these are most relevant for those firms who export little, the fringe players in export markets (Tybout, 2003). But what are these changes that produce waves of entry and exit? The three contributors most often discussed are exchange rates, policy innovation and agglomeration effects.

---

\(^7\) See for example Rodriguez and Rodrik (2000) and Greenaway et al. (2002) and see López (2005) for an evaluation of micro and macro evidence.
2.1 Exchange Rates

Macroeconomic evidence on the effect on trade of exchange rate levels and volatility suggests effects that are either significant but small in magnitude, or insignificant (Pozo, 1992; Chowdhury, 1993; Parley and Wei, 1993). This implies that exchange rate movements play little or no role as a sunk cost. The micro evidence suggests however that these results are a product of aggregation and exchange rates are important. In the presence of sunk-costs the export responsiveness of exchange rate changes is likely to be higher amongst current exporters compared to non-exporters. That is, changes in exchange rates are more likely to lead to changes in the intensive rather than extensive margin. Bernard and Jensen (2004b) for example, study the export response of US manufacturing plants to dollar depreciation in the 1980s, and report that 87% of the expansion was from increased export intensity and 13% from entry of new firms. A similarly strong correlation is reported by Bugamelli and Infante (2002) and Bernard and Jensen (2004a).

Whilst useful for future comparative work, this approach does not provide a complete explanation of micro responses for three reasons. First, Das et al. (2004) find significant cross-industry variation in the effects of exchange rate movements. Simulating a 20% devaluation for three Colombian industries they report that the magnitude of industry response depends on previous export exposure, homogeneity of expected profit flows between firms and their proximity to the export market entry threshold. Ten years after devaluation the industry level effect varies between 14 and 107% (although unfortunately they do not break this into that generated by new entrants and that from existing exporters).

Second, devaluation can also lead to substantial exit. According to Blalock and Roy (2007) the 2 to 1 devaluation of the Indonesian rupiah against the US dollar between 1996 and 1998 did not lead to an aggregate export boom. Deeper analysis showed that although there was an expansion of export activity by established exporters and new entry by non-exporters, new activity was offset by cessation of exporting by previous exporters. Bernard and Jensen (2004b) also find evidence of exit for the US. Blalock and Roy (2007) offer an explanation: firms that ceased exporting were no more likely to report liquidity constraints, or infrastructure problems, compared to firms that continued to export and were no less productive; they were however less likely to be foreign and less likely to have made R&D or training investments. These same variables predicted which firms would become new exporters.

An alternative explanation can be found in Maloney and Azevado (1995), where in a model in which firms export to diversify revenue streams fitted to Mexican data, exchange rate volatility and the co-movement of domestic and

---

8 This contrasts with the large estimated currency union effects of Rose and Stanley (2005).
foreign demand shocks can lead to counter-intuitive movements in export volumes following changes in exchange rates. Finally, as we also note below, all of the detailed micro level analysis of exchange rate movements has been of episodes during which the domestic currency depreciated. It is not known whether the effect of appreciation is symmetric.

2.2 Policy Innovation

Export decisions are likely to be influenced by the environment in which the firm operates, where policy changes may impact on both intensive and extensive margins. For example, were policy to lead to within firm improvement in productivity perhaps because of increased competition or reduced costs of intermediate imports, it may be more likely that non-exporters enter export markets, but also easier for current exporters to increase export sales to existing or new markets. Unfortunately however we have little evidence on what aspects of policy are important for export volumes. In fact the evidence is concentrated in just five studies across two types of policy, trade liberalisation and export promotion, the results for which are summarised in table 1.9

Evidence on trade liberalisation suggests an effect on both intensive and extensive margins.10 Blalock and Gertler (2004) find that liberalisation in Indonesia between 1990 to 1996 doubled the number of exporters, while in their study of the effects of NAFTA on Canadian firms, Baldwin and Gu (2003) report increases in both the number of exporters (the share of plants that export increased from 37 to 53% between 1984 and 1990) and export intensity (in 48% of exporters). Using more sophisticated econometric techniques, they find the effect of policy on the export entry decision to be substantial. The 4.5% reduction in Canadian-US tariffs that occurred increased the probability of exporting by 63%.

---

9 We concentrate on evidence of trade liberalisation on export volumes at the firm level. There is a larger literature, see for example Pavcnik (2002), Roberts and Tybout (1996) or Tybout (2003) for references, that discusses the productivity impacts of such changes and Head and Ries (1999) and Roberts and Tybout (1991) for the effect on firm size. Given the link between exports, firm size and productivity these might be seen as indirect evidence of the export effect of policy changes.

10 The table does not include the results from Blalock and Gertler (2004) because of a lack of formal econometric evidence in the paper.
Table 1: Evidence on Policy Intervention and Firm Export Responses

<table>
<thead>
<tr>
<th>Authors</th>
<th>Sample</th>
<th>Policy Intervention</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bernard and Jensen (2004a)</td>
<td>US, 1984-92</td>
<td>State expenditures on export promotion</td>
<td>Insignificant effect on export market participation</td>
</tr>
<tr>
<td>Görg et al. (2007)</td>
<td>Ireland, 1983-98</td>
<td>Capital grants, training grants, rent subsidies, employment grants, feasibility study grants, technology acquisition grants, loan guarantees, research and development grants</td>
<td>In a matched sample large grants lead to additional exports. No evidence of additional entry. Withdrawal of grants does not lead to exit.</td>
</tr>
</tbody>
</table>

Export promotion is pervasive, and most governments intervene in one way or another, ranging from providing infrastructure support to offering direct export subsidies. Empirical evidence is again mixed, although this may be a result of both the question asked and level of detail available. Both Bernard and Jensen (2004a) and Alvarez (2004) find an insignificant effect from export promotion schemes, the former for exporters versus non-exporters; the latter for permanent versus sporadic exporters. Alvarez (2004) does however find differences in detail. Trade missions and trade shows do not increase the probability that a firm will become a permanent exporter, whereas market studies and arranged meetings with clients, authorities and experts do, even when controlling for other firm and industry determinants. Finally, it is worth noting the evidence of self-selection when evaluating export promotion schemes, a problem thus far not dealt with. Alvarez (2004) finds that established exporters are much more likely to have used public instruments for export promotion than sporadic exporters.

More detailed information on the payment of grants to firms is available for Ireland, as discussed by Görg et al. (2007). Using matching to control for selection problems, the authors find only limited success from intervention; large grants can induce existing exporters to expand overseas sales further but fail to encourage additional entry from those that did not previously export.
2.3 Agglomeration

Compared to the scrutiny of productivity spillovers, where some 40 studies were evaluated in Görg and Greenaway (2004), the literature on export spillovers is limited. It also concentrates on spillovers from the presence of other multinational firms within the same industry or region. As can be seen from table 2 only Aitken et al. (1997), Clerides et al. (1998), Bernard and Jensen (2004a) and Greenaway and Kneller (2003) consider spillovers from other exporters and only Greenaway and Kneller (2003), Sjoholm (2003) and Kneller and Pisu (2007) allow for spillovers from outside the region or industry.

In line with evidence of spillovers more generally, results are somewhat mixed. Some studies identify strong positive spillover effects (Aitken et al., 1997; Kokko et al., 1997; Greenaway et al., 2004; Greenaway and Kneller, 2003) others have either found none and in some cases negative impacts (Bernard and Jensen, 2004a; Sjoholm, 2003; Barrios et al., 2003; Ruane and Sutherland, 2005). Kneller and Pisu (2007) and Swenson (2005) find mixed evidence, depending on the channel considered. Beyond country specific differences there is no obvious pattern to these inconsistencies. This is best seen from a comparison of Greenaway et al. (2004), Barrios et al. (2003) and Ruane and Sutherland (2005) which all focus on European countries, measure foreign presence in the same way, and use a similar methodology.

Greenaway et al. (2004) measure foreign presence in the UK as the sum of industry employment or output and, in an attempt to separate competition from information effects, add exports from foreign multinationals as a proportion of total exports in the industry. They find both the likelihood of exporting and export share are increasing in the industry-level foreign presence index, even controlling for firm and industry level characteristics. They report less clear results for the index measuring export activities of foreign firms, this being positive and weakly significant for the export decision and positive and insignificant in the decision of how much to export. By contrast, Barrios et al. (2003) for Spain find no evidence of an effect on the export decision from MNEs or the export share.

Ruane and Sutherland (2005) also use a Heckman selection model to account for interdependence between export participation and export share decisions, but with contrasting results. They find positive effects from foreign presence of multinationals and negative effects from their export share on both export and export share decisions, with a suggestion the latter is due to US multinationals. They attribute this to the use of Ireland as an export platform to the EU. They argue export spillovers are unlikely where the country is an export platform because competition with domestic firms in local markets is limited. The use of spillovers from other exporters does not appear to improve this. Aitken et al. (1997) and Bernard and Jensen (2004a) find no effect from such measures, whereas Greenaway and Kneller (2003) do.
### Table 2: Evidence on Agglomeration and Firm Export Responses

<table>
<thead>
<tr>
<th>Authors</th>
<th>Sample</th>
<th>Measure of agglomeration</th>
<th>Export Participation*</th>
<th>Export Share</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alden et al. (1997)</td>
<td>Mexico, 1986-89</td>
<td>Foreign MNE share of exports by state &amp; industry</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>State industry share of national exports</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Bartos et al. (2003)</td>
<td>Spain, 1990-98</td>
<td>Foreign MNE share of exports by industry</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Foreign MNE share of R&amp;D by industry</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>Bernard and Jenson (2004a)</td>
<td>US, 1984-92</td>
<td>No. of exporters in region</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No. of exporters in region &amp; industry</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Corderos et al. (1998)</td>
<td>Colombia, Mexico</td>
<td>Exporters per industry or region</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td>and Morocco</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Foreign MNE share of employment, by industry</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Kneller and Pisu (2005)</td>
<td>UK, 1988-98</td>
<td>Foreign MNE share of exports by industry</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Horizontal industry-region domestic sales</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>Kalleberg and Sutherland (2007)</td>
<td>Ireland, 1991-98</td>
<td>Foreign MNE share of employment, by industry</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Foreign MNE share of exports by industry</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Spohner (2003)</td>
<td>Indonesia, 1980-91</td>
<td>Foreign MNE share of output by region</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>No. of multinational firms in city and industry</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exports by multinationals in a city</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Exports by multinationals in a city and industry</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Relator transaction density in a city</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>

*Notes: + = the effect is positive and significant, − = the effect is negative and significant, 0 = the effect is insignificant and/or changes sign and/or significance through the paper.

†These regressions relate to the $24$-age Probit regressions reported in Table 3 and excluding natural resource intensive sectors.
While positive and insignificant effects are relatively easy to explain in this context, negative effects are more puzzling. Ruane and Sutherland (2005) explain theirs by Ireland being an export platform, thus multinationals have less contact with indigenous firms. It is not clear however why this makes Irish firms less likely to export. Perhaps more plausible is the congestion argument of Swenson (2005): competition with multinationals raises prices in product markets forcing domestic firms up their average cost curves for example; or, perhaps higher costs result from congestion of local infrastructure.

2.4 Consequences of Export Market Entry

Entry can have a number of different impacts on the firm and aggregate economy. Some have provoked less discussion than others. For example there is widespread evidence of an aggregate productivity effect through resource reallocation (Bernard and Jensen, 2004a; Hansson and Lundin, 2004; Falvey et al., 2004). The area given greatest attention however, is direction of causality between exporting and within-firm changes in productivity. We focus on that, although other important effects might relate to survival probability of exporters (Bernard and Wagner, 1997; Bernard and Jensen, 1999).

At the simplest level this literature can be seen as a test between self-selection and learning, and indeed this was explicit in the earliest studies. The umbrella label learning in fact contains three separate channels. First, interaction with foreign competitors and customers provides information about process and product reducing costs and raising quality, which can be interpreted as learning by exporting. Second exporting allows firms to increase scale. Finally increased competition in foreign markets forces firms to be more efficient and stimulates innovation. However this fails to recognize how the hypothesis under test has evolved, to one of a bi-causal relationship. Self-selection is important, but leads also to endogenous changes in productivity either as a result of learning by exporting or learning to export.

In the earliest literature the hypothesis under test was clearly one of self-selection versus learning. The arguments in favour of the former are most powerfully put by Bernard and Jensen (1999, 2004b). In their study of US plants they found productivity growth of exporters was not significantly different from non-exporters, independent of whether productivity was measured as labour productivity or TFP. This implies that the productivity distribution of firms in any given industry does not widen continuously over time, or put differently the growth effects from learning are not permanent. They also provided evidence that out of the pool of non-exporters, new exporters were already among the best and differed

11 Evidence from Tybout and Westbrook (1995) suggests that this may be an unimportant source of efficiency change.
significantly from the average non-exporter. Whilst there is some country specific sensitivity in the magnitude of any difference in performance, a reasonable summary would be that the results of Bernard and Jensen (1999) for the US are replicated for most other countries (see Table 3). Export market entry is associated with significant changes in performance around the point at which export sales begin.

This argument for self-selection is therefore based on a comparison between established exporters and non-exporters and a difference in the performance of new export firms around the point of entry which is not permanent. Future entrants have many of the right characteristics that make them likely to export and faster productivity growth than non-exporters when they do. But, after a short period they become indistinguishable from other exporters. The strong conclusions reached by Bernard and Jensen (1999) in favour of self-selection led quickly to an adaptation of the hypothesis being tested to one of self-selection versus a bi-causal relationship. Recognising that new exporters appeared to already have many of the right characteristics to become exporters one can test whether the surge in productivity associated with entry was explained by the decision to become an exporter, or whether the productivity surge led to the export decision. As a consequence of the change in focus, methodology also evolved, with attempts to control for self-selection using either instrumental variable or matching techniques (alone or in combination with difference in differences). As argued in Van Biesebroeck (2005) not controlling for self-selection will overstate evidence of learning for new exporters in the data.

Instrumental variable approaches have usually been estimated using GMM; see for example Van Biesebroeck (2005); Baldwin and Gu (2003). Whilst they have the advantage of being relatively easy to estimate one faces the perennial question of instrument validity. By contrast, matching attempts to reduce heterogeneity between new and non-exporters by using observable firm characteristics. It has the disadvantage of removing observations from the data set and requiring specific assumptions about non-observable factors such as managerial ability. Establishing causality is probably the most challenging issue facing researchers in this area. Our view is that matching offers the sounder foundation, but we leave arguments to which of these methodologies should be preferred to Blundell and Costa Dias (2000) and focus instead on results from each.

The impact of applying these alternative techniques has been largely to confirm self-selection is more important than learning. For example, comparisons of new exporters and non-exporters without controlling for selection in Germany (Bernard and Wagner, 1997) and the UK (Girma, Greenaway and Kneller, 2004) shows significant pre-entry differences in performance, whereas differences are not

---

12 The evidence for Sweden (Hansson and Lundin, 2004; Greenaway, Gullstrand and Kneller (2005) and Slovenia (Damijan et al., 2007) are exceptions.
Firm Heterogeneity, Exporting and Foreign Direct Investment

evident with methods controlling for selection. Yet whilst evidence of post-entry productivity changes are reported for the UK (Girma et al., 2005b) they are not for Germany (Wagner, 2002). Indeed whilst both GMM and matching advance on simply comparing new exporters with all non-export firms, they do not guarantee post-entry productivity changes will be observed. As table 3 shows, more studies report evidence for learning than fail to find such effects, although it is perhaps worth noting these tend to be studies that use matching.

So what explains this divergence? Two issues have been explored, heterogeneity and timing. Some have argued that learning is likely to be specific to some firms, such as those that are young (Delgado et al. 2002; Fernandes and Isgut, 2005), or highly exposed to export markets (Kraay, 1999; Castellani, 2002; Girma, Go¨rg and Strobl, 2004; Damijan et al., 2007). Others have found post-entry changes depend on existing industry characteristics, productivity changes are lower in industries in which current exposure to foreign firms (through arms length trade and FDI) is high (Greenaway and Kneller, 2003). While it is difficult to conclude against such effects, heterogeneity should not be allowed to become an easy excuse for inconsistencies across studies. To establish heterogeneity will require evidence that the same mechanisms (such as age or foreign market exposure) are important across countries.

The learning by exporting hypothesis attributes part of the change in productivity to the endogenous decision to start López (2004) and exporting. More recently Alvarez and López (2005) have questioned the timing issue, arguing that productivity changes occur after the decision to start exporting, that is they may pre-date the point at which export sales begin.13 Firms invest in new technologies leading to pre-entry changes in productivity: they learn to export rather than learn by exporting. This takes the view that learning effects are neither inevitable nor automatic but require investments in domestic technology (Keller, 2004). While this might be seen by some as an unfair shift of the goalposts, it is consistent with a test of exogenous versus endogenous changes in productivity associated with exporting. It has also existed as an idea within the case study literature for some time (see the review by Pack, 2000) and a number of studies report anecdotal evidence (López 2004; Alvarez and López, 2005; Van Biesebroeck, 2005; and Blalock and Gertler, 2004). Empirical testing of this using micro data sets becomes more difficult owing to the unobservable nature of the time at which the decision to start to export is made, and the likelihood that preparation time varies across firms.

13 Alvarez and López (2005) label pre-entry effects as learning to export compared to learning by exporting for post-entry effects. The common element between these is the effect of the decision to export on the firms productivity.
Table 3: Evidence on Export Market Entry Effects and Firms

<table>
<thead>
<tr>
<th>Authors</th>
<th>Sample</th>
<th>Methodology</th>
<th>Pre-entry difference</th>
<th>Post-entry difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self/Selection vs. Learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aze et al. (2003)</td>
<td>Korea, 1985-95 and Taiwan (China), 1981-91</td>
<td>New Exporters vs. non-exporters</td>
<td>5% TFP Taiwan</td>
<td>6% TFP Taiwan</td>
</tr>
<tr>
<td>Baldwin and Gu (2003)</td>
<td>Canada, 1974-99</td>
<td>New Exporters vs. non-exporters</td>
<td>3% ALP, 3% ATFP</td>
<td>3% ALP, 3% ATFP</td>
</tr>
<tr>
<td>Bernard and Jensen (1999)</td>
<td>US, 1984-92</td>
<td>New Exporters vs. non-exporters</td>
<td>0% TFP, 7-8% LP</td>
<td>1% ATFP, 1-2% ALP-long run</td>
</tr>
<tr>
<td>Bernard and Wagner (2007)</td>
<td>Germany, 1978-82</td>
<td>New Exporters vs. non-exporters</td>
<td>5% LP, 0% ALP</td>
<td>3% ATFP, 3% ALP</td>
</tr>
<tr>
<td>Castellani (2002)</td>
<td>Italy, 1998-94</td>
<td>Exporters vs. non-exporters</td>
<td>+ ATPF, 0 ATFP</td>
<td>0% TFP</td>
</tr>
<tr>
<td>Daniljan et al. (2009)</td>
<td>Slovenia, 1994-2002</td>
<td>Exporters vs. non-exporters</td>
<td>0% TFP</td>
<td>0% TFP when export to non-OECD countries, 11% TFP when export to OECD countries</td>
</tr>
<tr>
<td>Delgado et al. (2002)</td>
<td>Spain, 1991-96</td>
<td>New Exporters vs. non-exporters</td>
<td>+ TFP</td>
<td></td>
</tr>
<tr>
<td>Hahn (2004)</td>
<td>Korea, 1990-98</td>
<td>New Exporters vs. non-exporters</td>
<td>4% TFP</td>
<td>10% increase in exports = 1% TFP, 0% LP</td>
</tr>
<tr>
<td>Hanson and Lundin (2004)</td>
<td>Sweden, 1990-99</td>
<td>New Exporters vs. non-exporters</td>
<td>0% ATFP, 3% ALP</td>
<td>5% ATFP, 3% ALP</td>
</tr>
<tr>
<td>Inglot (2005)</td>
<td>Colombia, 1991-94</td>
<td>New Exporters vs. non-exporters</td>
<td>2% LP, 4% ATFP</td>
<td>1% increase in exports = 1% TFP, 2% LP, 5% ALP</td>
</tr>
<tr>
<td>Liu et al. (1999)</td>
<td>Taiwan, 1999-98</td>
<td>New Exporters vs. non-exporters</td>
<td>0% ALP</td>
<td>0% ATFP</td>
</tr>
<tr>
<td>Self/Selection with Endogenous Productivity Change</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arnold and Hauinger (2005)</td>
<td>Germany, 1992-99</td>
<td>Matched D4D</td>
<td>+ ATFP</td>
<td>0% ATFP matched sample</td>
</tr>
<tr>
<td>Baldwin and Gu (2003)</td>
<td>Canada, 1974-99</td>
<td>GMM</td>
<td>3% LP, 0% TFP non-matched sample</td>
<td>5% LP, 1.7% TFP non-matched sample</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3 Continued: Evidence on Export Market Entry Effects and Firms

<table>
<thead>
<tr>
<th>Authors</th>
<th>Sample</th>
<th>Methodology</th>
<th>Pre-entry difference</th>
<th>Post-entry difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balock and Gertler (2004)</td>
<td>Indonesian firms, 1990-96</td>
<td>1. Fixed effects 2. IV-OE &amp; LP 3. timing</td>
<td>1.5% TFP 2. 2-5% TFP</td>
<td></td>
</tr>
<tr>
<td>Gentiles et al. (1998)</td>
<td>Colombia 1981-91, Mexico, 1986-96 and Morocco 1984-91</td>
<td>GMM</td>
<td>3. 0% ΔTPP 4. 0% ΔALP</td>
<td>Colombia + LP 5. 4% ΔTPP 6. Mexico + LP Morocco + LP</td>
</tr>
<tr>
<td>De Loecker (2004)</td>
<td>Slovenia, 1994-2000</td>
<td>Matched D+D</td>
<td>22% TFP 4% ALP</td>
<td>22% TFP 4% ALP 2% 3. 4% ΔTPP 4% ΔALP</td>
</tr>
<tr>
<td>Girim et al. (2003)</td>
<td>UK, 1988-98</td>
<td>Matched D+D</td>
<td>0% ΔTPP, 0% ΔALP in matched sample</td>
<td>0% ΔTPP, 0% ΔALP 1% ΔTPP 0% ΔALP</td>
</tr>
<tr>
<td>Greenaway and Kneer (2005)</td>
<td>UK, 1989-2002</td>
<td>Matched D+D</td>
<td>1% ΔTPP 0% ΔALP in matched sample</td>
<td>1% ΔTPP 0% ΔALP 1% ΔTPP 0% ΔALP</td>
</tr>
<tr>
<td>Greenaway, Gelbran and Kneer (2005)</td>
<td>Sweden, 1996-97</td>
<td>Matched D+D</td>
<td>0% ALP</td>
<td>0% ALP 0% ΔTPP 0% ΔALP</td>
</tr>
<tr>
<td>Wagner (2006)</td>
<td>Germany, 1978-89</td>
<td>Matched D+D</td>
<td>0% ΔTPP</td>
<td>0% ΔTPP 0% ΔALP</td>
</tr>
<tr>
<td>Self-Selection with Endogenous Productivity Change</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alvarez and López (2005)</td>
<td>Chile, 1966-86</td>
<td>Matched D+D</td>
<td>0% ΔTPP, 0% ΔALP</td>
<td>0% ΔTPP, 0% ΔALP 0% ΔTPP</td>
</tr>
<tr>
<td>López (2004)</td>
<td>Chile, 1966-86</td>
<td>New Exporters vs. non-exporters</td>
<td>6% ΔTPP 0% ΔALP matched sample</td>
<td></td>
</tr>
</tbody>
</table>

Notes: Where possible the results refer to a comparison of new exporters versus non-exporters.

- TFP = total factor productivity, LP = labour productivity, Δ = growth
- + the difference relative to the control group is positive and significant, − the difference relative to the control group is negative and significant, 0 the difference relative to the control group changes sign and/or significance through the paper.

These results refer to firms that survive in export markets, as reported in Table 10 and for value added per worker.

As López (2004) notes however, without information on timing of the decision, the time path of an endogenous change in productivity is likely to look similar to that of an exogenous change and it becomes harder to conclude that observed productivity changes are orthogonal to the export entry decision.

Using an econometric approach Aw et al. (2006) study the evolution of productivity and R&D for exporters in Taiwanese electronics. They find that those that do not invest in R&D have lower productivity growth than those that just export, which in turn is lower than those firms that invest in both.\(^\text{14}\) They argue these findings are consistent with an interpretation that R&D investments are necessary for firms to benefit from their exposure to international markets. López (2004) develops the same idea for domestic sales and investment. He finds investment and productivity rises in the pre-entry period but domestic sales are flat and argues this is consistent with investment in technology for sales to foreign but not domestic markets.

Endogenous pre-entry changes in productivity offer an interesting possibility for future research, though current analysis raises questions. First, a simple growth accounting approach suggests that if investment rises and output remains flat, productivity should fall. Simultaneous increases in investment and productivity would therefore seem an unlikely combination, unless of course there are reductions in other inputs. Here more detailed data on equipment and R&D investment would help. Second, how are we to interpret evidence of post-entry changes in productivity? The most obvious explanation is overlap between the benefits to new technology with the point at which sales start, perhaps due to lags in their effects due to learning. An alternative might be a difference between firms that are passive and active in their export decision. Discussions with those involved in export promotion in the UK suggest both occur frequently. For those firms that are passive, no pre-entry investments are made and productivity changes are likely to occur with the start of export sales.

Ultimately perhaps issues surrounding timing of the decision and investment in new plant, equipment or personnel are difficult to answer with available data, which offers insufficient detail. While case studies offer one solution, perhaps a more interesting approach is that used by Baldwin and Gu (2004) who combine micro data with questionnaires about export behaviour. They find evidence consistent with changes in scale, increased efficiency through competition and learning. Canadian exporters used more foreign technologies, were more likely to have R&D collaboration with foreign firms and improved the flow of information

\(^{14}\) A number of papers have found that exporters have higher levels of R&D but do not establish the direction of causality, see for example Bleaney and Wakelin (2002) and Roper and Love (2002) for the UK, Bernard and Jensen (1995) for the US, Aw et al. (2006) for Taiwan and Baldwin and Gu (2004) for Canada.
about foreign technologies to Canadian firms. That also led to increased innovation and investments in absorptive capacity.

2.5 Determinants and Consequences of Exit

As with export market entry, the literature on exit splits into determinants and consequences. A reasonable expectation would be that exit should be symmetric to entry. To some extent this is so. Exit from export markets is correlated with similar firm level variables as entry: it is less likely the larger, more productive and more human capital intensive the firm, and the lower the ratio of exports to domestic sales; see for example Greenaway and Kneller (2003) and Blalock and Roy (2005). Industry determinants have been less well researched. For example, research that focuses on the effect of exchange rate changes considers periods of domestic currency depreciation, when exports are likely to expand (Bernard and Jensen, 2004b, Das et al., 2004; Blalock and Roy, 2005). Thus far no one has considered whether the effect of appreciation is symmetric, although evidence of substantial export market exit in the presence of a depreciation of the Indonesia rupiah by Blalock and Roy (2005) suggests it is not.

The set of industry variables is extended by Greenaway and Kneller (2003) to include import penetration and intra-industry trade, as well as industry sunk costs. Conditional on firm level variables they find exit is more likely in industries with low sunk-costs, (because re-entry is easier) and those with high levels of intra-industry trade. No role for import penetration was found which is consistent with Melitz (2003), where self-selection is driven not by an increase in imports but the pull of export markets.

The literature on consequences of exit is somewhat larger. As with entry, self-selection appears to be important. Export quitters tend to have lower productivity compared to firms that continue (Aw et al., 2000; Baldwin and Gu, 2003; Girma et al., 2003) and no significant difference from, or in some cases, lower productivity (growth) than non-exporters (Bernard and Jensen, 1999; Hansson and Lundin, 2004; Hahn, 2004). Firms seem to self-select out of export markets just as they do into them. One caveat might be made from an often overlooked feature of the data, the comparison of new exporters with entrants: evidence presented across studies comparing entrants and quitters suggests the latter have higher productivity.

As with entry the effect of exit on productivity produces mixed results. Of those not conditioning for self-selection Hansson and Lundin (2004) and Hahn, (2004) find no obvious post-exit productivity changes, whereas Girma et al. (2003) and Blalock and Gertler (2004) report similar results conditioning on self-selection. By contrast, for the US Bernard and Jensen (1999, 2004b) report post-exit changes, not controlling for self-selection. On balance, it would seem that self-selection is important, weaker firms are likely to exit, but unlike entry there is little impact on productivity of this choice.
3. Exporting and Foreign Direct Investment

3.1 Exports versus FDI

At the simplest level, exports and FDI are substitute channels for firms globalising. The conditions for foreign production become more favourable relative to exporting as the size of the foreign market increases and costs of exporting increase; and less favourable as costs of setting up foreign production grow. This is the proximity-concentration trade-off explained by Brainard (1993). The contribution of Helpman et al. (2004) to this is analogous to Melitz (2003) contribution to the basic model of trade with representative firms. Adding heterogeneity allows this choice to differ across firms within the same industry and thus determines which firms export and which become multinational. The interesting properties of the model in this regard are generated through the assumptions of different costs (largely fixed) associated with serving domestic and foreign markets (through FDI or exports), along with heterogeneity in productivity across firms.

As we have seen sunk-costs of exporting are typically thought to include fixed costs of research into product compliance, distribution networks, advertising and so on. Goods exported are also subject to transportation costs. The fixed costs of FDI are the duplication of costs in establishing domestic production facilities. They are assumed to be greater than those of exporting, FDI eliminates variable transport costs, but involves higher fixed costs. Heterogeneous productivity then ensures self-selection. Only the most productive firms become multinationals; firms whose productivity falls in an intermediate range export and the least productive only sell domestically.

Helpman et al. (2004) assume the decision to establish foreign production facilities is based purely on considerations of market access. All FDI is horizontally motivated. Head and Ries (2003) demonstrate that when there are factor price and market size differentials, firms invest abroad for vertical motives also: the ordering of the productivity distribution between multinationals and non-multinationals can even be reversed. If the foreign country is small and offers some cost advantage, for a certain range of the parameter of the model, the least productive firms locate abroad whereas more productive ones produce at home. In this case, low productivity enterprises have a greater incentive to pay the FDI sunk costs because they use more intensively the factor whose overseas price is low.

15 We concentrate here on the evidence at the level of the firm. The issue of complementarity and substitution between exports and FDI has been studied at many other levels of aggregation, a summary of the evidence for which can be found in the Head and Ries (2004).
Empirical tests of the heterogeneous firm model have generally followed one of two lines. First, testing within industries for substitution between exports and FDI related to productivity differences. Second, testing the cross-industry/country predictions – the volume of exports relative to FDI we might expect. Whilst there is a large literature comparing productivity levels of multinationals against non-multinationals and exporters against non-exporters, there are only a small number of studies that compare exporters and multinationals. In part this is because it is a relatively new question, in part because for many countries information on which domestic firms export and which are multinational is not available. As can be seen from table 4 two basic approaches to this question are evident. The first follows Head and Ries (2003) in comparing mean values (in some cases conditional on other firm and industry characteristics), see for example Castellani and Zanfei (2007) and Kimura and Kiyota (2004). The second follows Girma et al. (2005a) in using Kolmogrov-Smirnov tests of stochastic dominance, see Girma, Görg and Strobl (2004), Arnold and Hussinger (2005b) and Wagner (2005). This approach compares the cumulative distribution of productivity for different types of firms and not just the mean. Despite the difference in methodology, the prediction with regard to exports versus FDI would appear to have strong support, Head and Ries (2003) being the exception), while ironically that between exporters and non-exporters less so. Whilst explaining differences across a small number of studies is never easy, several report a bias towards large firms, and therefore a bias against finding significant productivity differences, and there is a suggestion that this is most severe in Head and Ries (2003), who use information on publicly listed firms.

Table 4: Evidence on Relative Productivity of Exporters and Multinationals

<table>
<thead>
<tr>
<th>Authors</th>
<th>Sample</th>
<th>Methodology</th>
<th>Exporters vs. non-exporters</th>
<th>MNEs vs. exporters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arnold and</td>
<td>German, 1950-55</td>
<td>K-S tests of stochastic dominance</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Hussinger (2006b)</td>
<td>Germany, 1990-92</td>
<td>OLS</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>Castellani and</td>
<td>Italy, 1990-95</td>
<td>OLS</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>Zanfei (2007)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girma, Görg and</td>
<td>Ireland, 2000</td>
<td>K-S tests of stochastic dominance</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>Strobl (2004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girma et al. (2005a)</td>
<td>UK, 1990-95</td>
<td>K-S tests of stochastic dominance</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Head and Ries (2003)</td>
<td>Japan, 1989</td>
<td>OLS</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Kimura and</td>
<td>Japan, 1994-2000</td>
<td>OLS</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Kiyota (2004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: + the effect is positive and significant, – the effect is negative and significant, 0 the effect is insignificant and/or changes sign and/or significance through the paper.
Head and Ries do find predictions in support of the model for size characteristics.
The second strand of the literature concerns itself with proximity-concentration predictions, the relative level of exports to FDI. Helpman et al. (2004) predict FDI will be more common relative to exports, the greater the dispersion of productivity levels within an industry. The data requirements of such a test are demanding however, particularly with regard to foreign sales by domestic multinationals and measures of dispersion within an industry. They use US data and regress the ratio of exports to FDI (measured by sales of overseas affiliates) on traditional proximity-concentration variables, unit costs of trade and plant fixed costs, as well as a new variable, within industry dispersion. They consistently find that dispersion has the expected effect on relative sales: industries in which firm size is highly dispersed are associated with relatively more FDI than exports.

3.2 Exports by MNEs

Whilst in a single product world exports and FDI are substitutes, even if this choice is determined exogenously by productivity levels, in practice multinationals also export. Indeed many report that foreign multinationals contribute disproportionately to exports compared to employment or output shares (Baldwin and Gu, 2003; Kneller and Pisu, 2004). To some extent this should be expected, a well-established result is the superior performance of foreign owned firms with respect to employment, wages and productivity, all of which are important determinants of exports. Should the export decision of multinational firms be modelled as identical to that of domestic firms however? What little evidence there is suggests not. Kneller and Pisu (2004) find that even controlling for characteristics, foreign firms are more likely to export than indigenous ones, and export more intensively.

So what explains export decisions of multinationals? Modelling has developed along two lines: export platform FDI and complementarity, broadly distinguished by the number of product lines the firm is assumed to produce. Export platform FDI is typically defined as the establishment of foreign production facilities and allocation of part or all of the output to serve a third country. It therefore refers to exports of a single product line, where these are not to the home country. Complementarity refers instead to multi-product firms, to multiple stages of production and to export and FDI flows from the home to foreign countries: exports and FDI become positively correlated if there are horizontal or vertical complementarities across product lines.

Theories of export platform FDI have developed by adding more countries and stages of production to traditional theories of FDI and in more recent developments in cross-firm heterogeneity, FDI becomes complex. Vertical FDI occurs when the

16 Helpman (2005) takes a somewhat broader view of this question adding a discussion of the role of incomplete contracts for firms internationalisation and offshoring decisions.
stages of production are located in more than one country; and horizontal when the same stage is located in more than one country. Vertical FDI is factor seeking; horizontal, market seeking. When there are more than two countries and more than two stages of production, multinationals are likely to undertake more complex FDI choices which involve intra-firm trade and export platform FDI. The effect of adding more countries is to allow for the possibility of a horizontal motive for export platform FDI, adding more stages allows for a vertical motive.

Motta and Norman (1996), motivated by the observation that much FDI is between countries in regional trading blocks, consider three identical countries and a single stage of production. Costs of production do not differ between countries but costs of trading do (because two either enter a free trade agreement or raise external barriers against the third). If we start from an equilibrium where each firm exports to the other two countries from its home base, raising external barriers or creating a free trade area encourages the outside firm to set up production facilities inside the free trade area and export to the other country in the bloc. Where the outside country chooses to locate production in and export from is left undetermined. Again, because of identical costs neither of the inside countries choose export platform FDI as a strategy.

The conditions under which export platform FDI is likely have been analysed by Ekholm et al. (2003) where there are two identical countries in the North (A and B) one in the South, and multiple stages of production. Each firm produces intermediates and a final good. Firms must provide headquarter services from their home northern country but can choose where to produce intermediates as well as assembling the final product. Two of the countries, one northern (A) and one southern are members of a free trade area. The drivers of the model include assumptions about the size of the (marginal) cost advantage of southern firms and trading costs between different sets of countries. The free trade area between A and the Southern country means it is always optimal for the northern country to locate production in the South and export home (owing to the cost advantage from doing so). Therefore, unlike Motta and Norman (1996), when there are no vertical motives for FDI, the country inside the free trade area always has a motive to undertake export platform FDI.

For the other northern country (B) the model predicts three outcomes. First, no FDI: firm B produces at home and exports to the free trade area; second, export-platform FDI: firm B produces the good to be sold at home domestically, whereas the final product sold in the other northern country is produced in the South and exported; third, vertical FDI (hybrid MNE): firm B locates all production in the South and exports to both markets in the North. The last is hybrid because toward the home country, the firm undertakes vertical FDI whereas, toward the other Northern country, it undertakes a pure form of export platform FDI. Which strategy is adopted depends on the size of the (marginal) cost advantage to Southern firms, and trade costs. As the cost advantage of Southern firms increases we move from
the first equilibrium to the second and when the cost advantage of locating in the South becomes large enough all production moves there. Similarly as trade costs between the Southern and two Northern countries fall, the Northern firm outside the FTA finds it competitive to move from exporting to the FTA, to export platform FDI, to locating all production in the Southern country. This has similarities to Motta and Norman (1996).

The predictions of these models are driven primarily on cross-country differences in costs. Grossman et al. (2003), developing the complex FDI model of Yeaple (2003), show that firm characteristics may also be important. If firms in the same industry are heterogeneous in productivity they may make different choices, even though costs of exporting and FDI are the same. They assume three countries (two North and one South); firms must provide headquarter services, produce intermediates and assemble the final product. Their analysis allows for the coexistence in the same sector of a rich array of profitable FDI strategies. In brief, the general lesson is that least productive firms will not undertake FDI. More productive firms choose complex strategies that involve a mix of FDI and exports. In most situations these can be classified as neither purely horizontal nor purely vertical, and involve the export of intermediates and/or final products.

Models of export platform FDI simplify the analysis to a single product firm (albeit with multiple stages of production). An alternative set of models consistent with the idea that multinationals may also export comes from the literature on complementarity (Head and Ries, 2004). Again there are horizontal and vertical elements to this. In a multi-product firm, exports and FDI become positively correlated if there are horizontal or vertical complementarities across product lines. For example, in the case of horizontal complementarities increased demand for the good supplied by foreign production may lead to increased demand for all goods produced by that firm, some of which may be supplied through arms-length trade. For vertical complementarities the establishment of a plant in a foreign country to produce or assemble final goods will displace the exports of this product, but at the same time increase exports of intermediates from the home country. Net complementarity may arise if the displaced export of the final good is more than compensated by increased exports of intermediates.

Empirical evidence on the export decision of multinationals has concentrated largely on direction of correlation, whether positive or negative, rather than explanation. In all cases, at the firm level, this relationship has been found to be positive, for example Lipsy and Weiss (1984) for the US, Swedenborg (1985) for Sweden, and Lipsy et al. (2000) and Kiyota and Urata (2005) for Japan. Attempts at understanding the explanation for any correlation are limited to Head and Ries (2003), Kiyota and Urata (2005) and Girma et al. (2005a). The first two test for the effect of vertical FDI on exports using export demand equations for the firm (both for Japan) and find similar results. Head and Ries (2001) find complementarity between exports and FDI for the most vertically integrated firms and substitution.
can be found for the least integrated, whereas Kiyota and Utata (2005) find that intra-firm exports grow faster than total exports—with increased FDI some of the inter-firm exports shift to intra-firm exports. By contrast Girma et al. (2005b) test for export platform FDI for the UK. They find foreign multinationals tend to acquire domestic firms that export – they cherry-pick the best firms. However there are differences in the post-acquisition export trajectories of acquired firms according to whether they is inside or outside the EU. For firms outside, export intensity rises, whereas it falls for firms inside. This appears consistent with export platform motives as discussed by Motta and Norman (1996).

4. Future Research Issues and Policy Dimensions

4.1 Future Research Issues

A review of the tables associated with this evaluation and references appended confirm how rapidly the literature has grown. It has also generated genuinely new insights, particularly with regard to the determinants of exporting. However, it is also a progressive research agenda in the sense that there is both unfinished business and new research questions being raised.

As we have seen, some aspects of the export decision have received more attention than others. For example, while much is known about the characteristics of exporters and non-exporters and what happens when a firm enters export markets, relatively little empirical work has been conducted around the question of choices that firms make between exports and FDI. To a degree this is data driven, given the demanding requirements of the underlying models. Since little may change with respect to data availability, or at least change only slowly, this suggests that future empirical work is likely to continue along current lines, with some spread to questions where the data constraints are not so severe. Tests of export-FDI models are also likely to remain specific to more data rich countries such as the US, Japan and Sweden. A new strand of empirical analysis does appear to be emerging from the predictions of the heterogeneous firm models that provides some insight about the export-FDI choice of firms however. That is the dynamic consequences of changes in the costs of exports and FDI. Perhaps the earliest example of this is by Pavcnik (2002), who studies the within firm and between firm productivity effects of trade liberalisation in Chile.

Although the evidence base points unambiguously to the crucial role of sunk costs, little research has as yet focused on what these are, and how agglomeration, exchange rates and policy changes affect them. Whilst many researchers go through the motions of commenting on (for example) changes in product design, setting up distribution channels and so on as possible sources, that is generally as far as it goes. Sharper insights are needed if we really are to understand firm
heterogeneity. This will rely on merging datasets and/or firm and industry specific survey based enquiry. A recent example of the former, which investigates the role of access to credit is Greenaway, Guariglia and Kneller (2005). A fourth issue, which again depends on merging datasets is the role, if any, of the origin and destination of trade/FDI. As we saw in section 1 (extensions of the Melitz model to incorporate country asymmetries) and section 3 (North-South FDI models) origin and destination are likely to affect outcomes. Moreover, they may be key to understanding some of the empirical findings reported in section 2. For example, it may be that potential learning from exporting is fashioned by the markets into which one exports.

Finally, a new strand of research is being pioneered by Antras (2003) and Antras and Helpman (2004) exploring the implications of heterogeneity for the boundaries of the firm and strategies for outsourcing and insourcing of activities. This is a potentially rich vein of research, yielding new insights into globalisation and industrial organisation. Empirically however research here will be even more challenging given the need for disaggregated data on trade in intermediates, mapped on to firm specific information.

4.2 Policy Dimensions

Intervention to promote exports is very widespread – every WTO Trade Policy Review\textsuperscript{17} contains a chapter on Measures Directly Affecting Exports and there are always measures to report. These range from intervention to improve market intelligence (public support for trade missions), to sector specific fiscal intervention (tax concessions or duty drawbacks), to export processing zones (free zones).

Such a widespread commitment to a specific policy agenda is unusual and the commitment to export promotion has historically been driven by a presumption that export growth and output growth are positively correlated. Although theoretical models linking openness and economic growth are not unequivocal, a large empirical literature points to a positive correlation, even if the direction of causality is controversial. Be that as it may, the key point is that intervention is motivated by macro-econometric evidence. Does the microeconometric evidence we have reviewed reinforce or undermine a case for active promotion? López (2005) asks this question and concludes that it reinforces the macro evidence. He argues that even if self-selection is the key driver of export market entry, it may nevertheless be conscious self selection, especially in developing countries. What he means is that firms consciously improve their productivity with the international

\textsuperscript{17} The WTO’s Trade Policy Review Mechanism ensures that the trade policies of Members are audited on a regular basis. For the big three (US, EU and Japan) this means every two years; for the smallest Members, it takes place every seven years.
market in mind, rather than the best firms just starting to export. Policy intervention could then stimulate more conscious self-selection and deliver a productivity boost. Clearly if learning by exporting does occur, productivity gains are boosted further. Moreover, if there are spillovers, perhaps because non-exporting firms learn to export from other (domestic or multinational) exporting firms, the case is strengthened.

This is a plausible argument, though it could only underpin a case for general rather than targeted intervention. López (2005) himself stresses the importance of reducing (overseas) barriers to exports, which clearly aligns with other arguments for trade liberalisation. To this should be added internal barriers to export, chief among which is domestic import protection, since as the incidence of protection literature shows, import tariffs are taxes on exporting. If sunk costs are important, one can think of intervention to improve aspects of infrastructure as relevant – improving information flows, promoting clustering and so on. If policy makers wanted evidence to support intervention targeted at specific sectors or firms, that would require much more information than we have access to at present. For example, are entry costs higher for small firms? is access to credit a barrier? and so on. In the absence of more robust evidence, targeted intervention to support exporting firms is subject to the same risks as identifying so-called infant industries and the record on that front is not a good one.

5. Conclusions

This article has synthesized and evaluated a new literature linking firms, trade and cross-border investment. Its starting point was a well-known feature of the real world, firms that export and others that do not co-exist in the same industries. Until recently, this was not well explained by core trade models. This has changed with the development of heterogeneous firm models. These explain how firms that export are more productive and this, together with the reallocation of output which occurs as less productive firms contract or go out of business, points to a direct link between exporting and productivity. The framework has been extended to allow for the fact that some firms choose to produce overseas rather than export. The empirical literature has grown fast and as we have seen extends across a large number of industrialized, transitional and developing countries. Moreover this literature points to a number of regularities: exporting firms do tend to be larger and more productive than non-exporters; sunk costs appear to be important; multinational firms tend to be more productive than domestic firms. Other evidence is less conclusive however, such as that relating to learning by exporting. We have learned a lot in a remarkably short space of time, but as we saw in the last section, a rich research agenda has been thrown-up and this is a literature that will continue to grow.
References


