Regional Agreements and Welfare in the South: When Scale Economies in Transport Matter

CARRERE, Céline


DOI: 10.1093/jae/ejt029

Available at:
http://archive-ouverte.unige.ch/unige:45372

Disclaimer: layout of this document may differ from the published version.
Regional Agreements and Welfare in the South: When Scale Economies in Transport Matter

Céline Carrère*
Global Studies Institute - University of Geneva, CEPR and FERDI, Geneva, Switzerland

* Corresponding author: Carrère Céline. E-mail: celine.carrere@unige.ch

Abstract

By taking into account scale economies in transport, this paper challenges the accepted pessimistic view that regional preferential trade agreements (PTAs) between developing countries are harmful in welfare terms. In this paper, we assume the adoption of new transport technology when trade increases and show that, given the standard effect of a PTA on regional trade, the welfare would be higher than that usually claimed due to the induced effect on the regional transport network. Moreover, there is evidence that with such sunk costs in transportation, the sequence of trade liberalisation matters: the free trade achieved under a regional PTA would lead to permanently higher welfare than the one achieved under multilateral liberalisation. A standard model of inter- and intra-industry trade is used and augmented by a ‘hub-and-spoke’ transport network structure, where transport costs depend on the distance between trade partners, the volume of trade and the level of development. Under a plausible parameterisation for scale economies in transport, regional liberalisation will have persistent effect on trade flows through an irreversible effect on regional transport costs that improve welfare.

JEL classification: F12, F15, R4, O1

1. Introduction

Ever since Viner’s (1950) pioneering study, the ambiguous impact on the welfare of regional preferential trade agreements (PTAs) has been analysed
in terms of the relative magnitude of trade creation and diversion. Since 20 years, transport costs have been recognised among the factors that could influence this trade-off. Wonacott and Lutz (1989) first argued that regional PTAs are more likely to be welfare enhancing when formed among what they called ‘natural trading partners’, i.e. countries that are geographically close. Krugman (1991a, b) developed and popularised this idea in a monopolistically competitive framework showing that continental free trade areas (i) decrease welfare unambiguously with zero inter-continental transport costs and (ii) increase welfare unambiguously with prohibitive inter-continental transport costs. Relying on simulations where transport costs take continuous values between zero and prohibitive values, Frankel et al. (1996) conclude that all else constant, a preferential trade agreement is the more likely to be welfare enhancing (i) the more remote the continental trading partners are from the rest of the world (i.e. the larger inter-continental transport costs are) thereby limiting potential trade diversion and (ii) the more 'natural' (i.e. the closer in distance) trading partners are thereby fostering potential trade creation (see Baier and Bergstrand (2004) for a complete survey on simulation results).

The ‘natural trading partner’ argument potentially concerns 70% of existing PTAs.1 It is particularly relevant for PTAs between developing countries (or regional ‘South–South’ agreements), notably in Sub-Saharan Africa (SSA) where many PTAs are implemented between neighbouring countries that are quite remote from major world markets. Actually, though developing countries benefit from some recent technological advances that reduce transport costs, extensive documentation attests to the fact that they still face considerably higher transport costs than developed countries. Shipping costs, for instance, are dramatically higher for developing countries according to the price quotes from international freight forwarders (see Hummels, 2007; Limao and Venables, 2001 or Busse, 2003). Geographical impediments (such as landlockedness) and poor transport infrastructure (Limao and Venables, 2001) contribute to these high transport costs in developing countries. Moreover, as shown by Hummels and Schaur (2013), Clark et al. (2004) and Djankov et al. (2010), other cost-raising factors, such as time in shipping or custom clearance further increase transport costs for developing countries.2

1 On the actual 248 PTAs on goods in force in January 2013 (i.e. notified to the GATT/WTO), 170 are implemented between countries of a same region. Source: World Trade Organization secretariat and Author’s calculation.

2 Fifteen of the 28 landlocked developing countries are in SSA, the host of many regional PTAs. Conservative estimates put African freight costs at over twice the world average (see, for example, the estimates in African Development Bank, 2010, figure 1.5).
So far, the literature has characterised the environment describing developing countries by assuming ‘iceberg’ transport costs (à la Samuelson, 1954) which supposes that only a constant fraction of the quantity shipped actually arrives (as if ‘only a fraction of the ice exported reaches its destination as unmelted ice’). Virtually, all simulation models so far analysing the welfare of regional trade liberalisation have relied on this representation of transport costs, thus ignoring the potential effect of scale economies in transport (e.g. Frankel et al., 1996; Frankel, 1997; Spilimbergo and Stein, 1998; Baier and Bergstrand, 2004). Simply put, transport costs are assumed unaffected by equilibrium quantities traded. In this paper, I relax this constancy of transport costs with regard to aggregate quantity of trade.

There is now strong direct evidence of the importance of scale economies in shipping costs. Using a dataset covering the bilateral trade of six importers (Argentina, Brazil, Chile, Paraguay, Uruguay and the United States) with all exporters worldwide in 1994, Hummels and Skiba (2004) find that doubling trade quantities along a route reduces shipping costs by 12% for all countries on that route. The same order of magnitude is reported by Mori and Nishikimi (2002) and Skiba (2010). Fink et al. (2002), on studying the liner transport price on all US imports carried by liners from 59 countries in 1998, also find significant economies of scale with regard to traffic originating from the same port.

What are the sources of these scale economies in shipping? Sanchez et al. (2003) and Hummels and Skiba (2004) identify three main sources of reductions in transport costs as trade quantities increase. First, a densely traded route allows for an effective use of ‘hub-and-spoke’ shipping economies. Second, increased quantities traded encourage the introduction of specialised transport technologies along a route (as standardised containerised shipping for maritime transport). A third source of scale benefits lies in pro-competitive effect in pricing (limiting the monopoly mark-ups of, for instance, the ‘liner conferences’) to which I return shortly.

In the model developed in this paper, I focus on the second source of scale economies in transport: the adoption of new transport technology when trade increases (the first source is exogenously imposed by a pre-determined ‘hub-and-spoke’ transport network built into the model). As to the endogenous market structure, I take the extreme, but representative of many countries in Africa where transport is provided by a monopolist. Hence, in the model, according to the volume traded, a monopoly shipper decides whether to pay sunk costs (such as investment in infrastructure) in order to adopt a lower marginal cost transport technology.
The justification for this approach stems from the importance of the welfare-reducing effects of high transport costs associated with a low trade volume, and further exacerbated by low competition intensity. For example, the problem of ‘cargo reservation schemes’ and liner conferences, whereby only one shipping company will cover a route because of low traffic densities, leads to monopoly practices (see the evidence in Hummels 2007; Fink et al., 2002). In this environment, the adoption containerisation is delayed and containerisation has yet to spread to low-income countries.

Moreover, the set-up developed in this paper takes into account a critique that has been raised against the ‘natural trading partner hypothesis’, namely that differences in costs determined by comparative advantage could be an important factor weighting against the benefits of trade between close partners. As noted by Panagariya (1998, p. 294), ‘distant partners can be efficient suppliers of certain products due to other cost advantages despite the fact that they must incur higher transport costs.’ In a comment to a model by Frankel et al. (1996), Krugman (1998, p. 115) also notes that the restriction of identical economic size may not be innocuous and ‘surely makes a major difference when we try to model the effects of integration’. In this paper, I take the view that differences in costs related to economic size are sufficient to address the concerns raised by Panagariya (1998) since larger countries will have lower production costs, while maintaining the parsimony afforded by an otherwise symmetric modelling framework.

Based on this evidence and stylised description of the transport sector in developing countries, the paper addresses the issue of the welfare costs of regional PTAs by answering two questions. First, how is the standard welfare analysis of regional trade liberalisation affected by the endogeneity of transport costs (i.e. if trade quantities and transport costs are jointly determined)? By boosting bilateral trade among members, regional liberalisation exploits scale economies along regional routes (through the adoption of new transport technologies) and then leads to a reduction in transport costs. This joins the idea of Laussel and Riezman (2008) who conclude that, in the presence of fixed (but not sunk) transportation costs, some countries may require some kind of ‘big push’, for instance, some public investments in transport infrastructures and some coordination between governments, to get out of the autarkic trap.

Second, the sequencing of trade liberalisation matters. Freund (2000) evidences that, with sunk costs, the welfare level is higher when free trade is reached through expanding regional trading blocs than when free trade is accomplished by multilateral negotiations. In the same way, what are the consequences of endogenous transport costs for welfare if worldwide free trade is
achieved via PTAs rather than via multilateral trade liberalisation? Suppose that the long-run objective is worldwide free trade. With exogenous transport costs (i.e. transport costs independent of the level of trade), the welfare achieved under worldwide free trade is independent of the chosen path (i.e. via regionalism or multilateral liberalisation). Now, suppose the endogenous transport technology. Sequencing then matters because of sunk costs in transportation. Regional PTAs will then generate persistent effects on member countries’ trade flows through the investment in sunk costs in the regional hub and then on final welfare when they liberalise trade by regional route.

This paper is organised as follows. Section 2 adapts a standard monopolistic competition model with inter-industry trade in an agricultural good produced under constant returns to scale and a manufacturing sector under monopolistic competition borrowed from Spilimbergo and Stein (1998). An explicit ‘hub-and-spoke’ transport sector with a profit maximising monopolist choosing endogenously his transport technology completes this stylised representation of production and trade in a developing country. Section 3 applies the model to a four-continent world with two types of countries: North and South, which differ by size and economies of scale. As a start, Section 3 compares the welfare evolution according to the degree of tariff preference within symmetric regional bloc (i.e. blocs within neighbour countries of South–South and North–North type) with exogenous/endogenous transport costs. Section 4 then tackles the sequencing issue by contrasting welfare results when worldwide free trade is achieved through a regional path versus a multilateral non-discriminatory path. Section 5 studies how sensitive the results are to the ‘hub-and-spoke’ transport network assumption. Results are also extended to North–South agreements. Section 6 concludes.

2. Overview of the Model

2.1 Basic set-up

The model includes three sectors augmented by a transport sector developed in Section 2.2. As in Spilimbergo and Stein (1998), we assume three sectors: agriculture, intermediate inputs and manufacturing, and two factors of production: capital \( (K) \) and labour \( (L) \). We consider two types of countries, which differ only in their capital endowment. In ‘poor’ countries (subscript, p), each individual is endowed with one unit of capital, as well as

\[ 3 \text{ Appendix A.1 in Carrère (2007) provides a complete description of the model.} \]
one unit of labour. In ‘rich’ countries (subscript, r), each individual owns one unit of labour and \( k \) units of capital (where \( k > 1 \)). Imposing symmetry within groups and a similar model structure across country groupings improves significantly the tractability of the model while capturing, in a stylised way, the main features necessary to include transport costs and preferential trade policy.

A representative consumer in country \( i \) share a Cobb–Douglas utility function given by

\[
U_i = (C_{mi})^\alpha (C_{ai})^{1-\alpha}0 < \alpha \leq 1, \quad i = \{r; p\}. \tag{1}
\]

With \( C_{m(ai)i} \) the consumption of manufactures (agriculture) in country \( i \), \( \alpha \) \((1-\alpha)\) the share of consumer’s income spent on manufactures (agriculture).

In agriculture, a homogeneous good is produced under constant returns to scale with labour as the only input and with labour productivity set to unity. Production of the agricultural good in country \( i \) (denoted \( q_{ai} \)) is then given by

\[
q_{ai} = L_{ai}, \quad i = \{r; p\}, \tag{2}
\]

\( L_{ai} \) being the total number of workers in the agricultural sector of country \( i \).

Therefore, under perfect competition

\[
p_{ai} = w_i, \quad i = \{r; p\}. \tag{3}
\]

with \( p_{ai} \) the price of agriculture and \( w_i \) the wage in country \( i \).4

A final manufactured good is produced for domestic consumption \( (q_{mi}) \) under a Dixit–Stiglitz technology for intermediate inputs with constant returns to scale, each intermediate input entering symmetrically into its production

\[
q_{mi} = \left( \sum c_{ji}^\theta \right)^{1/\theta}, \quad 0 < \theta < 1 \quad m, j = 1, \ldots, n_i, \quad i = \{r; p\}, \tag{4}
\]

\( c_{ji} \) being the consumption of the \( j \)th variety produced in country \( i \), \( n_i \) is the number of intermediate input varieties produced in country \( i \) and \( \theta \) capturing the extent of product differentiation across intermediates of different origin (‘love of variety’). As \( \theta \) approaches 1, the elasticity of substitution \( \sigma = 1/(1-\theta) \rightarrow \infty \), and intermediates of different origin become perfect substitutes, intra-industry trade is eliminated and only inter-industry trade remains.

Intermediate inputs are produced with capital as input under monopolistic competition. Increasing returns to scale is captured by assuming a fixed cost,

4 The wage in a poor country, \( w_p \), is used as numéraire in the model.
\[ K_{ji} = \gamma + \beta x_{ji} \iff x_{ji} = \frac{K_{ji} - \gamma}{\beta}, \quad j = 1, \ldots, n_i, \quad i \in \{r; p\} \]  

where \( x_{ji} \) is the production of the \( j \)th variety in country \( i \) and \( K_{ji} \) is the total amount of capital used in the production of the \( j \)th variety in country \( i \). Profit maximisation combined with free entry gives output per variety

\[ x_{ji} = x = \frac{\theta \gamma}{\beta(1 - \theta)}, \quad j = 1, \ldots, n_i, \quad i \in \{r; p\}. \]  

Adding the capital constraint in each country to equation (6) implies that the number of varieties produced in equilibrium is determined by relative country-size here captured by relative endowments of capital

\[ n_i = \frac{K_i(1 - \theta)}{\gamma}, \quad i \in \{r; p\} \quad \text{and} \quad \frac{n_r}{n_p} = \frac{K_r}{K_p} = k. \]  

A larger number of varieties produced in countries well endowed in capital lead to lower unit production costs in these countries, thereby introducing indirectly the concern of factor endowment-based models that differ in costs matter.

2.2 Transport costs and geography

Following the literature, I consider a symmetric world divided into a number of continents, \( C \), equidistant from one another and comprising the same number of countries, regions and blocs. There are 4 continents (\( C = 4 \)) and 64 countries (32 rich countries spread over 2 continents and 32 poor countries over the other two continents). Each continent is decomposed into four regions. I assume that each PTA bloc is implemented between the four neighbouring countries of a same region. This allows us to concentrate on ‘North–North’ and ‘South–South’ blocs, leaving other alternatives (such as North–South trade) to later consideration. This stylised World is relevant for Sub-Saharan Africa (SSA), roughly comprising four regions: ECOWAS, CEMAC, SADC and COMESA\(^5\), quite remote from the Northern market. Moreover, this stylised World will imply overseas North–South agreements in Section 5, once again relevant for SSA and the ongoing implementation of

EPAs (Economic Partnership Agreements between the EU and African, Caribbean and Pacific group of countries).6

As in the recent literature on economic geography (e.g. Frankel et al. (1996); Spilimbergo and Stein, 1998; Fujita et al., 1999), I assume a ‘hub-and-spoke’ transport network. This is in accordance with the emergence, in recent decades, of transport hubs as a privileged network structure for many types of transport services, notably for freight and air.7 This is consistent with the assumption of scale economies in the transport sector, as it is precisely the search for lower unit transport costs that has generated the development of ‘hub-and-spoke’ transport networks (see, e.g. Mori and Nishikimi, 2002).

In this set-up, each country represents a ‘spoke’ and two levels of ‘hub’ are assumed: regional and continental. For instance, in the case of SSA and the four main regions defined above, the corresponding regional hubs in terms of port capacity would be Durban, Abidjan, Douala and Mombasa.8

As shown in Figure 1, three freight rates (in % of the quantity traded) characterise transport costs:

- $f_b$: intra-bloc or intra-regional (from spoke to spoke via the regional hub)—b stands for blocs;
- $f_c$: intra-continental (from a regional hub to another via the continental hub)—c stands for continent and
- $f_o$: overseas or inter-continental (from a continental hub to another)—o stands for overseas.

Trade between two countries in the same region involves two spokes and one regional hub, which implies transport costs equal to $f_b$. Similarly, in the case of trade between countries in different regions of the same continent, transport

---

6 The counterpart is that this stylised world restricts somewhat the applicability of results to cases, such as North America, Western and Eastern Europe, and East Asia.

7 In the case of maritime transport that largely dominates international trade, vessels size increased drastically in relation to the development of containerisation. Container traffic is moreover essentially concentrated in major hub ports. The 20 largest container ports handled more than 52% of all the traffic in 2002. Examples include the European hub of Rotterdam, as well as Asian hubs in Singapore and Hong Kong (see Review of Maritime Transport, UNCTAD, 2003). The extraordinary development of ‘hub-and-spoke’ networks is also observed in airlines.

8 Hub ports are large regional ports, with high volumes of direct large-vessel calls. They service a large catchment area, which also serves the smaller regional ports by transshipping containers and general cargo in smaller vessels—see map 2.1, p. 36 in the ADB (2010) report.
costs are equal to \((f_b + f_c)\) as two spokes, two regional hubs and one continental hub are implicated. Finally, across-ocean trade generates costs of \((f_b + f_c + f_o)\). Hence, implicitly, transport costs depend positively on distance.

I approach the modelling of transport costs in two ways: the traditional ‘iceberg’ approach where \((f_b, f_c, f_o)\) represent the fraction of output lost by the exporting country en route to its destination (as in, e.g. Frankel et al., 1996; Frankel, 1997; Spilimbergo and Stein, 1998; Baier and Bergstrand, 2004) and one where \((f_b, f_c, f_o)\) represent the freight rate charged by a monopolist (see below). These two alternatives are presented in Table 1. Since only relative transport costs matter, I assume \(f_b = f_c = 0\) for rich countries to reflect the fact that transport costs vary according to the development level of countries. For a given distance, North–North trade is less costly in terms of transport costs than North–South trade, which is in turn less costly than South–South trade. Finally, for simplicity, I assume equal transport costs for intermediate inputs and agriculture products.

Letting \(p_r(p)\) be the producer price in a rich (poor) country, these assumptions on geography and transport costs give rise to the cost, insurance and freight (c.i.f.) prices reported in Table 1.

Table 1 presents that transport costs: (i) increase the prices of foreign intermediate inputs faced by producers of manufactures and (ii) increase the difference in relative price of agriculture goods between rich and poor countries which in turn increase the wage gap between rich and poor countries. Tariffs are levied on c.i.f. prices.
### Table 1: c.i.f. Prices Under Alternative Transport Models.

<table>
<thead>
<tr>
<th>From</th>
<th>With endogenous transport costs (% of the traded quantity)</th>
<th>With ‘iceberg’ transport costs (% of output lost by the exporting country en route)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>In a rich country (1)</td>
<td>In a poor country (2)</td>
</tr>
<tr>
<td><strong>c.i.f prices of imports for intermediates purchased by producers of manufactures</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bloc members</td>
<td>$p_{br} = p_r$</td>
<td>$p_{bp} = p_p + f_b$</td>
</tr>
<tr>
<td></td>
<td>$p_{cr} = p_r$</td>
<td>$p_{cp} = p_p + f_b + f_c$</td>
</tr>
<tr>
<td>Other countries on the same continent</td>
<td>$p_{or} = p_r + f_o$</td>
<td>$p_{or} = p_r + f_o/2 + f_c/2 + f_o$</td>
</tr>
<tr>
<td>Across ocean rich countries</td>
<td>$p_{op} = p_p + f_b/2 + f_o/2 + f_c/2 + f_o$</td>
<td>$p_{op} = p_p + f_b + f_c + f_o$</td>
</tr>
<tr>
<td>Across ocean poor countries</td>
<td>$p_{ar} = p_p + f_b/2$</td>
<td>$p_{ar} = p_p + f_b/2 + f_c/2 + f_o$</td>
</tr>
<tr>
<td><strong>c.i.f. prices for agriculture good (imported by rich from poor countries)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Across ocean</td>
<td>$p_{ar} = p_p + f_b/2 + f_c/2 + f_o$</td>
<td>$p_{ar} = p_p/((1 - f_b)(1 - f_c)(1 - f_o))$</td>
</tr>
</tbody>
</table>
2.3 Transport sector with scale economies

As mentioned in the introduction, for maritime transport, many trade routes are serviced by a small number of liner companies organised in formal cartels called ‘liner conferences’ (see Hummels, 2007; Hummels et al., 2009). Moreover, a movement towards concentration has occurred which would not imply market power if transport routes were contestable. At least one study, by Fink et al. (2002), has found evidence that freight rates are sensitive to regulatory changes meant to constrain collusive behaviour by liner conferences, suggesting the exertion of market power. More recently, Hummels et al. (2009) show how market power leads to systematically higher shipping prices in the developing world and find that eliminating market power in shipping would boost trade volume by 15.2% for Latin America. More generally, for developing countries, several studies indicate that factors such as national policies that severely restrict competition for transport services have a major influence on the level of freight rates.

Modelling the transport sector as a monopoly presents two advantages. First, it captures the monopoly mark-up often observed in transport service prices on two types of routes: maritime (corresponding to transport between two continental hubs in our framework) and within the South continental hub. Second, investment in new transport technology can be easily introduced explicitly in the model as a function of the shipper’s profit.

As proposed by Hummels and Skiba (2004), I assume that a monopoly shipper takes decisions about how to price transport services and which

---

9 Only a dozen firms in the world share 80% of the container traffic (against 40%, 10 years ago). The two leaders, accounting for more than 23% of the traffic, reinforced their domination by taking over hub ports and signing agreements (as the TransAtlantic Container Agreement) thereby forcing loaders to deal with them (see Rodrigue et al., 2004).

10 For instance, much of SSA international transport is cartelised, reflecting the regulations of African governments intended to promote national shipping companies and airlines (see, for example, African Development Bank, 2010 and World Bank, 2009, who describe the ‘cargo reservation schemes’ and UNCTAD’s Liner Shipping Connectivity Index which takes into the account the cost-raising factors contributing to the high transport costs for SSA estimated in Portugal-Perez and Wilson, 2009).

11 Note that the monopoly assumption does not concern transport within the North continent as we have assumed \( f_B = f_C = 0 \) for rich countries. Actually, we have some insights that the market power of liner conferences is strongly attenuated on routes involving OECD countries (e.g. the United States Shipping Acts of 1984 and 1998 or the EU Regulation 4056/86 putting an end to the possibility for liner carriers to meet in conferences, fix prices and regulate capacities as of October 2008). Moreover, as I focus on South–South integration, the hypothesis on the transport market involving developed countries has no impact while allowing an appreciable simplification.
transport technology to use, maximising the following profit function, \( \pi \):

\[
\pi = f_b q_b + f_c q_c + f_o q_o - C_b - C_c - C_o. \tag{8}
\]

With \( q_b, q_c \) and \( q_o \) being the total traded quantities requiring, respectively, intra-regional, intra-continental and across ocean transport services and \( C_b, C_c, C_o \) being the cost functions associated with the production of transport \( f_b, f_c \) and \( f_o \), respectively.

Transport costs along a given route \((h \in b, c, o)\) decline with the increase in the volume of trade along that route by adopting the following technology with fixed (or sunk) costs, \( F_h \), and constant marginal costs, \( \kappa_h \) per unit shipped:

\[
C_h = F_h + \kappa_h q_h, \quad h = \{b; c; o\}. \tag{9}
\]

To produce transport services, without loss of generality, the monopolist uses labour from the poor country where labour costs are lower. Each transport technology is characterised by the combination of parameters \( \{F_h; \kappa_h\} \). The initial technology is assumed to require no fixed costs, \( F_h = 0 \), but has a high marginal cost per unit shipped. Then, as trade quantities along a route increase, the monopolist can choose to improve the transport technology used on that route, i.e. to purchase a reduction in the marginal cost of \( \Delta \kappa_h \) with an incremental fixed cost \( F_{h'} \) according to the following relation:

\[
F_h = e^{\mu \Delta \kappa_h} - 1, \quad \mu < 0, \quad h = \{b, c, o\}. \tag{10}
\]

In this set-up, changes in technology are discrete\(^\text{12}\), irreversible and occur only when the profit associated with the new technology surpasses the profit associated with the old one.\(^\text{13}\) This means that even if the monopolist would be able to anticipate a reverse in the trade trend in near future (due, for instance, to the end of the regional PTA), the monopoly invests today and would still make pure profit in the case of a decrease in trade as the fixed cost has been paid. Note that equation (10) assumes that a given reduction in marginal cost

\(^{12}\) As noted by Hummels and Skiba (2004), ‘one can think of this choice either as a single yes/no decision on, for example, port infrastructure or [ ... ] as a menu of ship sizes which the shipper can select’. And as described by Skiba (2010), the containerisation of cargo, or the use of larger capacity ships significantly lowers the unit cost of shipping but requires an incremental investment, more likely to be done for large shipping volumes.

\(^{13}\) Here, we are in the ‘low’ hypothesis: all investments in key fixed costs are done by the shipper thanks to increasing trade volume at the regional level. The ‘high’ hypothesis would be that there exists a regional cooperation between countries to implement provisions for investment in infrastructure (e.g. Bond, 2007; Laussel and Riezman, 2008). Under this hypothesis, the shipper will adopt new technologies faster (as part of the fixed costs will be not entirely financed by its profit) which reinforces the ‘virtuous circles’ when a PTA is implemented.
requires greater fixed costs when marginal costs are already small than when marginal costs are high and, to ease interpretation, a given investment generates a similar reduction in marginal costs ($\mu$ constant) whatever the selected route $h$ (regional, continental or inter-continental).

The parameters entering the cost function in equation (10) are calibrated using estimates in the literature as follows. Start with the most costly technology: $\kappa_h = 5\%$ and $F_h = 0$, $h = b, c, o$. To anticipate results of the simulations reported in Section 3 where regional integration starts from an initial situation with a non-discriminatory (i.e. MFN) tariff on imports of $t = 30\%$, the prices of transport services that maximise the shipper’s profit are the following: $f_b = 9.6$, $f_c = 10.1$, $f_o = 6.5\%$, which implies transport costs in the $10–20\%$ (of quantity traded) range for a representative poor country (see Table 2, column 2). These are in accordance with estimates on the level of transport costs sustained by developing countries (see Limao and Venables, 2001; Hummels, 2007; Hummels and Schaur, 2013).

Consider now economies of scale. According to remarkably similar econometric estimates for different regions of the world by Hummels and Skiba (2004) and Mori and Nishikimi (2002), a $1\%$ increase in trade volume along a route reduces freight rates by $0.12\%$ for all countries on that route. In the same order of magnitude, Skiba (2010), using the data of five Latin American importers from all exporters worldwide in 1994 and at the six-digit level of the harmonised system confirms that a $10\%$ increase in regional volume of shipping reduces transport cost by about $1.3–2.1\%$ in the long run. Thus, I assume that each investment in new technology induces a gain in marginal cost of $0.2$ point of percentage, and I determine the value of $\mu$ (and of the fixed costs) that constrains the monopoly shipper to reduce freight rates charged along a route by around $0.12\%$ for each $1\%$ increase in trade volume along that route. The value of $\mu$ that satisfies the preceding constraint is $-15$. Hence, starting from the initial technology $\kappa_h = 5\%$ and $F_h = 0$, the next technology corresponds to a marginal cost of $4.8\%$ requiring a fixed cost around $F_i = e^{(-15)} \times (-0.2\%) - 1 = 0.03$ which represents around $10\%$ of the monopoly profit in the initial situation (i.e. under MFN and with the initial technology). Figure 2 illustrates average transport costs for the shipper as a function of distance under this calibration.

### 2.4 Equilibrium

Profit, utility maximisation and free entry in the production of intermediates lead to a vector of production and consumption in each country and to the corresponding factor and product prices. Departing from earlier
Table 2: Transport Costs and Welfare under Different Scenarios ($k = 3$, $\alpha = 0.5$, $\sigma = 4$, $C = 4$, $N_r = N_p = 16$, $B = 4$)

<table>
<thead>
<tr>
<th>Column</th>
<th>Iceberg (Benchmark)</th>
<th>Endogenous transport costs</th>
<th>Worldwide FT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Column (1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Transport costs on each route component (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intra-regional $f_b$ (from spoke to spoke via the regional hub)</td>
<td>10.0</td>
<td>9.6</td>
<td>8.0</td>
</tr>
<tr>
<td>Intra-continental $f_c$ (from a regional hub to another via the continental hub)</td>
<td>10.0</td>
<td>10.1</td>
<td>10.2</td>
</tr>
<tr>
<td>Inter-continental $f_o$ (from a continental hub to another)</td>
<td>10.0</td>
<td>6.5</td>
<td>6.4</td>
</tr>
<tr>
<td>Total transport costs between (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Two poor countries-same bloc</td>
<td>10.0</td>
<td>9.6</td>
<td>8.0</td>
</tr>
<tr>
<td>Two poor countries-same continent</td>
<td>19.0</td>
<td>19.7</td>
<td>18.2</td>
</tr>
<tr>
<td>Two poor countries-different continents</td>
<td>27.1</td>
<td>26.2</td>
<td>24.7</td>
</tr>
<tr>
<td>One poor country and 1 rich country</td>
<td>18.8</td>
<td>16.4</td>
<td>15.6</td>
</tr>
<tr>
<td>Two rich countries-same continent</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Two rich countries-different continents</td>
<td>10.0</td>
<td>6.5</td>
<td>6.4</td>
</tr>
<tr>
<td>Increase in $W^p$ (base = MFN situation) (%)</td>
<td>–</td>
<td>–</td>
<td>+0.50</td>
</tr>
</tbody>
</table>
contributions, the model also determines the profit maximising transport technology by a monopoly shipper (whose profits are symmetrically distributed to the representative consumer across rich countries) and the corresponding freight rates charges. Given the values of the ad valorem tariff rate ($t$), the degree of intra-bloc preference ($d$), the difference in capital endowment ($k$) and the parameters describing preferences, technology for production and transport, together with the wage normalisation $w_p = 1$, each equilibrium yields a value for the welfare indicator of a representative individual in a country.$^{14}$ The focus of attention in the remainder of the paper is how individual welfare in a poor country, i.e. $W^p$, changes under a trade policy organised around a trading bloc relative to a non-discriminatory policy. The full system of equations describing the model is reported in Carrère (2007, appendix A.1).

### 3. Welfare Implication of Preferential Trade Agreements

We start with welfare implications of PTAs, and then turn to multilateral trade liberalisation in Section 4. The set-up throughout assumes 4 continents ($C = 4$), 2 with only rich countries and 2 with only poor ones, with 16

$^{14}$ Following Frankel et al. (1996) or Spilimbergo and Stein (1998) the value of the utility of the representative individual is used as a measure of welfare. Then, the Welfare of a Southern (Northern) country is measured by the optimised value of the utility of the representative Southern (Northern) consumers. The optimisation problems of consumer in a poor and a rich country are reported in Carrère (2007, appendix A.1).
countries per continent ($N_r = N_p = 16$). Each continent has four regions. I assume that blocs are implemented between the four neighbour countries of a region ($B = 4$), which then imply only blocs between countries of the same development level. All countries are assumed to levy the same tariff rate of 30% on imports from non-members ($t = 0.3$) and to levy an intra-bloc tariff of $t_B = (1 - d)t$ on imports from member countries, where $d$ represents the preference margin within the bloc ($0 \leq d \leq 1$). Half of consumer income is spent on agricultural goods ($\alpha = 0.5$) and the elasticity of substitution among intermediate goods, $\sigma$, is equal to 4 (i.e. $\theta = 0.75$).\(^{15}\)

### 3.1 Traditional ‘iceberg’ transport costs

Since several patterns hold under both endogenous and exogenous transport costs, we start with ‘iceberg’ transport costs. This also helps us to relate the results under endogenous transport costs to previous ones, which all assumed exogenous transport costs. Figure 3 (and others) shows how welfare for a representative poor country, i.e. $W_p$, varies when the preference margin $d$ in favour of the regional partners increases. For each set of parameter values, welfare is normalised to 1 under the initial MFN world ($W_p^0 = 1$).

Note first that the inverted U-shape for $W_p$ as preferential margins increase for all configurations and parameter values. This typical second-best result was first noted by Meade (1955) with a slightly different model. Here, as in the Meade model, the marginal benefits from reducing the wedge decrease whereas the marginal costs of creating a wedge by discriminating between trading partners increases.\(^{16}\)

Start then with a totally symmetric world ($k = 1$). This implies that only intra-industry trade occurs between countries (agriculture is not traded, as there is no comparative advantage). All countries being identical in terms

---

\(^{15}\) I have chosen parameters similar to Spilimbergo and Stein (1998) and Baier and Bergstrand (2004) in order to have directly comparable results. A sensitivity analysis is made in Carrère (2007, appendix A.2).

\(^{16}\) More concretely, the initial reduction in intra-bloc tariffs leads to a small amount of trade diversion following the shift away from foreign varieties that were consumed in similar proportions for $d = 0$ (and no transport cost). At the same time, trade creation effects are large because domestic varieties (with smaller marginal utility, as they are already consumed in large quantities) are replaced by the bloc members’ varieties. Approaching the last reduction in intra-bloc tariffs ($d = 1$) however, consumption of member and domestic varieties are equalised with a small marginal gain, while the marginal loss of reduction in foreign varieties is now large: welfare effects of trade creation are then negligible while trade diversion effects are large.
of economic size, relative factor endowments, trade, tariffs and transport costs, the specification is very close to the monopolistically competitive framework in a perfect symmetric world proposed by Frankel et al. (1996) and Frankel (1997).

In the absence of transport costs \((f_b = f_c = f_o = 0)\), \(W^p\) reaches a maximum value \((W^p_{\text{max}})\) for a degree of intra-bloc preference of around 7% (which implies an intra-bloc tariff \(t_B = 28\%\)) and \(W^p < W^p_0\) for \(t_B = 26.2\%\) \((d = 12.6\%)\). Figure 3 shows that the introduction of positive ‘iceberg’ transport costs changes the relative magnitude of trade creation and trade diversion effects and then \(W^p_{\text{max}}\) but does not challenge the overall inverted-U path of welfare. With positive inter-continental transport costs and zero intra-continental transport costs, relative inter-continental transport costs increase, diminishing the volume of trade with remote countries (on other continents). As expected, reduced trade with remote countries diminishes the costs of implementing sub-continental PTAs, and hence also greater utility gains. As shown in Figure 3, with \(f_o = 0.2\), \(W^p = W^p_{\text{max}}\) for \(d = 11\%\) and for \(d > 20.2\%\) we have what Frankel et al. (1996) call the ‘supernatural zone’ to describe a welfare-reducing PTA (i.e. \(W^p < W^p_0\)) among natural partners.

Consider now an asymmetric world \((k = 3)\) as in Spilimbergo and Stein (1998). Not surprisingly, the welfare path of a poor country shown in Figure 3 is very similar to the path under total symmetry as introducing inter-industry trade with the distant (Northern partner) in effect destroys the positive effects of having lower relative transport costs within the Southern trading bloc.
Total welfare costs of discrimination are higher because countries derive utility not only from product differentiation in consumption, but also because of differences in costs.

Figure 3 also reports a simulation with both an asymmetric set-up and non-zero intra-continental transport costs for poor countries. Hence, transport costs are now a function not only of distance but also of the trade partner’s development level, as presented in Section 2.2, Table 1. This corresponds to our stylised representation of the world and I refer to it as the ‘benchmark’ in the following discussion. I assume \( f_0 = 0.1 \) and for poor countries: \( f_b = f_c = 0.1 \). Table 2 column 1 reports the corresponding bilateral transport costs derived from the formulas given in Table 1. As expected from non-zero intra-continental transport costs, Figure 3 indicates that the negative return of regionalism for a representative poor country sets in later (\( W_p < W_p^0 \) for \( d \approx 32\% \), i.e. \( t_B \approx 20\% \)).

### 3.2 Endogenous transport costs

Traded quantities and transport costs are now jointly determined along the lines described in Section 2.3: a monopoly shipper (monopolist for short) combinedly chooses profit-maximising prices and transport technology. The implications of this approach to endogenous transport costs are studied in two steps: in the first step, the monopolist fixes transport service prices with a single transport technology, then in the second step the monopolist combinedly chooses prices and transport technology. Figure 4 reports the evolution of the welfare indicator as a function of the preferential margin under both scenarios for the same regional PTA considered earlier.

#### 3.2.1 Single transport technology

Under the high-cost single transport technology described in Section 2.3, \( \kappa_h = 5\% \) and \( F_h = 0 \), \( h = b, c, o \), the evolution of \( W_p^\# \) appears to be less favourable to PTAs than the one obtained with exogenous ‘iceberg’ costs in Section 3.1 (benchmark from Figure 3 reported in Figure 4).

Tariff reduction, through the reduction of the elasticity of transport demand, causes the monopolist to charge a higher mark-up over marginal costs, which lowers trade creation, a result also obtained by Hummels and

\[ \text{17 For a detailed discussion of the results with non-zero intra-continental transport costs see Frankel (1997, pp. 320–21), Baier and Bergstrand (2004, pp. 42–4).} \]
Skiba in partial equilibrium and the shipper increases his price from 9.6% under MFN ($d = 0$) to 10.2% under free trade area—FTA ($d = 1$).

### 3.2.2 Endogenous transport technologies

Figure 4 shows that for the selected parameterisation, $W_p$ never enters the welfare-reducing zone when a regional PTA is implemented. Actually, a ‘virtuous circle’ is generated: the additive intra-bloc trade (due to the decrease in intra-regional tariff) increases the demand for intra-regional transport services, which leads the monopolist to adopt lower marginal costs technologies on these routes and then to offer a lower intra-regional freight rate, $f_b$, which in turn boosts intra-bloc trade and positively affects trade creation.\(^{18}\)

This optimistic conclusion is partly due to the parameterisation, which does not impose high sunk costs to obtain marginal transport cost gains. With higher fixed costs per unit decrease in marginal costs, the ‘jump’ to the associated higher welfare curves (the dotted lines in Figure 4, normalised to 1 under MFN regime and the first technology could be called iso-technology welfare curves) would occur later. Then, with more costly technologies, poor

\(^{18}\) Note that intra and inter-continental transport services demands, $q_c$ and $q_o$, respectively, decrease due to trade diversion. Hence, no new technology is adopted on routes between two regional and two continental hubs, respectively. However, as all trade flows have to pass through a regional hub, the improvement on regional routes (and the corresponding decrease in $f_b$) generates positive externalities for all routes (see Table 2 column 3) that mitigate the negative effects of trade diversion.
countries may sometimes, temporarily, enter the welfare-reducing zone (until the adoption of the next technology).\textsuperscript{19}

This said, the welfare curve shown in Figure 4 is in accordance with the econometric assessments of economies of scale in transport reported previously. Between MFN ($d = 0$) and a full regional FTA ($d = 1$) status, import demand increases by 133\% while the price of intra-regional transport services ($f_b$) decreases by around 16\% (see Table 2 columns 2 and 3). This estimate corresponds to the estimation suggested by the econometric evidence reported earlier, namely that ‘doubling trade quantities along a route reduces shipping costs by a 12\% on that route’.

4. The Sequencing of Trade Liberalisation

I now consider the sequencing issue (or path dependence) of trade liberalisation, recalling that under the traditional exogenous transport cost assumption, reaching free trade under multilateralism or regionalism would yield the same final value. Under the assumption that new transport technologies are not reversed (i.e. sunk costs), columns 4 and 5 in Table 2 contrast resulting transport costs under worldwide free trade under the two alternative paths. Column 4 indicates the transport prices charged by the monopolist if worldwide free trade is achieved by the following sequence: first simultaneous implementation of North–North (N–N) and South–South (S–S) FTAs followed by a removal of tariffs between blocs. Column 5 reports final transport costs when worldwide free trade is achieved via multilateral tariff reduction.\textsuperscript{20}

Comparing the values of the welfare indicator at the bottom of the table indicates a higher welfare when free trade is achieved under the regional route. This is due to: (i) the adoption of improved transport technologies on intra-regional routes to satisfy increased regional trade resulting from preferential tariff elimination as shown in Figure 4; (ii) gains made thanks to the sequencing whereby moving to free trade starting with regional free trade only requires developing two routes (intra-continental and inter-continental), whereas moving to free trade multilateral requires spreading

\textsuperscript{19} Appendix A.2 in \textit{Carrère (2007)} shows that the results obtained under the set of benchmark parameter values are robust.

\textsuperscript{20} We do not compare welfare reached under FTA and under worldwide free trade. Actually, except in the case of N–S FTA or in the case of ‘independent’ routes and endogenous transport costs (see Section 5), the welfare reached under FTA is always significantly lower than that under worldwide FT (see Table 2). In this section, FTA is only considered as a step towards worldwide free trade, not an end in itself.
transport cost savings on the three routes.\textsuperscript{21} As shown at the bottom of the table, of the $0.61 = 2.00 - 1.39$, $83\%$ ($= 0.50/0.61$) of the gain is due to the reduction in transport costs associated with the development of regional routes under regional FTAs.

Hence, with scale economies in transport costs and sunk costs, a symmetric (N–N and S–S) regionalism path to free trade has a persistent effect on trade flows through a permanent effect on regional transport costs that improves poor country welfare compared with the alternative multilateral path, i.e. $(W_{MN,SS}^{FT} > W_{MN}^{MFN})$. As part of the sensitivity analysis, next section explores an alternative path with the implementation of North–South (N–S) FTAs.

5. Extensions

5.1 Independent transport routes

As an alternative to the ‘hub-and-spoke’ transport network structure, I now assume that each country operates under three independent routes (also called ‘point-to-point’ network), each corresponding to one of the three kinds of trade partners: regional, continental outside the regional PTA and across ocean. It turns out that the patterns discussed here are robust to this alternative modelling of transports as far as only the regional routes are concerned. There is no significant difference between the two transport structures during the implementation of an FTA.\textsuperscript{22} Concerning the multilateral liberalisation stage, this conclusion is strongly reinforced: with an ‘independent routes’ network, $W^P$ under FTA (but no worldwide liberalisation) is superior to $W^P$ under worldwide free trade reached from an MFN situation! This is because with a multilateral liberalisation from an MFN situation (with $t = 30\%$), trade is spread too thinly among all partners so that the improved shipping technology is never adopted. As in Skiba (2010), PTAs are superior to MFN liberalisation with a ‘point-to-point’ network.

5.2 North–south regional blocs

The distinguishing feature of the current wave of regionalism is that it is now overseas N–S (rather than N–N and S–S during the first wave of regionalism in the 1960s) as illustrated by the ongoing implementation of the EPAs

\textsuperscript{21} Recall that, as described in Section 2, transport technology is specific to each kind of route.

\textsuperscript{22} As shown in Carrère (2007, appendix A.2, figures A.3).
(Economic Partnership Agreements between the EU and African, Caribbean and Pacific group of countries). Figure 5 contrasts welfare under N–S regionalism with welfare under S–S/N–N considered earlier. The evolution of $W_p$ during the N–S bloc implementation (i.e. bloc between two poor and two rich countries) is close to the evolution of $W_p$ under multilateral liberalisation (but with still the inverted U-shape) as now the two sources of gains from trade, product variety and costs differences can be exploited within the bloc.

In terms of reduced transport costs, symmetric blocs lead to a gain of 2% in regional marginal transport costs, whereas N–S blocs, in promoting trade on the three routes (regional, continental and across ocean), lead to gains that are spread out over the three routes (gain of 1% on each marginal transport cost, which is smaller than that under worldwide free trade).

As far as multilateral liberalisation is concerned, $W_p$ under worldwide free trade when reached through N–S regionalism is (i) higher than that through MFN liberalisation due to a higher volume of trade and to the adoption of a better technology on all three routes, but (ii) smaller than that through symmetric blocs due to less advanced regional transport technology. Then, for the representative parameterisation adopted here, the ranking of paths towards free trade, the asymmetric bloc approach yields a welfare gain of an elimination of protection in-between the alternatives examined earlier, i.e. $(W_{FT}^{NN}, SS > W_{FT}^{NS} > W_{FT}^{MFN})$.

6. Conclusions

This paper has challenged the pessimistic view that PTAs between neighbouring developing countries are likely to be welfare reducing. South–South trade
agreements look more favourable once one takes into account scale economies in transport (and the associated changes in transport technology from a profit-maximising monopoly shipper). For plausible parameter values, when a PTA is implemented with a Southern regional partner, a ‘virtuous circle’ is set in motion: preferential trade increases regional trade flows which allow investing in new transport technologies, improve the regional transport network and reduce intra-regional transport costs, which in turn boosts intra-bloc trade, etc. Moreover, in the long run, a regional approach to trade liberalisation may be preferable to a multilateral approach given the presence of irreversible effects in terms of investments in regional transport technologies. Of course, the analysis provided in this paper is the second best exercise. An alternative policy would be to directly subsidise transport networks and/or induce additional competition on the networks. Hence, we would have the desired effect of encouraging investment without necessarily having to introduce the potential distortions of a preferential trade agreement.

While these results are at best suggestive, they provide support to several recent regional PTAs and many policy recommendations. Many South–South PTAs (e.g. MERCOSUR, Andean pact, SADC, COMESA, UEMOA) have included ‘transport and trade facilitation’ agreements as part of their regional integration initiatives, the economic partnership agreements (EPAs) currently under negotiation between the EU and ACP involve a N–S FTA built upon a prior South–South FTA. More directly, the New Partnership for Africa’s Development (NEPAD) puts emphasis on investments in regional infrastructure and transport networks. And a recent exhaustive review of growth prospects for SSA identifies that poor infrastructure is the main bottleneck to sustained growth in the region and recommends that ‘deep regional integration’ (i.e. integration extending beyond the elimination of tariffs and other policy-imposed barriers to trade) is the key to sustained growth (World Bank, 2009). The challenge is to quantify these beneficial channels of regional integration with greater accuracy.

Acknowledgements

This paper has greatly been improved by the comments and suggestions of Jaime de Melo. I also thank Olivier Cadot, Stéphane Calipel, Riccardo Faini, Patrick Guillaumont, Thierry Mayer and Alexandre Skiba for helpful discussions on a previous version. I thank the editor and two anonymous reviewers for their constructive comments on this version. The usual disclaimer applies.


