Rising Trade Costs? Agglomeration and Trade With Endogenous Transaction Costs

Gilles Duranton

University of Pennsylvania

Michael Storper

Follow this and additional works at: https://repository.upenn.edu/real-estate_papers

Part of the Economics Commons, and the Real Estate Commons

Recommended Citation


At the time of publication, author Gilles Duranton was affiliated with the University of Toronto. Currently, he is a faculty member at the Real Estate Department at the University of Pennsylvania.

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/real-estate_papers/36
For more information, please contact repository@pobox.upenn.edu.
Rising Trade Costs? Agglomeration and Trade With Endogenous Transaction Costs

Abstract
While transport costs have fallen, the empirical evidence also points at rising total trade costs. In a model of industry location with endogenous transaction costs that seeks to replicate features from the machinery industry, we show how and under which conditions a decline in transport costs can lead to an increase in the total cost of trade. The subtle relationship between (endogenous) transport costs and the sensitivity of trade to distance is also explored.

Alors que les coûts de transport ont chuté, les résultats empiriques montrent que les coûts totaux de transaction sont en hausse. A l'aide d'un modèle de localisation industrielle avec coûts de transaction endogènes qui tente d'approximer les caractéristiques de l'industrie de la machinerie, on montre comment et dans quelles conditions une chute des coûts de transport peut entraîner un accroissement des les coûts totaux de transaction. On explore aussi la relation subtile entre les coûts de transport (endogènes) et la sensibilité des flux de commerce au facteur distance.

Keywords
transaction costs, trade costs, transport costs, agglomeration, vertically linked industries

Disciplines
Economics | Real Estate

Comments
At the time of publication, author Gilles Duranton was affiliated with the University of Toronto. Currently, he is a faculty member at the Real Estate Department at the University of Pennsylvania.
Rising Trade Costs? Agglomeration and Trade with
Endogenous Transaction Costs

Gilles Duranton and Michael Storper
Abstract
While transport costs have fallen, the empirical evidence also points at rising total trade costs. In a model of industry location with endogenous transaction costs, we show how and under which conditions a decline in transport costs can lead to an increase in the total cost of trade.

Key words: Transaction costs, trade costs, transport costs, agglomeration, vertically linked industries
JEL classification: D23, D24, R12

This paper was produced as part of the Centre’s Globalisation Programme. The Centre for Economic Performance is financed by the Economic and Social Research Council.

Acknowledgements
We have benefited from comments and discussions with Pol Antràs, Vernon Henderson, Maureen Kilkenny, Niko Matouschek, Frédéric Robert-Nicoud, Shin-Kun Peng, Jacques Thisse, Tony Venables, and seminar and conference participants in Marseille and at the North-American Regional Science Association meetings. We are grateful to Keith Head and Thierry Mayer for sharing their data.

Gilles Duranton is an Associate of the Centre for Economic Performance, and a Reader in the Department of Geography and Environment, London School of Economics. He is also affiliated to CEPR. Contact: g.duranton@lse.ac.uk, http://cep.lse.ac.uk/~duranton. Financial support from the Leverhulme Trust is gratefully acknowledged.

Michael Storper is Professor of Urban Geography in the Department of Geography and Environment at the London School of Economics. He is also Professor of Economic Sociology at the Institut d’Etudes Politiques de Paris and Professor of Regional and International Development, Department of Urban Planning in the School of Public Affairs at UCLA, the University of California, Los Angeles. Contact: m.storper@lse.ac.uk; michael.storper@sciences-po.fr.
1. Introduction

The cost of shipping goods from one place to another has declined relentlessly over at least the last hundred years. Glaeser and Kohlhase (2003) document that the share of US GDP in transportation industries has declined more than twofold over the 20th century. They also show that the real dollar cost per ton-mile for US railroad shipping has decreased nearly tenfold over a century. French data exhibit a similar trend: Combes and Lafourcade (2005) document a decline of nearly 40% of the cost of road freight over 1978-1998. There is thus very little doubt that unit internal transport costs have declined very significantly. The evidence on international transport costs is more difficult to assess (Anderson and van Wincoop, 2004). Hummels (1999) documents a dramatic fall in air transport costs since at least the early 1970s and a less pronounced decline in sea transport costs since 1985 (where containerisation, which improved shipping quality and efficiency, led to higher prices between 1970 and 1985).

Given this steep decline in transport costs, it is perfectly logical to expect a decline in trade costs (the sum of all the costs incurred to deliver a good to its user in which transport costs figure prominently). Some analysts even went as far as predicting the ‘death of distance’ (e.g., Cairncross, 2001). Surprisingly, the existing evidence on trade costs goes in the opposite direction. In his estimation of the trade cost parameter of a New Economic Geography (NEG) model, Hanson (2005) finds that internal US trade costs increased between 1980 and 1990.

While Hanson (2005)’s estimation relies on a wage gradient, it is also possible to estimate trade costs from gravity equations. Indeed, the coefficient on the log distance in ‘augmented’ gravity equations can be interpreted as a trade costs parameter in monopolistic competition models with iceberg trade costs (Anderson and van Wincoop, 2003). Combes, Mayer, and Thisse (2004) estimate a theory-consistent gravity equation using bilateral trade data dating back to 1870. The results on the distance coefficient are plotted in Figure 1 (a). This coefficient exhibits a clear upward trend since 1950, pointing at increasing trade costs.1 This finding is not a fluke coming from particular data-set or a particular estimation method. Disdier and Head (2004), in their meta-analysis of all the estimations of this coefficient in the literature, confirm this: trade costs appear to have increased considerably since 1950, as made clear by Figure 1 (b), which plots all existing estimates

---

1This trend cannot be explained by massive manufacturing relocations to countries with low production costs. These relocations could only generate a relative increase in trade costs with respect to production costs. Instead, in a gravity estimation, the rise of low-cost producers should have a negative effect on the distance coefficient since they tend to be further away from their trading partners.
in the literature by year of data. Put together, these findings strongly suggest that trade costs have increased over time.

Such these diverging trends for trade and transport costs make for a surprising puzzle since transport costs are expected to be of considerable importance in the determination of trade costs. If trade costs have risen despite falling transport costs, it must be that some of the other components of trade costs have increased even more. What are these other components of trade costs? The literature typically distinguishes policy barriers (tariffs and non-tariff barriers), local distribution costs (wholesale and retail), and a myriad of other costs often referred to as ‘transactions costs’, including information costs, contract enforcement costs, costs associated with the use of different currencies, legal and regulatory costs, etc (Anderson and van Wincoop, 2004).

Trade barriers have declined enormously since 1950 (Krugman, 1995). With the development of supermarkets, large retail chains, and computerised inventory management, it seems unlikely that there have been large increases in retail and wholesale costs. There is also no evidence that the composition of international trade has shifted towards more distance sensitive industries (Berthelon and Freund, 2004). By a process of elimination, all this suggests that a large increase in transaction costs is the most likely explanation for these diverging trends for trade and transport costs. In this paper, we develop a model of industry location in which transaction costs endogenously respond.

---

2 It could be added that trade between adjacent countries remain a very large part of world trade (Leamer and Storper, 2001). All this is also consistent with more indirect evidence such as studies of agglomeration in Europe, which routinely fail to uncover much systematic changes in the location patterns of industries despite European integration (Midelfart-Knarvik, Overman, Redding, and Venables, 2000; Storper, Chen, and De Paolis, 2002).
to changes in transport costs.

At this point it should be noted that pointing to broadly defined transaction costs does not provide very much guidance about the possible microeconomic foundations of any model of endogenous trade costs. To sharpen our intuition and to provide a guide in our choice of assumptions, it is worth looking briefly at one key industry where transaction costs are likely to play an important role and for which some micro-description is available: the machine industry. There is an additional advantage to considering this industry since it is often recognised as a major contributor to the manufacturing fortunes of countries (Rosenberg, 1982; Sutton, 2003).

In the machine industry, the design, development, customisation, installation, start-up, servicing, and updating of complex pieces of equipment require frequent interactions between producers and end users. Users must also be trained to use their equipment. All these operations are time-consuming and involve numerous trip back and forth between the producer and the user. Another key feature of this industry is that purchasing complex pieces of machinery is also fraught with difficulties. It is usually impossible to contract *ex-ante* on the future quality of custom-made machines.

An excellent illustration of these features can be found in Gertler’s (2004) very detailed account of the industrial decline of US Midwest and Canada. He argues that this decline is to a large extent the consequence of the unsuccessful adoption of flexible manufacturing in the late 1980s and early 1990s. Having to modernise their equipment, manufacturing firms in the US and Canada faced a scarcity of local machinery producers. Instead, German producers were perceived to offer superior products. However, buying machinery from German producers turned out to be very difficult because of all the transactions costs mentioned above. Often, US and Canadian manufacturers ended up with machines that were well below what they expected even after incurring very large costs of installation, training, and servicing. German machine producers also ended up dissatisfied because, despite their efforts, their machines had much worse performance in North America than was realised at home with German manufacturers.

Building on these insights, we propose a model of vertically linked industries. Final producers produce under decreasing returns a homogenous final good using a unit of entrepreneurial labour, production labour, and a machine. The machine has to be bought from a machine producer. The higher is the quality of the machine, the higher is the marginal productivity of labour for final goods. Machine producers can produce higher quality machines at an increasing (labour) cost
to themselves. The two important features are that machine quality is not contractible and each machine is specific to its final user. This leads to a standard hold-up problem between each final producer and her machine producer. In absence of trade in machines between countries, final producers buy their machines locally. When we allow for international trade in machines, we also assume that the cost of producing quality is higher between countries than within countries because of transport costs. On the other hand, the capabilities of domestic machine producers are endogenous and depend on the agglomeration of machine production.

Our first key result is that when machines are imported (rather than sourced locally), lower transport costs can lead to higher unit trade costs. This result arises because, as transport costs decline, exporters find it profitable to produce higher quality machines that require more interactions between producers and users and hence more transport. When transport costs are sufficiently high, this quality effect more than offsets the direct effect of lower transport costs. This result is consistent with Gertler (2004)'s description of the import by North-American manufacturers of German machinery. Cheaper transatlantic transport led to the import of ever more sophisticated machines for which the increase in transactions costs led to many trips back and forth across the Atlantic.  

3 Our second key result is that when we introduce agglomeration effects in the machine industry and allow for its location to be endogenous, we find that for high transport costs, there is a unique equilibrium for which machine production is evenly spread between the two countries. By contrast, for low transport costs, multiple equilibria arise and machine production can agglomerate in one (richer) country. This result is reminiscent of the standard NEG results and in particular those of Krugman and Venables (1995). It is nonetheless of interest for three reasons. First the framework in which it is derived is very different from the standard NEG framework.  

4 Among other things, ex-ante product differentiation plays no role in our model (and ex-post product differentiation is an outcome). Furthermore, we assume decreasing returns in final production instead of increasing returns.
typically attempts to model a generic industry, but often borrows features from many and ends up creating an implausible hybrid. Third, many interesting extensions such as the make-or-buy decision can be readily introduced in our framework.

This model is also related to the literature dealing with incomplete contracting in a spatial framework (Rotemberg and Saloner, 2000; Almazan, de Motta, and Titman, 2003; Helsley and Strange, 2004; Matouschek and Robert-Nicoud, 2005) and to the literature on international outsourcing (McLaren, 2000; Antràs, 2003; Grossman and Helpman, 2005). The key differences with the location-and-incomplete-contracting literature is that we consider contracting interactions across locations (instead of within locations) but ignore how agglomeration reduces opportunism. The main difference with the literature on international outsourcing is that we adopt a simpler framework but allow for the location of activities to be fully endogenous. Finally this model contributes to the emerging literature on the micro-foundations of trade costs. Harrigan and Venables (2004) explore the importance of time costs in the determination of trade patterns, while Leamer and Storper (2001) examine the benefits of face-to-face contact. This paper is instead concerned with how incomplete contractibility affects trade, agglomeration, and productivity.

The next section highlights the general mechanism behind our argument. Section 3 presents the basic mechanics of our model. For this purpose, we first consider an economy in autarchy with exogenously set capabilities for machine production. The following section explores our model in a trade context. The location of machine production and the capabilities of this industry are endogenously determined. Finally, the last section concludes.

2. The argument

In the next sections, we develop a general equilibrium model with detailed microeconomic foundations. Theoretical consistency forces us to specify why countries trade and detailed microeconomic foundations for transaction costs. In order to keep the model tractable, fairly specific assumptions are necessary. Our modelling strategy is to accept these requirements although they reduce the scope of the model. For concreteness, we chose our assumptions so that our model is at least a good representation of a given industry (machinery). However, our core argument is fairly general and relies on a more limited set of assumptions. In this section we present a very simple sketch of how it works in a very simple partial equilibrium set-up.
Consider a buyer in a country who imports one unit of a good from another country. This good is characterised by its endogenously determined quality level, $K$. The inverse demand of the buyer for a good of quality $K$ is:

$$P(K) = K^{1-\epsilon}, \quad 0 < \epsilon < 1.$$ (A)

The quality of the good is linear in the amount of labour used to produce it in the other country. When a good is exported, a fraction $\tau$ of the labour employed by the producer is lost in transport. This assumption could lead to different interpretations. The literal reading is that part of the quality ‘melts’ during the shipping. This is of course the famous ‘iceberg’ assumption popularised by the NEG. This may be applicable for some agricultural and other perishable goods but is probably less appropriate for many other goods (although it should be remembered that insurance costs are usually ad-valorem). More subtle interpretations are nonetheless possible (especially for durable goods). First, the good may need to be adjusted to its end user. Second, the instructions to use it need to be produced, read and even possibly discussed between the producer and the end user. Third, heavy maintenance may be needed. The consequence of all that will be that the seller needs to visit the buyer and that the number of visits increases with the quality of the good. Following Gertler’s (2004) description, this is the interpretation we use below in the case of custom-made machinery. Finally, it could also be argued that ‘quality’ is somewhat country specific so that producing the same level of perceived variety for another country is in fact more costly than it first appears (Gertler, 2004, also insists on this). Regardless of the interpretation chosen, this assumption implies that the imported quality is given by:

$$K = (1 - \tau)L,$$ (B)

where $L$ is the quantity of labour. The profit of the machine producer is $\pi = P(K) - wL$ where $w$ is the wage in the exporting country. Inserting equations (A) and (B) into the profit of the exporter implies: $\pi = [(1 - \tau)L]^{1-\epsilon} - wL$. Profit maximisation with respect to the level of quality by the exporter then implies

$$L = \left[\frac{(1 - \epsilon)(1 - \tau)^{1-\epsilon}}{w}\right]^{1/\epsilon}.$$ (C)

The unit trade cost is the labour time spent in transport $\tau L$ valued at the wage in the exporting country, $w$. Using the above, we get after simplification:

$$TC(\tau) = \tau(1 - \tau)^{1/\epsilon}Q,$$ (D)
where $Q$ is a constellation of constants. This function is obviously bell-shaped. Starting from $\tau = 1$, lower transport costs imply first higher trade costs before leading to lower trade costs. The general intuition is that as transport costs decline, exporters find it optimal to produce higher quality goods, which in turn incur higher trade costs. For high transport costs, the indirect effect of higher trade costs caused by higher quality dominates the direct effect of lower transport costs, whereas for low transport costs, the direct effect dominates.

The argument given here relies on the fact the exporters have some market power over their seller. However, it would be equivalent to assume a competitive market with decreasing returns. The only crucial elements are that transport costs should be ad-valorem and that quality should increase sufficiently when transport costs decline.

This argument illustrates in a fairly general manner the thrust of the model exposed below but it is incomplete and leaves many questions unanswered: Why is the good imported rather than produced locally? What are the underpinnings of the demand function? What is the source of the seller’s market power? Why are wages fixed and how do endogenous wages affect the outcome?, etc. The rest of the paper will propose a complete model with a consistent answer to these questions in a full-fledged general equilibrium framework.

3. Autarchy

A. The model

Consider a two-tier production process. There is a single homogenous final good that is used as numéraire. Each competitive final producer requires one unit of entrepreneurial labour, one machine, and uses some amount of workers’ time. Each final producer operates under decreasing returns to scale. After sinking her labour endowment, each entrepreneur acquires her machine from a machine producer. Workers are hired after the machines are delivered. The profit function of final producer $i$ is given by:

$$\pi_i = k_{ij}^a l_i^\beta - w l_i - P_{ij},$$

where $\alpha + \beta < 1$ (and $(\alpha, \beta) > (0,0)$). Labour, $l_i$, is hired at the competitive wage $w$. Finally, $k_{ij}$ denotes the quality of the machine bought from producer $j$ at price $P_{ij}$. After inserting the
first-order condition for profit maximisation with respect to employment in equation (1), we find:

\[ \pi_i = (1 - \beta) \left( \frac{\beta}{w} \right)^{\frac{\beta}{1 - \beta}} k_{ij}^{1 - \beta} - P_{ij} \equiv Z_i - P_{ij}, \tag{2} \]

In turn, machines are produced by machine producers. Each machine producer requires one unit of entrepreneurial labour and uses some amount of workers’ time. As in final production, entrepreneurs in the machine sector must sink their labour endowment before being able to operate. For reasons that will become clear below, this sunk labour is best thought of as a training period. The production of one machine of quality \( k_{ij} \) by machine producer \( j \) sold to final producer \( i \) requires a sunk investment of \( l_{ij} \) units of labour such that \( k_{ij} = Al_{ij} \). The productivity shifter \( A \), which we refer to as the capabilities of machine producers, is temporarily taken to be exogenous. Below, it will become an endogenous variable. The profit of entrepreneur \( j \) is the sum of her operating profit across all the final producers she sells a machine to

\[ \pi_j = \sum_{i \in C(j)} \pi_j(i) = \sum_{i \in C(j)} P_{ij} - \frac{w}{A} k_{ij}, \tag{3} \]

where \( C(j) \) denotes the set of \( j \)'s buyers.

The investment made by the entrepreneur in the machine sector to produce a machine is not contractible ex-ante.\(^5\) Because of customisation, each machine is specific to its prospective buyer and cannot be used by any other final producer. This specific and non-contractible investment opens the door to ex-post renegotiation, which in turns precludes the existence of any market where machine producers can credibly compete on prices ex-ante. Instead, each final producer is randomly matched to a machine producer. After being matched with final producer \( i \), machine producer \( j \) makes her non-contractible investment in quality, \( k_{ij} \). After this, the two parties bargain on the price \( P_{ij} \).

Following usual practice in the incomplete contract literature (see Grossman and Hart, 1986; Hart, 1995, and their followers), we assume that the hold-up problem is resolved co-operatively. The machine is thus delivered to its buyer and the surplus is split following a Nash-bargaining solution. The ex-post surplus made by the machine producer is \( P_{ij} \).\(^6\) The surplus made by the final

\(^5\)For consistency, we also assume that profits and employment in the two sectors cannot be contracted upon ex-ante. Otherwise it might be possible to contract on final output or some inputs and implement an efficient investment. Note that this paper takes contract incompleteness as a given and does not attempt to add anything to the debate about its micro-foundations (Hart and Moore, 1999). Final production and machine production require their entrepreneurs to sink their labour endowment so that vertical integration is simply not an option. This assumption is relaxed below.

\(^6\)Since the investment made by the machine producer is both specific to final producer \( i \) and sunk, it must be ignored at this stage.
producer is given by (2). If \( a \) is the ‘bargaining power’ of final producers, the Nash-bargaining solution is such that it maximises the following expression:

\[
\max_{P_i} (Z_i - P_i) a P_i^{1-a}.
\]

(4)

This directly yields \( P_i = (1 - a) Z_i \) so that the profit of final producer \( i \) can be written as \( \pi_i = a Z_i \) and the operating profit made by the machine producer with her machine is \( \pi_j(i) = (1 - a) Z_i - wk_{i,j}/A \). Expecting this \textit{ex-post} level of operating profit with final producer \( i \), the \textit{ex-ante} profit-maximising investment for \( j \) is:

\[
k_{i,j} = \left[ \frac{(1 - a) a A} {a(1 - \beta)} \right]^{\frac{1-\alpha-\beta}{\alpha}}.
\]

(5)

Note that the quality produced by machine producers increases with the efficiency shifter, \( A \), while it decreases with the cost of labour, \( w \). Quality also declines with the bargaining power of final producers, \( a \).

Free-entry of entrepreneurs in final production implies that in equilibrium \( \pi_i = w \). Using (2), (5), and \( \pi_i = a Z_i \), we find after simplification:

\[
k_{i,j} = \frac{(1 - a) a A}{a(1 - \beta)}
\]

(6)

and

\[
w = [a(1 - \beta)]^{1-\alpha-\beta} [1 - \beta] [\frac{(1 - a) a A}{a(1 - \beta)}].
\]

(7)

Using \( l_i \) as given by the first-order condition of (1) with respect to employment, the labour requirement for quality \( l_{i,j} = k_{i,j}/A \), and equation (6), employment in each final producer and in the production of each machine can be derived as \( l_i = \frac{\beta}{a(1 - \beta)} \) and \( l_{i,j} = \frac{(1 - a) a A}{a(1 - \beta)} \).

Finally, if \( n \) denotes the share of final production entrepreneurs in employment and \( m \) is the share of machine entrepreneurs, the random matching of each final producer with a machine producer implies that the expected profit of any entrepreneur in the machine sector is given by \( E[\pi_j] = E[\text{Card}(C(j))] \pi_j(i) = \frac{n}{m} \pi_j(i) \). Assuming risk-neutrality, free-entry in machine production then yields \( E[\pi_j] = \frac{n}{m} [(1 - a) Z_i - wk_{i,j}/A] = w \). After normalising total employment to unity, labour market clearing writes \( n + nl_i + m + nl_{i,j} = 1 \). Together with the results above, these two expressions imply: \( n = a(1 - \beta) \) and \( m = (1 - a)(1 - \alpha - \beta) \).

Before extending this framework a few comments are in order. First note that mean income \( w \) in equation (7) is maximised when the bargaining power of final producers, i.e., their share of
surplus is such that \( a = \frac{1 - \alpha - \beta}{1 - \beta} \). This value of \( a \) (between 0 and 1) reflects a trade-off between the rents of entrepreneurs in final production and those of entrepreneurs in the machine sector. Machine producers invest efficiently when they can capture all the surplus, i.e., when \( a = 0 \). On the other hand, low values of \( a \) make entry in final production unattractive since a low \( a \) implies a low share of surplus for final producers. With few entrepreneurs in final production, each will have to employ a large number of workers, and this in turn will imply a very low marginal product of labour. The other key parameter in (7) is the capability of machine producers, \( A \), which maps directly into wages. This is because higher capabilities to produce machines lead to better machines in equilibrium and thus a higher marginal product of labour in final production and higher wages for workers and entrepreneurs in both sectors.

B. Welfare analysis

Because machine producers cannot capture the entire marginal returns of their investment without preventing entrepreneurs in final production from entering efficiently, the equilibrium outcome described above is inefficient. The efficient outcome would involve the ex-ante co-operative maximisation of the joint-profit by the two entrepreneurs: \( \pi = k^a \beta l^\alpha - w_l - w_k / A \). The un-constrained first-best also requires any entrepreneur in the machine sector to produce as many machines as possible. To avoid any degeneracy, it is useful to assume a limited span of control for entrepreneurs in the machine sector so that an entrepreneur cannot produce more than \( g \) machines. For simplicity, assume \( g \) to be large enough for this limit on the span of control not to be binding at the free-market equilibrium.\(^7\) The first-best wage can then easily be derived as

\[
 w^{FB} = \left[ \frac{g(1 - \alpha - \beta)}{g + 1} \right]^{1 - \alpha - \beta} A^\alpha. \tag{8}
\]

Simple algebra shows that this wage is higher than the equilibrium wage derived in equation (7).

C. Make or buy

So far we have considered that machine production had to be out-sourced to specialist machine producers. However, it is worth considering the case where final producers can be integrated

\(^7\)From previous algebra, it is sufficient to have \( g > (1 - \beta)/\alpha \).
and produce their own machine, albeit less efficiently than ‘specialist’ machine producers. The assumption that non-integrated machine producers have higher capabilities is justified by the fact that the labour endowments of entrepreneurs in the machine sector have been sunk in specialist training. Put differently, there are benefits to specialisation. Alternatively, superior efficiency of specialists could be due to higher bureaucratic costs associated with vertical integration. Formally, we assume that final producers can also produce their own machine according to \( k_i = A_f l_{i,i} \) with \( A_f < A \).

The profit of integrated final producers (i.e., those who decide to make their own machine) is then:

\[
\pi_i = k_i^\alpha l_i^\beta - w l_i - \frac{w}{A_f} k_i. \tag{9}
\]

The two first-order conditions for profit maximisation with respect to \( k_i \) and \( l_i \) imply that equilibrium profit is given by

\[
\pi_i = (1 - \alpha - \beta) A_f^\alpha \frac{\alpha^\beta}{\beta^\alpha} \frac{1}{w^{\alpha/(\alpha+\beta)}}. \tag{10}
\]

When all final producers are integrated and produce their own machine, free-entry (i.e., \( \pi_i = w \)) implies that the equilibrium wage is:

\[
w = (1 - \alpha - \beta)^{1 - \alpha - \beta} A_f^\alpha. \tag{11}
\]

Using equations (7) and (11), it is easy to verify that if

\[
A_f > \left[ \frac{a(1 - \beta)}{1 - \alpha - \beta} \right]^{1 - \alpha - \beta} (1 - a) A,
\]

the unique equilibrium involves only integrated producers whereas in the opposite case all machines are produced by specialised machine producers. In the case where the two sides of equation (12) are equal, there is a continuum of equilibrium with machines being produced by both integrated final producers and machine producers. The stability of these equilibria can be readily checked. When (12) is satisfied, no final producer finds it profitable to out-source the production of her machine. Conversely, when the inequality is reversed, no final producer finds it profitable to produce her own machine.\(^9\) Equation (12) highlights a trade-off between integration, which saves

---

8Note that we only consider the case where final producers can integrate ‘upstream’. Allowing machine producers to integrate ‘downstream’ by engaging in final production with a lower efficiency than specialised final producers, albeit straightforward, would only lead to a similar trade-off within a more complicated choice structure. We ignore this extension here.

9Note that when the bargaining power is such that it maximises wages (\( a = \frac{1 - \alpha - \beta}{1 - \beta} \)), condition (12) boils down to \( A_f > \frac{1}{1 - \beta} A \).
on set-up and transactions costs, and out-sourcing, which allows access to specialist producers with greater capabilities.

Despite very significant modelling differences, these results remind us of Grossman and Helpman (2002). Grossman and Helpman (2002) consider differentiated final producers operating under increasing returns. In their approach, there is an intermediate input which can be either of high or low quality. By contrast, in the framework proposed here, the final good is homogeneous and produced under decreasing returns, while there is a continuum of possible qualities for the machines. However, like Grossman and Helpman (2002), we find that the equilibrium market structure is characterised by pervasive out-sourcing when the difference in capabilities between specialised machine producers and integrated final producers is large. When this difference is small, the equilibrium implies instead pervasive integration. In our simple benchmark, as in that of Grossman and Helpman (2002), the equilibrium is unique and integrated and non-integrated final producers do not co-exist except in knife edge cases.

In relation to the rest of the literature, note that there is no role for property rights in our model since there are no productive assets that can be traded ex-ante. This contrasts with the theory of firm developed by Grossman and Hart (1986). In our model, as in Grossman and Helpman (2002), the difference between integrated firms and out-sourcing depends on the identity of the investor — the machine producer or the final producer — not on who holds property rights. This is illustrated by equation (4) where the reserve option of machine producers is always zero. This differs from Grossman and Hart (1986), where property rights play a role and affect the ex-ante incentives to invest (rather than the identity of the investor). It would be however feasible to extend our model and allow property rights to play a role.10

10We would need to assume that final producers also make an ex-ante investment (e.g., they hire their labour before getting a machine). We would also need to assume that machine production requires a physical asset (rather than an entrepreneur sinking her labour endowment). Finally, machines should have some value to other potential users rather than be entirely specific to their buyer. Under these assumptions, it is quite clear that the reserve option of machine producers and final producers would depend on the allocation of property rights. When machine producers own the asset, they can walk out of the bargaining with a machine that has some resale value whereas they walk out with nothing in absence of property right. Conversely, final producers are able to retain the machine when they formally own the asset for machine production, whereas they end up with nothing in absence of property right. The optimal allocation of property rights is then non-trivial. When dis-integration is the optimal structure, results would be very similar to those below for the open economy. Another possible extension would be to consider different degrees of contract incompleteness as in Grossman and Helpman (2005).
4. International trade in machines

A. Transport costs and endogenous determination of $A$

Consider two countries that are similar in every respect. Assume that when a machine is produced in country 1 for a final producer in country 2, the labour employed at machine production must spend a share $\tau$ of their time travelling back and forth between the two countries. This travel time is not otherwise productive. This corresponds to the evidence presented in Gertler (2004) where the ex-ante customisation and the ex-post adjustment of machines to obtain satisfactory performance involves extensive travelling between the producer and the buyer. Consequently when machines are exported, the quality production function for machinery is $k_{i(2),j(1)} = A_1(1 - \tau)l_{i(2),j(1)}$ where $A_1$ is the capability of machine producers in 1 while subscript $i(2)$ denotes final producer $i$ located in country 2 and subscript $j(1)$ denotes machine producer $j$ located in country 1. Exported machines are paid in final goods, which we assume to be perfectly tradable. Figure 2 represents the complete input-output structure.

Turning to the determination of the capabilities of machine producers, $A_1$ and $A_2$, note that

\[11\] Note that this is equivalent to the standard assumption of iceberg transport costs in NEG.
the empirical literature argues that both the number of firms and total employment in the sector boost productivity at a given location (Rosenthal and Strange, 2004). For simplicity and given that total employment in each country is normalised to unity, we consider that a country’s share of employment in the machinery sector has a positive effect on its capabilities, $A$. If we denote the share of employment in machinery in a country (workers and entrepreneurs) by $\lambda$, our assumption implies $A_1 = A(\lambda_1)$ and $A_2 = A(\lambda_2)$ with $A' > 0$. For consistency with the material above, we normalise $A((1 - a)(1 - \beta)) = A$ where $(1 - a)(1 - \beta)$ is the share of labour employed in machinery at the autarchy equilibrium.

These agglomeration economies can accommodate a variety of microeconomic foundations. We do not develop them here since their explicit modelling would only add a layer of complication without any further insight.\textsuperscript{12} We refer instead to Duranton and Puga (2004) for a detailed exposition of the mechanics of those microeconomic foundations.

**B. Symmetric equilibrium with no trade**

From the analysis above, it is easy to see that for any $\tau > 0$, there is always a stable symmetric equilibrium. This equilibrium replicates the autarchy equilibrium in both countries. The local stability of this equilibrium is also straightforward since a small positive shock in machinery employment in one country would lead to only a marginal increase in local productivity which should be pitched against a discrete cost of exporting. The conditions under which this equilibrium is unique are explored below.

**C. Asymmetric equilibrium with trade**

We now explore the asymmetric configuration where all machinery production takes place in country 1, which is also active in final production.\textsuperscript{13} As a result of agglomeration economies, machine producers in country 1 have greater capabilities than machine producers in country 2: $A_1 = A(\lambda_1) > A_2 = A(0)$.

\textsuperscript{12}Except for one specific case that we discuss in the conclusion.

\textsuperscript{13}Formally, the condition for this country not to be fully specialised will turn out to be $(1 - a)(1 - \beta) < 0.5$. Given that $\beta$ is the share of labour in final production for which 0.5 is possibly a lower bound, this condition is very unlikely to be binding empirically.

14
Equilibrium wage and machine quality in country 1 can be derived in exactly the same way as in equations (1)-(7) to obtain:

$$k_{i(1),j(1)} = \frac{(1-a)aA_1}{a(1-\beta)} \quad (13)$$

and

$$w_1 = [a(1-\beta)]^{1-\alpha-\beta} \beta^\beta [(1-a)aA_1]^\alpha. \quad (14)$$

Because of constant returns to scale in aggregate production, these results are similar to those in closed economy for this country, except of course for the higher productivity shifter in machine production. Nonetheless, the shares of employment in the different occupations will differ.

Turning to country 2, profit maximisation by final producers implies:

$$\pi_{i(2)} = (1-\beta) \left( \frac{\beta}{w_2} \right) \frac{\beta}{w_1} k_{i(2),j(1)} - P_{i(2),j(1)} \equiv Z_{i(2)} - P_{i(2),j(1)}. \quad (15)$$

Again, ex-post bargaining between final producers and their machine providers leads to $$P_{i(2),j(1)} = aZ_{i(2)}$$. With the quality production function for (exported) machinery being $$k_{i(2),j(1)} = A_1(1-\tau)l_{i(2),j(1)}$$, profit maximisation by exporting machine producers in country 1 implies:

$$k_{i(2),j(1)} = \left[ \frac{(1-a)aA_1(1-\tau)}{w_1} \right]^{\frac{1-\beta}{1+\tau-\beta}} \left( \frac{\beta}{w_2} \right)^{\frac{\beta}{1+\tau-\beta}}. \quad (16)$$

Free-entry in final production in country 2 then implies that labour should be indifferent between the different occupations: $$\pi_{i(2)} = w_2$$. After simplification we obtain:

$$k_{i(2),j(1)} = \frac{(1-a)a(1-\tau)^{1-\alpha} A_1}{a(1-\beta)} = (1-\tau)^{1-\alpha} k_{i(1),j(1)} \quad (17)$$

and

$$w_2 = (1-\tau)^{\frac{1}{1+\alpha}} [a(1-\beta)]^{1-\alpha-\beta} \beta^\beta [(1-a)aA_1]^\alpha = (1-\tau)^{\frac{1}{1+\alpha}} w_1. \quad (18)$$

Machine producers must customise the machines for their customers, install them and train their workers to use them. When machines are exported, all these operations involve a large amount of travelling back and forth between the two countries. Consequently quality is more costly to produce for exportation and thus exported machines are of lower quality than those sold at home. In turn, this lower quality has a negative effect on the marginal productivity of the labour employed in final production in the other country. This also implies that lower transport costs lead to better quality machines for exports and higher wages in the importing country.\[14\]

\[14\]Note that this contrasts with the Alchian and Allen (1964) result according to which it is the ‘good apples’ that are shipped away while the ‘bad apples’ are sold locally. The key difference between these two results is that Alchian and Allen (1964) consider an additive transport cost independent from the value of the goods being shipped. We consider instead an ad-valorem transport cost. Furthermore, in Alchian and Allen (1964), qualities are exogenously determined, while they are optimally chosen in our model.
These effects are consistent with the description given by Gertler (2004) of the failure of Canadian and American manufacturers to adopt efficiently the latest generation of machinery mostly imported from Germany. According to Gertler (2004), there was a very large cost of distance making it very costly for machine producers to fine-tune properly their equipment after delivery. In case of a problem, the machines took very long to be serviced because of the distance, etc.\footnote{Gertler (2004) actually goes further and argues that transaction costs are not driven solely by $\tau$, a pure transport cost parameter. Instead, cultural and language differences made the co-ordination of machine producers with their North-American customers much more difficult than it was with their local customers (which is consistent with the robustness of the common language dummy in gravity equations). Gertler (2004) also insists on institutional differences. German machines were designed to be used by workers who maintain them lavishly and make a heavy personal investment to understand the details of their workings. North-American workers are typically reluctant to make such long-term investment because of their much higher job turn-over.}

After replacement and using the first-order conditions for profit maximisation in final production, employment in final production is $l_1 = l_2 = \frac{\beta}{a(1-\beta)}$. Using (17), (6) and the production function for machines, simple algebra yields $l_{i(2),j(1)} = (1-\tau)^{\frac{\alpha}{\alpha+1}} l_{i(1),j(1)}$ and $l_{i(1),j(1)} = \frac{(1-a)\alpha}{a(1-\beta)}$. Freeproption in machine production in country 1 implies that the expected income of entrepreneurs in this sector is equal to the wage: $\frac{n_1}{m_1} \pi_{j(1)}(i(1)) + \frac{n_2}{m_2} \pi_{j(1)}(i(2)) = w_1$. Together with the labour market clearing equations in both countries, $n_1 + n_1 l_{i(1)} + m_1 + n_1 l_{i(1),j(1)} + n_2 l_{i(2),j(1)}$ and $n_2 + n_2 l_{i(2)} = 1$, we can determine total employment in machine production in country 1:

$$\lambda_1 \equiv m_1 + n_1 l_{i(1),j(1)} + n_2 l_{i(2),j(1)} = (1-a)(1-\beta) \left[ 1 + (1-\tau)^{\frac{\alpha}{\alpha+1}} \right]$$

The unit trade cost for each exported machine can be computed as the time spent in transport by labour employed in the production of a machine, $\tau l_{i(2),j(1)}$ valued at the wage in the country that produces it, $w_1$. After simplification, this trade cost is equal to:

$$TC(\tau) = \tau (1-\tau)^{\frac{\alpha}{\alpha+1}} QA \left( (1-a)(1-\beta) \left[ 1 + (1-\tau)^{\frac{\alpha}{\alpha+1}} \right] \right)^{\alpha}$$

where $Q \equiv \beta \alpha [1-a](1-\beta)^{-\alpha-\beta}$ is a constant.

The product of the first two terms of $TC(\tau)$ in (20) is obviously bell-shaped and reaches a maximum for $\tau = 1-a$. With $A' > 0$, the last term of $TC(\tau)$ is decreasing with $\tau$. This ensures that $TC(\tau)$ is decreasing at least over $[1-a,1]$. If $A'((1-a)(1-\beta))$ is not infinite, it is also easy to show that $TC(\tau)$ is increasing with $\tau$ in the neighbourhood of zero. All this implies that in a region of high transport costs, their decline leads to higher trade costs, while in a region of low transport costs, their decline leads to lower trade costs.
costs, their decline leads to lower trade costs. It should also be noted that lower transport costs also imply an increase in trade (in value).\textsuperscript{16}

As in the make-or-buy case, it can be checked that this equilibrium is stable as long as no machine producer finds it profitable to enter in country 2. Simple calculations show that this holds provided $\tau$ is below a sustain point $\tau_{sust}$ such that:

$$\left(1 - \tau_{sust}\right) \frac{1 - \alpha}{\beta} A \left(1 - a\right) \left(1 - \beta\right) \left(1 + \left(1 - \tau_{sust}\right) \frac{1 - \alpha}{\beta}\right)^{1 - \alpha} = A(0).$$

Since the term on the left hand-side is decreasing in $\tau$, it is easy to see that $\tau_{sust}$ is unique. It is also easy to show that because (i) $A' > 0$ and (ii) trade costs are proportional, there is no asymmetric equilibrium with active machine producers in both countries.

Besides, note that a decline in transport costs implies higher trade costs only when transport costs are high, so that it may be the case that for this parameter region, the asymmetric equilibrium is not stable. A sufficient condition for the asymmetric equilibrium to be stable in a region of increasing trade costs is that the asymmetric equilibrium is stable when $\tau = 1 - \alpha$, value for which we know that trade costs increase with falling transport costs. This sufficient conditions reduces to: $A(0) < \alpha^{\lambda(1-a)} A \left(1 - a\right) \left(1 - \beta\right) \left(1 + \alpha^{\frac{1 - \alpha}{\beta}}\right)$.

Figure 3 illustrates these results for the following parameter values: $A(\lambda) = \lambda^{0.1}$, $a = 0.5$, $\alpha = 0.5$ and $\beta = 0.4$. Note that for these parameter values $A(0) = 0$ so that the asymmetric equilibrium is stable for any $0 \leq \tau < 1$.

Turning finally to the welfare analysis, note that the inefficiencies associated with incomplete contracting, which are discussed above, are still present in the open economy case. Our agglomeration mechanism leads to another possible source of inefficiencies. For low enough trade costs, the symmetric equilibrium is Pareto-dominated by the asymmetric equilibrium where machine production agglomerates in one country. This is because of a standard co-ordination failure where no-one wants to deviate from the symmetric equilibrium although the asymmetric equilibrium would make machine production sufficiently more efficient to offset trade costs. The asymmetric equilibrium may also be inefficient. This can occur when $A(.)$ is very concave with say $A(0) = 0$ and $A'$ very small for $\lambda > (1 - a)(1 - b)$. In this case, the asymmetric situation is always in equilibrium because no-one wants to deviate and start producing machines in the other country.

\textsuperscript{16}The volume of trade at the asymmetric equilibrium is constant by assumption since each final producer in country 2 is restricted to import only one machine. With an elastic demand for machines (caused for instance by heterogeneous capabilities of entrepreneurs in final production), it would be straightforward to extend the model and get increasing volumes of trade together with higher trade costs as consequences of lower transport costs.
On the other hand, with $A'$ very small, the gains from agglomeration are very small and can be more than offset by the trade costs. The asymmetric equilibrium cannot be Pareto-dominated however because a country always gains from the agglomeration of machine production.

**D. Discussion**

In the determination of unit trade costs, the transport cost parameter, $\tau$, intervenes three times (equation 20). First, it intervenes directly since a fraction $\tau$ of the producer’s labour is lost in transport. Through this direct effect, lower transport costs push towards lower trade costs. This direct effect is counterbalanced by two indirect effects. When transport costs decrease, a lesser fraction of labour is lost in transport so that it is optimal for the producer to produce higher quality machines for exports. This indirect effect dominates the direct effect for high trade costs while it is dominated for low trade costs. There is also a general equilibrium effect where lower transport costs and the subsequent increase in quality leads to higher rents for machine exporters. Through the labour market equilibrium (where wages are equalised across occupations), these higher rents imply higher wages in the exporting economy. Since unit transports costs are paid in the labour of the exporting country, this general equilibrium effect also contribute to higher trade costs as a consequence of lower transport costs.

It should be clear from this discussion that the detailed specification for transport costs matters. We chose an ad-valorem specification rather than a fixed amount to be paid per unit. This choice
was dictated by the greater realism of this assumption for the machinery industry. As already noted, our specification is equivalent to the iceberg transport cost assumption, thus making comparisons with the existing results easier. An additive specification for transport costs may lead to similar results provided two conditions are met: (i) lower transport costs lead to a strong enough positive quality effect and (ii) higher quality is more costly to trade. Then, it can be the case that the increase in trade costs caused by higher quality dominates the direct effect of lower transport costs.

The other assumptions made in the model are far less central to our main result. For instance, our model is rooted in an incomplete contracting framework. We believe this is an important feature of the industry we model. The rent-splitting that occurs because of incomplete contracts also introduces some very tractable proportionalities, which allow us to solve the model easily and obtain closed-form solutions. However incomplete contracting is not a necessary condition for a non-monotonic evolution of the trade costs. To see this, note that the first-best quality of exported machines is given by the maximisation of the joint surplus of the two firms and is such that $k_{i(2),j(1)} = k_{i(2),j(1)} / (1 - a)^{1 - \beta}$. This implies that first-best employment to produce an exported machine, $l_{i(2),j(1)}$, will contain a term in $1 - \tau$. Hence the trade costs, $\tau l_{i(2),j(1)} w_1$, will contain a term in $\tau$ multiplied by a term in $1 - \tau$ and some wage terms affected negatively by $\tau$. Consequently, trade costs should initially increase when transport costs decline from 1 and decrease when trade costs approach 0.

To summarise, this model provides an illustration of how the puzzle raised in the introduction can be solved. Trade costs can increase following a decline in transport costs because of a simple feedback mechanism where lower transport costs induce the export of higher quality goods, which in turn are more costly to trade. The model was purposefully rather special so that it would fit well with a particular industry for which a good empirical account was available. However, as shown in Section 2, the underlying argument is quite general and can rather easily be applied to many other industries.

Beyond offering a possible solution to an intriguing puzzle, our model also yields interesting implications in light of the existing literature on trade and location, where the NEG figures prominently. These models, albeit very stylised, provide a compelling story about the determinants of industrial agglomeration at large spatial scales. The comparative statics of those models with respect to trade costs are now heavily used for policy and predictive purpose. The usual thought
experiment is to consider the effects of a decline in trade costs (which are not distinguished from transport costs) on the location of economic activities. Yet, the strong conclusions of these models have not received much convincing empirical backing (see Head and Mayer, 2004, for a recent and very thorough review). This weak empirical performance is increasingly interpreted as a failure of these models. This paper suggests a different interpretation. It is not the basic reasoning of NEG models which are flawed but the assumption that in the real world, declining transport costs have actually brought about a powerful reduction in total trade costs across the board.

5. Conclusion

This paper proposes a model of vertically linked industries in which (i) the quality of inputs is not contractible and (ii) providing a given level of quality to suppliers becomes more costly with distance. Then lower transport costs imply that higher quality inputs are traded in equilibrium. The effect of this higher quality can be such that trade costs increase despite lower transport costs. This main result is consistent with the stylised facts presented in the introduction.

Three key extensions of this model can be envisaged in further work. First, the productivity shifter in machine production, $A$, needs to be made endogenous. As argued above, adding an extraneous source of agglomeration economies would merely lead to an extra layer of complexity. It would be more interesting to make $A$ endogenous within the current framework and use more intensively the features of our incomplete contracting framework. For instance, we could allow for a thicker local market to reduce opportunistic behaviour and increase efficiency as in Helsley and Strange (2004). The second important extension would be to introduce some heterogeneity to make the equilibria less extreme than either no trade or instead complete agglomeration of machine production in one country. A natural way to do it would be to introduce an idiosyncratic value for each possible match between machine and final producers. Despite high transport costs, one would still expect some final producer to get a much better match abroad than at home and thus import their machinery while others would source locally. At some critical level of transport costs, this ‘interior’ equilibrium may become unstable and machine production would then agglomerate in one country. These results would then be close to those of Krugman and Venables (1995).

The third extension would be to consider more than two goods and two countries. The multi-country extension is important because the stylised facts mentioned in the introduction are established for many countries for which relative locations matter, a feature not captured by our...
two-country framework. The multi-industry extension is even more important because trade costs and trade flows in different industries may be related in subtle ways through the trade balance.

References


Humels, David. 1999. Have international transportation costs declined? Processed, Purdue University.


<table>
<thead>
<tr>
<th>Paper Number</th>
<th>Authors</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>682</td>
<td>Carlo Rosa, Giovanni Verga</td>
<td>Is ECB Communication Effective?</td>
</tr>
<tr>
<td>680</td>
<td>Stephen Machin, Olivier Marie</td>
<td>Crime and Police Resources: the Street Crime Initiative</td>
</tr>
<tr>
<td>679</td>
<td>Alan Manning, Barbara Petrongolo</td>
<td>The Part-Time Pay Penalty</td>
</tr>
<tr>
<td>678</td>
<td>Andrew Clark, Fabien Postel-Vinay</td>
<td>Job Security and Job Protection</td>
</tr>
<tr>
<td>677</td>
<td>Eran Yashiv</td>
<td>Evaluating the Performance of the Search and Matching Model</td>
</tr>
<tr>
<td>676</td>
<td>Alex Bryson, Rafael Gomez, Tobias Kretschmer</td>
<td>Catching a Wave: the Adoption of Voice and High Commitment Workplace Practices in Britain: 1984-1998</td>
</tr>
<tr>
<td>675</td>
<td>Nick Bloom, Mark Schankerman, John Van Reenen</td>
<td>Identifying Technology Spillovers and Product Market Rivalry</td>
</tr>
<tr>
<td>674</td>
<td>Lorraine Dearden, Howard Reed, John Van Reenen</td>
<td>The Impact of Training on Productivity and Wages: Evidence from British Panel Data</td>
</tr>
<tr>
<td>673</td>
<td>Giulia Faggio, Stephen Nickell</td>
<td>Inactivity Among Prime Age Men in the UK</td>
</tr>
<tr>
<td>672</td>
<td>Chiara Criscuolo, Ralf Martin</td>
<td>Multinationals and US Productivity Leadership: Evidence from Great Britain</td>
</tr>
<tr>
<td>671</td>
<td>Roberto Torrini</td>
<td>Profit Share and Returns on Capital Stock in Italy: the Role of Privatisations Behind the Rise of the 1990s</td>
</tr>
</tbody>
</table>
Fiscal Discipline and the Cost of Public Debt Service: Some Estimates for OECD Countries

Can Comparative Advantage Explain the Growth of US Trade?

Is Poland the Next Spain?

Accounting for Cross-Country Income Differences

Designing Target Rules for International Monetary Policy Cooperation

Entrepreneurship: Can the Jack-of-All-Trades Attitude be Acquired?

Sinking the Blues: the Impact of Shop Closing Hours on Labor and Product Markets

Productivity Growth and Employment: Theory and Panel Estimates

Financial Globalization and Exchange Rates

Do Job Security Guarantees Work?

Unions, Performance-Related Pay and Procedural Justice: the Case of Classroom Teachers

How Special is the Special Relationship? Using the Impact of R&D Spillovers on UK Firms As a Test of Technology Sourcing