Trade, complexity and distance: an empirical investigation

Sandrine Noblet, Antoine Belgodere [1,2]

Very preliminary version - april 2016

Abstract

The aim of this paper is to link the complexity of production process and the geographical dimension of fragmentation. More specifically, we address the following questions: 1) how does complexity affect import volumes? and 2) how does complexity affect the sensitivity of trade flows to the distance? To do so, we construct an indicator of complexity from US input-output tables. This indicator is then introduced in a gravity equation (using Comtrade Data) as a regressor and as an interaction term with the distance. Our results show that: imports are reduced by complexity and that more complex production processes are related to higher distance coefficient (in absolute value).

Key words:

complexity, distance, trade, gravity equation

JEL Classification:

[1] Assistant professor, Université de Corse - CNRS UMR LISA 6240. Email: noblet@univ-corse.fr

[2] Assistant professor, Université de Corse - CNRS UMR LISA 6240. Email: belgodere@univ-corse.fr

Introduction

One of the most important features of globalization these last decades is the increasing fragmentation of production process (Krugman (1995)). This fact has been extensively documented in the literature and especially through the importance of IG flows in global trade flows (Jones and Kierzkowski (2005), Yi (2003) among others).

Another aspect of this increasing international fragmentation is the increasing complexity of production process (Krugman (1995)). In the early nineties Kremer (1993) emphasized the role of complexity related to development and labor economics issues. About fifteen years later, Hidalgo and Hausmann (2009) stress that the level of complexity was a good predictor of a country income level. Costinot (2009), Ma et al. (2012) Minondo and Requena-Silvente (2013) link complexity with trade structure. They show that richer countries have a comparative advantage in complex goods [3].

Fragmentation has also been studied with respect to its geographic extent. The main question was to determine if fragmentation is taking place on regional or in international basis. Answers to this question are mixed. According to Johnson and Noguera (2012), distance plays an important role in trade even if its magnitude depends on sectors, time, countries an partners. Their results suggest that regional dimension is greater than international dimension. Baldwin and Lopez-Gonzalez (2014) show that the regionalization of trade is particularly true for trade in intermediates rather than trade in final goods.

[3] Notice that despite both contributions explain trade specialization with complexity concept the channel is different. For Costinot (2009), it is a matter of institutional quality whereas for Minondo and Requena-Silvente (2013) it is a matter of labour productivity (See discussion in Minondo and Requena-Silvente (2013)).

However, Los et al. (2015) found evidence suggesting a shift from regional production systems to international ones.

The aim of this paper is to link the complexity issue and the geographical dimension of fragmentation. More specifically, we address the following question: how does complexity affect the volume of trade and the distance of trade. To do so, we construct an indicator of complexity from US input-output table. Indicator is then introduced in a gravity equation. Our results suggest that more complex production processes are related to higher distance coefficient (in absolute value).

The remaining of the paper is organized as follows. The next section provides a review of literature regarding 1) the complexity measures and 2) the link between fragmentation, complexity and proximity. In section (2) we discuss our empirical methodology. Then results are presented in section (3). The last section concludes.

1 The complexity issue

This section provides a review of literature on complexity issues. First, we present different measures of complexity. Then, we collect some evidence within the literature regarding the evolution of complexity and link complexity issues with fragmentation and geographical proximity.

1.1 Measuring complexity

There are several attempts in the literature to capture the level of complexity. We consider 3 categories of measure. The first one captures the level of fragmentation of goods. The second one is the Rauch' classification initially used to distinguished homogeneous goods from differentiated ones and later on applied to measure the level of complexity (Berkowitz et al. (2006)). Finally, sophistication indexes can be used to proxy the level of complexity (Hidalgo and Hausmann (2009)).

Fragmentation can be used as a proxy of the level of complexity. Within this literature we distinguish three ways to capture complexity: 1) the number of input sectors embodied in a final good (Clague (1991) from input-output data, Kremer (1993)) or the number of tasks from OES (Romalis (2004), Minondo and Requena-Silvente (2013)), 2) the dispersion of input over sectors captured by an Herfindhal from I-O data (Clague (1991), Fernandes (2015), 3) Average Propagation Lenght (APL hereafter) methodology (Dietzenbacher et al. (2005)) which consists in measuring the economic distance between sectors. This distance or length is inferred from I-O data by looking the number of steps for an exogenous change in a particular sector to propagate to another sector. These indicators provide information on the level of fragmentation inasmuch as they depend on the number of inputs or stages required to produce one final good. They are in line with the definition of complexity from Kremer (1993).

The Rauch's classification (Rauch (1999)) has also been used as a proxy of complexity (Berkowitz et al. (2006)). This classification consists in distinguishing homogeneous goods from differentiated ones. There are 3 kinds of goods:

1) goods traded on organized exchange, 2) goods with referenced price and 3) the other goods. This classification has been used to show that the level of differentiation of goods impacts the effect of distance on trade flows. Berkowitz et al. (2006) links this classification to the level of complexity of products. The first category of goods is considered as having a low level of complexity while the last one as having a high level of complexity (Berkowitz et al. (2006)).

Finally, sophistication indexes and especially the one proposed by Hidalgo and Hausmann (2009) can be used as a proxy of the level of complexity of goods. They introduce two indicators: the first one denotes the export diversification of a country, ie. the number of varieties exported for each countries and the second one measures the number of countries exporting a particular good. They show that when countries have diversified exportations, these goods are exported by a limited number of countries. This result suggests that these goods are more complex. Here again, the level of complexity is not captured but inferred from the fact that not all countries can trade complex goods. The smaller is the number of countries trading a good, the higher is the probability that this good is complex.

1.2 Fragmentation, complexity and proximity

From the latter definitions of complexity, it follows that the deeper is the level of fragmentation, the higher is the level of complexity of goods. This increasing complexity comes from: 1) the use of more specialized inputs and 2) the greater variety of inputs used in production process (Kremer (1993), Krugman (1995)).

Failing to have direct evidences of the complexification trend, we can draw some indirect evidences from the existing literature and especially from literature about the evolution of occupations. Spitz (2004) shows that an important feature of the fragmentation of production processes is the decreasing share of routine tasks performed by workers, whatever their skill level. More precisely, among these non-routine tasks, he highlights a particular increase in interactive tasks. In our view, this switch from routine to non-routine and interactive tasks might reflect the increasing complexity of production processes.

Sturgeon and Memedović (2011) showed that *customized* IG, defined as produced for few final goods, have gained importance in the top 50 traded IG flows. This fact may reflect the increasing complexity of production process coming from a higher specialisation.

Whatever the source of complexity ie. higher specialization or higher number of stages, we argue that complexity requires proximity. In previous papers, we proposed a modeling of coordination costs based on both geographical distance between upstream and downstream firms and on the number of intermediate goods entering in the production process. Here the complexity lies on the production process whereas in papers using the Rauch' classification (Berkowitz et al. (2006), Briant et al. (2014)), the complexity is embodied in the traded good. In both cases, complexity requires geographical proximity for coordination purpose or more broadly for information transfer needed between suppliers and customers.

2 Empirical strategy

Before presenting our empirical strategy to measure complexity, it seems useful to describe a first best strategy, that would be adopted in an ideal world where every desired data would be available. More precisely, in such a world, a very detailed input-output table would be available at the world level. It would describe, for each sector of use in each country, the volume of each kind of input, defined by type of intermediate and country of production. With such data, the complexity of sector of use would simply be the number of intermediate goods used by the sector. This level of complexity of sector of use would be a regressor in a gravity equation explaining trade for each type of intermediate good.

Alas, we do not live in this ideal world. Such database does not exist. More precisely, Dietzenbacher et al. (2013) provide such database but with only 15 sectors which is a very precious informations for many topics but not sufficiently disaggregated to measure complexity. Moreover, Comtrade database does not indicate the sector of use of traded goods. In the same line, OECD provides a database on IG trade flows precising the sector of use but here only 9 of them.

Here is our second best strategy. We use input-output data on US economy [4] to define the average complexity of the sector of use of a given product. More precisely, let q_{ij} be the quantity of intermediate goods of type i used by sector j. We define the *average complexity of sector of use* of good i as $acsu_i \equiv \frac{\sum_{j=1}^{n} comp_j q_{ij}}{\sum_{j=1}^{n} q_{ij}}$, where n is the number of sectors and $comp_j$ the indicator of complexity of sector j. $comp_j$ is just the number of non-null input rows of the sector in the input-output matrix (Clague (1991)).

This strategy leads to a loss of information at three levels: first, sectoral complexities are computed for the US industry, and assumed constant in every

[4] Data for 2007 are provided by BEA and available at: `http://www.bea.gov/industry/io_annual.htm`

country, whereas one could easily think that complexity can vary across countries. Second, BEA IO table only distinguishes 385 goods, which can be considered as not sufficiently desegregated to correctly apprehend complexity. Third, using the average complexity of sector of use is a clear loss compared to the ideal situation in which trade flows would be available for each sector of use.

However, this strategy being our second best, we adopt it. Table 1 presents the 15 less and 15 more complex sectors in US economy, based on the number of intermediate goods. Figure 1 presents a histogram of *acsu* accross US industrial sectors, which varies between 103 and 237.

[Table 1 about here.]

[Fig. 1 about here.]

3 Results

[Table 2 about here.]

Table 2 presents the results of main regressions. Regressions I and II are classical gravity regressions, with no account for complexity (For all settings, odd numbers are regressions without importer-specific dummies, even numbers are regressions with such dummies). They serve as benchmarks. Distance coefficients are close to $-.6$ ($-.496$ and $-.576$ respectively).

In regressions III and IV, we introduce complexity ($acsu_i$) as a regressor, with no interaction with distance. Other coefficients are barely affected by this add. It appears that complexity has an important impact on trade. An increase in complexity from first to last decile (126 to 194) would reduce trade by about a quarter *ceteris paribus* [5].

In regressions V and VI, we introduce an interaction variable between complexity and distance. In order to compare the distance coefficients to those of regressions I to IV, we replace *acsu* by its average value in the sample, which is 158 (indeed, the distance coefficient becomes the proper distance coefficient plus the coefficient of the interaction times *acsu*). They remain in the same magnitude ($-.5052$ and $-.5845$ respectively). The same calculus is done for the complexity coefficient, computed for the average log of distance (8.1 in the sample). Again, they remain in the same magnitude, although they slightly increase in absolute value. The coefficient of the interaction between (log-)distance and complexity is statistically significant (p-value reported as .000 by stata, as for every coefficient, except one of the country-specific dummies) and of the expected sign (-). Since 10% of the sectors have a *acsu* bellow 126 and 10% above 194, we check the importance of this interaction coefficient by computing the impact of a 68 increase in *acsu* on the coefficient of distance: the result is $-.163$ in both regressions. This figure is quite important, compared to the average value of distance coefficients which are close to $-.55$. The same calculus is done to measure the impact of distance on the complexity coefficient. Log of distance between trade partner is bellow 6.5 for 10% of pairs of country, 10% are above 9.4. A 2.9 increase in distance adds .0068 to the absolute value of complexity coefficients, which are .0040 and .0034 on average. Here, the impact is even more striking. Complexity coefficient is multiplied by 12 when distance increases from first to last decile.

Figure 2 illustrates those results by showing the predicted volume of trade

[5] With the coefficients of regression III, this increase in complexity would imply a multiplication of trade, *ceteris paribus*, by a factor $e^{-.0044\times(194-126)} \approx .74$. The same calculus gives .77 with the coefficient of regression IV.

between two virtual countries both having the sample average gdp (27 in log), and for 5 different values of $acsu_i$: the sample minimum, the first decile, the mean, the last decile and the sample maximum (103, 126, 158, 194, 237). Distance is presented in log from 5 to 10 ($\approx$ 150 to 22'000 km). Notice that, because of the different slopes, the traded volumes are identical for $\ln dist = 5$ and they spread when distance increases. The plotted function is:

$$\ln(imp_{i,r,p}) = -18.6 + .5\,((\ln y_r + \ln y_p)) - (.126 + .0024 \times acsu_i) \ln d_{r,p} + .0154 \times acsu_i$$

[Fig. 2 about here.]

4 Conclusion

The paper was an attempt to test the assumption that complexity of production processes requires geographical proximity between trade partners. As we mentioned before, this question is implicit within international trade literature. More precisely, we have tried to capture complexity at the level of production process. The underlying assumption is that a sector having many intermediate goods to coordinate is willing to have its suppliers nearby. This assumption is different from the one in Rauch (1999) who assumes this is the complexity (or more precisely the level of differentiation) of the traded good itself that has an impact on trade.

Our main difficulty is data availability. Indeed, COMTRADE data do not mention the sector of use of a good. Our strategy is to proxy the level of complexity of a sector of use by the average complexity of sectors of use of a traded good, based on desegregated IO table for US economy. With all its obvious shortcomings, this strategy suggests that complexity of sector of use increases the coefficient of distance in gravity equations. An increase in complexity from first to last decile would increase coefficient of distance by roughly 1.6 (in absolute value).

These results represent only a first step in our understanding of the impact of complexity on trade flows. Our further research on the topic will address the following questions:

- How to improve the measure of complexity of sector of use with the available data?
- How to account for the complexity of traded good itself, in addition to the complexity of sector of use?
- How complexity and its impact on trade evolve over time?

References

Baldwin, R. and J. Lopez-Gonzalez (2014). Supply-chain trade: A portrait of global patterns and several testable hypotheses. *The World Economy*.

Berkowitz, D., J. Moenius, and K. Pistor (2006). Trade, law, and product complexity. *The Review of Economics and Statistics 88*(2), 363–373.

Briant, A., P.-P. Combes, and M. Lafourcade (2014). Product complexity, quality of institutions and the pro-trade effect of immigrants. *The World Economy 37*(7192), 63–85.

Clague, C. K. (1991). Factor proportions, relative efficiency and developing countries' trade. *Journal of development economics 35*, 357–380.

Costinot, A. (2009). On the origins of comparative advantage. *Journal of*

International Economics 77, 255–264.

Dietzenbacher, E., B. Los, R. Stehrer, M. Timmer, and G. De Vries (2013). The construction of world input–output tables in the wiod project. *Economic Systems Research 25*, 71–98.

Dietzenbacher, E., I. R. Luna, and N. S. Bosma (2005). Using average propagation lengths to identify production chains in the andalusian economy. *Estudios de economía aplicada 23*(2), 405–422.

Fernandes, A. (2015). Institutions and the sectoral organization of production.

Hidalgo, C. A. and R. Hausmann (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences 106*(26), 10570–10575.

Johnson, R. C. and G. Noguera (2012, June). Fragmentation and trade in value added over four decades. Working Paper 18186, National Bureau of Economic Research.

Jones, R. and H. Kierzkowski (2005). International fragmentation and the new economic geography. *North American Journal of Economics and Finance 16*(1), 1–10.

Kremer, M. (1993). The o-ring theory of economic development. *The Quarterly Journal of Economics 108*, 551–575.

Krugman, P. (1995). Growing world trade: Causes and consequences. *Brookings Papers on Economic Activity 26*, 327–377.

Los, B., M. P. Timmer, and G. J. de Vries (2015). How global are global value chains? a new approach to measure international fragmentation. *Journal of Regional Science 55*(1), 66–92.

Ma, Y., B. Qu, and Y. Zhang (2012). Complex goods' exports and institutions: Empirics at the firm level. *Review of International Economics 20*(4), 841–853.

Minondo, A. and F. Requena-Silvente (2013). Does complexity explain the structure of trade? *Canadian Journal of Economics 46*(2), 928–955.

Rauch, J. E. (1999). Networks versus markets in international trade. *Journal of International Economics 48*(1), 7–35.

Romalis, J. (2004, March). Factor proportions and the structure of commodity trade. *American Economic Review 94*(1), 67–97.

Spitz, A. (2004). Are skill requirements in the workplace rising? stylized facts and evidence on skill-biased technological change. Technical report, ZEW Discussion Paper No. 04-33.

Sturgeon, T. J. and O. Memedović (2011). *Mapping global value chains: Intermediate goods trade and structural change in the world economy*. United Nations Industrial Development Organization.

Yi, K.-M. (2003). Can vertical specialization explain the growth of world trade? *Journal of Political Economy 111*(1), 52–102.

List of Figures

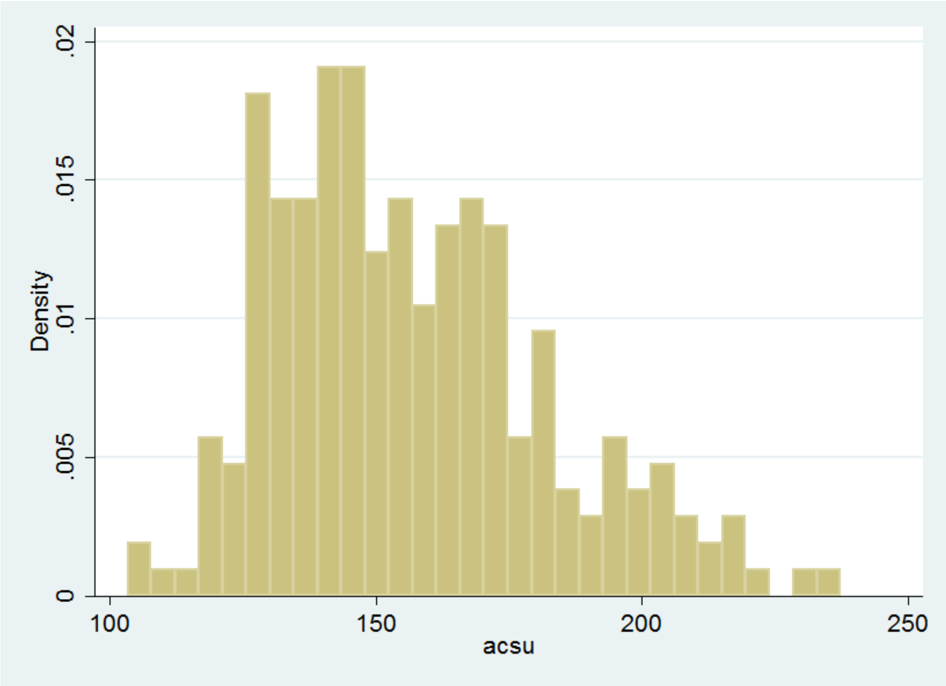

Fig. 1. Distribution of acsu between US industrial sectors

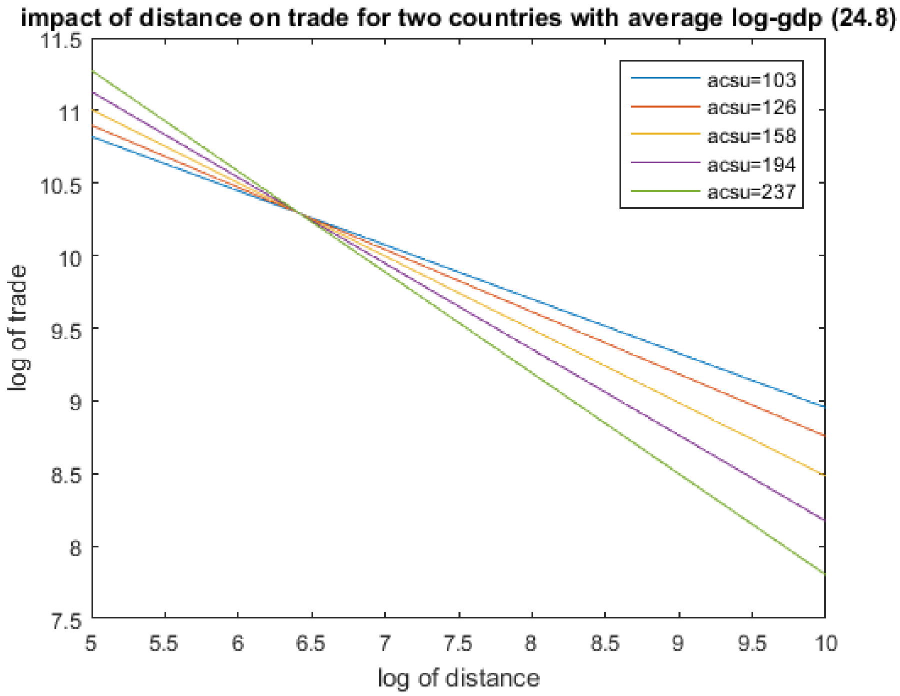

Fig. 2. Impact of distant or trade for different complexity values

List of Tables

rank	sector	complexity
1	Distilleries	68
2	Funds, trusts, and other financial vehicles	70
3	Beef cattle ranching and farming, including feedlots and dual-purpose ranching and farming	71
4	Federal electric utilities	74
5	Postal service	78
6	Forestry and logging	79
7	Primary smelting and refining of nonferrous metal (except copper and aluminum)	84
8	Flour milling and malt manufacturing	84
9	Primary smelting and refining of copper	86
10	Coffee and tea manufacturing	86
11	Carbon and graphite product manufacturing	87
12	Greenhouse, nursery, and floriculture production	87
13	Printing ink manufacturing	87
14	Lessors of nonfinancial intangible assets	89
15	Primary battery manufacturing	91
...	...	...
374	Architectural, engineering, and related services	206
375	All other food and drinking places	210
376	Wired telecommunications carriers	210
377	Full-service restaurants	212
378	Federal general government (nondefense)	213
379	Accommodation	218
380	Other amusement and recreation industries	220
381	Federal general government (defense)	222
382	Management of companies and enterprises	226
383	Hospitals	226
384	Junior colleges, colleges, universities, and professional schools	242
385	Wholesale trade	250
386	Other retail	253
387	Scientific research and development services	269
388	State and local general government	274

Table 1
15 less and 15 more complex sectors of US economy

	I	II	III	IV	V	VI
Constant	-16.83 (.0332)	-17.194 (12.96)	-15.601 (.0397)	-14.252	-18.6644 (.081)	-11.1037 (7.64)
$\ln(y_r)$	.538 (.0008)	-	0.53 (.0009)	-	.530 (.0009)	-
$\ln(y_p)$	.603 (.0009)	-	0.593 (.001)	-	.592 (.0010)	-
$\ln(dist_{rp})$	-.496 (.0015)	-.576 (.002)	-.509 (.0016)	-.592 (.0023)	-.126 (.0091)	-.2053 (.0091)
$acsu_i$			-.0042 (7e-5)	-.0037 (6e-5)	.0154 (.0005)	.0160 (.0004)
$\ln(dist_{rp})\times acsu_i$					-.0024 (6e-05)	-.0024 (5e-05)
Importer and exporter dummies	no	yes	no	yes	no	yes
pairwise dummies	yes	yes	yes	yes	yes	yes
$coefdist(\overline{acsu})$	-.496	-.576	-.509	-.592	-.5052	-.5845
$\Delta coefdist$					-.163	-.163
$coefacsu\left(\overline{\ln dist}\right)$			-.0042	-.0037	-.004	-.0034
$\Delta coefacsu$					-.0068	-.0068
$R2$	.1781	.2274	.1794	.2290	.1798	.2294

Table 2
The independent variable is the log of imports (by sector) in 2012. Standard errors are in brackets. P-values are not given because all the coefficients are significant at .1% in every regression. Number of observations is 3'587'251. $coefdist(\overline{acsu})$ is the coefficient of the distance for a sector having the average $acsu$ (158). $\Delta coefdist$ indicates the change in the coefficient of distance due to an increase in $acsu$ from the first to the last decile (126 to 194). Similarly, $coefacsu\left(\overline{\ln dist}\right)$ is the coefficient of $acsu$ for an average $lndist$ (8.1). $\Delta coefacsu$ indicates the change in the coefficient of distance due to an increase in ln_dist from the first to the last decile (6.5 to 9.4).