Public Disclosure Authorized

Public Disclosure Authorized

WPS6478

Policy Research Working Paper 6478

How Many Dimensions Do We Trade In?

Product Space Geometry and Latent Comparative Advantage

Jean-François Arvis

The World Bank
Poverty Reduction and Economic Management Network
International Trade Department
June 2013



Policy Research Working Paper 6478

Abstract

This paper proposes a new quantitative implementation of Balassa's idea that export composition and revealed comparative advantage inform the relationship between endowments in domestic factors of production and exports. It proposes that the export composition of countries is close to a low-dimensional manifold or "Product Space" within the space of export composition, which has as many dimensions as product lines. The Product Space corresponds to a few latent endowments explaining the structure of the trade matrix. The model uses non-linear techniques to identify the product space from the 2010 export matrix of 128 countries and 61 products, and to estimate the latent factors of endowments by country. It formalizes a concept of latent comparative advantage, which has practical country specific applications, relevant for "trade competitiveness" policies. Compared with classical revealed comparative advantage, the model assesses how well countries are matching their potential implied by the latent variables, and also identifies products for which the latent advantage is not yet revealed (extensive margin). The data suggests that the degree of overlap between latent and revealed advantage is a metric of "trade competitiveness."

This paper is a product of the International Trade Department, Poverty Reduction and Economic Management Network. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org. The author may be contacted at jarvis1@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Produced by the Research Support Team

How Many Dimensions Do We Trade In? Product Space Geometry and Latent Comparative Advantage

Jean-François Arvis[1] [2]

JEL C14, F11, F15, O24, O25

Keywords : Trade, Export Competitiveness, Comparative Advantage, Product Space, Non-linear
Sector Board : Economic Policy

---

[1] Jarvis1@worldbank.org, International Trade Department, the World Bank, Washington, D.C.

[2] The author would like to thank the following World Bank colleagues for useful inputs, comments and encouragements: Daniel Saslavski, Najy Benhassine, Olivier Cadot, Amir Fouad, Mona Haddad, Bernard Hoekman, Claire Hollweg, Daniel Lederman, Anasuya Raj, Cordula Rastogi, José Guillerme Reis, Benjamin Shepherd, and Daria Taglioni.

“Si nous comparons les nations entre elles, par quels signes certains constaterons-nous les progrès ou la décadence de leur prospérité ? “ (*)

Antoine-Augustin Cournot “Recherches sur les principes mathématiques de la théorie des richesses” (1838, chapter 5)

(*)When comparing nations among themselves, by which sure signs shall we notice progress or decadence of their wealth?

1 Introduction: What You Export Matters

What countries produce and eventually export depends upon a number of production inputs (labor, physical and human capital) and also domestic factors that impact productivity such as the business environment, institutions or governance (Hausmann, Hwang, & Rodrik, 2006), (Nunn, 2007), (McMillan & Rodrik, 2011). Differences in endowments of factors largely determine the composition of exports. Rich countries tend to export more sophisticated products. However, even at the same level of development, as measured for instance by GDP per capita, countries may be involved in production and exportation of more or less diversified baskets of products.

This phenomenon is generally referred to by policy makers and the institutions advising them as “export competitiveness”. Although the concept of competitiveness is intuitive to policy makers and practitioners (WEF, 2012), there is no consensus among experts on a formal definition of competitiveness embedded in economic theory ((Leamer, 1993)(J. P. Neary, 2003; J. Neary, 2006)); an ordinal concept of absolute competitiveness may not be easily reconciled with classical trade theory (comparative advantage).

Even loosely defined, competitiveness may be related to cross-sector and cross-country productivity changes that depend on factors such as cost of labor, capital, skills, logistics, innovation, quality standards, infrastructure (availability and quality), quality of institutions, corruption, etc. Understanding the relationship between countries’ endowments in those factors and the potential for growth through diversification of production and exports has been a major policy concern in rich and emerging economies, but also increasingly in poor countries willing to reduce their dependence on a relatively small basket of commodities.

Unsurprisingly, it is an area of very active policy work, especially by organizations whose role is to provide support to developing countries and advise on growth strategies (Reis & Farole, 2012). Disentangling competitiveness and understanding the linkages between factors of competiveness on the one hand, and export content and its diversification on the other hand, has been high on the practical policy research agenda in recent years. Several groups have made substantive contributions at both the conceptual and advisory levels, including multinational organizations (for example, the World Bank, International Trade Center, and World Economic Forum) and academia (including the Kennedy School of Economics and MIT). This paper is a contribution to the World Bank effort under the leadership of its International Trade Department.

The main challenge is to adequately relate trade and production outcomes by sector (i.e. export composition) with variables that capture quantitative inputs or performance metrics of these sectors at the country level. Such metrics would best capture the impact on production of a set of

domestic endowments or of relevant policies. Fortunately most of the relevant variables are now covered by established datasets of indicators based on statistics or, very often, surveys. To meet the expectations of policy makers to understand the differences of dynamics by industry, the analysis had to be disaggregated at the sector level. This provides a "meso-economic" description of how differences in factor endowment may be more conducive of certain products or exports.

The insight that trade outcomes can reveal information on sources of competitiveness or comparative advantage goes back half a century, when Balassa introduced the concept of revealed comparative advantage. This paper follows this tradition. It proposes that the structure of the global export matrix can be essentially explained by a few latent factors of endowments that determine comparative advantage. Thus the position of the problem is "inverse": country and product factor variables are deduced from the export outcome data, in contrast to the traditional econometric approach whereby export outcomes are regressed against actual endowment variables.

A nonlinear dimensionality reduction procedure (exponential-PCA or E-PCA) analyzes and produces the few latent factors. Although based on off-the-shelf mathematical tools and intensive computations, the main outputs consist in country specific information of practical value, including the latent composition of trade and latent comparative advantage ratios. Ultimately the paper proposes a scientific solution to the old problem of estimating the potential for diversification of countries based on existing data and trade theory.

The remainder of the paper is organized as follows.

The first part of the paper (section 2) introduces the key concepts and surveys the literature.

The second part of the paper develops an alternative concept of product space (section 3) rooted in neoclassical trade theory (section 4). It introduces the factor supply elasticity, which is a measure of the productivity gains brought about by changes in the supply of factors across countries. The estimation of the model requires non-linear techniques -Poisson Pseudo Maximum Likelihood (PLM)-, in contrast to previous implementations, and corresponds to a non-linear projection of the product space onto a one-dimensional manifold (section 5). Implementation to trade data and various factors of competitiveness is commented on (section 6).

Finally, the third part of the paper proposes to push the model further by applying a consistent procedure of dimensionality reduction to the product space to identify its dimensions. It combines non-linear estimation (Poisson PLM) with principal component analysis (E-PCA) (section 7). Country and product coordinates are estimated through this non-linear "projection". The significance of the principal axes in this reduction and their relationship with known factors of competitiveness are investigated (section 8), along with comparisons of latent against revealed comparative advantage by country (section 9).

The research dataset is available at www.worldbank.org/trade -> Data

2 Composition Space and Revealed Comparative Advantage: A Literature Review

Following the tradition of most empirical work, this paper analyzes export data. Indeed, export data are available across countries at a detailed sector level, more reliably than production data. Furthermore, trade data may be more relevant for assessing cross-country competitiveness since, by definition, this notion reflects competition between exporters from different countries.

However, the focus on export data may omit important features of production and trade. For instance, the ability to participate in global value chains and production sharing drives much of developing countries' diversification towards more complex production. An analysis based solely on export data, although already quite complex, may not capture fully the consequences of the cross-border nature of value chains or the importance of international backward and forward linkages (Baldwin, 2010).

Composition Space and Revealed Comparative Advantage

At an abstract level, changes—or rather relative changes—in factor endowments will correspond to trajectories of countries in the composition space of exports, where coordinates are the shares of exports by product. Composition space has many dimensions; as many as distinct product lines in the data. A change of composition in exports—resulting from a change over time (or across countries) of factors endowment—is naturally represented by a trajectory in a space with as many dimensions as products, where the coordinates are the relative export composition by country. The concept transposes to represent the position of cross-sections of countries with different endowments in factors of production. The composition coordinates are export shares:

$x_{ai} = \frac{X_{ai}}{X_a}$, where X_{ai} is the exports of product i from country a.[3]

Therefore, the composition space is the hyper-plane of co-dimension one (Fig 1) where the "movement" takes place:

$\sum_i x_{ai} = 1$ and $x_{ai} \geq 0, \forall i$

A perspective going back at least to Balassa (Balassa, 1979, 1986) is that the position of the problem can be reverted. Trajectories in composition space can "reveal" information about the importance or relative importance of factors of production for (export) competitiveness.

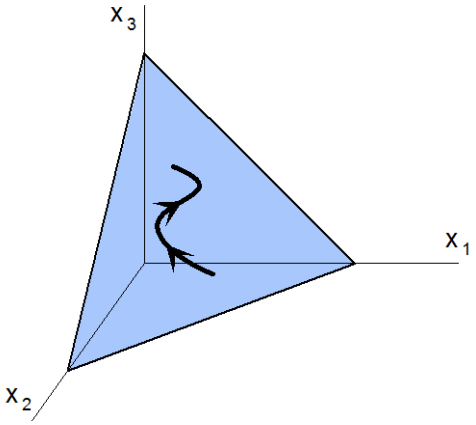
Balassa proposed the revealed comparative advantage (RCA) as a more meaningful representation of the position of countries in Composition Space than the composition vector. The RCA is the ratio of the country's export share of a particular product against the global

[3] The paper refers to a dot subscript as sums by countries or products (Einstein's convention).

$$X_a = X_{a.} = \sum_i X_{ai}, \quad X_i = X_{.i} = \sum_a X_{ai}, \quad X = X_{..} = \sum_{a,i} X_{ai}$$

export share of the same product, $RCA_{ai} = \frac{X_{ai} X_{..}}{X_{a.} X_{.i}}$

Fig 1. Composition space (with three products)



Beyond the observation of the RCAs, a precise description of the geometry of the composition space and the dynamics of products and countries would help inform the relationship between endowments and composition, and ultimately how factors of "competitiveness" influence the volume of exports across countries and across products. This identification would be of high practical relevance for understanding the potential for change of composition (synonymously diversification) and to assess how well countries are making use of their comparative advantage. However, this identification is quite complex because of the high dimensionality of the product space, the non-linear nature of the relationship and the fact that many factors are relevant to explain export compositions.

PRODY and Revealed Factor Intensity

The end of the last decade saw a fresh wave of interest in the problem, stimulated by the renewed focus on competitiveness policies. Practical methods have been proposed to understand the dependence of specific industries or export commodities on the endowments of different factors. This approach yields the "PRODY" (Hausmann & Klinger, 2006), generalized in the Revealed Factor Intensity (RFI) (Shirotori, Tumurchudur, & Cadot, 2010), available for each individual endowment factor that is measured by a macro-index available by country, say f_a. The revealed factor intensity by product (or industry) $\overline{f_i}$, is representative of the factor endowment f_a of countries that have the greatest comparative advantage of producing this product i. Thus the factor supply intensity by product $\overline{f_i}$ is a weighted average of countries' factor index values f_a, with a weight proportional to the RCA of country a in product i

$$\bar{f}_i = \frac{\sum_a \frac{X_{ai}}{X_a} f_a}{\sum_a \frac{X_{ai}}{X_a}}$$

The original PRODY referred to the level of development classically measured by GDP per capita. Other implementations extended the concept to RFI of capital or labor per unit of production. The RFI concept can essentially apply to any endowments for which a cross-country metric is available, such as the competitiveness related indicators available in the databases of the WEF or the World Bank (Doing Business, Logistics Performance, Governance).

HHR product space as a network: economic complexity

Haussman-Hidalgo and Rodrick (HHR) put forward a more comprehensive proposal (Hidalgo & Hausmann, 2009) that does not start from actual endowments but instead tries to understand the structure of exports. They introduce the concept of product space and economic complexity through a procedure akin to data analysis and dimensionality reduction. HHR views countries and products as nodes of two dual networks or graphs. A country and a product are linked when their RCA is more than one. This concept provides a simplified structure, thought of as a sort of "skeleton" of the composition space. The country and product associations can be graphically plotted. The implementation of network analysis tools, like the eigenvalue centrality (behind the Google search engine), produces an index of economic complexity by product or country (referred to as an atlas of economic complexity).

Other authors (Barigozzi, Fagiolo, & Garlaschelli, 2010) have similarly and independently implemented linear, automatic classification and data clustering analysis tools to provide a discrete, network-like description of products and countries as nodes or branches in networks or dendrograms.

However, this "skeleton" approach based on a series of simplifications does not directly provide country-specific information on non-revealed potential or on the linkages between composition and endowments.

Author's Comments

The PRODY/RFI concepts as well as product space and economic complexity have become rather popular. However, they have some limitations stemming from their intuitive and heuristic—as opposed to model-based—nature.

First is their relation to trade theory. Factor allocations are naturally defined at the country level (e.g. wages, investment per worker, policy variables), as they enter into the production functions, which typically vary among products. At the micro level, industries are not characterized by constant factor allocations worldwide but by production functions with product-dependent parameters (e.g. elasticities in a Cobb Douglas production function). The relationship between a product and a typical or optimal factor allocation, as in the RFI, is a statistical outcome "averaging" over different factor allocations across countries within the same production function. Implicitly, this association of an optimal level of a factor to a product relates to a life cycle explanation of export competitiveness. The relative export level for a given product

increases initially and then decreases beyond a point where comparative advantage has grown higher and the export of the product more intensive in the factor (Wells 1961). However, the optimal or typical factor supply for an industry is a dynamical outcome resulting from differentials in productivity across countries, products and global market demand (section 3); it is not a primary mechanism.

The RFI does not provide an explicit predictive model that could relate change in endowments to changes in RCAs or export composition. The "skeleton" approach in the product space decomposition is also based on a series of simplifications and does not provide either country-specific information on its non-revealed potential or the linkages between composition and endowments.

Both models are univariate. The RFI takes one factor at a time, even though many factors enter into a production function simultaneously. Economic complexity is also one dimensional. Neither the RFI nor the product space decomposition include a metric on how well they explain the structure of the composition matrix—the equivalent of an R^2 goodness of fit in a linear regression—making it difficult to assess quantitatively the explanatory value of the theory.

Another problem in interpreting the RFI/PRODY is that, being a weighted average of factor variable by country, it is essentially the first iteration towards a trivial fixed point, which is the simple average of the factor index value across countries. In fact for actual trade data it is easy to check that the rate of convergence is relatively fast.

3 What Export Composition Tells about Factor Endowments: Product Space and Latent Comparative Advantage

Although the original composition space has many dimensions—more than sixty if one retains a classification with two digits—it is expected that trade composition is determined by a much smaller number of factors of endowments. Hence the points representing countries in the composition space should be close to a manifold (or sub-space) with relatively few dimensions within the much larger composition space (Fig. 2).

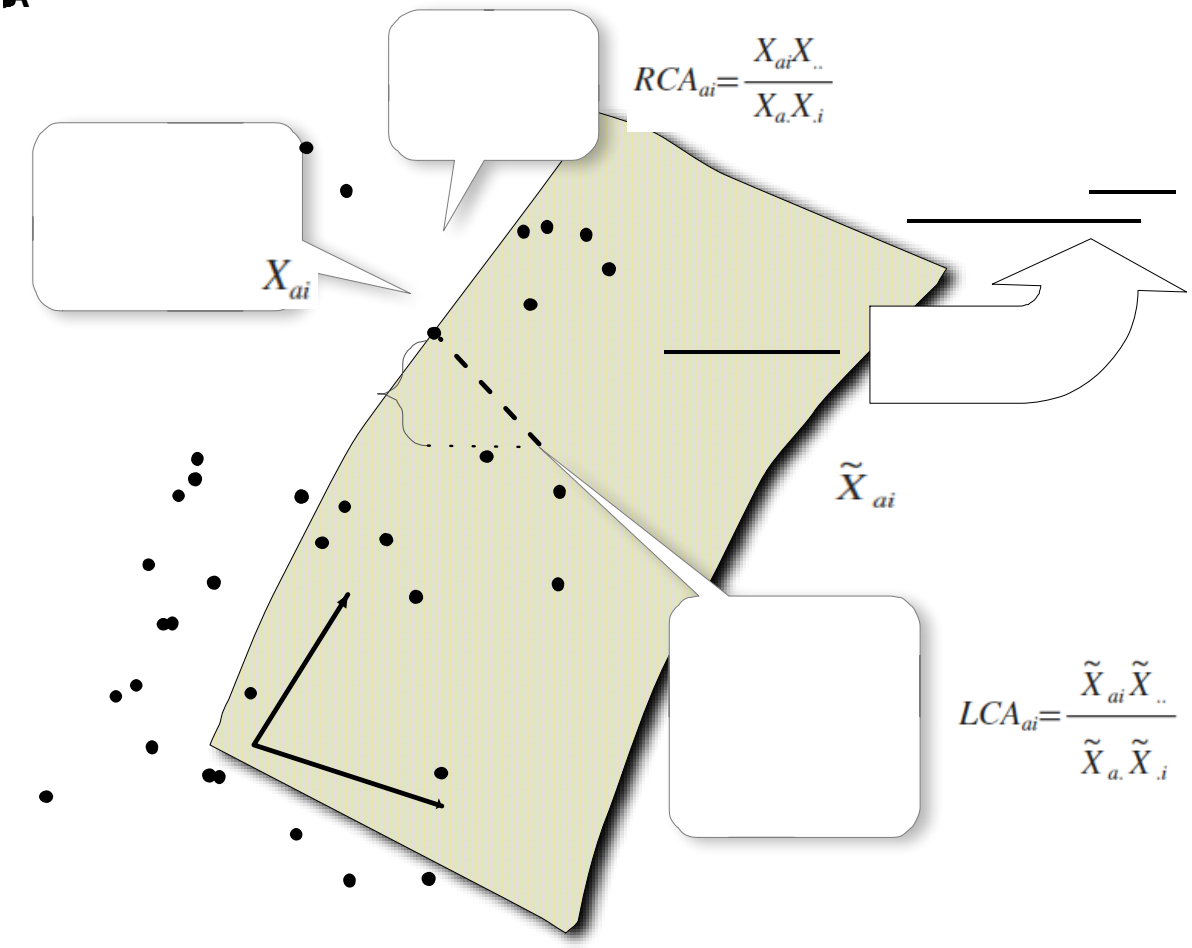
This manifold, dubbed “Product Space”,[4] is identified from the trade data through a geometric concept of dimensionality reduction. Modern econometric techniques yield a minimal set of latent factors for each country that best explain the structure of the trade matrix. The implementation of Product Space and latent factors is formalized in the forthcoming sections.

Countries have coordinates in the product space that represent latent factors and correspond to their projection from the composition space onto the product space. Products have dual coordinates in the same space. Thus the problem exhibits symmetry between countries and products. These product coordinates are export elasticities to each of the country’s factor endowment variables.

[4] The author chose to use the same vocabulary as HHR to refer to the geometrical object capturing the structure of export composition However, the parameterization of product space is very different in the two models: “skeleton” network in HHR vs. a continuous low dimensional manifold here.

The low-dimensional product space still incorporates most of the information from the entire high-dimensional composition space and the full trade matrix. Hence the projection of countries (Fig. 2) onto the product space encapsulates the information about its endowments and comparative advantage. Ultimately the paper follows an "inverse problem" approach to trade and deduces endowments from trade outcomes.

Fig 2. Product Space and Latent Comparative Advantage



The implementation of the model is made possible by combining two known tools in econometric and data analysis:

- The principal component analysis (PCA) is a powerful dimensionality reduction tool. It decomposes a cloud of data points with many dimensions into a few principal components or dimensions that capture most efficiently the variance in the data. The PCA is widely used in economics and social science.
- The second tool is the Poisson (pseudo) maximum likelihood. This choice is known to be the most natural when fitting actual flows on a network, such as trade data, against predicted values (Silva & Tenreyro, 2006).

There are several potential benefits with the proposal. The first is that the "projection" technique is itself embedded in quantitative trade theories of product differentiation. The second is that the

model has predictive value. For example, the expected change in export composition due to a change in endowments in the product space can be computed. Furthermore, the projection of countries in the low-dimensional product space corresponds to a reference latent export composition or latent comparative advantage (LCA), $\tilde{X}_{ai}$.

The comparison between latent $\tilde{X}_{ai}$ and actual X_{ai} compositions, or between revealed and latent comparative advantages, carries significant country information and may help answer practical questions. For example, how well a country realizes the potential of its position in the product space, or whether there are opportunities of production with significant latent comparative advantages not yet reflected in current comparative advantages.

4 Product Space in Trade Theory: Factor Supply Elasticities

This section formalizes the expected dependence of export composition on factors of endowments and product specific elasticities, which spans the Product Space.

Effects on production or exports from changes in domestic endowments are expected to happen through industry specific changes in productivity and eventually comparative advantage. In other words, the export from country a in product i would depend on the individual (or the set of) factor(s) f in the following way,

$$(\text{Comparative Advantage})_{ai} \sim F_i(f_a \ldots)$$

where F_i is a positive function summarizing the impact of the factor(s) f on the advantage to produce in industry i in country a. Furthermore, the export flow X_{ai} is dependent on the size of the market for product i and on the size of the exporting economy a, yielding a bi-proportional structure of exports

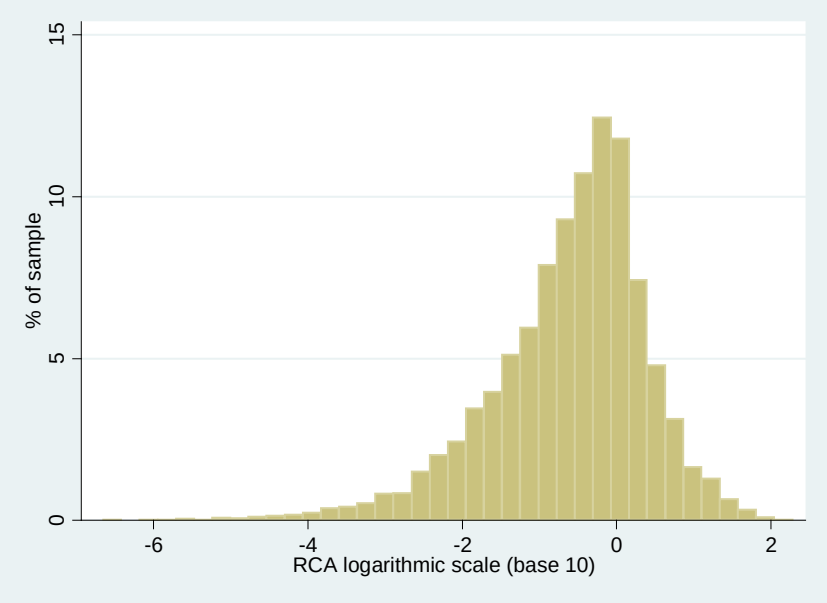
$$X_{ai} = \text{export potential}_a \times \text{market potential}_i \times (\text{Comparative Advantage})_{ai},$$

or $X_{ai} = A_a B_i F_i(f_a \ldots)$,

where the multiplicative fixed effects A_a and B_i account, respectively, for export country a's size and product i's market demand.

Revealed comparative advantages are positive numbers distributed over several orders of magnitude (Fig 3). Effects of different factors on productivity and comparative advantage are intuitively expected to be multiplicative. It is therefore natural to look for log-linear dependence where the expected impact of factor endowments is to enhance (or decrease depending on the direction of the effect) production and trade with a product dependent elasticity, hereon referred as **Factor Supply Elasticity** (FSE) α_i.

Fig 3. the distribution of RCAs (128 countries, 61 products)



Hence the exports by country a of product i take the form

$X_{ai} = A_a B_i exp(\alpha_i f_a)$ for one factor and

$X_{ai} = A_a B_i exp(\sum_k \alpha_i^k f_a^k)$, for several independent factors

This intuitive relationship is also embedded in modern quantitative trade theory exposed by many authors (J.E. Anderson, 2010; James E Anderson & van Wincoop, 2003; Armington, 1969). Essentially neoclassical quantitative models of trade flows suppose some degree of country differentiation (Armington's hypothesis) and make use of CES preferences by importers. They thus produce the generic multiplicative structure.

$$X_{ai} \sim (P_{ai})^{1-\sigma_i} \frac{X_a X_i}{X\bar{P}},$$

where P_{ai} represents the price of exports of product i by country a, σ_i is the Armington elasticity of substitution for product i, and $\bar{P}$is an average price over P_{ai}. Since the price would be inversely related to factor productivity Π_{ai} of country a's exports of product i, then the export flow X_{ai} comes as

$$X_{ai} = A_a B_i (\Pi_{ai})^{\sigma_i - 1},$$

a formula already in the original Armington paper (Armington 1969). If factor productivity comes in the classical Cobb-Douglas functional form of the indexes f of factor endowments,

$$\Pi_{ai} \sim exp(\sum_k \beta_i^k f_a^k),$$

then the export flows indeed take the expected form with factor supply elasticities related to factor productivity through $\alpha_i = (1-\sigma_i) * \beta_i$.

Elasticities of substitution are known to be relatively large (seven to eight is often retained in practical trade empirical research). Hence even relatively modest elasticities of firm productivity can lead to much larger effects in the trade matrix, as measured by the factor supply elasticities.

When the elasticities by products and factors are known, A and B are determined up to a factor by the market clearance requirement in row and columns,

$$X_a = \sum_i A_a B_i exp(\sum_k \alpha_i^k f_a^k)$$

and $X_i = \sum_a A_a B_i exp(\sum_k \alpha_i^k f_a^k)$.

Although the concept is straightforward, and the theoretical foundation known for a long time, the actual econometric implementation is not entirely trivial, given, for instance, the presence of zeros in the matrix and the inclusion of fixed effects. The problem is formally similar to that of the fixed effect spatial interaction gravity equation, where the bilateral interaction coefficient - or trade costs - takes the place of the productivity coefficient in the current problem (Anderson 2002). Recent advances in implementing the econometrics of fixed effect models of spatial gravity models can be transposed to the empirical analysis of the export composition space.

Properties of Factor Supply Elasticities

Annex 4 includes the proof of two properties relevant to the linkage between RCA and FSE.

1. Change in factor supply: an improvement in factor supply by δf_a for country a improves the RCA only in the export products i where the factor elasticity α_i is above the country weighted average elasticity $\bar{\alpha}_a$. Indeed, an increase in endowment for a factor should increase more the exports of products using this factor more intensively.
2. Approximate linear relationship between Revealed Factor Intensity (RFI) and FSE: for small values of elasticities α_i,

   $$\bar{f}_i = \text{(simple) average}(f_a) + \alpha_i \text{variance}(f_a).$$

   The RFI is thus not only dependent on the elasticity but also on the distribution of factors among countries. This relationship is also intuitive. Both the elasticity and the RFI are trying to measure the same phenomenon, namely how much the export of a given product are dependent on a given factor of endowment.

Invariances

The export matrix is invariant to a series of transformations of country and product market coefficients, as well as factors and elasticities. Again, consider the case of one endowment: $X_{ai} = A_a B_i exp(\alpha_i f_a)$. The following invariances apply (independently for each endowment):

1. Rescaling of As by the same coefficient and its inverse applied to the Bs.

$A_a \to \lambda A_a$ and $B_i \to \frac{B_i}{\lambda}$

2. Constant shift of all the factors f_a by f and corresponding change for the Bs

$f_a \to f_a + f$ and $B_i \to B_i exp(-f\alpha_i)$

3. Constant shift of all the elasticities α_i by α and corresponding change for the As

$\alpha_i \to \alpha_i + \alpha$ and $A_a \to A_a exp(-f_a \alpha)$

Given these invariances, what is the meaning of the country and product coefficients A and B? How do A and B differ from total export and import markets? Indeed, in a flat world without specialization, the simple formula

$$X_{ai} \sim A_a B_i$$

would hold, and A (respectively B) would be proportional to total exports Xa (respectively Xi). In general the proportionality does not hold and A and B are adjusted from exports and imports by a form of the "multilateral resistance" coefficient,

$$A_a = \frac{1}{\sum_i B_i exp(\alpha_i f_a)} X_a.$$

A Toy Model: Dynamic Life Cycle Effects and the Connection between Factor Supply Elasticities andRrevealed Factor Intensity

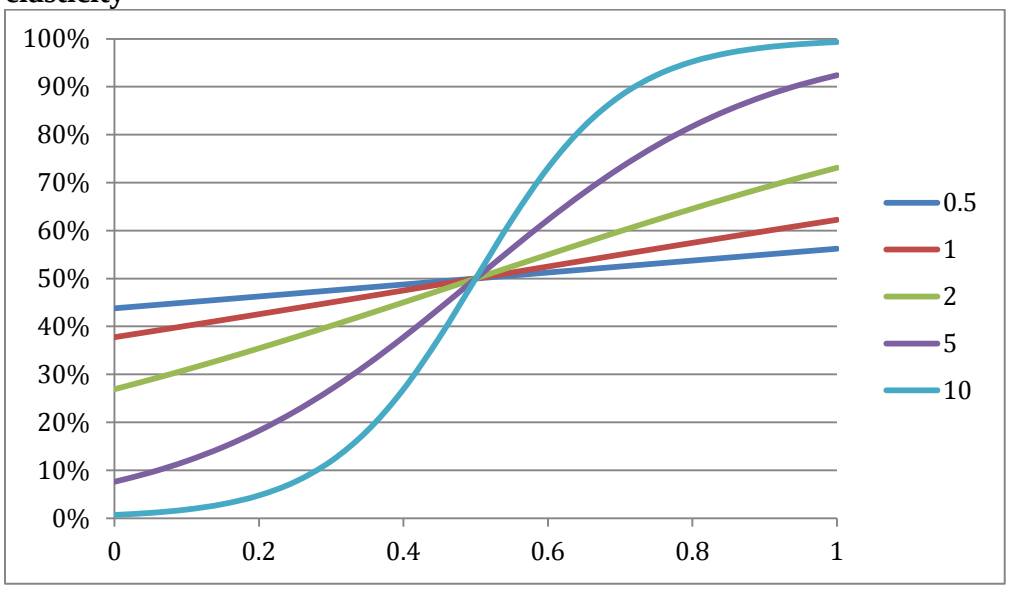
Life cycle effects, where products are apparently associated to a typical level of factor endowments as in HHR, are built in the model when countries have unequal factor supply. This comes as a consequence on the one hand of unequal factor supply between countries and on the other hand of market clearance constraints. Countries with the highest factor will concentrate their exports in the sector with the highest FSE.

Let us consider a minimalist **toy model** example with two sectors, a continuum of countries with factor f varying from 0 to 1, equal total exports per country, and equal demand for each of the two sectors. Let the elasticity be α for product 1 and zero for product 2 (α represents the difference in elasticity with respect to f in the two sectors)·

Then the relative strength of exports of product 1 as a function of f is proportional to $exp(\alpha(f - \frac{1}{2}))$, and the share of product 1 in the exports of a country with endowment f has an S shape (Fig 4)

$$x(f) = \frac{1}{2}\{1 + tanh(\frac{\alpha}{2}(f - \frac{1}{2}))\}.$$

Fig 4. Share of product 1 as a function of the factor endowment for different values of the elasticity

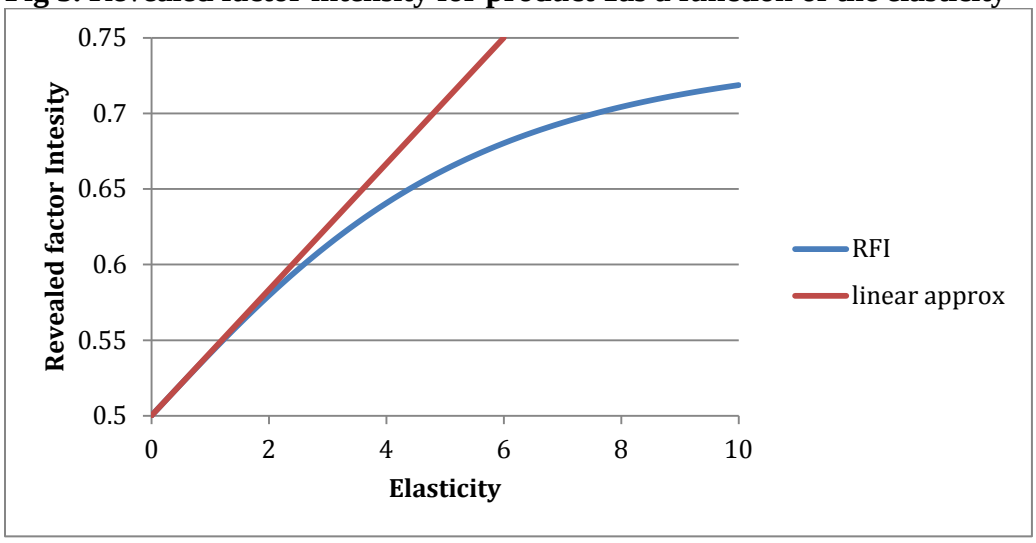


Under the assumption that countries' total exports are the same, the RFI of product 1 will be

$$RFI_1(\alpha) = \frac{\int_0^1 f * x(f)df}{\int_0^1 x(f)df} = 0.5 + \int_{-0.5}^{0.5} u * tanh(\frac{\alpha}{2}(u - \frac{1}{2}))du \quad .$$

This number ranges from 0.5 to 1. Fig 5 plots this function along with the linear approximation of the previous section. The toy model shows that the linear relationship RFI/ FSI may hold even for not-so-small elasticities (up to two or three).

Fig 5. Revealed factor intensity for product 1as a function of the elasticity



5 Product Space Geometry and Information Metric

This section tackles the empirical problem of fitting the original trade data X_{ai} against their predicted value of in the Product Space, or latent composition, described by the log-linear formula of section 4

$$\tilde{X}_{ai} = A_a B_i \, exp(\sum_k \alpha_i^k . f_a^k)$$

It uses standard econometric (Poisson Pseudo Maximum Likelihood) to estimate:

- The elasticities when trade and actual variables for the factors are known (section 5-6).
- Both elasticities and latent country factors variables from only the trade matrix (section 7 onward).

The procedure leads to a natural information metric of distance of countries to their projection (predicted value) onto the Product Space.

Poisson Pseudo Maximum Likelihood and its Information Metric

There is a strong theoretical and practical case (Silva & Tenreyro, 2006) (J. Arvis & Shepherd, 2013) to implement the Poisson pseudo maximum likelihood (PML) regression to this problem. It is widely accepted that the Poisson pseudo log-likelihood problem is the adequate econometric implementation to the similar gravity equation. It accepts zero values that are common in the trade matrix and behaves well in the presence of heteroskedasticity (higher variance for smaller flows). It also guarantees that the margin total in rows and columns are preserved and is the only PML with this property: $\sum_i \tilde{X}_{ai} = \sum_i X_{ai}$, and $\sum_a \tilde{X}_{ai} = \sum_a X_{ai}$.

Margin conservation is key to the consistent estimation of the fixed effect coefficients in the model (coefficients A and B). It is also essential to be able to use the predicted values of the model in further calculation, for instance to compare directly comparative advantages in the original matrix (RCAs) with those to be estimated from the predicted export matrix, hereafter referred to as latent comparative advantages (LCAs).

$$LCA_{ai} = \frac{\tilde{X}_{ai}\tilde{X}_{..}}{\tilde{X}_{a.}\tilde{X}_{.i}} = \frac{\tilde{X}_{ai}X_{..}}{X_{a.}X_{.i}} = \frac{\tilde{X}_{ai}}{X_{ai}} RCA_{ai}$$

The Poisson PML compares the original trade matrix X to the latent composition $\tilde{X}$

$$PML = -\sum_{a,i} \{X_{ai} Log(\frac{\tilde{X}_{ai}}{X_{ai}}) + X_{ai} - \tilde{X}_{ai}\} = -I$$

The PML is a negative or null number (it is null only when the predicted matrix equals the original). Its opposite I is an information metric (Kullback-Leibler (KL) information distance).In the reference case where no country has a comparative advantage ("flat world"), the trade matrix would be simply the product of shares in line and column

$$\tilde{X}^0_{ai} = \frac{X_{a.}X_{.i}}{X_{..}}$$

The information distance (Kullback Leibler distance) between the flat matrix and the actual one measures how non trivial the structure of export is, with strong specializations of countries.

$I^0 = -\sum_{a,i} X_{ai} Log(\frac{\frac{X_{a.}X_{.i}}{X_{..}}}{X_{ai}}) = \sum_{a,i} X_{ai} Log(RCA_{ai})$, while the KL distance between actual and latent export composition is $I = \sum_{a,i} X_{ai} Log(\frac{LCA_{ai}}{RCA_{ai}})$

According to McFadden's interpretation, the quality of a PML regression leading to the predicted value $\tilde{X}$ can be quantitatively measured by the relative improvement in the log-likelihood or, in this paper's context, the share of the information in the export matrix explained by the regression, namely:

$$\text{Pseudo-}R^2 = 1 - \frac{I}{I^0} = \frac{I - I^0}{I^0}$$

To What Degree do Countries Capture their Comparative Advantage?

The information metric I can be broken down according to the (relative) contribution of individual countries

$I = \sum_a X_a I_a$, with $I_a = \sum_i \frac{X_{ai}}{X_a} Log(\frac{\tilde{X}_{ai}}{X_{ai}}) = \sum_i \frac{X_{ai}}{X_a} Log(\frac{LCA_{ai}}{RCA_{ai}})$,

and for the "flat" composition

$I^0 = \sum_a X_a I^0{}_a$, with $I^0{}_a = -\sum_i \frac{X_{ai}}{X_a} Log(RCA_{ai})$.

The information metric I_a measures how close a country's export composition is to the latent composition estimated from the model. The smaller I_a is compared to the initial I^0_a, the better.

Econometric Implementation

The parameters to be estimated are the fixed effect coefficients, A_a and B_i, and the elasticities α_i. The maximization conditions yield three series of equalities.

$\sum_i X_{ai} = \sum_i \tilde{X}_{ai}$ for all a (maximization in A_a),

$\sum_a X_{ai} = \sum_a \tilde{X}_{ai}$ for all i (maximization in B_i), and

$\sum_a f_a{}^k X_{ai} = \sum_a f_a{}^k \tilde{X}_{ai}$ for all i (maximization in α_i) and each factor k.

These conditions are equivalent to saying that the estimator conserves i) the total in row and column, or margins, of the trade matrix (Arvis 2011), as well as ii) the weighted average of the country factors.

However, given the invariance properties of the estimators (section 3), there are two degrees of indetermination in the problem:

- First the A and B are determined up to a scale factor irrelevant to elasticities

$$A_a \to \lambda A_a \text{ and } B_i \to \frac{B_i}{\lambda}.$$

- In the numerical implementation we relieve the indetermination by imposing that the sums of A and B are equal.
- Then the elasticities are known up to a constant. To relieve this indetermination, in the following we choose the convention that the weighted average of elasticities is zero:

$$\langle \alpha_i^k \rangle_i = \frac{1}{X} \sum_i X_i \alpha_i^k = 0.$$

These conventions have no relevance to the predicted values of the model $\tilde{X}_{ai}$, but are needed for the implementation of the numerical algorithm. This algorithm recursively:

- Estimates the next A and B using the line and row conditions for a given set of elasticities α_i;
- Recomputes the next approximation for the elasticities α_i using Newton's approximation to solve the equation of conservation of average factor.

6 Application: Estimating the Factor Supply Elasticities for Known Endowments

This section applies the model to actual trade data when factors of endowments are known, and determines the factor supply elasticities by product.

The Poisson pseudo maximum likelihood is implemented for one year (2010) starting with 61 sectors at the 2-digit SITC rev 3 level. The independent variables used in the model are indicators typically used to explain trade volumes or trade costs.

The following calculations have been implemented:

1. Computation of the factor supply intensity and the revealed factor intensity for each of the variables (Table 1).
2. Computation of multivariate factor supply intensity for a combination of independent variables (Tables 2-3).

The main findings are:

1. Individual indicators taken separately do not have a very strong explanatory power (max 15%). That is, no single endowment can explain the export matrix well.
2. The FSE and RFI are as expected (section 4) strongly correlated (Fig 7 and Table 5).
3. Results 1 and 2 imply that RFI/PRODY explains a small share of the information in the export matrix. The explanatory power of such univariate indicators is more limited than previously thought.
4. Including several variables improves the information explained by the model but not too strongly: five variables cannot explain 50% of the information. Furthermore, the multivariate elasticities are expectedly quite dependent on the other variables included in the model.

Table 1. Information explained by individual selected "competitiveness" variables

Variable	Source	Countries	Information in model
GDP per capita (current US$) *log*	WDI	121	0.149
GDP, PPP (current international $) *log*	WDI	119	0.0557
Liner shipping connectivity index 2010	UNCTAD	103	0.145
Time required to start a business (days)	WDI	125	0.062
Natural Capital, $ per worker *log*	WDI	100	0.136
Physical Capital Stock per Worker *log*	UNCTAD	111	0.145
Governance: Political Stability and Absence of Violence 2011	WDI	128	0.113
Governance: Government Effectiveness 2011	WDI	128	0.118
Governance: Regulatory quality 2011	WDI	128	0.123
Governance Rule of Law 2011	WDI	128	0.136
Governance: Control of Corruption 2011	WDI	128	0.130
Governance: Voice and Accountability 2011	WDI	128	0.167
lpi score 2010	WDI	119	0.119
tons equivalent per unit of gdp *log*	WDI	110	0.076

Table 2. Multivariate FSE estimates (two variables)
First variable is log GNI per capita.

Second Variable	countries	Information content
Liner shipping connectivity index	96	0.274
Natural Capital, $ per worker	95	0.289
Physical Capital Stock per Worker	106	0.203
Governance: Political Stability and Absence of Violence 2011	121	0.163
Governance: Government Effectiveness 2011	121	0.186
Governance: Regulatory quality 2011	121	0.175
Governance: Control of Corruption 2011	121	0.180

Table 3. Multivariate FSE estimates (three variables or more).
First variable is log GNI per capita.
The second variable is liner shipping connectivity index (maximum value in 2004 =100).

# Variable (third and more)	count	information
3. Governance: Political Stability and Absence of Violence 2011	96	0.290
3. tons equivalent per unit of gdp	89	0.328
3. Physical Capital Stock per Worker	88	0.322
3. Physical Capital Stock per Worker 4. Governance: Political Stability and Absence of Violence 2011 5. tons equivalent per unit of gdp	83	0.382
3. Natural Capital, $ per worker	78	0.367
3. Natural Capital, $ per worker 4. Physical Capital Stock per Worker 5. tons equivalent per unit of gdp	69	0.455
3. Natural Capital, $ per worker 4. Governance: Political Stability and Absence of Violence 2011 5. tons equivalent per unit of gdp	72	0.411

Table 4. Correlation of Revealed Factor Intensity vs. Factor elasticity

GDP per capita (current US$)	0.536
GDP, PPP (current international $)	0.605
Liner shipping connectivity index	0.657
Time required to start a business (days)	0.101
Natural Capital, $ per worker	0.283
Physical Capital Stock per Worker	0.540
Governance: Political Stability and Absence of Violence 2011	0.597
Governance: Government Effectiveness 2011	0.645
Governance: Regulatory quality 2011	0.530
Governance Rule of Law 2011	0.584
Governance: Control of Corruption 2011	0.604
Governance: Voice and Accountability 2011	0.465
LPI score 2010	0.751
tons equivalent per unit of gdp	0.814

7 Dimensionality Reduction and Generalized Principal Component Analysis

This section extends the model of sections 4 and 5 to the case where the factors of endowments are not known but are latent variables to be determined from the trade data.

Dimensionality reduction is one of the most powerful concepts in science in order to address the complexity of problems with many degrees of freedom. It helps read data and understand patterns. The most popular techniques of dimensionality reduction are linear techniques such as principal component analysis (PCA) or factor analysis.

Standard linear techniques such as PCA cannot be applied directly for two rather obvious reasons. The first is that the impact of factors is expected to be log linear as apparent in the previous sections, not linear as in factor analysis or PCA. The second is that export matrices are rather sparse with a large number of zeros, so log-linearization of trade data will not work as is. Fortunately, there is a consistent way to implement the PCA concept with the previous FSE and Poisson regression techniques. This non-linear version of the PCA is related to the exponential-PCA (E-PCA) introduced by Collins et al. (Collins 2002) (Collins, Dasgupta, & Schapire, 2002) for pattern recognitions algorithm. This might be the first application of this idea to economics. This technique, in spite of the apparent algorithmic complexity, offers much insight into the structure of the product space.

Latent competitiveness variables

What this model does is essentially invert the generic trade model of section 3 and deduce from the actual trade flows the product elasticities and country factor indices for a few latent variables that explain most of the information in the trade matrix. Namely, we want to transform and decompose the trade matrix so it is ultimately explained by a series of separable factors of competiveness f_a^k by country and with elasticities by product α_i^k: $\tilde{X}_{ai} = A_a B_i exp(\sum_{k=1}^{d} \alpha_i^k f_a^k))$.

The decomposition should provide a small number d of significant dimensions, where countries and products are represented in two dual d dimensional spaces, where the fs are the coordinates of countries and the αs are the coordinates of products (Fig 2). It happens that this is a well-conditioned problem that can be solved by iteration. The iteration extracts sequentially the latent variables according to their decreasing contribution to the information on the trade matrix as measured by the log-likelihood.

As before, the problem has under-determination, since it is invariant to a constant shift in factors or elasticity as well as to an inversely proportional change of scale between factors and elasticities. Conventions similar to that of the PCA naturally address this issue (centered and normalized factors and centered elasticity):

$\text{weighted mean}[\alpha^k]=0$ is the same choice as in section 4 to fix the constant for elasticities;

$Mean[f^k]=0$ fixes the constant for the factors; and

$Var[f^k] = 1$ fixes the scale for both factors and elasticities since the product $(\sum_{k=1}^{d} \alpha_i^k f_a^k)$ is invariant by $\alpha \rightarrow \lambda\alpha$ and $f \rightarrow f / \lambda$.

Iterations

The E-PCA (Annex 3) generates a series of predicted matrices $\tilde{X}^l$ with an increasing number l of dimensions/variables, beginning with the "flat world" no-comparative advantage value:

$$\tilde{X}_{ai}^0 = \frac{X_a X_i}{X}$$

Each stage adds dual sets of factors and elasticities that provide the best fit with the original trade matrix. The estimator with l number of already determined latent variables is

$$\tilde{X}_{ai}^l = A_a^l B_i^l exp(\sum_{k=1}^{l} \alpha_i^k f_a^k)$$

and should minimize: $I^l = -L^l = -\sum_{ai} X_{ai} Log \frac{\tilde{X}_{ai}^l}{X_{ai}} - \tilde{X}_{ai}^l + X_{ai}$

The information improvement brought by the variable l is the improvement in log-likelihood from l-1 to l, or

$$J_l = L^l - L^{l-1} = \sum_{ai} X_{ai} Log \frac{\tilde{X}_{ai}^l}{\tilde{X}_{ai}^{l-1}}$$

The iteration from step l to l+1 proceeds as follows.

The impact on trade and comparative advantage of the first l variables comes from the coefficient:

$K_{ai}^l = exp(\sum_{k=1}^{l} \alpha_i^k f_a^k)$, with for l= 0 and $K_{ai}^0 = 1$. K_{ai}^l is known from the previous iteration up to stage l, and the predicted value at stage l+1 is

$$\tilde{X}_{ai}^{l+1} = A_a^{l+1} B_i^{l+1} exp(\alpha_i^{l+1} f_a^{l+1}) K_{ai}^l$$

The problem is formally quasi-identical to the determination of the FSE in section 5, except that α^{l+1} and f^{l+1} have to be determined simultaneously. This is achieved at each stage l+1 by:

- Seeding a randomly distributed f^{l+1} centered and normalized (zero mean and unit variance).
- Computing the corresponding elasticities α^{l+1} as per the algorithm in section 5, and centering them.
- From the elasticities, computing with the same algorithm a new iteration of f^{l+1}.

- Iterating until convergence, which is experimentally rather fast.

This procedure seems to produce stable results independent of the seed. The sign of factors and elasticities may be simultaneously changed from one trial to another, which has no relevance.

The number of dimensions to be retained is determined by procedures similar to the linear principal component analysis such as the scree plot.

8 Results: Latent Variables and Interpretation of Latent Factors

The model of section 7 is implemented for one year (2010) starting with the 2-digit SITCS rev 3 level. Energy trade is excluded. Twelve principal components are computed in order to estimate the number of relevant dimensions.

Number of dimensions in product space

As in PCA, the determination of the significant number of dimensions can be done by looking at the contribution to the increase of the log-likelihood L^n or the reduction in unexplained information, at stage n, I^n.

By construction the informational content of each successive variable $J^n = I^n - I^{n+1}$ is decreasing

$$J^1 > J^2 > \ldots > J^n > \ldots$$

hence the convex shape of the information curve and the decreasing one for the Js (see the scree plot in Fig 9). Table 5 below provides, for the first 12 component factors, the values of the contribution of component i as well as the cumulated contribution and the unexplained information in the matrix.

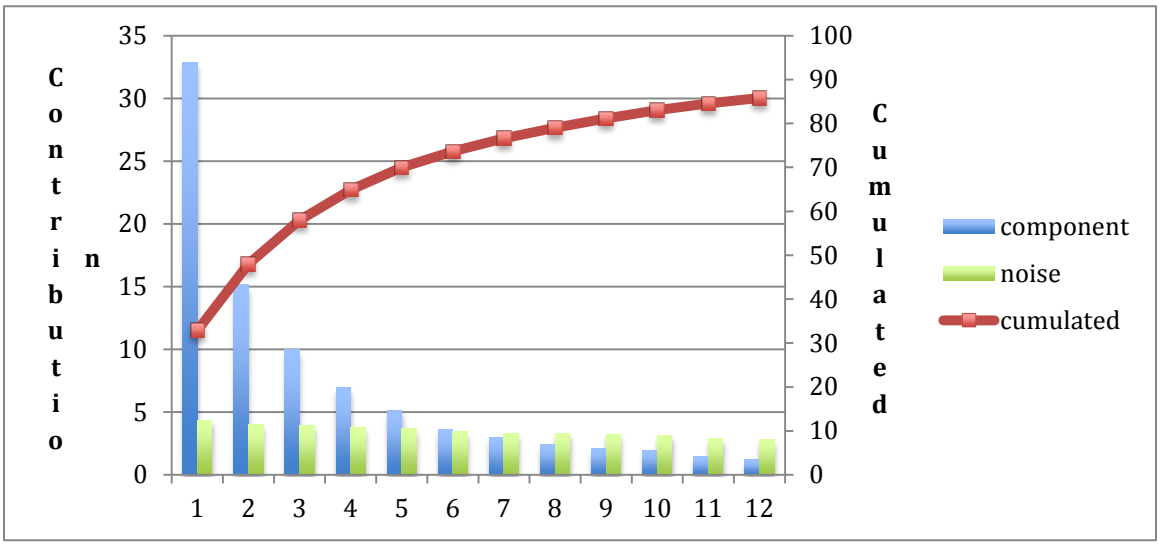
The number of dimensions relevant to the problem corresponds to the "bottom" of the scree where the relative information or improvement in the log-likelihood J^{n+1} brought by the next variable becomes significantly less than with the previous J^n, and could be considered as "noise". Unfortunately, as often is the case, visual inspection (Fig 6) is not enough to identify the "bottom" of the scree. A more rigorous criterion is to compare the contribution with the one corresponding to a pure random "noise" in the data. In such a case, the contributions are in a geometric series with a ratio close to one. The noisy principal components are easily generated with computer algebra software for the same number of countries and products.

Table 5. Improvement in information brought by the first 20 factors.

	Information			
	Unexplained I	Contribution J	Explained 100-I	"Noise" level
Factor	**100.00**			
1	**67.12**	32.88	32.88	4.36
2	**51.99**	15.13	48.01	4.01
3	**42.04**	9.95	57.96	3.95
4	**35.12**	6.92	64.88	3.75

5	**29.98**	5.13	70.02	3.65
6	**26.39**	3.59	73.61	3.41
7	**23.38**	3.01	76.62	3.32
8	**20.97**	2.41	79.03	3.26
9	**18.84**	2.13	81.16	3.23
10	**16.93**	1.91	83.07	3.14
11	**15.46**	1.47	84.54	2.92
12	**14.21**	1.24	85.79	2.81

Fig 6. Scree plot and explained information in the trade matrix.



In this case, the first two dimensions explain about half of the information in the trade matrix. However, there is no sharp drop in the contribution for the following dimensions. The determination of the bottom of the scree is somewhat arbitrary, in between the third and sixth dimensions. We retain five dimensions, as from the sixth dimension onward the contribution of the component is not distinguishable from the "noise" value.

Latent Variables and their Interpretation

The latent variables and factors for countries and products are available in the Excel files available in the online Annex. Unsurprisingly, compared with known variables (section 5), the latent variables perform much better at explaining the export matrix. Unfortunately, the interpretation of the variables is not totally obvious. The correlation between the known and latent variables is limited. This should be expected since section 5 showed that none of the typical variables explains much of the information structure of the actual matrix (10-20%).

The level of development, logistics and connectivity variables are better associated with the first two factors, as well as the variables measuring government effectiveness (the latter more for the second factor). However, the Doing Business variables do not appear as important to explaining the structure of the export matrix. This result probably stems from the fact that trade is carried

out not by entrant SMEs but by large established firms, while the Doing Business indicator on entry is relevant to the former.

Table 6. Pairwise correlation of known variables vs. latent variables

	f1	f2	f3	f4	f5
GDP per capita (current US$)	-0.375	-0.434	-0.072	-0.102	-0.002
GDP, PPP (current international $)	-0.378	-0.288	-0.009	-0.125	0.112
Liner shipping connectivity index (maximum value in 2004 =100)	-0.325	-0.371	-0.025	-0.135	0.162
Time required to start a business (days)	0.020	0.026	-0.023	0.085	0.075
Natural Capital, $ per worker	-0.001	-0.217	-0.044	-0.130	0.058
Physical Capital Stock per Worker	-0.390	-0.405	-0.001	-0.089	-0.069
Average Years of Schooling for 25 years and over	-0.333	-0.256	0.015	-0.039	-0.075
Arable Land hectares per person	0.191	-0.031	-0.007	-0.123	-0.108
Arable Land hectares per worker	0.212	0.008	-0.002	-0.100	-0.097
Governance: Political Stability and Absence of Violence 2011	-0.177	-0.374	-0.097	-0.008	-0.112
Governance: Government Effectiveness 2011	-0.309	-0.458	-0.004	-0.074	-0.123
Governance: Regulatory quality 2011	-0.294	-0.447	0.007	-0.105	-0.196
Governance Rule of Law 2011	-0.298	-0.454	-0.006	-0.047	-0.137
Governance: Control of Corruption 2011	-0.199	-0.423	-0.019	-0.027	-0.152
Governance: Voice and Accountability 2011	-0.187	-0.376	0.005	-0.056	-0.249
energy exports as % total exports 2009	0.153	0.140	-0.180	0.010	0.392
energy exports as % total exports 2011	0.090	0.115	-0.041	-0.147	0.254
lpi score 2007	-0.399	-0.505	-0.058	-0.112	-0.132
lpi score 2010	-0.377	-0.482	0.010	-0.193	-0.159
tons equivalent per unit of gdp	0.290	0.354	-0.146	0.079	0.191

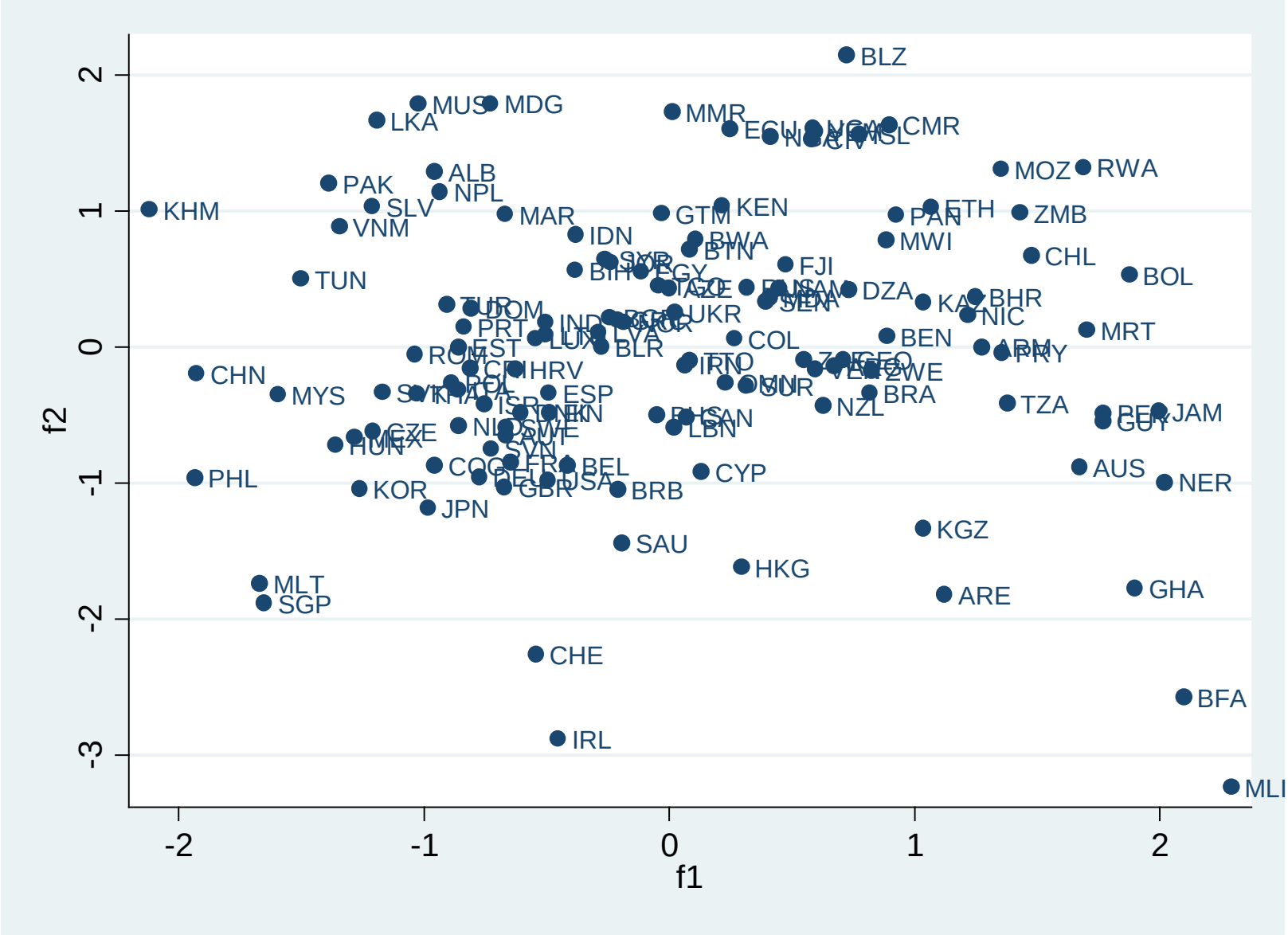
A number of patterns emerge from the pairwise correlation and the inspection of the correlation matrix as well as the reading of the distribution of countries and products in the first two dimensions:

- The first axis essentially distinguishes between labor intensive vs. resources intensive productions.
- The second axis distinguishes a greater degree of sophistication of manufacturing and the need for investment.

Fig 7. Two First Axis Products



Fig 8. Two first axis countries



9 Latent Comparative Advantage

This section builds on the results of the previous one, and proposes tools with practical policy value. It follows the route sketched out in the introduction and exploits the concepts of latent trade composition and latent comparative advantage, to propose for instance, **tools to identify the latent potential for diversification for any country**. It also looks at how to interpret the separation between actual (or revealed) values and latent values. **The main hypothesis tested here is that competitiveness is intimately associated to how well countries reveal their latent advantages.**

The concept of dimensional reduction developed in section 7 determined a limited number of latent factors that explain most of the export composition. The projection of a country onto the low-dimensional product space provides a direct identification of exports for which a country has a revealed or just-latent comparative advantage, given by its position in the product space and its endowment in factors explaining competitiveness. Indeed the projected composition

$$\tilde{X}_{ai} = A_a B_i exp(\sum \alpha_i^k f_a^k)$$

is what the export profile would be (with unchanged global demand by product), if exports were only dependent upon latent competiveness variables independently of the "noise" that makes the actual trade matrix X_{ai} deviate from the estimator. For instance, the estimator is always positive while actual trade has many zeros.

Hence $\tilde{X}_{ai}$ indicates for which products country a has a latent but not necessarily revealed comparative advantage. It is natural to introduce the latent comparative advantage as a match to Balassa's revealed comparative advantage.

$$LCA_{ai} = \frac{\tilde{X}_{ai}\tilde{X}_{..}}{\tilde{X}_{a.}\tilde{X}_{.i}} = \frac{\tilde{X}_{ai}X_{..}}{X_{a.}X_{.i}}$$, while $$RCA_{ai} = \frac{X_{ai}X_{..}}{X_{a.}X_{.i}}$$

Since the estimator $\tilde{X}_{ai}$ is smoother product-wise than the initial matrices, the LCAs have less dispersion than the RCAs. A direct pairwise comparison of LCAs and RCAs is not informative. However, there are meaningful ways to compare actual and latent trade data by country.

Latent and Actual Composition Profiles by Country

It is proposed to retain the following classical indicators to compare the latent and actual trade composition by country:

1. Theil Index of actual composition: $\text{Theil actual}_a = -\sum_i \frac{X_{ai}}{X_a} \log \frac{X_{ai}}{X_a}$
2. Theil Index of latent composition: $\text{Theil latent}_a = -\sum_i \frac{\tilde{X}_{ai}}{\tilde{X}_a} \log \frac{\tilde{X}_{ai}}{\tilde{X}_a}$

3. Information (Kullback Leibler) distance between the latent and actual export compositions (sections 5-7). $\text{KL Distance}_a = \sum_i \frac{X_{ai}}{X_a} \log \frac{\tilde{X}_{ai}}{X_a}$
4. The degree of overlap, or similarity indicator, between the latent and actual export compositions. $\text{overlap}_a = \frac{\sum_i Min(\tilde{X}_{ai}, X_{ai})}{X_{a.}}$

The latent Theil value is expected to be higher than the actual one because the latent composition is spread over product lines as compared to the original one. The KL distance is a positive number, which is zero for identical composition and increases when the composition diverges, because the latent composition is more spread out as compared to the original one. . The overlap or similarity measure has an inverse behavior, decreasing from one for a perfect match of latent and actual to zero for no overlap.

Cross-matrix of Latent versus Revealed Comparative Advantage by Country: Identifying the Potential for Diversification

Another informative and intuitive way to look at the same problem is to do a double-typology of products for each country according to whether their RCA and LCA are greater than one.

The four-category typology is the typical BCG interpretation:

Type one (RCA>1, LCA>1) represents products for which the latent comparative advantage is realized.

Type two (RCA<1, LCA>1) are those products for which the latent advantage is not revealed. The category is eventually an implementation of the concept of discovery (HHR) or intensive margin.

Type three (RCA>1, LCA<1) corresponds to products where the revealed comparative advantage is not supported by the position in the reduced product space, hence a possible interpretation is sectors with declining competitiveness.

Type four (RCA<1, LCA<1) corresponds to products with no particular advantage, which is typical and can represent a substantial fraction of exports.

The straightforward implementation of this typology leads to a simple tool, a “discovery” matrix that maps for each country the products according to this typology in four categories. The matrix is available in Annex 4. About half of the country-product pairs with a latent advantage are not revealed (Table 8).

Table 7. Product typology and statistics (% of country-product pairs)

	RCA > 1	**RCA < 1**
LCA > 1	Type 1 19.2%	Type 2 19.8%
	Revealed potential	Untapped potential of diversification
LCA < 1	Type 3 7.1%	Type 4 53.9%
	Declining sectors?	Marginal sectors

For each country the following derived indicators can be produced:

1. The number of products in each of the four categories: Type 1-4 #
2. The share of latent trade in each of the four categories Type 1-4 %

Competitiveness as the Realization of the Latent Advantage

The data file provided in Annex 5 includes:

- Indices by country: Theil actual, Theil latent, KL Distance, Overlap, Type 1-4 #, Type 1-4 %,
- The "discovery" matrix with the product typology by country.

The main observations are the following.

LCA metrics. The country-level indices of KL distance, overlap, and GL are expectedly related as they all measure how close latent and actual export compositions are. As expected, distance or overlap between actual and latent are also related to the relative importance of the untapped potential for diversification (Type 2).

Most significantly, the KL distance and the overlap indices seem to capture much of the export "competitiveness", with the KL distance being slightly better:

i. Both indicators are highly correlated with commonly accepted indicators of competitiveness such as the WEF Global Competitiveness index (Fig 9).
ii. The ranking of countries in the sample is consistent with the general knowledge of countries' exports (Table 10), despite a few unintuitive cases such as the Nordic countries.
iii. The KL distance is also associated to a higher degree of diversification of their actual trade for countries close to their latent potential, as measured by the Theil index (Fig 10).
iv. The metric of distance between latent and actual trade is better associated with generally accepted competitiveness outcomes than the diversification (Theil) index.

Table 8. Pairwise correlations

	Theil actual	KL distance	overlap	WEF CGI
KL distance	-0.6177	1		
overlap	0.4045	-0.8787	1	
WEF CGI	0.4976	-0.6096	0.5723	1
GDP per capita 2010	0.4875	-0.5569	0.4937	0.8635
Shipping Connectivity	0.308	-0.4389	0.5233	0.5204
Capital per worker	0.5091	-0.592	0.4937	0.7952
Government effectiveness	0.3734	-0.4957	0.4981	0.891
LPI 2010	0.5577	-0.6233	0.6014	0.8715

Table 9. Ranking of countries according to the information metric (KL Distance in increasing value)

country	rank	country	rank	country	rank	country	rank	country	rank	country	rank
DEU	1	IRL	23	ZMB	45	CRI	67	MLT	89	FJI	111
CHN	2	SVN	24	TUN	46	NOR	68	MDG	90	DZA	112
USA	3	CHL	25	HRV	47	JAM	69	COL	91	AZE	113
KOR	4	DNK	26	ARG	48	TZA	70	CYP	92	JOR	114
JPN	5	AUS	27	PHL	49	NPL	71	MAR	93	NGA	115
GBR	6	HUN	28	BOL	50	UGA	72	OMN	94	MMR	116
CAN	7	SWE	29	FIN	51	MOZ	73	LBN	95	BLZ	117
CZE	8	THA	30	BIH	52	BWA	74	BTN	96	ETH	118
POL	9	TUR	31	ISR	53	PRY	75	MLI	97	CMR	119
ITA	10	RUS	32	ECU	54	YEM	76	NAM	98	SUR	120
SGP	11	VNM	33	IRN	55	LKA	77	NER	99	MUS	121
MEX	12	ROM	34	UKR	56	SLV	78	MDA	100	ZWE	122
AUT	13	CHE	35	GTM	57	PAK	79	GEO	101	GHA	123
NLD	14	VEN	36	LVA	58	NIC	80	MRT	102	CIV	124
FRA	15	EST	37	PAN	59	BRB	81	GUY	103	KGZ	125
BEL	16	LTU	38	KEN	60	SYR	82	TGO	104	BEN	126
ESP	17	BRA	39	EGY	61	ARE	83	HKG	105	BHS	127
IND	18	GRC	40	NZL	62	DOM	84	SEN	106	MWI	128
MYS	19	BGR	41	PER	63	ALB	85	TTO	107		
SVK	20	KAZ	42	BHR	64	ISL	86	COG	108		
ZAF	21	LUX	43	SAU	65	RWA	87	BFA	109		
PRT	22	IDN	44	ARM	66	BLR	88	KHM	110		

Discovery matrix (Annex 4). Under visual inspection, the matrix is consistent with expert knowledge of countries, including developing countries targeted by World Bank assistance.

The data for European countries also provide some illustration of the differences between countries. Table 10 below illustrates the case of Germany, France, Italy, Netherlands, Spain, and the United Kingdom.

All countries are relatively close to their latent structure as measured by the indicators of distance and overlap. However, Germany clearly stands out by being closer and having a very small extensive margin of products to be revealed. Conversely, France and the Netherlands have a rather large extensive margin. At least in the case of France, this observation and the fact that its indicators are about the same as Spain cannot but be associated to the current debate in France about the competitiveness and the future of its industry (Giraud & Weil, 2013).

Table 10. Latent comparative data for selected EU countries

countries	DEU	ESP	FRA	ITA	GBR	NLD
Theil actual	3.33	3.42	3.47	3.51	3.35	3.59
KL Distance	0.02	0.12	0.12	0.09	0.07	0.11
overlap	0.93	0.80	0.83	0.85	0.86	0.80
Type 1#	16	25	14	22	18	19
Type 2#	3	15	14	9	9	16
Type 3#	2	3	7	4	0	4
Type 4#	40	18	26	26	34	22

Fig 9. Distance between actual and real vs. WEF Global Competitiveness Indicator.

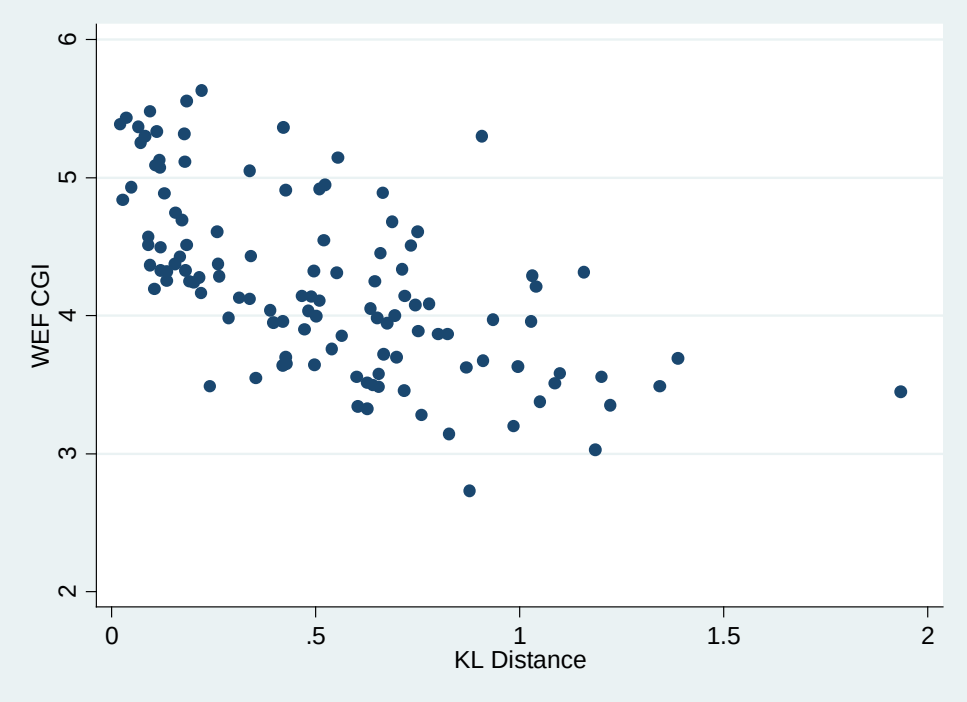
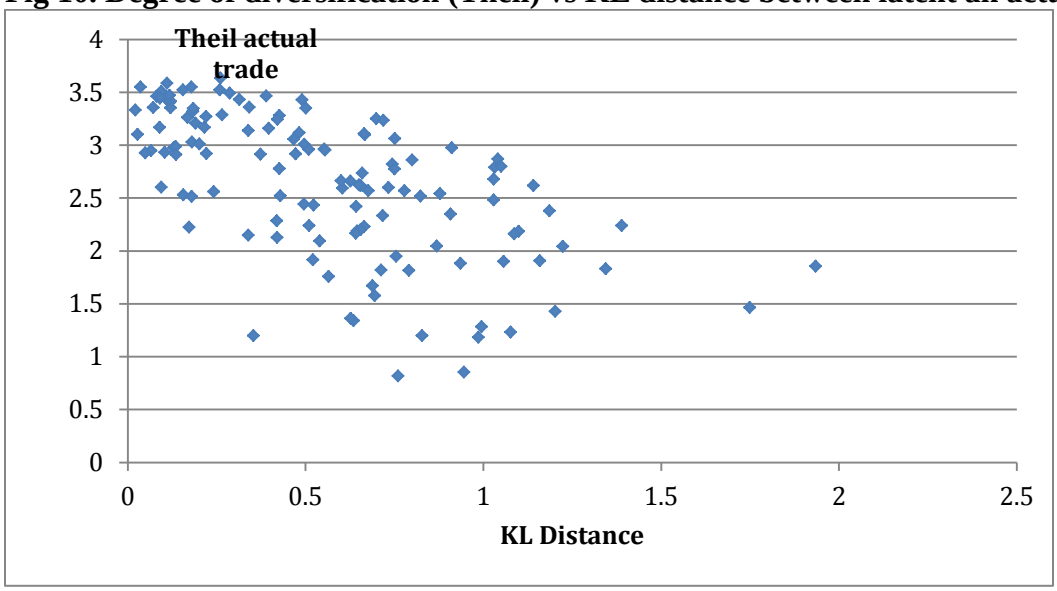


Fig 10. Degree of diversification (Theil) vs KL distance between latent an actual



10 Conclusions

The model is a thorough implementation of dimensionality reduction of the export matrix. It provides a quantitative description of dual product spaces where countries and products have coordinates numerically estimated. Furthermore, the model provides a latent composition and latent comparative advantage, a new concept representing the reference export composition implied by the position of the country in the product space. This position not only depends upon what a country actually exports (revealed comparative advantage) but also upon the export composition of all other countries. The comparisons provide some practical indication of the potential for diversification.

Main findings

1. The product space has relatively few dimensions (five explain 70% of the trade matrix).
2. Typical competiveness variables are weakly correlated to individual latent factors (coordinates in the product space).
3. Latent comparative advantages are available and comparisons with RCA inform on the extensive margins and the potential for diversification.
4. Typical competiveness variables are related to distance between latent and revealed (actual) positions in the product space.
5. Economies with less distance, or making the most of their latent advantage, are expectedly more “competitive” and more diversified.
6. The currently accepted concepts such as PRODY or RFI have, in fact, limited informational power.

The model presents a number of advantages. It is not only a descriptive data analysis tool but it is also embedded in classical trade theory. Therefore it is quantitative and eventually predictive. For instance, it allows estimates of how trade will change a country’s latent variables and endowments based upon known country factors. The associated information metric measures how much of the trade structure is explained by the model (about 70%). Furthermore the model

does not rely on arbitrary choices of parameters or of functional structure. The only partly subjective but not ad hoc interpretation is the reading of the scree plot, as in a standard PCA, to determine the number of relevant dimensions of the product space.

Improvements

Several improvements are relatively straightforward:

- The model has been implemented at 2-digits. Implementation at 3- or 4-digits should be considered.
- Improvements in the E-PCA implementation and further test of robustness.
- More detailed comparisons between latent variables and known indicators of competitiveness.
- Inclusions of several years in the model.

Extensions

The following areas have not been considered so far.

- Inclusion of import composition as well as export composition in order to better incorporate a value chain description of trade.
- Inclusion of geographical information in the model and how to incorporate spatial gravity modeling.

References

Anderson, J.E. (2010). The gravity model. *The World Economy*, (March).

Anderson, James E, & Van Wincoop, E. (2003). Gravity with Gravitas: A Solution to the Border Puzzle. *American Economic Review*, *93*(1), 170–192.

Armington, P. S. (1969). Theory of Demand. *Staff Papers-International Monetary Fund*, 159–178.

Arvis, J. J. F., & Shepherd, B. (2011). The air connectivity index: measuring integration in the global air transport network. *Bank Policy Research Working Paper No.*, (June).

Arvis, J., & Shepherd, B. (2013). The Poisson quasi-maximum likelihood estimator : a solution to the " adding up " problem in gravity models The Poisson quasi-maximum likelihood estimator : a solution to the " adding up " problem in gravity models, (August 2012), 37–41.

Balassa, B. (1979). THE CHANGING PATTERN OF COMPARATIVE ADVANTAGE. *The Review of Economics and Statistics*, *61*(2), 259–266.

Balassa, B. (1986). Comparative Advantage in Manufactured Goods : A Reappraisal. *The Review of Economics and Statistics*, *68*(2), 315–319.

Baldwin, R. (2010). Relocating the Value Chain: Offshoring and Agglomeration in the Global Economy.

Barigozzi, M., Fagiolo, G., & Garlaschelli, D. (2010). Multinetwork of international trade: A commodity-specific analysis. *Physical Review E*, *81*(4).

Collins, M., Dasgupta, S., & Schapire, R. E. (2002). A generalization of principal component analysis to the exponential family. *Advances in neural information processing systems* (Vol. 1, pp. 617–624). Citeseer.

Giraud, P.-N., & Weil, T. (2013). Comment faire revenir l'emploi industriel en France. *Le Monde*, (03/19/2013).

Hausmann, R., Hwang, J., & Rodrik, D. (2006). What you export matters. *Journal of Economic Growth*, *12*(1), 1–25.

Hausmann, R., & Klinger, B. (2006). *Structural transformation and patterns of comparative advantage in the product space*. Center for International Development at Harvard University. Retrieved from

Hidalgo, C. a, & Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(26), 10570–5. doi:10.1073/pnas.0900943106

Leamer, E. E. (1993). Factor-supply differences as a source of comparative advantage. *The American Economic Review*, *83*(2), 436–439.

McMillan, M. S., & Rodrik, D. (2011). Globalization, structural change and productivity growth. *America*.

Neary, J. (2006). Measuring competitiveness. *Economic and Social review*, *37*(2), 197–213.

Neary, J. P. (2003). Competitive versus comparative advantage. *The World Economy*, *26*(4), 457–470.

Nunn, N. (2007). Relationship-specificity, incomplete contracts, and the pattern of trade. *The Quarterly Journal of Economics*, (May), 569–600.

Reis, J., & Farole, T. (2012). *Trade Competitiveness Diagnostic Toolkit*. (T. W. Bank, Ed.). Washington DC.

Shirotori, M., Tumurchudur, B., & Cadot, O. (2010). REVEALED FACTOR INTENSITY INDICES at the product level. *POLICY ISSUES IN INTERNATIONAL TRADE AND COMMODITIES STUDY SERIES No. 44*, (44).

Silva, J. M. C. S., & Tenreyro, S. (2006). The Log of Gravity. *Review of Economics and Statistics*, *88*(4), 641–658.

WEF. (2012). *Global Competitiveness Report 2012-2013*. Geneva.

Annex 1 Results

Three MS Excel files are available at

www.worldbank.org/trade -> data

Annex 1 corresponds to the RFI and FSE and contains the variables used.

Annex 2 corresponds to section 8 and gives the latent variables and elasticities.

Annex 3 corresponds to section 9 and provides the "discovery matrix" and the latent comparative variables.

The core data work has been programmed in Mathematica®. The notebooks can be shared upon request.

Annex 2 Properties of the Model and Connection between Factor Supply Elasticity and Revealed Factor Intensity

Improvement in factor supply and change in comparative advantage

Consider a small change, ceteris paribus, of the factor supply for a given country a,

$$f_a \to f_a + \delta f_a$$

Then the change in the matrix is

$X_{ai} = A_a B_i exp(\alpha_i f_a) \to X_{ai}(1 + \alpha_i \delta f_a)$, or a change

$\delta X_{ai} = X_{ai} \alpha_i \delta f_a$, yielding the changes for total in rows and columns

$\delta X_a = (\sum_i X_{ai} \alpha_i) \delta f_a$, and $\delta X_i = X_{ai} \alpha_i \delta f_a$

and for the RCA

$$RCA_{ai} \to RCA_{ai}\{1 + \frac{\delta X_{ai}}{X_{ai}} - \frac{\delta X_{a.}}{X_{a.}} - \frac{\delta X_{.i}}{X_{.i}} + \frac{\delta X_{..}}{X_{..}}\}$$

$$\delta RCA_{ai} = RCA_{ai}\{\alpha_i - \sum_j \frac{X_{aj}}{X_{a.}} \alpha_j - \frac{X_{ai}}{X_{.i}} \alpha_i + \sum_j \frac{X_{aj}}{X_{..}} \alpha_j\} \delta f_a$$, or

$$\delta RCA_{ai} = RCA_{ai} * \{1 - \frac{X_{ai}}{X_{a.}}\}(\alpha_i - \lambda \bar{\alpha}_a) \delta f_a$$, where

$\bar{\alpha}_a = \sum_j \frac{X_{aj}}{X_{a.}} \alpha_j$ is the weighted average elasticity for country a and the coefficient

$\lambda = \frac{1 - \frac{X_{.i}}{X_{..}}}{1 - \frac{X_{ai}}{X_{a.}}}$, is in practice close to one: $\lambda \approx 1$,

Thus an improvement in factor supply by δf_a for country a improves the RCA only in export products i where the factor elasticity α_i is above the country weighted average of elasticity $\bar{\alpha}_a$.

Relationship between PRODY/Revealed Factor Intensity and Factor Supply Elasticity

In this sub-section we show how the model explains the significance of the revealed factor intensity introduced in the literature, and how it relates to factor supply elasticities. Consider a trade matrix with relatively mild effects of factor productivity (i.e. first order deviation from a situation of no comparative advantage). Then

$$X_{ai} = A_a B_i exp[\alpha_i f_a]$$

Elasticities and factors can be normalized by convention with zero weighted averages without loss of generality

$$\sum_i B_i \alpha_i = 0, \text{ and } \sum_a A_a f_a = 0.$$

At first order:

$$X_{ai} \approx A_a B_i (1 + \alpha_i f_a), \text{ and } X_a \approx A_a \sum_i B_i \ , \ X_i \approx \sum_a A_a B_i$$

Thus the RCA is

$$RCA_{ai} = 1 + \alpha_i f_a$$

$$\sum_a RCA_{ai} = N + \alpha_i (\sum_a f_a)$$

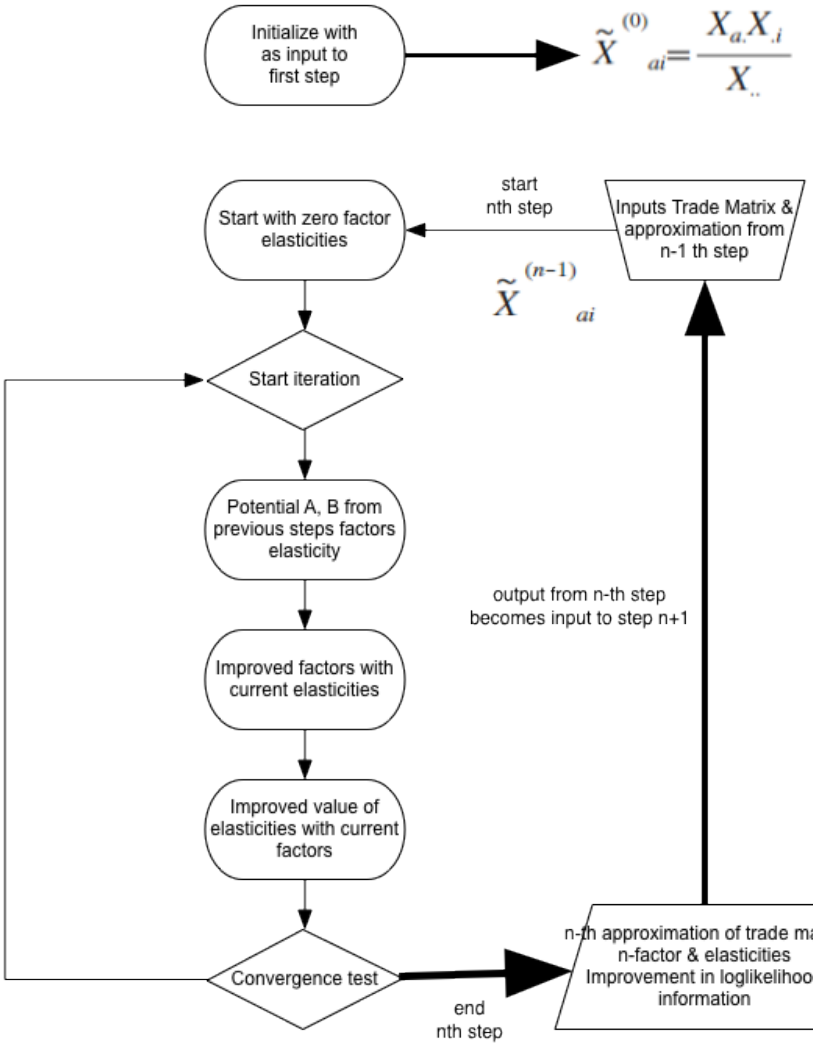
And the RFI is $\bar{f}_i = \sum_a \frac{RCA_{ai}}{\sum_b RCA_{bi}} f_a = \frac{\sum_a f_a + \alpha_i (\sum_a f_a^2)}{N(1 + \alpha_i \sum_a \frac{f_a}{N})}$

Or at the first order in the elasticity

$$\bar{f}_i = \sum_a \frac{f_a}{N} - \alpha_i (\sum_a \frac{f_a}{N})^2 + \alpha_i \sum_a \frac{f_a^2}{N},$$

yielding a simple linear relationship between RFI and elasticity for small values of elasticities: $\bar{f}_i = (\text{simple}) \text{ average}(f_a) + \alpha_i \text{variance}(f_a)$.

Annex 3 The Exponential PCA algorithm



Annex 4 The "Discovery" Matrix (extraction)

The numbers in the matrix stands for the combination of RCA and LCA.

Type 1 RCA and LCA>1	Type 2 RCA <1 LCA>1	Type 3 RCA >1 LCA<1	Type 4 RCA <1 LCA<1

	ARG	AUS	AUT	BEL	BGR	BRA	CAN	CHL	CHN	CIV	COL	CRI	CZE	DEU	EGY	ESP	FRA	GBR	GEO	HUN	IDN	IND	ISR
0-Live animals	2	1	2	2	1	1	1	2	4	4	2	2	3	4	4	1	1	1	3	3	4	4	4
1-Meat	1	1	1	1	1	1	1	1	4	4	2	2	4	4	4	1	1	2	2	3	4	3	4
2-Dairy prods	1	1	1	1	1	2	4	2	4	4	2	1	4	3	3	2	1	2	4	4	4	4	4
3-Seafood	1	4	4	4	2	2	1	1	4	2	1	1	4	4	2	1	4	4	2	2	1	3	4
4-Cereals	1	1	4	2	1	1	1	2	4	2	2	2	4	4	1	2	3	4	2	3	2	1	4
5-Veg&fruit	1	4	2	1	1	1	1	1	4	1	1	1	4	4	1	1	4	4	1	4	2	2	3
6-Sugar&honey	1	2	4	2	1	1	2	2	4	2	1	1	4	4	1	2	4	4	2	3	2	1	4
7-Coffee,tea, …	2	2	4	1	1	1	4	2	4	1	1	1	4	4	1	2	4	4	2	4	1	1	2
8-Animal food	1	1	4	1	1	1	2	1	4	2	2	2	4	4	1	2	3	4	2	3	2	3	4
9-Eadible prods	1	3	1	1	2	2	1	4	4	3	1	1	4	4	3	1	1	1	3	4	1	4	4
11-Bevrages	1	1	1	2	2	2	2	1	4	4	2	2	4	2	4	1	1	1	3	4	4	4	4
12-Tobacco	1	2	4	1	1	1	2	2	4	1	2	2	3	3	1	2	2	4	2	4	1	1	2
21-Hides, skins	2	1	2	4	2	2	1	2	4	4	2	2	4	4	4	1	3	2	3	2	4	4	4
22-Oil seeds	1	2	4	4	1	1	1	2	4	2	2	2	4	4	1	4	4	4	4	3	2	3	4
23-Crude rubber	2	4	4	3	2	2	4	4	4	1	2	2	4	4	2	4	4	4	4	4	1	4	4
24-Cork&wood	2	1	3	4	1	1	1	1	4	1	2	2	3	4	2	2	4	4	1	4	1	4	4
25-Pulp	2	2	4	4	1	1	1	1	4	4	2	2	4	4	2	1	4	4	2	4	1	4	4
26-Textile fibers	1	1	4	2	2	1	2	2	4	1	2	2	4	4	1	2	2	4	2	4	1	1	2
27-Fertilizers	2	2	3	1	1	1	1	1	4	2	2	2	4	4	1	1	4	4	1	4	2	1	1
28-Metal	1	1	4	2	1	1	1	1	4	2	1	4	4	4	2	4	4	4	1	4	3	3	4
29-Crude anim&veg	2	4	4	1	2	1	4	3	4	2	1	1	4	4	1	1	4	2	3	4	2	1	1
41-Animal oils	2	1	2	4	2	2	1	1	4	4	2	2	4	4	3	1	3	1	1	4	4	4	4
42-Veg oils	1	4	4	4	1	1	3	4	4	1	1	1	4	4	2	3	4	4	4	4	1	4	4
43-Animal oil	1	4	4	3	4	3	4	4	4	1	2	1	4	4	1	4	4	4	4	4	1	4	2
51-Org chems	4	4	4	1	4	4	4	4	4	4	4	4	4	4	4	4	2	1	4	4	4	3	2
52- Inorg chems	4	4	4	2	1	2	1	3	4	2	2	4	4	4	1	4	3	4	2	4	2	2	1
53-Dyeing mat	4	4	2	1	4	4	4	4	4	4	3	4	4	1	4	1	2	1	4	4	4	1	2
54-med&pharm	4	4	1	1	4	4	4	4	4	4	4	4	4	1	4	1	1	1	4	3	4	3	1
55-Essential oils	3	4	2	1	3	4	4	4	4	3	3	2	4	1	3	1	1	1	4	4	4	4	2
56-Fertlizers	2	2	4	3	1	2	1	1	4	2	2	4	4	4	1	4	4	4	1	4	2	2	3
57-Explosives	4	4	4	1	4	2	4	4	4	4	1	4	4	4	3	3	4	4	4	4	4	4	2
58-Artif. resins	4	4	1	1	4	4	2	4	4	4	3	3	1	1	4	1	2	1	4	4	4	4	3
59-Chem mats	3	4	2	1	4	4	4	4	4	4	3	2	4	1	3	2	1	1	4	4	4	4	1
61-Leather	1	2	3	2	2	1	4	2	4	2	1	1	4	4	1	1	4	4	2	4	2	1	2
62-Rubber manuf	4	4	2	4	2	4	1	4	2	4	4	1	1	1	2	1	3	4	4	1	3	4	4
63- Cork manuf	2	4	1	3	1	3	1	3	3	1	2	2	1	4	2	1	4	4	4	1	1	2	4
64-Paper	2	4	1	3	2	2	1	2	4	4	1	1	2	1	1	1	1	2	2	2	3	4	4
65-Textile yarn	4	4	4	4	1	4	4	4	1	4	3	4	4	4	1	4	4	4	4	4	1	1	2
66- Mineral man	4	4	2	1	1	4	4	4	4	2	1	4	4	4	1	1	2	1	2	4	2	1	1
67-Iron&steel	4	4	1	3	1	3	2	4	4	4	1	4	3	4	1	1	2	4	1	4	4	1	2
68-Non-ferrous	2	1	1	3	1	2	1	1	4	2	2	4	4	4	1	2	4	4	2	4	1	1	4
69- Metal manuf	4	4	1	2	2	4	4	4	1	4	4	2	1	1	3	1	2	2	4	2	4	2	4
71-Power gen.	4	4	1	4	4	4	1	4	4	4	4	4	1	1	4	1	1	1	4	1	4	4	4
72-Specialized	4	4	1	2	4	4	4	4	4	4	4	4	2	1	4	4	2	1	4	4	4	4	4
73-Metalworking	4	4	1	2	4	4	4	4	4	4	4	4	1	1	4	4	2	2	4	4	4	4	4
74-Industrial	4	4	1	2	4	4	4	4	4	4	4	4	1	1	4	2	1	1	4	2	4	4	4
75-Office	4	4	4	4	4	4	4	4	1	4	4	1	1	4	4	4	4	4	4	2	4	4	4
76-Telecom	4	4	4	4	4	4	4	4	1	4	4	4	1	4	4	4	4	4	4	1	4	4	3
77-Electrical	4	4	4	4	4	4	4	4	1	4	4	3	1	4	4	4	4	4	4	1	4	4	3
78-Road vehicle	3	4	2	3	4	4	1	4	4	4	4	4	1	1	4	1	1	1	4	1	4	4	4
79-Other transp.	4	4	2	2	4	4	3	4	4	3	4	4	4	1	4	4	1	4	3	4	4	1	1
81-Sanitary	4	4	1	4	1	4	4	4	1	4	4	2	1	1	3	2	2	2	4	1	4	4	4
82-Furniture	4	4	1	4	1	4	3	4	1	4	4	2	1	4	3	2	4	4	4	1	3	4	4
83-Travel goods	4	4	4	4	4	4	4	4	1	4	4	4	4	4	4	4	3	4	4	4	4	1	2
84-Clothing	4	4	4	4	1	4	4	4	1	4	3	2	4	4	1	3	4	4	4	4	1	1	2
85-Footwear	4	4	4	3	1	3	4	4	1	4	4	2	2	4	2	1	4	4	4	4	1	1	2
87-Professional	4	4	4	4	4	4	4	4	3	4	4	3	2	1	4	4	2	1	4	1	4	4	3
88-Photographic	4	4	4	2	4	4	4	4	4	4	4	4	4	2	4	4	2	2	4	4	4	4	2
89-Manuf artcls	4	4	1	2	4	4	4	4	1	4	4	1	1	2	4	2	1	1	4	4	4	1	2

	ITA	JPN	KEN	KOR	LVA	MAR	MEX	MOZ	NIC	NLD	PAK	PER	PHL	POL	ROM	RUS	SGP	SWE	THA	TUN	TUR	TZA	USA	VNM
0-Live animals	4	4	2	4	1	4	3	4	1	1	4	2	4	1	3	4	4	2	2	4	4	2	2	4
1-Meat	4	4	2	4	2	4	4	4	1	1	4	2	4	1	4	4	4	4	3	4	4	2	1	4
2-Dairy prods	1	4	2	4	1	4	4	4	1	1	4	2	4	1	4	4	4	4	4	4	4	4	2	4
3-Seafood	4	4	1	4	1	1	2	1	1	2	3	1	3	1	2	1	4	1	1	1	2	1	4	1
4-Cereals	3	4	2	4	1	2	4	2	1	2	3	2	4	2	3	1	4	2	1	4	3	1	1	3
5-Veg&fruit	1	4	1	4	2	1	3	1	1	1	3	1	3	1	2	4	4	2	1	1	1	1	3	1
6-Sugar&honey	4	4	1	4	2	2	3	3	1	2	3	2	4	4	4	2	4	4	1	4	4	2	2	4
7-Coffee,tea, …	4	4	1	4	2	2	4	2	1	1	2	1	4	3	4	2	4	4	4	2	4	1	4	1
8-Animal food	4	4	2	4	1	3	4	4	1	1	4	1	4	4	4	4	4	4	1	4	4	1	1	4
9-Eadible prods	3	4	1	4	1	4	4	4	2	1	4	4	4	1	4	4	3	1	1	3	3	4	1	4
11-Bevrages	1	4	1	4	1	4	3	4	1	1	4	4	4	2	4	4	4	2	4	4	4	4	2	4
12-Tobacco	2	4	1	4	2	2	4	3	1	1	4	2	3	1	3	1	4	4	2	2	3	1	2	1
21-Hides, skins	4	4	2	4	1	4	4	4	1	2	4	2	4	1	4	4	4	2	4	4	4	1	1	4
22-Oil seeds	4	4	2	4	1	4	4	3	1	2	4	2	4	4	3	4	4	4	2	4	4	1	1	4
23-Crude rubber	4	4	2	3	2	2	4	4	4	2	2	4	2	4	2	1	4	4	1	2	4	2	4	1
24-Cork&wood	4	4	2	4	1	2	4	1	2	4	4	2	4	1	1	1	4	1	2	4	2	1	1	1
25-Pulp	4	4	4	4	2	1	2	2	2	4	4	2	4	2	2	1	4	1	2	4	4	2	1	4
26-Textile fibers	4	4	1	3	2	2	4	3	2	2	1	1	4	4	4	2	4	4	3	4	4	1	1	4
27-Fertilizers	2	4	1	4	2	1	4	2	4	4	1	3	4	4	2	1	4	4	4	1	1	1	4	2
28-Metal	4	4	2	4	1	3	4	2	2	4	4	1	4	4	3	1	4	3	4	4	4	1	3	4
29-Crude anim&veg	2	4	1	4	1	1	4	4	2	1	3	3	3	2	4	4	4	4	2	2	4	3	2	2
41-Animal oils	4	4	2	4	2	3	4	2	2	1	4	1	4	1	4	4	4	2	4	4	4	1	1	3
42-Veg oils	4	4	1	4	2	2	4	4	1	1	2	4	1	4	4	3	4	4	2	1	4	3	4	2
43-Animal oil	4	4	1	4	4	2	4	4	2	1	1	4	2	4	4	4	4	3	2	2	3	3	4	2
51-Org chems	4	3	4	1	4	4	4	4	4	1	4	4	2	4	4	3	1	4	4	4	4	4	1	4
52- Inorg chems	4	2	3	2	2	1	4	4	4	3	2	4	3	4	2	1	4	4	2	1	2	2	1	4
53-Dyeing mat	1	3	4	4	3	4	4	4	4	1	4	3	4	4	4	4	4	3	4	4	4	4	1	4
54-med&pharm	1	4	4	4	3	4	4	4	4	4	4	4	4	4	4	4	4	3	4	4	4	4	3	4
55-Essential oils	1	4	3	4	4	4	4	4	4	2	4	4	4	3	4	4	3	4	3	4	4	4	1	4
56-Fertlizers	4	4	2	4	2	1	4	2	4	3	2	2	4	4	1	1	4	4	2	1	2	1	4	2
57-Explosives	4	2	4	1	4	4	4	4	4	1	4	4	2	4	4	4	1	4	1	4	4	4	1	4
58-Artif. resins	1	1	4	1	4	4	4	4	4	4	4	4	4	3	4	4	4	2	4	4	1	4	1	4
59-Chem mats	2	3	4	4	4	4	4	4	4	1	4	4	4	4	4	4	4	4	4	4	4	4	1	4
61-Leather	1	4	1	4	2	1	4	2	1	2	1	4	4	4	3	2	4	4	3	1	2	2	4	1
62-Rubber manuf	2	1	4	2	2	2	2	4	4	2	4	4	4	1	1	4	4	2	1	2	1	4	2	1
63- Cork manuf	2	4	2	4	1	2	2	4	4	2	2	4	3	1	1	3	4	1	2	2	1	4	2	2
64-Paper	1	4	4	4	2	2	2	4	4	2	4	4	4	1	2	1	4	1	2	3	2	3	1	4
65-Textile yarn	1	4	4	3	3	2	4	4	4	4	1	4	4	4	1	4	4	4	4	1	1	3	4	1
66- Mineral man	1	4	1	4	4	2	4	4	4	4	1	4	4	4	4	1	4	4	3	2	1	3	4	4
67-Iron&steel	1	1	4	1	3	2	4	2	4	4	2	4	4	4	1	1	4	3	4	2	1	4	4	4
68-Non-ferrous	4	4	2	4	2	2	4	1	2	4	4	1	4	3	2	1	4	2	4	4	2	2	4	4
69- Metal manuf	1	4	4	4	1	4	2	4	4	2	4	4	4	1	1	4	4	1	4	1	1	4	4	2
71-Power gen.	1	1	4	4	2	4	1	4	4	4	4	4	4	1	2	3	4	1	4	4	4	4	1	4
72-Specialized	3	1	4	1	4	4	4	4	4	3	4	4	4	4	4	4	3	1	4	4	4	4	1	4
73-Metalworking	1	1	4	1	4	4	4	4	4	4	4	4	4	4	4	4	4	1	4	4	4	4	3	4
74-Industrial	1	1	4	4	4	4	1	4	4	4	4	4	4	2	1	4	4	1	3	4	2	4	1	4
75-Office	4	4	4	4	4	4	1	4	4	1	4	4	1	4	4	4	1	4	1	2	4	4	4	2
76-Telecom	4	2	4	1	4	4	1	4	4	3	4	4	2	1	1	4	2	1	2	1	4	4	4	1
77-Electrical	4	1	4	1	4	3	1	4	4	2	4	4	1	4	3	4	1	4	1	3	4	4	2	4
78-Road vehicle	2	1	4	1	2	4	1	4	4	4	4	4	4	1	1	4	4	1	3	4	3	4	1	4
79-Other transp.	2	1	4	1	4	4	4	4	4	4	4	4	4	4	3	2	4	4	4	4	2	4	4	4
81-Sanitary	1	4	4	4	1	4	1	4	4	2	4	4	4	1	1	4	4	1	4	2	1	4	4	2
82-Furniture	1	4	4	4	1	2	1	4	4	2	4	4	4	1	1	4	4	1	2	2	1	4	4	1
83-Travel goods	1	4	4	4	4	4	4	4	4	4	2	4	4	4	4	4	4	4	4	1	2	4	4	1
84-Clothing	1	4	1	4	4	1	4	4	4	4	1	3	3	2	1	4	4	4	4	1	1	4	4	1
85-Footwear	1	4	1	4	2	1	4	4	4	4	2	4	4	2	1	4	4	4	4	1	2	4	4	1
87-Professional	4	1	4	1	4	4	3	4	4	4	4	4	2	4	4	4	1	4	2	4	4	4	1	4
88-Photographic	3	1	4	1	4	4	4	4	4	4	4	4	3	4	4	4	1	4	3	4	4	4	2	4
89-Manuf artcls	1	4	4	4	4	4	4	4	4	1	3	4	2	2	4	4	3	4	1	2	4	4	1	3

Annex 5 Latent comparative advantage indicators by country

country	Theil actual	KL Distance	Type 1#	Type 2#	country	Theil actual	KL Distance	Type 1#	Type 2#	country	Theil actual	KL Distance	Type 1#	Type 2#
ALB	2.57	0.68	11	10	FRA	3.47	0.12	14	14	NER	1.81	0.79	7	11
ARE	2.23	0.66	4	8	GBR	3.35	0.07	18	9	NGA	2.80	1.05	15	11
ARG	3.16	0.40	17	13	GEO	2.52	0.82	8	15	NIC	2.62	0.66	15	9
ARM	2.09	0.54	8	15	GHA	1.43	1.20	2	7	NLD	3.59	0.11	19	16
AUS	2.52	0.18	13	9	GRC	3.49	0.29	20	21	NOR	2.95	0.56	11	17
AUT	3.44	0.11	18	12	GTM	3.12	0.48	16	18	NPL	2.59	0.60	10	14
AZE	2.79	1.03	14	13	GUY	2.05	0.87	6	12	NZL	2.96	0.51	15	7
BEL	3.40	0.12	17	16	HKG	2.35	0.91	5	8	OMN	2.78	0.75	9	17
BEN	2.24	1.39	10	20	HRV	3.47	0.39	17	15	PAK	2.20	0.65	7	10
BFA	1.18	0.99	3	4	HUN	3.03	0.18	9	6	PAN	2.44	0.50	15	11
BGR	3.43	0.31	26	13	IDN	3.36	0.34	15	12	PER	2.24	0.51	9	11
BHR	1.91	0.52	3	21	IND	3.35	0.12	19	5	PHL	2.28	0.42	3	7
BHS	1.46	1.75	3	17	IRL	2.53	0.16	11	4	POL	3.44	0.09	19	8
BIH	3.28	0.43	18	16	IRN	3.06	0.47	16	12	PRT	3.52	0.16	21	13
BLR	3.25	0.70	17	22	ISL	1.67	0.69	5	15	PRY	2.17	0.64	11	12
BLZ	1.23	1.08	5	22	ISR	2.78	0.43	7	17	ROM	3.27	0.22	14	9
BOL	2.13	0.42	13	12	ITA	3.51	0.09	22	9	RUS	3.01	0.20	14	7
BRA	3.29	0.27	17	14	JAM	1.76	0.57	8	16	RWA	1.58	0.69	8	16
BRB	2.73	0.66	9	16	JOR	2.87	1.04	8	22	SAU	2.43	0.52	5	12
BTN	1.95	0.75	6	7	JPN	2.95	0.07	12	4	SEN	2.97	0.91	13	15
BWA	1.34	0.64	3	7	KAZ	2.15	0.34	9	6	SGP	2.60	0.09	6	1
CAN	3.46	0.08	22	7	KEN	3.01	0.50	16	15	SLV	2.62	0.65	9	12
CHE	2.92	0.22	11	3	KGZ	1.83	1.34	3	10	SUR	2.62	1.14	10	24
CHL	2.22	0.17	14	11	KHM	1.28	1.00	2	10	SVK	2.99	0.13	11	3
CHN	3.10	0.03	11	1	KOR	2.93	0.05	12	2	SVN	3.26	0.17	15	9
CIV	2.04	1.22	9	14	LBN	3.06	0.75	12	21	SWE	3.32	0.18	15	10
CMR	2.19	1.10	8	18	LKA	2.19	0.65	10	10	SYR	3.11	0.66	16	20
COG	0.85	0.95	1	5	LTU	3.63	0.26	26	13	TGO	2.54	0.88	9	11
COL	3.23	0.72	13	21	LUX	3.14	0.34	12	11	THA	3.35	0.18	12	15
CRI	2.96	0.55	14	22	LVA	3.43	0.49	16	21	TTO	1.88	0.94	5	9
CYP	2.60	0.73	10	14	MAR	2.82	0.74	10	18	TUN	2.91	0.37	13	13
CZE	3.17	0.09	14	4	MDA	2.86	0.80	15	15	TUR	3.21	0.19	12	12
DEU	3.33	0.02	16	3	MDG	2.33	0.72	12	10	TZA	2.66	0.60	15	8
DNK	3.55	0.18	25	9	MEX	2.93	0.10	9	6	UGA	2.66	0.63	15	12
DOM	3.10	0.67	14	17	MLI	0.82	0.76	1	1	UKR	2.92	0.47	15	10
DZA	2.48	1.03	10	12	MLT	1.82	0.71	2	6	USA	3.55	0.04	25	10
ECU	2.52	0.43	13	17	MMR	1.90	1.06	8	17	VEN	2.56	0.24	8	8
EGY	3.35	0.50	20	10	MOZ	1.36	0.63	4	10	VNM	3.17	0.22	14	9
ESP	3.42	0.12	25	15	MRT	1.20	0.83	5	8	YEM	2.42	0.64	13	13
EST	3.52	0.26	19	18	MUS	1.91	1.16	5	16	ZAF	2.91	0.14	15	10
ETH	2.16	1.09	8	21	MWI	1.86	1.93	10	18	ZMB	1.20	0.35	2	8
FIN	3.24	0.42	17	10	MYS	2.97	0.13	8	8	ZWE	2.38	1.19	10	18
FJI	2.68	1.03	11	17	NAM	2.57	0.78	11	18					