HEC Montréal

The Proportionality Assumption in Factor Content Computations – Measuring bias using Genetic Algorithms

by

Peter Puszko

Decision Sciences

Thesis submitted in fulfilment of the requirements of

the degree of Master of Science Business Analytics

September 2020

ABSTRACT

Certain policies and areas of research require production inputs to be traced all the way up the production chain and assigned to consumers to estimate their impact. The contexts range from factor content of trade, carbon dioxide (CO_2) content of trade, trade in value added, R&D content of intermediates, wage structures, and material offshoring effects – all of which need domestic and imported input-output data to be computed. Unfortunately, the granularity of data required for accurate computations is not available for all countries as it requires a tremendous collection effort. As a result, researchers use a proportionality assumption where import trade shares are identical in all end uses for each sector.

Winkler and Milberg (2009), Puzello (2012), and Feenstra and Jensen (2012) have demonstrated the assumption's shortcomings on factor content of trade by using the few countries that collect the necessary data, yet, those countries are too few to understand the potential effects of the assumption in a globalized world.

This paper broadens the scope by measuring the limitations of the proportionality assumption in factor content of trade computations for 140 countries through simulations, though the methodology employed can be applied to all other contexts that make use of input-output linkages and trade amongst countries. In the framework of factor content of trade, the maximum potential bias found averages 208% through Monte-Carlo simulations and 330% through genetic algorithms for all countries across all factors, confirming that restrictions on data are not resolved with the proportionality assumption.

TABLE OF CONTENTS

LIST OF TABLES, FIGURES, AND ALGORITHMS

ACKNOWLEDGEMENTS

I would like to thank my thesis supervisor, Justin Caron, for helping me work through a plethora of initial topics and filling gaps in my understanding of economics. My mother for making sure I never go hungry. My father and brother for always finding distractions whenever I needed them the least. My friends for proofreading through countless review cycles and keeping me sane.

INTRODUCTION

As globalisation expands and trade increases, economies become more intertwined as they use imports not only to satisfy consumption, but largely to satisfy production. Trade, mainly consisting of intermediates, makes itself complex to understand and analyze. Goods, rarely crossing the boarder only once, are found zigzagging from country to country as they go through intricate supply chains bringing new elements of trade to the forefront. Trade is no longer just exports minus imports; it is trade in value added (Johnson and Noguera, 2012; Nagengast and Stehrer, 2014; Johnson, 2018). Intermediates are not just commodities or raw materials, so the notion of R&D content of intermediates takes shape (Nishioka and Ripoll, 2012). As trade brings new possibilities to some countries, others start to observe impacts on their wage structures (Feenstra and Hanson, 1996, 1999; Hummels et al., 2001) and material offshoring (Feenstra and Jensen, 2012). Considering the back and forth of supply chains not only raises questions on consumers carbon footprint, but also on carbon dioxide (CO2) content of trade (Peters and Hertwich, 2008; Caron et al., 2017).

To identify the linkages in trade, Leontief's input-output technique provides a framework to represent quantities of intermediary products needed by unit of a commodity. Coefficients are calculated in input-output matrices, which track flows between sector and country pairs. Input-output matrices consist of domestic and imported input-output data. As the former are broadly available from national statistical agencies, the contrary applies to the latter, where very few countries report – too few to extrapolate to the rest of the world. The standard imputation approach applied is the proportionality assumption, where imported products are distributed proportionally to all destination sectors and final demand according to domestic demand by using bilateral trade vectors and the imported input-output matrices. For example, if 30% of wood products in the United States are imported from Canada, it is inferred that 30% of the pulp used to make Dunder Mifflin's paper comes from Canadian wood[1].

[1] The application of the proportionality assumption in estimating factor content of trade is limited to sectors. The application to industry is an extension of the proportionality assumption for illustrative purposes.

Empirically, the standard proportionality assumption is limiting as it does not consider differences in patterns of world trade present in country-sector pairs that might be explained by 'North-North', 'South-South', or North-South' trade. Some industries, like aircraft, may also solely depend on domestic inputs, as Trefler (1995) describes under home bias.

In cases where more data is available, such as the Asian Input-Output (AIO) tables, which provide more precise assessments of input-output structures for China, Indonesia, Korea, Japan, Malaysia, Philippines, Singapore, Taiwan, Thailand and the United States, Puzello (2012) showed that using the proportionality assumption underrates the relative use of imported inputs in the context of factor content of trade, particularly in significant sectors, implying that a country's comparative advantage depends on end-use and not factor endowments. Winkler and Milberg (2009) performed an analysis on data from Germany and found that in the case of services offshoring, a direct measure and the proportionality assumption had coefficients of opposite signs in many cases. A similar study on the USA by Feenstra and Jensen (2012) compared firm-level imports to the proportionality assumption and found a correlation of 0.68.

Assuming proportionality results in many biases by understating uses of imported inputs (Puzello, 2012) and overstating domestic factors in countries techniques (Zhu and Trefler, 2005). With current tendencies in offshoring and their impacts on labour, the National Research Council (2006) referred to the proportionality assumption issue as being a considerable restriction of present data collection and analysis. Yet, regardless of its shortcomings, the proportionality assumption continues to be used by academic researchers and policymakers without appropriate consideration for the effects the assumption holds on results.

Previous research has centered around other direct measures to identify the bias and implications of using the standard proportionality assumption (Winkler and Milberg, 2009; Puzello, 2012; Feenstra and Jensen, 2012). The approach used in this paper is different in that instead of calculating the true bias for the few countries for which data is available, it estimates the potential bias in the context of factor content of trade for all countries following the global supply chain of all sectors. The potential bias could add confidence to the current method or pinpoint areas where the proportionality assumption is an issue.

The methodology employed in this paper consists in comparing the proportionality assumption to three different approaches[2]: a random local maximum that maximizes trade flows with respect to volume, Monte-Carlo simulations that generate solutions at random, and genetic algorithms that generate improved results by transferring from one solution to the next.

The random local maximum uses trade volumes between country-sector pairs to determine trade flows. Larger trade relationships are emphasized to simulate industry competitors obtaining inputs from the same source. Such arrangements yield an average bias of 100% for all countries across all factors compared to the proportionality assumption.

Monte-Carlo simulations follow a maximization framework which aims to find the maximum potential bias through a multi-start method. The largest bias found averages 208% for all countries across all factors compared to the proportionality assumption.

Genetic algorithms, heuristics formed using neo-Darwinian theory of evolution, are developed in line with the maximization framework, but differ from Monte-Carlo simulations as good results from one iteration are used to build the next, which allows to find a bias averaging 330% for all countries across all factors compared to the proportionality assumption.

[2] Python code for measuring bias of the proportionality assumption in factor content computations can be downloaded from a GitHub repository, which is available at https://github.com/pszko/FCT-proportionality_bias.

TRACING BACK FACTOR CONTENT OF TRADE

Competitive advantage enables the understanding of trade patterns where exporters have superior abilities (or lower costs) in relation to importers. Such relations allow to model international economies with the help of Ricardian and Heckscher-Ohlin theories, and in turn, those models provide insights to researchers and policy makers. However, when put to the test, these trade theories showcase limitations which put in question the models extracted from them.

The Ricardian theory can be used to isolate explanatory variables that interpret trade flows in addition to explaining why engaging in trade is mutually beneficial. Following Ricardo, the pattern of trade is governed by input requirements ratios. With more than two commodities, the makeup of imports and exports is revealed by international differences in production functions – comparative factor productivities and demand. However, difficulties arise when applying Ricardo's one-factor model to the actual multi-factor world. Alternatives focus on the role of comparative differences in production functions or factor productivities in terms of a chosen factor[3].

On the other hand, the Heckscher-Ohlin theory is neither limited to one factor, unlike the Ricardian model, nor does it make factor supply peripheral in establishing patterns of trade with the assumption of constant returns to scale. The standard Heckscher-Ohlin model assumes two factors and presumes that comparative advantage is determined by international differences in factor endowments, where a country's exports use intensively the country's abundant factors.

The model yields different predictions if production functions for specific industries differ by country. These could be balanced by consumption preferences, which can have an offset in relative prices elsewhere. Another exclusion appears in production functions that are identical but are intensive in different factors, which effectively has the same impact as different

[3] With a_1 and a_2 representing the output-factor ratios for country A in activity 1 and 2 and b_1 and b_2 for country B, with a $a_1/a_2 > b_1/b_2$ relation, country A will export commodity 1 and import commodity 2. Also written as $a_1/b_1 > a_2/b_2$, asserts the comparative factor productivities (Bhagwati, 1964).

production functions. Both scenarios refute the Heckscher-Ohlin proposition, for which theorists have had to establish a set of conditions to keep it true[4]. With these added conditions, commodities can be ranked by comparative advantage with the introduction of demand just as the Ricardian model, in terms of a chosen factor.

Bhagwati (1964), brought to surface the lack of scalability of these theorems to the real world. By considering factor-intensities of products, with three factors, there are three different ways of arranging one production function in relation to another depending on which factor is used as a reference input. For example, using land as a refence input, land can be more or less capital intensive, more or less labour intensive, and more or less labour intensive with reference to capital. The arrangements get more complex with added factors, and their interpretability becomes even more difficult.

Vanek (1968) filled the void with *The Factor Proportions Theory* for cases with N factors, in which countries are net exporters of services relatively intensive in factors that they are endowed with. Shifting the reference from products to amounts of factor-services embodied in goods traded allows for scalability of the Heckscher-Ohlin theorem. A model of international price equalization has led Vanek to coining the concept of factor content of trade, where both goods and factor prices are equalized internationally under conditions of competition in goods and factor markets, free international arbitrage, common constant returns to scale technologies, and adequate restrictions on the distribution of world endowments (Davis and Weinstein, 2003). Within a general equilibrium framework, the idea of factor content of trade supports the evaluations of interactions between endowments, production, absorption and trade, helping to confirm trade theories and to address national policies. The Ricardian theory is less effective in this perspective as it is difficult to reconcile industrial mixes without the implication of factor supply.

---

[4] Conditions are dependant on the definition of factor abundance – either physical or price abundance. For physical abundance, conditions are: international identity of production functions, non-reversibility of factor-intensities, constant returns to scale and diminishing returns along isoquants in each production function, and identity of the consumption pattern between countries at each relevant commodity price ratio. For price abundance, the condition of identity of the consumption patterns is relaxed, allowing for an easier use of the definition; however, it comes with explanatory limitations.

From Vanek's proposition, country factor content of trade can be predicted using endowments and shares of world consumption as follows:

$$F_i = V_i - s_i V_w \tag{1}$$

where F_i is country i's factor content of trade vector, V_i is country i's endowment vector, s_i is country i's portion of world consumption, and V_w is the world endowment vector. This expression of the problem broadens the logic behind the Heckscher-Ohlin theorem in which factor content of trade may be determinate even though the patterns of trade may be indeterminate (Davis and Weinstein, 2003).

With endowment data not being available globally and patterns of trade largely consisting of trade in intermediates, factor content of trade takes on an enlarged scope to incorporate intermediate inputs and outputs in addition to final demand. Leontief's input-output technique facilitates the exercise but requires data to do so.

COMPUTING FACTOR CONTENT OF TRADE

Vanek presented factor content of trade as a direct relation between country i's endowment and country i's portion of world consumption considering the world endowment. Reimer (2006) and Trefler and Zhu (2010) were able to substantiate the direct relation by bringing out algorithms with a relevant definition of F based on Deardorff's (1982) 'actual' factor content of trade, as well as correct assumptions used to derive F related to technology. Earlier studies on factor content of trade had used the same technology matrix for all countries or used the producers' technology vector for both inputs and outputs. These simplifications have made the calculation of factor content of trade easier at the cost of disregarding the impact of indirect factor inputs, making the Vanek proposition (equation 1) violate the linear dependency of F_i, V_i, V_w.

Reimer (2006) and Trefler and Zhu (2010) calculate country i's factor content of trade using:

$$F_i = D\,(I - B)^{-1}\,T_i \qquad (2)$$

Where D is the matrix of direct factor unit requirements, B is the input-output matrix, $(I - B)^{-1}$ is the Leontief inverse, and T_i is country i's net trade, which is defined as the difference between exports and imports and represents direct and indirect factor inputs by $X_i - M_i$.

The Leontief inverse allows to compute utilization of outputs as inputs. As sectors are co-dependent where they need more of themselves and other sectors, production can be broken-down into three parts:

1. Proportion of a sector used to produce more of that sector
2. Proportion of other sectors used to produce more of that sector
3. Proportion of a sector used for final demand and trade

The three parts can be transformed into a system of equations consisting of a technology matrix (B) based on required resources for production and of demand and trade vectors based on final demand and trade constraints.

The complexity of the input-output matrix B, which is a square matrix, grows exponentially as countries and industries are added. In the case where B includes 3 sectors and 3 countries, B takes the following shape:

$$B = \begin{bmatrix} b_{1111} & b_{1112} & b_{1113} & b_{1121} & b_{1122} & \cdots & b_{1133} \\ b_{1211} & b_{1212} & b_{1213} & b_{1221} & b_{1222} & \cdots & b_{1233} \\ b_{1311} & b_{1312} & b_{1313} & b_{1321} & b_{1322} & \cdots & b_{1333} \\ b_{2111} & b_{2112} & b_{2113} & b_{2121} & b_{2122} & \cdots & b_{2133} \\ b_{2211} & b_{2212} & b_{2213} & b_{2221} & b_{2222} & \cdots & b_{2233} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{3311} & b_{3312} & b_{3313} & b_{3321} & b_{3322} & \cdots & b_{3333} \end{bmatrix} = \left(b_{igjh}\right)$$

indices i, g, j, h represent the output country, output sector, input country and input sector. The input-output matrix B can be viewed as an aggregation of domestic and imported input-output matrices. For domestic matrices, where $i = j$, values are broadly available from national statistical agencies that report such data. Using the U.S. as an example, it takes the following form:

$$B_{USA} = \begin{bmatrix} b_{USA1USA1} & b_{USA1USA2} & \cdots & b_{USA1USAh} \\ b_{USA2USA1} & b_{USA2USA2} & \cdots & b_{USA2USAh} \\ \vdots & \vdots & \ddots & \vdots \\ b_{USAgUSA1} & b_{USAgUSA2} & \cdots & b_{USAgUSAh} \end{bmatrix} \ \forall\, i, j = USA$$

For imported matrices, where $i \neq j$, values are scarcely available, with certain exceptions like the AIO tables. To calculate the factor content of trade or to perform the multiple other computations that require the input-output matrix as presented earlier, values need to be assumed. The required data for these assumptions is available through bilateral trade vectors and the imported input-output matrix, i.e. amounts of imports by sector purchased by firms.

This data forms constraints on the imported input-output matrices for every country, which can be represented mathematically as follows:

$$B_{USA} = \begin{array}{c} \begin{bmatrix} b_{11USA1} & b_{11USA2} & \cdots & b_{11USAw} \\ b_{12USA1} & b_{12USA2} & \cdots & b_{12USAw} \\ \vdots & \vdots & \ddots & \vdots \\ b_{1yUSA1} & b_{1yUSA2} & \cdots & b_{1yUSAw} \end{bmatrix} \\ \Sigma b_{1gUSA1} \quad \Sigma b_{1gUSA2} \quad \dots \quad \Sigma b_{1gUSAw} \end{array} \begin{array}{c} \Sigma b_{11USAh} \\ \Sigma b_{12USAh} \\ \vdots \\ \Sigma b_{1yUSAh} \\ \end{array} \quad \begin{array}{c} \forall\, j = USA \; and \; i = 1 \\ Where \; g = 1, \dots, y \\ and \; h = 1, \dots, w \end{array}$$ [5]

Where columns are subject to the sums of all imported product g's from all country i's for a sector in $j = US$ and rows are subject to bilateral trade between country i and the US and all sector h's.

As imported input-output data in the technology matrix is not widely reported, a standard proportionality assumption is used as a workaround, where imported products, per sector, are distributed proportionally to destination sectors (final or intermediate) according to domestic demand. Mathematically, this is defined by:

$$b_{igjh} \; = \; \frac{m_{gjh}}{\sum_{h \in H} m_{gjh}} \times k_{igj} \quad for \; i \neq j \tag{3}$$

where m represents entries in the imported input-output matrix, k represents the bilateral trade flows, and indices i, g j, h represent the output country, output sector, input country, and input sector.

Although this method satisfies constraints and its application is straight-forward, there is no well-founded reason to believe that a model based on that assumption is an accurate reflection of reality. The possible values that b_{igjh} can take in the technology matrix while satisfying constraints of imported products and bilateral trade are numerous, which raises the question of how biased the standard proportionality assumption can be.

[5] Where y is the last output sector and w is the last input sector.

EARLIER RENDITIONS OF FACTOR CONTENT OF TRADE

There are multiple definitions of the factor content of trade that have been previously used, but none represent it as the amount of factors used worldwide to produce a country's trade flows[6] (Trefler and Zhu, 2010). Trefler (1993, 1995), Davis and Weinstein (2001), Conway (2002), and Debaere (2003) have all used variations of:

$$\bar{A}_i = D_i \, (I - \bar{B}_i)^{-1} \, (X_i - M_i) \tag{2}$$

where $\bar{B}_i$, the national input-output matrix, which includes national and international sources, assumes $\bar{B}_i = \bar{B}_{USA}$, denoting the same choice-of-technique as the USA for all country i's. Moreover, $D_i = \Lambda_i^{-1} D_{USA}$, where Λ_i is a diagonal matrix in which entries hold values of the productivity of factors in country i in relation to the United States. Unfortunately, this approach does not allow for international technology differences.

One definition that does allow for international choice-of-technique differences was first introduced by Helpman and Krugman (1985) and later presented by Davis and Weinstein (2001):

$$F_i^{DW} = \bar{A}_i X_i - \Sigma_{j \neq i} \bar{A}_j M_{ij} \tag{3}$$

It uses country j's technology ($\bar{A}_j$) to evaluate country j's output (M_{ij}). Trefler and Zhu (2010) demonstrated how without further restrictions on the input-output matrix, there are no intermediates traded between country i and country j – to which Feenstra and Hanson (1996, 1999), Hummels et al. (2001) and Yi (2003) would tend to refute given their research in global vertical production networks, making this definition of factor content of trade non Vanek-relevant as F_i is not defined as being equal to $V_i - s_i V_w$.

[6] Previous definitions used $\bar{A}_i$ instead of A_i.

Trefler and Zhu (2000) presented another definition that allows for international choice-of-technique differences:

$$F_i^T = \bar{A}_i X_i^c - \Sigma_{j \neq i} \bar{A}_j M_{ij}^c + \bar{A}_i \left(X_i^y - M_i^y \right) - s_i \Sigma_j \bar{A}_j (X_j^y - M_j^y) \qquad (4)$$

It considers both country i's exports of consumption goods under X_i^c and country i's imports of consumption goods produced in country j under M_{ij}^c, as well as country i's exports of intermediate inputs under X_i^y and country i's imports of intermediate inputs under M_i^y. Under the Vanek null ($F_i^T = V_i - s_i V_w$), it is a Vanek-relevant definition, but it is ambiguous how to interpret F_i^T if it does not equate to $V_i - s_i V_w$, rendering it not particularly economically meaningful.

In this paper, the focus is on the impact the proportional imputation of the input-output matrix B has on the calculation of factor content of trade. The criteria of international choice-of-technique variation, Vanek-relevant F_i, and economical meaningfulness are satisfied with the use of datasets that account for technological differences between countries and the use of the formula presented by Reimer (2006) and Trefler and Zhu (2010) (equation 2).

DATA

This paper uses the GTAP 9 database gathered by the Global Trade Analysis Project at Purdue University. National input-output, trade, macroeconomic, and protection data tables assembled from the World Bank and IMF macroeconomic and Balance of Payments statistics, United Nations Commodity Trade Statistics (Comtrade) Database, and national statistical agencies make up the GTAP database. In the GTAP 9 release, data is available for 244 countries broken down into 134 individual countries and 6 aggregated regions representing the remaining countries[7]. For each country, data on 57 sectors[8] which includes 18 agriculture and natural resources sectors, 24 manufacturing sectors, and 15 services sectors, is available.

Computing factor content of trade requires the domestic input-output (provided by GTAP as the VDFM parameter), bilateral trade flows (provided by GTAP as the VXMD parameter), imported input-output (provided by GTAP as the VIFM parameter), and factor demand matrices (provided by GTAP as the VFM parameter). After separating final demand, including consumption, government purchases, and investments from VDFM, all 4 datasets are transformed to obtain:

1. $VDFM_{igjh}$ – domestic input-output matrices[9]
2. $VXMD_{igj}$ – bilateral trade flow vectors[10]
3. $VIFM_{gjh}$ – imported input-output vectors[11]
4. VFM_{fjh} – factor demand matrices[12]

---

[7] See Appendix A for a detailed list of countries and regions

[8] See Appendix B for a detailed list of sectors

[9] VDFM is transformed from a 3-dimensional array table with dimensions of output sector, input sector, and output/input country, into $VDFM_{igjh}$, a collection of domestic input-output matrices for each country.

[10] VXMD is transformed from a 3-dimensional array table with dimensions of output sector, output country, and input country, into $VXMD_{igj}$, a collection of bilateral trade flow vectors for each country.

[11] VIFM is transformed from a 2-dimensional array table with dimensions of output sector, input sector, and input country, into $VIFM_{gjh}$, a collection of import input-output vectors for each country.

[12] VFM is transformed from a 3-dimensional array table with dimensions of factor, sector, and country, into VFM_{fjh}, a collection of factor demand matrices for each country.

These matrices and vectors can then be used to calculate $F_i = D\ (I - B)^{-1}\ T_i$ where D is the aggregate of factor demand matrices VFM_{fjh}, T_i is calculated from bilateral trade vectors VXMD_{igj} by subtracting imports from exports $(X_i - M_i)$, and B is the aggregate of domestic input-output matrices VDFM_{igjh} and imputed import matrices.

POTENTIAL BIAS OF THE STANDARD PROPORTIONALITY ASSUMPTION

The proportionality assumption, having been outlined as a potentially considerable restriction of present data collection and analysis, is evaluated here from a different angle. Instead of looking at a few countries where relevant data has been collected, the potential bias of the entire global supply chain for all countries and all factors is analyzed through simulations and algorithms. As data required for the construction of a technology matrix is available, b_{igjh} entries can be simulated while satisfying the applicable constraints. To do so, concepts from operational research, a field of study that has helped the Allies win the Second World War, are borrowed.

Given an objective function, a simple optimization problem would be resolved by finding its maximum at the intersection of the applicable constraints. The construction of matrix B, however, represents a combinatorial optimization problem (COP) where intersections are interdependent and dynamic as any updated b_{igjh} entry in the respective imported input-output matrix modifies the applicable constraints for the following imputations:

$$B_{USA} = \begin{bmatrix} b_{11USA1} & b_{11USA2} & \cdots & b_{11USAh} \\ b_{12USA1} & b_{12USA2} & \cdots & b_{12USAh} \\ \vdots & \vdots & \ddots & \vdots \\ b_{1gUSA1} & b_{1gUSA2} & \cdots & b_{1gUSAh} \end{bmatrix} \forall\, j = USA \; and \; i = 1$$

for instance, b_{11USA1} is constrained by the sum of entries in the first column and first row. Under the maximization framework presented in *Brief algorithm 1*, the order in which b_{igjh} is imputed has an impact on the values in the same row and column since predecessors alter constraints dynamically. This is because entries of a given row must all be less than the sum of the entries of the given row and the entries of a given column must all be less than the sum of the entries of the given column. Therefore, they have a constraint in common. This is mathematically expressed by:

$$b_{igjh} = min[\sum_{ig \in IG} \overline{b_{\imath g \jmath h}} - \sum_{ig \in IG} c_{ig} \, , \sum_{jh \in JH} \overline{b_{\imath g \jmath h}} - \sum_{jh \in JH} r_{jh} \;] \tag{7}$$

where $c \in \; C_{igjh}$ is the list of immediate predecessors of b_{igjh} in column jh and $r \in \; R_{igjh}$ is the list of immediate predecessors of b_{igjh} in row ig where rows and columns are specified for every imported input-output matrix of every country.

As Matrix B is a collection of 140 domestic matrices, where $i = j$, and 19,460 imported matrices, where $i \neq j$, the collection can be visualized as:

$$B = \underbrace{\left. \begin{bmatrix} \overbrace{\begin{bmatrix} b_{1;1;1;1} & \cdots & b_{1;1;1;57} \\ \vdots & \ddots & \vdots \\ b_{1;57;1;1} & \cdots & b_{1;57;1;57} \end{bmatrix}}^{57} & \cdots & \begin{bmatrix} b_{1;1;140;1} & \cdots & b_{1;1;140;57} \\ \vdots & \ddots & \vdots \\ b_{1;57;140;1} & \cdots & b_{1;57;140;57} \end{bmatrix} \\ \vdots & \ddots & \vdots \\ \begin{bmatrix} b_{140;1;1;1} & \cdots & b_{140;1;1;57} \\ \vdots & \ddots & \vdots \\ b_{140;57;1;1} & \cdots & b_{140;57;1;57} \end{bmatrix} & \cdots & \begin{bmatrix} b_{140;1;140;1} & \cdots & b_{140;1;140;57} \\ \vdots & \ddots & \vdots \\ b_{140;57;140;1} & \cdots & b_{140;57;140;57} \end{bmatrix} \end{bmatrix} \right\} 7980}_{7980}$$

where domestic matrices are on the diagonal of matrix B and matrix B contains all possible combinations of 140 countries and 57 sectors, resulting in a matrix of size 7980 by 7980. Constraints of imported input-outputs are unique to every output sector, g, as well as every input country and input sector, jh. Bilateral trade constraints, on the other hand, are unique to every output country and output sector, ig, as well as every input country, j. Following the maximization framework, every iteration equates either a row or column constraint, implying that one can obtain at most $g + j - 1$ non-zero values in every imported matrix[13]. The brief pseudo code to assign non-zero values reads:

[13] See equation 7 for mathematical representation.

BRIEF ALGORITHM 1: MAXIMIZATION FRAMEWORK FOR IMPORTED MATRIX[14]

```
B_matrix ← empty matrix of size ig × jh
WHILE bilateral trade constraints AND imported input-output constraints > 0
   list_igjh ← list of all possible permutations of i, g, j, h
   i, g, j, h ← random selection from values in list
   IF bilateral constraints[igj] < imported input-output [gjh]
      B_matrix[igjh] += bilateral trade constraint [igj]
      bilateral trade constraint [igj] ← 0
      imported input-output constraint [gjh] −= bilateral trade constraint [igj]
      list_igjh −= i, g, j ∀ h
    ELSE
      B_matrix[igjh] += imported input-output [gjh]
      imported input-output constraint [gjh] ← 0
      bilateral trade constraint [igj] −= imported input-output constraint [gjh]
      list_igjh −= g, j, h ∀ i
```

With the GTAP 9 dataset, this implies choosing 113 out of 3249 values, giving $\left(\frac{3249!}{(3249-113)!}\right) = 9.373 \ldots \times 10^{395}$ possible permutations for every imported matrix, considering the order in which b_{igjh}s are imputed. Each permutation represents a local maximum, and the only way to guarantee a global maximum is by calculating every possible permutation, computationally unfeasible due to the complexity. As such, to simplify the search of a maximum potential bias, genetic algorithms are developed in this paper to find a good solution without trying every permutation.

The bias in this paper is assessed by comparing the mean absolute percent error (MAPE) of the factor content of trade calculated using a proportionally imputed matrix B and a version of matrix B imputed through a maximization framework:

$$MAPE = \left(\frac{1}{n}\sum_{m=1}^{n}\left|\frac{fct_p - fct_m}{fct_p}\right|\right) \times 100 \qquad (8)$$

[14] See Appendix C for detailed pseudo code of ALGORITHM 1: MAXIMIZATION FRAMEWORK FOR IMPORTED MATRIX

where fct_p is the factor content of trade calculated using the proportionality assumption and fct_m is the factor content of trade calculated using the maximization framework.

RANDOM LOCAL MAXIMUM

As a first step, a random local maximum is calculated by arranging the order in which entries b_{igjh} are imputed, starting with entries representing the largest trade relations based on volume obtained from their respective constraints of imported products and bilateral trade. In the case of paper pulp, if Canada is the largest supplier of wood products to the USA and the paper products industry in the USA is the largest importer of wood products, then the random local maximum imputes the corresponding entry in the B matrix first. The random local maximum simulates a likely representation of trade flows where competitors obtain inputs from the same source. Random ordering of entries with the same maximum potential volume creates the randomness in this simulation.

The random local maximum serves two main purposes: analyzing the impact of a naïve method of maximization on different factors available in the GTAP dataset, which is used to benchmark following efforts, and highlighting natural areas of interest based on trade relations[15]. The naïve method renders the results presented in *Table 1* where the naïve method is compared to the proportionality assumption by measuring the MAPE for all 140 regions.

TABLE 1: RANDOM LOCAL MAXIMUM MEAN ABSOLUTE PERCENT ERROR ACROSS ALL REGIONS

FACTOR	**AVERAGE**	**MAXIMUM**	**MAXIMUM REGION**
AGRICULTURAL AND OTHER LOW-SKILLED WORKERS	57.66	2680.36	CIV
CAPITAL	82.47	4266.69	NOR
CLERKS	96.48	5959.66	CIV
LAND	61.32	2120.62	NAM
NATURAL RESOURCES	233.30	28926.07	XNF
OFFICERS AND MANAGERIAL PROFESSIONALS	60.64	1178.09	BHR
SERVICE AND SHOP FLOOR WORKERS	160.79	11965.78	LAO
TECHNICALLY SKILLED PROFESSIONALS	48.24	694.91	CIV
MEAN ($\boldsymbol{\mu}$)	100.11	7224.02	

From *Table 1*, the value for the agricultural and other low-skilled workers factor varies by 2680.36% for Côte d'Ivoire when measuring the variation in absolute terms by applying the

[15] See Appendix D for TABLE 2 LARGEST RANGES FOR OUTPUT COUNTRY – SECTOR TO INPUT COUNTRY – SECTOR

MAPE formula for the random local maximum and the proportionally imputed matrix B to derive sector inputs and outputs.

MONTE-CARLO SIMULATIONS

As a second step, Monte-Carlo simulations are applied to generate B matrices in order to examine the extent of bias a multi-start method can have on factor content computations. The structure of the simulation model centers around the order in which every b_{igjh} is imputed. In the context of measuring the maximum potential bias, expectations on results are limited as they are dependent on chance. However, we do obtain insights on the potential types of distributions for every factor and an estimated value of maximum potential bias.

To obtain a $(1-\alpha)$ confidence interval for the potential maximum bias, the number of Monte-Carlo simulations required is estimated using:

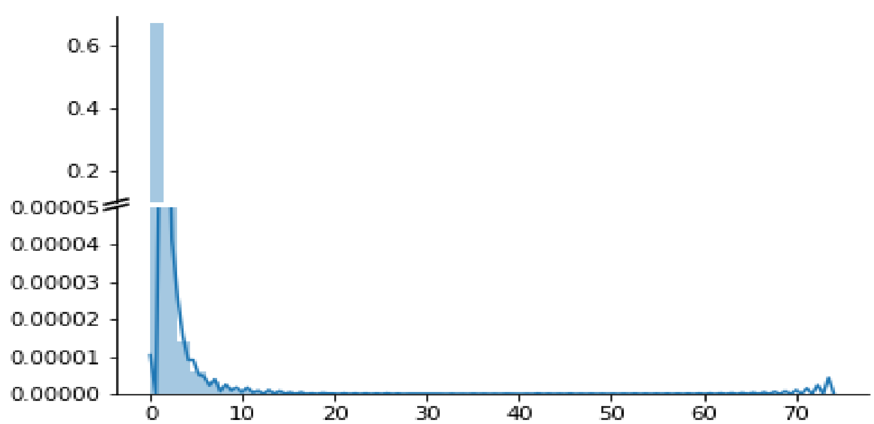
$$n = \left[\left(s \times Z\alpha_{/2}\right)/E\right]^2 \qquad (9)$$

where s is the estimated standard deviation of the mean obtained from simulations, $Z\alpha_{/2}$ is the critical value of the normal distribution for $\alpha/2$, and E is the precision intended based on the shrinking ratio of the upper and lower confidence limits (U, L):

$$U = \bar{x} + Z\alpha_{/2} s_{\bar{x}} \qquad (10)$$

$$L = \bar{x} - Z\alpha_{/2} s_{\bar{x}} \qquad (5)$$

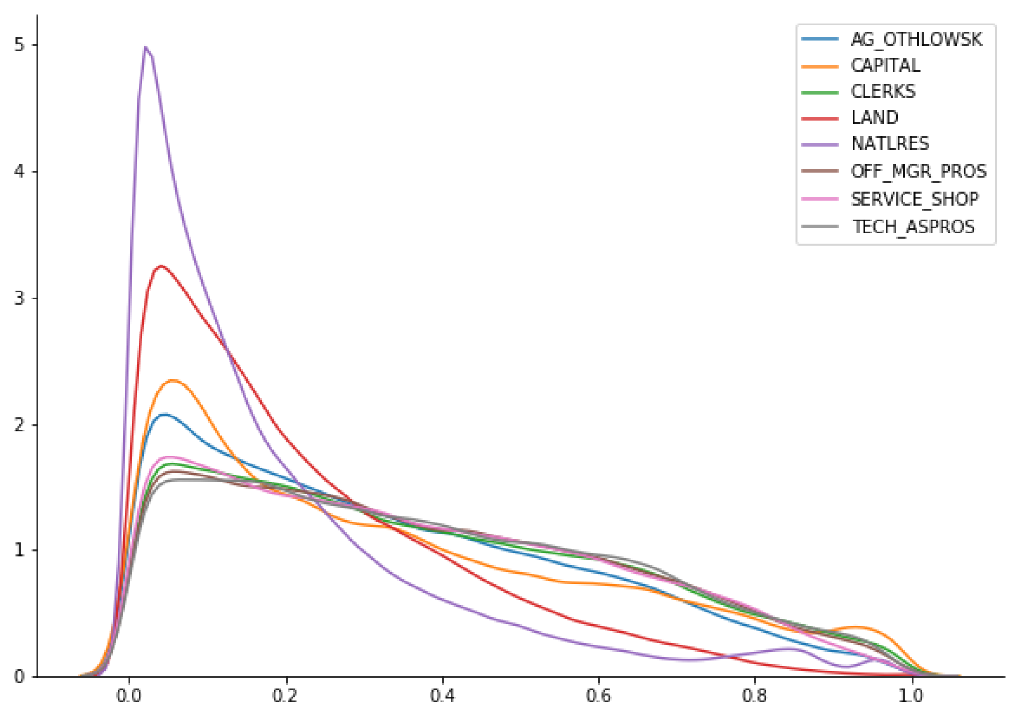
FIGURE 1: PROBABILITY DENSITY PLOT OF TRADE RELATIONS BASED ON VOLUME



After running 3300 simulations, enough to construct a 90% confidence interval and a 5% error[16], the bias results obtained from the Monte-Carlo method are greater than the random local maximum in all but one simulation. With such a result, we can conclude that the aggregate of smaller trade relations has a greater impact on factor content of trade than larger trade relations as Monte-Carlo simulations do not set a prioritization based on trade volume as the random local maximum does. Instead, they are initiated at random, with a greater likelihood of maximizing smaller trade relations as they are more frequent, as observed in *Figure 1: Probability density plot of trade relations based on volume*.

After transforming MAPE from Monte-Carlo simulations to $(0,1)$[17] to understand the distribution of factors, the aggregate of errors all exhibit properties one would expect from log-

FIGURE 2: KERNEL DENSITY OF ESTIMATED MAPE FROM MONTE-CARLO SIMULATIONS



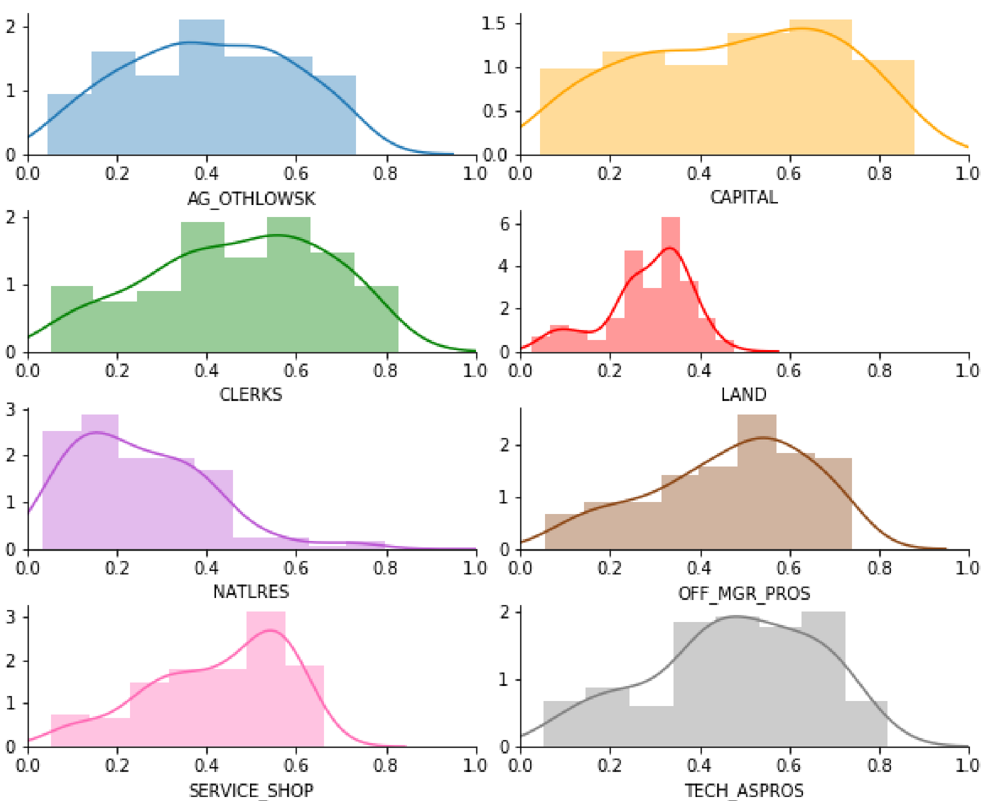
[16] Based on factor-country pair requiring most simulations. All other factor-country pairs hold a higher confidence percentage and a lower error percentage at that amount of simulations.
[17] Transforming through $x' = ((x - a')) / ((b' - a'))\ where\ a' = x(1)$ is the minimum and b'= x(n) is the maximum for every country and every factor.

normal distributions with varying standard deviations, as illustrated in *Figure 2: Kernel density of estimated MAPE from Monte-Carlo simulations.*

Nonetheless, when disaggregated, some of the observed distributions at the country level have a coefficient of variation greater than 0.33, which implies reasonable suspicion on their normality, as presented in *Figure 3: Distribution plots of coefficient of variation for factors by country.*

With a lack of normality across all factors and countries, the prospect of estimating likely results for the factor content of trade is constrained as in many cases multiple values would have the same or similar probability.

FIGURE 3: DISTRIBUTION PLOTS OF COEFFICIENT OF VARIATION FOR FACTORS BY COUNTRY



In addition to obtaining insights on the types of distributions, Monte-Carlo simulations showcase an estimated value of maximum potential bias from a multi-start method, as presented in *Table 2: MAPE from most biased Monte-Carlo simulation across all regions*.

TABLE 2: MAPE FROM MOST BIASED MONTE-CARLO SIMULATION ACROSS ALL REGIONS

FACTOR	**AVERAGE**	**MAXIMUM**	**MAXIMUM REGION**
AGRICULTURAL AND OTHER LOW-SKILLED WORKERS	318.81	29236.90	CIV
CAPITAL	302.55	21237.62	IRL
CLERKS	199.54	14932.51	CIV
LAND	103.90	1112.23	CHE
NATURAL RESOURCES	254.49	30564.56	XNF
OFFICERS AND MANAGERIAL PROFESSIONALS	100.46	1524.93	BHR
SERVICE AND SHOP FLOOR WORKERS	270.83	18675.09	LAO
TECHNICALLY SKILLED PROFESSIONALS	116.00	3594.05	IRL
MEAN (μ)	208.32	15109.74	

GENETIC ALGORITHMS

As a third step, heuristics are developed to narrow the search for the global maximum bias. Heuristics have proven to be more effective approaches to combinatorial optimization problems than multi-start methods, such as Monte-Carlo simulations, at finding a maximum. While the computational complexity of the problem is not reduced and there is no guarantee of reaching the global maximum, by exploiting the idea of learning by transferring knowledge from one solution to the next, heuristics explore the search space more productively, rendering better results.

For the purpose of measuring bias in the proportionality assumption of factor content of trade, genetic algorithms (GAs) are developed using neo-Darwinian theory of evolution, where learning takes the form of mutation and selection.

The development and implementation of GAs began in the 1960s under different names. In Germany, Ingo Rechenberg and Hans-Paul Schwefel developed *Evolutionsstrategie* (evolution strategy, in English). Bremermann, Fogel and others in the USA worked on *evolutionary programming*. As for the term “genetic algorithm”, it was first used in John Holland’s book *Adaptation in Natural and Artificial Systems* in 1975. The common theme in all this research is the notion of mutation and selection for adaptation, which later showcased its capabilities when applied to the traveling salesman and to the gas pipeline optimization problems.

Starting with a discrete search space X and a function:

$$f : X \mapsto \mathbb{R} \tag{12}$$

A genetic algorithm can be designed to find the maximum bias associated with the proportionality assumption as such:

$$\arg\max_{x \in X} f \tag{13}$$

where x is a vector of decision variables and f is the objective function. A distinct characteristic of the genetic algorithm approach is to allow for a departure from the actual variables of the original formulation into their biological interpretation.

The biological analogy starts with a population of feasible solutions. Parents from the population are selected to breed new child solutions by crossing over genes. Genes are then mutated to obtain stochasticity in the child solutions, a safeguard against getting stuck in sub-optimal areas of the solution space by ensuring diversity in the population. Child solutions are then assessed against the objective function and replace parents in the population if they perform better, becoming parents themselves. Breeding stops once a set number of breeds fail to be admitted to the population, indicating a maximum being reached given the stopping criteria.

In the development, a lot of flexibility is left in the hands of the researcher, all with their own sets of trade-offs: initial population, termination, selection, crossover, and mutation. Finding the goldilocks efficiency and efficacy becomes a function of problem understanding and testing. Deciding on an initial population that is too large would impede on the efficiency of finding a solution in a reasonable time. Too small of a population would constrain the search space, limiting the likelihood of finding an optimal solution. Ensuring every combination could be built from the initial population by crossover only could narrow the selection of an initial population. However, such considerations omit computational requirements of certain problems. Selection, crossover, and mutation all have very important impacts on one another. Selecting multiple parents influences the distributional and positional bias of crossovers. Mutations, on the other hand, have a bearing on the effects of selection. Certain genetic algorithms are built with only crossovers, some with only mutations, while others combine both. All variations come with compromises, and all have presented success in specific problems (Gendreau and Potvin, 2010).

A first GA combining both crossovers and mutations to search for the maximum bias of the proportionality assumption of factor content of trade is developed, for which the brief pseudo code reads:

BRIEF ALGORITHM 2: GENETIC ALGORITHM WITH CROSSOVERS AND MUTATIONS [18]

```
1  tz_fct = leontief_inverse(VFM, prop_B, trade_tz)
2  pop, scores ← initial population of matrix B, corresponding sorted list of bias scores
3  WHILE count < x AND scores[x] = scores[x+1] ∀ x ∈ scores
4    child = cross&mutate
5    child_fct = leontief_inverse(VFM, child_B, trade_child)
6    fit = MAPE (tz_fct, child_fct)
7    IF fit > scores[0]
8      pop[0], scores[0] = child, fit
9      pop, scores = sort(pop, scores)
10     count = 0
11   ELSE
12     count +=1
```

The algorithm starts off by calculating the factor content of trade ($F_i = D\ (I - B)^{-1}\ T_i$) for all countries using the proportionally imputed matrix B (line 1). Next, an initial population of feasible B matrices is selected[19] and each parent is scored based on fit using the MAPE for all countries and all factors, representing total average bias between the factor content of trade calculated using the proportionally imputed matrix B and each feasible B matrix (line 2).

The following step consists of defining the stopping criteria, genetic algorithms are stochastic search methods that could continue indefinitely otherwise. The stopping criteria are ideally based on a probabilistic principle that all viable combinations have been assessed; however, more practical proxies are applied. For this application, two stopping criteria are defined; sub-optimal crossovers and mutations in a row reaching a count equal to the number of Monte-Carlo simulations required for a 90% confidence interval and a 5% error, and a drop in the diversity of the population to a pre-set threshold of $1e - 09$ (line 3). The first criterion ensures a thorough assessment of search space X with a small likelihood of finding values beyond the upper limit, especially considering the counter restart with changes to the upper limit (line 10). The second criterion takes into account the population diversity in which a drop would indicate homogeneous solutions or multiple solutions with the same fit, suggesting a maximum.

[18] See Appendix E for detailed pseudo code of ALGORITHM 2: GENETIC ALGORITHM WITH CROSSOVERS AND MUTATIONS

[19] The initial population can be selected from the best results of the Monte-Carlo simulations for an enhanced initial population which limits future search iterations.

The next step selects parents from the population, crosses their genes, and mutates a portion of them. For a seamless exchange, genes are defined by segments of input countries $j \; \forall \; i, g, h$, where each segment is subject to independent constraints, avoiding the use of computational resources to test for feasibility and repairing unfeasible genes (line 4). More granular definitions of genes following limits specified by constraints of imported input-outputs and bilateral trade offer limited value when considering added computational time of crossovers and mutations.

FIGURE 4: CROSS AND MUTATE

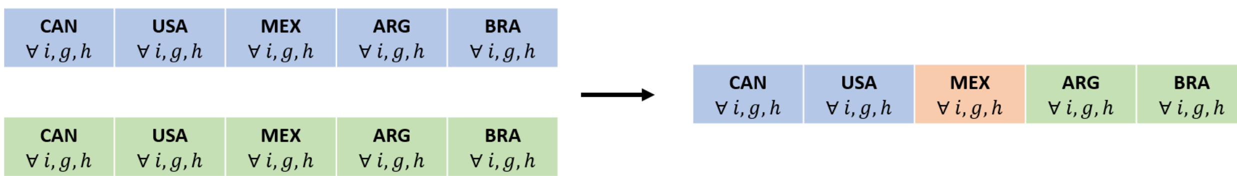


Figure 4: Cross and mutate illustrates a simplified cross of two parents (left) to breed a child solution (right) where genes for CAN and USA are taken from parent 1 and genes ARG and BRA are taken from parent 2. Gene for MEX is mutated, meaning it is not taken from either parent 1 or parent 2, it is generated through the maximization framework. To avoid positional bias of one-point crossovers and distributional bias of multi-point crossovers, the crossover and mutation sequence is generated at random for every child solution.

The factor content of trade is then calculated for the child solution for which a matrix of $63{,}680{,}400$ values is inverted and trade vectors T_i are computed to reflect flows from matrix B (line 5)[20].

Next, the fit is assessed using the mean absolute percent error for all countries and all factors between the factor content of trade calculated using the proportionally imputed matrix B and the child solution (line 6). If the fit value is greater than the lowest fit in the population, the child replaces the parent and the count restarts (lines 7-10). If the fit value is lower than the lowest fit in the population, the counter increases by 1 and lines 3 through 12 are repeated.

[20] Matrix B inversion and trade vectors T_i computation is required for the proportionality assumption, the random local maximum, and Monte-Carlo simulations as well.

Various permutations of initial population, selection, crossover, and mutation have been tested with populations ranging from 2 to 20, selection varying from 2 to the total population count, crossovers reflecting stochastic and proportional sampling based on fit, and mutations shifting between a single gene up to half of the phenotype.

The most efficient and effective genetic algorithm proved to be the algorithm with a population of 2 and a single mutated gene. Any increase in parent or mutation count lowered the convergence rate incrementally, offering limited expectations from additional testing. As for different types of crossovers, they showed no significant impact on the obtained results.

TABLE 3: MAPE FROM MOST BIASED CROSSOVER AND MUTATION GENETIC ALGORITHM ACROSS ALL REGIONS

FACTOR	**AVERAGE**	**MAXIMUM**	**MAXIMUM REGION**
AGRICULTURAL AND OTHER LOW-SKILLED WORKERS	370.69	34250.65	CIV
CAPITAL	300.99	19670.91	IRL
CLERKS	221.29	18181.91	CIV
LAND	110.64	2931.12	BLR
NATURAL RESOURCES	246.88	29252.51	XNF
OFFICERS AND MANAGERIAL PROFESSIONALS	109.01	1571.75	TWN
SERVICE AND SHOP FLOOR WORKERS	252.84	17084.76	LAO
TECHNICALLY SKILLED PROFESSIONALS	115.55	3386.63	IRL
MEAN ($\boldsymbol{\mu}$)	215.99	15791.28	

With large initial populations and numerous mutations, genetic algorithms for the application of searching for the maximum bias of the proportionality assumption applied to factor content of trade shared more with a multi-start method than with a heuristic[21]. Narrowing the scope of transformations presented a way for the genetic algorithm to learn progressively and transmit the information to the rest of the population, rendering improved results to any other genetic algorithm or Monte-Carlo simulation in less time and iterations.

The success of the simplest genetic algorithm could be attributed to *interactome* – indirect interactions among genes, or rather lack thereof. Constrained interactions are caused by the definition of genes which are segments of input countries $j \,\forall\, i, g, h$. Adding interactome

[21] Large population and numerous mutations limit the transfer of learning from one solution to the next.

requires breaking down genes beyond their constraints of imported input-outputs and bilateral trade. However, given the sparsity of imported matrices, their crossover would rarely follow a maximization framework, rendering sub-optimal solutions.

A second GA with only mutations to search for the maximum bias of the proportionality assumption of factor content of trade is developed, for which the brief pseudo code reads:

BRIEF ALGORITHM 3: GENETIC ALGORITHM WITH MUTATIONS

```
1  tz_fct = leontief_inverse(VFM, prop_B, trade_tz)
2  p_fct = leontief_inverse(VFM, parent_B, trade_parent)
3  fit_p, col_fit_p, bias_p = MAPE(tz_fct, p_fct), CMAPE(tz_fct, p_fct), MMAPE(tz_fct, p_fct)
4  WHILE count < x
5     child_B = shaking(list_ig, list_jh, VXMD_igj, VIFM_gjh, VDFM_igjh)
6     c_fct = leontief_inverse(VFM, child_B, trade_child)
7     fit_c, col_fit_c, bias_c = MAPE(tz_fct, c_fct), CMAPE(tz_fct, c_fct), MMAPE(tz_fct, c_fct)
8     FOR z in range(len(list_j))
9        IF col_fit_c[z] > col_fit_p[z]
10          fit_p[z]= fit_c[z]
11    col_fit_p, bias_ c = mean(fit_p, axis = 0), mean(fit_p)
12    IF bias_p == bias_ c
13       count +=1
14    ELSE
15       bias_p = bias_ c
16       count =0
```

The algorithm starts off by calculating the factor content of trade ($F_i = D\,(I - B)^{-1}\,T_i$) for the proportionally imputed matrix B and for a parent solution (line 1-2). Next, the mean absolute percent error for all countries and all factors, the mean absolute percent error for all countries, and the overall mean absolute percent error between the proportionally imputed matrix B and the parent solution are computed (line 3). The following step consists of the stopping criteria, which is simply the number of Monte-Carlo simulations required for a 90% confidence interval and a 5% error with the same intuition as the GA with crossovers and mutation (line 4). The next step generates a new matrix B in a function defined as shaking which follows the instructions defined under algorithm 1 maximization framework for imported matrix and adding values from domestic input-output matrices (line 5). Then, the factor content of trade ($F_i = D\,(I - B)^{-1}\,T_i$) and all associated fit and bias measures are computed for the child solution (line 6-7). The following steps consist of iterating over each country and comparing the fit between the parent gene and child gene. If the child gene has a greater fit, it replaces the parent gene in the solution for the given country (line 8-10). Next, the mean absolute percent error for all countries and the overall mean absolute percent error between the proportionally imputed matrix B and the new parent solution are computed (line 11). If the bias between the old parent and the new parent remains the same, the count increases, otherwise the count restarts at 0 (line 12-16).

TABLE 4: MAPE FROM MOST BIASED MUTATION GENETIC ALGORITHM ACROSS ALL REGIONS

FACTOR	**AVERAGE**	**MAXIMUM**	**MAXIMUM REGION**
AGRICULTURAL AND OTHER LOW-SKILLED WORKERS	444.26	36115.36	CIV
CAPITAL	369.85	23131.86	IRL
CLERKS	291.50	18289.81	CIV
LAND	343.46	7660.99	TTO
NATURAL RESOURCES	297.68	33045.13	XNF
OFFICERS AND MANAGERIAL PROFESSIONALS	187.71	2737.96	TWN
SERVICE AND SHOP FLOOR WORKERS	533.71	42328.63	LAO
TECHNICALLY SKILLED PROFESSIONALS	175.62	3867.88	IRL
MEAN ($\boldsymbol{\mu}$)	330.47	20897.20	

With only mutations of a single parent, the GA resembles a hill-climbing method, where it keeps all previous solutions with the exception of new improvements. This method proves to be very efficient in finding the maximum bias associated with the proportionality assumption compared to Monte-Carlo simulations and a GA with crossovers and mutations, given the results obtained.

CONCLUSION

With a lack of available data, simulations provide a tangible method to estimate the potential bias resulting from the use of the proportionality assumption in the calculation of factor content of trade. Even though the real bias can only be measured with actual data, simulations confirm the National Research Council's address of the proportionality assumption issue as being a considerable restriction of present data collection and analysis. The research presented in this paper finds a maximum potential bias averaging 330% for all countries across all factors using genetic algorithms and an average bias of 152.72% for all countries across all factors using Monte-Carlo simulations in the maximization framework.

In this regard, simulations, and an emphasis on constraints of imported input-outputs and bilateral trade can support policy makers and researchers in cautious assessments of carbon dioxide (CO_2) content of trade, trade in value added, R&D content of intermediates, wage structures, and material offshoring, as well as other areas that depend on linkages in trade.

Certain countries are more exposed than others to the bias of the proportionality assumption, as shown in *Table 5: Average maximum bias for all factors of most and least exposed countries*:

TABLE 5: AVERAGE MAXIMUM BIAS FOR ALL FACTORS OF MOST AND LEAST EXPOSED COUNTRIES

MOST EXPOSED COUNTRIES	AVERAGE MAXIMUM BIAS	LEAST EXPOSED COUNTRIES	AVERAGE MAXIMUM BIAS
CIV	7365.42	**BEN**	8.90
LAO	5348.70	**NGA**	16.97
XNF	4205.56	**NPL**	18.78
IRL	4052.23	**KGZ**	20.61
NOR	1704.94	**GEO**	21.73

where country trade flows and factor demand influence the average maximum bias. Countries with trading partners that have greater heterogeneity in factor demands (VFM_{fjh}) could have a larger potential bias, especially if it is paired with larger volumes of bilateral trade. The opposite would apply to countries with a low average maximum bias.

Overall, heuristics cannot guarantee finding the global maximum bias of the proportionality assumption; however, their use gives a sufficiently good solution considering the complexity of the combinatorial optimization problem. On the other hand, Monte-Carlo simulations allow for a better understanding of the search space by showcasing the various possible distributions for factors and countries. Through this exercise, one can extract findings otherwise only obtainable with more data, which is rare.

In the context of factor content of trade, bias arising from the proportionality assumption confirms the possibility of understating or overstating the relative use of imported inputs, suggesting that a range of probable error based on this assumption would be meaningful for researchers and policymakers, who are trying to understand sources of comparative advantage.

REFERENCES

Aguiar, A., Narayanan, B. and McDougall, R., 2016. An overview of the GTAP 9 data base. *Journal of Global Economic Analysis*, *1*(1), pp.181-208.

Bhagwati, J., 1964. The pure theory of international trade: A survey. *The Economic Journal*, *74*(293), pp.1-84.

Casella, B., Bolwijn, R., Moran, D. and Kanemoto, K., 2019. Improving the analysis of global value chains: the UNCTAD-Eora Database. TRANSNATIONAL CORPORATIONS, 26(3), pp.115.

Caron, J., Metcalf, G.E. and Reilly, J., 2017. The CO 2 content of consumption across US regions: a multi-regional input-output (MRIO) approach. *The Energy Journal*, *38*(1).

Conway, P.J., 2002. The case of the missing trade and other mysteries: Comment. *American Economic Review*, *92*(1), pp.394-404.

Davis, D.R. and Weinstein, D.E., 2001. An account of global factor trade. *American Economic Review*, *91*(5), pp.1423-1453.

Davis, D.R. and Weinstein, D.E., 2003. The factor content of trade. Handbook of international trade, 1, pp.119-145.

Deardorff, A.V., 1982. The general validity of the Heckscher-Ohlin theorem. *The American Economic Review*, *72*(4), pp.683-694.

Debaere, P., 2003. Relative factor abundance and trade. *Journal of Political Economy*, *111*(3), pp.589-610.

Feenstra, R.C. and Hanson, G.H., 1996. *Globalization, outsourcing, and wage inequality* (No. w5424). National Bureau of Economic Research.

Feenstra, R.C. and Hanson, G.H., 1999. The impact of outsourcing and high-technology capital on wages: estimates for the United States, 1979–1990. *The Quarterly Journal of Economics*, *114*(3), pp.907-940.

Feenstra, R.C. and Jensen, J.B., 2012. Evaluating estimates of materials offshoring from US manufacturing. *Economics Letters*, *117*(1), pp.170-173.

Forkman, J. and Verrill, S., 2008. The distribution of McKay's approximation for the coefficient of variation. *Statistics & probability letters*, *78*(1), pp.10-14.

Gendreau, M. and Potvin, J.Y. eds., 2010. Handbook of metaheuristics (Vol. 2, p. 9). New York: Springer.

Helpman, E. and Krugman, P.R., 1985. *Market structure and foreign trade: Increasing returns, imperfect competition, and the international economy*. MIT press.

Hummels, D., Ishii, J. and Yi, K.M., 2001. The nature and growth of vertical specialization in world trade. *Journal of international Economics*, *54*(1), pp.75-96.

Johnson, R.C., 2018. Measuring global value chains. *Annual Review of Economics*, *10*, pp.207-236.

Johnson, R.C. and Noguera, G., 2012. Accounting for intermediates: Production sharing and trade in value added. *Journal of international Economics*, *86*(2), pp.224-236.

Lenzen, M., Kanemoto, K., Moran, D. and Geschke, A., 2012. Mapping the structure of the world economy. Environmental science & technology, 46(15), pp.8374-8381.

Milberg, W. and Winkler, D.E., 2010. Errors from the 'Proportionality Assumption' in the Measurement of Offshoring: Application to German Labor Demand. *Available at SSRN 1635800*.

Nagengast, A. and Stehrer, R., 2014. Collateral imbalances in intra-european trade? Accounting for the difference between gross and value added trade balances.

National Research Council, 2006. *Analyzing the US content of imports and the foreign content of exports*. National Academies Press.

Nishioka, S. and Ripoll, M., 2012. Productivity, trade and the R&D content of intermediate inputs. *European Economic Review*, *56*(8), pp.1573-1592.

Peters, G.P. and Hertwich, E.G., 2008. CO2 embodied in international trade with implications for global climate policy.

Puzzello, L., 2012. A proportionality assumption and measurement biases in the factor content of trade. *Journal of International Economics*, *87*(1), pp.105-111.

Reimer, J.J., 2006. Global production sharing and trade in the services of factors. *Journal of International Economics*, *68*(2), pp.384-408.

Reeves, C.R., 1999, July. Predictive measures for problem difficulty. In *Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406)* (Vol. 1, pp. 736-743). IEEE.

Thomopoulos, N.T., 2012. *Essentials of Monte Carlo simulation: Statistical methods for building simulation models*. Springer Science & Business Media.

Trefler, D., 1993. International factor price differences: Leontief was right!. *Journal of political Economy*, *101*(6), pp.961-987.

Trefler, D., 1995. The case of the missing trade and other mysteries. *The American Economic Review*, pp.1029-1046.

Trefler, D. and Zhu, S.C., 2000. Beyond the algebra of explanation: HOV for the technology age. *American Economic Review*, *90*(2), pp.145-149.

Trefler, D. and Zhu, S.C., 2010. The structure of factor content predictions. *Journal of International Economics*, *82*(2), pp.195-207.

Vanek, J., 1968. The factor proportions theory: The n—factor case. *Kyklos*, *21*(4), pp.749-756.

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

Zhu, S.C. and Trefler, D., 2005. Trade and inequality in developing countries: a general equilibrium analysis. *Journal of international Economics*, *65*(1), pp.21-48.

APPENDIX A: LIST OF COUNTRIES AND REGIONS

Code	Country / Region Description
AUS	Australia
	- Australia
	- Christmas Island
	- Cocos (Keeling) Islands
	- Heard Island and McDonald Islands
	- Norfolk Island
NZL	New Zealand
XOC	Rest of Oceania
	- American Samoa
	- Cook Islands
	- Fiji
	- French Polynesia
	- Guam
	- Kiribati
	- Marshall Islands
	- Micronesia, Federated States of
	- Nauru
	- New Caledonia
	- Niue
	- Northern Mariana Islands
	- Palau
	- Papua New Guinea
	- Pitcairn
	- Samoa
	- Solomon Islands
	- Tokelau
	- Tonga
	- Tuvalu
	- United States Minor Outlying Islands
	- Vanuatu
	- Wallis and Futuna Islands
CHN	China
HKG	Hong Kong, Special Administrative Region of China
JPN	Japan
KOR	Korea, Republic of
MNG	Mongolia
TWN	Taiwan
XEA	Rest of East Asia
	- Korea, Democratic People's Republic of
	- Macao, Special Administrative Region of China
BRN	Brunei Darussalam
KHM	Cambodia
IDN	Indonesia
LAO	Lao PDR
MYS	Malaysia
PHL	Philippines
SGP	Singapore
THA	Thailand
VNM	Viet Nam
XSE	Rest of Southeast Asia
	- Myanmar
	- Timor-Leste
BGD	Bangladesh
IND	India
NPL	Nepal
PAK	Pakistan
LKA	Sri Lanka
XSA	Rest of South Asia
	- Afghanistan
	- Bhutan
	- Maldives
CAN	Canada
USA	United States of America
MEX	Mexico
XNA	Rest of North America
	- Bermuda
	- Greenland
	- Saint Pierre and Miquelon
ARG	Argentina
BOL	Bolivia
BRA	Brazil
CHL	Chile
COL	Colombia
ECU	Ecuador
PRY	Paraguay
PER	Peru
URY	Uruguay
VEN	Venezuela (Bolivarian Republic of)
XSM	Rest of South America

	- Falkland Islands (Malvinas)
	- French Guiana
	- Guyana
	- South Georgia and the South Sandwich Islands
	- Suriname
CRI	Costa Rica
GTM	Guatemala
HND	Honduras
NIC	Nicaragua
PAN	Panama
SLV	El Salvador
XCA	Rest of Central America
	- Belize
DOM	Dominican Republic P
JAM	Jamaica
PRI	Puerto Rico
TTO	Trinidad and Tobago P
XCB	Rest of Caribbean
	- Anguilla
	- Antigua and Barbuda
	- Aruba
	- Bahamas
	- Barbados
	- British Virgin Islands
	- Cayman Islands
	- Cuba
	- Dominica
	- Grenada
	- Haiti
	- Montserrat
	- Netherlands Antilles
	- Saint Kitts and Nevis
	- Saint Lucia
	- Saint Vincent and Grenadines
	- Turks and Caicos Islands
	- Virgin Islands, US
AUT	Austria
BEL	Belgium
CYP	Cyprus
CZE	Czech Republic
DNK	Denmark
EST	Estonia
FIN	Finland
	- Aland Islands
	- Finland
FRA	France
	- France
	- Guadeloupe
	- Martinique
	- Réunion
DEU	Germany
GRC	Greece
HUN	Hungary
IRL	Ireland
ITA	Italy
LVA	Latvia
LTU	Lithuania
LUX	Luxembourg
MLT	Malta
NLD	Netherlands
POL	Poland
PRT	Portugal
SVK	Slovakia
SVN	Slovenia
ESP	Spain
SWE	Sweden
GBR	United Kingdom
CHE	Switzerland
NOR	Norway
	- Norway
	- Svalbard and Jan Mayen Islands
XEF	Rest of European Free Trade Association
	- Iceland
	- Liechtenstein
ALB	Albania
BGR	Bulgaria
BLR	Belarus
HRV	Croatia
ROU	Romania
RUS	Russian Federation
UKR	Ukraine
XEE	Rest of Eastern Europe
	- Moldova
XER	Rest of Europe
	- Andorra

	- Bosnia and Herzegovina
	- Faroe Islands
	- Gibraltar
	- Guernsey
	- Holy See (Vatican City State)
	- Isle of Man
	- Jersey
	- Macedonia, Republic of
	- Monaco
	- Montenegro
	- San Marino
	- Serbia
KAZ	Kazakhstan
KGZ	Kyrgyztan
XSU	Rest of Former Soviet Union
	- Tajikistan
	- Turkmenistan
	- Uzbekistan
ARM	Armenia
AZE	Azerbaijan
GEO	Georgia
BHR	Bahrain
IRN	Iran, Islamic Republic of
ISR	Israel
JOR	Jordan
KWT	Kuwait
OMN	Oman
QAT	Qatar
SAU	Saudi Arabia
TUR	Turkey
ARE	United Arab Emirates
XWS	Rest of Western Asia
	- Iraq
	- Lebanon
	- Palestinian Territory, Occupied
	- Syrian Arab Republic (Syria)
	- Yemen
EGY	Egypt
MAR	Morocco
TUN	Tunisia
XNF	Rest of North Africa
	- Algeria
	- Libya
	- Western Sahara
BEN	Benin
BFA	Burkina Faso
CMR	Cameroon
CIV	Côte d'Ivoire
GHA	Ghana
GIN	Guinea
NGA	Nigeria
SEN	Senegal
TGO	Togo
XWF	Rest of Western Africa
	- Cape Verde
	- Gambia
	- Guinea-Bissau
	- Liberia
	- Mali
	- Mauritania
	- Niger
	- Saint Helena
	- Sierra Leone
XCF	Rest of Central Africa
	- Central African Republic
	- Chad
	- Congo
	- Equatorial Guinea
	- Gabon
	- Sao Tome and Principe
XAC	South Central Africa
	- Angola
	- Congo, Democratic Republic of the
ETH	Ethiopia
KEN	Kenya
MDG	Madagascar
MWI	Malawi
MUS	Mauritius
MOZ	Mozambique
RWA	Rwanda
TZA	Tanzania, United Republic of
UGA	Uganda
ZMB	Zambia
ZWE	Zimbabwe
XEC	Rest of Eastern Africa
	- Burundi

	- Comoros
	- Djibouti
	- Eritrea
	- Mayotte
	- Seychelles
	- Somalia
	- Sudan
BWA	Botswana
NAM	Namibia
ZAF	South Africa
XSC	Rest of South African Customs Union
	- Lesotho
	- Swaziland
XTW	Rest of the World
	- Antarctica
	- Bouvet Island
	- British Indian Ocean Territory
	- French Southern Territories

APPENDIX B: LIST OF SECTORS

Code	Sector Description
PDR	Paddy rice
WHT	Wheat
GRO	Cereal grains nec
V_F	Vegetables, fruit, nuts
OSD	Oil seeds
C_B	Sugar cane, sugar beet
PFB	Plant-based fibers
OCR	Crops nec
CTL	Bovine cattle, sheep and goats, horses
OAP	Animal products nec
RMK	Raw milk
WOL	Wool, silk-worm cocoons
FRS	Forestry
FSH	Fishing
COA	Coal
OIL	Oil
GAS	Gas
OMN	Minerals nec
CMT	Bovine meat products
OMT	Meat products nec
VOL	Vegetable oils and fats
MIL	Dairy products
PCR	Processed rice
SGR	Sugar
OFD	Food products nec
B_T	Beverages and tobacco products
TEX	Textiles
WAP	Wearing apparel
LEA	Leather products
LUM	Wood products
PPP	Paper products, publishing
P_C	Petroleum, coal products
CRP	Chemical, rubber, plastic products
NMM	Mineral products nec
I_S	Ferrous metals
NFM	Metals nec
FMP	Metal products
MVH	Motor vehicles and parts
OTN	Transport equipment nec
ELE	Electronic equipment
OME	Machinery and equipment nec
OMF	Manufactures nec
ELY	Electricity
GDT	Gas manufacture, distribution
WTR	Water
CNS	Construction
TRD	Trade
OTP	Transport nec
WTP	Water transport
ATP	Air transport
CMN	Communication
OFI	Financial services nec
ISR	Insurance
OBS	Business services nec
ROS	Recreational and other services
OSG	Public Administration, Defense, Education, Health
DWE	Dwellings

APPENDIX C: ALGORITHM 1 MAXIMIZATION FRAMEWORK FOR IMPORTED MATRIX

ALGORITHM 1 MAXIMIZATION FRAMEWORK FOR IMPORTED MATRIX

```
PROCEDURE imputing b_igjh's
B_matrix ← empty matrix of size ig × jh
i_count ← length i
i_list ← list of numbers 0 to i
WHILE i_count >= 1
  i_coor ← random integer between 0 and i_count
  i ← item in i_list corresponding to i_coor
  i_list[i_coor], i_list[i_count - 1] = i_list[i_count - 1], i_list[i_coor]
  i_count −= 1
  g_count ← length g
  g_list ← list of numbers 0 to g
  WHILE g_count >= 1
    g_coor ← random integer between 0 and g_count
    g ← item in g_list corresponding to g_coor
    g_list[g_coor], g_list[g_count - 1] = g_list[g_count - 1], g_list[g_coor]
    g_count −= 1
    j_count ← length j
    j_list ← list of numbers 0 to j
    WHILE j_count >= 1
      j_coor ← random integer between 0 and j_count
      j ← item in j_list corresponding to j_coor
      j_list[j_coor], j_list[j_count - 1] = j_list[j_count - 1], j_list[j_coor]
      j_count −= 1
      IF bilateral constraints [igj] = 0
        continue
      ELSE
        h_count ← length h
        h_list ← list of numbers 0 to h
        WHILE h_count >= 1
          h_coor ← random integer between 0 and h_count
          h ← item in h_list corresponding to h_coor
          h_list[h_coor], h_list[h_count - 1] = h_list[h_count - 1], h_list[h_coor]
          h_count −= 1
          IF imported input-output [gjh] = 0
            continue
          ELSE
            IF bilateral constraints [igj] < imported input-output [gjh]
              B_matrix[igjh] += bilateral trade constraint [igj]
              bilateral trade constraint [igj] ← 0
              imported input-output constraint [gjh] −= bilateral trade constraint [igj]
              list_i, list_g, list_j −= i, g, j
              break
            ELSE
              B_matrix[igjh] += imported input-output [gjh]
              imported input-output constraint [gjh] ← 0
              bilateral trade constraint [igj] −= imported input-output constraint [gjh]
              list_g, list_j, list_h −= g, j, h
```

APPENDIX D: LARGEST RANGES FOR OUTPUT COUNTRY-SECTOR TO INPUT COUNTRY-SECTOR

LARGEST RANGES FOR OUTPUT COUNTRY – SECTOR TO INPUT COUNTRY – SECTOR

China – Electronic equipment	→	United States of America – Electronic equipment
Japan – Machinery and equipment nec	→	China – Machinery and equipment nec
Saudi Arabia – Oil	→	United States of America – Petroleum, coal products
Taiwan – Electronic equipment	→	China – Electronic equipment
Australia – Minerals nec	→	China – Ferrous metals
Canada – Oil	→	United States of America – Petroleum, coal products
Australia – Minerals nec	→	China – Metals nec
Nigeria – Oil	→	United States of America – Petroleum, coal products
Korea, Republic of – Machinery and equipment nec	→	China – Machinery and equipment nec
Saudi Arabia – Oil	→	Japan – Petroleum, coal products
Mexico – Electronic equipment	→	United States of America – Electronic equipment
China – Machinery and equipment nec	→	United States of America – Machinery and equipment nec
Japan – Machinery and equipment nec	→	United States of America – Machinery and equipment nec
Mexico – Machinery and equipment nec	→	United States of America – Machinery and equipment nec
Germany – Machinery and equipment nec	→	China – Machinery and equipment nec
Japan – Motor vehicles and parts	→	United States of America – Motor vehicles and parts
Mexico – Motor vehicles and parts	→	United States of America – Motor vehicles and parts
Canada – Motor vehicles and parts	→	United States of America – Motor vehicles and parts
United Arab Emirates – Oil	→	Japan – Petroleum, coal products
Taiwan – Machinery and equipment nec	→	China – Machinery and equipment nec
Canada – Chemical, rubber, plastic products	→	United States of America – Chemical, rubber, plastic products
Canada – Chemical, rubber, plastic products	→	United States of America – Education, Health
China – Electronic equipment	→	United States of America – Business services nec
Mexico – Electronic equipment	→	United States of America – Business services nec
Germany – Machinery and equipment nec	→	United States of America – Machinery and equipment nec
Ireland – Chemical, rubber, plastic products	→	United States of America – Chemical, rubber, plastic products
Ireland – Chemical, rubber, plastic products	→	United States of America – Education, Health
Korea, Republic of – Chemical, rubber, plastic products	→	China – Chemical, rubber, plastic products
Germany – Machinery and equipment nec	→	United States of America – Construction
Mexico – Machinery and equipment nec	→	United States of America – Construction
China – Machinery and equipment nec	→	United States of America – Construction
Japan – Machinery and equipment nec	→	United States of America – Construction
China – Chemical, rubber, plastic products	→	United States of America – Construction
China – Chemical, rubber, plastic products	→	United States of America – Chemical, rubber, plastic products
Venezuela (Bolivarian Republic of) – Oil	→	United States of America – Petroleum, coal products
Saudi Arabia – Oil	→	Korea, Republic of – Petroleum, coal products
Russian Federation – Oil	→	Germany – Petroleum, coal products
Japan – Electronic equipment	→	China – Electronic equipment
Germany – Motor vehicles and parts	→	United States of America – Motor vehicles and parts
Japan – Chemical, rubber, plastic products	→	China – Chemical, rubber, plastic products
Mexico – Oil	→	USA- Petroleum, coal products
Rest of Western Asia – Oil	→	IND- Petroleum, coal products
Korea, Republic of – Electronic equipment	→	China – Electronic equipment
Brazil – Minerals nec	→	China – Ferrous metals
Brazil – Minerals nec	→	China – Metals nec
China – Electronic equipment	→	Japan – Electronic equipment
United States of America – Machinery and equipment nec	→	China – Machinery and equipment nec
Saudi Arabia – Oil	→	IND – Petroleum, coal products
South Central Africa – Oil	→	China – Petroleum, coal products
Germany – Motor vehicles and parts	→	China – Motor vehicles and parts

APPENDIX E: ALGORITHM 2 GENETIC ALGORITHM WITH CROSSOVERS AND MUTATIONS

ALGORITHM 2 GENETIC ALGORITHM WITH CROSSOVERS AND MUTATIONS

```
PROCEDURE GA for max bias search with crossovers and mutations

FUNCTION leontief_inverse(VFM, IO, trade)
   rate = 1/(VFM_jh[22]+IO_jh[23]) ∀ jh
   factors[24], B = VFM ∗ rate, IO * rate
   return factors ∗ (I − B)^−1 ∗ trade

FUNCTION cross&mutate(pop)
   child ← empty matrix of size ig-jh
   FOR country in j
      r ←random integer between 0 and pop+1
      IF x = pop+1
         child[j] = maximization(j)
      ELSE
         child[j] = pop[r]
   return child

tz_fct = leontief_inverse(VFM, prop_B, trade_tz)
pop, scores ← initial population of matrix B, corresponding sorted list of bias scores
WHILE count < x AND scores[x] = scores[x+1] ∀ x ∈ scores
   child = cross&mutate
   child_fct = leontief_inverse(VFM, child_B, trade_child)
   fit = MAPE (tz_fct, child_fct)
   IF fit > scores[0]
      pop[0], scores[0] = child, fit
      pop, scores = sort(pop, scores)
      count = 0
   ELSE
      count +=1
```

*where VFM_{jh} is the sum of values in column jh of VFM

[22] where VFM_{jh} is the sum of values in column jh of VFM

[23] where IO_{jh} is the sum of values in column jh of IO

[24] A bias due to rounding is introduced during the rate conversion, however, it is negligible on results as it is present in the factor content of trade calculated using the proportionality assumption and the maximization framework.