High Performance Export Portfolio: Design Growth-Enhancing Export Structure with Machine Learning

Tascha Che, Xuege Zhang

September 16, 2021

Abstract

This paper studies the relationship between export structure and growth performance. We design an export recommendation system using a collaborative filtering algorithm based on countries' revealed comparative advantages. The system is used to produce export portfolio recommendations covering over 190 economies and over 30 years. We find that economies with their actual export structure more aligned with the recommended structure have better growth performance, both in terms of higher GDP growth rate and lower growth volatility. The results are overall robust. These findings demonstrate that export structure matters for achieving high and stable growth. Our recommendations and empirical analysis can serve as a practical tool for policymakers seeking actionable insights on their countries' export potentials that may be complex and hard to quantify.

Keywords: export diversification, comparative advantage, machine learning, collaborative filtering, economic growth, international trade

JEL Codes: F1, F4, O1, O4

1 Introduction

Many success stories of growth and income convergence of the past decades are exports-led, most prominently, China and several other East Asian emerging economies. Some of these countries have governments that actively pursue industrial policies that foster strategic export industries, others let the market take the lead. Either way, there is no denying that export diversification and industrial structural change are important for growth (e.g. Aiginger and Rodrik, 2020). Given the relevence of the topic, there is, however, surprisingly little guidance from the economic literature regarding what type of export structures are growth-enhancing, and what are concrete products, if any, a country could consider to diversify.

Classical trade theory suggests that countries should export what they are relatively good at producing, i.e. following comparative advantages. But how exactly does one ascertain comparative advantages? Trade theories predict that developing countries tend to have comparative advantages in labor-intensive exports and should stay away from capital-intensive industries. But in reality, comparative advantages contain far more dimensions than capital and labor. Some of these dimensions are linearly quantifiable, others are not.

The matter becomes even more complicated when we consider the fact that comparative advantages evolve as a country grows. How should the export structure change as a result? General theories don't go very far in providing country-specific, practical insights in guiding the structural change in exports.

Che (2020) proposes a novel method to operationalize the concept of comparative advantage and its evolution. It uses collaborative filtering algorithms in machine learning most commonly applied to product recommendations in e-commerce, to produce export diversification recommendations that reflect a country's *latent comparative advantages* and potentials in export structure. Section 3 will go over the details of the methodology. But the broad idea is that a country is likely to have comparative advantages in products that are highly related to the products that it is currently exporting (i.e. its revealed comparative advantages), where the "relatedness" between any two products is measured by the similarities of countries that are the main exporters of the two products.

According to Che (2020), the rationale for such an export recommendation algorithm comes from two observations. First, products that require similar production inputs and know-how tend to show up in an export portfolio together. For example, a country that has successfully exported beef can branch into, with some effort, dairy. A country that has mastered the trade of exporting desktop computer hardware is in a better position to produce and export cellphones, than otherwise. Therefore, the products in a country's existing export portfolio contains valuable information regarding what other products the country can get good at producing. Secondly, countries with similar comparative advantages tend to export similar products. Bangladesh and Vietnam are both successful in exporting garments because of the countries' shared abundance in low cost labor. New Zealand and Uruguay both specialize in cattle exports partly because of the high availability of pasture land. In other words, related products to a country's existing exports and export portfolios of *similar countries* contain information about the country's latent comparative advantages, even though the latter cannot always be neatly expressed quantitatively.

Che (2020) found that the export structures recommended by the "Product-based KNN" algorithm successfully predict the evolution of actual export structure for several high-growth countries including China, India, Chile and Poland. Here the *export structure* is measured by the number of Standard International Trade Classification (SITC) 4-digit products recommended by the algorithm that belong to each of the 10 SITC 1-digit sectors, as a share of the total number of recommended products.

It is important to note that the *export portfolio* is a related, but different concept from *export diversification* as commonly understood. A country can double the number of products it exports, i.e. diversification in numbers, without changing its export structure at all, if the sectoral distribution of its exports stays the same. In contrast, if a country used to export 100 products all in the food-stuff sector, but now changes to export 50 products in the food sector and 50 in the machinery sector, it hasn't "diversified" in numbers but export structure has changed. A country's export portfolio can be improved by diversification in the number of export products as well as adjusting of export portfolio. Though both aspects need to revolve around the country's comparative advantages. And the machine learning based export recommendations may provide useful guidance on both.

In this paper, our goal is to test the hypothesis that export product recommendations based on a collaborative filtering algorithm indeed reflect what a country's export structure should be at any given time.

Specifically, we use the product-based KNN algorithm similar to Che (2020) to make annual export product recommendations in the SITC 4-digit product space, for 194 countries over three decades. We then compare the recommended sectoral structure of exports with the actual export structure of each country (see Section 3 for details on methodology). If the export recommendations produced by the algorithm indeed capture countries' latent comparative advantages, we should observe that countries whose export structure closely aligns with the recommended structure would have better growth performance. Here we define "better" as higher growth and lower growth volatility.

A preliminary look at the data appears to support our hypothesis. Figure 1 plots the cross-country correlation for between average real GDP growth per capita over 1985-2015 and the average *similarity score* between a country's actual export structure and recommended export structure produced by the produce-based KNN algorithm.[1] Figure 2 plots the correlation between the 5-year standard deviation of annual growth rate and the similarity score. The charts indicate that countries with an export structure closer to the recommended structure enjoy higher growth and lower growth volatility. The same can be discerned from Figure 3, which presents a positive correlation between similarity score and "risk adjusted" growth, i.e. 5-year average growth divided by standard deviation of growth.

[1] The similarity score is calculated as the Pearson correlation between actual and recommended export structures. Thus it has a theoretical range of [-1, 1]. See Section 3 for details.

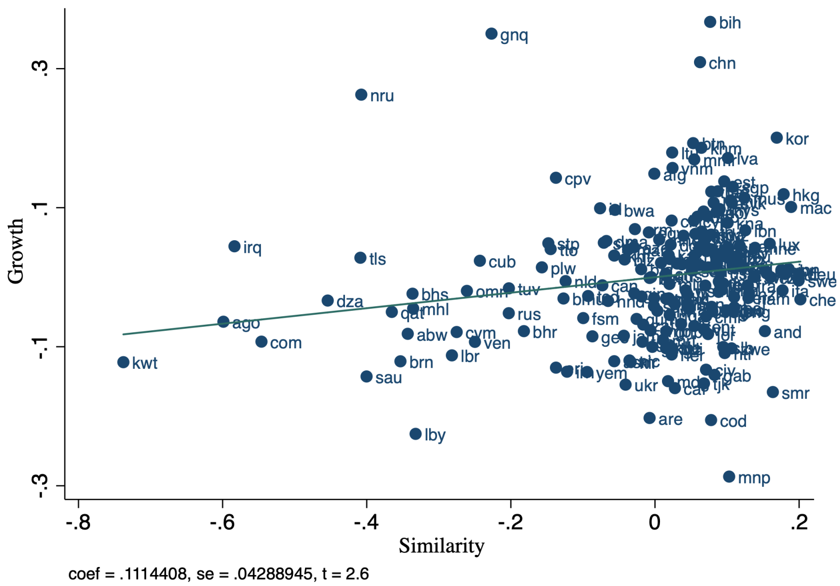

Figure 1: Relationship between "similarity score" and growth

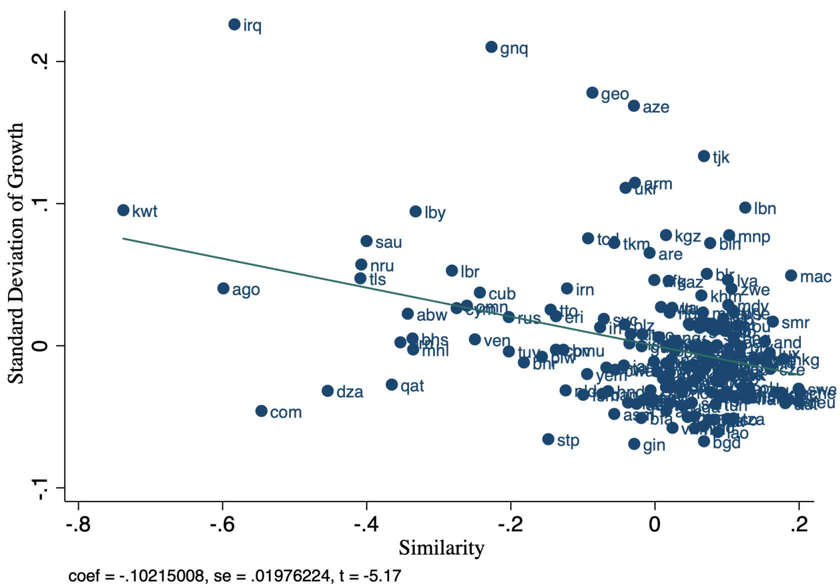

Figure 2: Relationship between "similarity score" and growth volatility

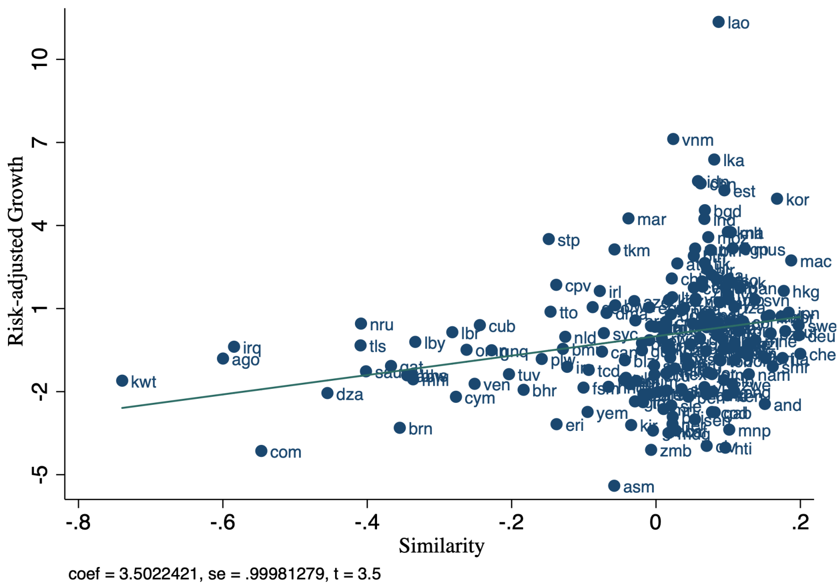

Figure 3: Relationship between "similarity score" and volatility-adjusted growth

It's interesting to look at some country examples as well. Figure 4 plots the evolution of the similarity score between algorithm-recommended export structure and actual export structure for China, Singapore, South Korea, and Germany. Since the 1990s, the similarity score for China has increased significantly, from below the world average to top 3% of the world sample. The magnitude of increase for Singapore is similar. For South Korea, thought the similarity score has dropped somewhat overtime, it is still quite high (top 15%). Likewise, Germany has one of the highest similarity scores in the world, which is unsurprising given the country's diversified and dynamic industrial export base.

Figure 5 plots the evolution of the similarity score for several developing countries with lower growth– Honduras, Kuwait, Libya, and Venezuela. For Libya and Kuwait, the similarity scores are particularly low. Though the score for Kuwait has increased in the past two decades, it stands at around 0.2, compared to the world average of 0.83. For Honduras and Venezuela, the similarity score is higher, but is still below the world level and has dropped significantly in the more recent period, likely reflecting a decline in diversification and manufacturing capacity.

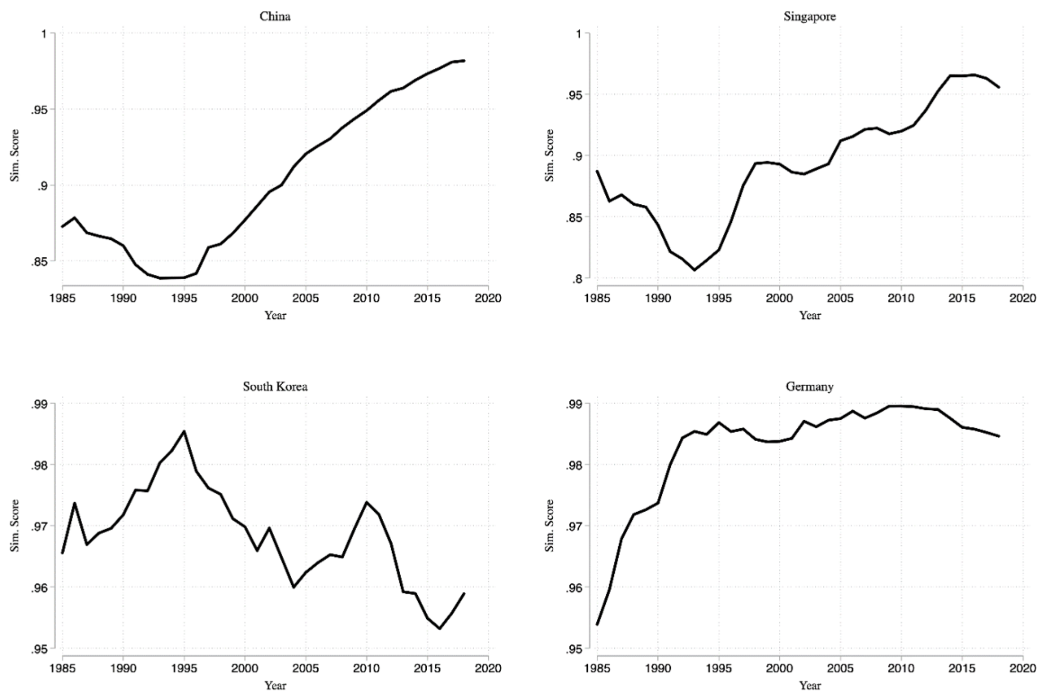

Figure 4: Similarity scores for select high-growth & developed countries

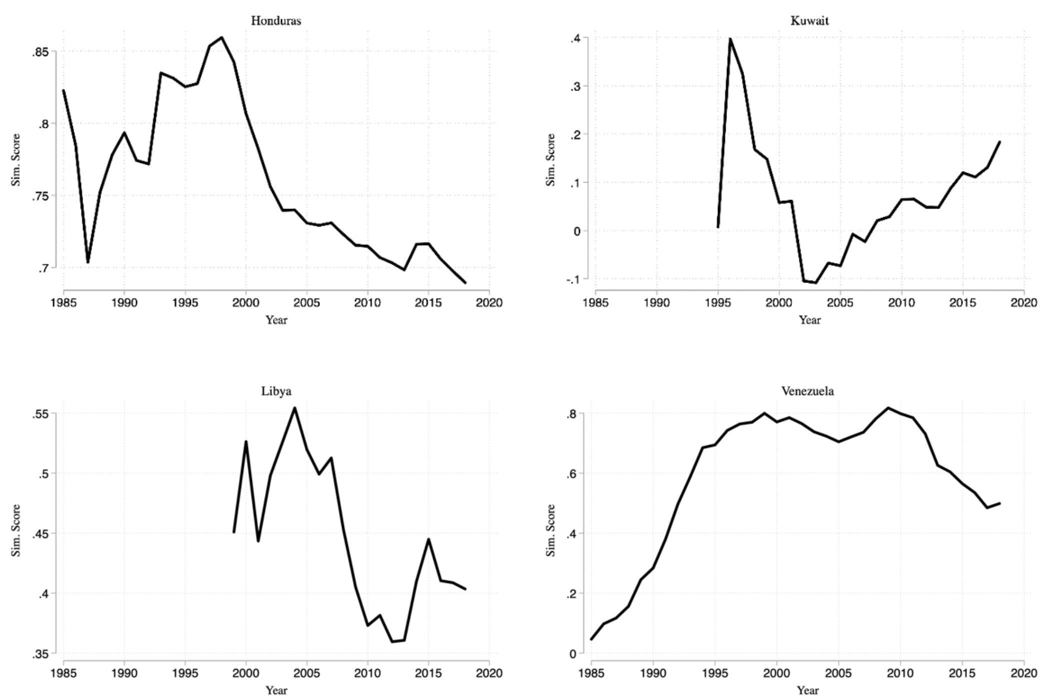

Figure 5: Similarity scores for select low-growth & fragile states

The rest of the paper is organized as follows. Section 2 presents related literature on export structure and diversification. Section 3 describes the product-based KNN algorithm and our empirical methodology. Section 4 describes the data. Sections 5 and 6 present the main empirical results and extensions. Section 7 concludes.

2 Related literature

The literature closest to the present paper is the studies on the so-called *product space* and its implication for diversification and growth (e.g. Hausmann & Klinger 2007, Hidalgo & Hausmann 2009). Like the current paper, this strand of research seeks to understand a country's export structure by looking at the *relatedness* among products.

But there are two key differences. The first is information efficiency. The product-space literature uses a probability formula to represent the relatedness, or *proximity* between two products.[2] While this approach features a clean, easy-to-understand formula and makes the subsequent analysis computationally simpler, it is at the cost of not fully employing the information contained in the data matrix of country-product exports. In contrast, the product-based KNN algorithm in the present paper makes more efficient use of the data to detect the unique blend of characteristics of countries and products. This leads to potentially better recommendations. To be sure, it is at the cost of requiring more computational resources and forsaking the easily comprehensible linear formula. This is a common drawback of many machine learning algorithms– the nonparametric nature of the approach can make some results seem to come out of "magic", harder to explain with linear logic.

The second, more important, difference is one of perspectives. The product-space literature makes specific value judgments about the worthiness of different products for diversification purpose. A product's diversification value is seen to broadly depend on 1) how "complex" it is, meaning, how much sophisticated knowledge is required to product the product, and 2) how

[2]Specifically, the proximity between product A and product B is defined as the probability that a country exports product A given that it exports product B, or vice versa. For example, suppose that 17 countries export wine, 24 export grapes and 11 export both, all with revealed comparative advantage. Then, the proximity between wine and grapes is 11/24 = 0.46.

closely related the product is to other more complex products. Each product is assigned a complexity level as such. The rationale for doing so is a reasonable one– more complex products have higher value-added, use more human capital, face less global competition, and products that are "bridges" to more complex products may be a pathway for a country to move up the international value chain. Some empirical evidence shows that diversifying into these products is supposed to be better for growth (Hausmann, 2007). However, several issues emerge when this model is used for recommending export products to specific countries. First, there is an underlining tension between this line of thinking and the framework of comparative advantages that the product-space analysis is built on. By assigning each product a score of virtue (e.g. industrial products are good, agro commodities are bad), it leads to a tendency to recommend products that the model deems universally worthy to countries of drastically different fundamentals. In the extreme– though improbably– scenario where all countries internalize the same worthiness ranking of products for developing their export structure, there would be no comparative advantages to speak of. Secondly, to come up with a tractable, universally applied scoring system for "product complexity", strong assumptions need to be made that reduce the feature dimensions of reality and throw away valuable country- and product- specific information, which may limit the model's usefulness in producing realistic export recommendations for individual countries.

In contrast, the approach of the current paper is agnostic regarding the diversification value of any specific product. Instead, we seek to fully exploit the information contained in the country-product space, and make realistic export recommendations off of a country's current revealed comparative advantages. One implication is that countries do not necessarily need to chase the "complex" exports to achieve better growth performance. As Section 5 shows, countries whose export structure closely aligns with the algorithm-recommended structure have higher and more stable growth, even though the algorithm's recommendations do not make any judgment regarding product complexity, and are solely based on information from a country's currently revealed comparative advantages.

The paper is also related to the literature on the relationship between export diversification and countries' economic performance. Existing research asserts that export diversification is a key element in the economic development process, particularly for developing and emerging market countries trying to catch up with their advanced peers. Various studies provide evidence of a positive association between export diversification and economic development (e.g. Imbs and Wacziarg, 2003; Klinger and Lederman 2004 and 2011; Cadot et al., 2011). Numerous country studies also supports the benefits of export diversification. For example, Feenstra and Kee (2008) use data from a large set of countries exporting to the US, to show that a sustained increase in export diversification results in increases in productivity and a notable increase in the GDP of the exporters. IMF (2014) finds that diversification in exports and in domestic production has been conducive to faster economic growth in LICs. Al-Marhubi (2000) provides similar findings within a set of developing economies. Balaguer and Cantavella-Jorda (2004) find that export variety plays a key role in Spain's economic development. And Herzer and Danzinger (2006) report a positive impact of export diversification on economic growth of Chile. Research also points to a positive association between export diversification and macroeconomic stability (e.g. IMF, 2014).

However, not all types of diversification are created equal, and diversification for its own sake is hardly a recipe for sustainable growth. A foundational idea of the classical international trade theory is that under free trade, countries will tend to export what they are relatively good at producing, i.e. products they have a comparative advantage in. "Diversifying" into industries that are misaligned with a country's current endowment fundamentals, as the former Soviet-block nations did after World War II through industrial policies that aimed to accelerate industrialization, has negative growth consequences (see e.g. Lin, 2009). On the other end of the spectrum, delayed industrialization also leads to negative growth outcomes, as the experience of many resource-rich countries that are entrenched in their over-dependence on commodity exports has shown (e.g. Frankel, 2010).

A difference in focus between the current paper and the export diversification literature is that the latter sees diversification as mostly in increasing the number of export products, while the current paper emphasizes on adjusting the structure of exports. Our algorithm does provide a list of recommended products for each country, which provides useful insights for countries looking to increase the number of export items. But our econometric exercise focuses on the growth impact of the right export structure, i.e. sectoral distribution of exports.

3 Methodology

Our goal in this paper is to answer the question of whether our algorithm-based export recommendations can produce an growth-enhancing export structure, in the sense that countries that follow the recommendations could achieve better growth performance. We go about answering this question in the following steps:

- **STEP 1.** Choose the number of SITC 4-digit products to recommend for each country (see Section 3.1). This number is derived from a country's size and development level.
- **STEP 2.** Generate a list of recommended export products for each country-year in the sample using a product-based KNN algorithm (see Section 3.2).
- **STEP 3.** Calculate the *similarity score* between the export structure implied by the list of recommended export products and the actual export structure, for each country-year (see Section 3.3).
- **STEP 4.** Estimate the impact of the similarity score on growth and volatility of growth (see Section 3.4).

An important concept used throughout the paper is *Revealed Comparative Advantage* (RCA). The RCA indicator, first introduced by Balassa & Noland (1965), is a popular measure to calculate the relative importance of a product in a country's export basket. Formally, the RCA score of country i in product j can be calculated as:

$$RCA_{ij} = \frac{E_{ij}/E_i}{E_j/\sum_{i' \in I} E_{i'}}$$

where E_{ij} is the export value of product i from country j, E_i is the total export values of country i, E_j is the total exports of product j from all countries around the world, and $\sum_{i' \in I} E_{i'}$ is the total world exports.

Throughout the paper, a *high-RCA product* for country i is defined as a product with its $RCA_{ij} > 1$. Mathematically, it means that the product's share in the country's export portfolio is greater than its share in the total world exports, which can be seen as an indication that the country has a comparative advantage in the product. For example, vehicle exports were about 12 percent of total world exports in 2017, while they constituted 22 percent of total exports from Mexico. Therefore, $RCA_{ij} = 22/12 = 1.8$ for Mexico's vehicle exports in 2017. Since it is $>$ 1, according to our criteria, Mexico has a revealed comparative advantage in automobiles. Or to put it another way, automobiles is a high-RCA product for Mexico. The recommendation algorithm that will be introduced in Section 3.2 essentially simulates a hypothetical RCA score for each country-product, and pick the top n products with the highest hypothetical scores as the recommended export portfolio for country i.

3.1 Choosing the number of recommended products

Examining the export data by SITC 4-digit industry.[3] reveals the following empirical regularities. First, more developed economies tend to have a larger number of high RCA products. Figure 6[4] regresses the number of high RCA exports of each country on its real GDP per capita relative to the US level, controlling country size. Secondly, bigger countries tend to have a larger number of high RCA exports. This is unsurprising, as population size correlates highly with the number of firms, the amount of human capital and the amount of other production resources a country may have, enabling the country to viably export a wider range of products. In addition, some industries and products need a minimum scale to be sustainable. Figure 7 plots this positive relationship between number of high RCA exports and country population.

There are obviously other factors that determine how many high RCA exports a country has.

[3]See Section 4 for a more detailed description of the underlining data.

[4]This chart reproduces Figure 3 of Che (2020).

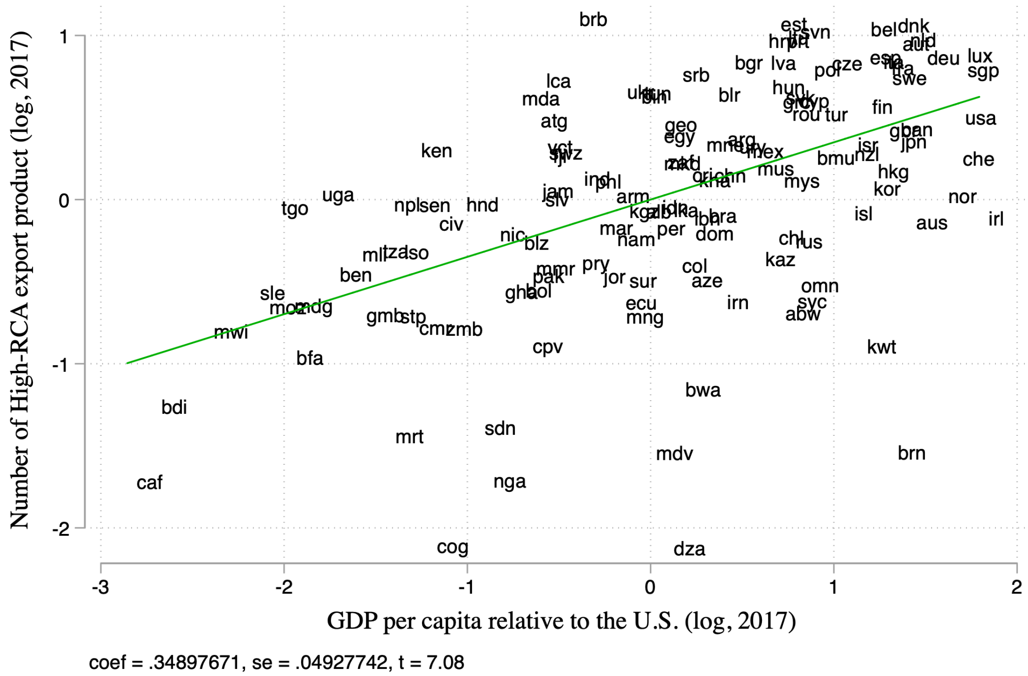

Figure 6: Number of High-RCA Exports v.s. Income Level, Partial Regression Plot

But since we are focusing on exploring export structure adjustment instead of just expanding the number of export products in this paper, we pick the number of recommended export products for each country as predicted by the country's size and development stage. Specifically, we run the following estimation:

$$N_{rca,it} = \beta_1 GDP_{it} + \beta_2 POP_{it} + \gamma_t + \epsilon_{it} \tag{1}$$

where $\hat{N_{rca,it}}$ is the estimated number of high RCA exports that country *i* is expected to have in year *t*. *GDP* is GDP per capita and *POP* is population size. We add a time fixed effect γ_t in the regression, as the average number of high RCA exports tends to rise overtime around the world, with the increase in product variety brought about by economic growth.

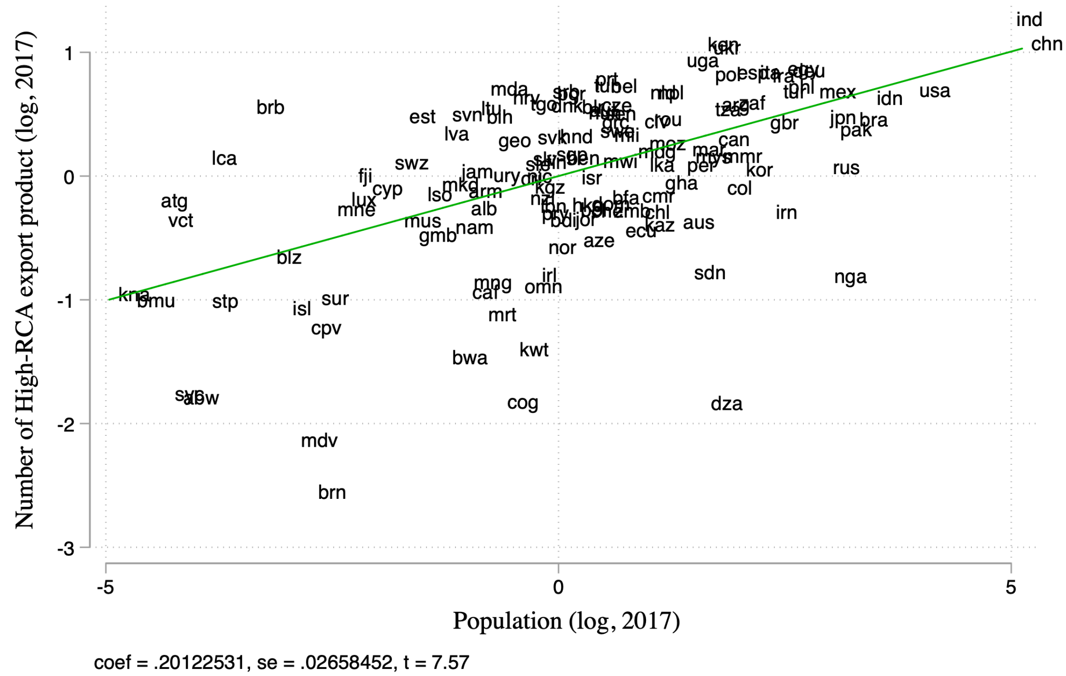

Figure 7: Number of High-RCA Exports v.s. Population, Partial Regression Plot

3.2 The recommendation algorithm

Our export product recommendation system employs a product-based K-nearest neighbor (KNN) algorithm that is widely used in the collaborative filtering recommendation systems of online commerce.[5] The goal of the exercise is to generate, for each country-year, a list of "top-N recommendations", i.e. N products that a country should export the most of. The algorithm produces the list by estimating the hypothetical RCA scores (which we call *“recommendation scores”* later) of different products for the underling country, using the training dataset of export values by country and SITC 4-digit product, and recommending the N products with the highest recommendation scores. Here, we set N as equal to the $\hat{N_{rca,it}}$ estimated in Section 3.1 for each country i in year t.

The underlining data used in the recommendation algorithm can be represented as a $m \times n$ matrix R, where m is the number of countries in the database, and n is the total number of SITC 4-digit products. The content of R, i.e. r_{ij}, is country i's RCA score in product j. R is a

[5]See Che (2020) for detailed explanations of other similar algorithms.

sparse matrix due to the fact that each country only exports a subset of the products in the SITC universe. In the case that country i does not export any product j, $r_{ij} = 0$. If in running the algorithm, multiple years of export data are used as the training set, then each country-year is a row in R, i.e. $m = c \times y$, where $c =$ the number of countries in the dataset, and $y =$ the number of years included. In practice, we set $y = 1$, i.e. when we're generating export recommendations for country i in 2017, only the cross-country export data for 2017 is included in the training set.[6]

KNN is one of the most frequently used methods in solving classification and pattern recognition problems, and is a popular approach in constructing recommender systems. The basic idea of KNN is learning by analogy– classifying the test sample by comparing it to the set of training samples most similar to it. Different KNN implementations vary in terms of their choices of how the similarity between input vectors is calculated. In the present paper, the cosine similarity score is used as the similarity measure.

The intuition behind the product-based KNN implementation is simple– first look at what products a country already has a revealed comparative advantage in, and then recommend other products that are "related" to those products. To explain the approach in more details, let's first rewrite the RCA score matrix R as:

$$R = \left[\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_n\right]$$

where $\mathbf{p}_j$ is a vector of length m that represents the RCA scores of product j for all the m countries in the sample:

$$\mathbf{p}_j = \begin{bmatrix} r_{1j} \\ r_{2j} \\ . \\ . \\ r_{mj} \end{bmatrix}$$

[6] We experimented with including multiple years of data in the training set, but found no significant improvement in the results, while the model took longer to compute as the size of m increases.

In machine learning terminology, each product in the sample has m features. The cosine similarity between products i and j is equal to $\frac{\mathbf{p}_j \cdot \mathbf{p}_l}{\|\mathbf{p}_j\| \|\mathbf{p}_l\|}$, which ranges from -1, when the two vectors are the exact opposite, to 1, when the two are exactly the same. The intuition behind this is that by comparing the two sets of countries that export i and j, and how important the products are in the countries' export baskets, information can be inferred regarding how closely related the two products are.

The implementation of the product-based KNN recommender for country i in year *t* involves the following steps:

1. Represent each product in the SITC 4-digit product space as a vector of RCA scores, $\mathbf{p}_j$.
2. Select the set of products that country i has a revealed comparative advantage, i.e. $r_{ij} > 1$. Let's call it the high-RCA product set of country i.
3. For each $j \in [1, n]$, calculate the predicted value of r_{ij} as a weighted average RCA score of the high-RCA product set, weighted by the cosine similarity between product j and the products in the country's high-RCA set.
4. The recommended products for country i are the $\hat{N_{rca,it}}$ products with the highest predicted r_{ij} values (i.e., recommendation scores), where $\hat{N_{rca,it}}$ comes from the estimation in Section 3.1.

We repeat the above steps for each country-year to generate the recommended export portfolio in terms of SITC 4-digit products for every country in each sample year.

3.3 Calculating similarity scores

For the next step, we compute the similarity between the actual export portfolio of a country and the recommended export portfolio.

We define the portfolio structure of country *i*'s actual exports in time *t* as the the number of high RCA exports (on SITC 4-digit level) that belong to each SITC 1-digit sector,[7] as a share of

[7]See appendix Table 17 for the full list of SITC 1-digit sectors

total number of high RCA exports. In other words, let $N_{k,it}^{actual}$ be the number of high RCA exports in sector *k*, and

$$s_{k,it}^{actual} \equiv \frac{N_{k,it}^{actual}}{\sum_l N_{l,it}^{actual}}$$

is the share of the number of high RCA exports that belong to sector *k* in the total number of high RCA exports. Country *i*'s export structure, S_{it}^{actual}, is thus defined as a $K \times 1$ vector: $\left[s_{k,it}^{actual}\right]_{K\times1}$, where $K = 10$, is the number of SITC 1-digit sectors.

Similarly, we define the recommended export structure $S_{it}^{rec} \equiv \left[s_{k,it}^{rec}\right]_{K\times1}$, as the vector for the number of recommended products that belong to each SITC 1-digit sector as a share of the total number of recommended export products.

The similarity score between the actual and the recommended export portfolios for country *i* at time *t* is then calculated as the similarity between the two vectors of actual and recommended structures:

$$Sim_{it} \equiv \frac{\left(S_{i,t-\Delta t}^{rec} - \bar{S}_{i,t-\Delta t}^{rec}\right) \cdot \left(S_{it}^{actual} - \bar{S}_{i,t}^{actual}\right)}{\|S_{i,t-\Delta t}^{rec} - \bar{S}_{i,t-\Delta t}^{rec}\| \|S_{it}^{actual} - \bar{S}_{i,t}^{actual}\|} \tag{2}$$

Che (2020) found that recommendations given by the product-based KNN algorithm are to some extent forward-looking, in that they match the export portfolios of several high-growth countries in their future years better than in the current year. Therefore, we include a time lag, Δt, in calculating the similarity scores, to account for the fact that it takes time for an export structure to evolve. In our baseline estimation, we set Δt = 5 years. Alternative assumptions for the time window are also adopted as robustness checks (see Section 6).

We calculate the annual Sim_{it} for all countries in the sample, and then incorporate the scores into the growth/volatility regressions that will be specified in the following section.

3.4 Growth and volatility estimations

Our main econometric exercise aims to investigate the impact of export structure on growth and volatility of growth. Our hypothesis is that countries with an export structure highly aligned with their latent comparative advantages– indicated by a high similarity score as defined in Section 3.3– should see higher and more stable growth over time.

To examine the impact of export structure on growth, we specify the following estimation model:

$$g_{it} = \beta_0 + \beta_1 y_{i,t-\Delta t} + \beta_2 Sim_{it} + \gamma \mathbf{X}_{it} + \epsilon_{it} \tag{3}$$

where g_{it} the average annual growth in GDP per capita for country i from $t-\Delta t$ to t. $y_{i,t-\Delta t}$ is the lagged real GDP per capita in log form. Sim_{it} is the similarity score calculated as in Section 3.3. $\mathbf{X}_{it}$ is a set of controls, including investment-to-GDP ratio, human capital growth, TFP growth, and world GDP growth in some specifications. We also include country and time fixed effects in $\mathbf{X}_{it}$ for some of the regression specifications (see Section 5). The similarity scores, as well as most control variables, are annual averages over the Δt time window. In our baseline estimation, we set Δt = 5 years. The regressions are run with non-overlapping Δt as the time unit. Our main parameter of interest is β_2.

Similarly, we can look at the impact of export structure on the volatility of growth with the following model:

$$vol_{it} = \beta_0 + \beta_1 vol_{i,t-\Delta t} + \beta_2 Sim_{it} + \gamma \mathbf{X}_{it} + \epsilon_{it} \tag{4}$$

where vol_{it} is the standard deviation of annual growth of real GDP per capita during the Δt time period. And $Vol_{i,t-\Delta t}$ is the lagged dependent variable. Controls ($\mathbf{X}_{it}$) are broadly the same as in the growth regression, except we replace world growth with the growth volatility of world GDP, to control for the level of external volatility.

Alternatively, we can combine the information on the left-hand side of Equations 3 and 4, and estimate the impact of export structure on countries' "risk-adjusted growth". Here we define country i's risk-adjusted growth, g_{it}^{ra}, as the deviation of country i's growth from the world average growth rate, $g_{it} - \overline{g}_t$, divided by its standard deviation σ_{it}, over the Δt time period. We then estimate the following equation,

$$g_{it}^{ra} = \beta_0 + \beta_1 g_{i,t-\Delta t}^{ra} + \beta_2 Sim_{it} + \gamma \mathbf{X}_{it} + \epsilon_{it} \tag{5}$$

Controls ($\mathbf{X}_{it}$) are the same as growth and volatility regressions. Note that in system GMM regressions, except we now include the average risk-adjusted growth across countries in $\mathbf{X}_{it}$.

For each equation, estimation is done using simple OLS, country and time fixed-effect estimator, and a system GMM estimator following Arellano and Bond (1991). The system GMM estimator is employed to address the endogeneity issue introduced by having the lagged dependent variable on the right hand side, which likely affects the consistency of OLS and fixed-effect estimators. In the system GMM estimation, the lagged dependent variable and country-level controls are treated as endogenous and instrumented as such. Time fixed effect and world-level controls are treated as exogenous. Section 5 presents results from all three estimators for each regression.

4 Data

The country-product level export data, including the RCA scores, come from the Atlas of Economic Complexity Dataverse (2020 version), which in turn sourced the data from UN COMTRADE. The macroeconomic variables comes from World Bank and Penn World Table[8]. Summary statistics for the main variables used in the regressions are shown in Table 1. The data is on annual frequency covering 1980-2018.

The *similarity score* is calculated at country-year level, following the steps described in Section 3.3. In the appendix, we show summary statistics for RCA scores and recommendation scores used to calculate the similarity score (see Table 18 and Table 19), as well as the box plots for the distributions of RCA scores and recommendation scores by SITC 1-digit sector (see Figure **??** and

[8]Version 10.0, see Feenstra et al. 2015 for metadata details.

Table 1: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Similarity score	0.829	0.196	-0.403	0.995	1202
GDP per capita	8.390	1.505	5.129	11.663	1414
Investment rate	0.219	0.109	-0.479	0.942	1398
TFP growth	0.002	0.028	-0.184	0.222	868
Human capital growth	0.01	0.007	-0.025	0.043	1107
GDP per capita growth	0.017	0.037	-0.247	0.367	1324
Growth volatility	0.015	0.014	0.001	0.142	1177
Risk-adjusted growth	0.705	4.474	-75.438	50.674	1175

Figure 9). Figure 8 shows the similarity score distribution around the world in 2018.

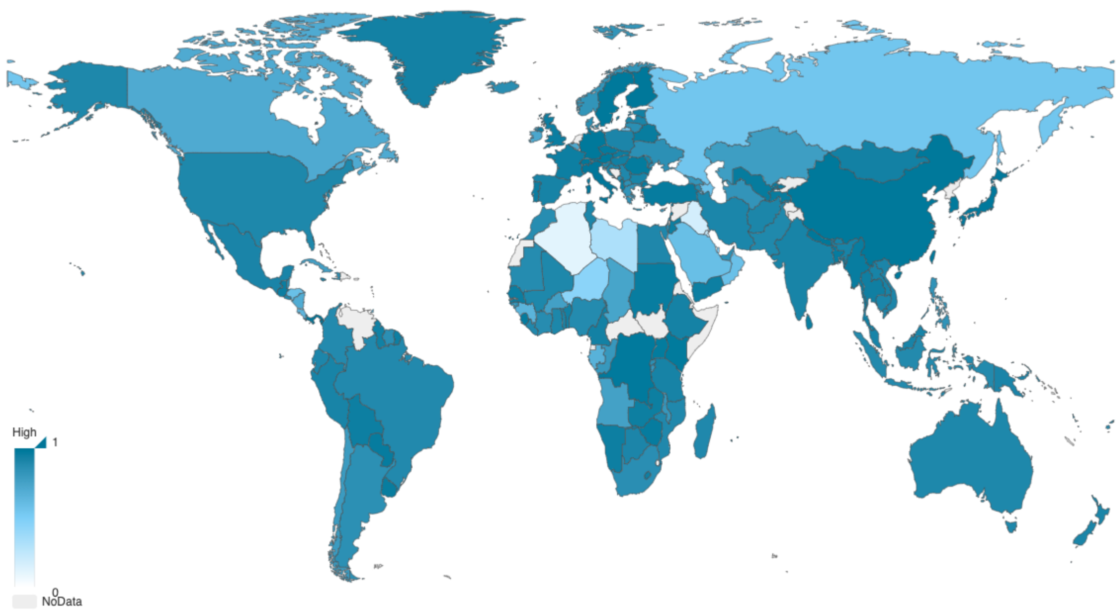

Figure 8: World Similarity Scores

5 Estimation Results

5.1 Impact on growth

We show baseline results in Table 2 below. Column (1) and (2) are for OLS without and with controls respectively (with robust standard errors). Column (3) is for fixed effects with with clustered standard errors. Column (4) and (5) shows system GMM results with year dummies and average world GDP growth respectively. According to the baseline results, a 0.1 increase in similarity score level (Sim_{it}) is associated with a 0.17 percentage points increase in real GDP per capita annual growth. The coefficient is statistically significant at 1% level. An example for a 0.1 increase in similarity score level is the change of China's similarity score from around 2001 (the accession to WTO) to 2018. This is also equivalent to a move from the median to 90 percentile of the similarity distribution. We explored $\Delta t = 3, 7$ for robustness in Section 6.

5.2 Impact on growth volatility

Volatility regression results are shown in Table 3. Similar with the structure in growth regression results table, column (1)-(3) are for OLS and fixed effects estimation. And column (4)-(5) are system-GMM results, with year dummies and with average world GDP growth respectively. The last column suggests that a 0.1 increase in similarity score level is associated with 0.0011 decrease in standard deviation in a country's growth rates in a 5-year rolling window. This is statistically significant at 5% level. In robustness section, we explored $\Delta t = 3, 7$. The signs and significance are all preserved.

5.3 Impact on risk-adjusted growth

Risk-adjusted growth regression results are shown in Table 4. Similar with the structure in growth regression results table, column (1)-(3) are for OLS and fixed effects estimation. Same with growth and volatility regressions, column (4)-(5) are system-GMM results, with year dummies and with average world GDP growth respectively. The last column suggests that a 0.1 increase in similarity score level is associated with 26.57 percentage points increase per standard deviation in annual growth rates. Another way to look at the magnitude of impact is that an increase of 0.8 in the similarity score will move a country from the world medium in risk-adjusted growth to the 75 percentile level. This is statistically significant at 5% level. In robustness section, we explored $\Delta t = 3, 7$. The signs and significance are all preserved. Results with more foresight time horizon can be found in the appendix.

6 Robustness

6.1 Changing time interval

First, we explore the results with different time horizons by varying Δt. The following tables show results with $\Delta t = [3, 7]$ under OLS, FE and system GMM specifications.

According to the growth regression results shown in Table 5, the signs and significance for the coefficient of interest (Sim_{it}) are preserved from the baseline case ($\Delta t = 5$). When $\Delta t = 3$, a 0.1 increase in the similarity score is associated with 0.21 percentage points increase in growth, while this value is 0.16 when $\Delta t = 7$. Details for the case of $\Delta t = 3, 7$ can be found in Table 8 and Table 9 in the appendix.

Volatility regression results are shown in Table 10 and Table 11. A 0.1 increase in similarity score is associated with 0.0014 and 0.0013 decrease in standard deviation when $\Delta t = 3, 7$ respectively. These are almost consistent with the baseline result.

Adjusted-growth regression results are shown in Table 12 and Table 13. When $\Delta t = 3$, a 0.1 increase in similarity score is associated with 21.11 percentage points increase in risk-adjusted annual growth rate at 1% significance level. However, this value decreases to 11.17 when $\Delta t = 7$, and is not statistically significant. We think this could be a reflection of the validity of time horizon in product-RCA based collaborative filtering. When time horizon increases, there might be more significant changes in a country's fundamentals and structural changes. Furthermore, shocks and uncertainties in both domestic and global market may affect the denominator for risk-adjusted growth measure thus the coefficients turns out to be insignificant, although still positive. Quantifying the relative importance of the growth resistance and the volatility resistance for risk-adjusted growth is beyond the scope of this research. But we would like to reserve it for future studies.

6.2 Winsorization

We also conduct a set of regression where we winsorize the dependent variables by trimming the top 5% and bottom 5%. In general, signs and significance are mostly preserved for the coefficients of Similarity Score variable.

In Table 14, we show results of growth regression with real growth rates winsorized. According to column (5), a 0.1 increase in similarity score (Sim) is associated with a 0.18 percentage points increase in real GDP per capita growth, similar to the baseline result.

In Table 15, we show results of volatility regression with volatility winsorized. According to column (5), a 0.1 increase in similarity score is associated with a 0.06 decrease in standard deviation of real GDP growth over a 5-year period, similar to the baseline result.

In Table 16, we show results of adjusted-growth regression with real growth rates winsorized. According to column (5), a 0.1 increase in similarity score is associated with a 0.0834 percentage point per standard deviation risk increase in annual GDP per capita growth. This value is much smaller than the baseline regression result (0.2657). Statistical significance level also changes from 1% to 10%.

7 Conclusion

One of the frequently voiced complaints from economists and policy makers regarding the use of machine learning algorithms in empirical studies is the seeming opaque nature of the algorithms. The human cognitive system can differentiate a picture of a dog from that of a cat easily. But there is very little theory, i.e. a linear and logical explanation, behind why such discernment can be reliably made. Many machine learning algorithms share the same characteristics. These algorithms are very effective in making realistic pattern-recognition judgements, but an articulated rationale of such judgements is often lacking. On the other hand, traditional parametric econometric studies are under-pined by economic theories with easy-to-understand trains of thought. But the typical linear regression models are drastic simplifications of reality, which may reduce their usefulness in guiding practical decisions.

In this paper we try to combine the best of both worlds to shed some light on the importance of export structure evolution in the growth and income convergence process. We leveraged machine learning methodology to characterize the complex patterns in countries' latent comparative advantages and issue export recommendations accordingly. We then use a standard linear regression model to evaluate the merits of the recommendations by asking whether the countries' growth performance is better if they had de facto "followed" these recommendations.

Specifically, we used a product-based K-nearest neighbor (KNN) algorithm to provide annual export product recommendations at the SITC 4-digit level for over 190 economies, from 1980 to 2018. We then test whether more alignment between a country's recommended export structure and its actual export structure has any impact on growth and growth volatility.

Our results confirm the value of such algorithm-based export recommendations. They show that economies with a higher *similarity score* between recommended and actual export portfolios achieve better growth performance. In our baseline estimation, a 0.5 increase in the similarity score is associated with a 0.85 percentage point increase in the annual growth of real GDP per capita, and a 0.005 decrease in the standard deviation of growth rate over 5-year time windows. These results are overall robust with respect to changing time horizons and removing outliers.

It's worth noting that although we believe the algorithm-produced export recommendations can be a useful tool for policy makers to evaluate industrial policy options and for private investors entering new markets, they are no substitutes for detailed and multidimensional analyses of the viability of any industry in a country. In addition, it goes without saying that knowing which industries a country may have a comparative advantage in does not automatically translate into specific policy recommendations. Neither are we advocating for direct policy interventions in shaping a country's export structure. How a country can best support the growth of tradable sectors that leverage its comparative advantages is likely a case-by-case discussion and depends on various country-specific factors. Nonetheless, We believe that knowing what a country's ideal export structure may look like and comparing it with the reality is a valuable exercise for policy makers to identify potential policy gaps and reform areas to focus on for achieving better growth.

References

Adeniyi, David Adedayo, Zhaoqiang Wei, and Y Yongquan. "Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method". In: *Applied Computing and Informatics* 12.1 (2016), pp. 90–108.

Aiginger, Karl and Dani Rodrik. "Rebirth of Industrial Policy and an Agenda for the Twenty-First Century". In: *Journal of Industry, Competition and Trade* (), pp. 1–19.

Bacha, Edmar L and Albert Fishlow. "The recent commodity price boom and Latin American growth: More than new bottles for an old wine?" In: *The Oxford Handbook of Latin American Economics*. 2011.

Balaguer, Jacint and Manuel Cantavella-Jorda. "Structural change in exports and economic growth: cointegration and causality analysis for Spain (1961–2000)". In: *Applied Economics* 36.5 (2004), pp. 473–477.

Cadot, Olivier, Céline Carrère, and Vanessa Strauss-Kahn. "Export diversification: what's behind the hump?" In: *Review of Economics and Statistics* 93.2 (2011), pp. 590–605.

Che, Tascha. "Intelligent Export Diversification: An Export Recommendation System with Machine Learning". In: *Tascha Labs* 2020 (2020).

Clickstreams, Multi-faceted Web. "Workshop Notes". In: (2005).

Feenstra, Robert and Hiau Looi Kee. "Export variety and country productivity: Estimating the monopolistic competition model with endogenous productivity". In: *Journal of international Economics* 74.2 (2008), pp. 500–518.

Frankel, Jeffrey A. *The natural resource curse: a survey*. Tech. rep. National Bureau of Economic Research, 2010.

Giri, Rahul, Mr Saad N Quayyum, and Rujun Yin. *Understanding Export Diversification: Key Drivers and Policy Implications*. International Monetary Fund, 2019.

Hausmann, Ricardo, Jason Hwang, and Dani Rodrik. “What you export matters”. In: *Journal of economic growth* 12.1 (2007), pp. 1–25.

Hausmann, Ricardo and Bailey Klinger. “The structure of the product space and the evolution of comparative advantage”. In: *CID Working Paper Series* (2007).

Herzer, Dierk and Felicitas Nowak-Lehnmann D. “What does export diversification do for growth? An econometric analysis”. In: *Applied economics* 38.15 (2006), pp. 1825–1838.

Hidalgo, César A and Ricardo Hausmann. “The building blocks of economic complexity”. In: *Proceedings of the national academy of sciences* 106.26 (2009), pp. 10570–10575.

Imbs, Jean and Romain Wacziarg. “Stages of diversification”. In: *American Economic Review* 93.1 (2003), pp. 63–86.

IMF. “Sustaining Long-Run Growth and Macroeconomic Stability in Low-Income Countries-The Role of Structural Transformation and Diversification”. In: *IMF Policy Paper* (2014).

Klinger, Bailey and Daniel Lederman. *Discovery and development: An empirical exploration of “new” products.* The World Bank, 2004.

— “Export discoveries, diversification and barriers to entry”. In: *Economic Systems* 35.1 (2011), pp. 64–83.

Koren, Yehuda, Robert Bell, and Chris Volinsky. “Matrix factorization techniques for recommender systems”. In: *Computer* 42.8 (2009), pp. 30–37.

Lathia, Neal, Stephen Hailes, and Licia Capra. “kNN CF: a temporal social network”. In: *Proceedings of the 2008 ACM conference on Recommender systems.* 2008, pp. 227–234.

Lin, Justin Yifu and Feiyue Li. *Development strategy, viability, and economic distortions in developing countries.* The World Bank, 2009.

Al-Marhubi, Fahim. “Export diversification and growth: an empirical investigation”. In: *Applied economics letters* 7.9 (2000), pp. 559–562.

Paterek, Arkadiusz. “Improving regularized singular value decomposition for collaborative filtering”. In: *Proceedings of KDD cup and workshop.* Vol. 2007. 2007, pp. 5–8.

Prebisch, Raul. “The economic development of Latin America and its principal problems”. In: *Economic Bulletin for Latin America* (1962).

Sarwar, Badrul et al. *Application of dimensionality reduction in recommender system-a case study.* Tech. rep. Minnesota Univ Minneapolis Dept of Computer Science, 2000.

— “Incremental singular value decomposition algorithms for highly scalable recommender systems”. In: Citeseer.

Singer, Hans W. “The distribution of gains between investing and borrowing countries”. In: *The Strategy of International Development*. Springer, 1975, pp. 43–57.

Singer, HW. “The Distribution of Gains between Investing and Borrowing Countries”. In: *The American Economic Review* 40.2 (1950), pp. 473–485.

Table 2: Growth Regression (baseline)

	(1) OLS	(2) OLS-X	(3) FE(t),X	(4) GMM-(t)	(5) GMM-X(t)
Lagged GDP Per Capita	-0.001**	-0.008***	-0.018***	0.000	0.001
	(-2.126)	(-3.140)	(-4.213)	(0.166)	(0.544)
Similarity Score	0.028***	0.023**	0.018*	0.017**	0.017***
	(3.471)	(2.364)	(1.728)	(2.480)	(2.659)
Inv. Rate		0.099***	0.092***	0.129***	0.125***
		(3.753)	(3.434)	(4.160)	(4.274)
TFP Growth		0.725***	0.693***	0.764***	0.771***
		(10.289)	(11.548)	(12.294)	(12.953)
Human Capital Growth		0.423***	0.422***	0.591***	0.591***
		(3.276)	(3.115)	(3.738)	(3.509)
World Growth		0.446**			0.341
		(2.189)			(1.632)
No. of Obs.	1168	732	732	732	732
AR1 (p-value)				0.00	0.00
AR2 (p-value)				0.01	0.02
Hansen-J (p-value)				0.78	0.76

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Volatility Regression (baseline)

	(1) OLS	(2) OLS-X	(3) FE(t),X	(4) GMM-(t)	(5) GMM-X(t)
Lagged Growth Volatility	0.456***	0.273***	0.272***	0.517***	0.500***
	(6.249)	(3.763)	(4.329)	(7.101)	(7.094)
Similarity Score	-0.006*	-0.019	-0.019	-0.010***	-0.011***
	(-1.795)	(-1.508)	(-1.582)	(-2.636)	(-2.745)
Inv. Rate		0.000	0.000	0.001	-0.000
		(0.019)	(0.037)	(0.072)	(-0.012)
TFP Growth		-0.017	-0.019	-0.023	-0.020
		(-0.462)	(-0.762)	(-0.921)	(-0.815)
Human Capital Growth		0.124	0.130	0.106	0.080
		(1.593)	(1.500)	(1.504)	(1.027)
World Volatility		0.644***			0.495**
		(3.346)			(2.572)
No. of Obs.	966	614	613	614	614
AR1 (p-value)				0.00	0.00
AR2 (p-value)				0.58	0.55
Hansen-J (p-value)				0.44	0.41

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Risk-adjusted Growth Regression (baseline)

	(1) OLS	(2) OLS-X	(3) FE(t),X	(4) GMM-(t)	(5) GMM-X(t)
Lagged Adj. Growth	0.306***	0.065	0.079	0.063	0.046
	(2.733)	(1.474)	(1.635)	(1.124)	(0.903)
Similarity Score	2.971***	1.504	1.617	2.665***	2.657***
	(5.263)	(0.985)	(1.005)	(2.953)	(2.743)
Inv. Rate		15.894***	17.117***	14.728***	12.872***
		(4.214)	(4.454)	(3.812)	(2.963)
TFP Growth		35.374***	40.265***	37.469***	29.930***
		(3.814)	(3.390)	(3.754)	(3.239)
Human Capital Growth		3.496	-9.245	-25.568	-18.241
		(0.122)	(-0.339)	(-0.828)	(-0.548)
World Adj. Grwoth		0.559***			0.573**
		(2.617)			(2.437)
No. of Obs.	964	613	612	613	613
AR1 (p-value)				0.02	0.02
AR2 (p-value)				0.35	0.12
Hansen-J (p-value)				0.59	0.45

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Robustness: Growth

	(1)	(2)	(3)
	+3	+5	+7
Lagged GDP Per Capita	-0.000	0.001	-0.000
	(-0.033)	(0.544)	(-0.204)
Similarity Score	0.021***	0.017***	0.016**
	(3.120)	(2.659)	(2.101)
Inv. Rate	0.130***	0.125***	0.123***
	(5.321)	(4.274)	(3.806)
TFP Growth	0.806***	0.771***	0.773***
	(10.979)	(12.953)	(8.777)
Human Capital Growth	0.650***	0.591***	0.380
	(3.799)	(3.509)	(1.578)
World Growth	0.433***	0.341	0.606*
	(4.764)	(1.632)	(1.889)
Constant	-0.039***	-0.040***	-0.032**
	(-3.027)	(-3.068)	(-2.068)
No. of Obs.	1258	732	523
AR1 (p-value)	0.00	0.00	0.01
AR2 (p-value)	0.87	0.02	0.18
Hansen-J (p-value)	1.00	0.76	0.01

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Robustness: Volatility

	(1)	(2)	(3)
	+3	+5	+7
Lagged Volatility	0.439***	0.506***	0.260***
	(6.893)	(7.093)	(3.558)
Similarity Score	-0.014***	-0.011***	-0.014***
	(-3.770)	(-2.788)	(-2.891)
Inv. Rate	-0.011	0.001	-0.003
	(-1.175)	(0.066)	(-0.277)
TFP (g)	-0.122***	-0.020	0.082**
	(-3.789)	(-0.800)	(2.440)
Human Capital (g)	0.036	0.090	-0.017
	(0.309)	(1.151)	(-0.182)
World Volatility	0.432***	0.399**	0.420*
	(2.928)	(2.141)	(1.908)
Constant	0.015***	0.008	0.014**
	(2.935)	(1.605)	(2.157)
No. of Obs.	1231	614	409
AR1 (p-value)	0.00	0.00	0.00
AR2 (p-value)	0.05	0.56	0.39
Hansen-J (p-value)	1.00	0.39	0.21

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Robustness: Risk-adjusted Growth

	(1)	(2)	(3)
	+3	+5	+7
Lagged Adj. Growth	0.309***	0.077	0.469***
	(5.295)	(1.286)	(6.555)
Similarity Score	2.356***	2.968***	1.124
	(3.681)	(3.030)	(1.188)
Inv. Rate	11.502***	13.434***	6.415**
	(5.035)	(3.432)	(2.029)
TFP Growth	24.697***	15.026**	27.794**
	(4.242)	(2.248)	(2.320)
Human Capital Growth	30.285	-31.868	7.815
	(1.359)	(-0.837)	(0.145)
World Adj. Grwoth	0.494***	0.600**	0.676***
	(2.971)	(2.243)	(2.713)
Constant	-4.268***	-4.484***	-2.346*
	(-5.340)	(-3.497)	(-1.822)
No. of Obs.	1231	614	409
AR1 (p-value)	0.00	0.04	0.00
AR2 (p-value)	0.13	0.01	0.29
Hansen-J (p-value)	1.00	0.43	0.01

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Growth Regression (fwd+3)

	(1) OLS	(2) OLS-X	(3) FE(t),X	(4) GMM-(t)	(5) GMM-X(t)
Lagged GDP Per Capita	-0.001	-0.008***	-0.016***	-0.000	-0.000
	(-1.582)	(-3.744)	(-3.703)	(-0.311)	(-0.033)
Similarity Score	0.029***	0.037***	0.033***	0.019***	0.021***
	(4.462)	(3.924)	(2.703)	(2.753)	(3.120)
Inv. Rate		0.127***	0.119***	0.130***	0.130***
		(6.280)	(4.718)	(4.642)	(5.321)
TFP Growth		0.766***	0.739***	0.789***	0.806***
		(14.833)	(9.806)	(10.541)	(10.979)
Human Capital Growth		0.504***	0.502***	0.674***	0.650***
		(4.108)	(2.855)	(3.169)	(3.799)
World Growth		0.454***			0.433***
		(4.572)			(4.764)
No. of Obs.	2030	1258	1258	1258	1258
AR1 (p-value)				0.00	0.00
AR2 (p-value)				0.92	0.87
Hansen-J (p-value)				1.00	1.00

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Growth Regression (fwd+7)

	(1) OLS	(2) OLS-X	(3) FE(t),X	(4) GMM-(t)	(5) GMM-X(t)
L.l_gdppc	-0.001**	-0.012***	-0.025***	-0.001	-0.000
	(-2.060)	(-4.004)	(-4.882)	(-0.367)	(-0.204)
Similarity Score	0.026***	0.025*	0.018	0.015**	0.016**
	(3.656)	(1.921)	(1.505)	(2.120)	(2.101)
Inv. Rate		0.109***	0.100***	0.123***	0.123***
		(5.490)	(4.571)	(3.880)	(3.806)
TFP Growth		0.680***	0.612***	0.771***	0.773***
		(8.461)	(6.263)	(8.532)	(8.777)
Human Capital Growth		0.522***	0.537***	0.445**	0.380
		(2.957)	(2.787)	(2.068)	(1.578)
World Growth		1.077***			0.606*
		(3.603)			(1.889)
No. of Obs.	838	523	522	523	523
AR1 (p-value)				0.01	0.01
AR2 (p-value)				0.24	0.18
Hansen-J (p-value)				0.02	0.01

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Volatility Regression (fwd+3)

	(1) OLS	(2) OLS-X	(3) FE(t),X	(4) GMM-(t)	(5) GMM-X(t)
L.sd of growth_gdppc	0.452***	0.310***	0.335***	0.459***	0.432***
	(10.127)	(4.913)	(5.917)	(7.659)	(6.840)
Similarity Score	-0.010***	-0.002	-0.004	-0.013***	-0.014***
	(-3.371)	(-0.218)	(-0.362)	(-3.588)	(-3.729)
Inv. Rate		-0.008	-0.010	-0.012	-0.012
		(-0.841)	(-1.035)	(-1.212)	(-1.256)
TFP Growth		-0.114***	-0.118***	-0.122***	-0.117***
		(-3.787)	(-4.179)	(-3.859)	(-3.650)
Human Capital Growth		0.025	0.056	0.108	0.037
		(0.278)	(0.526)	(0.776)	(0.364)
World Volatility		0.672***			0.525***
		(5.655)			(3.823)
No. of Obs.	1948	1231	1231	1231	1231
AR1 (p-value)				0.00	0.00
AR2 (p-value)				0.03	0.05
Hansen-J (p-value)				1.00	1.00

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Volatility Regression (fwd+7)

	(1) OLS	(2) OLS-X	(3) FE(t),X	(4) GMM-(t)	(5) GMM-X(t)
L.sd of growth_gdppc	0.470^{***}	-0.010	-0.012	0.265^{***}	0.257^{***}
	(4.693)	(-0.104)	(-0.206)	(3.772)	(3.508)
Similarity Score	-0.006	-0.011	-0.009	-0.013^{***}	-0.013^{***}
	(-1.501)	(-0.735)	(-0.576)	(-2.917)	(-2.814)
Inv. Rate		-0.012	-0.013	-0.006	-0.002
		(-1.205)	(-1.274)	(-0.482)	(-0.199)
TFP Growth		0.067	0.071^{*}	0.087^{***}	0.079^{**}
		(1.591)	(1.765)	(2.622)	(2.363)
Human Capital Growth		0.166	0.135	-0.028	-0.015
		(1.317)	(0.862)	(-0.267)	(-0.158)
World Volatility		0.868^{***}			0.500^{**}
		(3.407)			(2.198)
No. of Obs.	645	409	402	409	409
AR1 (p-value)				0.00	0.00
AR2 (p-value)				0.42	0.39
Hansen-J (p-value)				0.22	0.22

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Risk-adjusted Growth Regression (fwd+3)

	(1) OLS	(2) OLS-X	(3) FE(t),X	(4) GMM-(t)	(5) GMM-X(t)
L.SR	0.405^{***}	0.157^{***}	0.166^{***}	0.274^{***}	0.260^{***}
	(8.205)	(3.711)	(4.360)	(4.602)	(4.493)
Similarity Score	2.655^{***}	2.123^{*}	2.761^{**}	2.111^{***}	2.111^{***}
	(6.366)	(1.943)	(2.282)	(2.773)	(3.174)
Inv. Rate		15.103^{***}	15.564^{***}	13.142^{***}	12.885^{***}
		(6.345)	(6.275)	(5.605)	(5.175)
TFP Growth		35.935^{***}	37.509^{***}	40.165^{***}	38.193^{***}
		(7.664)	(4.661)	(5.183)	(4.759)
Human Capital Growth		58.730^{***}	42.019^{**}	22.448	40.206^{*}
		(3.045)	(2.165)	(0.992)	(1.691)
World Adj. Grwoth		0.444^{***}			0.540^{***}
		(3.200)			(3.610)
No. of Obs.	1946	1230	1230	1230	1230
AR1 (p-value)				0.00	0.00
AR2 (p-value)				0.16	0.14
Hansen-J (p-value)				1.00	1.00

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 13: Risk-adjusted Growth Regression (fwd+7)

	(1) OLS	(2) OLS-X	(3) FE(t),X	(4) GMM-(t)	(5) GMM-X(t)
L.SR	0.230**	0.125**	0.145***	0.310***	0.303***
	(2.158)	(2.328)	(3.184)	(5.333)	(5.702)
Similarity Score	3.411***	-1.672	-2.279	0.766	1.117
	(4.973)	(-0.793)	(-1.054)	(0.790)	(1.213)
Inv. Rate		12.183***	11.982***	8.355***	8.917***
		(3.582)	(3.413)	(3.188)	(3.563)
TFP Growth		44.303***	49.137***	65.266***	56.841***
		(3.438)	(3.051)	(4.732)	(4.512)
Human Capital Growth		44.823	26.236	58.728	58.364
		(0.874)	(0.506)	(1.430)	(1.374)
World Adj. Grwoth		0.830***			0.700***
		(4.246)			(3.897)
No. of Obs.	643	408	401	408	408
AR1 (p-value)				0.00	0.00
AR2 (p-value)				0.05	0.02
Hansen-J (p-value)				0.15	0.16

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 14: Growth Regression (winsorized)

	(1) OLS	(2) OLS-X	(3) FE(t),X	(4) GMM-(t)	(5) GMM-X(t)
Lagged Growth	-0.000	-0.009***	-0.020**	-0.002*	-0.002*
	(-0.635)	(-2.990)	(-2.606)	(-1.851)	(-1.846)
Similarity Score	0.025***	0.030***	0.025*	0.017**	0.018**
	(4.299)	(3.417)	(2.352)	(2.149)	(2.386)
Inv. Rate		0.127***	0.119***	0.131***	0.133***
		(8.610)	(6.668)	(8.008)	(7.439)
TFP Growth		0.678***	0.648***	0.700***	0.697***
		(10.656)	(6.765)	(9.392)	(10.489)
Human Capital Growth		0.336***	0.320**	0.429***	0.397***
		(2.904)	(3.006)	(2.786)	(2.673)
World Growth		0.369**			0.353**
		(2.138)			(2.126)
Constant	-0.002	0.024	0.138*	-0.009	-0.017
	(-0.266)	(0.966)	(1.966)	(-0.693)	(-1.444)
No. of Obs.	931	631	627	631	631
AR1 (p-value)				0.00	0.00
AR2 (p-value)				0.03	0.06
Hansen-J (p-value)				0.88	0.78

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 15: Volatility Regression (winsorized)

	(1) OLS	(2) OLS-X	(3) FE(t),X	(4) GMM-(t)	(5) GMM-X(t)
Lagged Volatility _gdppc	0.334***	0.137**	0.125***	0.314***	0.316***
	(7.614)	(1.984)	(2.880)	(8.250)	(8.890)
Similarity Score	-0.003	-0.001	0.001	-0.006**	-0.006*
	(-1.077)	(-0.068)	(0.126)	(-2.007)	(-1.851)
Inv. Rate		-0.011	-0.011	-0.005	-0.004
		(-1.386)	(-1.093)	(-0.513)	(-0.363)
TFP Growth		-0.015	-0.016	-0.001	-0.002
		(-0.505)	(-0.679)	(-0.048)	(-0.093)
Human Capital Growth		0.034	0.023	-0.015	-0.019
		(0.614)	(0.432)	(-0.285)	(-0.343)
World Volatility		0.618***			0.450***
		(4.116)			(2.654)
World Growth		0.111			
		(1.199)			
Constant	0.010***	0.016**	0.011**	0.014***	0.007*
	(4.533)	(1.978)	(2.521)	(4.600)	(1.882)
No. of Obs.	816	530	524	530	530
AR1 (p-value)				0.00	0.00
AR2 (p-value)				0.34	0.33
Hansen-J (p-value)				0.33	0.30

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 16: Risk-adjusted Growth Regression (winsorized)

	(1) OLS	(2) OLS-X	(3) FE(t),X	(4) GMM-(t)	(5) GMM-X(t)
Lagged Adj. Growth	0.434***	0.225***	0.273***	0.398***	0.333***
	(11.279)	(4.367)	(5.834)	(6.722)	(5.526)
Similarity Score	1.546***	-0.079	0.023	0.845*	0.890**
	(3.909)	(-0.071)	(0.020)	(1.892)	(2.090)
Inv. Rate		6.083***	7.245***	3.527*	1.901
		(2.650)	(2.836)	(1.778)	(1.005)
TFP Growth		29.157***	32.236***	38.550***	33.544***
		(4.216)	(3.589)	(4.449)	(5.239)
Human Capital Growth		22.108	12.873	23.103	30.058
		(1.204)	(0.777)	(0.888)	(1.226)
World Adj. Growth		0.397***			0.380***
		(3.629)			(3.856)
Constant	-0.886***	-2.157**	-1.162	-1.954***	-1.340**
	(-2.668)	(-2.585)	(-1.101)	(-2.892)	(-2.182)
No. of Obs.	804	522	517	522	522
AR1 (p-value)				0.00	0.00
AR2 (p-value)				0.79	0.21
Hansen-J (p-value)				0.24	0.29

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 17: Classification according to the SITC 1 -Section

SITC Code	**Sector Name**
0	Food and live animals
1	Beverages and tobacco
2	Crude materials, inedible, except fuels
3	Mineral fuels, lubricants and related materials
4	Animal and vegetable oils, fats and waxes
5	Chemicals and related products, n.e.s.
6	Manufactured goods classified chiefly by material
7	Machinery and transport equipment
8	Miscellaneous manufactured articles
9	Commodities and transactions not classified elsewhere in the SITC

Table 18: Summary of Actual RCA by year

year	N	Mean	Q1	Median	Q3
1985	61160	0.136	0.024	0.158	0.844
1990	69130	0.133	0.024	0.156	0.788
1995	84213	0.144	0.027	0.168	0.798
2000	96925	0.138	0.027	0.163	0.780
2005	100931	0.128	0.024	0.155	0.744
2010	104069	0.120	0.023	0.152	0.735
2015	101789	0.117	0.022	0.143	0.708

Table 19: Summary of Recommendation Scores by year

year	N	Mean	Q1	Median	Q3
1985	133722	0.269	0.140	0.328	0.648
1990	134504	0.308	0.161	0.370	0.686
1995	152281	0.350	0.202	0.403	0.712
2000	159444	0.365	0.217	0.438	0.754
2005	160218	0.348	0.198	0.419	0.749
2010	160784	0.342	0.195	0.416	0.747
2015	160576	0.337	0.194	0.393	0.702

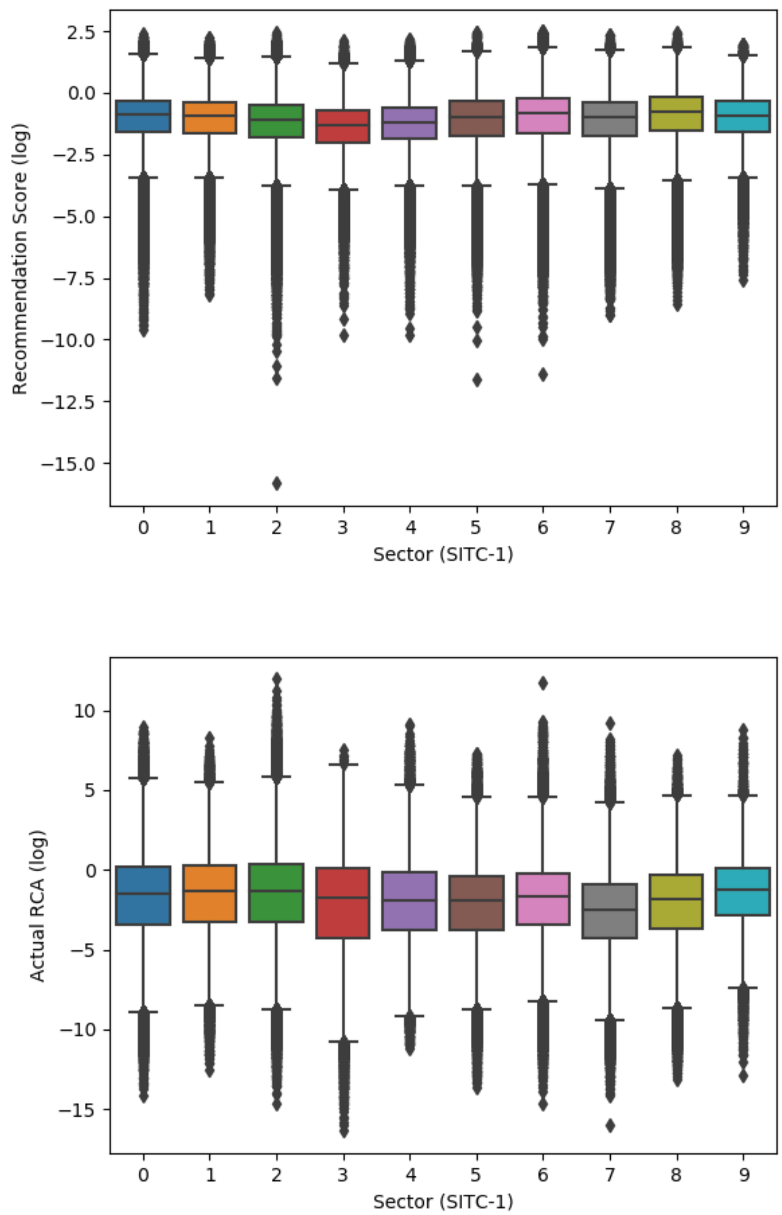

Figure 9: Distribution of Actual RCA and Recommendation Scores by Sector