

Practical Applications of Economic Complexity Analysis

Development Policy Research Unit

FINDING & UNDERSTANDING THE DATA

Using Economic Complexity and Related Measures

There are 3 possible paths one can take in order to get these measures:

1. Use the data provided to us by the CID (thanks to Sebastian Bustos)
 - Dropbox
 - USB
2. Generate complexity measures using Stata package developed by Sebastian Bustos and Muhammed Yildirim (*ecomplexity*)
 - net install ecomplexity, from("<https://raw.githubusercontent.com/cid-harvard/ecomplexity/master/>") force
 - See documentation on ecomplexity by typing 'help ecomplexity'
3. Reverse engineer using the code provided by CID (<https://github.com/cid-harvard/atlas-data>)
 - Useful if one wants to advance the methodology

Using Economic Complexity and Related Measures

Generate complexity measures using Stata package developed by Sebastian Bustos and Muhammed Yildirim (*ecomplexity*)

- In Stata:

```
net install ecomplexity,  
from( "https://raw.githubusercontent.com/cid-harvard/ecomplexity/master/" ) force
```

- Download BACI data (<http://www.cepii.fr/CEPII/en/welcome.asp>). If your institution has access to UN Comtrade, then you should be able to get access.
- In order to run the *ecomplexity* command, you need country-level export data at the product-level by year (also an option to include population data)
 - E.g. `ecomplexity export_value, i(ccode_x) p(h0) t(year)`
 - Restrict estimations to 128 Atlas countries (see Section 7 of Atlas of Economic Complexity book)
 - Choose nomenclature (HS or SITC) and level of aggregation – CID advised using the 4-digit level
 - Collapse data to exporter-product-year level
 - Rectangularise the dataset
 - Exclude products that account for tiny share of world trade (sort products by share and include products that account for 99.9% of world trade)
 - Run command and it will generate Complexity variables
- Let's look at some data

Familiarising yourself with the datasets

Important to note that you may require different datasets depending on your research question, but some examples we can look at include:

H0 Data

- All countries
- ECI
- PCI
- Opp. Gain
- Value of Exports
- Products

WB Class'n

- All countries
- Regions
- Income Groups

GDP Data

- All countries
- GDP data

Lall Class'n

- All products
- Lall classification
- Technology and product categories

Using Economic Complexity and Related Measures

Complexity variables in the H0 dataset:

- *year* – period of interest
- *exporter* – use ISO 3-digit code
- *commoditycode* – choose nomenclature and level of aggregation
- *export_value* – value of exports for product *p* from country *c* in period *t*
- *import_value* and *population* (not in self-generated dataset)
- *rca* – revealed comparative advantage index for product *p* from country *c* in period *t*
- *mcp* – $M=1$ if $RCA>1$ (for product *p* from country *c* in period *t*)
- *eci* – Economic complexity index (for country *c* in period *t*)
- *pci* – Product complexity index (for product *p* in period *t*)
- *diversity* – number of products exported with $RCA>1$ by country *c* in period *t* (not in CID dataset)
- *ubiquity* – number of countries exporting product *p* in period *t* with $RCA>1$ (not in CID dataset)
- *distance* – distance to the product (distance=1/density) (for product *p* from country *c* in period *t*)
- *coi* – Complexity Outlook Index (or Opportunity Value Index) for country *c* in period *t*
- *cog* – Complexity Opportunity Gain for country *c* and product *p* in period *t*

Using (or generating) the Product Space

There are three possible avenues to generating a product space diagram:

- Copy from website and annotate
 - See <http://atlas.media.mit.edu/en/> or <http://atlas.cid.harvard.edu/>
 - Save as .png file
- Use the Stata ProductSpaceParser.exe written by Cesar Hidalgo that generates GML files of product space maps that network software (such as [Cytoscape](#)) can read.
 - We have tried this and it works
 - However, need more knowledge of network analytics in order to generate useful graphics
- Sebastian Bustos and Muhammed Yildirim write the *genproximity* Stata programme and one can use this to generate the input needed for network software such as Cytoscape

FURTHER APPLICATIONS

H. Bhorat, R. Kanbur, C. Rooney, and F. Steenkamp (2017) – [click here](#)

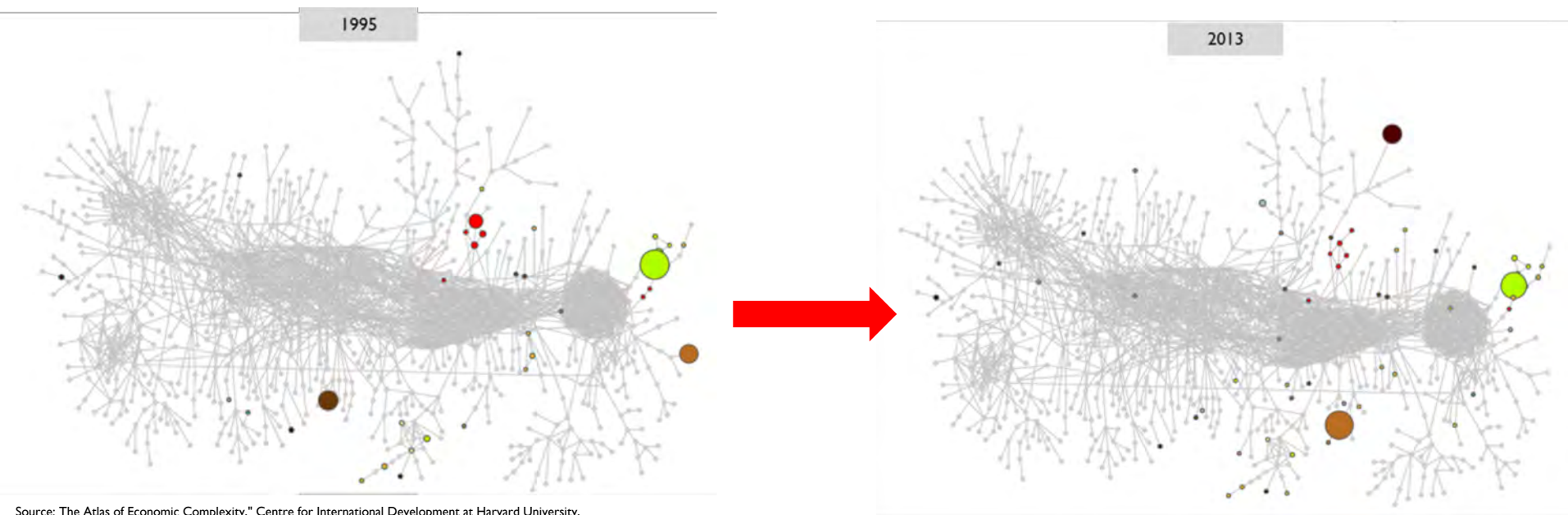
SUB-SAHARAN AFRICA'S MANUFACTURING SECTOR: BUILDING COMPLEXITY

Research questions arising

- Why have we seen a lack of economic development in Sub-Saharan Africa, compared to the East Asian countries?
- What evidence of structural transformation have we seen in these regions, and what does this mean for employment in the future?

Evidence of Structural Transformation in Africa: The Product Space and Manufacturing in Africa

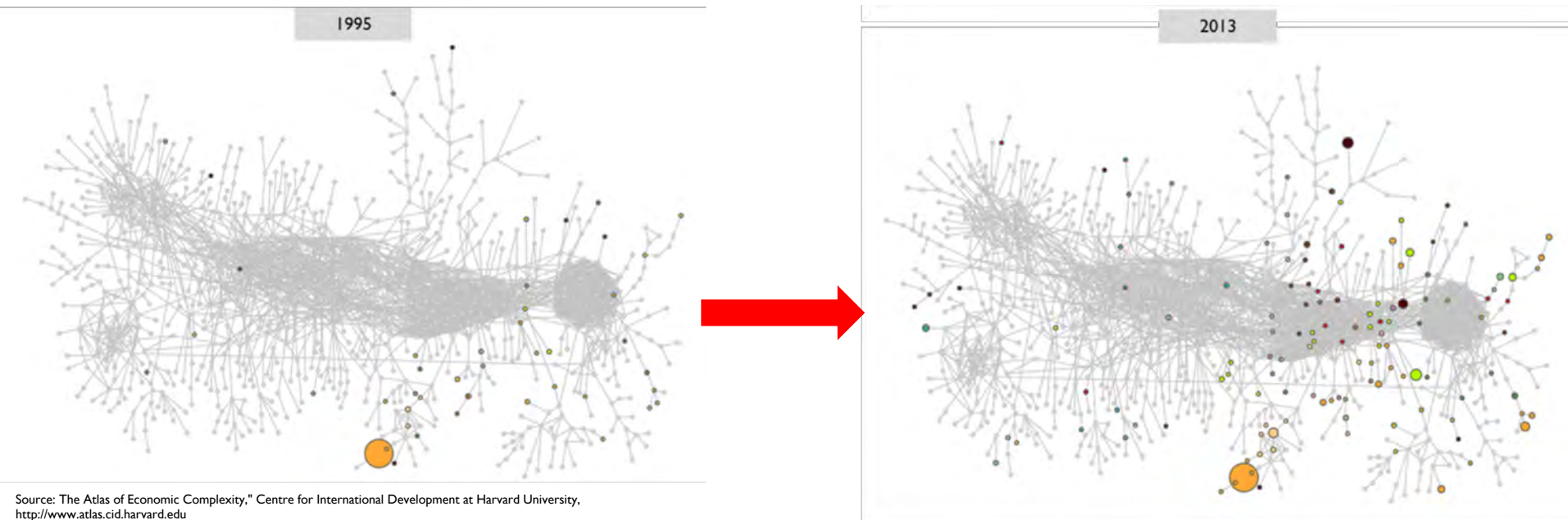
Product Space Analysis Ghana, 1995 & 2013



- Ghana represents the ‘aggregate African product space’
- Little evidence of manufacturing-led structural transformation
- Implications for future structural transformation

Evidence of Structural Transformation in Africa: The Product Space and Manufacturing in Africa

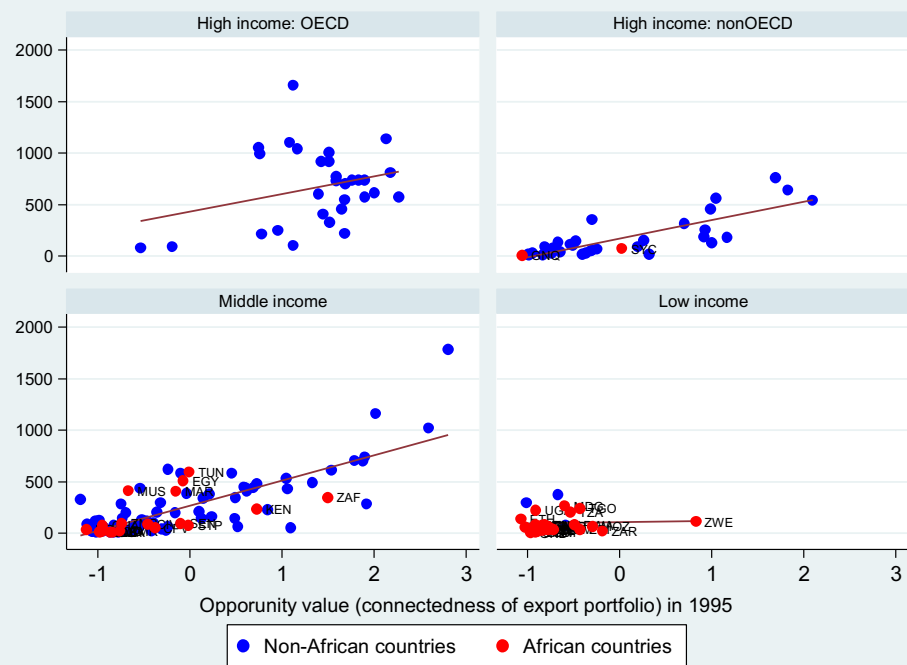
Product Space Analysis Uganda, 1995 & 2013



- BUT, evidence of heterogeneity in ‘African’ product space
- Uganda – ‘manufacturing frontier economy’ – emblematic of a number of other African countries (e.g. Kenya, Mauritius)

Evidence of Structural Transformation in Africa: The Product Space and Manufacturing in Africa

Opportunity Value in 1995 – No. Manuf. Exports (RCA \geq 1) in 2013



- Determine whether a country's initial export structure, and the productive capabilities and connectedness associated with that export structure, impacts on its ability to undergo structural transformation, particularly, a shift toward more complex manufactured products.
- Low income – peripheral nature of productive structure offers little potential
- Middle income – initial productive structure allowed for subsequent diversification

Source: Own calculation using data from The Economic Complexity Observatory (Simoes & Hidalgo, 2011)
 Note: Dropped Germany from High income: OECD sample since it was an outlier.

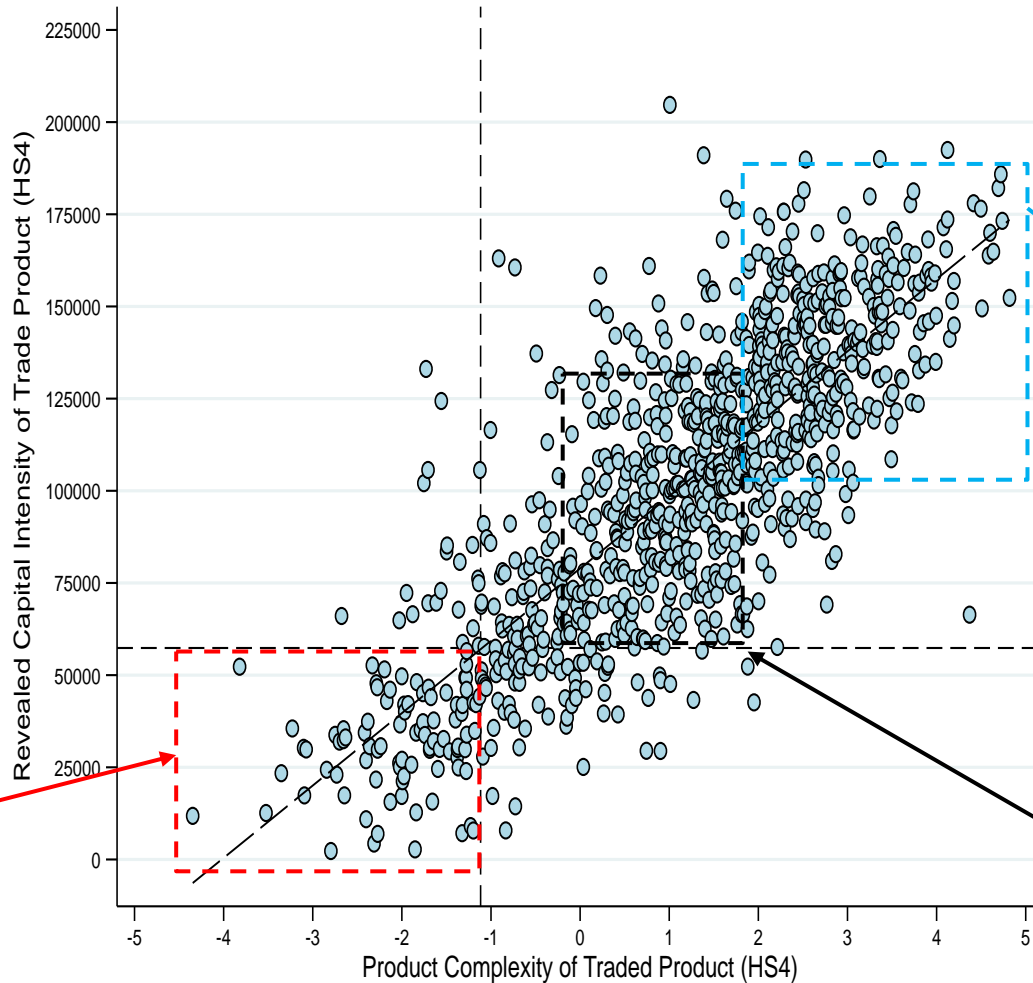
Estimating the Determinants of Africa's Manufacturing Performance

Dependent variable = Ln of product count of TM exports	FE (1)	FE (2)
Log of fixed capital per worker	0.119* [0.063]	0.113* [0.065]
Total factor productivity	0.073 [0.103]	0.095 [0.102]
Total natural resources rents (% of GDP)	0.002 [0.002]	0.002 [0.002]
Africa		
Economic complexity index	-0.029 [0.066]	-0.053 [0.057]
Opportunity value index	0.150*** [0.034]	
Opportunity value index * Low income country dummy		0.333 [0.226]
Opportunity value index * Middle income country dummy		0.149** [0.064]
Opportunity value index * High income OECD country dummy		0.136*** [0.031]
Opportunity value index * High income non-OECD country dummy		0.154 [0.098]
Constant	6.037*** [0.719]	6.104*** [0.744]
Observations	1,750	1,750
Number of groups	104	104
R-squared	0.312	0.318
Country FE	YES	YES
Year FE	YES	YES

Notes: 1. Robust standard errors in brackets. 2. *** p<0.01, ** p<0.05, * p<0.1 3. 'Total natural resource rents' is used a proxy for natural resource abundance in a country. 4. 'Africa' is a dummy variable controlling for whether a country is an African country. 5. The 'total factor productivity', the variable used to control for technology, and is measured using current PPPs with USA=1. 6. Dependent variable is log of product count of TM exports

Product Complexity and Development

Approach
Taken: A
Quick
Guide



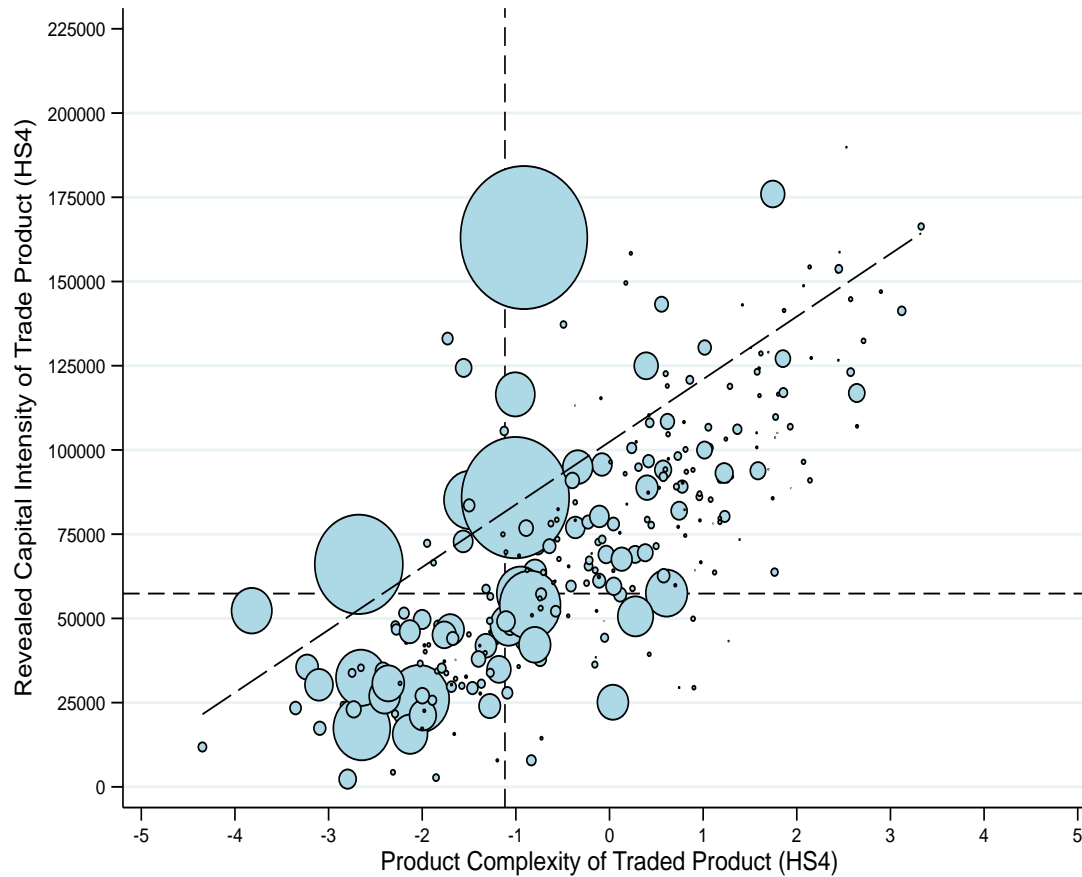
e.g. Jute Woven Fabric; Tanned sheep hides; Knit t-shirts; Felt hats; Basketwork; Non-knit men's undergarments; Knit women's coats; Knotted carpets; Coconut & other vegetable fibres; Heavy pure woven cotton

e.g. Electric motor parts; Office machine parts; Electric locomotives; Semiconductor devices; Polymer ion-exchangers; Lifting machinery; Electrical parts; Medical instruments; Microscopes; Antibiotics; X-ray equipment

e.g. Sulphates; Electric batteries; Synthetic fabrics; Plastic lids; Eyewear; Cast iron pipes; Stranded copper wire; Mill Machinery; Fishing ships; Dental products; Clock cases & parts; Water & gas generators; Buses; Sewing machines; Padlocks; Drilling machines; Hot-rolled iron bars; Refrigerators; Safes

Product Complexity and Development

Evolution of Sub-Saharan Africa's Export Portfolio – Existing Products, 1995-2013



Source: Own calculations using trade data from BACI data (HS 4-digit, revision 1992) to create product complexity measure, and revealed factor intensity data developed by Shirotori et al. (2010).

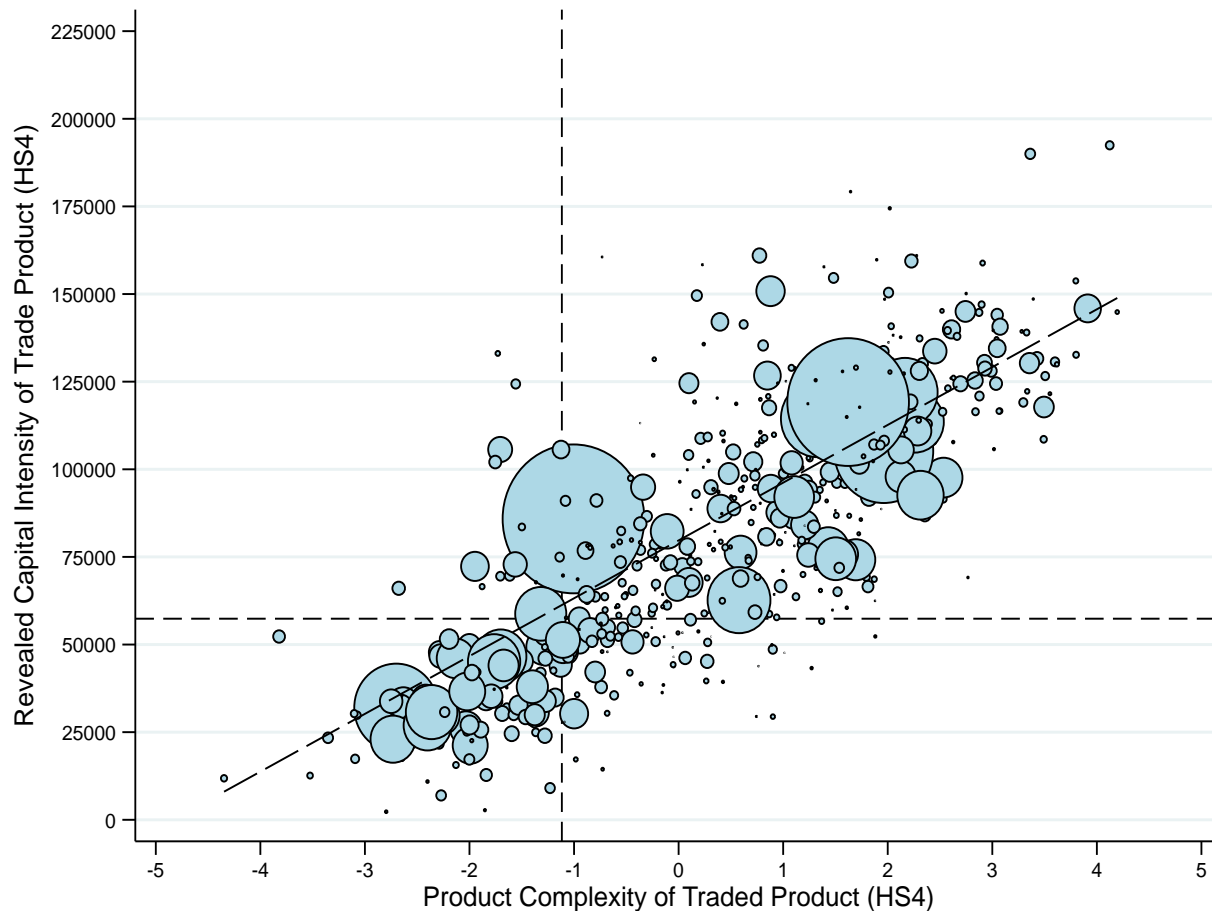
Notes: 1. Traded products are classified at the 4-digit level of the Harmonised System (HS), with each bubble representing a 4-digit product line. 2. The size of each bubble represents the share of that product in total exports in the final period, 2013.

3. The horizontal and vertical lines in each scatter plot represent the average revealed capital intensity and the average product complexity for low-technology manufactures falling within the fashion cluster of the Lall (2000) classification (i.e. clothing and textiles).

4. Trade flows are restricted to products in which at least one country within a region has a revealed comparative advantage. 5. Trade flows restricted to manufacturing products.

Product Complexity and Development

Evolution of East and South Asia's Export Portfolio – Existing Products, 1995-2013



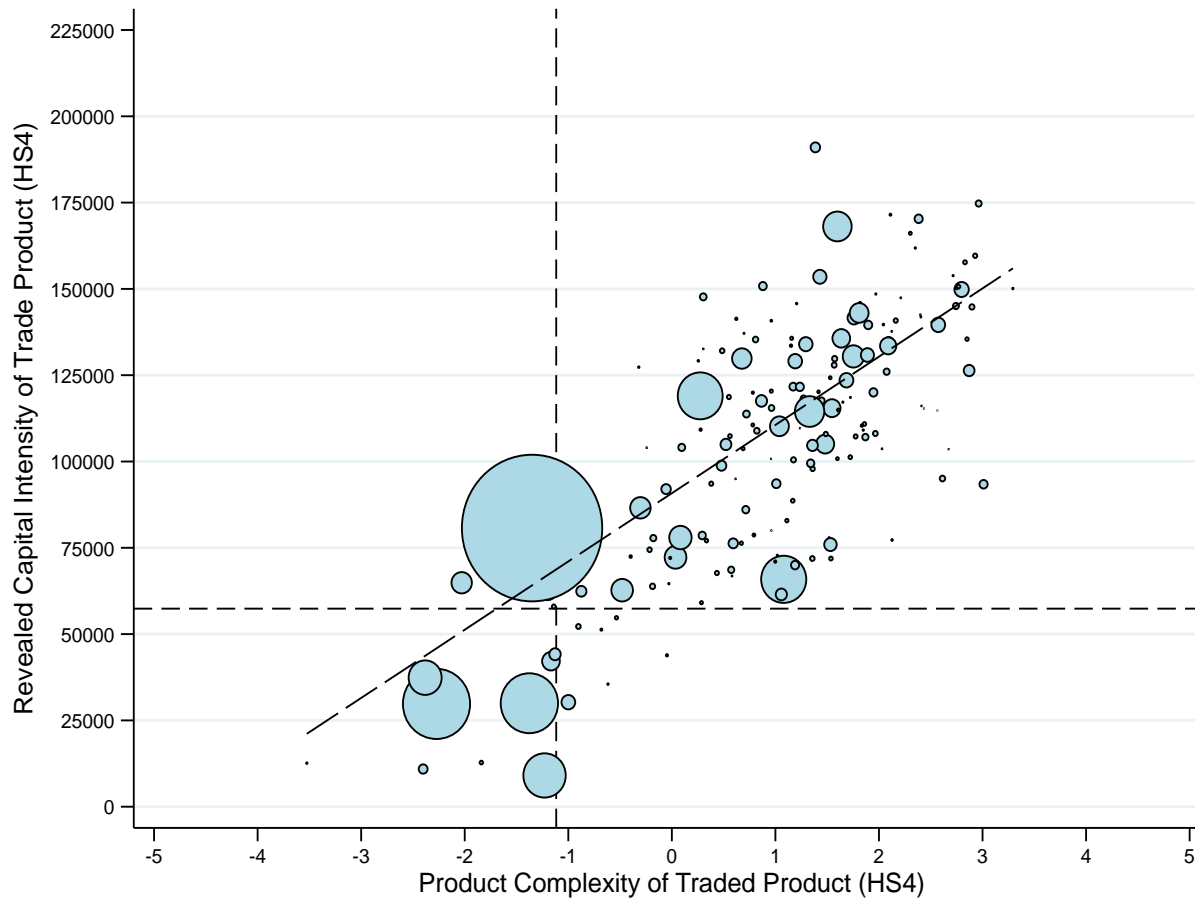
Source: Own calculations using trade data from BACI data (HS 4-digit, revision 1992) to create product complexity measure, and revealed factor intensity data developed by Shirotori et al. (2010).

Notes: 1. Traded products are classified at the 4-digit level of the Harmonised System (HS), with each bubble representing a 4-digit product line. 2. The size of each bubble represents the share of that product in total exports in the final period, 2013.

3. The horizontal and vertical lines in each scatter plot represent the average revealed capital intensity and the average product complexity for low-technology manufactures falling within the fashion cluster of the Lall (2000) classification (i.e. clothing and textiles). 4. Trade flows are restricted to products in which at least one country within a region has a revealed comparative advantage. 5. Trade flows restricted to manufacturing products.

Product Complexity and Development

Evolution of Sub-Saharan Africa's Export Portfolio – Entry into New Products in 2013



Source: Own calculations using trade data from BACI data (HS 4-digit, revision 1992) to create product complexity measure, and revealed factor intensity data developed by Shirotori et al. (2010).

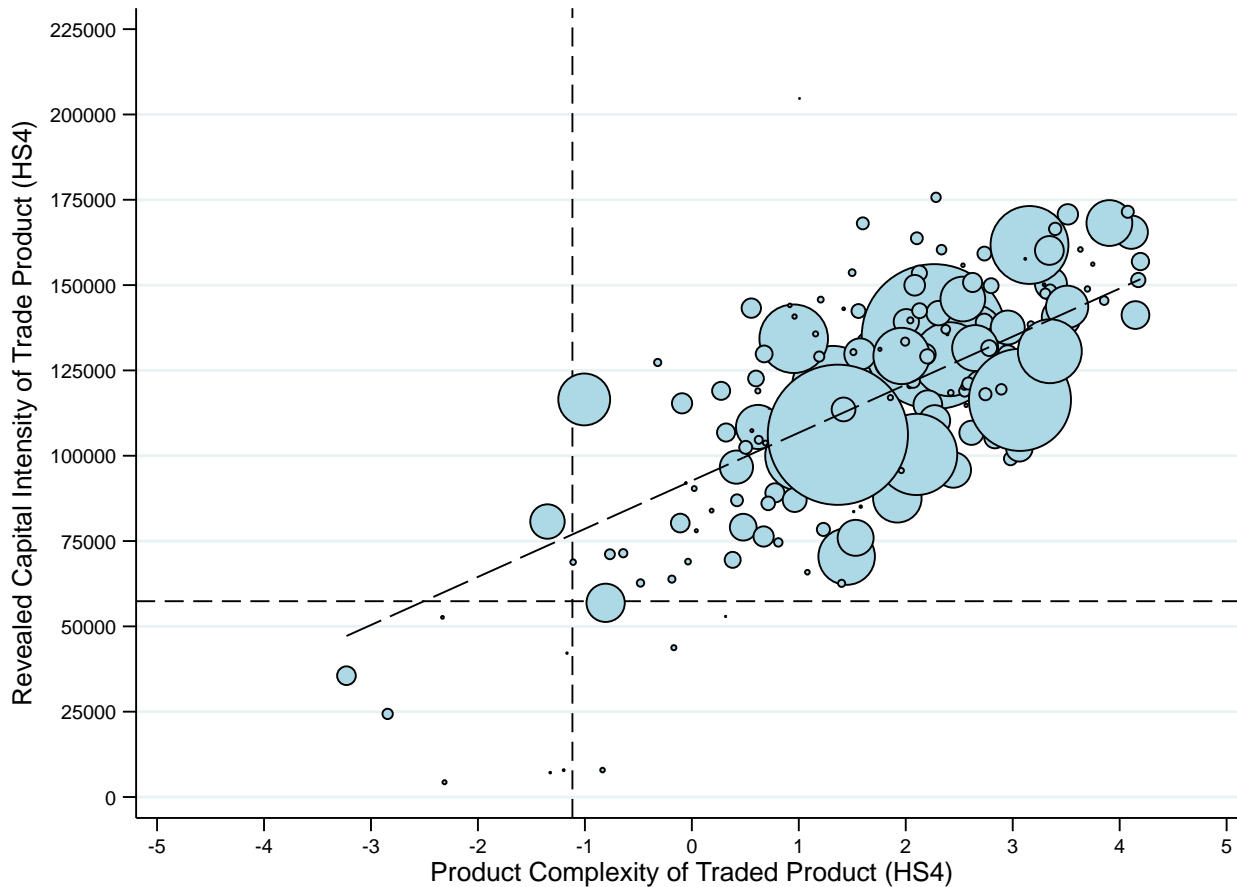
Notes: 1. Traded products are classified at the 4-digit level of the Harmonised System (HS), with each bubble representing a 4-digit product line. 2. The size of each bubble represents the share of that product in total exports in the final period, 2013.

3. The horizontal and vertical lines in each scatter plot represent the average revealed capital intensity and the average product complexity for low-technology manufactures falling within the fashion cluster of the Lall (2000) classification (i.e. clothing and textiles).

4. Trade flows are restricted to products in which at least one country within a region has a revealed comparative advantage. 5. Trade flows restricted to manufacturing products.

Product Complexity and Development

Evolution of East and South Asia's Export Portfolio – Entry into New Products in 2013

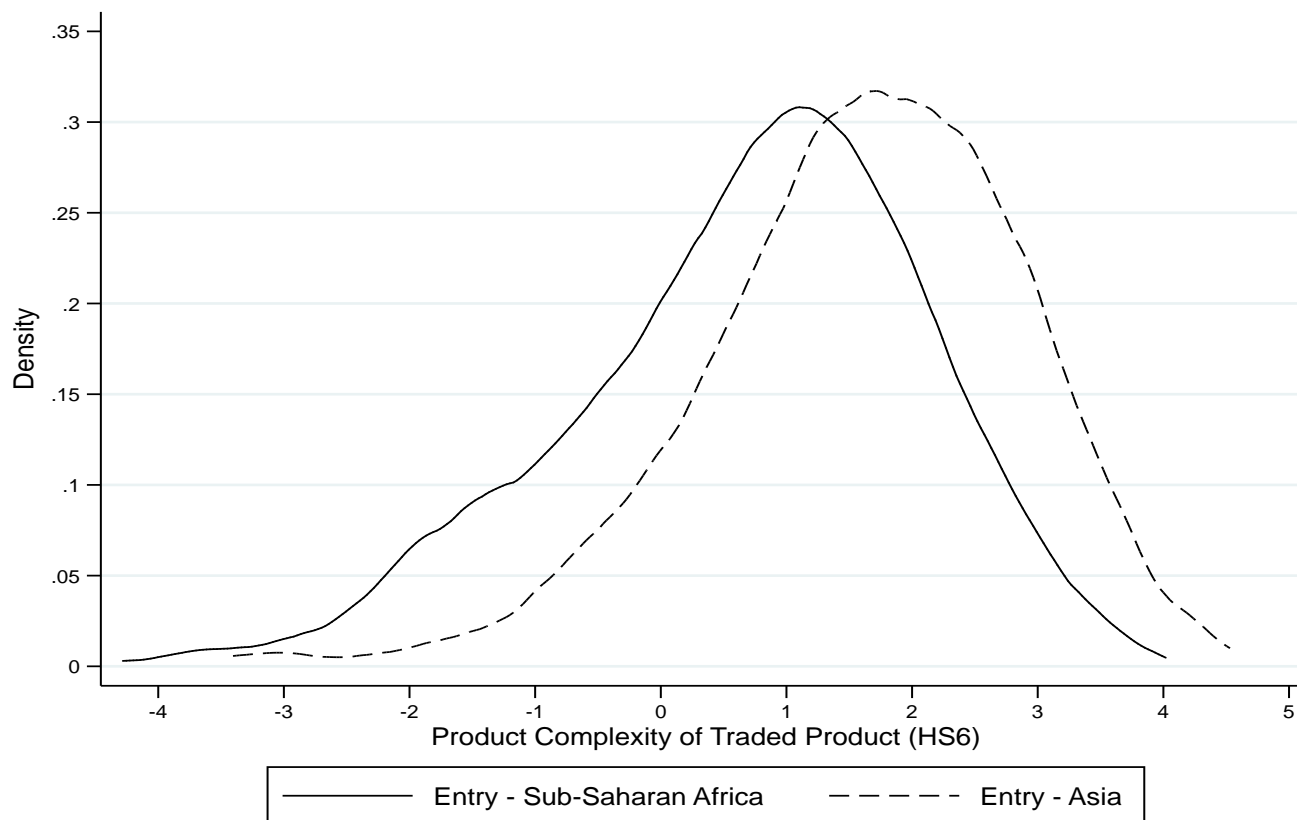


Source: Own calculations using trade data from BACI data (HS 4-digit, revision 1992) to create product complexity measure, and revealed factor intensity data developed by Shirotori et al. (2010).

Notes: 1. Traded products are classified at the 4-digit level of the Harmonised System (HS), with each bubble representing a 4-digit product line. 2. The size of each bubble represents the share of that product in total exports in the final period, 2013. 3. The horizontal and vertical lines in each scatter plot represent the average revealed capital intensity and the average product complexity for low-technology manufactures falling within the fashion cluster of the Lall (2000) classification (i.e. clothing and textiles). 4. Trade flows are restricted to products in which at least one country within a region has a revealed comparative advantage. 5. Trade flows restricted to manufacturing products.

Product Complexity and Development

Distribution of Entries by Region



Source: Own calculations using trade data from BACI data (HS 6-digit, revision 1992) to create product complexity measure, and revealed factor intensity data developed by Shirotori et al. (2010).

Notes: 1. Trade flows are restricted to products in which at least one country within a region has a revealed comparative advantage. 2. Trade flows restricted to manufacturing products.

J. Gao, B. Jun, A. Pentland, T. Zhou & C. Hidalgo (2017) – [click here](#)

COLLECTIVE LEARNING IN CHINA'S REGIONAL ECONOMIC DEVELOPMENT

Research questions arising

- How do different areas of China specialize in different products?
- How could inter-regional learning have played a role in China's economic development?

The industry space

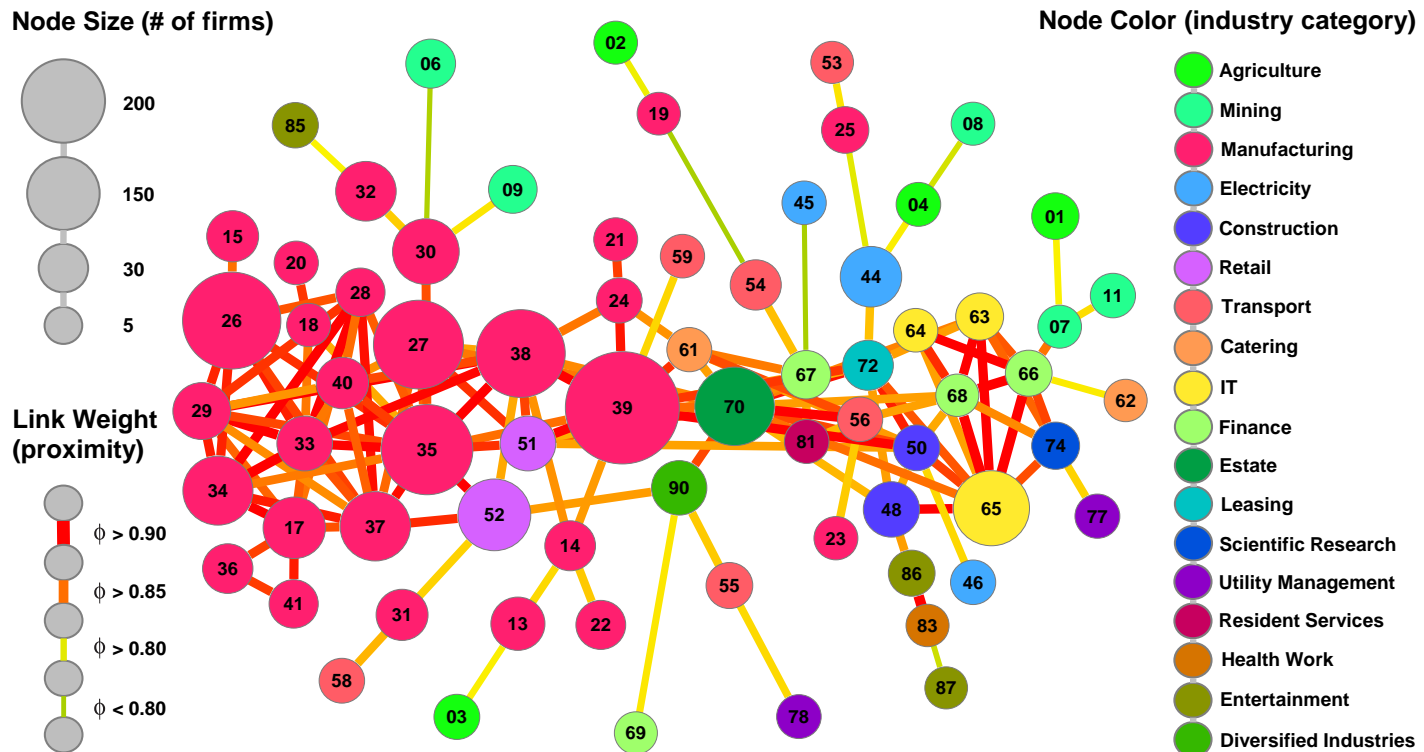


Figure 1: Network representation of China's industry space in 2015. Nodes (circles) represent industries. Links connect industries that are likely to locate in the same province. Nodes are classified into 70 subcategories and colored according to 18 sectors. The size of each node is proportional to the number of firms in that industry. The color and weight of links correspond to the proximity value (ϕ) between two industries.

Development of new industries

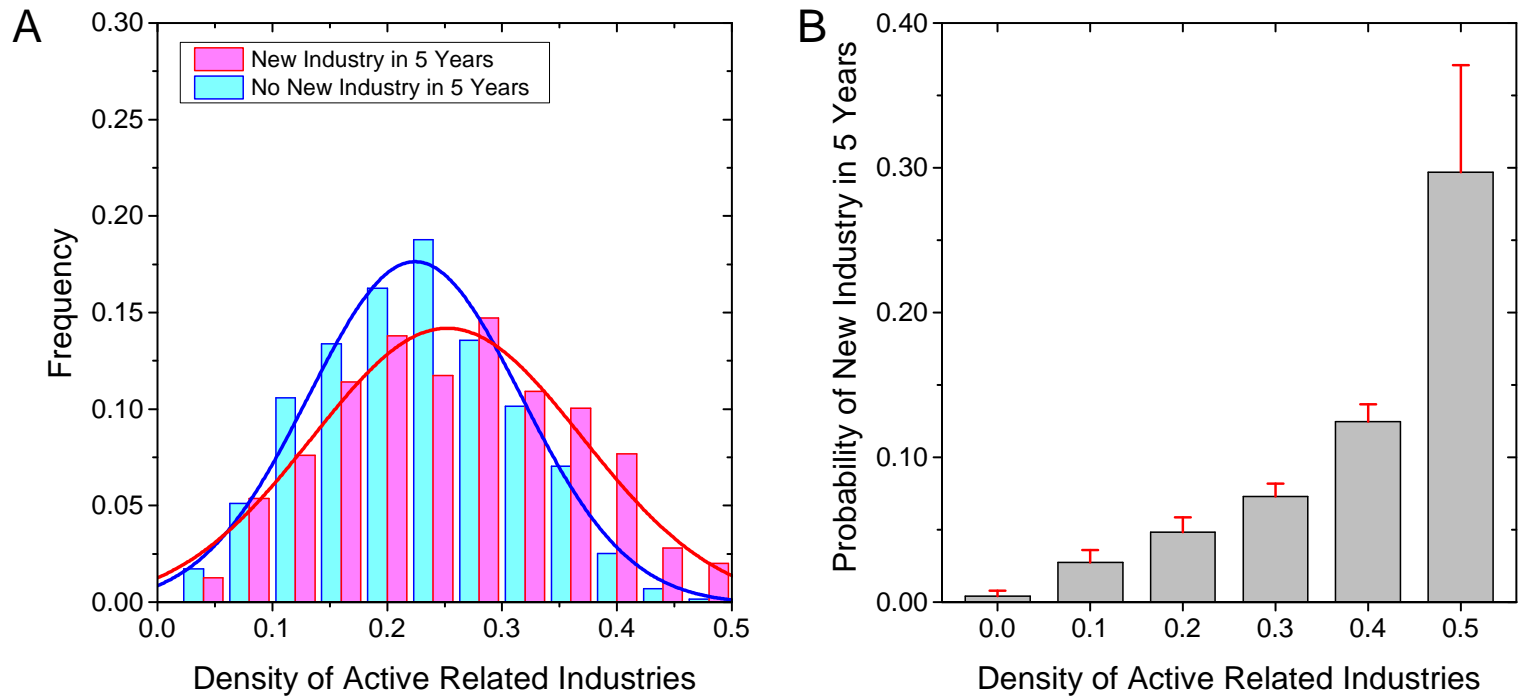
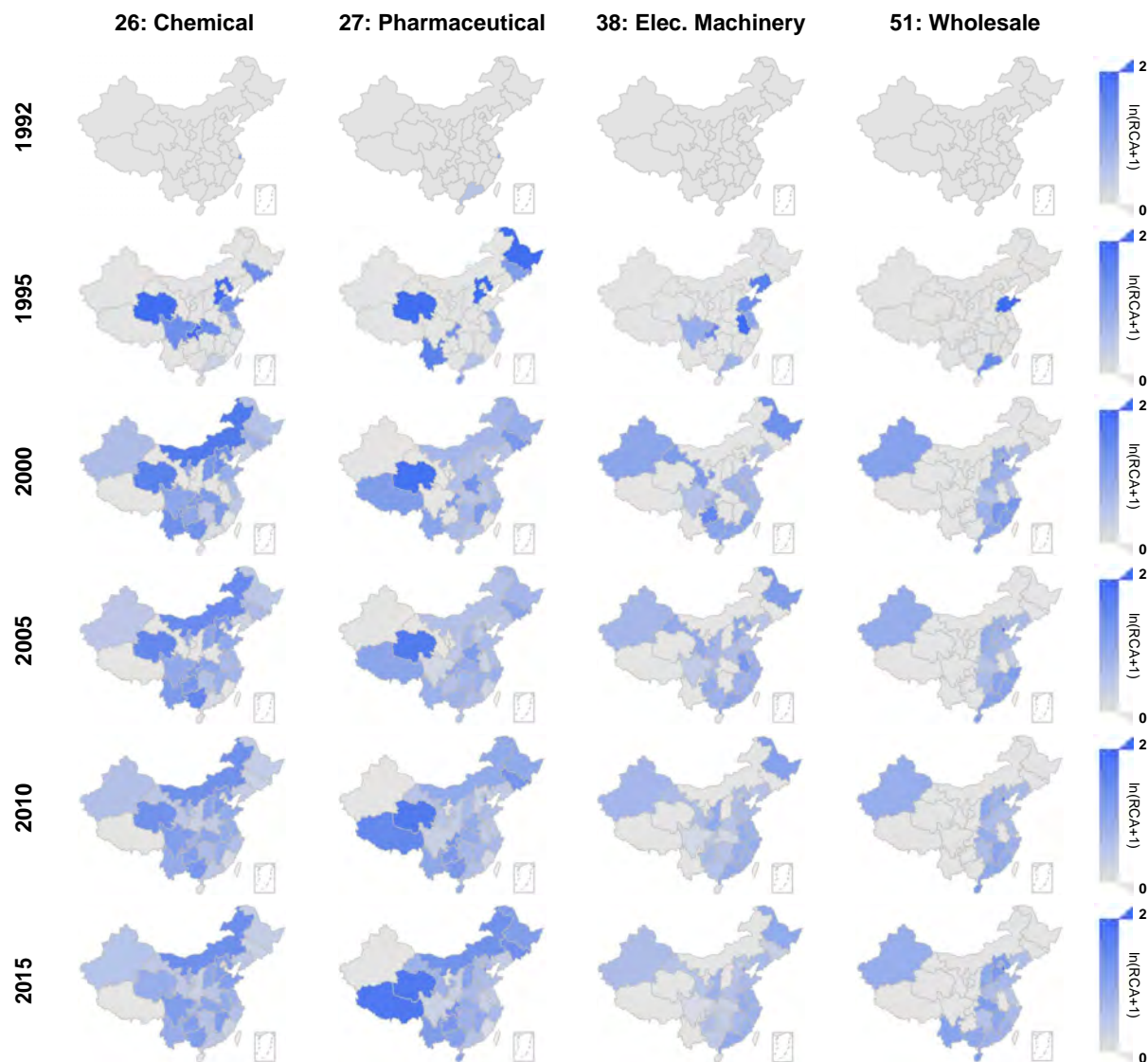


Figure 3: Inter-industry learning. (A) Distribution of the density of active related industries for each pair of provinces and industries. The pink distribution focuses only on pairs of provinces and industries that developed revealed comparative advantage in the next five years. The blue distribution is for the pairs of industries and provinces that did not develop revealed comparative advantage. The mean of the pink distribution is significantly larger than that of the blue distribution (ANOVA p -value= 2.1×10^{-40}). (B) Probability that a new industry will appear in a province as a function of the density of active related industries (ω). Bars indicate average values and error bars indicate standard errors. Results show averages for 2001-2015 using five-year intervals. In all calculations, densities were calculated for the base year.

Inter-regional learning in China



This diagram effectively shows how a particular area/province having $RCA > 1$ could lead to inter-regional learning, and the development of industries in neighbouring areas with $RCA > 1$ in the future.

Industrial similarity and geographic distance

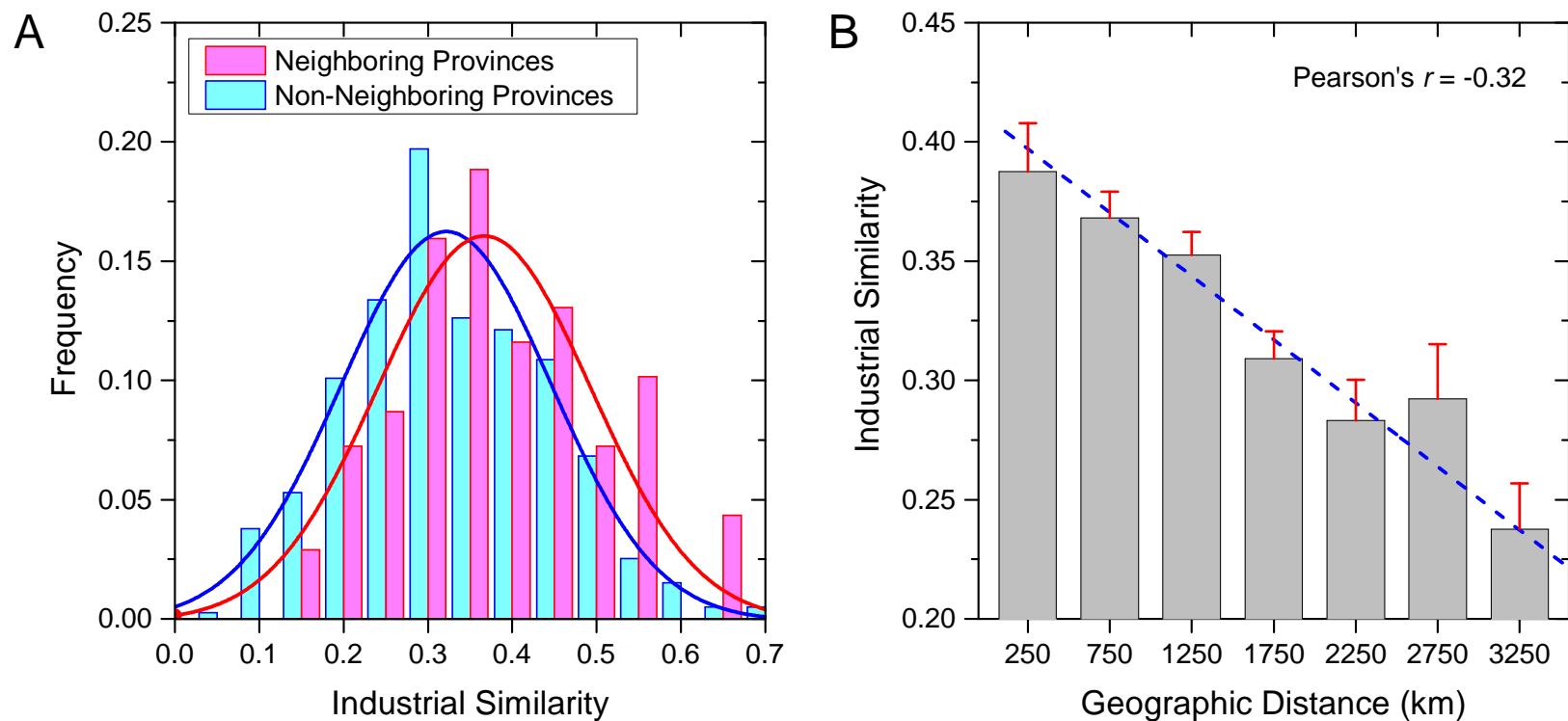


Figure 5: (A) Distribution of industrial similarity between pairs of neighboring provinces (in pink) and non-neighboring provinces (in blue). The red and blue curves are, respectively, normal fits for the distributions for neighboring and non-neighboring province pairs. (B) Industrial similarity between all pairs of provinces as a function of their geographic distance. Bars correspond to the average industrial similarity (φ) of pairs of provinces at that distance and error bars correspond to standard errors. The blue dash line represent a linear fit of the unbinned data. Pearson's correlation between industrial similarity and geographic distance is $r = -0.32$.

The probability of developing new industries

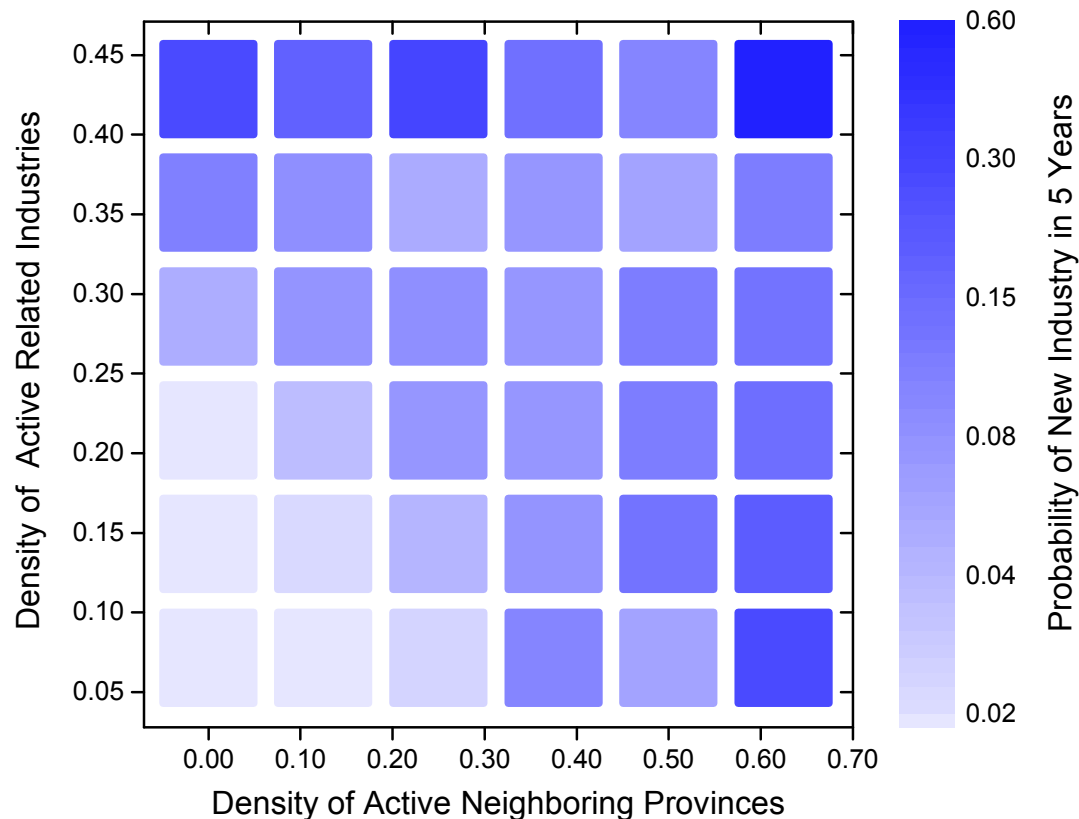


Figure 7: Joint probability of a province developing revealed comparative advantage in a new industry in a five-year period given the density of active neighboring provinces (Ω) in horizontal-axis and the density of active related industries (ω) in vertical-axis.

The high-speed rail as an instrument for provincial connectedness

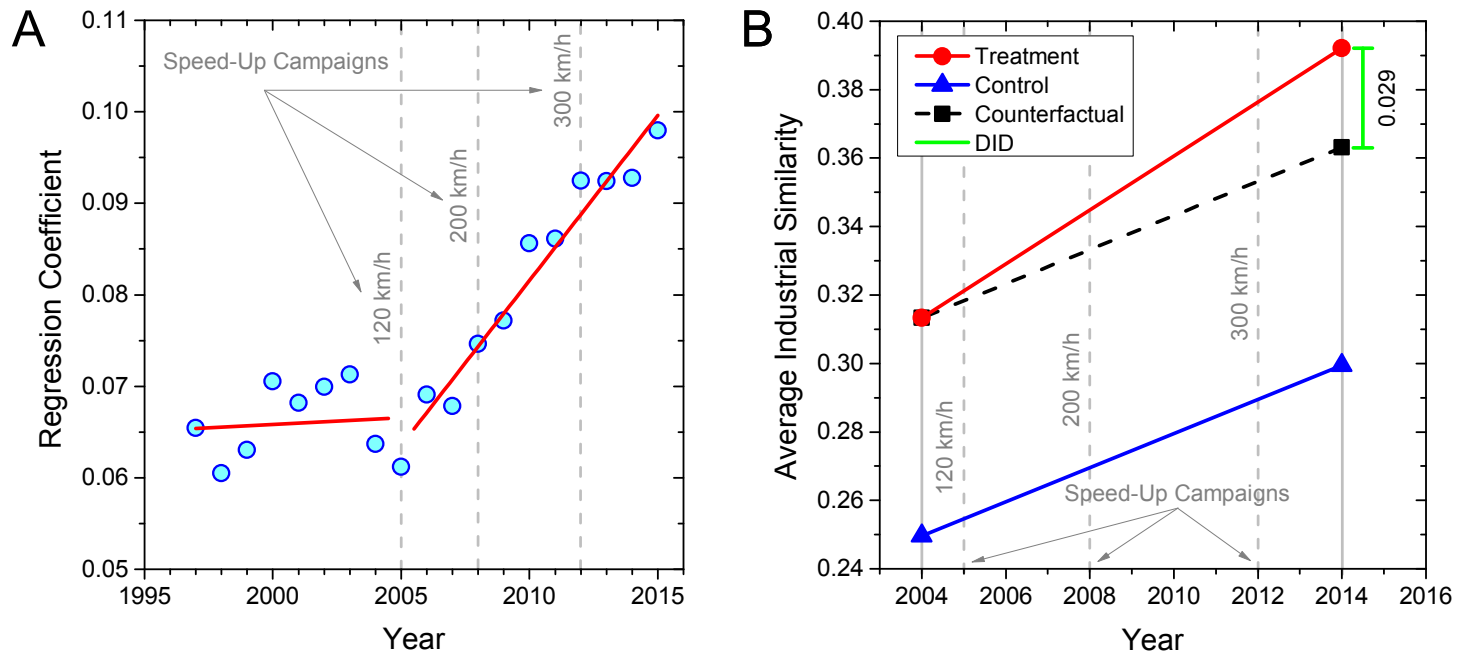


Figure 8: Industrial similarity and the introduction of high-speed rail. (A) Event study results. The y-axis shows the regression coefficient (β_k in Eq. (9)) as a function of the year, after regressing the industrial similarity of pairs of provinces that were eventually connected by high-speed rail against the entry of high-speed rail. Red lines are linear fits for 1997-2005 and 2005-2015. (B) Differences-in-differences (DID) results. The y-axis is the average industrial similarity of all pairs of provinces connected by high-speed rail (in red) or not connected by high-speed rail (in blue). The value of DID (in green) is 0.029, and it is statistically significant. Vertical dash lines mark the years after speed-up campaigns, besides which the approximate average speeds of high-speed rail are shown.

The high-speed rail as an instrument for provincial connectedness

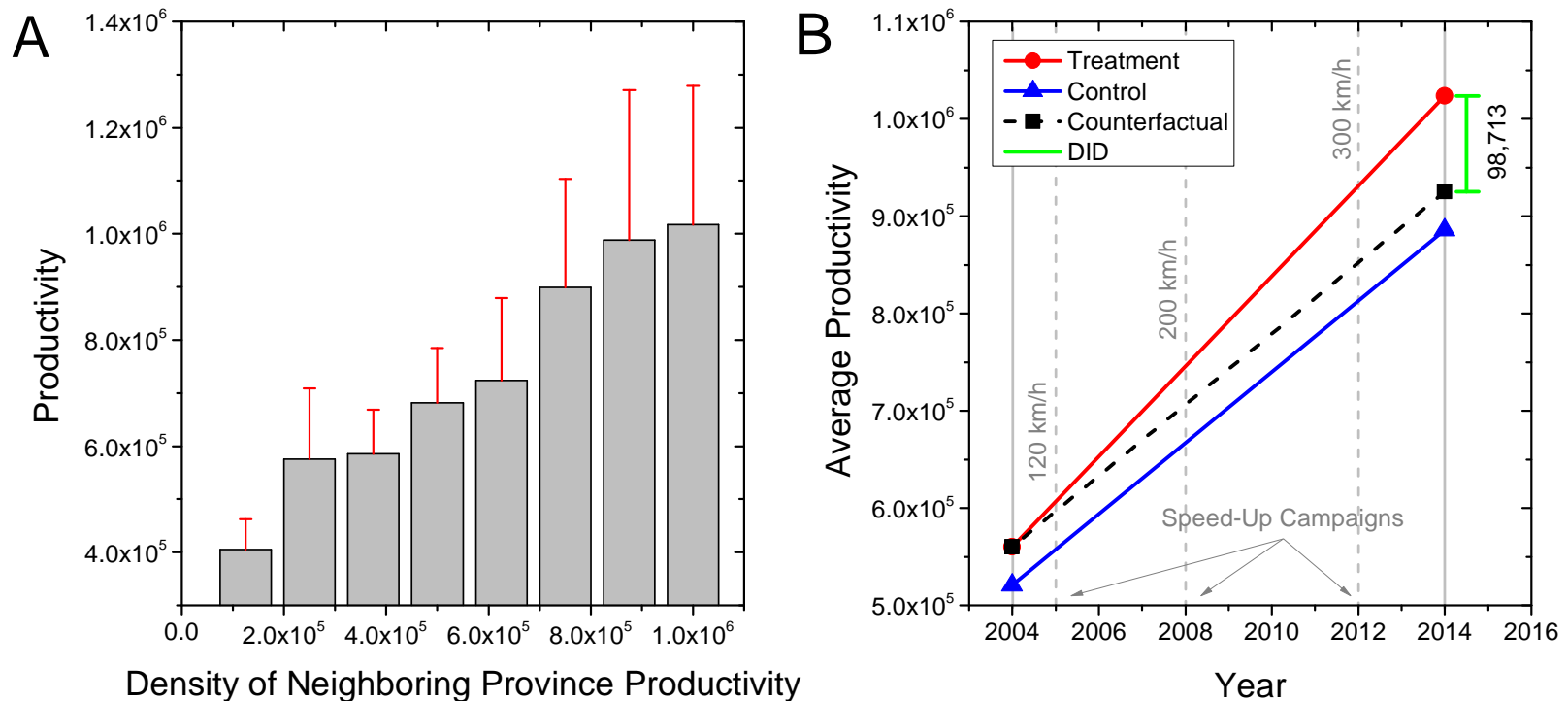


Figure 9: (A) Productivity of provinces as a function of the density of neighboring province productivity five years before. Bars indicate average values and error bars indicate standard errors. Results show averages for 2005-2014 using five-year intervals. (B) Average productivity of province pairs connected with high-speed rail (treatment, in red) and without high-speed rail (control, in blue). The differences-in-differences (DID, in green) is CNY 98,713 (~USD 15k). Vertical dash lines mark the years after speed-up campaigns, besides which the approximate average speeds of high-speed rail are shown.

The high-speed rail as an instrument for provincial connectedness

Table 4: DID regressions considering the effect of high-speed rail entry on the industrial similarity and the productivity of industries.

Independent Variables	DID Regressions Using OLS Model					
	Industrial Similarity			Productivity		
	(1)	(2)	(3)	(4)	(5)	(6)
High-speed Rail Entry	0.0290* (0.0152)	0.0266* (0.0150)	0.0268* (0.0152)	98713*** (27649)	107343*** (27211)	105636*** (26011)
Treatment Group	0.0637*** (0.0107)	0.0565*** (0.0110)	0.0588*** (0.0108)	39135** (16240)	30463* (17033)	26796 (17379)
After Entry	0.0498*** (0.0091)	0.0466*** (0.0091)	0.0506*** (0.0090)	364939*** (17603)	376791*** (17362)	361501*** (16524)
Δ Population (log)		-0.0204*** (0.0049)			-6881 (8767)	
Δ GDP per capita (log)		-0.0207** (0.0081)			109114*** (17389)	
Δ Urbanization			0.0160*** (0.0127)			213686*** (33900)
Δ Trade (log)			-0.0068*** (0.0024)			20877*** (4615)
Observations	930	930	930	930	930	930
Robust R^2	0.1628	0.1833	0.1689	0.4980	0.5223	0.5548
RMSE	0.1109	0.1097	0.1106	2.10×10^5	2.00×10^5	2.00×10^5

Notes: Data are for the year 2004 (before high-speed rail entry) and 2014 (after high-speed rail entry). Significant level: * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

D. Hartmann, M.R. Guevara, C. Jara-Figueroa, M. Aristarán & C.A. Hidalgo (2017) – [click here](#)

LINKING ECONOMIC COMPLEXITY, INSTITUTIONS, AND INCOME INEQUALITY

Research questions arising

- Is there a link between the productive structure of a country's economy and their ability to generate and distribute income?
- Can we explain inequality trends in the world through the use of economic complexity to model productive structures?

ECI and Inequality?

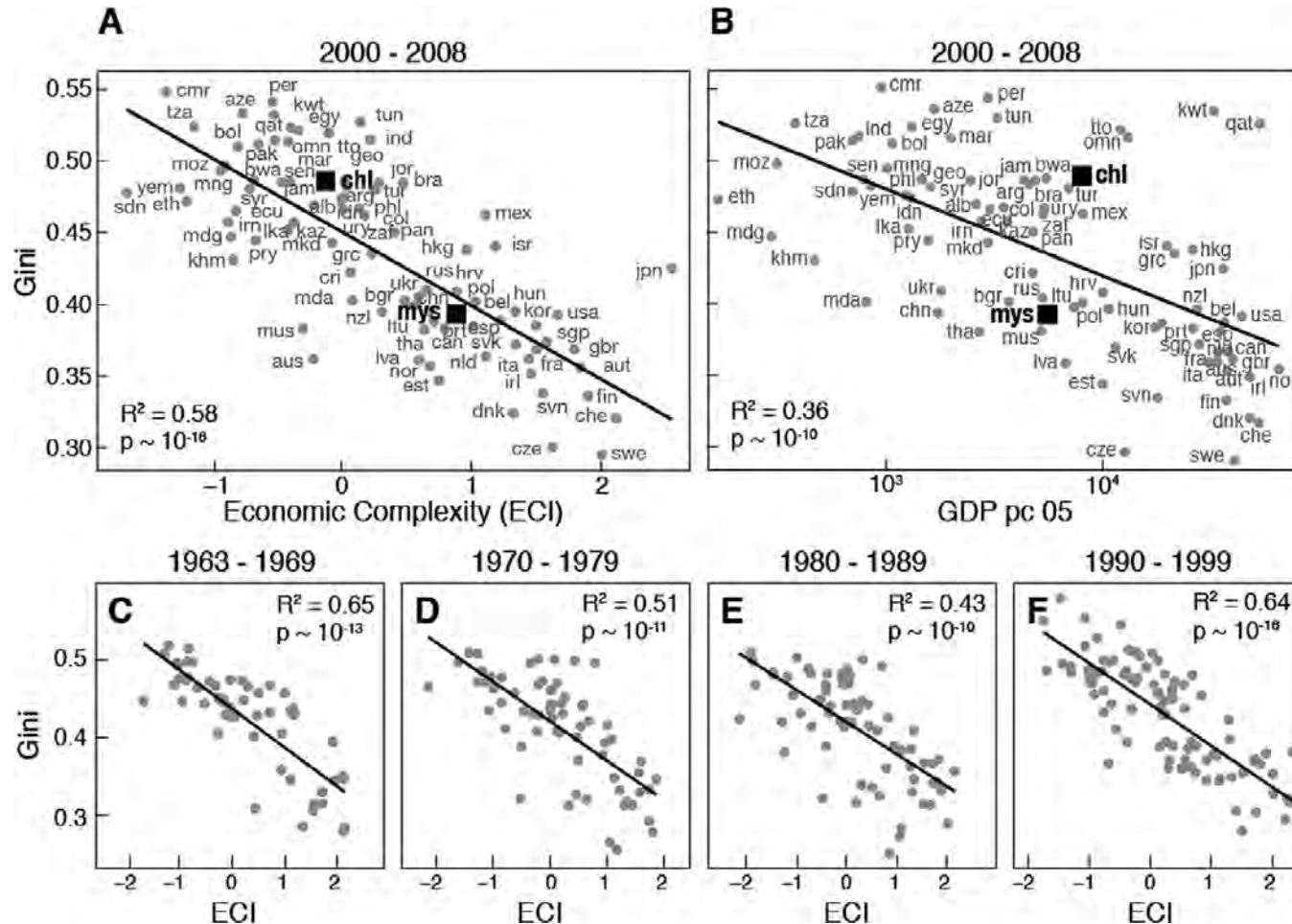
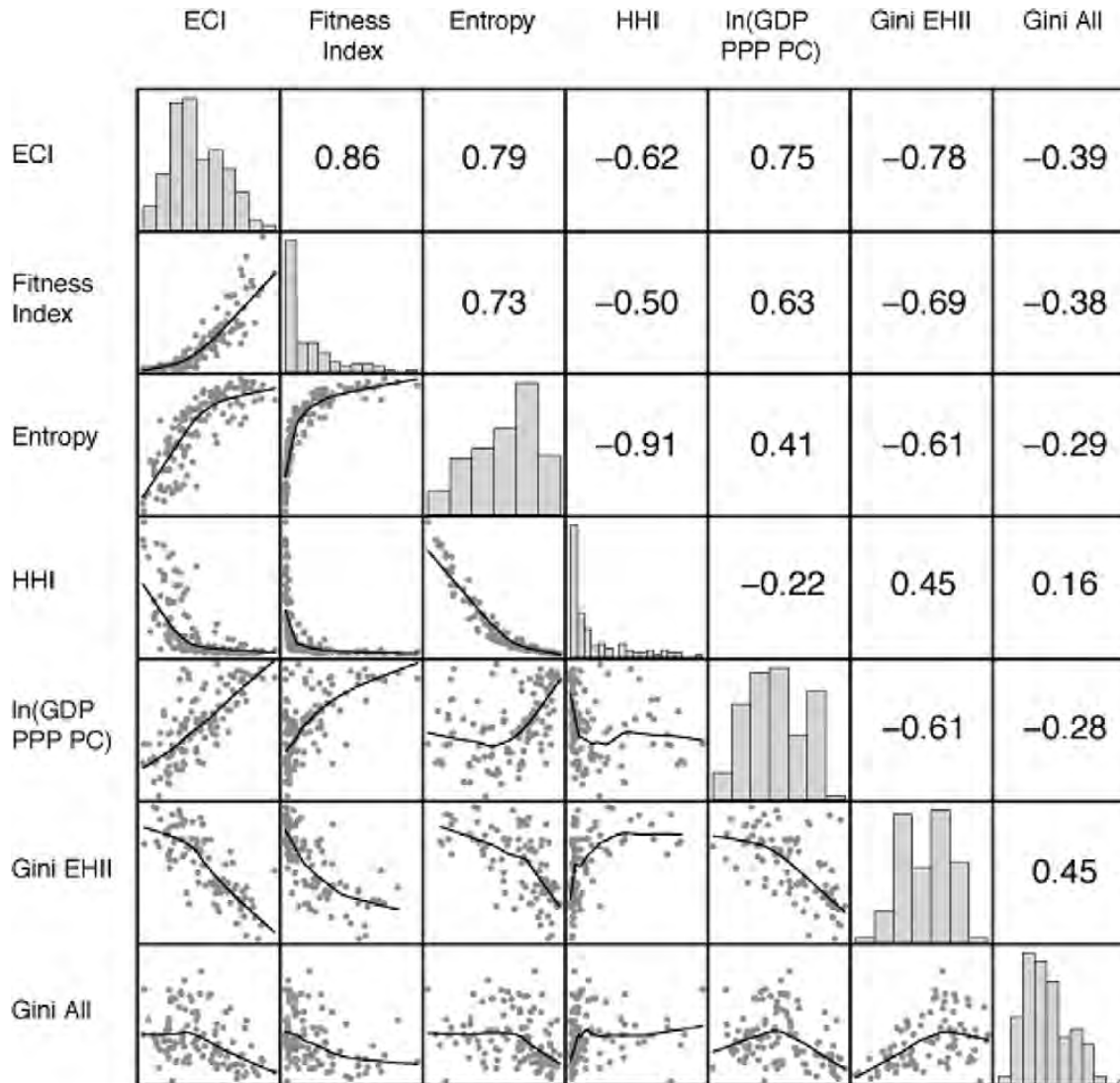


Figure 2. Bivariate relationships between economic complexity, income, and income inequality. Notes: All figures show that R^2 and all p-values are less than 10^{-10} . (A) ECI versus GINI EHII in 2000–08. (B) Natural logarithm of GDP per capita (constant 2005 US\$) versus GINI EHII. (C) ECI versus GINI EHII in 1963–69, (D) 1970–79, (E) 1980–89, and (F) 1990–99.

ECI and Inequality?



ECI and its relationship to the Gini

Table 2. Pooled OLS regression models

	Dependent variable: Gini					
	(I)	(II)	(III)	(IV)	(V)	(VI)
ECI	-0.040*** (0.007)		-0.037*** (0.007)	-0.046*** (0.007)	-0.033*** (0.006)	-0.044*** (0.006)
ln(GDP PPP pc)	0.067** (0.028)	0.059* (0.032)		0.060** (0.029)	0.056* (0.028)	0.075*** (0.025)
ln(GDP PPP pc) ²	-0.004** (0.002)	-0.004* (0.002)		-0.003* (0.002)	-0.003* (0.002)	-0.004*** (0.001)
Schooling	-0.005*** (0.002)	-0.009*** (0.002)	-0.004** (0.002)		-0.006*** (0.002)	-0.005*** (0.002)
Ln Population	0.007** (0.003)	0.0001 (0.003)	0.005* (0.003)	0.008*** (0.003)		0.009*** (0.002)
Rule of law	-0.013 (0.013)	-0.016 (0.014)	-0.016 (0.013)	-0.015 (0.013)	-0.013 (0.013)	
Corruption Control	0.011 (0.013)	0.027* (0.014)	0.009 (0.013)	0.016 (0.013)	0.007 (0.013)	
Government Effectiveness	0.002 (0.017)	-0.022 (0.018)	0.003 (0.017)	0.006 (0.017)	0.010 (0.017)	
Political stability	-0.010 (0.006)	-0.017** (0.007)	-0.009 (0.006)	-0.009 (0.006)	-0.017*** (0.006)	
Regulatory quality	-0.006 (0.012)	-0.012 (0.014)	-0.0002 (0.012)	-0.010 (0.012)	-0.012 (0.012)	
Voice and accountability	0.001 (0.008)	0.006 (0.009)	0.001 (0.008)	-0.004 (0.008)	0.003 (0.008)	
Constant	0.083 (0.130)	0.286** (0.141)	0.391*** (0.050)	0.068 (0.134)	0.244** (0.114)	0.016 (0.121)
Observations	142	142	142	142	142	142
R ²	0.717	0.639	0.701	0.699	0.704	0.704
Adjusted R ²	0.693	0.612	0.681	0.676	0.681	0.693
Residual std. error	0.035 (df = 130)	0.039 (df = 131)	0.035 (df = 132)	0.035 (df = 131)	0.035 (df = 131)	0.035 (df = 136)
F-statistic	29.916*** (df = 11; 130)	23.208*** (df = 10; 131)	34.413*** (df = 9; 132)	30.458*** (df = 10; 131)	31.165*** (df = 10; 131)	64.656*** (df = 5; 136)

ECI and its relationship to the Gini

Table 3. *Fixed-effects panel regression*

	<i>Dependent variable: GINI</i>						
	I	II	III	IV	V	VI	VII
ECI	-0.031*** -0.007	-0.033*** -0.007	-0.024*** -0.007	-0.026*** -0.007		-0.030*** -0.007	-0.029*** -0.007
ln(GDP PPP pc)		-0.038 -0.028	-0.042 -0.027	-0.017 -0.029	-0.032 -0.03		-0.053* -0.03
ln(GDP PPPpc) ²		0.003* -0.002	0.002 -0.002	-0.00003 -0.002	0.0005 -0.002		0.004** -0.002
Schooling			0.010*** -0.002	0.014*** -0.003	0.015*** -0.003	0.010*** -0.002	
Ln population				-0.024** -0.011	-0.016 -0.011	-0.022** -0.01	0.014* -0.008
Observations	338	338	338	338	338	338	338
R ²	0.077	0.123	0.198	0.213	0.165	0.196	0.134
Adjusted R ²	0.055	0.087	0.139	0.149	0.116	0.138	0.094
F-statistic	20.13*** (df = 1; 240)	11.14*** (df = 3; 238)	14.63*** (df = 4; 237)	12.80*** (df = 5; 236)	11.74*** (df = 4; 237)	19.36*** (df = 3; 238)	9.15*** (df = 4; 237)
Country fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

These seven fixed-effects panel regression models explore whether changes in a country's level of economic complexity are associated with changes in income inequality (column I), also controlling for the effects that other socioeconomic factors like income (column II), human capital (column III) and population (column IV) have on income inequality. Columns V–VII control the variance explained by the model when ECI, income, or schooling, are excluded from the analysis. The numbers in parenthesis are standard errors (SEM).

The Product Gini Index (PGI)

- The PGI is a way of decomposing income inequality at the product level.
- It estimates the average inequality (as measured through the Gini) of the countries which export a particular product.
- Mathematically:

$$PGI_p = \frac{1}{N_p} \sum_c M_{cp} s_{cp} Gini_c$$

M is 1 if a country exports product p with $RCA > 1$, and 0 otherwise; s is the share of a country's exports made up by product p . N_p is a normalizing factor to ensure that PGI_p is a weighted average of the Ginis only.

Translating the Gini into the product space

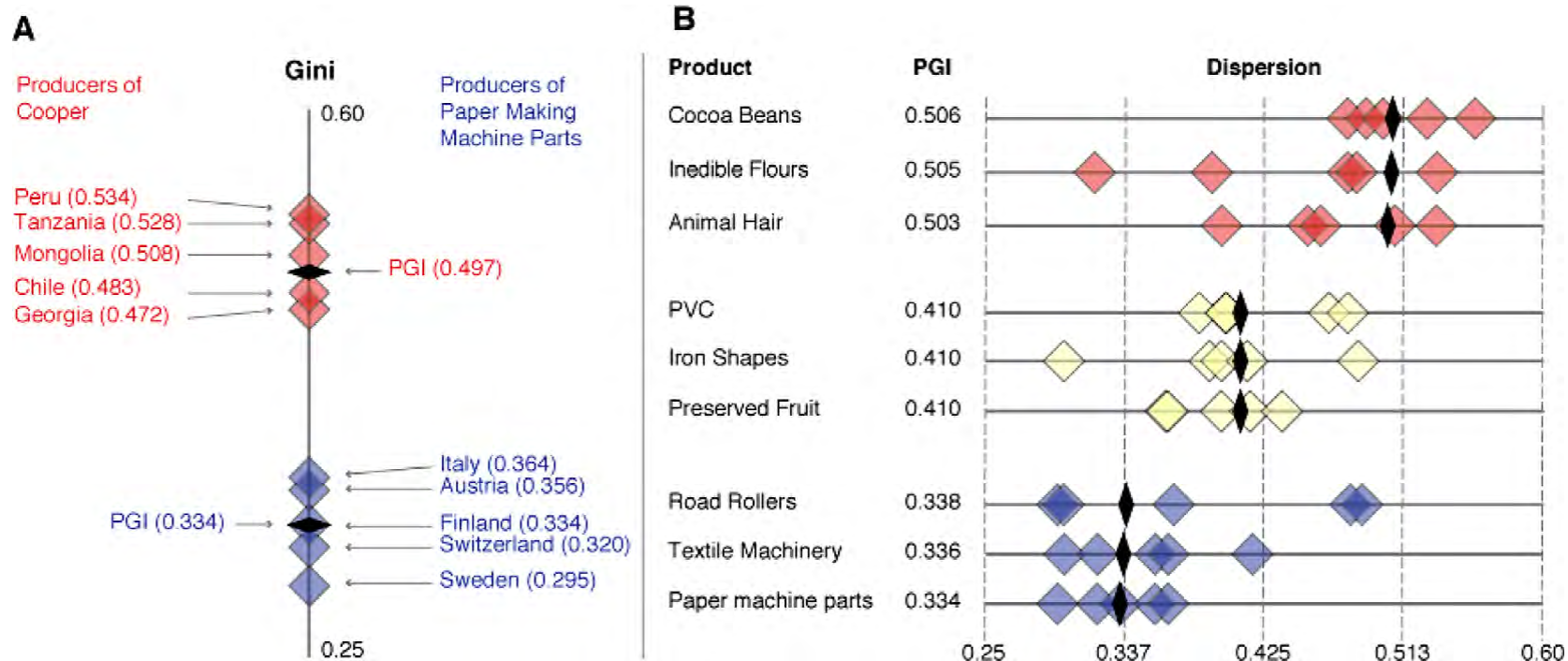


Figure 4. *The Product Gini Index (PGI)*. Notes: (A) The Product Gini Index (PGI) is a weighted average of the Gini coefficients of the countries that export a product. The Gini coefficients of five copper exporters are shown in red. In blue, we show the Gini coefficients of exporters of paper-making machine parts. (B). Top three, middle three, and bottom three products by PGI values. The PGI value is indicated with a black diamond. The Gini values of the five countries that contribute the most to each of these PGI are shown using diamonds. All values are measured using data from 1995 to 2008. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The product space and inequality

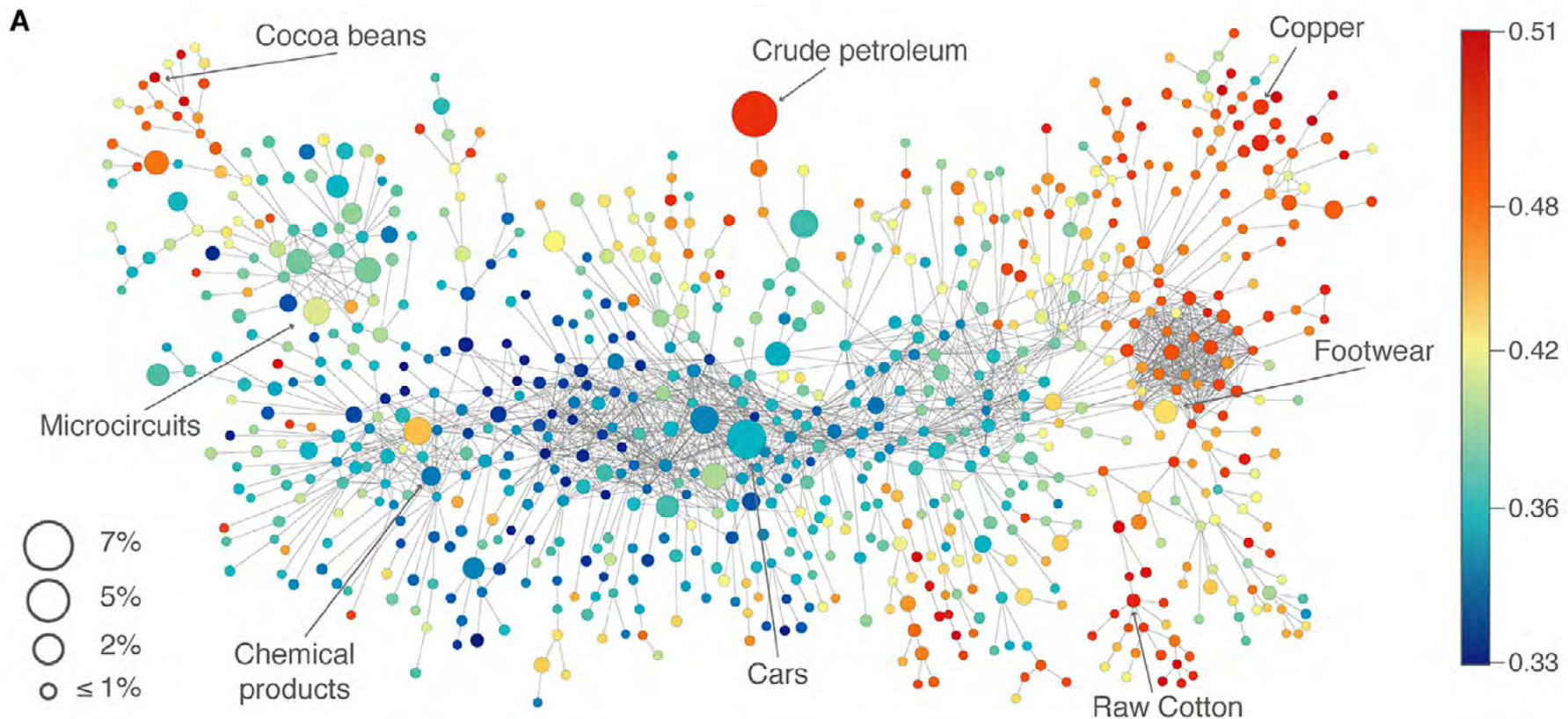


Figure 5. *The product space and income inequality. (A) In this visualization of the product space nodes are colored according to a product's PGI as measured during 1995–2008. Node sizes are proportional to world trade during 2000–08. The networks are based on a proximity matrix representing 775 SITC-4 product classes exported during 1963–2008. The link strength (proximity) is based on the conditional probability that the products are co-exported.*

What else can be done?

- As we have seen, economic complexity can be utilized in a number of diverse economic investigations.
- It is a powerful way of representing or visualising a country's economy, which is very flexible in its application to further research.
- Perhaps you could spend some time thinking about how your own research may benefit from this inclusion of the theory of economic complexity.

Thank you