Beyond the Theory of Complexity: Empirical Study of the Economic Complexity Index

Martin Saveski
MIT Media Lab
msaveski@mit.edu

Abdullah Almaatouq
MIT Media Lab
amaatouq@mit.edu

ABSTRACT

We show that ECI captures multitude of dimensions that characterizes a country's economy and that it is a robust predictor of the economic growth even when matching on a countrys background characteristics (income, population, life expectancy, years of education). We find that increase in ECI (low to high) results in 14.4% GDP growth in the subsequent five years. We also test the out-of-sample predictive power of ECI and show that it is better predictor of long-term growth. Finally, we perform simulations to help policymakers understand how ECI affects both short and long term growth. Specifically, we find that an increase in ECI leads to small short-term growth (1% in the next year) and large long-term growth (23% in the next 20 years).

INTRODUCTION

Hidalgo and Housmann [6] recently developed the *Method of Reflection (MR)*, which provides an original approach to measure the economic and product complexities based on a network view of international trade data. As a result of their methodology, two measures are presented: (*i*) the Economic Complexity Index (*ECI*); and (*ii*) the Product Complexity Index (*PCI*). The main idea is based on the argument that the economies can develop products close in terms of required capabilities to those they currently produce [7]. This theory accounts for the process by which countries accumulate capabilities; indicating that development efforts should focus on generating the conditions that would allow complexity to emerge to generate sustained growth and prosperity. The authors suggest that these complexity indicators (i.e., *ECI* and *PCI*) can capture information not only related to the countries' current income, but also can predict future growth.

Several follow-up studies investigated alternative methods, which correct conceptual weaknesses of the *MR* [3, 4, 13] or as a mean for measuring knowledge diffusion [1]. However, rigorous empirical validation of the indicators properties and developing frameworks and guidelines for policy-makers to better design industrial policies, remains an understudied area.

In this work, we investigate the main features of the *ECI* and show that this single measure can capture many information regarding capabilities. We also find that the *ECI* index is consistently associated with higher income growth and that this relationship persists even when we include detailed controls for the country's population, life expectancy, years of education (i.e., primary and secondary education) and the economy's size. Therefore, the robustness of the measure should make the indicator more appealing to empirical economists. In addition, we show that the *ECI* can best predict long-term growth rather than short-term one, which can be relevant to policy-makers.

PRELIMINARIES

This section explains how the *ECI* is computed followed the procedure proposed by [6] and [5]. We also describe the data used in order to conduct this work.

Computing the Economic Complexity Index

The first step in computing the *ECI* is to use Balassa's revealed comparative advantage [2], denoted by $RCA_{c,p}$, which is computed as follows:

$$RCA_{c,p} = \frac{x_{c,p}}{\sum_P x_{c,p}} \Big/ \frac{\sum_C x_{c,p}}{\sum_{c,p} x_{c,p}}, \tag{1}$$

where $x_{c,p}$ is the export value of product $p \in P$ by country $c \in C$. Therefore, the $RCA_{c,p}$ can be seen as the importance of a product p in country c total exports relative to the importance of the product across countries. After computing the $RCA_{c,p}$, it is possible to construct a matrix of countries (rows) and products (columns), denoted by $M_{c,p}$. Hidalgo et al. [6] uses $R^* = 1$ as a threshold for the $RCA_{c,p}$, where the matrix entry $M_{c,p} = 1$, when $RCA_{c,p} \geq R^*$ and $RCA_{c,p} = 0$ otherwise. This means that we will only consider those products that have a higher or equal weight in the country's export than in the global trade.

From the $M_{c,p}$ matrix, one can compute the complexity of a product k_p and the complexity of a country's economy k_c.

$$k_{p,n} = \frac{1}{k_{p,0}} \sum_{c=1}^{N_c} M_{c,p}.k_{c,n-1}, \tag{2}$$

$$k_{c,n} = \frac{1}{k_{c,0}} \sum_{p=1}^{N_p} M_{c,p}.k_{p,n-1}, \quad (3)$$

where n is the number of iterations used to define the indicators. This recursive relationship would ensure getting rid of distortionary effects. Therefore, $k_{c,n}$ converges to economic complexity index ECI and $k_{p,n}$ converges to the Product Complexity Index PCI, for large n values.

Conceptually, equation 2 captures the intuition that the number of countries that can make a product can tell us something about the variety of productive knowledge that a product requires. Similarly, equation 3 indicates that the number and type of products that a country makes can reveal something about the productive knowledge that a country has. This insight greatly facilitates the analysis because, while it is difficult to measure productive knowledge directly, it is easy to measure the number of products that a country makes, and how many countries make a given product. Therefore, the recursive definition in equations 3and 2 ECI can be simplified and mathematically defined as the eigenvector associated with the second largest eigenvalue a of a matrix connecting countries to countries, which is a projection of the matrix connecting countries to the products they export. Further mathematical derivation and definitions can be found in [7, 6, 5].

In this work, we focus on deepening our understanding of the ECI only, and examine its association with short and long term economic growth.

Datasets

All auxiliary data comes from the World Development Indicators (WDI) as reported by the World Bank [1], except data on per capita GDP at Purchasing Power Parity (PPP) for which we use the International Monetary Fond (IMF) [2].

DIMENSIONS OF ECONOMIC COMPLEXITY

Previous work [11, 3] has indicated the difficulty of interpreting the ECI for high iterations. As the MR method used to compute the ECI from the trade data gets rid of distortionary effects and reach an evaluation of economic complexity, one would expect it to be explained by different kinds of capabilities. Figure 1 shows the results of OLS estimation with ECI as a dependent variable and including covariates that can account for diverse economic dimensions, namely:

- *Tertiary enrollment*: the total enrollment in tertiary education, regardless of age, expressed as a percentage of the total population. One would expect that high economic complexity will generate new industries and encourage people to participate in tertiary education.

- *Industry value added*: the net output of a sector after adding up all outputs and subtracting intermediate inputs. This would capture the accumulated capabilities expressed in the country's industrial composition.

- *Exports of goods and services*: computed as a percentage of the GDP and includes the value of merchandise, freight, insurance, transport, and other services, such as communication, construction, financial, personal, and government services. The measure excludes compensation of employees and investment income. As the ECI is computed based on export data, one would expect that the exports of goods and services would explain a large proportion of the variance.

- *Natural resources rents*: the sum of oil rents, natural gas rents, coal rents (hard and soft), mineral rents, and forest rents [10]. Previous work (i.e., [6, 5]) showed that countries with high ECI tend to have comparative advantage on non-natural resource products [7]. Therefore, one would expect negative coefficient on the natural resource rents.

- *Population*: Finally, the logarithm of the population as a control for the size of the economy.

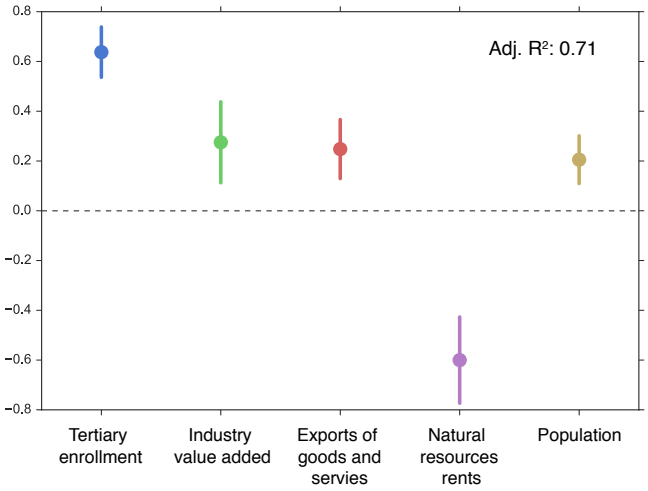


Figure 1: Economic Complexity Index (ECI) captures many dimensions of economic complexity. The plot shows the mean effect size of the covariates (solid circles) and the 95% confidence intervals (bars). The table of the results is included in the Appendix

We find that the ECI is remarkably capturing rich information regarding an economy's complexity. This result provide evidence supporting the hypothesis that ECI captures many more dimensions of economic complexity since all covariates are significant and yet part of the ECI variation is explained by unobservables. Figure 1 shows that human capital and manufactured capital (expressed by the tertiary enrollment and industry value added, respectively) are both positive and statistically significant ($p < 0.05$). On the other hand, the natural capital is negative and significant, as one would expect from the ECI theory. This result provides an empirical evidence that the theory captures the specialization a country has. In the traditional theory of growth [12], a country that is rich with natural resource (i.e., low ECI) has the same potential of economic growth of a country with more industrial reliance. Therefore, a country like Saudi Arabia –one

[1] http://data.worldbank.org/; accessed on 20 April 2015
[2] http://www.imf.org/external; accessed on 25 March 2015.

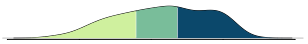
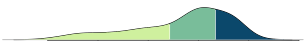
Variable	Min	Median	Max	Distribution
Economic Complexity Index	-2.78	0.04	2.41	
GDP at purchasing power parity per capita (log)	5.18	8.41	10.92	
Population (log)	12.8	16.1	21.0	
Life expectency	40.8	69.4	81.1	
Years of education	9.0	12.0	13.0	

Table 1: Statistics of the Economic Complexity Index (treatment variable) and four control variable capturing the background characteristics of a country. The different colors in the distribution plots corresponded to the coarsened values used for matching. The data consist of four 5-year increments between 1985 and 2005, resulting in total of 446 observations.

of the world's largest producer of petroleum– should be deincentivised to diversify its economy. However, as we proceed to show, the type of specialization captured by the *ECI* is a significant and robust indicator of future growth.

ROBUSTNESS OF ECI

Hidalgo and Hausmann [6] have shown that the ECI is correlated with country's level of income and that deviations from this relationship are predictive of future growth. In this section we study this hypothesis more thoroughly: (*i*) we test whether the predictive power of ECI is robust to the inclusion of different control variables, and (*ii*) we perform matching on country's background characteristics to eliminate model dependence and shed some light on a possible casual relationship.

Simple Regression

We start by replicating Hidalgo and Hausmann's findings, focusing on four 5-year periods between 1985 and 2005 (85-90, 90-95, 95-00, 00-05). We regress the log of the income growth on the ECI, controlling for the initial level of GDP per capita:

$$\log\left(\frac{\text{GDP}(t+\Delta t)}{\text{GDP}(t)}\right) = \beta_0 + \beta_1 \log(\text{GDP}(t)) + \beta_2 \text{ECI}(t)$$

We find that ECI is highly significant predictor of the rate of economic growth ($p \ll 0.01$) and that larger ECI corresponds to higher growth rate (Table 3, Model 1). This is consistent with [6].

Including Control Variables.

Next, we test whether ECI is still a robust predictor of economic growth when controlling for different background characteristics of a country. We consider: population size, life expectancy, and years of education. Table 1 shows the distributions and summary statistics of these variables. Again, we regress the log of the income growth on the ECI, now controlling for the initial values of all variables:

$$\begin{aligned}\log\left(\frac{\text{GDP}(t+\Delta t)}{\text{GDP}(t)}\right) = \beta_0 &+ \beta_1 \text{ECI}(t) + \beta_2 \log(\text{GDP}(t)) \\ &+ \beta_2 \log(\text{population}(t)) + \beta_3 \text{life exp}(t) \\ &+ \beta_4 \text{years of education}(t).\end{aligned}$$

ECI is still a significant predictor of the rate of economic growth ($p = 0.03$), with a positive coefficient (Table 3, Model 2). Also, life expectancy is a significant predictor. At the end of this section, we show how that influences the quantity of interest.

Matching

CEM. To control for the confounding influence of pretreatment control variables we perform Coarsened Exact Matching [8]. CEM rest on the observation that continues variables have natural breaking points and that they can be coarsened so that substantively indistinguishable values are grouped and assigned the same numerical value. Then, exact matching (i.e., matching treatment and control units that have the exact same values of the covariates) is applied to the coarsened data to determine the matches and to prune unmatched units. Finally, the coarsened data are discarded and the original (uncoarsened) values of the matched data are retained.

Coarsening. We first choose a reasonable coarsening for each variable. Table 1 column distribution shows the distribution of each variable, where the colors correspond to coarsened regions. We divide ECI (the treatment variable) in there equal frequency bins: low [-2.8, -0.6), medium [-0.58, 0.4), and high [0.4, 2.4). For GDP (log) we use the following cutpoints: [5.18, 7.58), [7.58, 8.73), [8.73, 10.92). We divide the population size (log) in three bins: [12.8, 15.6), [15.6, 16.8), [16.8, 21.0). For life expectancy we use: [40.8, 64.3), [64.3, 73.4), [73.4, 81.1). Finally, we divide years of education in two bins: (9, 11] and (12, 13]. The initial multidimensional imbalance is $\mathcal{L}_1 = 0.906$.

After running CEM with the above coarsening, we obtain $m_{\text{ECI=low}} = 54$ units with low ECI, matched with $m_{\text{ECI=medium}} = 107$ units with medium ECI, and $m_{\text{ECI=high}} = 67$ units with high ECI. This results in total of 228 matched and 218 unmatched units. The multidimensional imbalance after matching is $\mathcal{L}_1 = 0.895$, corresponding to 1.23% imbalance reduction.

Estimation. Next, we estimate the causal effect under CEM. We compute the FSATT (Feasible Sample Average Treatment on the Treated), using the CEM-matched units via linear regression using all variables in the data set. We obtain the following estimates: $\tau_{low \to medium} = 0.104$ (i.e., 10.4%

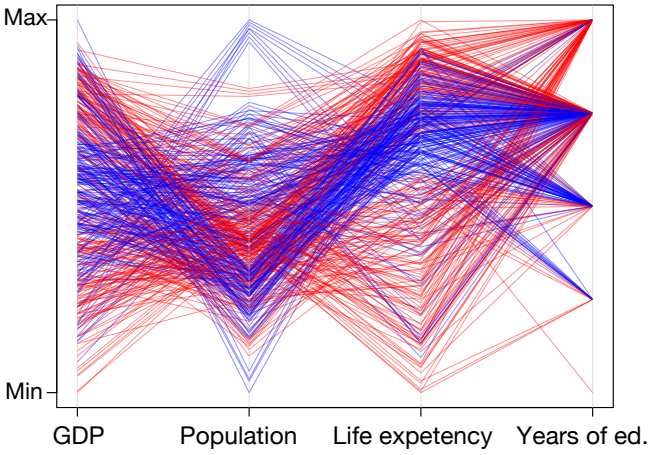


Figure 2: Parallel plot of numerical variables for CEM matched units (blue) against unmatched units (red).

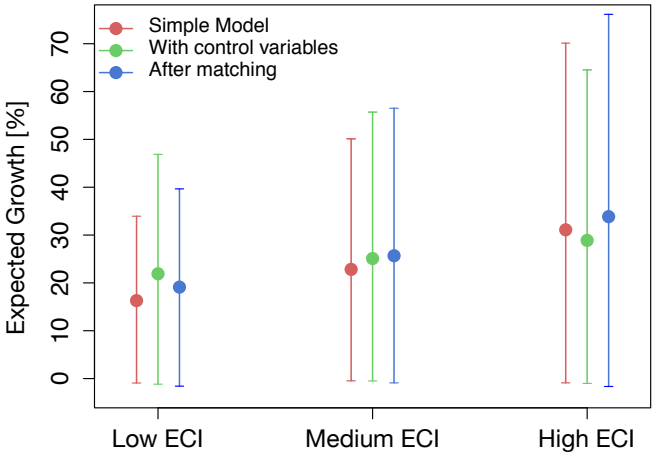


Figure 3: Expected growth of a median country with low, medium, and high ECI. The different lines correspond to different models. Red: Growth ECI + GDP (Simple Model). Blue: Growth ECI + GDP + Population + Life expectancy + years of education (With control variables). Red: applying the the same model, but using Coarsened Exact Matching as a preprocessing step (After matching).

growth), $\tau_{medium \to high} = 0.0376$ (i.e., 3.76% growth), and $\tau_{low \to high} = 0.144$ (i.e., 14.4% growth). Table 3 (Model 3), shows the results of the regression; both ECI-medium and ECI-high are statistically signifiant.

Defining the FSATT. Following [8], we define the estimands via a parallel plot (Figure 2). We represent each observation in the original data set by a single line that traces out its values on each of the continuous variables (horizontally) between its minimum and maximum values (vertically). The matched units are colored blue, while the unmatched units are colored red. The estimands are the average treatment effects for countries with different levels of GDP, population size, and years of education, but mostly high levels of life expectancy.

Comparison between models. We compare the estimates of expected growth for a median country produced by the different models. We set x to the median values for GDP, population, life expectance and year of education, and we vary the ECI, setting it to the mean value of each category (low: -1.13, medium: -0.05, high: 1.22). We perform 1000 simulations to incorporate the estimation and fundamental uncertainty. Figure 3 shows the expected values and the 5% confidence intervals. There are two things to note: (*i*) the estimates of all models are very similar, even without including control variable and matching we obtain reliable estimates; (*ii*) in all cases larger ECI results in increased economic growth, confirming our findings in the previous section.

OUT OF SAMPLE PREDICTION OF GROWTH

In the Section "Robustness of ECI", we have shown that the ECI is a significant predictor of the economic growth in the subsequent five years, even when controlling for different variables and matching. In this section, we test the out-of-sample predictive power of the ECI over different time increments.

We consider time increments between one and twenty years in the period from 1985 to 2005. For each increment we create a dataset of all (non-overlapping) occurrences of the increment. For instance, to build the dataset for ten year increments, we consider the growth between 1985-1995, and between 1995-2005. We perform 10-fold cross-validation on each of these datasets as follows. (*i*) We randomly partition the dataset in ten equal size subsets, we pick one subset and we set it aside, (*ii*) using the remaining nine subsets, we build a linear regression model of the economic growth (between the start and the end of the period) as a function of the ECI at beginning of the period; (*iii*) we use the model to predict the remaining subset and we measure the prediction error, repeating this procedure for every subset.

To measure the prediction error, we use *Root Mean Square Error*:

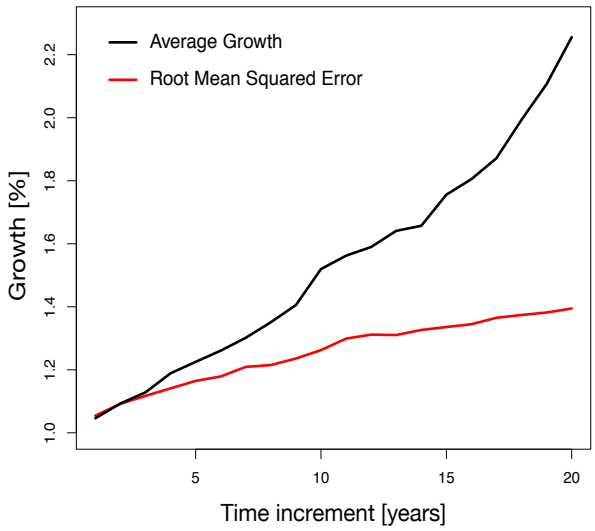
$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(p_i - y_i)^2}{n}},$$

where p_i is the predicted growth, y_i is the observed growth, and n is the number of observations. However, RMSE is not comparable across time increments, due to differences in magnitude of growth. For instance, the average growth of a 20-year increments is much higher then the 5-year increments, and although the RMSE of the 5-year increments maybe smaller, the error maybe more significant relative to the average growth. Thus, to compare the errors between different time increments, we compute the coefficient of variation (CV) of the RMSE, i.e., we normalize by the average growth of the increment:

$$\text{CV(RMSE)} = \frac{\text{RMSE}}{\bar{y}}$$

where $\bar{y}$ is the average growth.

We find that the ECI is more predictive of the economic growth in the long run and has limited predictive power for smaller time increments (Figure 4, left). When predicting



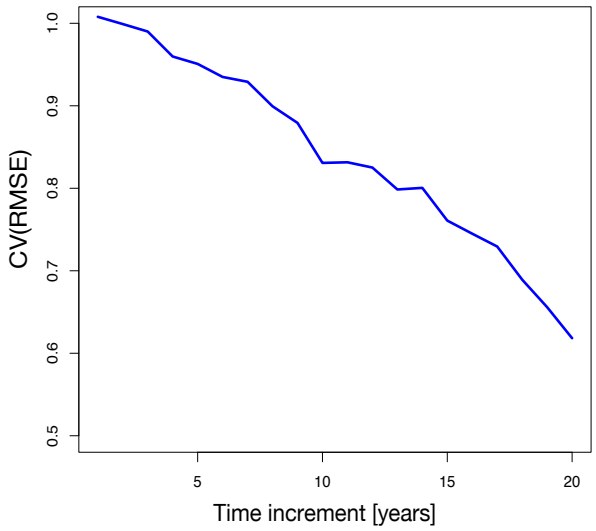


Figure 4: *Left*: Root Mean Squared Error (RMSE) and Average Growth; *Right*: Coefficient of Variation of the RMSE (i.e., RMSE / Average growth) for different time increments. The CV(RMSE) measure the out-of-sample predictive power of ECI and is comparable across time increments.

the growth in the next year the ECI maybe off by 5% (Figure 4, right), as much as the average growth for one year increments. In contrast the prediction error for five and ten year increments is 16% and 26%, or 95% and 83% of the average growth, respectively. Finally, ECI is most predictive on the twenty year increments being off by only 39% or 62% of the average growth.

Note that the errors reported are obtained using only the ECI as the explanatory variable.

RETURN OF INVESTMENT

Policymakers are often faced with the choice of which products or industries their country should invest in to generate sustained economic growth and prosperity. While, it is easy to predict how investing in a new product will increase the ECI of a country, it is unclear how this investment will influence its future growth. In this section, we perform simulations to help policymakers understand how increase in ECI affects both long and short term economic growth.

We collect data on the economic complexity and background characteristics of all countries in the period between 1985 and 2005 (see Subsection Datasets) and we consider time increments between one and twenty years. For each time increment, we build a linear regression model of the economic growth between the start and end of the period as a function of the initial values of GDP, population size, life expectancy, and years of education.

We focus on the expected economic growth of a median country. We set the initial values of GDP, population size, life expectancy, and years of education to the median values for the time increment and we compute the first differences between low (20% quantile) and high (80% quantile) ECI. Using Zelig [9], we perform 1000 simulations to account for the fundamental and estimation uncertainty.

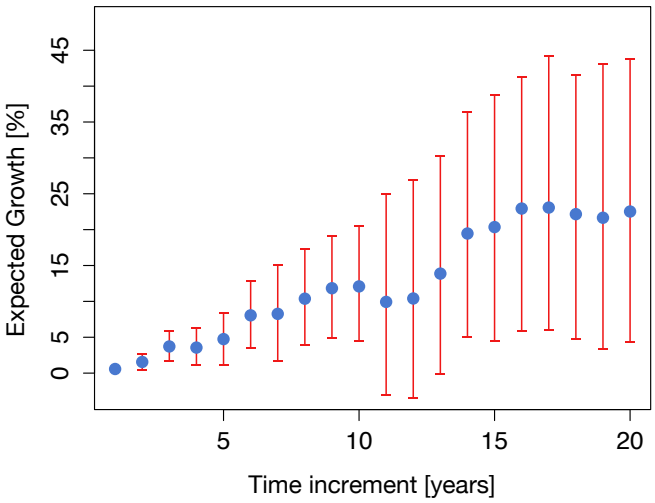


Figure 5: Estimate of the first differences in expected growth between high and low ECI for a median country. The bars correspond to 5% confidence intervals obtained from 1000 simulations.

Figure 5 shows the estimations of the expected growth and the 5% confidence intervals. As one may expect, the expected economic growth in the next year is small (i.e., 1%). An increase in the ECI results in expect growth of 5% and 12% in the next five and ten years, respectively. In the long term, an increase in the economic complexity is expected to generate a growth of 20% and 23% in the following 15 and 20 years, respectively.

Note that time increments bigger than ten years have larger confidence intervals. This is because we have less observations for these increments (e.g., there is one 15-year time increment between 1985 and 2005, but there are four 5-year increments).

Summary and Discussion

Many published work have discussed the relationship between a country's current production status and how this effects the long term growth possibilities. More recently, Hidalgo and Housmann [6] has introduced a very robust and useful tool to compute the *economic complexity* of a country and the sophistication of a product. In this work, we investigate the main features of the *ECI* and show that this single measure can capture many information regarding capabilities. We also find that the economic complexity measure significantly explain current per capita GDP. Through rigorous tests of the robustness of the measure, we find this significant statistical relationship to persists even when we include detailed controls for the country's population, life expectancy, years of education and the size of the economy. Finally, we find that the complexity measure highly predicts future long-term growth. This conclusion is supported by a wide body of literature in economic growth theory which states that the relationship between production and growth materializes over the long term.

Finally, combining our results with the previous work in the literature can provide policy-makers with detailed insights to better design industrial policies that help economies diversify in a systematic way and accumulate important capabilities, which would generate sustained growth and prosperity.

ACKNOWLEDGMENTS

The authors would like to thank James M. Sappenfield for his many comments and suggestions. We also highly appreciate the technical help and support received by Stephen Pettigrew and Soledad Prillaman. Finally, we are very grateful for sincere and valuable guidance and encouragement extended to us by Prof. Gary King. In addition, the authors thank the Center for Complex Engineering Systems (CCES) at KACST and MIT and the Media Lab at MIT for their support.

REFERENCES

1. Bahar, D., Hausmann, R., and Hidalgo, C. A. Neighbors and the evolution of the comparative advantage of nations: Evidence of international knowledge diffusion? *Journal of International Economics 92*, 1 (2014), 111–123.
2. Balassa, B. Trade liberalisation and revealed comparative advantage1. *The Manchester School 33*, 2 (1965), 99–123.
3. Caldarelli, G., Cristelli, M., Gabrielli, A., Pietronero, L., Scala, A., and Tacchella, A. A network analysis of countries export flows: firm grounds for the building blocks of the economy. *PloS one 7*, 10 (2012), e47278.
4. Cristelli, M., Gabrielli, A., Tacchella, A., Caldarelli, G., and Pietronero, L. Measuring the intangibles: A metrics for the economic complexity of countries and products. *PloS one 8*, 8 (2013), e70726.
5. Hausmann, R., and Hidalgo, C. A. *The atlas of economic complexity: Mapping paths to prosperity*. MIT Press, 2014.
6. Hidalgo, C. A., and Hausmann, R. The building blocks of economic complexity. *Proceedings of the National Academy of Sciences 106*, 26 (2009), 10570–10575.
7. Hidalgo, C. A., Klinger, B., Barabási, A.-L., and Hausmann, R. The product space conditions the development of nations. *Science 317*, 5837 (2007), 482–487.
8. Iacus, S. M., King, G., and Porro, G. Causal inference without balance checking: Coarsened exact matching. *Political analysis* (2011).
9. Imai, K., King, G., and Lau, O. Zelig: Everyones statistical software. *R package version 3*, 5 (2009).
10. Jarvis, M., Lange, G.-M., Hamilton, K., Desai, D., Fraumeni, B., Edens, B., Ferreira, S., Li, H., Chakraborti, L., Kingsmill, W., et al. *The changing wealth of nations: measuring sustainable development in the new millennium*. 2011.
11. Ourens, G., et al. Can the method of reflections help predict future growth? *Université catholique de Louvain, Institut de Recherches Economiques et Sociales (IRES)* (2013).
12. Ramsey, F. P. A mathematical theory of saving. *The economic journal* (1928), 543–559.
13. Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A., and Pietronero, L. Economic complexity: conceptual grounding of a new metrics for global competitiveness. *Journal of Economic Dynamics and Control 37*, 8 (2013), 1683–1691.

Appendix

Dimensions of Economic Complexity Table

	Model 1
(Intercept)	0.05 (0.06)
tertiary_education_2005	0.64 (0.06)***
industry_value_added_2005	0.28 (0.10)**
exports_of_goods_and_services_2005	0.25 (0.07)***
natural_resources_rents_2005	−0.60 (0.10)***
population_2005	0.21 (0.06)***
R^2	0.73
Adj. R^2	0.71
Num. obs.	95

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 2: Economic Complexity Index (ECI) captures many dimensions of economic complexity. The table shows the effect size of the covariates and Standard Errors in brackets for 95 countries in 2005.

Matching Regressions

	Model 1	Model 2	Model 3
(Intercept)	0.467 (0.068)***	0.317 (0.182)	0.073 (0.356)
GDP per capita (log)	−0.031 (0.008)***	−0.061 (0.011)***	−0.049 (0.017)**
Population (log)		0.005 (0.005)	0.020 (0.008)*
Life Expectancy		0.008 (0.001)***	0.008 (0.003)**
Years of Education		−0.017 (0.010)	−0.030 (0.018)
ECI	0.050 (0.010)***	0.023 (0.011)*	
ECI Medium			0.083 (0.026)**
ECI High			0.089 (0.029)**
R^2	0.058	0.131	0.204
Adj. R^2	0.054	0.121	0.183
Num. obs.	446	446	228

***$p < 0.001$, **$p < 0.01$, *$p < 0.05$

Table 3: Robustness of the Economic Complexity Index, where the dependent variables is the log of the economic growth. See Section “The Robustness of ECI”.