9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

Quantifying the technological foundations of economic complexity

Economic complexity, Synergy, Information theory, Innovation, Resilience

Extended Abstract

The transformative process of turning a set of inputs into an output is prevalent in various socioeconomic and physical systems. In socioeconomic systems, it often takes the shape of a production process, and different disciplines analyze it through distinct theoretical frameworks and quantitative tools. For instance, management sciences specialize in supply chains and global value chains, human and economic geography utilize systems analysis, economists study input-output models and fit production functions, systems engineering construct design structure matrices, and innovation scholars focus on combinatorial models and network analysis. Independently of the domain of application, it is commonly agreed that technological sophistication is a building block in the study of economic complexity. Thus, quantifying the degree and structure of technological sophistication is critical to understand the evolution, performance, fragility, and resilience of production systems.

By production process, we refer to the procedure through which a certain technology transforms a set of inputs into an output. Due to the focus of this paper, we use the terms *technological, industrial, productive, and economic sophistication* interchangeably, to refer to the capacity of a production process to generate novel outputs from a set of inputs. Thus, an industry is considered more sophisticated if it is produces more novelty than others when using the same inputs. In this paper[1], we develop a method to quantify the technological sophistication of production processes. We test our method using two major input-output datasets[2, 3], and validate it through independent export database[4] and well-established economic complexity indices[5, 6, **?**] that are commonly used as proxies for technological sophistication at a national level. Broadly speaking, we leverage the PID framework[7] to estimate the mutual information between pairs of inputs in a given industry (from IO tables), and decompose their contribution to the output into different modes of information sharing. We focus on a particular mode known as *synergistic information*, i.e., the information about the output that cannot be obtained from any of the inputs alone but only exists as a virtue of the interaction between the inputs. Using this information, we measure the degree of synergy between the inputs and estimate a *synergy score*. A higher synergy score means that input interactions produce more novel information during the production process. This interpretation aligns with the established notion of a recombinant process generating novel outputs. Thus, the central thesis of the work claims that the synergy score quantifies technological sophistication. In other words, an industry that exhibits higher synergy scores should have a more sophisticated technology underlying its production process. We find that our synergy score predicts these indices (see Figure 1), and that the inferred synergistic interaction networks have non-trivial topologies that characterize various production technologies. To the extent of our knowledge, this is the first framework that quantifies technological sophistication by empirically inferring the nature of input-input and input-output relationships (as opposed to assuming them). Our approach is non-parametric, so it does not require the *ex ante* assumption of specific production functions or design structure matrices. Instead, by exploiting the mutual information between input data in an output

signal, it allows to infer the structure of synergistic interactions. This is a major innovation as it facilitates both the overall quantification of technological sophistication across firms, industries, sectors, and countries, as well as the estimation of the interaction networks living at the heart of production processes; currently considered a black box by economic complexity scholars. Furthermore, it provides an empirical basis to justify the specification of certain production functions in input-output models. The method is of general purpose as it can adapted to other contexts where transformative processes may be important (both in socioeconomic and physical systems).

References

[1] Hardik Rajpal and Omar A Guerrero. Quantifying the technological foundations of economic complexity. *arXiv preprint arXiv:2301.04579*, 2023.

[2] Manfred Lenzen, Daniel Moran, Keiichiro Kanemoto, and Arne Geschke. Building Eora: A Global Multi-Region Input–Output Database at High Country and Sector Resolution. *Economic Systems Research*, 25(1):20–49, March 2013.

[3] Joaquim Guilhoto, Geoffrey Hewings, Nick Johnstone, Colin Webb, and Norihiko Yamano. Exploring changes in world production and trade: Insights from the 2018 update of oecd's icio/tiva database. 2019.

[4] Guillaume Gaulier and Soledad Zignago. BACI: International Trade Database at the Product-Level. The 1994-2007 Version. Working Paper, CEPII research center, October 2010.

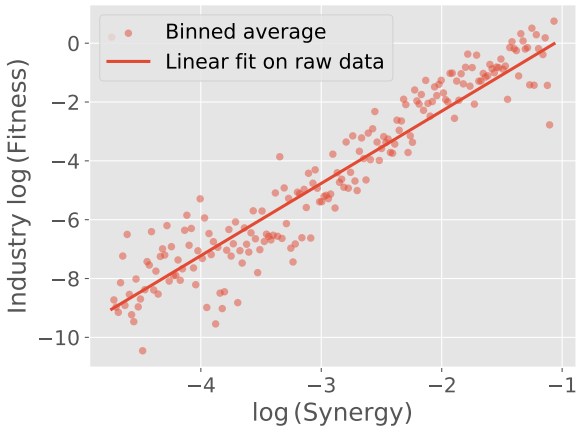
[5] C. Hidalgo and R. Hausmann. The Building Blocks of Economic Complexity. *Proceedings of the National Academy of Sciences*, 106(26):10570–10575, 2009.

[6] M. Cristelli, A. Gabrielli, A. Tacchella, G. Caldarelli, and L. Pietronero. Measuring the Intangibles: A Metrics for the Economic Complexity of Countries and Products. *PLOS ONE*, 8(8):e70726, 2013.

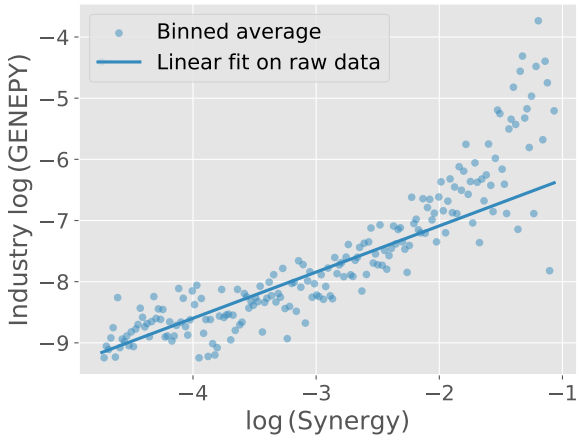
[7] Paul L Williams and Randall D Beer. Nonnegative decomposition of multivariate information. *arXiv preprint arXiv:1004.2515*, 2010.

Figure 1: Synergy scores and their association to industry sophistication

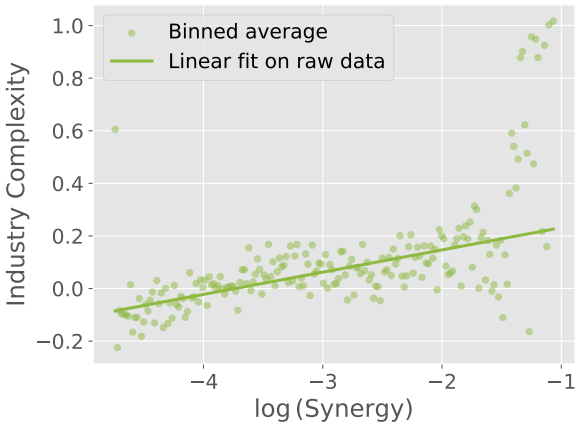
(a) Economic fitness index



(b) GENEPY index



(c) Economic complexity index



Notes: Synergy scores across all industries and countries are found to be correlated to their respective industry level indices of technological sophistication. The data is binned by synergy scores for visualization purposes, the linear fit is calculated on the entire dataset.