CESifo WORKING PAPERS

Classifying Top Economists Using Archetypoid Analysis

Sabine Gralka, Klaus Wohlrabe

CES**ifo**

Classifying Top Economists Using Archetypoid Analysis

Abstract

Updating the study by Seiler and Wohlrabe (2013) we use archetypoid analysis to classify top economists. The approach allows us to identify typical characteristics of extreme (archetypal) values in a multivariate data set. In contrast to its predecessor, the archetypal analysis, archetypoids *always* represent actual observed units in the data. Using bibliometric data from 776 top economists we identify four archetypoids. These types represent solid, low, top and diligent performer. Each economist is assigned to one or more of these archetypoids.

Sabine Gralka
Faculty of Business and Economics
TU Dresden / Germany
ORCID: 0000-0001-8303-4297
sabine.gralka@tu-dresden.de

Klaus Wohlrabe*
Center for Macroeconomics and Survey
ifo Institute – Leibniz Institute for Economic
Research at the University of Munich
Munich / Germany
ORCID: 0000-0002-5651-2641
wohlrabe@ifo.de

*corresponding author

Article Highlights:
- Archetypoid analysis to categorize top economists

- Based on seven bibliometric indicators for 776 economists obtained from the RePEc database

- Identification of four archetypoids, which are characterized as the low, solid, top and diligent performer

1. Introduction

Within this paper archetypoid analysis is used to classify top economists. The classification provides a straightforward overview of the various top researchers working in the area of economics. In times of an increasing number of scholars across all research fields (Bloom et al. 2020) such an easy tool for classification is valuable. The present paper builds upon the studies by Seiler and Wohlrabe (2013), evaluating economists and Wohlrabe and Gralka (2020a), categorizing economics faculties and authors. With a similar goal, the former study used archetypal analysis to categorize scientists. A drawback of this approach is that identified archetypes might be artificial, i.e., no entity from the employed dataset can fully be assigned. In line with the latter study we therefore employ the archetypoid analysis. In contrast to its predecessor, the archetypal analysis, archetypoids always represent actual observed units in the data. Hence, we are able to identify those top economists which represent identified archetypoids perfectly. Parallel to both studies we employ RePEc (Research Papers in Economics) data for our investigation, focusing on the top economists worldwide. While the data used by Seiler and Wohlrabe (2013) refers to 2011, our dataset includes all years up to 2019. The RePEc database has evolved tremendously both with respect to quantity and quality over time. We therefore think that the update is necessary and useful.

Archetypal analysis has aroused the interest of researchers working in various fields, such as astrophysics (Chan et al. 2003), machine learning (Morup and Hansen 2012, Seth and Eugster 2016), and sports (Seth and Eugster 2016). The same applies for the archetypoid analysis, which was used for the evaluation within fields such as sports (Vinué and Epifanio 2017) and the financial stock market (Moliner and Epifanio 2018).

2. Methodology and Data

Due to space constraints we refer the reader to Wohlrabe and Gralka (2020b) for a short introduction of archetypoid analysis and corresponding references. Vinué et al. (2015) provide a detailed description of the approach including computational issues. For our analysis we used the R package *Anthropometry* version 1.13 by Vinué (2017).

We employ seven bibliometric indicators from the RePEc (Research Papers in Economics) website (http://www.repec.org).[1] RePEc offers currently 40 rankings of registered authors based on bibliometric data. However, these rankings are often highly correlated, see Zimmermann (2013) for further details. In line with Seiler and Wohlrabe (2013) and Wohlrabe and Gralka (2020a) we use the following seven ones:

- Published work, which includes working papers, books, software codes, and chapters
    - Number of distinct works, unweighted
    - Number of distinct works, weighted by simple impact factor

---

[1] RePEc data has often been used in bibliometric studies. See, inter alia, Rath and Wohlrabe (2016), Meyer and Wohlrabe (2018), Sommer and Wohlrabe (2017), Bornmann et al. (2018) or García-Suaza et al. (2020).

- Citations, which represent the impact of an author
  - Number of citations, unweighted
  - Number of citations, weighted by the simple impact factor
- Pages, which account for the published articles
  - Number of pages, unweighted
  - Number of pages, weighted by the simple impact factor
- Number of downloads, which shows the access

For each of our m = 7 bibliometric indicators we collected the corresponding economists for which the scores were publicly available in January 2020.[2] The scores refer to the December 2019 ranking. We ended up with N = 776 economists. A full list can be found in Wohlrabe and Gralka (2020b). It is important to note that due the construction of the data set we only consider top economists which are among the top 5% worldwide in each category.

The descriptive statistics for the original indicators can be found in Table 1. We report the mean, median as well as the minimum and maximum score. On average an economist in our sample has published around 152 articles which can either be working papers or journal articles. These publications contain around 1,600 pages in (refereed) journals and were downloaded around 1,000 times. The top economists gather 4,800 citations on average. The most cited economist is Andrei Shleifer from the Harvard University. It is thereby important to bear in mind, that the numbers describe stock levels (up to December 2019), which explains the high values of the indicators. The standard deviations are quite large, at least half of the mean. This shows that the class of top economists are quite heterogenous with respect to bibliometric indicators.

Table 1: Descriptive statistics

776 Observations	**Mean**	**Std. Dev.**	**Min**	**Max**
Number of Distinct Works	152	80	75	1,102
-, Weighted by Simple Impact Factor	4,187	2,757	1,133	23,228
Number of Citations	4,794	4,520	921	46,645
-, Discounted by Citation Age	55,125	57,056	8,930	568,616
Number of Pages	1,644	832	737	6,967
-, Weighted by Simple Impact Factor	36,481	26,261	10,759	253,971
Number of Downloads	1,063	829	349	8,057

---

[2] For each ranking only the top 5% of all registered economists are shown on the RePEc website.

3. Results

A priori it is not clear how many archetypoids are reasonable given the seven indicators for each economist. In Figure 1 we show the scree-plot of the RSS, which is used to determine the number of archetypoids to retain using the so-called elbow criterion. Based on the scree-plot we perform an analysis with four potential archetypoids.[3]
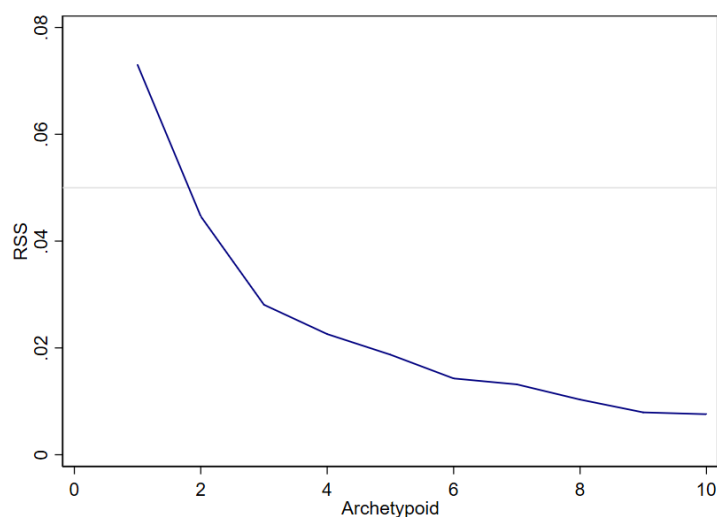
Figure 1: RSS-plot

Figure 2 shows the bar-plots representing percentiles for four archetypoids. Bar-plots thereby serves as a different graphical representation of the convex hull. The height of each bar-plot denotes the share of the convex hull relative to the maximum in the category. Hence, the y-axis denotes the relevance of the indicator for the archetype. Table 2 provides additional context for the percentiles shown in Figure 2. The table reports the archetypoid value for each bibliometric indicator, when the reported percentiles are applied to the original dataset. To give an example: while archetypoid 1 published 146 works, archetypoid 2 published 75 articles.

The four archetypoids can be interpreted as follows:

• Archetypoid 1 represents the *low performer* with a relatively high number of working papers and articles, but a relatively low number of citations and downloads as can be seen in Table 2. The weighted impact is the smallest of all archetypoids

• Archetypoid 2 denotes the *solid performer* with a relatively low number of works and downloads compared the first archetypoid. However, the quality-weighted impact of this work is higher.

• Archetypoid 3 represents the *top performer* among the top-level authors, given the other archetypoids. The author performs well across all indicators, even if the indicators are

---

[3] Since a potential flattening can also be detected for 3 and 5 archetypoids we tested how the results change, if a different number of archetypoids is considered. The change mainly affects the performers between the top and low class most. Simply put, if 3 archetypoids are considered, the solid and diligent performer are summed up to one group. If 5 archetypoids are considered, the solid performer is further distinguished into a (relatively) top and low class. The detailed results can be provided by the authors upon request.

quality-weighted. Table 2 shows that the corresponding values for citations are by far larger than those of the other archetypoids.

- Archetypoid 4 denotes the *diligent performer*, one between the previous two extremes, with a relatively high number of published work. However, compared to archetypoid 3, the quality adjusted indicators are lower as well as the citations and downloads. Still, they are larger than the figures of archetypoid 1 and 2.

Figure 2: Identified archetypoids
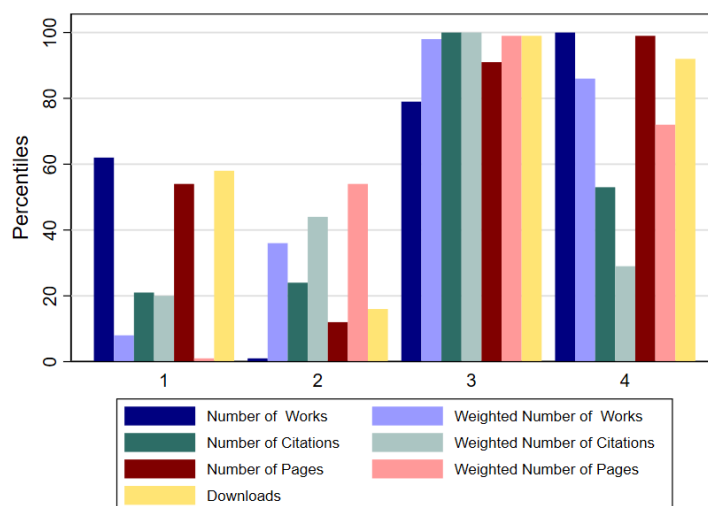
Nevertheless, as we extracted the best economic authors from RePEc, the terminus low performer for the first archetypoid must be interpreted cautiously. Compared to all authors listed in RePEc every author in our sample can be classified as top, nevertheless, compared among each other some perform worse than other ones.
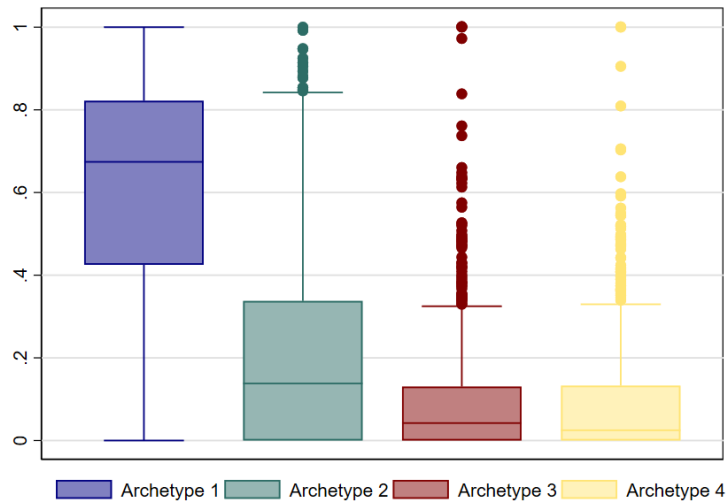
Table 2: Percentile values for each archetypoid

	Number of Works	Weighted Number of Works	Number of Citations	Weighted Number of Citations	Number of Pages	Weighted Number of Pages	Downloads
1	146	1,622	1,957	16,972	1,493	11,116	891
2	75	2,791	2,094	30,090	936	30,810	479
3	193	12,374	46,645	568,616	2,664	121,392	4,331
4	1,102	6,393	3,518	21,153	4,973	43,422	2,103

Besides the aggregated analysis it is also possible to look at the relative share of each archetypoid for each author. In practice, percentages of all four archetypoids are allocated to the authors. This implies, that each author is assigned four values, which add up to one. In Figure 3 we show the box-plots for all percentage shares of the four identified archetypoids. It displays that most of the authors in our data set are characterized by Archetype 1, the low performer, as it represents the largest relative shares. In comparison, the Archetype 2, the solid

performer, is less frequent. Of particular interest is the third archetypoid, the top performer. As one would suspect, the majority of authors in our sample have a low share of this archetypoid and only some – the true top performers – are assigned a large share of this archetypoid. The share of Archetype 4, the diligent performer, is similarly low.

Figure 3: Barplots of archetypoid shares allocated to the economists

In Table 3 we report the top five authors for each archetypoid, i.e. authors with the largest relative share for the respective archetypoid. The values for each author of the sample can be found in Wohlrabe and Gralka (2020b). A value of 1.000 denotes that the respective author represents the identified archetypoid perfectly. By assumption, this is the case for at least one author for all four archetypoids. Andrei Shleifer, James J. Heckman and Robert J. Barro represent the Archetypoid 3, the top performer. In contrast, Peter Nijkamp and Michael McAleer represent the forth one, the diligent performer. Five authors represent the low performer, which is plausible since the type is most common among the three archetypoids. Nevertheless, in most cases, authors are a mixture of the four different archetypoids.

Table 3: Economists top five for each archetypoid

Archetypoid	1	2	3	4
Francis Bloch	**1.000**	0.000	0.000	0.000
Eleonora Patacchini	**1.000**	0.000	0.000	0.000
William F. Maloney	**1.000**	0.000	0.000	0.000
Laurent Gobillon	**1.000**	0.000	0.000	0.000
Xin Meng	**1.000**	0.000	0.000	0.000
Thomas R. Palfrey	0.000	**1.000**	0.000	0.000
Debraj Ray	0.000	**0.992**	0.008	0.000
Larry G. Epstein	0.000	**0.948**	0.052	0.000
Martin L. Weitzman	0.000	**0.947**	0.053	0.000
Bruce D. Smith	0.000	**0.925**	0.000	0.075
Andrei Shleifer	0.000	0.000	**1.000**	0.000
James J. Heckman	0.000	0.000	**1.000**	0.000
Robert J. Barro	0.000	0.000	**1.000**	0.000
Daron Acemoglu	0.000	0.000	**0.973**	0.028
Jean Tirole	0.000	0.000	**0.839**	0.162
Peter Nijkamp	0.000	0.000	0.000	**1.000**
Michael McAleer	0.000	0.000	0.000	**1.000**
Hans-Werner Sinn	0.000	0.095	0.000	**0.905**
Mohsen Bahmani-Oskooee	0.191	0.000	0.000	**0.809**
Bruno S. Frey	0.182	0.000	0.113	**0.706**

4. Conclusions

In this paper we applied archetypoid analysis to the evaluation of top economists. The approach allows to extract typical characteristics (archetypoids) within a multivariate data set. We evaluate 776 scientists obtained from the RePEc database. We obtained seven bibliometric scores for each author, spanning over various measures of (quality-weighted) number of published work, citations and access statistics. We identified four main archetypoids, which are characterized as low, solid, top and diligent performer. Our focus was on research related indicators and we left out aspects as teaching or acquisition of third-party funding. We leave this for future research.

References

Bloom, N., Jones, C. I., Van Reenen, J., & Webb, M. (2020). Are ideas getting harder to find? *American Economic Review*, *110*(4), 1104–44.

Bornmann, L., Butz, A., & Wohlrabe, K. (2018). What are the top five journals in economics? A new meta-ranking. *Applied Economics*, *50*(6), 659–675.

Chan, B. H., Mitchell, D. A., & Cram, L. E. (2003). Archetypal analysis of galaxy spectra. *Monthly Notices of the Royal Astronomical Society*, *338*(3), 790–795.

García-Suaza, A., Otero, J., & Winkelmann, R. (2020). Predicting early career productivity of PhD economists: Does advisor-match matter? *Scientometrics*, *122*(1), 429–449.

Meyer, J., & Wohlrabe, K. (2018). Standing on the shoulder of giants: the aspect of free-riding in RePEc rankings. *Applied Economics Letters*, *25*(4), 223–228.

Moliner, J., & Epifanio, I. (2018). Bivariate functional archetypoid analysis: an application to financial time series. In *Mathematical and Statistical Methods for Actuarial Sciences and Finance* (pp. 473–476). Springer.

Mørup, M., & Hansen, L. K. (2012). Archetypal analysis for machine learning and data mining. *Neurocomputing*, *80*, 54–63.

Rath, K., & Wohlrabe, K. (2016). Recent trends in co-authorship in economics: evidence from RePEc. *Applied Economics Letters*, *23*(12), 897–902.

Seiler, C., & Wohlrabe, K. (2013). Archetypal scientists. *Journal of Informetrics*, *7*(2), 345–356.

Seth, S., & Eugster, M. J. (2016). Probabilistic archetypal analysis. *Machine learning*, *102*(1), 85–113.

Sommer, V., & Wohlrabe, K. (2017). Citations, journal ranking and multiple authorships reconsidered: evidence from almost one million articles. *Applied Economics Letters*, *24*(11), 809–814.

Vinué, G., & Epifanio, I. (2017). Archetypoid analysis for sports analytics. *Data Mining and Knowledge Discovery*, *31*(6), 1643–1677.

Vinué, Guillermo. (2017). Anthropometry: An R package for analysis of anthropometric data. *Journal of Statistical Software*, *77*(6), 1–39.

Vinué, Guillermo, Epifanio, I., & Alemany, S. (2015). Archetypoids: A new approach to define representative archetypal data. *Computational Statistics & Data Analysis*, *87*, 102–115.

Wohlrabe, K., & Gralka, S. (2020a). Using archetypoid analysis to classify institutions and faculties of economics. *Scientometrics*, *123*(1), 159–179.

Wohlrabe, K., & Gralka, S. (2020b). Classifying top economists using archetypoid analysis. *CEPIE Working Paper*, *01/20*, 1–32.

Zimmermann, C. (2013). Academic rankings with RePEc. *Econometrics*, *1*(3), 249–280.