Does Drinking Coffee Improve Research Output? Evidence from the Paris School of Economics

Thomas Blanchet*

February 24, 2018

Abstract

Although people routinely consume coffee to get them through their workday, little is actually known on how coffee consumption truly impacts productivity. I address this question for the first time by matching a unique individual-level administrative dataset of coffee consumption at the Paris School of Economics with public bibliometric data from RePEc. I find that coffee consumption is positively and significantly associated with better metrics of research output: a 10% increase in weekly coffee consumption corresponds to a 0.5 points increase in the H-index, and a 0.9 points increase in the i10-index.

*Paris School of Economics. Although this paper is written as a joke, all the data and regressions are genuine. Replication codes and anonymized data are available at `https://github.com/thomasblanchet/coffee-consumption-research`.

Introduction

Does coffee make us more productive? Despite the widespread consumption of coffee in the workplace, the answer to that question remains unclear. There are some anecdotal cross-country correlations with GDP per hours worked (Charpentier, 2014) and the number of researchers per inhabitants (Hossenfelder, 2013), but no studies that have been able to address that question at the individual level.

In this paper, I match an administrative dataset on weekly coffee consumption at the Paris School of Economics with bibliometric data from RePEC (Research Papers in Economics) and its companion service CitEc (Citations in Economics). Hence, I am able to study how coffee consumption is related to the academic output of researchers. I do find a positive and significant impact of drinking coffee on the main citation-weighted metrics of academic production: a 10% increase in weekly coffee consumption is associated to a 0.5 points increase in the H-index and a 0.9 points increase in the i10-index, a result that holds even after controlling for the researcher's time spent in activity.

1 Data and Methodology

Members of the Paris School of Economics can get coffee from two self-service machines. Access to the machines is open, but people are asked to contribute proportionally to their use of the machines. To that end, they must nominatively report their weekly coffee consumption on a public online spreadsheet.

The spreadsheet contains the name and coffee consumption of a total of 154 people. I matched that name to the "short-ID" of researchers that uniquely identifies them on RePEc and related services. The matching was done automatically on the researcher's names via a greedy fuzzy matching algorithm based on the Jaro–Winkler distance (Jaro, 1989; Winkler, 1990).[1] In the end, 54 people were successfully matched. The rest corresponds to people not registered on RePEc.

I retrieved the bibliometric information for the matched researchers from CitEc, a RePEc companion project, using web scraping. I focus on the two synthetic indexes of academic output publicly available on this service: the H-index and the i10-index. A H-index equal to k means that the researcher has k papers with at least k citations. The i10-index corresponds to the number of papers with at least 10 citations, and was introduced by the Google Scholar project.

Table 1 shows the characteristics of the sample. It skews toward more senior researchers who are more likely to be registered on RePEc. People in the sample have spent on

[1] Informally, the Jaro distance between two strings is the minimum number of single-character transpositions required to change one string into the other. The Jaro–Winkler distance uses a prefix scale p which gives more favorable ratings to strings that match from the beginning for a set prefix length ℓ.

Table 1: Descriptive statistics

	mean	standard deviation	min	max	obs
weekly coffee consumption	9.48	3.83	3	20	54
years in activity	16.50	9.66	1	43	54
H-index	9.00	8.39	1	38	54
i10-index	10.43	14.51	0	59	54

average 16 and a half years in academia (measured as the number of years since their first recorded paper). They have an average H-index of 9 and an average i10-index of 10.5. And they consume an average of 9.5 cups of coffee per week.

2 Results

I use two model specifications for each outcome variable: one with weekly coffee consumption as the sole regressor, the other with the time the researcher has been in activity as an additional explanatory variable. This second variable can control for age and seniority effects which may be correlated with both coffee consumption and academic output.

Table 2: Regression results

	(1) H-index	(2) H-index	(3) i10-index	(4) i10-index
log(weekly coffee consumption)	6.109**	5.261**	10.28**	9.005**
	(2.556)	(2.026)	(4.493)	(3.721)
log(years in activity)		4.717***		7.084***
		(0.936)		(1.831)
constant	-4.201	-14.36***	-11.79	-27.04***
	(5.058)	(4.953)	(8.869)	(9.293)
observations	54	54	54	54
R^2	0.105	0.346	0.099	0.281

Robust standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

As shown in table 2, weekly coffee consumption is always positively and significantly associated with better metrics. In the first specification, a 10% increase in weekly coffee consumption is associated to a 0.6 points increase ($p = 0.021$) of the H-index and a 1 point increase ($p = 0.026$) of the i10-index. The results hold when we add the years spent in activity to the models: the impact of coffee consumption is only slightly weaker, though not significantly so. A 10% increase in weekly coffee consumption is now associated with a 0.5 points increase ($p = 0.012$) in the H-index and a 0.9 points increase ($p = 0.019$) in the i10-index.

References

Charpentier, Arthur (2014). "Coffee and Productivity", Freakonometrics, URL: http://freakonometrics.hypotheses.org/16010

Jaro, M. A. (1989). "Advances in record linkage methodology as applied to the 1985 census of Tampa, Florida". Journal of the American Statistical Association. 84 (406): 414–20. doi:10.1080/01621459.1989.10478785.

Hossenfelder, Sabine (2013). "Researchers and coffee consumption", Backreaction, URL: http://backreaction.blogspot.fr/2013/08/researchers-and-coffee-consumption.html

Winkler, W. E. (1990). "String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage". Proceedings of the Section on Survey Research Methods. American Statistical Association: 354–359.

Appendix
Alternative Model Specifications

Table 3: Alternative specification: linear-linear

	(1) H-index	(2) H-index	(3) i10-index	(4) i10-index
weekly coffee consumption	0.803**	0.668**	1.367**	1.164**
	(0.328)	(0.287)	(0.594)	(0.538)
log(years in activity)		4.586***		6.847***
		(0.920)		(1.802)
constant	1.383	-8.988***	-2.533	-18.02***
	(2.757)	(3.296)	(5.002)	(6.280)
observations	54	54	54	54
R^2	0.134	0.360	0.130	0.298

Robust standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 4: Alternative specification: log-log

	(1) log(H-index)	(2) log(H-index)	(3) log(i10-index)	(4) log(i10-index)
log(weekly coffee consumption)	0.525*	0.391**	0.845**	0.761***
	(0.281)	(0.177)	(0.363)	(0.243)
log(years in activity)		0.746***		0.904***
		(0.105)		(0.117)
constant	0.675	-0.932**	-0.0269	-2.221***
	(0.592)	(0.430)	(0.783)	(0.586)
observations	54	54	48	48
R^2	0.063	0.556	0.103	0.532

Robust standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$