Exploring the Effect of Software Ecosystem Health on the Financial Performance of the Open Source Companies

Joost Dijkers
Utrecht University
j.j.g.dijkers@students.uu.nl

Rowan Sincic
Utrecht University
r.s.i.sincic@students.uu.nl

Nicole Wasankhasit
Utrecht University
n.n.wasankhasit@students.uu.nl

Slinger Jansen
Utrecht University
slinger.jansen@uu.nl

ABSTRACT

Background. It is currently unknown how software ecosystem health affects the financial performance of open source companies. This is a problem, because open source is becoming increasingly popular and more knowledge is necessary for companies on how to capitalize on this phenomenon. *Objectives.* With this paper, insight is developed into the relation between software ecosystem health and financial performance of the open source company that nurtures it. *Method.* A case study on two open source companies, Cloudera and Hortonworks, is performed. The software ecosystem health and financial performance of both companies are assessed. *Results.* Cloudera is healthier in terms of robustness and niche creation, while Hortonworks is healthier in terms of productivity. Financially, Cloudera performs better than Hortonworks. *Conclusion.* The following hypotheses are formulated. Software ecosystem health has an expected positive relation with financial performance. Niche creation health is the main contributor to this relation, robustness health is a minor reason for this relation, and software ecosystem productivity health has little to no influence.

KEYWORDS

software ecosystem health, OSEHO, open source software, financial performance

ACM Reference Format:
Joost Dijkers, Rowan Sincic, Nicole Wasankhasit, and Slinger Jansen. 2018. Exploring the Effect of Software Ecosystem Health on the Financial Performance of the Open Source Companies. In *SoHeal'18: SoHeal'18:IEEE/ACM 1st International Workshop on Software Health , May 27-June 3, 2018, Gothenburg, Sweden.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3194124.3194126

1 INTRODUCTION

Open source software is described as software for which the source code is publicly available, typically having one person or body in charge of choosing one subset of all source code as the main release and distributing it [1]. One distinguishing characteristic of open source software is that it is not necessarily developed by paid developers, but also by voluntary developers and maybe even consumers [2]. Furthermore, open source software can be based on or use other open source software. An open source software project is thus part of a software ecosystem (SECO), which is defined as: "a set of actors functioning as a unit and interacting with a shared market for software and services, together with the relationships among them. These relationships are frequently underpinned by a common technological platform or market and operate through the exchange of information, resources and artifacts" (p. 18) [3].

The performance of a SECO is measured by the ecosystem health [4], which is also defined as the capability for growth and the sustainability thereof [5]. A more healthy ecosystem performs better and has more propensity for growth. Consequently, an organization that sells open source software (referred to as an open source company) supposedly will be more successful if its SECO is healthy. This success is characterized by an increase in revenue, an improved market position, or an increase in development [6].

Multiple attempts have been made to operationalize SECO health [5, 7, 8], but validation of these methods and validation of the importance of SECO health in general is still lacking [9, 10]. Furthermore, what little validation work has been done provides no conclusive answers on the relation between SECO health and financial performance. For example, in one study the SECO health of content management systems plays no role in the use thereof [11], and in a different case study a company that scored poorly in terms of SECO health reported a large total revenue indicating market interest [12]. But in another case study the health of the Ruby ecosystem was assessed and used to formulate managerial strategies [13]. More research is thus necessary on the relation between SECO health and financial performance, so that more conclusive insights can be made from SECO health assessments. The following research question is formulated:

Research question: *What is the relation between software ecosystem health and the financial performance of the open source company that nurtures it?*

This research aims to find an answer to the research question by conducting a case study on two open source companies. Open source companies were chosen because gathering data from multiple companies for the same metrics has proven to be difficult in the past [14]. A SECO health assessment and a financial performance assessment is done for both companies, which are subsequently

SoHeal'18, May 27-June 3, 2018, Gothenburg, Sweden

ACM ISBN 978-1-4503-5730-2/18/05…$15.00
https://doi.org/10.1145/3194124.3194126

analyzed to find relations between SECO health and financial performance.

This paper is structured as follows: in section 2 related work is studied on how SECO health is measured. One of the identified frameworks is the OSEHO framework [5], which is used in this paper to measure SECO health. Then, in section 3 the research methodology is described. In section 4 the results of the research are shown and analyzed. The analysis indicates that there is a positive relation between SECO health in financial performance. In section 5 the benefits and the limitations of this research are discussed. Finally, in section 6 the main findings of this paper are summarized and an answer is given to the research question. Given the limitations, the positive relation is formulated as a hypothesis. Furthermore, suggestions for future research are given in this section, such as researching a way to quantify a SECO health assessment.

2 LITERATURE STUDY

In this section three frameworks for measuring SECO health are discussed. The conceptual "Open Source Ecosystem Health Operationalization" (OSEHO) framework categorizes SECO health metrics in three categories: productivity, robustness, and niche creation [5]. Iansiti and Levien [15] define these categories as follows: productivity is the ability to transform inputs to outputs, robustness refers to the stability of the ecosystem in face of change and innovation, and niche creation is about the meaningful diversity within an ecosystem.

Another framework to evaluate the health of a SECO is the HEAL-ME framework [7]. In addition to productivity, robustness, and niche creation, sustainability and diversity are used as categories. It is important to mention that the OSEHO framework addresses diversity through niche creation and sustainability through all categories [5]. The added value of the HEAL-ME framework lies in the automation of gathering data.

A third framework to measure SECO is currently being developed [8], which aims to apply a socio-technical approach to measuring SECO health. One aspect that differs from the previous health measurements [5, 7] is the social perspective, taking into account the relations between the members in a SECO.

Because the third framework is still being developed, it is not used in this paper. Additionally, the HEAL-ME framework is neither as complete or as tested as the OSEHO framework, resulting in a preference for the OSEHO framework. Finally, the OSEHO framework is specifically made for open source SECOs, making it a good fit for this research.

3 METHOD

3.1 Software ecosystem health metrics

The metrics used to measure the SECO health are based on the OSEHO framework [5] and can be found in Table 1. Multiple metrics for each category (productivity, robustness, and niche creation) exist and inclusion of metrics depends on the goals of the health assessment [5]. Therefore, for each metric that is measured, the reason for why it is an interesting metric is discussed below.

Productivity is measured by three metrics [5]. New projects (Met01) is defined as the evolution of the number of new projects belonging to an ecosystem over time. New projects measures the productivity of an ecosystem because it indicates the rate of growth over time. Commits for projects (Met02) is the number of times a project has been changed by a developer. Commits measure productivity because the metric indicates which projects are often changed. Number of forks (Met03) is a metric that measures the number of times repositories are forked. A fork is a copy of a repository, typically used to experiment with the repository or to save the repository as a backup. Number of forks measures productivity because it indicates developer, in the sense that developers are experimenting with a repository or that developers are interested in the continued existence of the repository.

Table 1: Software ecosystem health metrics

Metric (label) [health category]	Source	Date gathered
New projects (Met01) [productivity]	GitHub	4th of January 2018
Commits for projects (Met02) [productivity]	GitHub	4th of January 2018
Number of forks (Met03) [productivity]	GithHub	28th of December 2017
Number of active projects (Met04) [robustness]	GitHub	28th of December 2017
Interest (Met05) [robustness]	Google Trends	4th of January 2018
Potential for niche creation (Met06) [niche creation]	Company website	2nd of January 2018
Actual niche creation (Met07) [niche creation]	GitHub	28th of December 2017

Robustness is measured by two metrics [5]. Number of active projects (Met04) defines how many projects are still being updated, and measures the robustness of an ecosystem because it shows what projects are still being updated while the ecosystem undergoes change. The concept of active projects is arbitrarily defined as three months. Interest (Met05) is defined as a metric for robustness as measured by the public interest in the ecosystem [16]. Google Trends was reported to be a good way to measure interest [11], and is thus used in this paper.

Niche creation is measured by two metrics [5]. Potential for niche creation (Met06) and actual niche creation (Met07) are derived from the metric variety in projects. Potential for niche creation measures in which possible areas the SECO can support niche creation, whereas actual niche creation shows in which areas the niche creation is actually occurring. Both measure niche creation, because if there is a plentiful amount of either potential or actual niche creation, then there is opportunity for a niche player to survive in the ecosystem. This metric is measured by looking at the thirty most valued repositories on GitHub. Repositories are mapped to the categories of the ecosystem as defined on the company websites.

3.2 Financial performance

There are a myriad of ways to measure financial performance [17]. The measures that were used in this paper can be found in Table 2, and are explained below.

Table 2: Financial performance measures

Financial performance (label)	Source	Date gathered
Share price (Fin01)	Amigobulls	28th of December 2017
Return on equity (Fin02)	Balance sheet and income statement	28th of December 2017
Earnings per share (Fin03)	Income statement	28th of December 2017
Total revenue (Fin04)	Income statement	28th of December 2017
Altman's Z"-score (Fin05)	Balance sheet and income statement	28th of December 2017

Share price (Fin01) was chosen because maximizing the value of equity is the primary goal of a business [17]. Following the same reasoning, return on equity (Fin02) and earnings per share (Fin03) were chosen as a financial performance measures. Return on equity is calculated as the net income divided by the total equity, in other words it indicates how much profit a company has generated using the money invested by shareholders. Earnings per share is calculated as the net income divided by the number of shares outstanding, and thus measures the profit made per share.

Total revenue (Fin04) was used as a financial performance measure because it indicates the size of the company, because it was a factor in choosing the subjects for the case study, and because literature on SECO health suggests that a healthy SECO increases total revenue [6].

The Altman's Z"-score for emerging markets (Fin05) indicates the possibility that a company goes bankrupt within the next two years [18]. It was chosen because it is calculated from financial ratios indicating the profitability, liquidity, and solvency of a company. Thus, the Altman's Z"-score is semantically rich when it comes to measuring financial performance. The revised Z"-score is used, for which the formula is Z" = 6.56 (A) + 3.26 (B) + 6.72 (C) + 1.05 (D) + 3.25, where A = working capital / total assets, B = retained earnings / total assets, C = earnings before interest and tax / total assets, and D = market value of equity / total liabilities. A Z"-score of below 1.23 indicates that a company is at risk of going bankrupt, a Z"-score of above 2.9 indicates that a company is safe, and a Z"-score between 1.23 and 2.9 represents a gray area for which no prediction can be made.

3.3 Case study

In order to find an answer to the research question, observational case studies were performed and compared. Wieringa defines an observational case study as: "a study of a real-world case without performing an intervention" (p. 225) [19]. An observational case study allowed a real world application of a SECO health framework and allowed the researchers to relate the results to actual financial performance. Selection criteria for the subjects of the observational case studies were as follows. The product of the company must be open source, the company must have gone public so that financial information is readily available, and the SECO surrounding the open source product must be active. The latter was determined by doing an initial check for the metrics that were measured. Two observational case studies were performed: one on Cloudera and one on Hortonworks. Both these companies satisfied the selection criteria. Furthermore, both companies operate in the same market, which eliminates the effect of the market as an independent variable on both SECO health and financial performance. Cloudera[1] and Hortonworks[2] offer products and solutions building on Apache Hadoop, a big data framework. Apache Hadoop, Cloudera, and Hortonworks are published as open source software with the Apache License 2.0. Additionally, Cloudera and Hortonworks were chosen as companies because they have reported a similar revenue: Cloudera has reported a revenue of 166.04 million dollars[3] and Hortonworks has reported a revenue of 184.46 million dollars[4]. Finally, previous research on the Hadoop ecosystem indicates that Cloudera and Hortonwork are the keystone players [20].

3.4 Data gathering and analysis

The dates on which data was gathered for the SECO health metrics can be found in Table 1 and the dates on which data was gathered for the financial performance metrics can be found in Table 2. Data from GitHub was gathered from the company pages on GitHub[5], using a combination of the scraping tool Import.io[6] and the GHTorrent project [21]. GHTorrent collects repository related data from GitHub through the GitHub API. All other data has been gathered manually from the given sources. For the analysis of data, developments in SECO health and financial performance were judged on desirability. Desirability is defined as improvement. That is, if the measure for a certain metric has improved over time, then the health for that metric is changing in a desirable way. The following categories are used: '+' to indicate a positive desirability, '-' to indicate a negative desirability, and '0' to indicate stagnation.

4 RESULTS AND ANALYSIS

4.1 Productivity

New projects (Met01). The total number of projects have been calculated over a time period from the 5th of June in 2009 to the 7th of December in 2017. This range fits the earliest and latest found values. The first project for Cloudera was shared on the 5th of June in 2009 and Cloudera has on average been sharing 13.3 new projects per year from the start of 2010 until the end of 2016. The first project that Hortonworks was shared on the 13th of February in 2012. On average, 33.6 new projects per year were shared by Hortonworks from the start of 2013 until the end of 2016. Cloudera

[1] https://www.cloudera.com/
[2] https://hortonworks.com/
[3] https://amigobulls.com/stocks/CLDR/income-statement/annual?f=pg
[4] https://amigobulls.com/stocks/HDP/income-statement/annual?f=pg
[5] https://github.com/cloudera and https://github.com/Hortonworks
[6] https://www.import.io/

had more existing projects than Hortonworks until the 9th of march in 2016, when Hortonworks surpassed Cloudera. The number of new projects for Cloudera is still increasing, but this growth is stagnating. For Hortonworks on the other hand the number of new projects is increasing relatively fast. As such, the desirability of the development of this metric is '0' for Cloudera and '+' for Hortonworks.

Commits for projects (Met02). The four Cloudera projects with the most commits account for 75.1% of all commits, meaning the other 105 repositories account for 24.4% of all commits. In total, the Cloudera projects have 40134 commits. The top four Hortonworks projects with the most commits account for 48.6% of all commits, whereas the other 149 projects account for 51.4% of all commits. In total, the Hortonworks projects have 54070 commits. An overview of this data can be found in Table 3. Cloudera's highest number of commits are mainly centered around four repositories. This implies that developers are highly interested in just a few repositories. For Hortonworks on the other hand, the commits are more spread out over repositories, which indicates that developer interest is allocated relatively evenly. As such, the desirability of the development of this metric is '0' for Cloudera and '+' for Hortonworks.

Number of forks (Met03). The highest percent of repositories (35%) from Cloudera have between ten to forty-nine forks, followed by one to nine forks (28% of the repositories). Most of the repositories from Hortonworks have between one and nine forks (38% of the repositories), followed by zero forks (35% of the repositories). Comparatively, 1% of the Cloudera repositories have zero forks. An overview of this data can be found in Table 3. Developer experimentation occurs in almost all repositories in varying degrees for Cloudera. However, for Hortonworks there are many neglected repositories and many others with just a few forks. As such, the desirability of the development of this metric is '+' for Cloudera and '-' for Hortonworks.

4.2 Robustness

Number of active projects (Met04). As mentioned in the method, active projects are repositories that have been updated in the past three months. Cloudera has 51 active projects which corresponds to 46.8% of all their projects being active projects. Hortonworks has 89 active projects which means that 58.6% of all their projects are active projects. The higher number of Hortonworks can be partially explained due to the higher rate at which new projects are created (Met01), so the development is considered the same for both Cloudera and Hortonworks. These are both healthy developments, so desirability for both Cloudera and Hortonworks is '+'.

Interest over time (Met05). Interest was calculated in two different ways: independent of each other and dependent on each other's search frequencies (to show relative interest). The scores represent search interest for the search term, related to the highest number in the diagram for a specific region and time. The score hundred represents the highest interest for the search term, fifty represents half of the interest of the highest, and zero represents less than 1% interest compared to its peak.

Cloudera has quite a stable interest except for a peak on the 23rd of April, which is likely related to Cloudera going public on the 28th of April. The mean of all scores of their year 2017 is sixty, showing some interest above half of the highest score. Hortonworks has a high interest throughout the year, with a mean of the score seventy-three. Hortonworks has a sudden peak on 11th of June, however a reason for this could not be found. In addition, the interest in both companies dependent on each other's search frequency has been analyzed. The new mean for Hortonworks' score is 24.5, while Cloudera's new values are almost indistinguishable from the old values. In other words, interest in Cloudera is relatively higher than in Hortonworks. Interest for both companies is stagnating, but for Cloudera at a higher level than for Hortonworks. As such, the desirability of the development for this metric is '0' for both Cloudera and Hortonworks, where Cloudera is doing better than Hortonworks.

4.3 Niche creation

Potential for niche creation (Met06). Cloudera and Hortonworks have visualized their own SECO from which different niches can be inferred [22, 23]. Cloudera has categorized their SECO in four ways: process, analyze, and serve; unified services; store; and integrate. These four categories in total have thirteen subcategories. Hortonworks has categorized their SECO in six ways: governance integration, tools, data access, data management, security, and operations. These six categories in total have sixteen subcategories. A more in depth comparison indicates that this difference in number of categories and subcategories mainly comes from a different separation of concerns. For example, security is a separate category for Hortonworks, but it is part of of the unified services category from Cloudera. As both Cloudera and Hortonworks are mainly based on Apache Hadoop, but are working with more than the main modules, they both have a high potential for variety in projects. Right now there are 205 available Apache projects [8], meaning that the companies in the future can integrate even more projects into their platforms. Additionally, the companies have their GitHub pages where they interact with other projects outside their main Apache Hadoop ecosystem. The desirability of the development of this metric is thus '+' for both Cloudera and Hortonworks.

Actual niche creation (Met07). The thirty most starred categories on the GitHub company pages were analyzed and put in a niche category from potential for niche creation (Met06). For Cloudera there is a relatively big interest in the process, analyze, and serve category. Moreover, repositories for each category are shared and only three subcategories did not have a repository. For Hortonworks there is a relatively big interest in the tools and operations categories. However, eleven subcategories did not have a repository. Because Cloudera has repositories intended for most of the ecosystem categories, the development of this metric is categorized as '+'. For Hortonworks however, the development of this metric is categorized as '-' because many subcategories are not represented.

4.4 Financial performance

Share price (Fin01). Cloudera went public on the 28th of April in 2017 with an opening share price of $17.8. The opening share price steadily climbed to near $23 until the beginning of June in the

[8] https://projects.apache.org/projects.html

Table 3: Measures for Commits for projects (Met02) and Number of forks (Met03)

Metric	Measures	Cloudera	Hortonworks
Commits for projects (Met02)	Top four projects	75.1%	48.6%
	Total number	40134	54070
Number of forks (Met03)	Repositories with zero forks	1%	35%
	Repositories with one to nine forks	28%	38%
	Repositories with ten to forty-nine forks	35%	18%
	Repositories with fifty to ninety-nine forks	10%	3%
	Repositories with 100 to 499 forks	19%	4%
	Repositories with 500 to 1000 forks	4%	2%
	Repositories with more than 1000 forks	3%	1%

same year, after which it started to decline. Since then the opening share price has been fluctuating around $16-$17. Hortonworks went public on the 23th of December in 2014 with an opening share price of $24. The opening share price fluctuated heavily throughout 2015 between $20-$28, and plummeted down to $10 in early 2016 due to the news of a second initial public offering[10]. Since late 2016 the opening share price has been steadily climbing from around $7 to $20.48 on the 27th of December in 2017. Because the share price of Cloudera shows neither an increase of decrease over time, the development of this metric is '0'. For Hortonworks, the steady increase indicates that the development of this metric can be categorized as '+'.

Return on equity (Fin02). The return on equity for both Cloudera and Hortonworks can be found in Table 4. Missing values mean that data is not available for that quarter. For both companies, whenever the return on equity is negative it is due to a reported net loss, whereas if the return on equity is positive it is due to a reported net loss and equity deficit. Such a positive return on equity is the worst situation a company can be in. A negative return on equity is also undesirable. For Hortonworks, the stakeholder's equity has been declining since 2016 (eventually turning negative), but the net loss has slowly been decreasing. Cloudera reported an unusually large net loss in the first quarter of 2017, so the return on equity for the second and third quarter of 2017 are more indicative of the financial performance, which indicate that the negative return on equity is improving. The development for this metric is thus '-' for both Cloudera and Hortonworks.

Earnings per share (Fin03). The earnings per share for both Cloudera and Hortonworks can be found in Table 4. Both Cloudera and Hortonworks continually have negative earnings per share due to both companies consistently reporting a negative net income. A negative earnings per share is bad, because it may decrease the value of the firm and thus the share price. However, when looking at the development of the earnings per share over time, a trend towards a positive value indicates that the company is on its way to become profitable. This is the case for Hortonworks, which has experienced a steady decrease in the net loss per share since 2016. Cloudera has reported a large net loss per share in the first quarter of 2017, likely due to a big investment in that quarter, but reported a relatively low net loss per share in the second and third quarter of 2017. The development for this metric can thus be categorized as '-' but improving for both Cloudera and Hortonworks.

Total revenue (Fin04). The total revenue for both Cloudera and Hortonworks can be found in Table 4. Both Cloudera and Hortonworks have reported an increasing total revenue for all quarters for which data is available. Cloudera experiences strong fluctuations in the increase of total revenue, for example 12.1% revenue growth in the second quarter of 2017 but then 5.98% revenue growth in the third quarter of 2017. Hortonworks experiences a more steady increase in revenue growth, for example 10.47% in the second quarter of 2017 and 11.6% in the third quarter of 2017. The development of this metric can thus be categorized as '+' for both Cloudera and Hortonworks.

Altman's Z"-score (Fin05). The Altman's Z"-score for both Cloudera and Hortonwork for the first three quarters in 2017 can be found in Table 4. The relatively low Altman's Z"-score of Cloudera in the first quarter can be explained by several factors. But first the formula for the Altman's Z"-score is repeated: Z" = 6.56 (A) + 3.26 (B) + 6.72 (C) + 1.05 (D) + 3.25, where A = working capital / total assets, B = retained earnings / total assets, C = earnings before interest and tax / total assets, and D = market value of equity / total liabilities. As explained when showing the results of the return on equity (Fin02), Cloudera reported an unusually large net loss in the first quarter of 2017, which affects B through retained earnings and C through earnings before interest and taxes (EBIT). Additionally, in the second quarter the current assets increased drastically, which affects A through working capital and B and C through total assets. Finally, the shares outstanding increased drastically in the second quarter, which affects D through the market value of equity. As such, the Altman's Z"-score for the second and third quarter give a more accurate indication of the risk of bankruptcy for Cloudera. The Altman's Z"-score for Cloudera in those two quarters is unusually large, showing that Cloudera is not at risk for bankruptcy. For Hortonworks the Altman's Z"-score indicate no risk of bankruptcy in the first and third quarter, while it is just barely in the gray area in the second quarter. The development of this metric is therefore categorized as '+' for both Cloudera and Hortonworks.

4.5 Summary of SECO health and financial performance assessments

The development in the SECO health and financial performance metrics is summarized in Table 5. Additionally, the companies are

[10]https://siliconangle.com/blog/2016/01/20/hapless-hortonworks-shares-plunge-22-on-news-of-secondary-ipo/

Table 4: Financial metrics results

Quarter	Return on equity		Earnings per share		Total revenue (millions $)		Altman's Z"-score	
	Cloudera	Hortonworks	Cloudera	Hortonworks	Cloudera	Hortonworks	Cloudera	Hortonworks
Q1 2016		-0.531	-1.2	-1.151	56.49	41.34		
Q2 2016		-0.777	-1.068	-1.112	64.46	43.64		
Q3 2016		-1.445	-1.203	-1.095	67.26	47.52		
Q4 2016		-5.022		-0.939		51.96		
Q1 2017	0.433	3.545	-5.776	-0.868	79.6	55.97	1.6	3
Q2 2017	-0.181	1.296	-0.478	-0.892	89.83	61.83	13.51	2.89
Q3 2017	-0.168	0.888	-0.4	-0.676	94.57	69	12.09	6.71

compared on SECO health. When comparing Cloudera and Hortonworks per metric, it is at first sight inconclusive which company is more healthy than the other. With regards to new projects (Met01) and commits for projects (Met02), Hortonworks performs better. On the other hand, Cloudera performs better on number of forks (Met03), interest (Met05), and actual niche creation (Met07). On the number of active projects (Met04) and the potential for niche creation (Met06) both companies perform similarly. However, when looking at the bigger picture it is noticeable that when Hortonworks is performing better, Cloudera still generally performs well, whereas when Cloudera is performing better, Hortonworks is performing poorly. As such, Cloudera has a more healthy ecosystem than Hortonworks.

The companies are also compared on financial performance. As can be inferred from this summary and also the previous discussion, Cloudera is performing better in terms of return on equity (Fin02). In addition, Cloudera is considered performing better in terms of earnings per share (Fin03) and Altman's Z"-score (Fin05) due to the better values. Furthermore, Cloudera and Hortonworks perform similarly well in terms of total revenue. Finally, Hortonworks performs with regards to share price. In conclusion, Cloudera overall performs better financially than Hortonworks.

4.6 Relation between SECO health and financial performance

The assessments of SECO health and financial performance (summarized in Table 5) reveal some patterns that indicate relations between SECO health and financial performance.

When looking at the companies individually, a general relation between SECO health and financial performance is not immediately apparent. Cloudera has a healthy SECO in the sense that all metrics except for commits for projects (Met02) and interest (Met05) are developing in a desirable way. Likewise, financial performance of Cloudera is developing in a desirable way (except for share price), even if Cloudera is not performing well on all financial performance metrics. For Hortonworks however, the ecosystem is healthy for some metrics and unhealthy for others. The financial performance of Hortonworks does not follow a similar pattern, it is largely developing in a desirable way except for the return on equity.

Table 5: Assessment of SECO health metrics and financial performance

Metric	Cloudera	Hortonworks	The better company
New projects (Met01)	+	+	Hortonworks
Commits for projects (Met02)	0	+	Hortonworks
Number of forks (Met03)	+	-	Cloudera
Number of active projects (Met04)	+	+	Same
Interest (Met05)	0	0	Cloudera
Potential for niche creation (Met06)	+	+	Same
Actual niche creation (Met07)	+	-	Cloudera
Share price (Fin01)	0	+	Hortonworks
Return on equity (Fin02)	- but improving	-	Cloudera
Earnings per share (Fin03)	- but improving	- but improving	Cloudera
Total revenue (Fin04)	+	+	Same
Altman's Z"-score (Fin05)	+	+	Cloudera

When the companies are compared to one another, a relation between SECO health and financial performance becomes more apparent. As concluded in section 4.6, Cloudera is performing better financially than Hortonworks. By comparing the SECO health of Cloudera and Hortonworks, relative importance of the SECO health categories (productivity, robustness, and niche creation) can be

inferred. The comparison between Cloudera and Hortonworks per SECO health metric was discussed in section 4.6.

Productivity is measured by the new projects (Met01), commits for projects (Met02), and the number of forks (Met03). Hortonworks is more healthy in terms of new projects (Met01) and commits for projects (Met02), whereas Cloudera is more healthy in terms of number of forks (Met03). Thus, overall Hortonworks performs better on productivity.

Robustness is measured by the number of active projects (Met04) and by interest (Met05). Cloudera and Hortonworks are similarly healthy in terms of the number of active projects (Met04). The interest (Met05) in Cloudera and Hortonworks is stagnating for both companies, however at a relatively higher level for Cloudera. Cloudera thus overall performs slightly better on robustness.

Niche creation is measured by the potential for niche creation (Met06) and the actual niche creation (Met07). Cloudera and Hortonworks have a similar potential for niche creation (Met06). Cloudera is healthy in terms of actual niche creation (Met07), whereas Hortonworks is not. Cloudera thus performs better on niche creation.

In this research it is assumed that the difference in financial performance is due to the difference in SECO health. Thus, because Cloudera is not healthier in terms of productivity, there does not seem to be a relation between productivity and financial performance. Furthermore, because Cloudera is slightly healthier in terms of robustness, there seems to be a small, positive relation between robustness and niche creation. Finally, because Cloudera is healthier in terms of niche creation, there seems to be a positive relation between niche creation and financial performance.

5 DISCUSSION

5.1 Contributions to research

According to a proposal for operationalizing SECO health, a healthy SECO results in an increase in revenue, an improved market position, or an increase in development [6]. Since that proposal, SECO health has been operationalized into several frameworks [5, 7, 8]. But there is not enough research validating these frameworks and SECO health in general [9, 10]. Furthermore, there is no research focused on the relation between SECO health and financial performance.

This paper tried to fill this gap by investigating the relation between SECO and financial health for open source companies. SECO health was measured in three ways: productivity, robustness, and niche creation. Results indicate that SECO health in terms of productivity has no relation with financial performance. Instead, the most clear relation between SECO health and financial performance was found with SECO health in terms of niche creation. A small relation was found between SECO health in terms of robustness and financial performance. The lack of relation between SECO health in terms of productivity and financial performance is especially interesting considering that productivity is supposedly the most important indicator for the health of an ecosystem [15]. Furthermore the clear and positive relation between SECO health in terms of niche creation and financial performance provides validation to the often stated, positive effect of niche creation on the success of a SECO [10]. Finally, these findings contradict previously made observations, i.e. that SECO health has no impact in the use of the underlying platform [11] and that SECO health is not apparently related to total revenue [12].

Practitioners can use this knowledge to manage their SECO in such a way that it supports and motivates innovation in meaningful niches. Furthermore, practitioners should take precautions to ensure their SECO is robust to change. Strategic approaches to do so have already been formulated in other research [4]. Ensuring robustness can be done by building communities to connect niche players and by providing these niche players marketing and sales support. Niche creation can be stimulated by attracting and supporting niche players, which can be done in two ways. First, by applying the aforementioned robustness improvement mechanisms. This connection between robustness and niche creation is also echoed in other research [10]. Second, by building a strong reputation and lowering entry barriers.

5.2 Limitations

There are a few limitations concerning this paper. Most importantly, the relation between SECO health and financial performance was not tested with statistical tests. This was done intentionally, because some SECO health metrics such as actual niche creation (Met07) were difficult to quantify or difficult to transform into developments over time (which would be a prerequisite for correlation tests). Second, due to the lack of statistical tests a qualitative approach to data analysis was used. Objectivity was key and of paramount importance, but the potential exists that the descriptive nature of the analysis and results have led to subjective conclusions. Third, other indicators of financial performance, such as corporate social responsibility [24], were not considered. Fourth, due to little historical financial performance data (especially for Cloudera), drawing conclusions was more challenging. Finally, only two subjects were compared, which raises concerns on the generalizability of the conclusions to the population.

6 CONCLUSION

After an assessment was made on the software ecosystem health and financial performance of each company individually, the relation between software ecosystem health and financial performance was analyzed both by looking at the companies individually and by comparing the companies. When looking at the companies individually, a positive relation was found at one company but no clear relation was found at the other company. When comparing the companies, no clear relation was found for software ecosystem health in terms of productivity, while a minor positive relation was found for software ecosystem health in terms of robustness, and a positive relation was found for software ecosystem health in terms of niche creation. Given the small sample size and the absence of statistical tests, there is no hard evidence for these conclusions. As such, the main conclusions of this research are considered as hypotheses. That is, software ecosystem health has an expected positive relation with financial performance, software ecosystem niche creation health is the main contributor to this relation, software ecosystem robustness health is a minor reason for this relation, and software ecosystem productivity health has little to no influence on this relation.

Because the main conclusions of this research are hypotheses, more research is necessary to validate these hypotheses. Possible ways to validate these hypotheses are by studying a larger sample size, by including subjects for which a longer history of financial information is available, and by comparing subjects which are less similar in terms of financial performance (e.g. a subject with much market share and a subject with little market share). Validation of these hypotheses should ideally be done using the method from this research, so that analyses and conclusions can be compared. Furthermore, once more validation work has been done, more research is necessary into the precise aspects of the relationship. That is, whether specific software ecosystem metrics truly cause a change in financial performance, and if so, to what degree. Finally, research on how to quantify the assessment of a SECO's health is an important research direction, because one of the challenges encountered during this research was interpreting the results from the metrics.

REFERENCES

[1] Audris Mockus, Roy T. Fielding, and James D. Herbsleb. 2002. Two case studies of open source software development: Apache and Mozilla. ACM Transactions on Software Engineering and Methodology 11, 3 (2002), 309–346. DOI:http://dx.doi.org/10.1145/567793.567795

[2] Sonali K. Shah. 2006. Motivation, Governance & the Viability of Hybrid Forms in Open Source Software Development. Management Science 52, 7 (2006), 1000–1014. DOI:http://dx.doi.org/10.2139/ssrn.898247

[3] Slinger Jansen and Michael Cusumano. 2013. Defining software ecosystems: a survey of software platforms and business network governance. In Slinger Jansen, Michael Cusumano, & Sjaak Brinkkemper, eds. Software Ecosystems: Analyzing and Managing Business Networks in the Software Industry. Edward Elgar Publishers.

[4] Ivo Van Den Berk, Slinger Jansen, and Lútzen Luinenburg. 2010. Software ecosystems: a software ecosystem strategy assessment model. Proceedings of the Fourth European Conference on Software Architecture Companion Volume - ECSA 10 (2010), 127–134. DOI:http://dx.doi.org/10.1145/1842752.1842781

[5] Slinger Jansen. 2014. Measuring the health of open source software ecosystems: Beyond the scope of project health. Information and Software Technology 56, 11 (2014), 1508–1519. DOI:http://dx.doi.org/10.1016/j.infsof.2014.04.006

[6] Konstantinos Manikas and Klaus Marius Hansen. 2013. Reviewing the health of software ecosystems–a conceptual framework proposal. Proceedings of the 5th International Workshop on Software Ecosystems (IWSECO) (2013), 33–44. DOI:http://dx.doi.org/10.3403/30209330

[7] Iuri Carvalho, Fernanda Campos, Regina Braga, Jose Maria N. David, Victor Stroelle, and Marco Antonio Araujo. 2017. HEAL ME - An Architecture for Health Software Ecosystem Evaluation. Proceedings of the Joint 5th International Workshop on Software Engineering for Systems-of-Systems and 11th Workshop on Distributed Software Development (2017), 59–65. DOI:http://dx.doi.org/10.1109/jsos.2017.13

[8] Tom Mens, Bram Adams, and Josianne Marsan. 2017. Towards an interdisciplinary, socio-technical analysis of software ecosystem health. arXiv:1711.04532. Retrieved from https://arxiv.org/pdf/1711.04532.pdf

[9] Sami Hyrynsalmi, Marko Seppänen, Tiina Nokkala, Arho Suominen, and Antero Järvi. 2015. Wealthy, Healthy and/or Happy—What does 'ecosystem health'stand for?. International Conference of Software Business (2015), 272-287.

[10] Konstantinos Manikas and Klaus Marius Hansen. 2013. Software ecosystems–A systematic literature review. Journal of Systems and Software 86, 5 (2013), 1294-1306.

[11] Sonny van Lingen, Adrien Palomba, and Garm Lucassen. 2013. On the software ecosystem health of open source content management systems. 5th International Workshop on Software Ecosystems (IWSECO 2013), 38-49.

[12] Garm Lucassen, Kevin Van Rooij, and Slinger Jansen. 2013. Ecosystem Health of Cloud PaaS Providers. Lecture Notes in Business Information Processing Software Business. From Physical Products to Software Services and Solutions (2013), 183–194. DOI:http://dx.doi.org/10.1007/978-3-642-39336-5_18

[13] Jaap Kabbedijk and Slinger Jansen. 2011. Steering Insight: An Exploration of the Ruby Software Ecosystem. International Conference of Software Business (2011), 44–55. DOI:http://dx.doi.org/10.1007/978-3-642-21544-5_5

[14] Daniel Alami, María Rodríguez, and Slinger Jansen. 2015. Relating health to platform success: exploring three e-commerce ecosystems. Proceedings of the 2015 European Conference on Software Architecture Workshops - ECSAW 15 (2015), 43–49. DOI:http://dx.doi.org/10.1145/2797433.2797478

[15] Marco Iansiti and Roy Levien. 2004. Strategy as ecology. Harvard business review 82, 3 (2004), 68-81.

[16] Paul Van Vulpen, Abel Menkveld, and Slinger Jansen. 2017. Health Measurement of Data-Scarce Software Ecosystems: A Case Study of Apple's ResearchKit. Lecture Notes in Business Information Processing Software Business (2017), 131–145. DOI:http://dx.doi.org/10.1007/978-3-319-69191-6_9

[17] David Hillier, Iain Clacher, Stephen Ross, Randolph Westerfield, and Bradford Jordan. 2014. Fundamentals of corporate finance (2nd. ed.). McGraw-Hill Education, Maidenhead, United Kingdom

[18] Edward I. Altman. Predicting financial distress of companies: revisiting the Z-Score and ZETA models. Handbook of Research Methods and Applications in Empirical Finance, 428–456. DOI:http://dx.doi.org/10.4337/9780857936097.00027

[19] Roel J. Wieringa. 2014. Design Science Methodology - for Information Systems and Software Engineering. Springer, New York.

[20] Yates J. Monteithm John D. McGregor, and John E. Ingram. 2013. Hadoop and its evolving ecosystem. Hadoop and its evolving ecosystem. 5th International Workshop on Software Ecosystems (IWSECO 2013), 50-61

[21] Georgios Gousios. 2013. The GHTorrent dataset and tool suite. Proceedings of the 10th working conference on mining software repositories (213), 233-236

[22] Cloudera. Apache Hadoop ecosystem. Retrieved from https://www.cloudera.com/products/open-source/apache-hadoop.html.

[23] Hortonworks. Apache Hadoop ecosystem and open source big data projects. Retrieved from https://hortonworks.com/ecosystems/.

[24] Marc Orlitzky, Frank L. Schmidt, and Sara L. Rynes. 2003. Corporate social and financial performance: A meta-analysis. Organization studies 24, 3 (2003), 403-441.