Università degli Studi di Cagliari

Dottorato di Ricerca in Economia e Gestione Aziendale

Ciclo XXVII

Technological Cycles, Meta-Ranking and Open Access Performance

Settore scientifico disciplinare di afferenza:

SECS-P/10 – Organizzazione Aziendale

Coordinatore Dottorato: Prof. Andrea Melis

Presentata da: Gianfranco Ennas
Tutor Scientifico: Prof.ssa Maria Chiara Di Guardo

Esame finale anno accademico 2013 – 2014

University of Cagliari

Doctoral Thesis

Technological Cycles, Meta-Ranking and Open Access Performance

Author:
Gianfranco Ennas

Supervisor:
Prof. M. Chiara Di Guardo

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in

Business Administration

May 2015

"Some people never go crazy. What truly horrible lives they must lead."

Charles Bukowski

Abstract

This thesis consists of three essays, linked by innovation, classification and change. In the first paper, I analyze a theoretical problem regarding the re-emergence and affirmation of a technological paradigm over the others; in the second article, I propose a framework to aggregate journal rankings and classify academic journals; in the third essay I analyze the performance of Open Access journals, considered an innovative form of publishing, with the aim of identifying the main features of top-rated ones.

More specifically, the first essay deals on the technological life cycle which explains how the battles between competing technologies sooner or later end with the dominance of one over the others, or, under certain conditions, with their co-existence. However, the practice points out that, sometimes, beaten technologies can re-emerge in the market. Firms dealing with technology investment decisions need to completely understand the competing technologies dynamics, because the emergence of an alternative and potentially superior technology does not necessarily mean the failure of the incumbent, and different scenario would be traced. Starting from the analysis of the microprocessor market and considering the relationships with complementary companies, I show how the battle for dominance between two rival technologies can be reopened with a new era of ferment. While factors of dominance have been explored by a great amount of literature, little has been said on this question. In particular, I find a non-conventional S-curve trend and I seek to explicate its managerial implication.

The second chapter deals with ranking academic journals, an issue that during the years received several contribute from literature of Business and Management [DuBois and Reeb, 2000, Franke et al, 1990, Serenko and Bontis, 2004, Tüselmann et al, 2015, Werner, 2002]. Ranking journals is a longstanding problem and can be addressed quantitatively, qualitatively or using a combination of both approaches. In the last decades, the Impact Factor (*i.e.*, the most known quantitative approach) has been widely questioned, and other indices have thus been developed and become popular. Previous studies have reported strengths and weaknesses of each index, and devised meta-indices to rank journals in a certain field of study. However, the proposed meta-indices exhibit some intrinsic limitations: (*i*) the indices to be combined are not always chosen according to well-grounded principles; (*ii*) combination methods are usually unweighted; and (*iii*) some of the proposed meta-indices are parametric, which requires assuming a specific underlying data distribution. I propose a data-driven methodology that linearly combines an arbitrary number of indices to produce an aggregated ranking, using different learning techniques to estimate the combining weights. I am also able to measure correlations and distances between indices and meta-indices in a vector space, to quantitatively evaluate their differences.

The goal of the third essay, is to identify the features of top-rated gold open access (OA) journals by testing seven main variables: languages, countries, years of activity and years in the DOAJ repository, publication fee, the field of study, whether the journal has been launched as OA or converted, and the type of publisher. A sample of 1,910 gold OA journals has been obtained by combining SCImago Journal & Country Rank (SJR) 2012, the DOAJ, and data provided by previous studies [Solomon, 2013]. I have divided the SJR index into quartiles for all journals' subject areas. First, I show descriptive statistics by combining quartiles based on their features. Then, after having converted the quartiles into a dummy variable, I test it as a dependent variable and in a binary logistic regression. This work contributes empirically to better understanding the gold OA efficacy, which may be helpful in improving journals' rankings in the areas where this is still a struggle. Significant results have been found for all variables, except for the types of publishers, and for born or converted journals.

Contents

List of Figures

List of Tables

Chapter 1

Technological Cycle and S-Curve: Managing a Non-Conventional Trend

1.1 Abstract

The technological life cycle explains how the battles between competing technologies sooner or later end with the dominance of one over the others, or, under certain conditions, with their coexistence. However, the practice points out that, sometimes, beaten technologies can re-emerge in the market. Firms dealing with technology investment decisions need to completely understand the competing technologies dynamics, because the emergence of an alternative and potentially superior technology does not necessarily mean the failure of the incumbent, and different scenario would be traced. Starting from the analysis of the microprocessor market and considering the relationships with complementary companies, I show how the battle for dominance between rival technologies can be reopened with a new era of ferment. While factors of dominance have been explored by a great amount of literature, little has been said on this question. In particular, I find a non-conventional S-curve trend and I seek to explicate its managerial implication.

1.2 Introduction

The success of highly innovative companies such Intel, Microsoft or Google is due to the capacity they had to impose their standards. A lens to analyze these fascinating stories is the technological life cycle, which permits to understand the emergence of a technology over the others. Life cycle theory has been applied since ages to describe the behavior of technologies and its involvement in decision making, and in order to understand the specific implications for managers. The “macro view” of technology life cycle [Anderson and Tushman, 1990] consists of individual technology cycles beginning with a period of *technological discontinuity*, during which process or product advancements gradually guide to a *period of ferment*, when the competition between technologies takes place. Then, when a new technology prevails, it will substitute the older one and the design competition will begin. Only when a technology is largely adopted and modifies the nature of competition within the industry, the *dominant design* emerges and becomes the industry standard. An era of *incremental evolution* of the selected technology follows, characterized by evolutionary, continuous and incremental changes, until a further technological discontinuity, when a new cycle begins and this *creative destruction* process restart [Schumpeter, 1934].

However, the innovation adoption process does not follow rigid schemes, in fact, the dominant design is not even the best available, but *the most widely accepted*. For this reason, it is worth understanding why a particular design succeeds over the others: technology characteristics, firm level as well as environmental factors can influence the competition outcome [Murmann and Frenken, 2006, Suarez, 2004]. Among these, network externalities are crucial in the battle for dominance [Anderson and Tushman, 1990, McIntyre and Chintakananda, 2014]. Furthermore, in case of network effects, the product and its sponsoring firm may finally lock in the market for a given good [McIntyre and Chintakananda, 2014]. From the literature, it is also know that exceptional situation may take place. In fact, the era of ferment may persist for up to twenty years before a technology prevails and is locked as a dominant design [Schilling, 2002]. Again, the coexistence of technologies occurs when rival technologies compete in the same market without exclude each other, due to technology and market factors. In this situation of "creative persistence", the period of ferment indefinitely extends, preventing the emergence of a clear winner or the exit of losers [Nair and Ahlstrom, 2003]. Again, technology can also re-emerge in the market after a period of failure, because of a number of influencing factors. Then, it seems that technologies tend to displace older ones, re-emerge, co-exist with and even come to dominate newer technologies: this process represents the creation and re-creation of products, organizations and communities identities [Raffaelli, 2013].

In this paper, I analyze the microprocessor market with the purpose to demonstrate an unconventional trend in the technological cycle and its consequences. In particular, I show how the battle for dominance between rival technologies can be reopened with a new era of ferment. While factors of dominance have been explored by a great amount of literature, little has been said on this question. Besides the academic relevance of these considerations, we have to acknowledge that microprocessors are widely used (*i.e.* in PC, server, smartphone, tablet and so on) and they operate with complementary goods such as Operating Systems (OS) and software in general. For this reason, I consider the relationships between companies who act as complementaries, providing coordinated products that jointly appeal to end customers. The effects of complementaries networks in the technological cycle evolution are important for several reasons: they confer resources critical to the success of contemporary high-technology ecosystems. Secondly, they involve coordinated product launches and mutual dependency for success. Last but not less important point, they provide insights into the technological cycle evolution [Venkatraman and Lee, 2004] The paper is structured as follows:

the second paragraph presents a literature review about the technology life cycle, the third paragraph is devoted to the study of the microprocessor market. The fourth part explores evidences from smartphone and tablet markets, followed by the discussion paragraph, which identifies the management implications and main limitations. Finally, conclusions propose an indication of possible developments for further research.

1.3 The technology life cycle: a literature review

Firms need to be able to position technologies within their life cycle, and to understand the specific implications of this for managerial decisions [Taylor and Taylor, 2012]. Even if a clear conceptualization of the life cycle of a technology is difficult, the Anderson and Tushman [1990] technology evolution model is a central perspective and represents the foundation of to the "macro view" of the technology life cycle. The macro view considers individual technology cycles, each of which begins with a period of technological discontinuity, characterized by advances in a process or in a product that immediately lead to a second cycle, the period of ferment. This era sees the competition among different variations of the original technology, and it is divided into two phases: substitution and design competition, once the superiority of the new technologies has been demonstrated, they rapidly substitute the older and the design competition begins [Anderson and Tushman, 1990]. Then, when a technology is widely adopted and associated with changes in the nature of competition within the corresponding industry, the design competition ends with the emergence of the dominant design. It usually involves a synthesis of available technologies, resolution of competing technological standards, and perceptions of closure by user groups [Pinch and Bijker, 2009]. This period could be followed by an era of incremental evolution of the selected technology, characterized by evolutionary, continuous and incremental changes, until a further technological discontinuity, when a new cycle begins. This cyclical process of technological change is what Schumpeter (1934) named "creative destruction".

Although there is a general agreement that the Anderson and Tushman's model concerns innovations of both products and processes, the emphasis changes between these during the cycle. Indeed, during the era of ferment the focus is on the product technology with the emergence of a dominant standard, while in the era of incremental change greater emphasis is placed on the development of processes that will improve the product technology [Taylor and Taylor, 2012]. The dominant design needs not to be the best available, it needs only to gain a widespread acceptance. An inferior one can win and, in this way, scholars have appealed to a variety of factors explaining why a particular design rather than other ones emerges as the dominant. In reviewing the dominant design literature, five groups of causal mechanisms have been classified [Murmann and Frenken, 2006]: the technological predominance among different functional characteristics of a technology; the economies of scale that can be realized with standardized products; network externalities and their effects (path-dependent processes); firms strategies; combination of historical, sociological, political and organizational dynamics. Among these, economies of scale and network externalities are the two conditions that create dynamic increasing returns and even the design with the small lead will inexorably win a dominant position if higher returns can be achieved with it. In particular, network externalities generates when the utility that a user derives from consumption of the good increases with the number of other agents consuming the good, who are in the same "network". The possible sources of network externalities could be direct physical effects, indirect effects (*e.g.* the hardware-software paradigm) and post-purchase services [Katz and Shapiro, 1985]. Studying the process by which a technology achieves dominance when battling against other technological designs, two broad groups of factors influencing the outcome have been classified [Suarez, 2004]: firm level factors and environmental factors. There are a number of examples regarding the emerging of a technology over another; among these, the most meaningful and cited are VHS versus Betamax [Besen and Farrell, 1994] and QWERTY versus other keyboards layout [David, 1985]. In the first case a better format usability, the additional time available for recording and the widespread diffusion of movie shops adopting the format increased the preference of VHS instead of the better quality that characterized the Betamax format. In the second case, the first product available with a new technology dominated most of the market; this is a good example of lock-in and path-dependence caused by dynamics that go beyond the behaviors of individuals, and show that, when a new technology is introduced and spread so largely and quickly, it is quite impossible to come back to the old one.

The market diffusion of a technology is plotted by the S-curve [Foster, 1988], whose common interpretation considers the cumulative adoption of the technology over time, envisioning a number of phases such as embryonic, growth, maturity and ageing. There are also alternative interpretations, but however plotted, the S-curve reach saturation at maturity, when a new disruptive technology may emerge to replace the old one. This period of technological discontinuity is characterized by competing technologies with their own S-curve, which could be connected or disconnected each other, in relation to the higher rate of performance. The resulting situation is a technology progression characterized by multiple S-curves or technology cycles occurring over time [Taylor and Taylor, 2012]. Some scholars pointed out that the period of ferment may indefinitely extend and not resolve with the dominance of a standard among others, but the rival technologies may coexists under certain conditions [Galvagno and Faraci, 2004, Nair and Ahlstrom, 2003]. The coexistence of technologies changes the linear and systematic course of the technology life cycle and it is generated when different competing technologies occur simultaneously in the same market, without exclude each other.

According to the literature, the technology complexity, regulatory regimes and factors connected with the intermediate and final markets demand, influence the interaction among competing technologies, preventing the emergence of a clear winner or the exit of losers. When such dynamics exist, the distinct features create product niches and consumer communities, gateway technologies, multichannel end systems, appropriability regime and persistency. In particular, a niche is defined as containing one consumer group or "class": since each class has a distinct preference set (*e.g.* a particular point in quality/price space), the number of potential market niches is determined by the number of consumer classes that are initialized by the modeler. It has been observed that the survival of the new technology requires the establishment of a protected space in which further development can be achieved [Rosenberg, 1982]. This can take the form of distinct niche or sub-niche in the market, which may be complementary to the established technology, or else take the form of public sector support, where users are often also contributors to the R&D process. The protection afforded by its niche has enabled the technology to be further developed and improved [Windrum and Birchenhall, 2005]. A practical case is given by different types of flash memory card [De Vries et al, 2011]. The coexistence thus is highly probable in any case of similarity between technologies. While the coexistence manifestation and duration is obviously different depending on the type of technology and on whether intervening factors, surely each of these factors can individually or simultaneously affect the duration of the competition between technologies, and determine the presence within the same market. In such situations, the creative destruction does not seem to be the rule. It is possible to assume a kind of "creative persistence" and a coexistence of different technological solutions [Galvagno and Faraci, 2004].

Another situation that moves away from the linearity of the technology cycle is the re-emergence case, that is the case in which a technology fails at one time period, exits the market, but later returns. Factors concerned with the re-emergence of a technology are: institutional shaping, competing alternatives, rate of learning, market characteristics, firm strategic positioning, key firm networks and firm age and size. Although new or discontinuous technologies tend to displace older ones, technologies can re-emerge, co-exist with, and even come to dominate newer technologies. This process seems to be the creation and re-creation of product, organization, and community identities [Raffaelli, 2013].

1.4 What is different in the microprocessor market?

The microprocessor market, is particularly interesting because characterized by battles for emerging and strategic alliances between big corporate such IBM, Microsoft, Intel and so on. Again, this market, presents several advantage in studying technological cycles because [Tegarden et al, 1999]: (*i*) support many design, (*ii*) there are high switching cost between rival and incompatible designs, due to hardware/software incompatibilities, (*iii*) presence of high network externalities, (*iv*) high growth in both customers and the number of (*v*) the introduction of the IBM PC effectively changed the nature of competition in the personal computer market by imposing a clear standard architecture.

To better understand the dynamics of microprocessor market it is important to point out there is reciprocal interdependence between CPU and OS; this means that the evolution of one of them influence the evolution of the other(s). In fact, since the beginning of PCs diffusion, combination between CPU architecture and OS were determinant. A practical example can be found in the middle 1970's, when Zilog Z80 processor and CP/M OS became the dominant CPU & OS combination of the period circa 1976 to 1983, and despite the great commercial success of the Apple II and its OS, Apple was forced to produce a compatible card that allow to install CP/M OS also in its computer. It can be said there are fundamentally two rival architecture designs in microprocessor: RISC (reduced instruction set computer) and CISC (complex instruction set). The question between them is longstanding, and there was an important concern in the 1980s and 1990s, when chip area and processor design were the primary constraints. In the past decades, the Intel and Advanced Micro Devices Inc (AMD) x86 (CISC CPU) has dominated desktops and servers markets, while the ARM Ltd (RISC CPU) were in the low-power embedded computing segment [Blem et al, 2013]. The companies have two different strategies: while ARM design and just sell licenses to producers (Mediatek, NVIDIA, Qualcomm and so on), INTEL and AMD design and produce their own products. Today, the x86 architecture is arguable the only chip which retains CISC architecture, though newer Intel's processors in some ways are hybrid and called "CRISC". RISC CPUs were considered superior for many technical points [Krad and Al-Taie, 2007] but, in the 1981 IBM launched the Personal Computer (PC), with Intel supplying the microprocessor and Microsoft the OS. As a group, this triad created the microcomputer format that, within a few years, drove competitors to the periphery of the market. IBM would not purchase a device unless it was made by at least two companies, so they would contract with other manufacturers to make their design. Having other companies manufactured this design, or compatible parts, also increased the market share of that architecture (i.e. cross-license agreement Intel/AMD). By 1985, the Intel microprocessor was embodied in the majority of personal computers shipped (55% or 175 out of 277 firms shipping personal computers used an Intel microprocessor) [Tegarden et al, 1999]. Notwithstanding, in the 1987 the cross-licensing agreement between AMD and Intel terminated, and the triad slowly fell apart, Microsoft and Intel went on to develop the powerful "Wintel" alliance, which established the dominant industry standard [Gomes-Casseres, 2003]. History and literature teach us that, when industries are characterized by network externalities, the installed base technology and the availability of complementary goods will play major roles in user adoption. An insufficient installed base or lack of complementary goods may result in technology lockout [Semmler, 2010]. As we have seen above, the reason why CISC processor has won is not due to a technical supremacy over RISC but, as happened in the previous examples (VHS vs Betamax and QWERTY keyboard), to various of factors.

In ICT industries network externalities are more pervasive than in other [Lin et al, 2011], in particular indirect network externalities exist. The value of a PC is influenced by the level and the variety of the supply of applications that is possible to utilize with it. From this statement we can easily understand why once a combination between OS and CPU architecture is established it generates high switching costs and then lock-in, because semiconductor manufacturers tend to produce unique and incompatible designs. Both PC software and drivers for peripherals must be designed around the microprocessor, and switching to another one can be extremely costly; it might involve extensive redesign of the product, or a total washout of costs incurred in the development of customized software [Choi, 1994]. Switching costs also go well beyond the product changes to include the costs associated with coordinating a product component change within the organization as well as between suppliers and customers. A firm attempting to modify a design will face costs due to modifying documentation, increased communication between marketing, engineering and production, obsolete inventory, and the lost time of key personnel which need to deal with the unknowns associated with quality and performance variations in their product [Tegarden et al, 1999]. In addition, the manufacturer must undertake search costs (both money and time, involving in some cases both suppliers and buyers), set up new external relationships, and face uncertainties in input quality [Garud and Kumaraswamy, 1993].

1.5 How smartphones and tablets are revolutionizing the market

From the statement in the previous paragraphs we could say the CPU market should continue to be dominated by CISC architecture until a new discontinue innovation will open a new era of ferment. But, innovation, by definition, is never linear nor follows schematic trends. Surprisingly, we will see RISC architecture is moving from the embedded market to the mainstream one and then how the equilibrium in the microprocessor market is changing with a "new era of ferment". In doing so, I have considered the two most important up-to-date technologies, smartphones and tablets, where a clear dominant standard is emerging, with the ARM-based CPUs that have achieved a more than 95% penetration (ARM Ltd annual reports & accounts 2013). I have particularly checked:

1) If incumbents (in the desktop and server markets) - Intel and AMD - recognize ARM as a challenge in their core business;

2) If new entrants - ARM - recognize the opportunity to enter other markets.

According to Intel (form 10-k 2012-2013), new competitors are joining traditional ones in their core PC and server business areas, where they are leading provider, while they face incumbent competitors in adjacent market segments they are pursuing, such as smartphones and tablets. In particular, Intel competitors include Advanced Micro Devices Inc. (AMD), International Business

Machines (IBM), Oracle Corporation, as well as ARM architecture licensees from ARM Limited, such as QUALCOMM Incorporated, NVIDIA Corporation, Samsung Electronics Co., Ltd. and Texas Instruments Incorporated. AMD (form 10-k 2012-2013), points out that Intel's dominant position in the microprocessor market and integrated graphics chipset market, its existing relationships with top-tier original equipment manufacturers (OEMs) and its aggressive marketing and pricing strategies could result in lower unit sales and average selling price for its products, which could have a material adverse effect on them. Additionally, it indicates that other competitors include companies providing or developing ARM-based designs as relatively low cost and low power processors for the computing market, including netbooks, tablets and thin client form factors, as well as dense servers, set-top boxes and gaming consoles. ARM Holdings designs and licenses its ARM architecture and offers supporting software and services. Its ability to compete with companies who use ARM based solutions depends on its ability to design energy-efficient, high-performing products at an attractive price point. In addition, Nvidia builds custom CPU cores based on ARM architecture to support tablets and small form factor PCs, servers, workstations and super computers. AMD states its willingness to transform the business to reach approximately 50% of revenue from high-growth markets by the end of 2015. AMD also states that they will sample their first ARM technology-based processor for servers in the first quarter of 2014. ARM (annual reports & accounts 2012-2013) on the other hand, confirm to keep over the 95% of the market share in the smartphones and tablets markets, with an increase by more than 100% year-on-year. ARM reported its customers shipped more than 10 billion ARM-based chips into everything from phones and tablets to smart sensors and servers, and it points out that it faces competition both from large semiconductor companies and from smaller companies. Regarding big competitors, Intel is developing processors for use in PCs and servers, and it is looking to deploy these chips in markets such as tablets, mobile phones and embedded markets, including the Internet of Things. Any success by its competition would result in a reduction in royalty revenue to ARM. ARM expects that its customers will continue to re-equip their R&D teams with the latest processors for existing product lines. In addition, ARM's technology is becoming increasingly relevant to growing markets such as sensors, computers and servers, leading to new customers acquiring their first ARM license. Additionally, Chromebook -a notebook shipped with Google OS- within its range of 17 different models, accounts 4 shipped with ARM CPU, 3 made by Samsung and 1 by HP. As of May 2013, the Samsung ARM Chromebook has led Amazon's list of best-selling laptop. With the launch of latest products, ARM has introduced additional features that are important to data center environments and is expected to challenge Intel in the server and PC markets.

1.6 Managing a non-conventional technological cycle

The above findings support the idea of "reopening for dominance between two rival technologies" and are particularly interesting from a managerial point of view. To understand the trend of CPU technological cycle, it is interesting and crucial to investigate the corporate strategies both for CPU and OSes. To start with order, I first look at CPU market leader strategies, than the incumbent one and finally the

OS maker one. As seen above, Intel is the leader in the desktop and server industries, and to keep its supremacy it has decided to exploit its technology; in fact, it "is innovating around energy-efficient performance", and it is "accelerating the process technology development for its Intel Atom processor product line to deliver increased battery life, performance, and feature integration". Intel recognizes to be a relatively new entrant in the tablet market segment, and it is trying to offer optimized architecture solutions for multiple operating systems and application ecosystems. It also recognizes that boundaries between the various segments are changing, as the industry evolves and new segments emerge. Conversely, AMD has ever had a smaller market share in the desktop and server markets, thus, it has decided to adopt an *ambidextrous strategy*. With this strategy AMD is trying to be able to both explore into new spaces as well as exploit their existing capabilities [O'Reilly and Tushman, 2008]. In fact, AMD is differentiating its strategies by licensing ARM, in addition to its x86 processors. Software makers have to be able to manage that innovation, in fact, Microsoft, as a leader in the desktop and notebook OS markets, has recognized the threat of new devices. In particular it states (form 10-k 2013) its system faces competition from various commercial software products and from alternative platforms and devices, mainly from Apple and Google. Consequently, it has adapted its strategy, releasing Windows 8, the first version of the Windows operating system that supports both x86 (CISC) and ARM (RISC) chip architectures (for non-embedded purposes). Conversely, software developed for the Android OS may run in every architecture because, simplifying, just like java, it uses a virtual machine to run software.

Considering these premises, it can be stated, with reasonable evidences that *the*

S-curve follows a different trend in this market, and almost three decades after the alliance between Intel and Microsoft that drove RISC processor out of PC and server markets and signed the emerging of the dominant paradigm, the *challenge is reopened*: the first phase has been the affirmation of CISC technology, followed by a long period of incremental improvement; meanwhile, the RISC technology gained lower adoption, up to the advent of smartphone and tablet, that caused a rapid rise of RISC architecture. Therefore, it can be assumed, that the S-curve might follow the trend proposed in Fig. 1.2, that is a different from the common interpretation Fig. 1.1, which considers that, once a technology prevails, keeps its supremacy until a new disruptive technology enter and defeats the market. Indeed, in the CPU industry two technologies have coexisted, the CISC dominating the market and the RISC relegated to the embedded segment, but with the advent of new complementary goods - tablets and smartphones - the adoption of RISC systems is experiencing a rapid growth with a sudden change in the curve concavity. According to the analysis presented above, they are currently facing a *new era of ferment*, and basically three future scenarios can be envisaged:

1) The CISC technology maintains its supremacy and follows the trend described by the yellow curve, while the RISC one follows a lower trend, described by the green curve.

2) The RISC technology imposes its own standard in the market segments currently dominated by CISC, and follows the trend described by the blue curve, while the CISC one proceed along the lower trend described by the red curve.

3) Both technologies coexists in different market segments, without exclude each other.

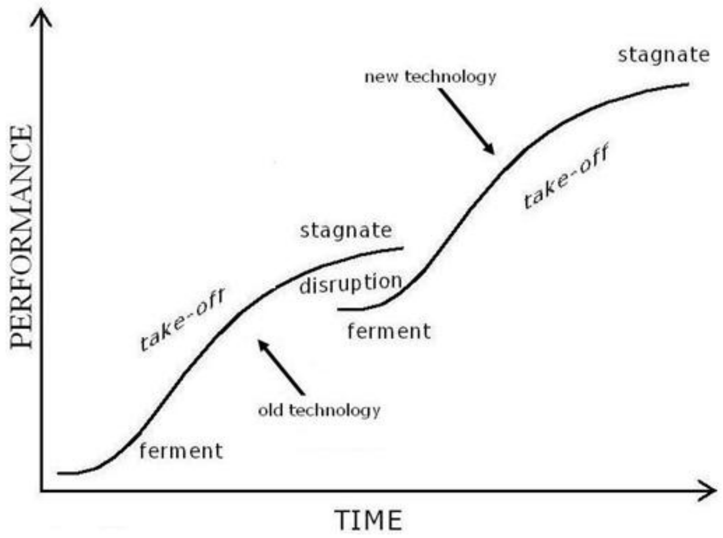

FIGURE 1.1: S-Curve.

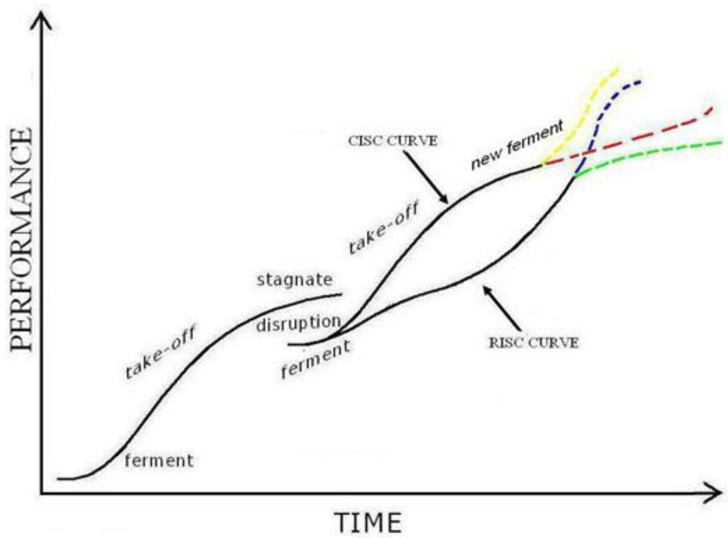

FIGURE 1.2: S-Curve and the unconventional trend.

Regardless of how things go actually, it is clear that this trend of S-curves is very different from what we know. From a strategic management point of view, my study reinforces the need to understand the possible alternative trends of the technological cycle and it offers some pointers for management practice. Just to quantify the importance of these trends, it must be considered that processor market generates a turnover of around 300 billion dollars and is moving earnings from a technology to another. Additionally, as previously pointed out, devices equipped with a CPU are complex systems, therefore implications will affect the software. Hence, suggestions I stated below can be applied also for other markets, in particular, the software market.

First of all, in battles for dominance *strategic alliances* between hardware and software makers and the subsequent indirect network externalities play a crucial role, because the amount of complementary products and services available can strongly contribute to the affirmation of a technology over another. Therefore, firms need to look inward to identify competencies they need in order to respond quickly to technological changes. In doing so, they have to consider that inter-organizational coordination of product architectures is not limited to microprocessor but appears to operate in many other settings. Coordination of complementary quotes a new competency for managers, named "network orchestration", focused on finding the balance between cooperation and competition across different types of resources over time and involving the coordination of multiple types of relationships.

Additionally, network orchestration may be a critical driver of superior performance: in periods of turbulent change a network can be a source of strength or a source of constraint and this competency entails managers' simultaneously focusing on the macro and micro logic of network structure, that is how relationships are structured for resource access and how the selection, cultivation, and dissolution processes contribute to maximal performance [Venkatraman and Lee, 2004]. Secondly, during the process by which standards are established firms crucially affect the competitive environment in which they will operate, and they must assess the extent to which industry profits will be dissoluted by competitors. In fact, firms choose to join a rival's network or to offer its technologies to rivals, and this choice depend not only on its views about how likely it is to prevail in each form of competition but also on the nature of the competition itself. When firms are similar, they will probably choose the same compatibility strategy: if all are willing to offer compatible products, standards are likely to emerge fairly easily; if all want compatibility, but only on their own preferred technologies, each will try to encourage its rivals to join its network; finally, if all try to establish propriety standards, an all-out standards battle will ensue. Conversely, when firms are dissimilar, conflicting strategies are more likely: *newcomers may prefer to join the network of an industry leader while the leader tries to prevent them from doing so.* Here, firms do not choose how to compete but fight over how to compete [Besen and Farrell, 1994].

Firms dealing with technology investment decisions need to completely understand the competing technologies dynamics, because the emergence of an alternative and potentially superior technology does not necessarily mean the failure of the incumbent, and different scenario would be traced. In particular, they have to take into account that bringing new technologies to market is a time and resource consuming process with the potential to lengthen the commercialization cycle. The possible extended commercialization is often due to *technical* challenges - associated with developing the technology - and *social* challenges - due to integrating a new system into an established one -. Additionally, market demand during the technology emergence has the potential to influence if commercialization becomes fully realized during a given cycle. These difficulties can inevitably lead to technology market failure [Gilbert et al, 2014].

1.7 Conclusions

In this paper I have explored the CPU market and its relationships between companies who act as *complementors*, finding that the era of ferment may restart after a long period of time, and technologies competing in distinct segments races each other. These results suggest that the S-curve may have a different trend and propose a non-conventional view of the technology adoption process. Again, I have analyzed the trend from a managerial point of view. However, I believe that in addition to these preliminary considerations, this research has thrown up many questions regarding the technology diffusion in need of further investigation. Although I have evidence from the microprocessor market, the insights of this study should be confirmed in other context to extend, generalize and eventually improve technological cycle literature. If it is true that not even the best technology wins, I have shown that this could be a dynamic position and the era of ferment may be re-opened.

Chapter 2

Data-driven Journal Meta-Ranking in Business and Management

2.1 Abstract

Ranking journals is a longstanding problem and can be addressed quantitatively, qualitatively or using a combination of both approaches. In the last decades, the Impact Factor (*i.e.*, the most known quantitative approach) has been widely questioned, and other indices have thus been developed and become popular. Previous studies have reported strengths and weaknesses of each index, and devised meta-indices to rank journals in a certain field of study. However, the proposed meta-indices exhibit some intrinsic limitations: (*i*) the indices to be combined are not always chosen according to well-grounded principles; (*ii*) combination methods are usually unweighted; and (*iii*) some of the proposed meta-indices are parametric, which requires assuming a specific underlying data distribution. I propose a data-driven methodology that linearly combines an arbitrary number of indices to produce an aggregated ranking, using different learning techniques to estimate the combining weights. I am also able to measure correlations and distances between indices and meta-indices in a vector space, to quantitatively evaluate their differences.

2.2 Introduction

Ranking academic journals is an issue that affects many players, especially in academia; *e.g.*, scholars choosing among potential outlets for their research, departments to measure their productivity and to ensure funding success. It also affects the non-academic world, such as publishers aiming to evaluate the quality of their journals, professional societies, practitioners and funding organizations [Falagas et al, 2008, Vokurka, 1996]. Obtaining a reliable journal ranking is a longstanding problem and although is a concern affecting all sciences, it has received several contribute from literature of Business and Management [DuBois and Reeb, 2000, Franke et al, 1990, Serenko and Bontis, 2004, Tüselmann et al, 2015, Werner, 2002]. Its intrinsic difficulty relies on the multidimensionality of the quality concept. In fact, research quality can be measured qualitatively, quantitatively, or using a combination of both approaches (referred to as *hybrid approach* in the following). Qualitative approaches consist of ranking journals according to their perceived quality and reputation, *e.g.*, by interviewing a qualified sample of experts to rate journals in a particular field of study [Mylonopoulos and Theoharakis, 2001, Peffers and Ya, 2003, Sellers et al, 2004]. Quantitative methods, instead, provide indices depending on the number of published articles in a journal, and the corresponding number of citations [DuBois and Reeb, 2000, Hodge and Lacasse, 2010, Seglen, 1997]; *i.e.*, they evaluate two main aspects called respectively *size* and *impact* [Leydesdorff, 2009]. Hybrid approaches combine experts' opinions and quantitative approaches to capture both perspectives.

Since its proposal, the Impact Factor (IF) has been widely used as a quantitative approach; however, several limitations related to the (mis)use of this index have emerged (Sect. 2.4.1) [Bornmann et al, 2012, Lancho-Barrantes et al, 2010, Seglen, 1997]. For this reason, alternative indices to the IF, such as the SCImago Journal Ranking (SJR) and the H-index, have been developed and become popular in academia, although they exhibit other kinds of limitations (Sect. 2.4.2) [Leydesdorff, 2009]. This has fueled the development of meta-indices to mitigate issues specific to the use of each base index. In reviewing previous work [Bador and Lafouge, 2010, Hodge and Lacasse, 2010, Theuβl et al, 2014, Tsai, 2014, Vanclay, 2008], three main issues have emerged: (*i*) indices to be combined are not always chosen according to well-defined guidelines, affecting the reliability of the corresponding meta-index; (*ii*) combination methods are usually unweighted; and (*iii*) some of the proposed meta-indices are parametric, requiring specific assumptions on the underlying data distribution (Sect. 2.4.3).

The contribution of this work is twofold: (*i*) to overcome the aforementioned limitations, I propose a data-driven methodology that learns a weighted, linear combination of an arbitrary number of indices (potentially also including experts' opinions), yielding a more principled, aggregated journal ranking; (*ii*) I am also able to measure correlations and distances between indices and meta-indices in a vector space, in order to compare their relative performance (Sect. 2.5). In doing so, I empirically validate this approach using two journal databases obtained by combining journals indexed in Business and Management from the Thomson Reuters Journal Citation Report (JCR) and from SCImago Journal & Country Rank (SJR) (Sect. 2.6). I finally discuss conclusions and future research directions in Sect. 2.7.

2.3 The Importance of journal ranking in Business and Management

Although studies attempting to rank and evaluate academic journals are common in various disciplines, historically the fields of Business and Management have significantly contributed to the development of the *science metrics* (*i.e.* bibliometrics, scientometrics, informetrics and so on) [DuBois and Reeb, 2000, Franke et al, 1990, Serenko and Bontis, 2004, Tüselmann et al, 2015, Werner, 2002]. On the other hand scientometricians construct indicators that can be used in policy and management contexts [Leydesdorff and Milojević, 2012, Martin and Irvine, 1984]. Actually these research fields assume crucial importance worldwide because they are used to evaluate Universities, Departments, Faculties and also researchers. In particular, in Italy, the Agency for the Evaluation of University and Research Systems (ANVUR) with the 2004–2010 VQR (Research Quality Evaluation) have used a hybrid peer-review/bibliometrics approach to ranking Universities. This ranking is now used to determine the allocation of financing for each University [Abramo and D'Angelo, 2015]. ANVUR hybrid indicators are also used in contest for Researcher and Professor positions in Italian Universities, they have been criticized by Academicians precisely for the methodology used to rank journals [Abramo and D'Angelo, 2015, Hicks et al, 2015]. Thus, the framework proposed in this chapter may be helpful in building indicators to improve the ANVUR research evaluation to provide a fairer reallocation of financial resources in Italian Universities and to evaluate candidates in public contest.

2.4 Background

In this section, I start analyzing the main limitations of the journals Impact Factor and present the SJR and H indices; then I discuss on the need to combine indices and briefly review literature related to "meta-indices".

2.4.1 Limitations of the Impact Factor

As stated in the introduction, the various indices used to measure research quality have limitations in nature or in capturing the different dimensions of the quality concept. In particular, the Impact Factor (IF) index, was developed by Eugene Garfield [Garfield, 2006] now is a product of Thomson Reuters Corporation and by definition is: *"the average number of times articles from the journal published in the past two years have been cited in the JCR year"*.

$$\mathrm{IF} = \frac{\text{citations in } Y_n \text{ of documents published in } Y_{n-1} + Y_{n-2}}{\text{citable items in } Y_{n-1} + Y_{n-2}} \quad . \tag{2.1}$$

It has been criticized for multiple reasons such: (i) small research fields tend to lack journals with high impact; (ii) citation rates of articles determine journal impact but not vice versa; (iii) IF is a function of the number of references per article in the research field; (iv) in some journals (*e.g.* Nature) letters and correspondence are considered citations and, of course, they inflate the index; (v) journal impact factors are not statistically representative of individual journal articles. In other words, different scientific areas, fields and micro-fields of study have different citation habits; (vi) IF have a limited coverage, in particular in the Social Sciences and Arts and Humanities [Bornmann et al, 2012, Lancho-Barrantes et al, 2010, Seglen, 1997]. Even, the academic journal Scientometrics dedicated a special issue in the 2012 aimed to evaluate IF problems also in comparison to its counterparts. Despite all these limitations, IF is widely used, for three reasons: the first is a *path dependence*, as since its introduction scholars and editors have learnt to use it, and without any strong alternative it has become a "*de facto* standard"; the second reason is that the computation method of this index is intrinsically easy to understand; and the third and last reason is that, due to its simplicity and growing popularity, it has been also (mis)used to measure the overall impact of journals. However, it is worth reminding that IF was originally thought to a different end, namely, to measure the impact of an average item published in a journal [Harter and Nisonger, 1997].

2.4.2 SJR and H indices

During the last ten years, two indices have become popular; they are the SJR index and the H-index. In 2004, Elsevier launched the Scopus database as an alternative to the Thomson Reuters ISI databases and soon become widely used in academia [Vieira and Gomes, 2009]. It includes more titles, from more countries and published in a greater variety of languages than the Thomson Reuters one [Leydesdorff et al, 2010].

In the Scopus database, is calculated the SCImago Journal & Country Rank index (SJR), it expresses the average number of weighted citations received in the selected year by the documents published in the selected journal in the three previous years [SJR, 2007]. Conceptually, this index is easy to understand but its calculation is not; in fact, it relies on an iterative algorithm that distributes prestige values among the journals until a steady-state solution is reached. The SJR algorithm starts giving an identical amount of prestige to each journal, then this prestige is redistributed in a process where journals transfer their achieved prestige to each other through citations. The process ends up when the difference between journals prestige values in consecutive iterations do not reach a minimum threshold value any more. The use of the SJR index reduce the influence of self-citations because they cannot be more than 33 per cent of the total; the prestige can be transferred to a journal by all other journals, but not by itself [González-Pereira et al, 2010]. Recently, it has been shown a correlation between SJR and IF Falagas et al [2008] show that half of the journals in the IF top 100 list are placed within a reasonable range of ranking places in the SJR indicator journals list. Furthermore, IF and SJR are directly comparable because use a similar time window, respectively 2 years the first and 3 years the second.

The H-index was proposed about a decade ago by Hirsch and rapidly gain a widespread acceptance. Its initial aim was to measure scholars productivity and citation impact: "a scientist has index h if h of his or her Np papers have at least h citations each and the other $(Np - h)$ papers have less than h citations each" [Hirsch, 2005]. Subsequently [Braun et al, 2006] show how that index can be successfully applied to journals instead of scientists. The first advantage is that H-index does not have an artificially fixed time horizon, theoretically it could be calculated since the creation of the journal, even if this should not be appropriate. Again, is insensitive of an excess of uncited/highly cited article, because is not based on mean scores, this could be seen also as a disadvantage [Braun et al, 2006, Leydesdorff, 2009]. Lastly, it has been demonstrated that normalizes for size a bit more strongly than the IF and is highly correlated with expert opinion [Hodge and Lacasse, 2010]. According to Leydesdorff [2009] SJR is more close to the size dimension than the IF, while the H-index attempts to capture both dimensions.

2.4.3 Why Combining?

Whereby different properties of various indices or experts opinion, the strategy to combine two or more rankings have become a widely used techniques to obtain a meta-index. Most studies to date has tended to focus on comparison between IF and H-index ranking for journals in a certain field of study. Vanclay [2008] reports data for IF and H-index for forestry journals. In social work, Hodge and Lacasse [2010] show correlation between IF, 5-year IF, H-index and expert opinions. Again, in literature there are studies attempting to combine IF and H-index with the aim to obtain a meta-ranking. In particular, Bador and Lafouge [2010] show four groups of journals divided per quartile according to their categorical combined score in IF and H-index. In that work it is assumed that the distribution of the citations random variable is a discrete Paretian distribution with finite expectation.

More recently, it has been applied an optimization-based consensus ranking method to a dataset of journals from the Harzing List with the aim to construct suitable aggregates of individual journal rankings [Theuβl et al, 2014]. As authors recognize, this method is not robust with the increasing of individual rankings, there are implications on the stability of the derived aggregate ranking. Again the solution tends to degenerate as the size of the journal list explodes. In a recent study, it has also been proposed a combination method (CombSUM) between IF and H-index to re-ranking journals in computer science [Tsai, 2014]. That technique of data fusion has the limitation to be not weighted. Finally, Tüselmann et al [2015] tried to place an order, proposing a framework to deal with missing values and parametric problem, using a meta-index based on a modified Data Envelopment Analysis (DEA) model. Despite their significant contribution, final ranking is populated by several equal positions (*e.g.* 4 journals in the first position, three in the ninth and so on). In the next section I propose a statistical approach to overcome all these limitations, able to combine an arbitrary number of indices, assigning to each a different weight without making any underlying assumption on the data distribution.

2.5 Learning Aggregated Indices for Meta-Ranking

In this section, I discuss a methodology to aggregate existing indices, aiming to capture the different dimensions characterizing aspects of research *quality* in a consistent manner. Although this model can be also used to combine the experts' opinion, in this paper, the analysis is limited to the combination of quantitative indicators. As mentioned in Sect. 2.4, the goal is to propose an index aggregation scheme that overcomes the limitations emerged from the state of the art. To this end, I consider a simple linear combination of indices whose weights can be determined based on specific (and potentially different) criteria. Note also that some of the previously-proposed meta-indices can be expressed in terms of a linear combination of indices, as discussed in the following.

Let us assume we are given a set of journals $\mathcal{D} = \{\boldsymbol{x}_i\}_i^n$, where $\boldsymbol{x}_i = (x_i^1, \dots, x_i^d) \in \mathbb{R}^d$ represent d different index values for the i^{th} journal; *e.g.*, x_1^1 and x_1^2 may respectively represent the H-index and the SJR for the first journal in the set $\mathcal{D}$. The goal is then to learn an aggregated index as:

$$f(\boldsymbol{x}) = \sum_{k=1}^{d} w^k x^k + b \tag{2.2}$$

where $\boldsymbol{w} = (w^1, \dots, w^d) \in \mathbb{R}^d$ is the d-dimensional vector of weights, each assigned to a different index, and b is a bias, to allow f to have a non-zero mean.[1] Different techniques can be exploited to learn $\boldsymbol{w}$ and b in the above scheme. For instance, one is the DEA model proposed by Tüselmann et al [2015], which learns a set of weights $\boldsymbol{w}$, while using a null bias b (see also Sect. 2.4.3). Furthermore, simple aggregation rules like CombSUM [Tsai, 2014] and Borda Count (specific for ranking) can be expressed in terms of this representation by assuming uniform weights (*i.e.*, $w^k = 1$, for $k = 1, \dots, d$), and normalized index values. In particular, for the CombSUM method, indices may be normalized using min-max or Z normalization, respectively as:

$$x_i'^k = \frac{x_i^k - \min_{j=1,\dots,n} x_j^k}{\max_{j=1,\dots,n} x_j^k - \min_{j=1,\dots,n} x_j^k}\,, \tag{2.3}$$

$$x_i'^k = \frac{x_i^k - \mu^k}{\sigma^k}\,, \tag{2.4}$$

[1] Note that, although the value of b is irrelevant when ranking journals according to $f(\boldsymbol{x})$, it may be helpful during the process of learning the weights $\boldsymbol{w}$, as the values of the considered indices do not typically have zero mean.

where $x_i'^k$ is the normalized value for the k^{th} index of the i^{th} journal, and μ^k and σ^k are the mean and standard deviation for the k^{th} index values of the journals in $\mathcal{D}$. For Borda Count, and similar ranking aggregation methods, I should consider as values of $x_i'^k$ the position of the i^{th} journal in the ranked list of the k^{th} index; in particular, if we are given $n = 100$ journals, and the i^{th} journal is ranked $r = 5^{\text{th}}$ using the k^{th} index, then $x_i'^k = n - r + 1 = 96$.

In general, to learn a linear combination function $f(\boldsymbol{x})$, *i.e.*, its parameters $\boldsymbol{w}$ and b, we are not restricted to the use of DEA or simple combination rules as the aforementioned ones. A set of different existing techniques proposed in the area of statistical data mining and machine learning can be exploited to this end [Bishop, 2007]. For instance, one may project the data $\mathcal{D}$ onto a reduced vector space using Principal Component Analysis (PCA), and consider as the weights $\boldsymbol{w}$ the values of the first component (eigenvector). This will capture the direction of the vector space along which data is maximally spread (*i.e.*, exhibiting the highest variance). PCA is an example of an *unsupervised* learning technique, as it projects data onto a subspace without exploiting any knowledge of a desired *target* value. Conversely, *supervised* learning techniques assume that, for each sample in $\mathcal{D}$, we are also given a target value y_i, and learn f by minimizing a functional of the form:

$$\min_{\boldsymbol{w},b} \frac{1}{n} \sum_{i=1}^{n} \ell\left(y_i, f(\boldsymbol{x}_i)\right) + \lambda\Omega(\boldsymbol{w}) \,, \tag{2.5}$$

where $\ell\left(y_i, f(\boldsymbol{x}_i)\right)$ is a loss function that penalizes values of $f(\boldsymbol{x}_i)$ which are different from the target value y_i, $\Omega(\boldsymbol{w})$ is a regularization term that penalizes high values of $\boldsymbol{w}$ to provide a more stable solution, and λ is a trade-off parameter. To be more concrete, let us give some examples. If we consider

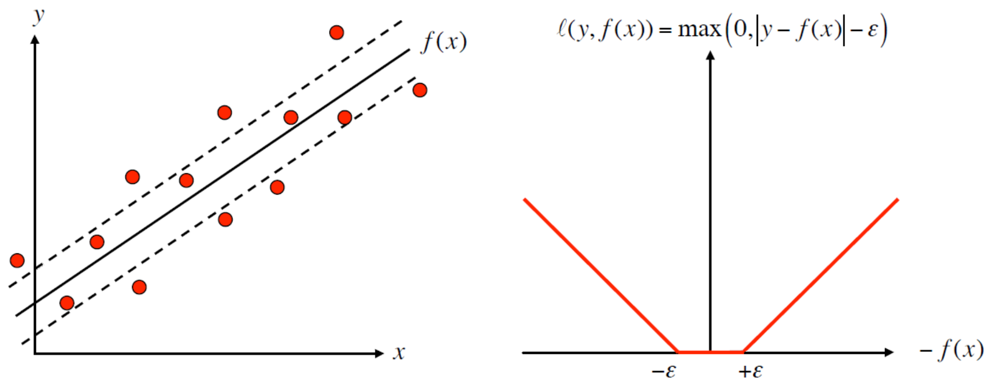

FIGURE 2.1: *Left:* SVR finds a linear function $f(\boldsymbol{x})$ that only penalizes y values outside of the tolerance band $[f(\boldsymbol{x}) - \varepsilon, f(\boldsymbol{x}) + \varepsilon]$ (dashed black lines). *Right:* The ε-insensitive loss.

$\ell\left(y_i, f(\boldsymbol{x}_i)\right) = (y_i - f(\boldsymbol{x}_i))^2$, without regularization, we yield the classical minimum mean square error (MMSE) linear regression problem. If we consider an additional regularization term $\Omega(\boldsymbol{w}) = \|\boldsymbol{w}\|_2^2 = \sum_{k=1}^{d}\left(w^k\right)^2$ (*i.e.*, the ℓ_2-norm of $\boldsymbol{w}$), and $\lambda > 0$, we yield *ridge* regression.

Support Vector Regression. Another very popular regression technique is Support Vector Regression (SVR) [Vapnik, 1995]. It minimizes a functional as that given in Eq. (2.5), where $\ell\left(y_i, f(\boldsymbol{x}_i)\right) = \max\left(0, |y_i - f(\boldsymbol{x}_i)| - \varepsilon\right)$ is the so-called ε-insensitive loss, and $\Omega(\boldsymbol{w}) = \|\boldsymbol{w}\|_2^2$. This essentially assigns a linear penalty to points for which y falls outside of a "tolerance" band $[f(\boldsymbol{x}) - \varepsilon, f(\boldsymbol{x}) + \varepsilon]$, as shown in Fig. 2.1. This technique can also be used to perform nonlinear regression tasks, by means of the so-called *kernel* trick, which allows one to write the function $f(\boldsymbol{x})$ as a linear combination of similarities (*i.e.*, kernel functions) computed between $\boldsymbol{x}$ and the so-called *support vectors* (*i.e.*, a subset of the training points in $\mathcal{D}$). This is why this technique is named support vector regression. It is however out of the scope of this work to provide further details about this technique, for which I refer the reader to Vapnik [1995] and Bishop [2007].

Defining the target values. In this case, assuming known values of y is equivalent to assuming that an *ideal* value of our aggregated index is already known for the journals in $\mathcal{D}$, which is clearly not the case (and it is indeed what I would like to achieve). However, I can somehow approximate the distribution of this value and make inference on that to learn $\boldsymbol{w}$ and b as discussed above, using the many supervised learning techniques that have already been proposed. To approximate the values of y, I leverage on a similar idea to that exploited by Bador and Lafouge [2010]. It amounts to producing a tied rank of journals by letting each single index vote whether a given journal should be in the first, second, third, or fourth quartile of the final distribution, and then summing up votes coming from the different available indices. Let us thus generally assume that the k^{th} index expresses a vote $q^k \in \{0, \ldots, m-1\}$, where $m-1$ denotes the highest ranking (*e.g.*, the first quartile would correspond to $q = 3$, as, for quartiles, $m = 4$). Then the value y_i for the i^{th} journal is simply given by $y_i = \sum_{k=1}^{d} q^k$. The corresponding y_i values will thus be in the set $\{0, \ldots, d(m-1)\}$, and several journals will have the same value of y, *i.e.*, the same rank in the final list. These ties can be broken to obtain an unambiguous final ranking by learning $\boldsymbol{w}$ and b using one of the aforementioned supervised learning techniques; *e.g.*, ridge regression and SVR. The regularization parameter λ in Eq. (2.5) is often estimated through a k-fold cross validation on the available data $\mathcal{D}$, to optimize performance while minimizing the risk of *overfitting*, *i.e.*, of learning functions that predict the training data with almost no error, but do not properly generalize on unseen data. This has to be especially accounted for in high-dimensional spaces, and when learning nonlinear functions. In this case, I exploit cross-validation to tune the parameter λ by testing different values and retaining the one that minimizes the mean absolute error. I then retain the score

$f(\boldsymbol{x})$ given by our method trained with the best value of λ to the points in the validation fold.

2.6 Meta-Ranking of Business and Management Journals

In this section I apply this analysis of journal meta-ranking using the two most important citation databases, *i.e.*, Thomson Reuters JCR and the Scopus SJR. I have selected two field of study: Business and Management. In building the datasets, I have decided to use the 5-year Impact Factor and the H-index to capture the stability, IF and SJR to capture the current trend. This mix of indices seems to be appropriate because SJR is more close to the size dimension than the IF, while H-index attempts to capture both dimensions [Hodge and Lacasse, 2010, Leydesdorff, 2009]. Again, according to previous studies, the H-index correlates highly with Thomson Reuters 5-year impact factors and its scores are similar to the experts' opinion [Hodge and Lacasse, 2010]. As already stated, SJR contains a large number of journals compared to JCR, hence to have consistent data, I have decided to use the Thomson dataset as "master". In particular, from that report, I have extracted IF and 5-year IF indices, whereas I have extracted the SJR and H-index from Scopus. In total, we have respectively $n = 173$ journals in Management and $n = 111$ in Business, corresponding to the whole journals indexed in Thomson Reuters for both subject areas. For both cases, I denote with $\mathcal{D} = \{\boldsymbol{x}_i\}_{i=1}^n$ the retrieved set of n journals, where each journal $\boldsymbol{x}_i$ is a four-dimensional vector characterized by the four index values IF, 5-year IF, SJR, and H-index.

Setup. For both journals in Management and Business, I have aggregated the four baseline indices into the following meta-indices: SVR, CombSUM, Borda Count and PCA. To define the ground-truth labels y required to train the SVR, I have used quartile-based voting for each baseline index, *i.e.*, I have set $m = 4$ (see Sect 2.5). The regularization parameter $\lambda \in \frac{1}{n}\{10^{-3}, 10^{-2}, \dots, 10^{3}\}$ of the SVR has been selected using a 5-fold cross-validation to maximize the mean absolute error on the validation fold.

Results. In Figs. 2.2 and 2.3, I show Pearson's correlations between each meta-index and each of the baseline indices. They are clearly highly correlated with each other, with correlation values higher than 0.9 almost for all pairs, except for those involving the Borda Count. The reason is that Borda Count is the only technique that does not exploit the numeric value of each of the baseline indices, but directly combines their rankings. Despite the high correlation between baseline and meta-indices, the corresponding rankings may exhibit significant variations, as one may appreciate from Tables 2.1 and 2.2, where I have considered the top 20 journals according to the SVR index, and how they are ranked by the other indices.

I finally consider a different projection to visualize each index in a compact, two-dimensional vector space, and evaluate again how similar they are to each other, similarly to the procedure adopted by Leydesdorff [2009]. In particular, I apply again PCA, but this time considering each index as a point, and the values

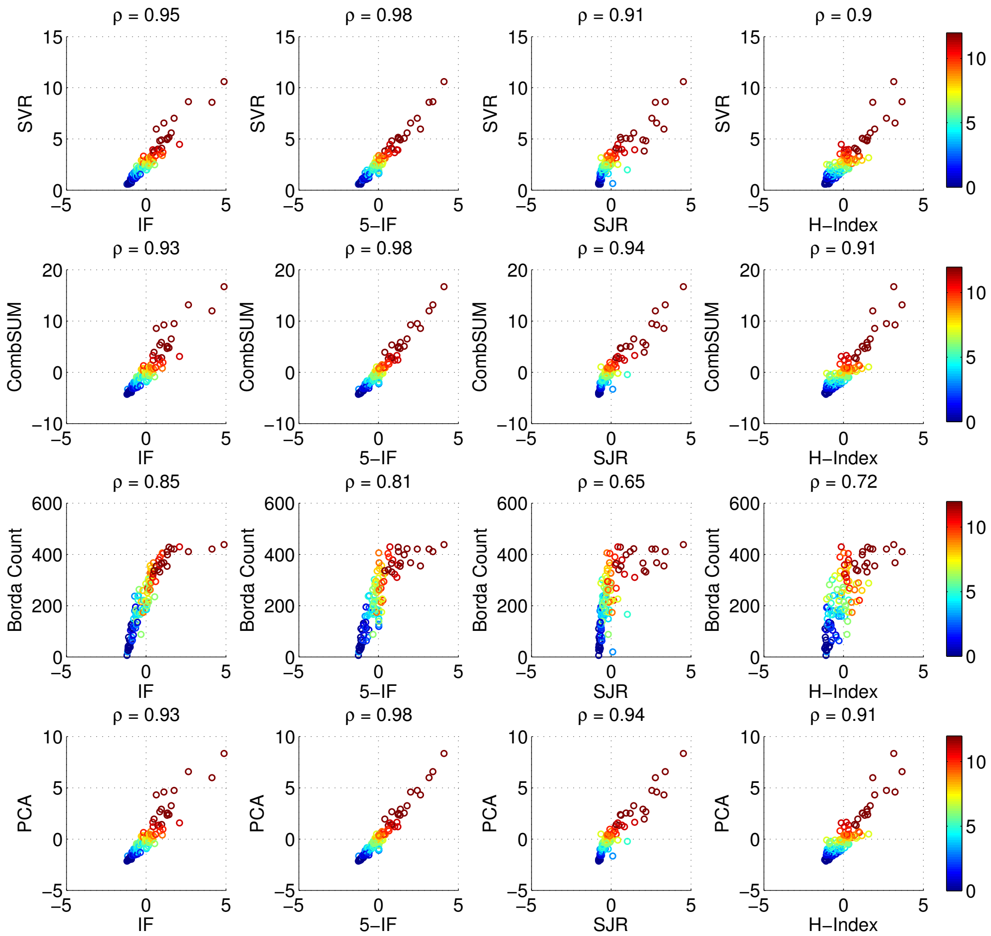

FIGURE 2.2: Pearson's correlations between each meta-index (rows) and each baseline index (columns) for journals in the Business area. Each point in the scatter plots represents a distinct journal, and its color denotes the corresponding ground-truth value y.

it assigns to each journal as its dimensions. The first two principal components of this projection are shown in Fig. 2.4, where it can be appreciated how almost all meta-indices (except for Borda Count) are close to each other and well-summarize the characteristics of the four combined baseline indices. Pearson's correlation and Euclidean distance values (computed in the non-reduced space, using all n journals as dimensions) are also reported in Table 2.3 for the sake of completeness.

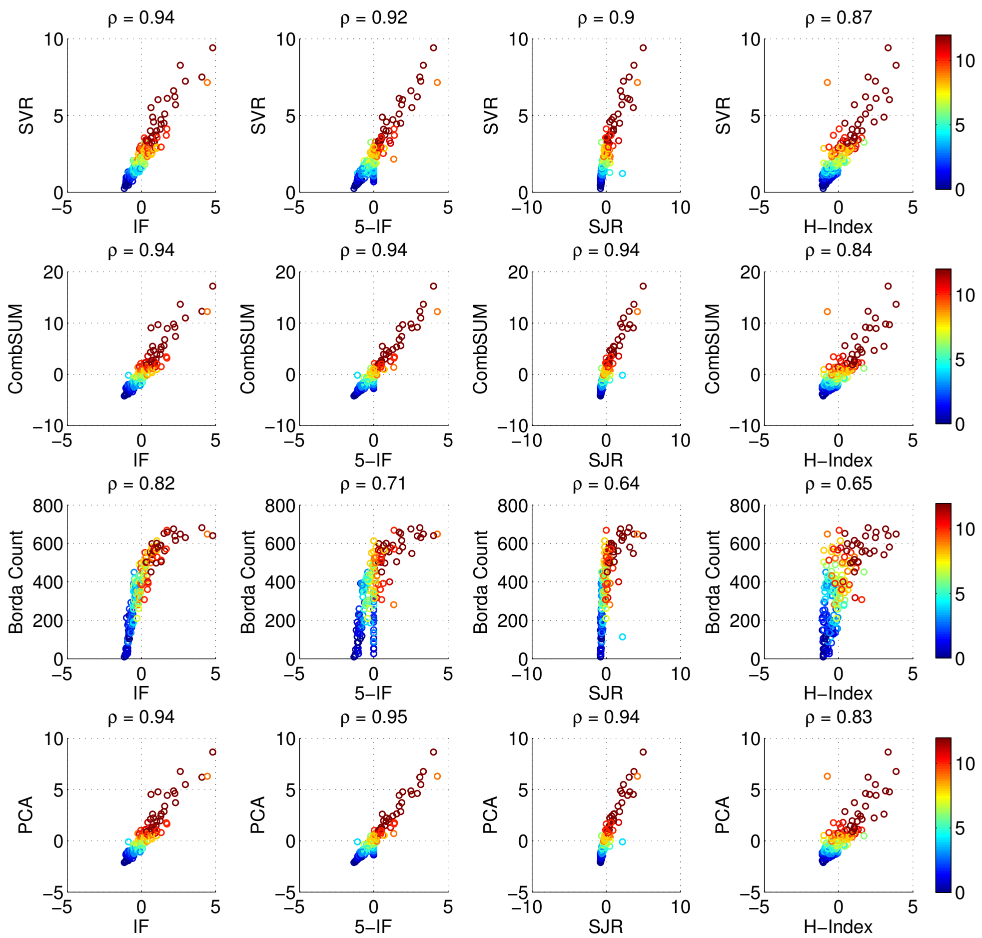

FIGURE 2.3: Pearson's correlations between each meta-index (rows) and each baseline index (columns) for journals in the Management area. See the caption of Fig. 2.2 for further details.

Journal / Index	IF	5-IF	SJR	H	SVR	CS	BC	PCA
Academy of Management Review	1	1	1	3	1	1	1	1
Academy of Management Journal	3	2	2	1	2	2	7	2
Journal of Management	2	3	5	6	3	3	5	3
Journal of Marketing	5	5	6	4	4	4	6	4
Strategic Management Journal	11	6	4	2	5	5	14	5
Administrative Science Quarterly	21	4	3	5	6	6	20	6
Journal of Int'l Business Studies	6	7	11	7	7	7	4	7
J. Academy of Marketing Science	7	12	18	9	8	12	3	12
Journal of Management Studies	8	8	14	14	9	10	11	10
Journal of Business Venturing	9	11	12	13	10	11	8	11
Journal of Organizational Behavior	10	10	17	12	11	13	21	13
Journal of Consumer Research	15	9	9	10	12	8	16	8
Journal of Marketing Research	18	17	7	11	13	9	26	9
Family Business Review	4	19	21	54	14	16	2	16
Academy of Management Perspectives	14	18	23	22	15	17	17	17
J. Env. Economics and Management	20	21	19	20	16	19	23	19
Entrepreneurship Theory and Practice	19	16	16	29	17	18	13	18
Long Range Planning	27	14	10	39	18	15	31	15
Int'l J. Management Reviews	17	13	24	46	19	20	18	20
Marketing Science	24	24	8	18	20	14	24	14

TABLE 2.1: Differences in ranking and meta-ranking for journals in the Business area. Journals are sorted here according to the SVR ranking. The corresponding rank according to each other index is reported in the corresponding column. IF and 5-IF stand for Impact Factor and 5-year IF, H for H-index, CS for CombSUM, and BC for Borda Count.

Journal / Index	IF	5-IF	SJR	H	SVR	CS	BC	PCA
Academy of Management Review	1	2	1	3	1	1	8	1
Academy of Management Journal	5	3	3	1	2	2	7	2
Journal of Management	3	5	6	10	3	3	1	4
MIS Quarterly: Manag. Inf. Systems	4	4	8	7	4	5	10	5
Academy of Management Annals	2	1	2	131*	5	4	6	3
Journal of Applied Psychology	8	8	12	4	6	6	2	6
Strategic Management Journal	18	9	5	2	7	7	17	7
Journal of Operations Management	7	6	10	11	8	8	9	8
Organization Science	9	13	7	6	9	10	4	10
Administrative Science Quarterly	34	7	4	9	10	9	30	9
Personnel Psychology	6	10	11	22	11	11	11	11
Journal of Int'l Business Studies	12	12	15	12	12	12	23	12
Management Science	29	28	17	5	13	14	31	15
Journal of Organizational Behavior	16	18	21	16	14	16	15	16
Organizational Research Methods	13	11	13	40	15	13	5	13
Journal of Management Studies	14	14	16	18	16	15	38	14
Research Policy	27	22	26	8	17	17	25	17
Omega	17	26	19	27	18	19	16	19
Org. Behavior and Human Dec. Proc.	20	23	22	19	19	20	13	20
Information Systems Research	37	21	18	14	20	18	27	18

TABLE 2.2: Differences in ranking and meta-ranking for journals in the Management area. See caption of Tab. 2.1 for further details. *This journal has low rank for the H-index, as its data is available only from 2011.

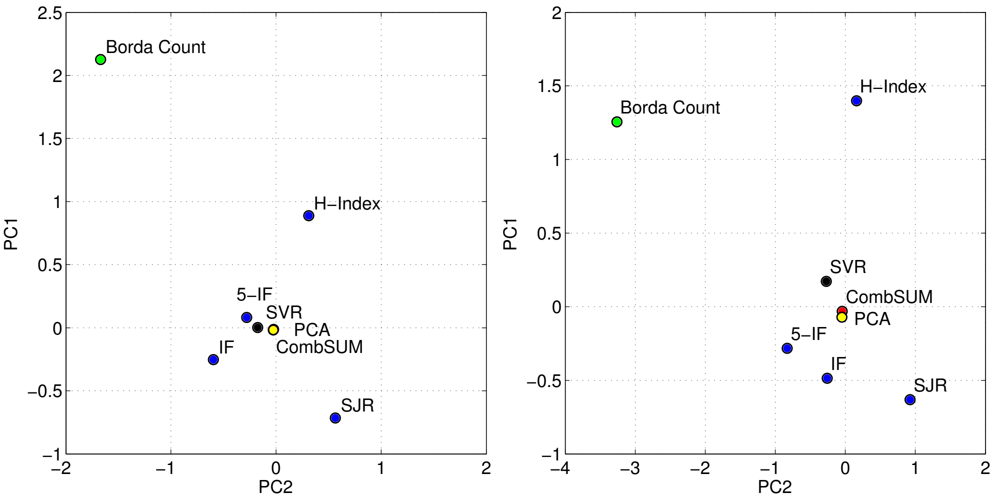

Figure 2.4: PCA-based projection on a two-dimensional space for Business (left) and Management (right) journals. Each point represents an index in the space of the first two principal components. Note that PCA and CombSUM meta-indices are overlapped in the left plot.

	IF	5-IF	SJR	H-Index	SVR	CombSUM	Borda Count	PCA
IF	0.00 (1.00)	0.89 (0.91)	1.28 (0.81)	1.47 (0.77)	0.57 (0.95)	0.70 (0.93)	3.64 (0.85)	0.70 (0.93)
5-IF		0.00 (1.00)	1.25 (0.90)	1.16 (0.85)	0.52 (0.98)	0.53 (0.98)	3.32 (0.81)	0.52 (0.98)
SJR			0.00 (1.00)	1.63 (0.81)	1.04 (0.91)	0.92 (0.94)	4.34 (0.65)	0.92 (0.94)
H-Index				0.00 (1.00)	1.02 (0.90)	0.97 (0.91)	3.39 (0.72)	0.98 (0.91)
SVR					0.00 (1.00)	0.19 (1.00)	3.56 (0.83)	0.19 (1.00)
CombSUM						0.00 (1.00)	3.59 (0.81)	0.01 (1.00)
Borda Count							0.00 (1.00)	3.59 (0.81)
PCA								0.00 (1.00)

IF	0.00 (1.00)	1.20 (0.90)	1.45 (0.85)	2.02 (0.68)	0.77 (0.94)	0.80 (0.94)	4.61 (0.82)	0.78 (0.94)
5-IF		0.00 (1.00)	1.80 (0.87)	2.00 (0.69)	0.99 (0.92)	0.93 (0.94)	4.29 (0.71)	0.91 (0.95)
SJR			0.00 (1.00)	2.18 (0.72)	1.41 (0.90)	1.16 (0.94)	5.55 (0.64)	1.14 (0.94)
H-Index				0.00 (1.00)	1.32 (0.87)	1.45 (0.84)	4.60 (0.65)	1.49 (0.83)
SVR					0.00 (1.00)	0.31 (0.99)	4.45 (0.79)	0.33 (0.99)
CombSUM						0.00 (1.00)	4.61 (0.77)	0.04 (1.00)
Borda Count							0.00 (1.00)	4.62 (0.77)
PCA								0.00 (1.00)

Table 2.3: Euclidean distance and Pearson's correlation (in parentheses) between each pair of indices for Business (top) and Management (bottom) journals.

2.7 Discussion and Conclusions

Producing a reliable, widely-approved journal ranking is a non-trivial task, mainly due to the inherent difficulty of selecting a proper set of baseline indices and combination technique (*i.e.*, meta-index) among the existing ones. In this paper, I have highlighted the need of combining various indices, by leveraging on the main limitations emerged from previous work. I have proposed and formalized an approach able to obtain a tied rank of journals, and to capture the different dimensions characterizing the aspects of research *quality* in a consistent manner. Firstly, I have determined if a given journal should be in the first, second, third, or fourth quartile of the final distribution, according to each index. Secondly, using different techniques, I have aggregated the votes coming from the different baseline indices. Finally, I have sorted journals for both Business and Management area, according to the considered indices and meta-indices. In order to complete the analysis, I have also evaluated the performance of each meta-index, finding a high correlation between indices and meta-indices. Moreover, I have evaluated their distance in a two-dimensional vector space to visualize how similar they are to each other. This analysis has shown that supervised and unsupervised learning techniques (in particular, PCA and SVR), CombSUM and, partially, also Borda Count, are all qualified tools to produce aggregate indices for journal ranking.

Although I have chosen the combined indices according to a well-motivated principle to balance the contribution of stability and of the current trend, it is still an open issue to quantitatively evaluate how and to what extent the proposed meta-indices can be retained properly representative of the aforementioned aspects. To this end, I envision the possibility of combining and comparing the proposed technique with qualitative approaches (*e.g.*, based on the analysis of experts' opinions). This can be definitely considered a promising research direction. As previous work has been mainly focused on defining novel combination methods to aggregate a set of given base indices, there is need of shedding more light on how to select a proper set of indices to be combined, also taking into account the given combination method. This is another relevant research direction that may be worth investigating in the future. I finally believe that this framework can provide useful results for many purposes, *e.g.*, for researchers, as a reference to choose their publication outlets, and for faculties, departments and editors to evaluate and compare the quality of their own journals.

Chapter 3

Features of top-rated gold open access journals: An analysis of the scopus database

3.1 Abstract

The goal is to identify the features of top-rated gold open access (OA) journals by testing seven main variables: languages, countries, years of activity and years in the DOAJ repository, publication fee, the field of study, whether the journal has been launched as OA or converted, and the type of publisher. A sample of 1,910 gold OA journals has been obtained by combining Scopus SJR 2012, the DOAJ, and data provided by previous studies [Solomon, 2013]. I have divided the SJR index into quartiles for all journals' subject areas. First, I show descriptive statistics by combining quartiles based on their features. Then, after having converted the quartiles into a dummy variable, I test it as a dependent variable and in a binary logistic regression. This work contributes empirically to better understanding the gold OA efficacy, which may be helpful in improving journals' rankings in the areas where this is still a struggle. Significant results have been found for all variables, except for the types of publishers, and for born or converted journals.

3.2 Introduction

Research quality has always played a crucial role for scholars, publishers, professional societies, and funding organizations [Falagas et al, 2008]. Authors compete for the opportunity to publish their research in high-quality and highly ranked journals in order to gain the largest diffusion possible. Research impact is the degree to which findings are read, used, applied, built-upon, and cited by users in their own further research and applications; again, it is a measure of the progress and productivity of studies [Harnad et al, 2004]. Understanding journals' performances in terms of the impacts they have is a significant challenge; bibliometrics works in this field, developing and studying indices and indicators and providing statistics of various types. In the last decade, open access (OA) has become an established and well-known phenomenon, and the number of journals and articles released in OA has grown rapidly. Although OA has not changed how research is conducted [Pinfield, 2005], it has upset the rules of publishing scholarly articles. A great amount of literature has attempted to measure OA's success and efficacy by looking at its metrics since the origin of OA [McVeigh, 2004]. In particular, previous works have studied some features (see the Tables below for a detailed list) of gold OA (where the publisher provides free online access) journals' rankings, which are intended as an impact factor or other similar indicator, in comparison to other forms of OA and to traditional publishing methods. Of course, these studies have helped to understand the OA phenomenon, but they present two limitations: First, single variables are combined with the journals' rankings and show, for most cases, only the descriptive statistics. Second, each work usually focuses on a small set of variables to achieve its purpose; therefore, understanding the features of top journals as a whole is not easy because they use different data-sets built in different years.

The aim of this paper is to understand what the relevant features of top gold OA journals are, and, in doing so, I use several descriptive statistics and a binary logistic regression model. With this type of regression method, on the one hand, it is possible to classify the journal features by their relationship with the ranking; yet, on the other hand, I can identify the features that a journal should have to raise its ranking. The main motivation is related to the fact that a high ranking can be interpreted as a success factor for a journal; therefore, understanding what is a determinant for reaching success can contribute to the improvement of OA efficacy in the field of studies where OA is still struggling.

I have built a sample of 1,910 gold OA journals, and I have used the journals' ranking as a dependent variable. In doing so, I have divided gold OA journals by the SJR index for each subject category. I have considered top journals (Q1) as those where the relative location is in the top 25% of the SJR distribution. I have decided to combine three databases; the main regard is Scopus because it contains a larger number of gold OA journals with an impact factor larger than other databases. It also includes titles from more countries that are published in a greater variety of languages [Leydesdorff et al, 2010]. Again, I have used the Directory of Open Access Journals (DOAJ), repository data, and other variables' information from previous studies' datasets [Solomon, 2013]. This paper is divided into five sections: in the next, I provide a literature review; in the third, I present the material and methods used; in the fourth, I present variables with some descriptive statistics and comparisons to other works; in the fifth, I discuss results of regression; and, in the sixth and last part, I present conclusions.

3.3 Background

A scientific publication represents the final stage of many months and sometimes years of meticulous planning, execution, and analyses of hundreds of experiments [Benos et al, 2005]. With the advent of the Internet, more and more researchers are making their research openly accessible by self-archiving it online to increase their visibility, usage, and citation impact. The publication of scientific content has been one of the areas to benefit most from the emergence of the Internet [Björk, 2004]. OA had its first formal definition and guidelines in 2002, with the Budapest Open Access Initiative, followed by the Bethesda statement on open access publishing in 2003 and the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities in 2003, which is usually known as the BBB definition. OA can be defined as free and unrestricted access on the public Internet to literature that scholars provide without expectation of direct payment [Prosser, 2003]. According to these definitions, publishing in OA means satisfying two conditions: The first is granting unrestricted access to anyone via the internet and the license to copy, use, distribute for non-commercial purposes, and make and distribute derivative works without any payments or restrictions. The aim is to remove barriers to literature in order to accelerate research, enrich education, and share learning. The second condition is to deposit a complete version of the work and all supplemental materials immediately upon initial publication in at least one online repository, including a copy of the permission in a suitable standard electronic format. Repositories consist of a physical space reserved for permanent or intermediate storage of archival material and can be searched and retrieved for later use [Hayes, 2005]. Actually, the DOAJ, which is managed by Lund University, is the largest repository, including scientific journals presenting quality controls, that allows free access. OA occurs in two variants [Harnad et al, 2004]: Green OA refers to publishing in any appropriate traditional journal, in addition to self-archiving the pre- or post-print paper in a repository. Gold OA refers to articles in fully accessible OA journals. The gold model uses a traditional journal publication system, and nothing is paid by the reader of a peer-reviewed article. Some journals require a fee, paid by the author's organization or the research funder and sometimes by the author. According to the DOAJ repository, in November 2013, there were 6,573 journals that did not require a processing charge and 2,652 that required a processing charge.

Regarding the benefits of the OA articles in terms of citations, there is not agreement in the literature. Early studies have claimed that OA articles are cited more often and published in less time in comparison to traditional publications. In particular, Lawrence [2001] reports that OA articles in computer science are cited more, and multidisciplinary works between diverse fields of study at varying stages of adoption of OA have confirmed that OA articles have a greater research impact than articles that are not freely available [Antelman, 2004]. In a longitudinal study of a cohort of OA and non-OA articles with an article-level approach, it has been demonstrated that there is direct and strong evidence for preferential or earlier citation of articles published originally as OA [Eysenbach, 2006]. However, some authors are critical about the cause and effect relationship between OA and higher citations, stating that the benefits of self-archiving may be uncertain and may vary between different fields of study [Craig et al, 2007]. Again, Moed [2007] has highlighted two points: the first is a self-selection bias, *i.e.*, authors tend to self-archive high-quality articles and thus receive more citations. The second is that many works do not take into consideration a wide time windows to evaluate the benefits of OA in comparison to traditional publishing methods. As can be noted in Table 3.1, during the last decade, several scholars have measured the impact of gold OA journals and have analyzed the main characteristics of them in comparison to green OA journals and to traditional journals. The common strategy has been to combine two or more databases. The first published study is a decade old and uses the Thomson Reuters database [McVeigh, 2004]. The results highlight that top OA journals are not equally distributed between different fields of study, with a prevalence of Physics, Engineering and Mathematics. McVeigh [2004] has also noted that over 55% of journals allow self-archiving; in regard to the geographical distribution, over one-third of OA journals were published in Asia-Pacific, while North America and Western Europe account for approximately 40% of OA titles. Again, she has shown that the overall mean percentile rank in terms of Journal Impact Factor was 39.8 percentile, while two-thirds of the journals were below the 50th percentile in rank. The mean percentile rank of OA journals by Immediacy Index was the 46 percentile. As a group, journals that have adopted an OA distribution model have not achieved a significantly greater citation impact. However, individual OA journals have been appearing among the highest ranked journals, even within a few years of their launch. More recently, Giglia [2010] has found a low presence of OA journals in JCR 2008, and she has confirmed, with some different results, that there are strong differences between disciplinary areas and impact, considering the best performances: in Medicine, there is a strong presence in the top twenty percentiles, 15.96%; 14.42% in Life Sciences; 12.63% in Mathematics, Physics, and Engineering; and 4.66% in Chemistry. Again, she has shown that there are many titles that rank low by Impact Factor but high by Immediacy Index. Regarding geographical aspects, she has also confirmed that nearly 71% of OA in JCR 2008 Science edition have come from Central and South America. Finally, she has stated a direct causal relationship between age, visibility and prestige, in terms of citations that cannot be straightforwardly inferred.

Subsequently, Miguel et al [2011] confirmed previous studies, but, regarding the impact, they found that, "*For the most part, belong to the fourth quartile regardless of the geographic area of origin*". In contrast, Gumpenberger et al [2013] have shown that gold OA journals' IF is increasing, and one-third of newly launched journals are indexed in JCR after a year. Again, in that work, they have shown a percentage lower than 20% of the journals in the first quartile and a concentration of 80% of the top gold OA journals in the UK and in the USA. Björk and Solomon [2012] have introduced another element to investigate: a comparison between OA journals that require publication fees (APCs) and journals that do not require payment. The result was that a funding mechanism is not related to the journals' quality. In the same paper, they have also shown that 70% of subscription journals are owned by organizations located in the four major publishing companies (the USA, the UK, the Netherlands and Germany). These results have been confirmed in the work of Solomon et al [2013], in which the Source-Normalized Impact per Paper Version 2 (SNIP2) for OA journals was combined with and without APCs during the period from 1999 through 2010. They have found that, regardless of the business model, articles are cited at a similar rate to subscription journals. Again, converted OA journals have a SNIP2 lower than born OA journals, although some differences exist between health science and other fields of study. Again, they have found that a high percentage of converted journals are not published in the English language. Finally, Solomon [2013] has introduced a classification between types of publishers (APCs) and platforms. Some results state that over one-third of the journals and 42% of the articles are owned by professional publishers, and the APCs are closely related to them. Moreover, in the same work, he has confirmed that a great number of OA journals across all disciplines are published outside the four major publishing countries. In Table 3.2, I have summarized, in the first column, author(s) and, in the other columns, variables studied: ranking field of study, geographical distribution, journals' age and publication fee are the most recurrent. In the following paragraphs, I present the sample and the variables and compare data to previous studies.

Author(s)	Purpose	Data source	Method	Variables	Principal Findings
McVeigh 2004	Overview of OA journals in Thomson Reuters database.	ISI Web of Knowledge DOAJ J-STAGE ScIELO	Descriptive statistics and graphics	Journals by subject Regions Impact Factor, Immediacy Index, Percentile Rank Journals' age	High-ranking OA journals are not evenly distributed among the subject fields. Geographical distribution differs strongly from the ISI citation databases as a whole. Two-thirds of the journals are below the 50th percentile in rank. The average number of years in OA varies a bit between fields. OA publishing is growing slowly, both by creation of new titles and by conversion.
Giglia 2010	Test the performance of OA journals with the most traditional bibliometric indicators.	JCR DOAJ	Descriptive statistics and graphics	Geographical distribution Macro disciplinary areas, Impact Factor, percentile, Immediacy index, 5-year Impact Factor Journals' age	Low presence of OA journals in JCR 2008. Nearly 71% of OA in JCR 2008 Science edition comes from Central and South America. IF performance is 38.62% in the 0-50 percentiles. The direct causal relationship between age and visibility and prestige in terms of citations cannot be straightforwardly inferred.
Miguel, Chinchilla-Rodriguez, & de Moya-Anegón 2011	H1. There is a stronger gold road in the social sciences and humanities. H2. There is a greater proportion of gold road journals in emerging economic regions. H3: These journals are not the most relevant and are, for the most part, from the last quartile.	SCOPUS DOAJ Sherpa-RoMEO	Descriptive statistics, graphics and Chi-square test	Gold, green and no OA Main Field of Study SJR Index Geographic region	The gold road has a greater proportion of work in journals belonging to Medicine; Biochemistry, Genetics and Molecular Biology; Areas Related to Medicine; and Earth and Environmental Sciences. The peripheral and emerging regions have greater proportions of gold road texts, and publications in green journals are almost nonexistent. The gold road journals have no visibility and, for the most part, belong to the fourth quartile regardless of the geographic area of origin.
Björk and Solomon 2012	Comparing the scientific impact of OA journals with subscription journals, controlling for journal age, the country of the publisher, discipline and their business model.	Combining Datasets: Ulrichsweb, SJR, DOAJ	Descriptive statistics and graphics	SJR, Countries, Journal launch period, Publication Fee	Seventy percent of subscription journals are from publishers in the four major publishing countries. Newer journals, particularly in medicine and health, are performing at about the same level as subscription journals. The funding mechanism of a journal is irrelevant in considering its quality. Almost half of OA journals started before 1996 and were not published in top publisher countries.
Gumpenberger, Ovalle-Perandones and Gorraiz	1. To identify the number of Gold Open Access journals	Ulrichsweb DOAJ	Descriptive statistics,	IF, SJR, SNIP, Countries,	The impact of top Gold OA journals is generally increasing for all of the analyzed indicators. One-third of the newly

2013	that have successfully taken the hurdle to be indexed in JCR and allocated to disciplines, countries and quartiles. 2. To analyze the temporal evolution of the gold OA journals' impact.	JCR SJR CWTS	graphics, and Linear regression Correlation	Timeline, Publisher distribution	launched OA titles were already indexed in JCR after 1 year, and 80 % of these received an IF at least within a 5-year interval. The percentage of Q1 titles is < 20 % in the UK, and the USA contributes to 80 % of the top Gold OA. Gold OA is still small compared to the total number of scholarly journals worldwide. Q1 titles are predominantly journals in English.
Solomon, Laakso, Björk 2013	1) Document the growth of OA journals and articles between 1999 and 2010. 2) Compare SNIP2 Citation Averages for OA journals with that of subscription journals during the period of 1999 through 2010.	SJR and DOAJ	Descriptive statistics and graphics	SNIP2, Journal Born or converted as OA, Publication fee, Sciences (health non health) Language of converted journals	High quality OA publishing is growing at a rapid rate and in the case of OA publishing, funded by APCs. There is evidence based on citation rates that these journals are on par with subscription journals. The bulk of converted OA journals are located outside the four major publishing countries. A high percentage of the journals that are converted to OA are also in languages other than English.
Solomon 2013	Types of organizations and their characteristics of OA journals' publishers in Scopus.	SJR and DOAJ	Descriptive statistics and graphics	Publisher Type, Discipline, Publication Fee	Professional, society, and university publishers own approximately 85% of the journals and articles published. Over 80% of the journals published by professional publishers charge APCs. There appears to be no clear delineation between funding models and types of publishers. The types of organizations publishing OA journals differed significantly across disciplines.

Table 3.1. Highlights of previous studies

	Ranking	Subject	Geographical distribution	Journals' age	Born/Converted as OA	Publication Fee	Type of Publisher	Language	Total
McVeigh	x	x	x	x					**4**
Giglia	x	x	x	x					**4**
Miguel et al	x	x	x						**3**
Gumpenberger	x	x	x	x				x	**5**
Björk & Solomon	x		x	x		x			**4**
Solomon et al	x	x			x	x		x	**5**
Solomon		x				x	x		**4**
Total	**6**	**6**	**5**	**4**	**1**	**3**	**1**	**2**	

Table 3.2. Features studied by previous works

3.4 Material and Method

The dataset consists of 1,910 gold OA journals. In building it, the first step was to decide how to compare journals' rankings. In fact, the problem concerning indicators to evaluate the research is longstanding; several metrics exist to measure the ranking of a journal, and the most known is the Thomson Reuters IF. It has been criticized for multiple reasons [Bornmann et al, 2012, Seglen, 1997]; the most critical reasons for our study are as follows: (*i*) small research fields tend to lack journals with high impact; (*ii*) citation rates of articles determine journals' impacts but not *vice versa*; (*iii*) IF is a function of the number of references per article in the research field; and (*iv*) journal impact factors are not statistically representative of individual journal articles. In other words, different scientific areas, fields and micro-fields of study have different citation habits [Lancho-Barrantes et al, 2010].

Recently, the Scopus SJR indicator began an alternative to Thomson Reuters IF; it also takes into account the "quality" of citations received by a journal, whereas the journal IF considers incoming citations only in a quantitative manner. Thus, the use of the SJR indicator allows for the estimation of a journal's impact, reducing the influence of self-citations such that they cannot be more than 33% of the total because prestige can be transferred to a journal by all other journals, but not by itself [González-Pereira et al, 2010]. Nonetheless, the introduction of the SJR indicator does not bring about radical changes in this regard. It has been shown that half of the journals in the IF top 100 journal list are placed within a reasonable range of 32 ranking places in the SJR indicator journal list [Falagas et al, 2008]. A common workaround to overcome index limitations is to consider each journal in its field, determine the quartile or the percentile, and then compare them with all of the journals. Using this strategy, it is possible to benefit from the advantage of working with categorical variables. Other scholars have preferred to use a normalized impact factor, the SNIP index; however, in the literature, there are many criticisms about that indicator [Lee and Shin, 2014, Leydesdorff and Opthof, 2010, Leydesdorff et al, 2013]. Having said this, I have used three sources. The first is the Scopus database; the total number of journals in Scopus is 20,544. SJR divides sciences into 310 subject areas and 27 subject fields. I have taken the database of all journals by subject area, and then I have merged the 307 datasheets (three were empty). In doing so, I have obtained the countries' variables and the quartile division. It is frequent that a journal is indexed in multiple subject areas and categories. To have manageable data, I have maintained individual journals in the highest quartile per subject area; for example, if the journal is in the second quartile in Nursing and in the third quartile in Health Professions, I have kept only Nursing; if it is in the same quartile, I have kept both. After that, I have merged the datasheet with the DOAJ dataset to identify gold OA and to obtain the following variables: publication fee, languages, age of the journals and the period they were added into the DOAJ repository. The third source is Solomon's database [Solomon, 2013], which is used to obtain the publisher type classification and to determine if the single journal has been born as OA or if it has been converted. Finally, I have cleaned the data, deleting a total of 100 observations because 45 journals have ended their activity before the end of 2012 and 55 were without quartile information. It is also important to note that this study covers all active gold OA journals in Scopus and over 20.70% of OA journals in the DOAJ.

3.5 Variables

3.5.1 Dependent Variable

Ranking: The total number of journals in the first quartile is 288, the number in the second quartile is 503, the number in the third quartile is 652, and the number in the fourth quartile 467. I have converted the journals' quartile variable into a dichotomic variable called ranking, assigning 0 for quartiles 2, 3 and 4 and 1 for quartile 1. Hence, the value 0 is for 1,622, and the value 1 is for 288 journals. Only approximately 15% of the journals are in the first quartile. Regarding the first quartile, similar results have been provided by Gumpenberger et al [2013], in contrast to what was stated by Miguel et al [2011].

3.5.2 Independent Variables

Below, all of the independent variables used for the model are presented; descriptive statistics are shown, where possible, with all quartile information in order to give additional advice.

3.5.3 Language

Language: The publishing language has been taken into consideration Solomon [2013] in reference to converted journals and by Gumpenberger et al [2013] in reference to Q1 journals. In agreement with Lobachev [2008], I think that understanding the diversity of the information universe represents an important point for determining current trends in global information production. Hence, I want to investigate these variables to find additional insight. Almost all of the titles in the sample, 1,718 out of 1,910 and 284 out of 288 regarding the top-ranked journals, are in English or in English plus some other languages, and 501 journals publish in more than one language. The top languages in overall worldwide scholarly production are English, German, Chinese, Spanish and French [Lobachev, 2008]. It is interesting to see that Chinese does not appear in the top positions. Conversely, Portuguese is not only at the third position but also has 13 journals in the first quartile.

	English	Multilingual	Spanish	Portuguese	French	German
1^{st} Quartile	284	24	10	13	8	6
2^{nd} Quartile	488	108	49	53	16	6
3^{rd} Quartile	579	192	132	94	30	16
4^{th} Quartile	367	177	149	62	25	21
Total	1718	501	340	222	79	49

Table 3.3: Top publishing languages.

If we consider non-multilingual journals, English not only dominates, but, as seen in Table 3.4 other languages have a marginal role; just two journals in Portuguese are in the first quartile.

3.5.4 Country

Country: Over 49% of the total number of journals and over 74% of the top ranked journals are concentrated in six countries. While the UK and the USA

	English only	Spanish only	Portuguese only	French only	German only
1^{st} Quartile	262	0	2	0	0
2^{nd} Quartile	381	8	5	0	0
3^{rd} Quartile	392	30	19	0	3
4^{th} Quartile	207	52	7	1	8
Total	1242	90	33	1	11

Table 3.4: Top Publishing Languages in Non-Multilingual Journals

own 38.19% and 22.92% of the total number of top ranked journals (288), respectively, others hold only a small percentage. Geographical distribution has been studied by McVeigh [2004], Giglia [2010] and Miguel et al [2011], but only statistics related to the continents have been shown. Björk and Solomon [2012] have analyzed the four major traditional publishers (the US, the UK, the Netherlands and Germany) versus others. The only paper showing statistics per country was produced by Gumpenberger et al [2013]; they compared the number of journals in Q1 in Ulrichsweb and in the DOAJ. With close results in both databases, they found the US to be in first place, the UK in second, Germany in third, Switzerland and Japan in fourth, and Canada in fifth. According to the DOAJ, the top publishing OA countries, in terms of the number of journals, are the United States with 12.37%, Brazil with 9.35%, the UK with 6.29%, India with 6.08%, and Spain with 5.31%. In the table below are the top publishing countries according to the sample:

In Table 3.6, the top 10 publishing countries according to the Scopus database are compared, and the percentage denotes the total number of journals present on 12/31/2012 (the entire database contains 20,554 entries). In comparison to those results, we can see the top 10 gold OA publishing countries according to our sample and the DOAJ repositories. Both in our sample and in the DOAJ

	USA	UK	Brazil	India	Spain	Japan	New Zealand
1^{st} Quartile	66	110	12	6	5	7	7
2^{nd} Quartile	71	80	58	35	22	23	21
3^{rd} Quartile	76	20	79	44	37	22	25
4^{th} Quartile	50	9	35	10	24	16	13
Total	263	219	184	95	88	68	66

Table 3.5: Top Publishing Countries

repository, the Netherlands and Germany, which are known as large publishers, are not in the first 10 positions. If we compare results from our sample to the DOAJ, we can say that the top five publishing countries are very close.

Scopus	Journals	%	Sample	Journals	%	DOAJ	Journals	%
USA	5605	27.28	USA	263	13.77	USA	1201	12.37
UK	5036	24.51	UK	219	11.47	Brazil	908	9.35
The Netherlands	1706	8.30	Brazil	184	9.63	UK	611	6.29
Germany	1213	5.90	India	95	4.97	India	590	6.08
China	538	2.62	Spain	88	4.61	Spain	516	5.31
France	487	2.37	Japan	67	3.51	Egypt	440	4.53
Japan	459	2.23	New Zealand	66	3.46	Germany	333	3.43
Italy	401	1.95	Turkey	62	3.25	Romania	297	3.06
Spain	393	1.91	Chile	60	3.14	Italy	287	2.96
India	369	1.80	Poland	59	3.09	Canada	262	2.70

Table 3.6: Comparison of Top Publishing Countries in Scopus and DOAJ

3.5.5 Type of publishers

Type of publishers: I have used Solomon's classification, and almost 84% of the journals are owned by three categories of publishers: Professional, Society and University. Hereinafter, I present the relative percentage of journals in the first quartile for all categories of publishers: 25.00% professionally published journals, 10.61% society journals, 7.19% university published journals, 22.22% independent scholar publishers, 14.29% government agencies, 15.50% other organizations, and

5.17% unknown publishers. However, independent scholar publishers possess only 36 journals, and the percentage in the first quartile is very close to that of professionally published journals.

Scopus	Journals	%	Sample	Journals	%	DOAJ	Journals	%
USA	5605	27.28	USA	263	13.77	USA	1201	12.37
UK	5036	24.51	UK	219	11.47	Brazil	908	9.35
The Netherlands	1706	8.30	Brazil	184	9.63	UK	611	6.29
Germany	1213	5.90	India	95	4.97	India	590	6.08
China	538	2.62	Spain	88	4.61	Spain	516	5.31
France	487	2.37	Japan	67	3.51	Egypt	440	4.53
Japan	459	2.23	New Zealand	66	3.46	Germany	333	3.43
Italy	401	1.95	Turkey	62	3.25	Romania	297	3.06
Spain	393	1.91	Chile	60	3.14	Italy	287	2.96
India	369	1.80	Poland	59	3.09	Canada	262	2.70

TABLE 3.7: Type of publishers

3.5.6 Temporal Variables

Regarding the temporal variables, I have tested the age of the single journal intended as the period between its foundation and 12/31/2012, as well as the years in the DOAJ repositories. McVeigh [2004] has found that journals adopting an OA distribution model have not achieved a significantly greater citation impact. More recently, Giglia [2010] has stated: "Direct causal relationship between age and visibility and prestige in terms of citations cannot be straightforwardly inferred". Finally, Solomon et al [2013] have found that the distribution ages for subscription and free journals are roughly equal. Giving diverse results by previous studies seems to be an important aspect to investigate for this variable, as is understanding if the permanence in the DOAJ repositories can influence the ranking.

Years of Activity is a continuous variable referring to how many years the journal has existed and been published. The min is 1, the max is 132, the mean is 12.09, and σ is 10.24.

Years in DOAJ is a continuous variable and reports how many years the journal has been indexed in the DOAJ repository. The min is approximately six months, the max is 10 years and 7 months, the mean is 6.19, and σ is 2.65.

3.5.7 Born or converted

Born or converted is a Boolean variable that is assigned a value of 0 if the journal was born as OA (864) and a value of 1 (1,030) if it has been converted. Although most of the journals have been converted, over 68% of the journals in the first quartile were born as OA journals. For 16 journals, information is missing. This variable has been studied only by Solomon et al [2013], as they correctly noted that OA experiences growth in two distinct ways: first, by conversion of existing journals and, second, by the birth of new journals. As widely demonstrated, OA articles are, in general, cited more; however, as we can see in the table below, it seems that converted journals tend to obtain a low ranking.

	Born	Converted	Total
1^{st} Quartile	197	91	288
2^{nd} Quartile	238	261	499
3^{rd} Quartile	247	398	645
4^{th} Quartile	182	280	462
Total	864	1030	1894

Table 3.8: Born or converted

3.5.8 Publication Fee

Publication Fee: Distribution of this variable has a strategic importance because "pay to publish" might suggest a deterrent for gold OA. In this regard, Solomon et al [2013] have demonstrated that journals without APCs have increased more rapidly than others. Unexpectedly, the number of articles published with APCs is higher than articles without a publication fee. This means that journals adopting a business model that requires a form of payment tend to publish a larger number of papers. Hence, investigation of the APCs journals and their relationship with the ranking is interesting. According to our data, despite the fact that 62.11% of the journals do not require a publication fee, 61.45% of the top ranked journals do require one

	No Pub fee	Pub fee	Total
1^{st} Quartile	96	177	273
2^{nd} Quartile	260	219	479
3^{rd} Quartile	445	172	617
4^{th} Quartile	327	120	447
Total	1128	688	1816

Table 3.9: Publication Fee

3.5.9 Subject area

Subject area: As previously stated, journals are frequently classified in more than one subject area; in the database, 475 out of 1,910 journals are in more than one. Hence, the total number of journals per subject area is 2,518 instead of 1,910. There is an enormous difference between the percentage of top OA journals in the first quartile and the subject area, varying from 0% in Dentistry to over 44% in Energy. Again, in the last column, we can see the percentage of gold OA journals in the entire Scopus database; here, the percentage varies from 2.45% in Energy to 28.40% in Multidisciplinary.

Subject area	OA	Q1	% on Q1	Total in Scopus	% of OA
Agricultural and Biological Sciences	247	22	8.91%	1692	14.60%
Arts and Humanities	74	12	16.22%	2102	3.52%
Biochemistry, Genetics and Molecular Biology	201	32	15.92%	1553	12.94%
Business, Management and Accounting	26	1	3.85%	854	3.04%
Chemical Engineering	35	1	2.86%	439	7.97%
Chemistry	42	3	7.14%	563	7.46%
Computer Science	84	8	9.52%	1084	7.75%
Decision Sciences	8	1	12.50%	178	4.49%
Dentistry	14	0	0.00%	119	11.76%
Earth and Planetary Sciences	81	16	19.75%	853	9.50%
Economics, Econometrics and Finance	36	1	2.78%	640	5.63%
Energy	9	4	44.44%	367	2.45%
Engineering	109	7	6.42%	1893	5.76%
Environmental Science	82	15	18.29%	889	9.22%
Health Professions	23	4	17.39%	218	10.55%
Immunology and Microbiology	61	7	11.48%	458	13.32%
Material Science	52	6	11.54%	706	7.37%
Mathematics	72	10	13.89%	963	7.48%
Medicine	705	133	18.87%	5478	12.87%
Multidisciplinary	23	6	26.09%	81	28.40%
Neuroscience	48	10	20.83%	337	14.24%
Nursing	32	7	21.88%	371	8.63%
Pharmacology, Toxicology and Pharmaceutics	91	12	13.19%	554	16.43%
Physics and Astronomy	51	5	9.80%	568	8.98%
Psychology	43	3	6.98%	717	6.00%
Social Sciences	236	27	11.44%	3413	6.91%
Veterinary	33	3	9.09%	177	18.64%
Total	2518	356		27267	

TABLE 3.10: Subject area

3.6 Results and discussion

As stated in the introduction, the goal of this paper is to find out which variables, within the whole range of variables available, contribute significantly to the journal ranking. With categorical-dependent variables, the logistic regression is a robust method, in this case the dependent variable is dichotomic, so I have used the binomial logistic regression [Agresti, 2013, Hosmer Jr and Lemeshow, 2004]. Let us denote with $\mathcal{D} = \{\boldsymbol{x}_i, y_i\}_{i=1}^n$ a dataset sampled from an underlying (though unknown) data distribution $p(\boldsymbol{X}, Y)$, with $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ik}) \in \mathbb{R}^k$ and $y_i \in \{0, 1\}$. Logistic regression assumes a parametric model to approximate the posterior probability p as:

$$p(y = 0|\boldsymbol{x}) = \frac{1}{1 + \exp\left(\beta_0 + \boldsymbol{\beta}^\top \boldsymbol{x}\right)} \quad , \tag{3.1}$$

$$p(y = 1|\boldsymbol{x}) = \frac{\exp\left(\beta_0 + \boldsymbol{\beta}^\top \boldsymbol{x}\right)}{1 + \exp\left(\beta_0 + \boldsymbol{\beta}^\top \boldsymbol{x}\right)} \quad , \tag{3.2}$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)$ and β_0 are the model parameters, and can be estimated using the classical maximum likelihood estimation (MLE) framework, by solving the following optimization problem:

$$\max_{\boldsymbol{\beta}, \beta_0} \prod_{i=1}^{n} p(y_i|\boldsymbol{x}_i) \quad . \tag{3.3}$$

For mathematical convenience, one can equivalently maximize the log-likelihood $\sum_{i=1}^n \log p(y_i|\boldsymbol{x}_i)$. We refer the reader to [Bishop, 2007, Ch. 4.3.2] for further details. Logistic regression differs from linear regression as it estimates the posterior distribution $p\ (y \mid x)$ of the binary variable y, *i.e.*, a Bernoulli distribution rather than a Gaussian distribution. In practice, it describes how the probability of a particular category depends on the values of the independent variables.

Results. Results of the logistic regression are shown in Table 3.11; the model has fit the data quite well. In fact, the chi-square test rejects the hypothesis of no explanatory power, and the model correctly predicted 83% of the observations. As expected,the subject area of Dentistry was automatically dropped by the software (STATA 13) because estimation is not possible when a covariate does not vary within the category of an independent variable [Long and Freese, 2006]. In fact, ln (0) is undefined; that is, the variable's distribution does not permit a finite coefficient, and, therefore, this does not bias the remaining coefficients in the model. The software automatically checks for multi-correlation with no evidence of it. Further tests have been performed, and they reported a mean variance inflation factor (VIF) of 1.88; again, the condition number reported, 26.58, is below the critical value of 30, so it can be said that the regression has non-significant multi co-linearity. To evaluate the effects of independent variables, the significance values (p) were analyzed. If the significant values are less than 0.05 (95% confidence interval), it can be said that the independent variables have an effect on the ranking. With caution, we can say that a positive regression coefficient means that the explanatory variable increases the probability of the outcome,while a negative regression coefficient means that the variable decreases the probability of that outcome.

	Odds ratio	Std. error	Z	$P > \mid z \mid$	[95% Conf. Interval]	
English*	6.1936530	3.7279520	3.03	0.002	1.9037490	20.150420
Multilingual*	0.2527771	0.1095607	-3.17	0.002	0.1080944	0.591116
Spanish**	0.3705111	0.1776948	-2.07	0.038	0.1447339	0.948489
Portuguese	2.3790430	1.3972540	1.48	0.140	0.7524483	7.521906
French	1.2528960	0.6785869	0.42	0.677	0.4334026	3.621915
German	2.9442690	1.7766910	1.79	0.074	0.9022522	9.607867
USA*	1.9782060	0.4611308	2.93	0.003	1.2527140	3.123859
UK*	5.3925550	1.3765240	6.6	0.000	3.2697430	8.893559
Brazil	1.2027000	0.6430883	0.35	0.730	0.4217124	3.430031
India	0.5466591	0.2600571	-1.27	0.204	0.2151703	1.388835
Spain	1.6865990	1.0016790	0.88	0.379	0.5265983	5.401871
Japan	0.7066355	0.3845526	-0.64	0.523	0.2432034	2.053152
New Zealand	0.7741365	0.3522113	-0.56	0.574	0.3173538	1.888389
Turkey	0.2353511	0.2462435	-1.38	0.167	0.0302775	1.829417
Poland	0.6641993	0.4281880	-0.63	0.526	0.1877397	2.349853
Years of activity	1.0115100	0.0082510	1.4	0.161	0.9954667	1.027811
Years in DOAJ*	1.0904830	0.0361491	2.61	0.009	1.0218850	1.163687
Pub Fee*	2.0841300	0.4892931	3.13	0.002	1.3154900	3.301887
Agricultural and Biological Sciences	1.4097910	0.4248569	1.14	0.254	0.7809715	2.544921
Arts and Humanities*	6.0241560	2.6731210	4.05	0.000	2.5245810	14.374840
Biochemistry Genetics and Molecular Biology	1.2153530	0.3374231	0.7	0.482	0.7053104	2.094232
Business Management and Accounting	0.3556165	0.3914741	-0.94	0.348	0.0411104	3.076183
Chemical Engineering	0.4376612	0.4640878	-0.78	0.436	0.0547696	3.497327
Chemistry	1.1955240	0.8105124	0.26	0.792	0.3165783	4.514767
Computer Science	0.9307175	0.4061676	-0.16	0.869	0.3956865	2.189195
Decision Sciences	0.9017095	1.0897060	-0.09	0.932	0.0844110	9.632391
Earth and Planetary Sciences*	3.6979950	1.4398090	3.36	0.001	1.7240600	7.931957
Economics Econometrics and Finance	0.3561789	0.3806080	-0.97	0.334	0.0438617	2.892349
Energy*	23.3672500	20.3649200	3.62	0.000	4.2342110	128.956400
Engineering	0.6220425	0.2978551	-0.99	0.321	0.2433525	1.590026
Environmental Science	1.3922260	0.5216103	0.88	0.377	0.6680350	2.901485
Health Professions	3.4356430	2.3889130	1.77	0.076	0.8793066	13.423810
Immunology and Microbiology	0.6372469	0.3066982	-0.94	0.349	0.2481053	1.636739
Material Science	1.6366410	0.9283956	0.87	0.385	0.5383982	4.975116
Mathematics	1.1482720	0.5166916	0.31	0.759	0.4753651	2.773720
Medicine**	1.7590100	0.3951165	2.51	0.012	1.1325770	2.731926
Multidisciplinary*	4.4293140	2.5456210	2.59	0.010	1.4359280	13.662810
Neuroscience**	2.6651550	1.2406300	2.11	0.035	1.0702560	6.636774
Nursing	1.9013670	1.0975180	1.11	0.266	0.6133817	5.893879
Pharmacology Toxicology and Pharmaceutics	1.5658660	0.6049308	1.16	0.246	0.7343711	3.338826
Physics and Astronomy	0.4956044	0.3217324	-1.08	0.280	0.1388542	1.768932
Psychology	1.7591720	1.2358480	0.8	0.421	0.4439409	6.970945
Social Science*	2.3886400	0.7482052	2.78	0.005	1.2927800	4.413434
Veterinary	1.2798170	0.8721241	0.36	0.717	0.3365899	4.866250
Born or Converted	0.7322029	0.1509078	-1.51	0.130	0.4888751	1.096642
Professionally published Journal	0.9218220	0.6249184	-0.12	0.904	0.2441200	3.480893
Society Journals	1.3829560	0.9093563	0.49	0.622	0.3811575	5.017785
University published journals	1.0703100	0.7224771	0.1	0.920	0.2850589	4.018692
Independent scholar publisher	1.5080790	1.2398400	0.5	0.617	0.3010398	7.554823
Government Agency	2.6355750	1.9610160	1.3	0.193	0.6131033	11.329670
Other Organization	1.8591310	1.3077210	0.88	0.378	0.4683509	7.379870
Constant	0.0048070	0.0043770	-5.86	0.000	0.0008069	0.028000

* Statistically significant at 1% significance level.

**Statistically significant at 5% significance level.

TABLE 3.11: Logistic regression results

Regarding the first variable, Language, it can be said that English is highly and positively related to the ranking; however, this is not surprising news, as we are studying scientific journals, and English is recognized as the "standard language" to share research. In section 3.5.3, I have shown that over a quarter of journals are

Log-Lik Intercept Only:	-764.173	Log-Lik Full Model:	-581.789
D (1747)	1163.578	LR(51):	364.768
		Prob > LR:	0.000
McFadden's R^2	0.239	McFadden's Adj R^2:	0.171
ML (Cox-Snell) R^2:	0.184	Cragg-Uhler (Nagelkerke) R^2:	0.321
McKelvey & Zavoina's R^2:	0.425	Efron's R^2:	0.233
Variance of y:	5.722	Variance of error:	3.29
Count R^2:	0.863	Adj Count R^2:	0.092
AIC:	0.705	AIC* n:	1267.578
BIC:	-11930.163	BIC':	17.476
BIC used by Stata:	1553.317	AIC used by Stata:	1267.578

Table 3.12: Measures of fit for logistic regression

published in more than one language; according to the results, it is interesting to see how this strategy is strongly and negatively related to the ranking. If we look more deeply at the data, we can observe that 476 out 501 multilingual journals contain English as a publishing language, hence it is possible to state with evidence that English is a discriminant, but only if the journal is not multilingual. Again, we have strong negative presumption about Spanish; for other languages, there are no signs of significance. The second category of analyzed variables is the country, as we have already seen in the descriptive statistics; the larger OA publisher countries are quite different in comparison to the entire Scopus database. Looking at results, we can say that journals owned by UK and US publishers have a very strong and positive relation to the ranking; other countries do not show signs of significance. Gumpenberger et al [2013] show a higher rate of success for Japanese JCR indexed OA titles; the number of journals in their study appears coherent with our sample, but the results are not significant. A temporal variable yields an important outcome because an answer in this regard has not yet clearly been given by previous studies. With the results of regression, it seems clear how journals benefit from the increase of the permanence in the DOAJ repositories, while the age of the journal has no significant relation with the ranking. Regarding publication fee, from previous studies [Solomon, 2013], we already know that APCs journals are lower in comparison to others. It can be said with a very strong presumption (p value lower than 1%) that journals adopting a business model requiring a form of payment to publish tend to become top rated more than others. This result is in contrast to the work of Björk and Solomon [2012], which states that a funding mechanism is not related to the journal's quality. The fifth category regards the field of study, as seen by looking at regression results; the situation is very varied, and an outcome of 7 out of 27 subject areas with a positive relation to the ranking appears to have good results. If we consider how recently the OA phenomenon occurred, results support its goodness for multiple fields of studies. A qualitative analysis may be helpful in understanding the low success of OA in some subject areas, but that extends beyond the scope of this paper. Although descriptive statistics show prevalence in the first quartile of born OA journals, the results of regression have not confirmed a significant relationship. A final word is devoted to the fact that no significant effect has been found for the types of publishers, despite the fact that the lower number of independent scholarly publishers in comparison to other types of publishers is still low; the probability of obtaining high visibility seems to be equal.

3.7 Conclusions

In this paper, I have provided an integrated and novel approach to understanding the features related to the ranking of top gold OA journals. I have provided several descriptive statistics, and I have successfully applied a binary logistic regression to test all seven of the variables by their relationship with the ranking. To summarize, I have found that English is significant with a positive sign, while Multilingual and Spanish are significant with negative signs, and there are no signs of significance for other tested languages. Again, I have found that journals owned by organizations located in the USA and in the UK are positively related to the ranking. Regarding the years of existence of the journals, I have found that this variable is not related to the rankings; conversely, years in the DOAJ repository are positively related with the rankings. In other words, *coeteris paribus*, increasing the permanence in the DOAJ repository increases a journal's ranking. I have also identified the fields of study where gold OA has reached positive results in terms of ranking. Again, I have not found a significant relationship between ranking and journals launched as OA or converted and types of publishers. Moreover, this research has brought to light an important question about funding models; because top ranking journals tend to require a fee to publish, this situation can limit *de facto* research sharing in gold OA for those who cannot support payment of publication fees.

Consider the Berlin OA declaration (2003) that states: "[...] mission of disseminating knowledge is only half complete if the information is not made widely and readily available to society [...]". It can be said that gold OA has reached important results for what concerns the diffusion, but it is still struggling to achieve widespread high ranking. I believe these findings might be helpful to the OA cause, particularly for the fields of study where journals have not yet reached high rankings. This paper has three limitations: First, it only takes into consideration journals indexed in the Scopus dataset. Second, the analysis is limited to journals indexed on 12/31/2012; a temporal evolution study could be helpful to avoid potential extemporaneous situations. Third, this research focuses only on gold OA journals; analyzing the same variables and comparing them to green OA and traditional publishing methods could offer more insight. Thus, further studies are required to improve research findings.

Bibliography

Abramo G, D'Angelo CA (2015) The vqr, italy's second national research assessment: Methodological failures and ranking distortions. Journal of the Association for Information Science and Technology

Agresti A (2013) Categorical data analysis. John Wiley & Sons

Anderson P, Tushman ML (1990) Technological discontinuities and dominant designs: A cyclical model of technological change. Administrative science quarterly pp 604–633

Antelman K (2004) Do open-access articles have a greater research impact? College & research libraries 65(5):372–382

Bador P, Lafouge T (2010) Comparative analysis between impact factor and h-index for pharmacology and psychiatry journals. Scientometrics 84(1):65–79

Benos DJ, Fabres J, Farmer J, Gutierrez JP, Hennessy K, Kosek D, Lee JH, Olteanu D, Russell T, Shaikh F, et al (2005) Ethics and scientific publication. Advances in Physiology Education 29(2):59–74

Besen SM, Farrell J (1994) Choosing how to compete: Strategies and tactics in standardization. The Journal of Economic Perspectives pp 117–131

Bishop CM (2007) Pattern Recognition and Machine Learning (Information Science and Statistics), 1st edn. Springer

Björk BC (2004) Open access to scientific publications–an analysis of the barriers to change? University of Lund Library

Björk BC, Solomon D (2012) Open access versus subscription journals: a comparison of scientific impact. Bmc Medicine 10(1):73

Blem E, Menon J, Sankaralingam K (2013) Power struggles: Revisiting the risc vs. cisc debate on contemporary arm and x86 architectures. In: High Performance Computer Architecture (HPCA2013), 2013 IEEE 19th International Symposium on, IEEE, pp 1–12

Bornmann L, Marx W, Gasparyan AY, Kitas GD (2012) Diversity, value and limitations of the journal impact factor and alternative metrics. Rheumatology international 32(7):1861–1867

Braun T, Glänzel W, Schubert A (2006) A hirsch-type index for journals. Scientometrics 69(1):169–173

Choi JP (1994) Irreversible choice of uncertain technologies with network externalities. The Rand Journal of Economics pp 382–401

Craig ID, Plume AM, McVeigh ME, Pringle J, Amin M (2007) Do open access articles have greater citation impact?: a critical review of the literature. Journal of Informetrics 1(3):239–248

David PA (1985) Clio and the economics of qwerty. The American economic review pp 332–337

De Vries HJ, De Ruijter JP, Argam N (2011) Dominant design or multiple designs: The flash memory card case. Technology Analysis & Strategic Management 23(3):249–262

DuBois FL, Reeb D (2000) Ranking the international business journals. Journal of International Business Studies 31(4):689–704

Eysenbach G (2006) Citation advantage of open access articles. PLoS biology 4(5):e157

Falagas ME, Kouranos VD, Arencibia-Jorge R, Karageorgopoulos DE (2008) Comparison of scimago journal rank indicator with journal impact factor. The FASEB journal 22(8):2623–2628

Foster RN (1988) Innovation: The attacker's advantage. Summit Books

Franke RH, Edlund TW, Oster F (1990) The development of strategic management: Journal quality and article impact. Strategic Management Journal 11(3):243–253

Galvagno M, Faraci RF (2004) La coesistenza fra tecnologie: Definizione ed elementi costitutivi (the coexistence of technologies: Definition and main components). Sinergie

Garfield E (2006) Citation indexes for science. a new dimension in documentation through association of ideas. International Journal of Epidemiology 35(5):1123–1127

Garud R, Kumaraswamy A (1993) Changing competitive dynamics in network industries: An exploration of sun microsystems' open systems strategy. Strategic Management Journal 14(5):351–369

Giglia E (2010) The impact factor of open access journals: data and trends. E-Prints in Library and Information Science

Gilbert BA, Taeube F, Migendt M, Brymer R (2014) Finding a place at the table: A study on the re-emergence of a radical technology. In: Strategic Management Society Conference

Gomes-Casseres B (2003) Competitive advantage in alliance constellations. Strategic organization 1(3):327–335

González-Pereira B, Guerrero-Bote VP, Moya-Anegón F (2010) A new approach to the metric of journals' scientific prestige: The sjr indicator. Journal of informetrics 4(3):379–391

Gumpenberger C, Ovalle-Perandones MA, Gorraiz J (2013) On the impact of gold open access journals. Scientometrics 96(1):221–238

Harnad S, Brody T, Vallières F, Carr L, Hitchcock S, Gingras Y, Oppenheim C, Stamerjohanns H, Hilf ER (2004) The access/impact problem and the green and gold roads to open access. Serials review 30(4):310–314

Harter SP, Nisonger TE (1997) Isi's impact factor as misnomer: a proposed new measure to assess journal impact. Journal of the American society for information science 48(12):1146–1148

Hayes H (2005) Digital repositories: Helping universities and colleges. JISC Briefing Paper: Higher Education Sector

Hicks D, Wouters P, Waltman L, de Rijcke S, Rafols I (2015) Bibliometrics: The leiden manifesto for research metrics. Nature 520(7548):429–431

Hirsch JE (2005) An index to quantify an individual's scientific research output. Proceedings of the National academy of Sciences of the United States of America 102(46):16,569–16,572

Hodge DR, Lacasse JR (2010) Evaluating journal quality: Is the h-index a better measure than impact factors? Research on Social Work Practice

Hosmer Jr DW, Lemeshow S (2004) Applied logistic regression. John Wiley & Sons

Katz ML, Shapiro C (1985) Network externalities, competition, and compatibility. The American economic review pp 424–440

Krad H, Al-Taie AY (2007) A new trend for cisc and risc architectures. Asian J Inform Techno 6(11):1125–1131

Lancho-Barrantes BS, Guerrero-Bote VP, Moya-Anegón F (2010) What lies behind the averages and significance of citation indicators in different disciplines? Journal of Information Science 36(3):371–382

Lawrence S (2001) Online or invisible. Nature 411(6837):521

Lee H, Shin J (2014) Measuring journal performance for multidisciplinary research: An efficiency perspective. Journal of Informetrics 8(1):77–88

Leydesdorff L (2009) How are new citation-based journal indicators adding to the bibliometric toolbox? Journal of the American Society for Information Science and Technology 60(7):1327–1336

Leydesdorff L, Milojević S (2012) Scientometrics. arXiv preprint arXiv:12084566

Leydesdorff L, Opthof T (2010) Scopus's source normalized impact per paper (snip) versus a journal impact factor based on fractional counting of citations. Journal of the American Society for Information Science and Technology 61(11):2365–2369

Leydesdorff L, de Moya-Anegón F, Guerrero-Bote VP (2010) Journal maps on the basis of scopus data: A comparison with the journal citation reports of the isi. Journal of the American Society for information Science and Technology 61(2):352–369

Leydesdorff L, Radicchi F, Bornmann L, Castellano C, Nooy W (2013) Field-normalized impact factors (ifs): A comparison of rescaling and fractionally counted ifs. Journal of the American Society for Information Science and Technology 64(11):2299–2309

Lin CP, Tsai YH, Wang YJ, Chiu CK (2011) Modeling it relationship quality and its determinants: A potential perspective of network externalities in e-service. Technological Forecasting and Social Change 78(1):171–184

Lobachev S (2008) Top languages in global information production. Partnership: the Canadian Journal of Library and Information Practice and Research 3(2)

Long JS, Freese J (2006) Regression models for categorical dependent variables using Stata. Stata press

Martin BR, Irvine J (1984) Foresight in science–picking the winners. Pinter, London

McIntyre DP, Chintakananda A (2014) Competing in network markets: Can the winner take all? Business Horizons 57(1):117–125

McVeigh ME (2004) Open access journals in the ISI citation databases: analysis of impact factors and citation patterns: a citation study from Thomson Scientific. Thomson Scientific

Miguel S, Chinchilla-Rodriguez Z, de Moya-Anegón F (2011) Open access and scopus: A new approach to scientific visibility from the standpoint of access. Journal of the American Society for Information Science and Technology 62(6):1130–1145

Moed HF (2007) The effect of "open access" on citation impact: An analysis of arxiv's condensed matter section. Journal of the American Society for Information Science and Technology 58(13):2047–2054

Murmann JP, Frenken K (2006) Toward a systematic framework for research on dominant designs, technological innovations, and industrial change. Research Policy 35(7):925–952

Mylonopoulos NA, Theoharakis V (2001) On site: global perceptions of is journals. Communications of the ACM 44(9):29–33

Nair A, Ahlstrom D (2003) Delayed creative destruction and the coexistence of technologies. Journal of Engineering and Technology Management 20(4):345–365

O'Reilly CA, Tushman ML (2008) Ambidexterity as a dynamic capability: Resolving the innovator's dilemma. Research in organizational behavior 28:185–206

Peffers K, Ya T (2003) Identifying and evaluating the universe of outlets for information systems research: Ranking the journals. Journal of Information Technology Theory and Application (JITTA) 5(1):6

Pinch TJ, Bijker W (2009) 8 "the social construction of facts and artifacts". TECHNOLOGY AND SOCIETY p 107

Pinfield S (2005) A mandate to self archive? the role of open access institutional repositories. Serials: The Journal for the Serials Community 18(1):30–34

Prosser D (2003) Institutional repositories and open access: The future of scholarly communication. Information services and Use 23(2):167–170

Raffaelli R (2013) Mechanisms of technology re-emergence and identity change in a mature field: Swiss watchmaking. In: Academy of Management Proceedings, Academy of Management, 1, p 13784

Rosenberg N (1982) Inside the black box: Technology and economics. Cambridge University Press

Schilling MA (2002) Technology success and failure in winner-take-all markets: The impact of learning orientation, timing, and network externalities. Academy of Management Journal 45(2):387–398

Schumpeter JA (1934) The theory of economic development: An inquiry into profits, capital, credit, interest, and the business cycle, vol 55. Transaction publishers

Seglen PO (1997) Why the impact factor of journals should not be used for evaluating research. Bmj 314(7079):497

Sellers SL, Mathiesen SG, Perry R, Smith T (2004) Evaluation of social work journal quality: Citation versus reputation approaches. Journal of Social Work Education 40(1):143–160

Semmler A (2010) Competition in the microprocessor market: Intel, amd and beyond. Univeristy of Teier pp 1–7

Serenko A, Bontis N (2004) Meta-review of knowledge management and intellectual capital literature: citation impact and research productivity rankings. Knowledge and process management 11(3):185–198

SJR (2007) - scimago journal & country rank. SCimago Retrieved in 2014, URL http://www.scimagojr.com

Solomon D (2013) Types of open access publishers in scopus. Publications 1(1):16–26

Solomon DJ, Laakso M, Björk BC (2013) A longitudinal comparison of citation rates and growth among open access journals. Journal of Informetrics 7(3):642–650

Suarez FF (2004) Battles for technological dominance: an integrative framework. Research Policy 33(2):271–286

Taylor M, Taylor A (2012) The technology life cycle: Conceptualization and managerial implications. International journal of production economics 140(1):541–553

Tegarden LF, Hatfield DE, Echols AE (1999) Doomed from the start: What is the value of selecting a future dominant design? Strategic Management Journal 20(6):495–518

Theuβl S, Reutterer T, Hornik K (2014) How to derive consensus among various marketing journal rankings? Journal of Business Research 67(5):998–1006

Tsai CF (2014) Citation impact analysis of top ranked computer science journals and their rankings. Journal of Informetrics 8(2):318 – 328

Tüselmann H, Sinkovics RR, Pishchulov G (2015) Towards a consolidation of worldwide journal rankings–a classification using random forests and aggregate rating via data envelopment analysis. Omega 51:11–23

Vanclay JK (2008) Ranking forestry journals using the h-index. Journal of informetrics 2(4):326–334

Vapnik VN (1995) The nature of statistical learning theory. Springer-Verlag New York, Inc., New York, NY, USA

Venkatraman N, Lee CH (2004) Preferential linkage and network evolution: A conceptual model and empirical test in the us video game sector. Academy of Management Journal 47(6):876–892

Vieira ES, Gomes JA (2009) A comparison of scopus and web of science for a typical university. Scientometrics 81(2):587–600

Vokurka RJ (1996) The relative importance of journals used in operations management research a citation analysis. Journal of Operations Management 14(4):345–355

Werner S (2002) Recent developments in international management research: A review of 20 top management journals. Journal of Management 28(3):277–305

Windrum P, Birchenhall C (2005) Structural change in the presence of network externalities: a co-evolutionary model of technological successions. Journal of evolutionary economics 15(2):123–148