Crowdsourced Digital Goods and Firm Productivity: Evidence from Open Source Software

Frank Nagle (University of Southern California)[1]

August 28, 2015

Abstract

As firms increasingly rely on crowdsourced digital goods, understanding their impact on productivity becomes critical. This study measures the firm-level productivity impact of one such good, non-pecuniary (free) open source software (OSS). The results show a previously unmeasured positive and significant return to the usage of non-pecuniary OSS that is not solely due to cost savings. Inverse probability weighting, instrumental variables, firm fixed effects, and management quality data add support for a causal interpretation. Across firms, a 1% increase in non-pecuniary OSS leads to a .073% increase in productivity or a $1.35 million increase in value-added production for the average firm in the sample. This effect is greater for larger firms and for firms in the services industry. These findings indicate that existing studies underestimate the amount of IT firms use and suggest that firms assuming the risks associated with non-pecuniary OSS gain benefits from collective intelligence and labor spillovers.

---

[1] frank.nagle@marshall.usc.edu. The author is grateful for helpful comments from Shane Greenstein, Carliss Baldwin, Yochai Benkler, Raj Choudhury, Anil Doshi, Marco Iansiti, Ohchan Kwon, Karim Lakhani, Kristina McElheran, Hart Posen, Scott Stern, Neil Thompson, Mike Toffel, Joel West, and Feng Zhu. Additional helpful comments were received from participants at ACAC 2014, AEA 2015, AOM 2014, AOM 2014 BPS Dissertation Consortium, CCC 2014, Charles River Conference 2014, DRUID 2014, HBS TOM DBA Seminar 2014, HBS TOM Alumni Conference 2014, NYU Engelberg Center Conference on Knowledge Commons 2014, OUI 2014, SMS 2014, and ZEW ICT Conference 2014. Helpful comments were also received from seminar participants at Bocconi University, Boston College, Carnegie Mellon University, Columbia Business School, Harvard Business School, IESE Business School, McGill University, Temple University, University College London, University of Maryland, University of Pennsylvania, and University of Southern California. All mistakes remain the author's own.

I. Introduction

As the digital age progresses, information goods are easier and easier to reproduce at costs that are rapidly approaching zero. Coupled with decreases in communication costs, this has made it easier for groups of individuals, frequently referred to as the crowd, to produce digital goods that are freely distributed to users who do not pay a monetary price. Wikipedia, the online crowdsourced encyclopedia, is a frequently cited example of this phenomenon, although there are many other examples including open source software (OSS), crowdsourced innovation tournaments, and the digitization of consumers' opinions via online review sites and social media. The same information cost decreases that enable the production of these goods also enable firms to use these crowdsourced goods as inputs into production. Recent research has shown that firms are increasingly relying on these types of goods to drive innovation and production (Baldwin and von Hippel, 2011; Lakhani, Lifshitz-Assaf, and Tushman, 2012; Corrado and Hulten, 2013; Altman, Nagle, and Tushman, 2014).

This trend is also widely discussed in the popular press as technology giants like Apple, Google, and Facebook increase their reliance on crowdsourced digital goods to grow their innovative and productive efforts (Sorkin and Peters, 2006; Asay, 2013; Finley, 2013). However, it is not only technology focused companies that are relying on the crowd - Ford, Pepsi, Walmart, and a host of other well-known non-IT brands use free inputs from the crowd to help drive their bottom line (Horovitz, 2013; McCue, 2013; Phipps, 2014). Additionally, these same crowd-based technologies are allowing small start-ups to have a large impact, even when they are capital constrained, due to a reliance on free crowdsourced digital goods as inputs. OSS, the empirical focus of this study, is a particularly important example of a crowdsourced digital good as more than 50% of firms now use or contribute to OSS (Black Duck, 2014) and billions of venture capital dollars are pouring into the OSS ecosystem (Black Duck, 2014; Forrest, 2014; Hamilton, 2014; Lunden, 2014). Further, due to the rise of mobile operating systems such as Android and iOS, more than 50% of all computing devices are now based on OSS (Yarow, 2013).

Despite the growing importance of crowdsourced digital goods as inputs into production, measuring the value they help create can be difficult. In a classic Schumpeterian creative destruction process (Schumpeter, 1942), these new goods destroy old business models while creating new opportunities for growth. For example, the introduction of Wikipedia destroyed much of the market for pecuniary encyclopedias (both paper and digital). At the same time, Wikipedia has provided great societal value. However, as with all crowdsourced digital goods, this value is difficult to measure for two primary reasons. First, because these goods are frequently free, standard productivity measures, which rely on price to reflect value, do not properly capture these increasingly critical inputs. Second, because such goods are often distributed under licenses that allow for unlimited copying, it is unknown exactly how widespread they are. Despite the increasing prominence of crowdsourcing, these measurement challenges have prevented researchers from analyzing how its impact varies across different firms and market environments. Further, it has been suggested that integrating such resources into the firms production process can be more costly than comparable non-crowdsourced inputs (Giera and Brown, 2004), and consequently their use could have a negative impact on productivity. Therefore, the goal of this paper is to answer the following question: what is the impact of non-pecuniary crowdsourced digital goods on firm productivity? After answering this broad question, the paper seeks to answer the related question: What are the firm-level determinants of the productivity impact of such goods?

As the production, and productive use, of such goods increases, the answer to these questions becomes more interesting and more important. Recent research has shown that the increased use of unpriced goods of both a digital (Brynjolfsson and Saunders, 2009; Greenstein and Nagle, 2014) and non-digital (Bridgman, 2013) nature may be an important factor in understanding recent trends in Gross Domestic Product (GDP). Non-pecuniary digital goods can cause standard GDP measures to greatly underestimate the true productivity of a nation and its firms. These same mismeasurement issues can lead firms and managers to underestimate the importance of including crowdsourced digital goods as key inputs into their productive and innovative processes. While some leading firms, like Google and Facebook, have embraced the crowd and the free labor and content it provides, others have shied away from relying on such inputs due to concerns about reliability, sharing with competitors, and the costs of restructuring business models to add the user directly into the production and innovation process.

In addition to productivity-related implications, the reliance on, and contribution to, crowdsourced goods also has implications for firm competitive strategy. In a world where a firm must rely on actors outside of its boundaries for valuable inputs, and at the same time must consider contributing internally developed code to the world, co-opetition (Brandenburger and Nalebuff, 1996; Afuah, 2000) becomes an increasingly important concept. As firms' competitors increase their reliance on crowdsourced digital goods, understanding how these goods contribute to productivity and for what types of firms they are the most useful becomes increasingly important to allow managers to make the right decisions regarding the crowd. Finally, understanding the productive implications of free digital goods scratches the surface of the broader issue of all digital goods, which essentially have a marginal cost of zero, and are therefore likely priced below their actual value.

To understand how usage of such non-pecuniary digital inputs affects firm productivity, this paper first discusses why such goods could have a positive or negative impact on productivity and then considers what firm characteristics are likely to determine the degree of this impact. To test the resultant competing hypotheses, it utilizes a dataset that measures the usage of one particularly important non-pecuniary crowdsourced digital good, open source software (OSS) operating systems. OSS is an important digital good that is produced by a community of tens of thousands of users and is frequently distributed free of charge. Thus it is exactly the type of non-pecuniary digital input that is uncounted in GDP and other productivity measures. This data is combined with firm financial data and productivity measures to allow for the application of a classic Cobb-Douglas production function analysis to understand the role of non-pecuniary IT inputs in firm-level productivity. This is a standard methodology for estimating the value of IT (Brynjolfsson and Hitt, 1996; Dewan and Min, 1997; Tambe, Hitt, and Brynjolfsson, 2012; Huang, Ceccagnoli, Forman, and Wu, 2013), although non-pecuniary OSS is normally not accounted for in such frameworks. Due to sample selection and endogeneity concerns, inverse probability weighting, a method similar to propensity score matching, is used to construct a setting more like that of an experiment. Panel fixed effects and instrumental variables are also utilized to allow for a more causal interpretation of the results. Further, for a sub-sample of the firms, data from the World Management Survey (Bloom, Sadun, and Van Reenen, 2012) is used to show that there is no correlation between firm management quality and the use of non-pecuniary OSS, indicating that the full sample results are not proxying for management quality.

The results show that firms that use non-pecuniary OSS have higher levels of productivity than those that do not. They also show that increased usage of non-pecuniary OSS has a positive and significant impact on firm productivity. This makes intuitive sense since firms that use non-pecuniary IT are able to tap into the collective intelligence of the crowd through spillovers from free labor. The primary effect is robust to various endogeneity concerns, allowing for a causal interpretation of the results. The estimates indicate that a 1% increase in the amount of non-pecuniary OSS used by a firm leads to a .073% increase in productivity when comparing firms against other firms. The average value added for the firms in the sample is $1.846 billion; this indicates that a 1% increase in the number of non-pecuniary OSS operating systems leads to a $1.35 million increase in value-added production (or profits) for the average firm. This effect size is more than double the size of the coefficient on traditional pecuniary IT capital. This effect is greater for larger firms and for firms in the services sector (versus those in the manufacturing sector). The main effect is of a similar order of magnitude as other IT-related inputs. Because the study measures only non-pecuniary OSS operating systems, it does not capture other firm investments in non-pecuniary OSS, thus the main effect is likely a lower bound for the true effect of all non-pecuniary OSS on productivity. Further, the results indicate that it is not only the lack of cost of such software that provides a benefit to the firm. Indeed, if the non-pecuniary OSS were assigned a cost similar to that of other pecuniary operating systems, it would still have a significant positive effect. Finally, the results indicate that current studies underestimate the amount of IT at the firm.

This paper seeks to add insights to two important bodies of literature: the user innovation literature and the returns to IT literature. The user innovation literature (e.g., von Hippel, 1986, Chatterji and Fabrizio, 2014), in particular that which is centered on OSS (e.g., Kogut and Metiu, 2001; Lerner and Tirole, 2002; Lakhani and von Hippel, 2003; West and Lakhani, 2008), focuses primarily on supply side questions, e.g. why do individuals and firms contribute time and resources to the development of OSS, with almost no literature focusing on the demand and usage side of the OSS market. At the same time, the literature on the returns to IT investment (e.g., Brynjolfsson and Hitt, 1996; Tambe and Hitt, 2012; Huang, Ceccagnoli, Forman, and Wu, 2013) focuses almost exclusively on IT investments of a pecuniary nature, completely missing investments in non-pecuniary IT, such as OSS. This paper contributes to both of these bodies of work by filling these important gaps in the literature and shedding light on the underestimation of IT used by the firm, and therefore the underestimation of the productivity impact of non-pecuniary IT. Understanding the impact of such goods on firm productivity not only helps to contribute to the broad literature on the determinants of productivity[2], but also shows that user innovation is no longer a rare phenomenon and is becoming a key input into firm productivity and innovation. Additionally, the paper offers insights for practitioners that can be utilized to increase the profitability of the firm's operations and gain competitive advantage by using crowdsourced goods as inputs. Finally, for policy makers, the results encourage policies that incentivize production of public digital goods as a method for increasing firm and, in turn, national productivity.

II. Crowdsourced Digital Goods and the Returns to Information Technology

One of the oldest and most successful crowdsourced digital goods is open source software and this will be the empirical setting of this analysis. Therefore, this section first reviews prior research on crowdsourced digital goods and user innovation as well as research on the returns to IT investments. In doing so, an important gap is identified at the intersection of these two literatures, motivating the primary research question. Then, this section gives a brief history of the development of the two most widely used OSS operating system, GNU/Linux and BSD, both of which play an integral part in today's modern IT ecosystem.

[2] See Syverson, 2011 for an over view of this literature.

II.A Free and Open Source Software as an Input into Productivity

As early as the 1980's, production by users has been a topic of interest in the management field (von Hippel, 1986). While such production is by no means limited to the digital world, it is here that user innovation is frequently studied, primarily in the realm of OSS. However, most of the academic work on OSS has been focused on exploring supply side mechanisms – why do users contribute to OSS (Benkler, 2002; Lerner and Tirole, 2002; West and Lakhani, 2008, Athey and Ellison, 2014), how do users join OSS projects (von Krogh, Spaeth, and Lakhani, 2003), how do users help each other contribute to OSS (Lakhani and von Hippel, 2003), and how do OSS communities organize to protect their intellectual property (O'Mahony, 2003) and to guard against free-riding (Baldwin and Clark, 2006). Research on the supply side has also been extended to better understand why firms release some of their proprietary code as OSS (Harhoff, Henkel, and von Hippel, 2003; von Hippel and von Krogh, 2003; Lerner, Pathak, and Tirole, 2006; Henkel, 2006; Fosfuri, Giarratana, and Luzzi, 2008; Lerner and Schankerman, 2010; Casadesus-Masanell and Llanes, 2011). Despite the abundance of literature on the supply side of OSS, there is almost no literature on the demand side of OSS[3] – who uses it, why do they use it, and are there productivity benefits to using it remain unanswered questions. This is despite the fact that OSS, and – more broadly – non-pecuniary, community-based user-production, has been identified as an increasingly important input into the business models of firms in both academic literature (Krishnamurthy, 2005; Baldwin and von Hippel, 2011; Lakhani, Lifshitz-Assaf, and Tushman, 2012; Altman, Nagle, and Tushman, 2014; Greenstein and Nagle, 2014) and popular literature (Howe, 2008; Shirky, 2008).

Although the productivity related value of OSS usage has not been directly investigated, there is a significant body of literature examining the impact of IT usage on productivity at both the firm and country levels. This literature has shown that the rate of return for investments in IT is positive and significant (Brynjolfsson and Hitt, 1996; Athey and Stern, 2002) and productivity boosts from investments in IT are frequently mistaken for intangible firm-specific benefits (Brynjolfsson, Hitt, and Yang, 2002; Syverson, 2011; Tambe, Hitt, and Brynjolfsson, 2011; Saunders and Brynjolfsson, 2013). Studies have also shown that IT-producing and using industries contributed a disproportionately large amount to the economic growth experienced in the US, particularly from 1995-2004 (Jorgenson, 2001; Jorgenson, Ho, and Stiroh, 2005). In addition to spending on IT capital, spending on IT labor has also been found to boost firm productivity (Tambe and Hitt, 2012). Further, participation in networks of practice adds IT related knowledge spillovers that increase productivity (Huang, Ceccagnoli, Forman, and

[3] The one notable exception is Lerner and Schankerman (2010), which explores the cross-country differences in demand for OSS usage. However, their analysis does not examine the returns to OSS usage and does not include the US.

Wu, 2013; Ceccagnoli, Forman, Huang, and Wu, 2014). Relatedly, investments in IT outsourcing have been shown to have a positive impact on productivity (Han, Kauffman, and Nault, 2011; Han and Mithas, 2013) However, it has been found that not all firms receive the same return on IT investment (Aral and Weill, 2007) and that the returns to IT investment are not as strong as they once were (Byrne, Oliner, Sichel, 2013). An important aspect of all such studies is that they measure IT investment via dollars spent on software, hardware, labor, or a combination of the three. Since most OSS does not have a price directly associated with it,[4] it is not properly factored into such calculations. This mismeasurement of "digital dark matter" has been shown to be on the order of billions of dollars for one piece of OSS in the US alone (Greenstein and Nagle, 2014) and the inclusion of intangibles[5] and non-pecuniary production have been shown to significantly alter GDP calculations (Corrado, Hulten, and Sichel, 2009; Bridgman, 2013). Because of this measurement issue, OSS is not properly included in current productivity calculations, and therefore the productive value of OSS is currently unknown.

Despite the vast literatures that exist in these two areas, there is a noticeable dearth of literature that addresses the intersection, leaving an open question this paper attempts to answer: What is the impact of OSS on firm productivity? After establishing a baseline answer to this question, the paper further considers the firm-level differences in extracting productivity value from OSS, allowing for a better understanding of the productivity implications of non-pecuniary crowdsourced digital goods.

II.B Institutional Context: The Free and Open Source Software Movement

Although the concept of free and open source software developed as part of the early computer culture, it was not formalized until 1983 when Richard Stallman founded the GNU Project[6] to create a computer operating system that gave users the freedom to share and modify the software, unlike the predominant operating system at the time, UNIX, which was proprietary and closed-source software. Two years later, Stallman founded the Free Software Foundation (FSF), a non-profit organization designed to encourage the creation and dissemination of software with unrestrictive licenses, including the GNU General Public License (GPL), which continues to be the most widely used software license for free software. The FSF emphasizes that it uses the word "free" to mean "liberty, not price", encapsulated in

---

[4] Although some literature exists analyzing the total-cost of ownership (TCO) when comparing open and closed source software (e.g., MacCormack, 2003; Varian and Shapiro, 2003; Russo et al, 2005; Wheeler, 2005; Fitzgerald, 2006), a consensus has not been reached and this literature does not explore the productivity implications of the two types of software, just the costs of employing it. The analysis in this study will control for the costs of employing either type of software by including labor and capital costs in the analysis. This allows for the measurement of the impact of the software itself even though the TCO question is not directly addressed.

[5] Intangible assets include intellectual property, user-generated content, organizational capital, and human capital.

[6] GNU is a recursive acronym for "GNU's Not UNIX".

the pithy slogan “free as in free speech, not as in free beer.”[7] However, the software released under this license is frequently also offered at a price of zero. This ambiguity later led to Eric Raymond’s call for the use of the term “open source” instead of “free” (Raymond, 1998).

As the GNU Project progressed, it was successful in creating most of the middle and upper layers (user interface) of the operating system. However, very little work had been finished for the lowest layers, known as the kernel, of the operating system. In 1991, Linus Torvalds released the Linux kernel to take the place of the incomplete GNU kernel. GNU developers rapidly latched on to the Linux kernel and the combination of the Linux kernel and GNU software on top of it became the basis for most free and open source operating systems in use today. The other main free and open source operating system is the Berkeley Software Distribution (BSD) operating system, which was initially proprietary until a variant of version 4.3 was released as open source in 1989 under the terms of the BSD License, which allowed for redistribution provided the BSD License was included. Both GNU/Linux and BSD rely on a community of mostly unpaid contributors to maintain and upgrade the code base.[8] From 2005 to 2013, nearly 10,000 developers contributed to the Linux Kernel (Corbet, Kroah-Hartman, McPherson, 2013). From 1993-2014, FreeBSD, one of the largest BSD distributions, had nearly 1,000 core developers and nearly 3,000 contributors (FreeBSD, 2014).

Since these early operating systems were released, there has been a flood of free and open source software projects that are either a variant of these operating systems or are applications that run on top of them, such as the vast array of projects maintained by the Apache Software Foundation. Although unrestricted non-pecuniary software is at the core of the free and open source software movement, many companies have structured profitable business models on top of this software. Common examples include Red Hat, which offers its own Linux distribution and charges for customer support, the IBM HTTP Server, which is built on the open source Apache HTTP Server and is included with the IBM WebSphere Application Server, and Apple’s Mac OS X, which is built on the FreeBSD operating system. Figure 1 gives various examples of operating systems and other software that fall on different dimensions of price and the openness of the code base.

[7] http://www.gnu.org/philosophy/free-sw.html, retrieved on February 23, 2014.
[8] Although historically such OSS projects relied primarily on unpaid contributors, larger projects are increasingly receiving contributions from coders who are paid by their company to contribute to the code base. However, from the perspective of the OSS project, these contributions are unpaid since the project does not pay the coders directly. Further, during the timeframe of the empirical setting in this paper, widespread corporate contributions to OSS were limited.

Insert Figure 1 Here

III. Theory and Hypothesis Development

As shown in Figure 1, when a firm considers a software investment, it must make decisions along two important dimensions: price and whether the code base is open or closed. Compared to closed and pecuniary software, using free and open source software can be risky, but it can also provide a number of additional benefits. This section discusses these risks and benefits and develops competing hypotheses about the baseline productivity impact of using non-pecuniary OSS as well as further hypotheses about the characteristics of the firm that moderate the main effect.

III.A Risks of Using Non-Pecuniary OSS

Compared to pecuniary and closed source software, non-pecuniary OSS can be a risky investment. This section discusses the largest of these risks, including the fact that free software is not costless, there is no guaranteed technical support or technical path, OSS has security concerns not present in closed source software, and there is no contractual relationship allowing for recourse if something goes wrong.

When considering implementing new software, the allure of "free" software can be great for any capital constrained firm. However, firms run the risk of assuming that implementing such software will be costless. The price of the software itself does not truly represent the total cost of ownership (TCO) of the investment. Indeed, although there is a diversity of opinions, the consensus in the literature on the TCO of software is that the actual cost for software is negligible when compared to the hardware and labor costs of implementing, using, and maintaining it (e.g., Varian and Shapiro, 2003; Russo et al, 2005; Wheeler, 2005; Fitzgerald, 2006). In a review of the literature on TCO, MacCormack (2003) finds that the one fact most TCO studies can agree on is that the purchase price of a piece of software represents less than 10% of all of the costs that go into using that software. Therefore, one of the most salient benefits of non-pecuniary OSS, may actually be misleading and may lead to long-term costs that are 5% to 20% higher than those of proprietary closed-source software (Giera and Brown, 2004).

In addition to the direct monetary costs of supporting it, non-pecuniary OSS[9] is often seen as riskier than pecuniary software for a number of reasons. First, because a collective of users, rather than a central producer, creates non-pecuniary OSS, there is rarely official technical support for the products. While some users do offer help by creating manuals or answering user questions (Lakhani and von Hippel, 2003), there is no guarantee that a user's question will ever be answered because they do not have a service agreement with any vendor (Woods and Guliani, 2005). Relatedly, although larger OSS foundations, like the Linux Foundation and the Apache Foundation, employ commons-based governance structures (Ostrom, 1990; O'Mahony and Ferraro, 2007), there is no guarantee that the OSS project will be continuously developed and supported. Likewise, even if the project is continuously maintained, there is no guarantee about the features and technical path of future versions (Kogut and Metiu, 2001).

From a security standpoint, the openness of the underlying code in OSS allows anyone to examine it for security vulnerabilities. Although Linus's Law[10] would predict that the open nature of the code would be a benefit from a security perspective, recent widespread vulnerabilities in OSS integral to the operation of the Internet and Linux have shown that these bugs are not always caught early in the development process.[11] Perhaps the most concerning risk of all is the lack of a contractual relationship between a firm using non-pecuniary OSS and any one entity responsible for the development of such software, which leaves the firm with no one to sue when something goes wrong. There are no service level agreements (SLAs) for non-pecuniary OSS, which means the use of such software is riskier than pecuniary software where such agreements exist.

The view of non-pecuniary OSS as a risky decision led to the commonly used phrase "No one ever got fired for buying Microsoft."[12] This phrase became popular in the technology industry as customers were increasingly willing to pay a premium for software from big name firms they could trust.

[9] The focus of this research is primarily on non-pecuniary OSS. The availability of pecuniary products, like Red Hat Linux, which build on non-pecuniary OSS is important, but the risks associated with these products is lower due to the contractual relationship a customer has with the vendor, which greatly mitigates these risks.
[10] Linus's Law is attributed to Eric Raymond (1999), but named after the founder of Linux, Linus Torvalds. Linus's Law states "Given enough eyeballs, all bugs are shallow," which implies that the more people who look at the code, the more likely bugs are to be found and fixed.
[11] The Heartbleed security bug was introduced into the OpenSSL cryptography library in December 2011, and was not noticed and fixed until April 2014. As of May 8, 2014, more than 300,000 public web servers were still vulnerable to the issue (Graham, 2014). The Shellshock security bug was introduced into the Bash Shell in 1992, and was not noticed and fixed until September 2014. The Bash Shell is used in nearly all Unix-style operating systems, including Linux and BSD, the latter of which is the basis of the Mac OS X operating system.
[12] This phrase actually started about IBM in the 1970's, long before OSS. However, it was ported to Microsoft in the 1990's as OSS started to gain traction in the marketplace. Interestingly, IBM later invested heavily in OSS and built some of its products on top of OSS. However, IBM but offered large support contracts and SLAs, removing many of the risks associated with the use of non-pecuniary OSS.

In aggregate, the various risks laid out above could have a negative impact on the productivity of the firm. Formally,

H1a: The usage of non-pecuniary OSS at a firm has a negative impact on firm productivity.

III.B Benefits of Using Non-Pecuniary OSS

Despite all of the risks discussed above, non-pecuniary OSS can also provide a number of benefits to the firms willing to take on these risks. These benefits include reduced upfront costs, collective intelligence of the crowd, and greater flexibility to alter and enhance the code base.

The most salient benefit of using non-pecuniary OSS is the free nature of the software. Although, as discussed above, the actual cost of software is minimal compared to the costs of implementing, the fact remains that firms using non-pecuniary OSS are paying less for their software than their competitors using pecuniary software. However, since this cost reduction is rather small, if there is a measurable positive effect of non-pecuniary OSS on firm productivity, it is likely that the free nature of the software is not the only mechanism driving this effect.

Beyond being free, the crowdsourced nature of non-pecuniary OSS can have an important effect on the quality of software development. A pithy quote from the technology industry helps to illuminate this potential benefit of non-pecuniary OSS – "No matter who you are, most of the smartest people work for someone else." This quote, known as Joy's Law, highlights the fact that regardless of how big and powerful a company is, it can never hire all of the best and brightest people.[13] This is the modern-day interpretation of earlier arguments by von Hayek (1945), who pointed out that knowledge is distributed throughout society and cannot be fully aggregated in one central body. In the software development world, this means that code developed within a closed firm cannot benefit from the intelligence of anyone outside of the firm (Kogut and Metiu, 2001; von Hippel and von Krogh, 2003). Non-pecuniary OSS projects address this problem by allowing anyone to contribute to the development of the underlying code base. Indeed, as mentioned above, nearly 10,000 individuals contribute to the Linux kernel, while less than 1,000 individuals contributed to all of Windows 7 (Schofield, 2008), and only one team of less than 40 people created the Windows 8 kernel (Sinofsky, 2011). Therefore, the use of OSS allows a firm to harness the labor efforts of a wide collective of individuals. Further, as individuals' motives for contributing are primarily intrinsic (Lerner and Tirole, 2002), any benefits by firms using the software can be seen as positive externalities via spillovers from the labor contributions of the crowd.

[13] This statement is from a speech Bill Joy, the co-founder of Sun Microsystems, gave in 1990, and was first mentioned in print by Gilder (1995).

Although collective intelligence and the wisdom of crowds is often associated with completing simple problems, recent research has shown that the crowd can also be successful in solving more complex problems (Woolley et al, 2010; Woolley and Fuchs, 2011; Yi et al, 2012), including software development (von Hippel and von Krogh, 2003). Further, collective intelligence represents an important mechanism for enhancing the knowledge inputs of the firm, which have been shown to contribute to productivity (Hulten, 2010).

The open nature of non-pecuniary OSS has the added benefit of allowing firms to avoid hold-up problems. If a firm relies on closed or pecuniary software built on OSS, it cannot control the path of development and is therefore subject to hold-up by the developer. However, if a firm relies on non-pecuniary OSS and they need a specific function, they can contribute the code themselves (Schwarz and Takhteyev, 2011). This freedom and flexibility allows for the firm to more efficiently use its software once it is deployed within the enterprise (Woods and Guliani, 2005). Further, the open nature of the software leads to a more modular architecture, which has been shown to allow for better integration (MacCormack, Rusnak, and Baldwin, 2006).

Like many investment opportunities a firm must make, the decision to invest in non-pecuniary OSS allows firms that are willing to take on higher levels of risk to obtain higher levels of reward. For many firms, the risks of relying on non-pecuniary OSS are too high and they therefore rely on pecuniary software. However, the firms that are willing to take on the risks associated with non-pecuniary OSS allows them to obtain the benefits of tapping into the collective intelligence of the crowd, leading to productivity spillovers from the free external labor and knowledge[14] that support the non-pecuniary OSS ecosystem as well as the more flexible nature of OSS. Therefore, firms that use non-pecuniary OSS should obtain a net positive effect on productivity:

H1b: An increase in the amount of non-pecuniary OSS used at a firm has a positive impact on firm productivity.

III.C Moderating Effect of Firm Size

Due to differences in capital constraints, it is likely that firm size will play a role in determining the productive impact of non-pecuniary OSS. For very small firms, non-pecuniary OSS can play a critical

[14] While it is true that some firms who use non-pecuniary OSS also contribute back to the creation of these products, even these firms benefit from the external labor contributed by other firms and individuals, which they do not pay for. A deeper analysis of this relationship is left for future research.

role in allowing the IT capability of the firm to ramp up quickly, without expensive outlays for pecuniary software. However, as firms grow, it is likely they will not be able to fully support a non-pecuniary OSS infrastructure themselves, and will therefore rely on external consulting firms to take the place of the support that comes with pecuniary software. On the other hand, larger firms have the capacity for greater economies of scale[15] and can therefore obtain greater returns from their IT investments as well as any consulting activities to help implement an OSS infrastructure. Together, this implies a U-shaped relationship between firm size and productivity returns to non-pecuniary OSS that is high for very small firms, drops for medium sized firms, and increases for larger firms. Due to data restrictions and the sample only consisting of public firms, it is only possible to test the latter portion of this relationship and the former is therefore left for future research. This leads to the following formal hypothesis:

H2: For public firms, the productivity impact of non-pecuniary OSS is more positive (less negative) for larger firms than for smaller firms.

III.D Moderating Effect of Industry

IT related inputs frequently require higher levels of human capital for value extraction. This is especially the case for software that is not supported by a vendor, as is the case with non-pecuniary OSS. Accordingly, prior research (Dewan and Min, 1997; Huang, Ceccagnoli, Forman, and Wu, 2013) has shown that the output elasticity of IT is lower in firms that are in the less human capital intensive manufacturing sector compared to those that in the services sector. Since non-pecuniary OSS is an important piece of the IT ecosystem, this relationship should hold for it as well.

H3: Compared to firms in the manufacturing sector, firms in the services sector will obtain higher (less negative) returns from the use of non-pecuniary OSS.

III.E Additional Moderating Effects

Although some research has speculated a labor-premium for IT workers who understand OSS, this has not yet been shown to be true in all cases.[16] However, since OSS is less frequently used than pecuniary software, the skills to operate and maintain OSS are more niche. Therefore, it is possible that IT workers who are capable of operating and maintaining OSS are of a higher quality than those who are not. Were this true, then the presence of OSS would indicate higher quality labor, which would result in

[15] There may be a concern that if larger firms disproportionately use non-pecuniary OSS, then the use of OSS could simply be proxying for economies of scale. However, it is possible to control for firm size when estimating the effect of OSS on productivity. Controlling for this effect should allow for it to be ruled out as an alternative explanation to the main effect of non-pecuniary OSS.

[16] Hann et al. (2002) and Hann, Roberts, and Slaughter (2013) show that not all participants in OSS receive higher wages in their jobs, but they do find that OSS contributors with managerial responsibilities in the OSS community receive up to an 18% increase in wages.

additional productivity as an indirect consequence of the use of OSS. However, estimating this effect is difficult due to the misattribution issues associated with non-pecuniary IT investments (Greenstein and Nagle, 2014). For example, comparing the elasticity of labor to productivity for firms who use OSS to those who do not may result in a higher return to IT labor for firms using OSS. However, these results would be observationally equivalent to the results if misattribution was the cause because the misattribution discussed above could result in the same shift in elasticity, but for a different reason (namely that the OSS is unaccounted for). To properly disentangle these effects, detailed data on IT labor inputs would be necessary. Such data is not currently available. Therefore, it is not possible test for this effect in the current setting.

Likewise, if non-pecuniary OSS were of a higher quality than its pecuniary counterpart, then firms using OSS would gain an increase in productivity due to the difference in quality of inputs. However, this too is difficult to disentangle from the misattribution effect. If this effect were driving the increase in productivity, comparing the elasticity of IT-software capital between firms who do and do not use OSS would again be observationally equivalent to the case where the misattributed value of OSS increases the coefficient for IT capital. Therefore, testing this relationship is left for future research.

IV. Empirical Methodology

This section describes the empirical methodology employed to test the hypotheses developed above. First, it describes the estimation model, which is consistent with other models of the productivity of IT, but accounts for non-pecuniary digital inputs. Then, it discusses identification concerns due to sample selection and endogeneity as well as the methodologies employed to address these concerns. These methods include inverse probability weighting, instrumental variables, and firm fixed effects.

IV.A Estimation Models

The dataset will measure capital, labor, and various IT inputs. Before describing this data in detail, it is useful to review the model and estimation approach of the paper. In the economics of IT literature, the standard method of estimation is the classic Cobb-Douglas Production function modified to include IT (Brynjolfsson and Hitt, 1996; Dewan and Min, 1997; Tambe and Hitt 2012; Tambe, Hitt, and Brynjolfsson, 2012; Huang, Ceccagnoli, Forman, and Wu, 2013):

$$Y_{it} = {K_{it}}^{\alpha} {L_{it}}^{\beta} IT_{it}^{\gamma} A_{it} \tag{1}$$

where Y_{it} is the production of firm i in time t, ${K_{it}}^{\alpha}$ is the amount of non-IT capital stock, and ${L_{it}}^{\beta}$ is the amount of non-IT labor. IT_{it}^{γ} is the amount of IT capital stock and A_{it} is a firm-specific efficiency multiplier that captures intangible assets such as management skill or institutional knowledge and learning. In earlier literature, IT capital and IT labor have been combined into a single variable; however, more recent literature has shown a differing effect of these two inputs (Tambe and Hitt 2012). Therefore, the primary specification separates the two, but a robustness check is performed with them combined.

$$Y_{it} = {K_{it}}^{\alpha}{L_{it}}^{\beta}ITK_{it}^{\gamma_1}ITL_{it}^{\gamma_2}A_{it} \tag{2}$$

Value-added productivity (VA_{it}) is substituted for sales as a measure of output to remove concerns about trends in the economy or demand shocks (Brynjolfsson and Hitt 2003) and then the log of each side is taken to obtain:

$$\ln(VA_{it}) = \alpha\ln K_{it} + \beta\ln L_{it} + \gamma_1\ln ITK_{it} + \gamma_2\ln ITL_{it} + \varepsilon_{it} \tag{3}$$

Taking the natural log of each side results in coefficients that are equivalent to a firm's output elasticity to a given input. This allows for an interpretation of the coefficients as the percentage change in VA_{it} for a one percent change in the value of the given input. Unobserved differences in firm-level efficiency are captured in the error term. This baseline model is consistent with the most current total-factor productivity models of productivity measurement that account for IT usage (e.g., Tambe and Hitt 2012; Tambe, Hitt, and Brynjolfsson, 2012; Huang, Ceccagnoli, Forman, and Wu, 2013). However, all of these models rely on the assumption that the price of the inputs reveals their importance into production. For example, one-hour of labor that costs \$15 will have less of an effect on output than one-hour of labor that costs \$20. What such models cannot account for is when the value of an input is priced at \$0 (such as non-pecuniary OSS). Such an input is essentially uncounted in such models and can lead to misattribution of production at the macro-level in a variety of ways (Greenstein and Nagle, 2014). To account for this properly, a measure of a firm's utilization of non-pecuniary open source software, $non\_pecuniary\_OSS_{it}$, in a given period is added to the specification. Non-pecuniary OSS must be separated from pecuniary OSS because the latter is already measured by current productivity methods since it has a price.[17] The measurement of

[17] As mentioned above, an important aspect of the OSS movement is the ability to build pecuniary software on top of non-pecuniary OSS. For example, Red Hat Enterprise Linux is built on the open source Linux kernel, but is not free due to the additional functionality and support Red Hat provides. Conversely, a product like Mandrake Linux is both open source and non-pecuniary. Therefore, pecuniary OSS is considered differently than non-pecuniary OSS.

non-pecuniary OSS is described in the data section below. To allow for consistent interpretation, the natural log of this measure is used. This results in the following equation:

$$\ln(VA_{it}) = \alpha \ln K_{it} + \beta \ln L_{it} + \gamma_1 \ln ITK_{it} + \gamma_2 \ln ITL_{it} + \gamma_3 \ln non\_pecuniary\_OSS_{it} + \varepsilon_{it} \quad (4)$$

Using equation 4 as the preferred estimation equation, an estimate of the impact of non-pecuniary OSS usage can be obtained.

IV.B Identification Strategy

In an ideal experiment, one would randomly assign firms from the full population of US firms to use or not use non-pecuniary OSS at varying levels of intensity. However, such an experiment is infeasible and therefore observational data, discussed in the next section, is used. Like all studies of the impact of IT on productivity using observational data, this analysis is subject to both sample selection bias and endogeneity. Sample selection is a potential threat to identification due to the fact that the dataset (discussed below) undersamples firms that use non-pecuniary OSS. This could result in incorrect estimation of coefficients for the population. A second threat to identification is the fact that firms endogenously decide whether or not to use non-pecuniary OSS. If firms that are, for example, better managed are both more likely to use non-pecuniary OSS and have higher levels of productivity, then the relationship between non-pecuniary OSS and productivity could not be interpreted as causal due to simultaneity bias. Further, this could lead to an incorrect estimation of the size of the effect. Both of these concerns prevent a complete answer to the primary question that can be used to make recommendations to managers. Additionally, to understand the determinants of how OSS impacts productivity, a believable baseline must be established. Therefore, the paper employs a number of methods that help to address both of these concerns. These methods allow for the coefficient on use of non-pecuniary OSS to be interpreted in a more causal manner. Further, the coefficient can be interpreted as the impact of not only the non-pecuniary OSS itself, but also the ecosystem of complementarities that are utilized when such software is employed. Such complementarities have been found to play an important role in the impact of IT on productivity (Bresnahan, Brynjolfsson, and Hitt, 2002; Aral, Brynjolfsson, and Wu, 2012; Brynjolfsson and Milgrom, 2012).

Inverse-Probability Weighting

First, inverse-probability weighting (IPW) (Horvitz and Thompson, 1952) is utilized to address the issue of sample selection bias. This increases the consistency of the estimator (Wooldridge, 2007) in a manner similar to Heckman correction (Heckman, 1976, 1979), but with fewer assumptions (Wooldridge, 2002; Young and Johnson, 2009). This is necessary because the dataset (discussed below) undersamples firms that use OSS, which can adversely affect the estimation procedure. IPW also helps address endogeneity concerns and allows for the results to be interpreted as causal, in a manner similar to matching, by balancing the dataset between treatment and control groups to identify the direct effect of the independent variable (Hirano, Imbens, and Ridder, 2003; Hogan and Lancaster, 2004; Cole and Hernan, 2008; Huber, 2013).

IPW is similar to propensity score matching, but allows for full use of all existing observations. This makes IPW more efficient than matching, which drops observations that do not have a close match. The first step is to predict the propensity of a firm to adopt non-pecuniary OSS based on observables. To do this, a Probit function is used to predict the likelihood of treatment (adoption of non-pecuniary OSS) based on observables. In addition to the four primary input variables ($ITK_{it}, ITL_{it}, K_{it}, L_{it}$), the model also uses two constructed variables estimating the number of pecuniary OSS operating systems and closed source operating systems at the firm ($pecuniary\_OSS_{it}$ and $closed_{it}$). These additional variables help to account for the amount of other operating systems used by the firm, which could be an important predictor of non-pecuniary OSS adoption. The propensity function looks as follows:

$$\Pr(T=1) = \alpha \ln K_{it} + \beta \ln L_{it} + \gamma_1 \ln ITK_{it} + \gamma_2 \ln ITL_{it} + \gamma_3 \ln pecuniary\_OSS_{it} + \gamma_4 \ln closed_{it} + \varepsilon_{it} \quad (5)$$

The coefficients from the propensity function are then used to predict the likelihood of a given firm to adopt non-pecuniary OSS, $\hat{T}$. This allows for the construction of a weighting such that firms who have adopted (are treated, $T=1$), are assigned a weight of the inverse of their propensity to adopt, $1/\hat{T}$, and firms who have not adopted ($T=0$), are assigned a weight of the inverse of 1 minus their propensity to adopt, $\frac{1}{1-\hat{T}}$. These weights are then used to adjust the regression results to account for the sample selection bias such that firms who adopt and do not adopt are equally weighted in the regression results. This is similar to a propensity score matching procedure where each adopting firm is matched with a non-adopting firm that has a similar likelihood of adopting, based on observables, but does not require dropping observations that do not have a good match. Therefore, the resulting estimation can be interpreted as a causal effect similar to that of a randomized experiment, but without actually randomizing adoption (Hirano, Imbens, and Ridder, 2003; Hogan and Lancaster, 2004; Cole and Hernan, 2008; Huber, 2013).

Instrumental Variables

Two instrumental variables that exogenously shift a firm's likelihood of using non-pecuniary OSS are used to further address endogeneity concerns. Both instruments are constructed based on the non-pecuniary OSS adoption habits of firms that are similar (in industry or geography) to the focal firm, but whose adoption decision is exogenous to the firm itself. Such firms face supply conditions similar to the focal firm and are therefore likely to be affected by similar shocks to supply. This is similar to instruments that have been used for other studies of the digital economy (e.g., Forman, Goldfarb, and Greenstein, 2005). Importantly, most firms in the sample were founded before OSS diffused widely. Therefore, the firm's decisions to operate in a specific industry and locate in a specific geography are independent of OSS adoption patterns.

The first instrument is a measure of the mean non-pecuniary OSS usage of other firms within a given firm's 2-digit Standard Industrial Classification (SIC) industry within the same year. The amount of non-pecuniary OSS usage by the firms in a firm's same industry exogenously affects that firm's propensity for using non-pecuniary OSS primarily through labor. Employees of firms in a given industry are likely to interact with other firms in their industry through conferences and job movement. Therefore, in industries where there is widespread use of non-pecuniary OSS, a given firm is more likely to use non-pecuniary OSS.

The second instrument is a measure of the mean non-pecuniary OSS usage by other establishments within a given firm's county within the same year. Similarly to industry, geographically close firms also face supply conditions similar to the focal firm. Specifically, the availability of IT labor familiar with OSS in a local area is likely to affect the firm's decision to adopt OSS. The availability of this labor is greater in areas where other firms are already using OSS. Therefore, the amount of non-pecuniary OSS usage by the firms in a firm's local geography may exogenously shift that firm's propensity for using non-pecuniary OSS, but does not directly affect the firm's productivity level.

Panel Data Methods

Finally, since the data is panel data, firm fixed effect models can be used to estimate the effect at individual firms. However, because an individual firm is likely to only change from not using non-pecuniary OSS to using it once, fixed effects are only used when looking at continuous adoption of non-pecuniary OSS. This helps identify the effect as it relies on within-firm variation in usage of non-pecuniary OSS rather than across firm variation. This method is not used as the primary identification approach because the changes from year to year within the firm are often not that great, and therefore the results are less well-identified than other methods. Further, to control for unobserved time and industry trends, the models uses year fixed effect and industry fixed effect at the 1-digit SIC level. The latter is only used when the 2-digit SIC instrument is not in use to avoid perverse instrumentation. The combination of these approaches helps eliminate unobserved firm, time, or industry effects that may bias the results. In aggregate, the identification strategy adds significant weight to a causal interpretation rather than just a correlational one.

V. Data

The data breaks into two primary areas: OSS usage and financial statements, both of which are at the firm level. Data on which firms are using OSS comes from the Harte Hanks IT Survey – a survey of IT usage by multiple sites at over 10,000 firms from 2000-2009. This database is used frequently in studies of the impact of IT on firm-level productivity (Brynjolfsson and Hitt, 2003; Forman, 2005; Forman, Goldfarb, and Greenstein, 2005; Forman, Goldfarb, and Greenstein, 2008; Tambe, Hitt, and Brynjolfsson, 2012; Huang, Ceccagnoli, Forman, and Wu, 2013; McElheran, 2014). The Harte Hanks survey asks site-level IT managers questions about the types of IT (both hardware and software) used at the site as well as the number of IT employees at the site. In cases where Harte Hanks does not interview all sites within a firm, the average values for sites that are interviewed is assigned to sites that are not interviewed. This allows for the construction of firm level values that account for all sites within the firm.

The Harte Hanks data is augmented with detailed firm financial data. In particular, firm expenditures on labor (IT and non-IT) and capital (IT and non-IT) as well as firm revenues and costs of materials. For public firms, this information is available via Standard and Poor's Compustat database. The firm's stock ticker symbol is used to match the Harte Hanks data to the Compustat data. In this manner, sites within the Harte Hanks database that are owned by different firms in different years (e.g., through mergers or acquisitions) will be associated with the correct parent firm and therefore the correct financial data. Although the Harte Hanks database contains information on over 10,000 firms, the final sample uses only public firms as the model requires additional financial information filed in the firm's 10-K. This reduces the sample size to 1,850 firms, and indicates that the results can best be applied to public firms. The sections below detail how these two datasets are used to construct the variables discussed in the previous section. All monetary values are converted to 2009 dollars using an appropriate deflation index and are reported in millions of dollars.

Value-Added (VA_{it})

The dependent variable is constructed using a method consistent with prior literature (e.g., Dewan and Min, 1997; Brynjolfsson and Hitt, 2003; Huang, Ceccagnoli, Forman, and Wu, 2013). First, yearly operating costs (XOPR in Compustat) are deflated by the BLS Producer Price Index by stage of processing for intermediate materials, supplies, and components. Then deflated IT labor and non-IT labor (defined below) are both subtracted from the operating costs. The result is then subtracted from yearly sales (SALE in Compustat) deflated by the BEA Gross Domestic Product Price Index for gross output for private industries.

IT Capital (ITK_{it})

Most prior literature in the field constructs a combined measure of IT Capital that includes both the value of IT hardware at the firm and three times the value of IT labor at the firm due to the importance of IT labor being used for internal software development efforts, the result of which is a capital good (Brynjofsson and Hitt, 1996; Hitt and Brynjofsson, 1996; Dewan and Min, 1997; Huang, Ceccagnoli, Forman, and Wu 2013).[18] However, recent literature has shown that IT labor can have a separate effect from IT capital (Tambe and Hitt, 2012). Therefore, the primary analysis uses separate IT capital and IT labor variables. Later, the combined variable is tested for robustness purposes and the results are shown to be consistent.

To calculate IT Capital, the market value of the IT stock is estimated by multiplying the number of PCs and Servers at the firm (from Harte Hanks[19]) by the average value of a PC or Server that year from The Economist Intelligence Unit Telecommunications Database. The BEA Price Index for computers and peripherals is then used to deflate this value. This method is consistent with prior work in this area (e.g., Brynjofsson and Hitt, 1996; Huang, Ceccagnoli, Forman, and Wu 2013). Because the costs of the IT Capital are being imputed, a robustness check using the raw number of PCs and servers will be run and shows that the results are consistent.

---

[18] Ideally, the portion of the IT budget that is spent on software in addition to hardware would be included. However, software expenditures are combined with other capital expenditures in firm 10-K reporting. Therefore, while purchased software cannot be separated from other firm purchases, the cost of such software is captured in the non-IT Capital variable. Further, internal software development efforts will be captured in the IT Labor variable. This methodology is consistent with prior literature (e.g., Brynjofsson and Hitt, 1996; Huang, Ceccagnoli, Forman, and Wu 2013). Additionally, the high correlation between purchased software and hardware expenditures helps to mitigate concerns about not having software expenditure data.

[19] For most firms, Harte Hanks only surveys a sample of the sites within the firm. In such cases, the average number of PCs and Servers at the sites that are in the survey is multiplied by the total number of sites in the firm to obtain the total number of PCs and Servers in the firm. The same procedure is used for calculating the number of IT employees and the number of each type of operating system at the firm.

IT Labor (ITL_{it})

The value of IT labor is calculated by taking the number of IT workers at each firm (from Harte Hanks[20]) and multiplying by the mean annual wage for all Computer and Mathematical Science Occupations[21]. The BLS Employment Cost Index for wages and salaries for private industry workers is then used to deflate this value. Because the cost of the IT labor is being imputed, a robustness check using the raw number of IT employees will be run and shows that the results are consistent.

Non-IT Capital (K_{it})

The K_{it} variable is constructed by taking the yearly Gross Total Property, Plant and Equipment (PPEGT in Compustat), deflating it by the BLS price index for Detailed Capital Measures for All Assets for the Private Non-Farm Business Sector, and then subtracting the deflated value of IT Capital (defined above).

Non-IT Labor (L_{it})

Non-IT Labor is constructed using the total number of employees at the firm (EMP in Compustat) and subtracting the number of IT employees (from Harte Hanks) to obtain the total number of non-IT employees. This is then multiplied by the mean annual wage of all occupations[22] that year. The BLS Employment Cost Index for wages and salaries for private industry workers is then used to deflate this result. This method of calculation is consistent with prior studies on IT productivity (Bloom and Van Reenen, 2007; Bresnahan, Brynjolfsson, and Hitt, 2002; Brynjolfsson and Hitt 2003). However, because the cost of labor is being imputed, a robustness check with the raw number of non-IT employees is run and shows that the results are consistent.

Non-Pecuniary Open Source Software Usage

To measure the intensity of non-pecuniary OSS usage at the firm, the number and type of operating systems used at the firm is measured. Although operating systems are certainly not the only non-pecuniary OSS used at the firm, they are important and frequently indicate the wider use of non-pecuniary OSS. Further, the Harte Hanks survey asks firms what type of operating systems they use, but does not always capture other types of non-pecuniary OSS. Because this only captures non-pecuniary OSS operating systems, the dataset necessarily underestimates the amount of non-pecuniary OSS used at the

[20] Harte Hanks reports the number of IT employees at each site as a range so the average value of the range is used. The ranges are 1-4, 5-9, 10-24, 25-49, 50-99, 100-249, 250-499, and 500 or More.
[21] Obtained from the Bureau of Labor and Statistics: http://www.bls.gov/oes/2009/may/oes_nat.htm#15-0000.
[22] Obtained from the Bureau of Labor and Statistics, for example the data for 2009 can be found here: http://www.bls.gov/oes/2009/may/oes_nat.htm#00-0000.

firm. Therefore, the estimates should be considered a lower bound on the impact of non-pecuniary OSS to the firm.

In addition to constructing a measure of non-pecuniary OSS operating systems, measures of pecuniary OSS and closed-source operating systems are also constructed for use in predicting the propensity of a firm to adopt non-pecuniary OSS. These three measures ($non\_pecuniary\_OSS_{it}$, $pecuniary\_OSS_{it}$, and $closed_{it}$) are constructed by calculating the total number of each type of operating system at the firm (from Harte Hanks). The Harte Hanks data does not report the precise number of operating systems in use at a given firm. It does, however, report the different types of operating systems used at each site at the firm. These operating systems are classified into three categories: non-pecuniary OSS, pecuniary OSS, or closed source. Table 1 shows the OSS operating systems in the dataset.[23] All other operating systems are labeled as "closed". Harte Hanks also reports whether each operating system is for a PC or a server as well as the total number of PCs and servers at each site. Therefore, for each site, the number of PC operating systems is evenly split over the total number of PCs at the site. The same is done for servers. This yields an estimate of how many instances of a given type of operating system exist at the site. This is then aggregated to the firm level and divided by the number of sites at the firm in the Harte Hanks database to obtain an average per site. Finally, this average is multiplied by the total number of sites in the firm to obtain a firm-wide imputation of the number of each type of operating system. As the resulting numbers are estimates, the analysis begins by only using a binary indicator of the presence of non-pecuniary OSS at the firm. The estimated number of operating systems will then allow for a more granular interpretation of the primary effect.

Insert Table 1 Here

Because the number of operating systems in any of the three categories can potentially be zero (e.g., that category of operating system is not in use at the firm), one is added to the number of operating systems in each category before taking the natural log as the natural log of zero is undefined. Although there are many firms that have zero non-pecuniary and pecuniary OSS operating systems, there is a high degree of skewness in these numbers (as shown in the descriptive statistics below). Therefore, adding a

[23] Although some non-pecuniary OSS operating systems, such as Debian, are offered at a nominal pecuniary price by third-party vendors for the convenience of the distribution being pre-loaded on a CD or DVD, they are included in the non-pecuniary column as the full distribution is downloadable for free via the distribution's website. Additionally, although Apple's Mac OS X is built on BSD, it behaves more like a closed operating system than one that is pecuniary, but built on OSS, like Red Hat. Robustness checks were run against this assumption with no change to the primary results.

one before taking the natural log should not significantly bias the results.

Table 2 shows the descriptive statistics of the firms in the dataset. There are 12,244 firm/year observations from 1,850 firms in the dataset.[24] The ranges vary greatly for all variables and demonstrate the breadth of the firms in the sample. This breadth allows for results that are more generalizable than many other studies of this kind, which only focus on Fortune 1000 companies. However, due to the Harte Hanks sampling methodology, larger firms are overrepresented in the sample and very small firms (e.g., startups) are not in the sample. Additionally, because of the reliance on 10-k data for financial information, all firms in the sample are public firms, which tend to be medium or large. For example, as shown in Table 2, the smallest company in the sample (Matec Corp.) had sales of $2.7 million in its lowest selling year. Comparatively, the largest firm (Exxon Mobil Corp.) had sales of $425 billion. Therefore, results should be interpreted as applying to medium and large firms. The firms in the dataset also have a wide range of the type and intensity of IT use. The mean number of closed source operating systems at a firm is 5,026.755 while the mean number of non-pecuniary OSS and pecuniary OSS operating systems are much lower at 182.253 and 181.172, respectively. Looking deeper into the data, there are 3,527 observations where firms use at least one non-pecuniary OSS operating system. For these 3,527 observations, the average number of non-pecuniary OSS operating systems is 632.635. 7,341 observations use no OSS (pecuniary or non-pecuniary) at all. Only 10 observations use exclusively OSS (pecuniary or non-pecuniary).

Insert Table 2 Here

Table 3 shows the correlation matrix. As to be expected, K_{it} and L_{it} have a fairly high correlation with value-added productivity since they are the primary inputs into the production function. Additionally, it is notable that the correlations between non-pecuniary OSS and the other two types of operating systems, pecuniary OSS and closed, are fairy low, while the correlation between pecuniary OSS and closed is comparatively high. Table 4 shows the breakdown of observations by industry. While 48% of the observations are from the manufacturing industry, there is also good representation from other key industries, such as finance (14%), services (14%), and trade (11%). Further, Table 4 shows the percentage of firms within the industry that use non-pecuniary OSS or any type of OSS operating system. The

[24] This results in an average of 6.6 observations per firm. The panel is unbalanced because Harte Hanks does not survey every firm in every year. However, this is still a large enough number of observations per firm to conduct a fixed effect analysis and does not adversely affect the pooled analysis.

percentage of firms in an industry using non-pecuniary OSS varies between 17.82% and 34.78%, with an average of 28.81% and has no major outliers. The percentage of firms in an industry using any OSS varies between 26.49% and 71.43%. However, this maximum should be considered an outlier because SIC 0 has a low number of observations. Therefore, the more realistic range is between 26.49% and 47.20%, with an average of 40.04%.

Insert Tables 3 & 4 Here

VI. Results and Discussion

This section presents the results of the empirical analysis and discusses the interpretation of these results in light of the hypotheses. First, basic three-factor productivity results are compared to those of other studies to confirm the consistency of the data and methods with prior research. Then, the results from the propensity score analysis, the first stage of the inverse-probability weighting method, are presented. These weightings are then used to obtain baseline regression results for the impact of non-pecuniary OSS on firm productivity. An instrumental variable approach is then employed to enhance the causal interpretation of these results. A number of moderator and split-sample analyses are then conducted to better understand the firm characteristics that are important determinants of the primary results. Finally, several robustness checks are considered to confirm that various assumptions are not driving the results.

VI.A Three-Factor Productivity Analysis

Before delving into the results on open source usage, the results of the baseline regression are presented to compare the elasticities of the three main productivity inputs with other existing studies. To properly achieve this comparison, the combined measure of IT Capital that is consistent with prior studies is used, rather than the separated measures used in the primary analysis. Table 5 shows the results of the basic three-factor productivity analysis. Models 1-3 use Ordinary Least Squares (OLS) regression with increasingly restrictive fixed effects, while Model 4 uses panel regression with firm fixed effects and Model 5 uses panel regression with random effects. For all models, the standard errors are robust and clustered by firm to account for any serial correlation in the error terms since the dataset contains multiple observations of the same firm over different time periods (Angrist and Pischke, 2009; Imbens and Kolesar, 2012). The high R^2 values are characteristic of such productivity studies. The confidence intervals of the coefficients in models 4 and 5 overlap with those of Huang, Ceccagnoli, Forman, and Wu (2013), whose methodology this study most closely resembles. However, the coefficients on non-IT capital are slightly higher than theirs, likely because their sample size is only companies in the Fortune 1000, while this study casts a wider net. Further, the column 4 coefficient on IT capital is very similar to that of Brynjolfsson and Hitt (2003) in their 1-year difference model with year and industry controls. The coefficients in column 4 are also very similar to the fixed effect estimate of Tambe and Hitt (2012), although the IT capital coefficient is slightly lower, likely because they are calculating their coefficient based solely on IT labor. These similarities help to add support to the validity of the dataset used in this study. The similarities also imply that if support is found for the hypotheses above, then the estimates in the prior literature are likely suffering from either attribution or omission bias.

Insert Table 5 Here

VI.B Propensity to Adopt Non-Pecuniary OSS

As discussed previously, propensity scores are used to estimate the likelihood a firm adopts non-pecuniary OSS based on observables. The presence of non-pecuniary OSS in a firm-year observation is predicted based on the four primary input variables ($ITK_{it}, ITL_{it}, K_{it}, L_{it}$) as well as the two constructed variables estimating the number of pecuniary OSS operating systems and closed source operating systems at the firm ($pecuniary\_OSS_{it}$ and $closed_{it}$). These additional variables help to account for the technology usage of the firm. This method relies on firm observables to predict the propensity to adopt non-pecuniary OSS. Traits of the firm that are unobservable through a firm's financial reports, such as management quality, may also have an impact on the firm's propensity to adopt. However, as will be shown in a robustness check in Section VI.F, for a subset of the firms in this study that are also in the World Management Survey dataset (Bloom, Sadun, and Van Reenen, 2012), management quality does not predict use of non-pecuniary OSS.

The results of the propensity estimation are shown in Table 6. These results show there is a significant negative coefficient on ITK_{it} indicating that firms who spend more on IT Capital are less likely to adopt non-pecuniary OSS. This supports the theory that non-pecuniary OSS is a substitute for other IT, rather than a complement. However, there is a positive and significant coefficient on ITL_{it}, indicating that firms with larger IT staffs are more likely to adopt non-pecuniary OSS. Although interesting, it is difficult to interpret these results as causal due to the inherent endogeneity and potential omitted variable bias. However, they allow for the construction of the inverse-probability weighting discussed above, such that the remaining results are adjusted for sample bias and can be interpreted in a more causal manner.

Table 7 shows the resulting improvement of the balance in the sample after applying the IPW. Panel A shows the covariate balance without weighting. The t-statistics indicate that the adopting firms in the sample are significantly different from those that are non-adopters when comparing the four primary production inputs. Panel B shows the covariate balance after weighting. Here, the balance is much better and for all inputs except IT Capital, the balance drastically improves. While the IT Capital balance is still concerning, the use of weighting is primarily to deal with sample selection. This motivates the additional use of an instrumental variable approach. Although IPW improves the ability to interpret the resulting coefficients as causal, the instrumental variable approach helps to diminish any concerns of the covariate balance in the weighted sample presenting a threat to causal identification.

Insert Tables 6 & 7 Here

VI.C Baseline Regression Results

Table 8 presents the estimation results using pooled OLS regressions without instrumental variables but with inverse-probability weighting. Columns 1 and 2 show the results when considering non-pecuniary OSS as a binary variable – do firms use non-pecuniary OSS or not. Column 1 shows a positive and significant coefficient of 0.059 on the use of non-pecuniary OSS. However, this effect becomes not significant when adding in the industry fixed effect in Column 2. These results are encouraging, although not conclusive due to the lack of granularity over how much non-pecuniary OSS a firm uses. Columns 3 and 4 show results for a similar analysis, but use a continuous measure of how many non-pecuniary OSS operating systems a firm uses. Here, the coefficient is slightly smaller than the binary coefficient, which makes intuitive sense, but it remains stable and significant when adding in the industry fixed effect. Columns 5 and 6 show a similar, although slightly larger, effect when considering only firms who have adopted at least one non-pecuniary OSS operating system. By only using firms that have adopted non-pecuniary OSS, the results in these two columns can be interpreted in a slightly more causal manner than the prior results as they compare firms who have all made the decision to adopt non-pecuniary OSS and therefore estimate the impact of the amount of non-pecuniary OSS adopted on productivity. However, caution must be applied in interpreting any of the results in Table 8 as causal as they only rely on IPW for dealing with endogeneity. The results in the following section use IPW as well as instrumental variables to additional add support for a causal interpretation.

Insert Table 8 Here

VI.D Instrumental Variable Regression Results

Having found a positive and significant result in the baseline regressions, the instrumental variables discussed above are now used in a two-stage least-squares framework to help further address endogeneity concerns. The results of this analysis are shown in Table 9. The first-stage F-statistics are above 10 for all models, adding support to the choice of instruments. Columns 1 and 2 show the results when pooling observations and considering adoption of non-pecuniary OSS in a binary manner. These columns show a larger coefficient on the binary usage of non-pecuniary OSS that is highly significant both when using only the industry instrument (column 1) and when using both instruments (column 2). Likewise, when considering adoption in a continuous manner, columns 3 and 4 show strong positive coefficients on the amount of non-pecuniary OSS used by the firm. Since the dependent variable is a natural log, the coefficient on $non\_pecuniary\_OSS_{it}$ in column 4 indicates that a 1% increase in the use of non-pecuniary OSS results in a .073% increase in productivity (as measured by value-added). The average value added for the firms in the sample is $1.846 billion; this indicates that a 1% increase in the number of non-pecuniary OSS operating systems leads to a $1.35 million increase in production output for the average firm. This effect is more than double the size of the coefficient on all IT capital found in columns 4 and 5 of Table 5. The negative coefficient on IT Capital (ITK_{it}) is characteristic of such analyses (Huang, Ceccagnoli, Forman, and Wu, 2013) due to the high level of correlation between IT related variables.[25] Column 5 reports the results when using a firm fixed-effect specification such that it is measuring the within firm variation of non-pecuniary OSS usage. The coefficient is again positive and statistically significant. Together, these results add significant support for H1b rather than H1a, indicating that the adoption of non-pecuniary OSS has a positive impact on firm productivity. Although the size of the coefficient may at first appear too large, it is important to recognize that the use of non-pecuniary OSS captures an ecosystem of complimentary organizational practices. The importance of such complementarities has been identified in the literature before (Bresnahan, Brynjolfsson, and Hitt, 2002; Aral, Brynjolfsson, and Wu, 2012; Brynjolfsson and Milgrom, 2012), although it is know to be difficult to fully tease them apart (Athey and Stern, 1998). Notably, the coefficients on non-pecuniary OSS are larger when using the IV methodology, indicating that overlooking the endogeneity concerns discussed above biases the baseline regression results towards zero. This is not surprising because of the geographic and

[25] This is especially the case when using the continuous measure of non-pecuniary OSS operating systems as the number of operating systems and the number of computers is highly correlated. Since the IT Capital variable is not instrumented in this estimation, it acts as a control and therefore the negative coefficient should not be interpreted as causal.

industry differences that can effect the technology decisions of the firm.

Insert Table 9 Here

VI.E Moderators and Split-Sample Analysis

After establishing the primary effect, the preferred specification (the pooled instrumental variable analysis with the continuous measure of non-pecuniary OSS) is used to calculate various moderator and split-sample results to better understand the determinants of the main effect. For specifications that include an interaction term, the interaction of the two instruments with the moderator is also used to ensure a causal interpretation is still plausible. Table 10 shows the results of this analysis. Column 1 shows the effect of using open source interacted with the size of the firm, measured by the natural log of yearly employees. A positive coefficient on the interaction term indicates a positive relationship between firm size and the effect of OSS usage on firm productivity.[26] This finding adds support for H2.

Columns 2 and 3 break down the analysis by industry showing the manufacturing sector (column 2) and the services sector (column 3). Consistent with H3, these results show that services firms have a much greater output elasticity for non-pecuniary OSS than manufacturing firms. Interestingly however, when lagging the use of non-pecuniary OSS by one year, the coefficient for manufacturing firms becomes positive, but not significant. When lagging usage by two years, the coefficient for manufacturing firms becomes positive and significant at the 10% level, indicating that non-pecuniary OSS can also have a positive impact on firms in the manufacturing sector, it just takes longer for these benefits to accrue.[27] Column 4 shows the analysis when removing firms in the finance industry (SIC code 6) as their financial reporting methods often differ from other types of companies. However, removing these firms does not significantly alter the main results, indicating that the main effect is not being driven by financial reporting methods. Column 5 shows the analysis when removing firms in the agriculture and mining industries, as their use of IT differs from most other industries. However, removing these firms does not significantly alter the main results.

Finally, columns 6-8 consider the importance of IT at the industry level. Jorgenson, Ho, and Stiroh (2005), show that the importance of IT to productivity is higher in industries that are either IT-

[26] As mentioned above, the dataset focuses on medium to large public firms, so small firms in this sample are still larger than many private firms or startups.
[27] The results of this lagged analysis are not included to save space. However, they are available from the author upon request.

producing or IT-using when compared to industries that are neither. Columns 6-8 separate the industries into these three categories based on the same industry classification as Jorgenson, Ho, and Stiroh (2005). The baseline analysis for this breakdown was inaccurately measured due to large standard errors, and therefore a one-year lag of the use of non-pecuniary OSS, as well as the instrumental variables, is used. The full impact of IT often takes longer than one year to materialize (Brynjolfsson and Hitt, 2003). This phenomenon is explored further in the next section. As seen by the coefficients in columns 6-8, non-pecuniary OSS has a strong effect on the productivity of IT-using and IT-producing industries, while it appears to have no effect on firms in neither of those groups. This is consistent with the findings in Jorgenson, Ho, and Stiroh (2005). Interestingly, the point estimate for the impact of non-pecuniary OSS is higher for firms in IT-using industries than it is for IT-producing industries. However, the confidence intervals overlap so it is difficult to interpret this in any meaningful way.

Insert Table 10 Here

VI.F Robustness Checks

As with any empirical estimation, the estimation strategy is founded on a number of assumptions that may affect the outcome of the analysis. Therefore, this section considers a number of robustness checks against some of these assumptions to ensure they are not directly leading to the results discussed above. Due to space constraints, only the results of the preferred specification (the pooled instrumental variable analysis with the continuous measure of non-pecuniary OSS) are shown for each robustness check in Table 11.

Production Input Assumptions

As mentioned in Section V, IT Labor and IT Capital are separated, rather than including them in a combined variable, as is standard in the economics of IT literature (Brynjofsson and Hitt, 1996; Hitt and Brynjofsson, 1996; Dewan and Min, 1997; Huang, Ceccagnoli, Forman, and Wu 2013). Therefore, to confirm the separation of these variables does not have an impact on the results, a combined IT variable consistent with the prior literature is considered. This variable consists of the deflated value of IT Capital plus three times the deflated value of IT Labor. Using this combined variable instead of the separate IT Capital and IT Labor variables, both the baseline and the IV regressions are re-estimated. In all cases, the results for the coefficient on non-pecuniary OSS were substantively similar. In all cases, the coefficient is consistently positive and significant, and in almost all cases the confidence interval of the coefficients overlaps when comparing the results for the combined IT variable and the separated variables. The results of this robustness check with the preferred specification are shown in column 1 of Table 11. This adds support to the robustness of the primary results against concerns that using the more granular separation of the two variables drove the results.

Average prices and wages for a given input in a given year are used to impute the costs of many of the primary input variables. As discussed in Section V, the IT Labor, non-IT Labor, and IT Capital variables are all imputed based on the raw number of IT employees, non-IT employees, and computers and the yearly average for IT worker wages, non-IT worker wages, and prices for PCs and servers, respectively. To confirm that the results are robust against the assumption that these averages apply to all firms in a similar manner, all regressions are re-run using only the raw numbers for the inputs, rather than the imputed cost of each input. Again, in all cases, the coefficient on non-pecuniary OSS is consistently positive and significant, and in most cases the confidence interval of the coefficient overlaps when comparing the results for the imputed cost variables with those of the raw input variables. The results of this robustness check with the preferred specification are shown in column 2 of Table 11. This adds support to the robustness of the primary results against concerns that imputing the cost drove the results.

The inclusion of non-pecuniary OSS operating systems as a raw number in the regressions makes comparing the size of the effect to other inputs un-intuitive, as the other inputs are all measured in dollars. Therefore, the price of a pecuniary operating system, Microsoft Windows, for that year is used to estimate the value of each non-pecuniary OSS operating system.[28] The BEA computer price index is then used to deflate this value. The cost of replacing the non-pecuniary OSS operating systems at each firm with this pecuniary alternative is then estimated in a method similar to that of Greenstein and Nagle (2014), who perform the same estimation for the non-pecuniary OSS web server Apache. Although there is wide variance in the functionality and quality of operating systems, this rough estimate allows for a comparison of dollars to dollars, rather than dollars to number of operating systems. The result is shown in column 3 of Table 11. The resulting coefficient is significant and positive and is greater than the coefficient for IT Capital found in the more restrictive models in columns 4 and 5 of Table 5. This is encouraging as it indicates that the value of non-pecuniary OSS is on a similar order to that of other IT-related inputs. However, its effect is greater than these less risky inputs, adding further support to the primary hypotheses.

[28] Prices for Microsoft Windows are based on the latest version of Windows in a given year and are gathered from various industry publications at the time of release.

Timing of OSS Factors

Prior studies have shown that the full effect of IT on productivity can take 5-7 years to be realized due to the organizational changes that must occur for the full effect of IT to be realized (Brynjolfsson and Hitt, 2003). Therefore, the analysis in column 4 shows the preferred specification with a 6-year lag of the amount of non-pecuniary OSS used. To account for this lag, the instruments are lagged by 6 years as well. The coefficient on lagged non-pecuniary OSS is larger than in the preferred specification, although the confidence intervals overlap. Similar results occur for lags up to 6 years, but are not show due to space constraints. These results indicate that investments in non-pecuniary OSS in year's past have an effect that spills over to the productivity of the current year.

Relatedly, the implementation of the instrumental variables is such that the instruments are constructed for the same year as the observation being estimated. It is quite possible that it is the adoption of non-pecuniary OSS in prior years by other firms in the same county or industry that influences the likelihood of a given firm to adopt. Therefore, a robustness check is run using a 1-year lag of both instruments, rather than the same year. The results for the preferred specification are shown in column 5 of Table 11. The resulting coefficient on non-pecuniary OSS is positive and significant and the confidence interval overlaps with that of the coefficient from the primary specifications. Therefore, the primary results are robust to this concern.

Estimation Methodology

There may be a concern that all results shown from the IV regressions have inverse probability weighting applied. To confirm that the results from the IV regressions are not only the result of the weighting, column 6 in Table 11 shows the results of the primary specification with no weighting, but with both instruments. The results show that the coefficient on non-pecuniary OSS is still positive and significant. Further, the confidence intervals of this coefficient overlap with those of the primary specification, indicating that the use of IPW is not interfering with the application of the instruments.

There may also be a possible concern that the results are driven by local industry agglomeration or knowledge spillovers, which have been shown to have an important effect on innovation (Jaffe, Trajtenberg, and Henderson, 1993; Furman, Porter, and Stern, 2002). This is of a particular concern as the second IV is based on county. Therefore, column 7 in Table 11 shows the results of the primary specification with a county-fixed effect and without the county IV. The coefficient on non-pecuniary OSS continues to be positive and significant, adding support to the robustness of the primary results against such concerns.

There may also be concerns with the use of IPW rather than a more standard matching methodology. Therefore, as a robustness check, I also use the nearest-neighbor matching methodology of Abadie and Imbens (2006). Using a nearest-neighbor match based on all observables used in the prior regressions, I construct a matched sample based on the binary use of non-pecuniary OSS. I then use this matched sample to estimate the sample average treatment effect (SATE) at 0.165 with a standard error of 0.025. This positive and statistically significant coefficient again offers support for the validity of my primary results.

Identification of OSS Effect

Concerns may arise that the effect found in the primary analysis is just that of an accounting nature, that the results are simply because non-pecuniary OSS is free and therefore it is not accounted for. While this may be true to some degree, the TCO literature discussed above has argued that the actual cost of software is so small compared to the implementation costs (hardware and labor), that it is almost negligible. Therefore, any residual effect found in this analysis should not be primarily due to an accounting issue, but instead to the firm benefiting from spillovers due to crowd intelligence. However, to further rule this alternative explanation out, an analysis is run that includes both non-pecuniary OSS and pecuniary OSS. One would expect that the coefficient on such a variable may be slightly smaller than non-pecuniary OSS alone as many of the risks, and likewise the benefits, associated with pecuniary OSS are lower. This is indeed what is found in column 8. The coefficient on the combined OSS is slightly lower than that on non-pecuniary OSS alone, although the confidence intervals overlap.

An additional concern may be that the use of non-pecuniary OSS is correlated with unobservable managerial practices that are likely to increase productivity. Although the primary data set does not allow ruling out such simultaneity bias, additional data from the World Management Survey (Bloom, Sadun, and Van Reenen, 2012) is used to confirm this is not driving the results.[29] The World Management Survey (WMS) asks a wide array of firms about their management practices every few years starting in 2004. 183 of the 1,850 firms from the main dataset for this paper appear at least once in the WMS dataset. Although this is far from a complete overlap, it does represent nearly 10% of the firms in the dataset. There are 247 firm/year observations that overlap from two datasets. To increase the amount of overlap, results from the WMS data are carried one year forward and one year backwards, except where the firm is actually surveyed in consecutive years. For example, the results from a firm surveyed in 2004 are carried to both

[29] The author is grateful to Nick Bloom, Raffaella Sadun, and John Van Reenen for allowing access to the WMS dataset.

2003 and 2005. This allows for the expansion of the number of firm/year observations to 650. Although this method assumes firm management practices do not change significantly within a one-year time window, this assumption is consistent with results from firms that were surveyed multiple times. The firms that appear in both datasets are used to test the correlation between management practices and the use of OSS (both pecuniary and non-pecuniary). The results indicate that an increase in the quality of a firm's management practices is uncorrelated with the decision to use non-pecuniary or pecuniary OSS.[30] This result is consistent when using the 247 firm/year direct observations or the 650 imputed observations. Further, it is consistent when examining the binary or continuous use of OSS, and when controlling for the production inputs of the firm ($ITK_{it}, ITL_{it}, K_{it}, L_{it}$). Indeed, when running a regression of the binary or continuous usage of OSS on production inputs and the WMS measure of management quality, the coefficient on the latter is negative, but not significant. This indicates that the quality of a firm's management is uncorrelated with the firm's decision to use OSS. Therefore, concerns of simultaneity bias due to management quality can be alleviated.

Insert Table 11 Here

VII. Conclusion

The results of this study show that the use of non-pecuniary OSS does indeed have an impact on the productivity of the firm, and that this impact is positive. The effect is consistently positive in all specifications that account for sample selection and endogeneity via inverse probability weighting, instrumental variable analysis, and firm fixed effects. This effect exists when considering the use of non-pecuniary OSS at both a binary and continuous level such that both the usage and the amount of non-pecuniary OSS used positively affect productivity. The effect is still positive and significant when considering within firm variation through a firm-fixed effect model. Because the use of non-pecuniary OSS is only measured via operating systems, other firm investments in non-pecuniary OSS are not captured. Therefore, the true effect of all non-pecuniary OSS is likely greater than the effect found in this study.

Digging further into the main effect by exploring various split sample analyses reveals that larger firms (based on employees) gain a larger benefit from increased usage of non-pecuniary OSS. However, due to the sample construction, even the smallest firms are still rather large. It is quite possible, even likely, that the use of non-pecuniary OSS has an even larger effect for firms that are very small and

[30] The full tables of results are not shown to save space, but are available from the author upon request.

therefore capital-constrained. However, due to data constraints, the effect of non-pecuniary OSS on small companies, technology related start-ups in particular, is left for future research. Finally, consistent with other literature on the productivity of IT, this study finds that services firms have a higher output elasticity of non-pecuniary OSS than manufacturing firms. These findings, as well as the risks associated with adopting non-pecuniary OSS discussed above, help explain why not all firms are using what, at first glance, appears to be a free input.

Although endogeneity is always a concern in productivity studies, this study takes many steps to help rule out this bias to allow for the results to be interpreted in a causal manner. All of the regression results use fixed effects for year. This helps to rule out alternative explanations due to trends over time. In all specifications inverse probability weighting is used to generate an analysis similar to that of a matched sample strategy. With this statistically rigorous matching method, the primary finding of a positive causal effect of non-pecuniary OSS usage on productivity holds. Additionally, in some specifications firm fixed effects are used so that a firm is compared with itself over time. Finally, the use of instrumental variables allows for a proper identification of the effect within this panel framework. As mentioned above, the complete identification strategy adds a significant amount of weight to a causal interpretation of the findings, rather than just a correlational interpretation.

The findings have important implications for researchers, practitioners, and policy makers. For researchers, the results draw additional attention to the mismeasurement that occurs when firms use non-pecuniary OSS (and, more generally, non-pecuniary crowdsourced digital goods) as inputs into production. The results indicate that current studies underestimate the amount of IT at the firm. Future studies of productivity, especially the productivity of IT, should account for these non-pecuniary inputs, rather than misattributing them to firm intangible effects. This is especially important as information costs are increasingly approaching zero and the amount of non-pecuniary crowdsourced digital inputs firms use is likely to rise in the coming years. For practitioners, the results indicate that firms of all sizes may enhance their productivity by increasing the amount of OSS they employ in their production process, although larger firms may benefit more than medium sized firms due to economies of scale. Similarly, firms in the services sector may benefit more than those in the manufacturing sector. For policy makers, the results indicate that federal funding of OSS and other publicly available digital goods could enhance the productivity of firms. While other studies have shown that federal investments in such goods can have a high rate of return based on the value of the goods themselves (Greenstein and Nagle, 2014), the results of this study indicate that such goods can also boost the productivity of the firms that use them. However, as shown in the moderator and split sample results, not all firms benefit to the same degree.

Figures and Tables

Figure 1: **Examples of Software on the Free/Open Spectrum**

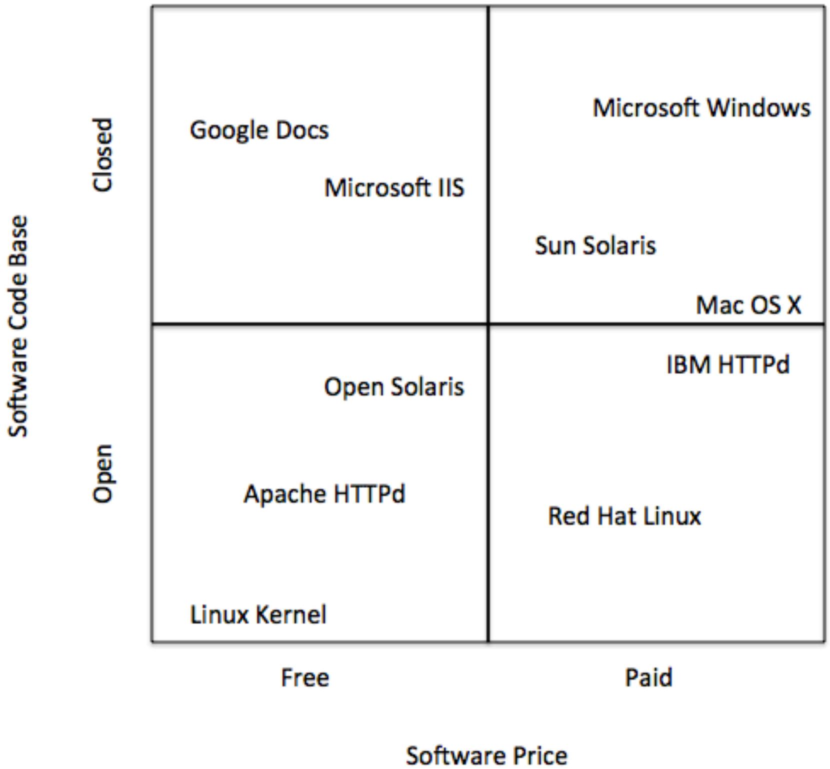

Table 1: **Open Source Operating Systems**

Pecuniary OSS Operating Systems	Non-Pecuniary OSS Operating Systems
Red Hat Linux	Berkeley Software Distribution (BSD)
SUSE Linux	Debian
SCO Linux	Conectiva
TurboLinux	Fedora
	FreeBSD
	Gentoo Linux
	Linux Kernel
	Mandrake Linux
	NetBSD
	OpenBSD
	Ubuntu

Table 2: **Descriptive Statistics**

Variable	Obs	Mean	Std. Dev.	Min	Max
$sales_{it}$	12244	5951.913	18793.42	2.694	425071
VA_{it}	12244	1845.747	5471.536	.006	154608
ITK_{it}	12244	8.279	48.687	.001	3165.154
$c\ K_{it}$	12244	4243.141	14840.1	.113	305797.1
L_{it}	12244	851.044	2838.818	.028	91149.09
$non\_pecuniary\_OSS_{it}$	12244	182.253	1264.606	0	65690
$pecuniary\_OSS_{it}$	12244	181.172	2983.13	0	207646
$closed_{it}$	12244	5026.755	18304.15	0	1176977

Values for monetary variables are in millions of deflated US dollars. Values for operating systems are in number of computers at the firm running operating systems in that category.

Table 3: **Correlation Matrix**

	VA_{it}	ITK_{it}	ITL_{it}	K_{it}	L_{it}	$nonpecuniary$ OSS_{it}	$pecuniary$ OSS_{it}	$closed_{it}$
VA_{it}	1.0000							
ITK_{it}	0.2989	1.0000						
ITL_{it}	0.4659	0.4444	1.0000					
K_{it}	0.7461	0.1910	0.3921	1.0000				
L_{it}	0.7846	0.1948	0.3561	0.4378	1.0000			
$nonpecuniary$ OSS_{it}	0.1846	0.0986	0.3264	0.1448	0.1541	1.0000		
$pecuniary$ OSS_{it}	0.1389	0.6848	0.2205	0.0594	0.0956	0.0384	1.0000	
$closed_{it}$	0.4089	0.9339	0.6007	0.3024	0.2758	0.1744	0.5472	1.0000

Table 4: **Industry Breakdown**

1-Digit SIC	Description	Frequency	Percent of all firms	Percent of firms using non-pecuniary OSS	Percent of firms using any OSS
0	Agriculture, Forestry, and Fishing	21	0.17	33.33	71.43
1	Mining and Construction	650	5.31	25.38	31.23
2-3	Manufacturing	5,879	48.02	31.25	43.35
4	Transportation, Communications, Electric, Gas, and Sanitary Services	927	7.57	27.83	37.32
5	Wholesale and Retail Trade	1,397	11.41	17.82	26.49
6	Finance, Insurance, and Real Estate	1,694	13.84	25.27	37.07
7-8	Services	1,676	13.69	34.78	47.20
9	Public Administration	0	0	0	0
All		12,244	100	28.81	40.04

Table 5: **Three-Factor Productivity Results**

DV: Value-Added (VA_{it})	1	2	3	4	5
Model	OLS	OLS	OLS	FE	RE
IT Capital (IT_{it})	.098*** (.008)	.066*** (.008)	.055*** (.008)	.030*** (.007)	.035*** (.006)
Non-IT Capital (K_{it})	.317*** (.012)	.314*** (.012)	.299*** (.012)	.082** (.034)	.270*** (.014)
Non-IT Labor (L_{it})	.631*** (.014)	.649*** (.014)	.671*** (.015)	.745*** (.035)	.699*** (.017)
Constant	.308*** (.040)	.234** (.045)	.298* (.163)	1.313*** (.169)	.379*** (.010)
Year fixed effect?	N	Y	Y	Y	Y
Industry fixed effect (SIC2)	N	N	Y	Y	Y
Number of firm/year observations	12244	12244	12244	12244	12244
Number of firms (groups)	1850	1850	1850	1850	1850
R^2 (between for panel)	0.898	0.913	0.917	0.903	0.930

***p<.01, **p<.05, *p<.1. All standard errors are clustered at the firm level. All variables are the natural log of the underlying variable.

Table 6: **Predicting Adoption of Non-Pecuniary OSS**

DV: Binary adoption of OSS	1
Model	Probit
IT Capital (ITK_{it})	-.426*** (.065)
IT Labor (ITL_{it})	.200*** (.020)
Non-IT Capital (K_{it})	.019 (.015)
Non-IT Labor (L_{it})	.029 (.021)
$pecuniary\_OSS_{it}$	.022** (.011)
$closed_{it}$	.431*** (.073)
Constant	-4.092*** (.073)
Number of firm/year observations	12244
Number of firms (groups)	1850
Pseudo - R^2	0.085
Wald chi^2	373.06

***p<.01, **p<.05, *p<.1. All standard errors are clustered at the firm level. All variables are the natural log of the underlying variable.

Table 7: **Covariate Balance**

	Panel A			Panel B		
	Unweighted Sample			Weighted Sample		
	Adopters	Non-Adopters	t-stat	Adopters	Non-Adopters	t-stat
IT Capital (ITK_{it})	10.567	7.354	3.31	3.935	10.191	4.27
IT Labor (ITL_{it})	37.044	12.810	27.08	12.523	17.086	1.35
Non-IT Capital (K_{it})	7231.559	3032.545	14.30	3129.700	3728.530	1.42
Non-IT Labor (L_{it})	1300.912	668.785	11.22	673.788	826.972	2.56
Number of firm/year observations	3,527	8,717		3,527	8,717	

Values reported are the means of the adopting or non-adopting firms. Panel A presents the unweighted OLS regression of the given variable on non-pecuniary OSS adoption. Panel B presents the weighted OLS regression of the given variable on non-pecuniary OSS adoption.

Table 8: **Baseline Regressions**

DV: Value-Added (VA_{it})	1	2	3	4	5	6
Model	Pooled OLS	Pooled OLS	Pooled OLS	Pooled OLS	Pooled OLS	Pooled OLS
Adoption Measure	Binary	Binary	Continuous	Continuous	Continuous	Continuous
IT Capital (ITK_{it})	0.017 (0.022)	-0.001 (0.015)	0.012 (0.022)	-0.005 (0.014)	0.003 (0.035)	-0.043 (0.024)
IT Labor (ITL_{it})	0.024 (0.018)	0.026** (0.013)	0.028 (0.018)	0.028** (0.013)	0.032 (0.025)	0.030 (0.019)
Non-IT Capital (K_{it})	0.303*** (0.023)	0.288*** (0.022)	0.302*** (0.023)	0.286*** (0.023)	0.297*** (0.036)	0.283*** (0.033)
Non-IT Labor (L_{it})	0.663*** (0.018)	0.694*** (0.019)	0.660*** (0.018)	0.695*** (0.019)	0.651*** (0.030)	0.708*** (0.031)
$non\_pecuniary\_OSS_{it}$	0.058* (0.031)	0.067** (0.034)	0.016*** (0.006)	0.016** (0.008)	0.021*** (0.008)	0.026** (0.011)
Constant	0.329** (0.147)	0.191 (0.124)	0.338** (0.152)	0.199 (0.127)	0.401 (0.241)	0.149 (0.225)
Year fixed effect?	Y	Y	Y	Y	Y	Y
Industry fixed effect (SIC2)?	N	Y	N	Y	N	Y
Number of firm/year observations	12244	12244	12244	12244	3530	3530
Number of firms (groups)	1850	1850	1850	1850	946	946
R^2	0.925	0.934	0.928	0.925	0.936	0.945

***p<.01, **p<.05, *p<.1. Standard errors are clustered at the firm level. All variables are the natural log of the underlying variable. All regressions are weighted with inverse-probability weightings based on the propensity of the firm to adopt non-pecuniary OSS. Columns 5 and 6 only use firms that have adopted non-pecuniary OSS as the sample.

Table 9: **IV Regressions**

DV: Value-Added (VA_{it})	1	2	3	4	5
Model	Pooled 2SLS	Pooled 2SLS	Pooled 2SLS	Pooled 2SLS	FE 2SLS
Adoption Measure	Binary	Binary	Continuous	Continuous	Continuous
IT Capital (ITK_{it})	-0.039 (0.030)	-0.010 (0.027)	-0.039 (0.032)	-0.020 (0.029)	-0.175** (0.078)
IT Labor (ITL_{it})	0.088*** (0.031)	0.055** (0.024)	0.080*** (0.029)	0.061** (0.025)	0.045*** (0.015)
Non-IT Capital (K_{it})	0.302*** (0.021)	0.302*** (0.022)	0.298*** (0.022)	0.299*** (0.022)	-0.128 (0.114)
Non-IT Labor (L_{it})	0.649*** (0.019)	0.656*** (0.018)	0.639*** (0.019)	0.647*** (0.018)	0.834*** (0.058)
$non\_pecuniary\_OSS_{it}$	0.813*** (0.276)	0.428** (0.173)	0.107*** (0.035)	0.073*** (0.025)	0.407** (0.200)
Constant	0.388*** (0.142)	0.434*** (0.144)	0.548*** (0.133)	0.529*** (0.137)	-
Year fixed effect?	Y	Y	Y	Y	Y
SIC2 Instrument (for $non\_pecuniary\_OSS_{it}$)	Y	Y	Y	Y	Y
County Instrument (for $non\_pecuniary\_OSS_{it}$)	N	Y	N	Y	N
Number of firm/year observations	12244	12244	12244	12244	12244
Number of firms (groups)	1850	1850	1850	1850	1850
First Stage F-test	26.74	22.73	28.64	19.15	9.80
R^2	0.898	0.918	0.906	0.913	0.478

***p<.01, **p<.05, *p<.1. Standard errors are clustered at the firm level for models 1-4 and are conventional GLS for model 5. All variables are the natural log of the underlying variable. All regressions are weighted with inverse-probability weightings based on the propensity of the firm to adopt non-pecuniary OSS.

Table 10: **Moderator and Split-Sample Regression Results**

DV: Value-Added (VA_{it})	1	2	3	4	5	6	7	8
Model	Pooled 2SLS	Pooled 2SLS	Pooled 2SLS	Pooled 2SLS	Pooled 2SLS	Pooled 2SLS	Pooled 2SLS	Pooled 2SLS
IT Capital (ITK_{it})	0.003 (0.034)	0.042 (0.053)	-0.173*** (0.054)	-0.039 (0.032)	-0.028 (0.029)	-0.027 (0.050)	-0.126 (0.077)	0.055 (0.040)
IT Labor (ITL_{it})	0.011 (0.027)	-0.008 (0.035)	0.137*** (0.046)	0.053** (0.027)	0.070*** (0.025)	0.056 (0.038)	0.068* (0.036)	-0.002 (0.018)
Non-IT Capital (K_{it})	0.287*** (0.030)	0.187*** (0.058)	0.179*** (0.049)	0.295*** (0.027)	0.269*** (0.022)	0.241*** (0.053)	0.292*** (0.025)	0.335*** (0.021)
Non-IT Labor (L_{it})	0.101 (0.168)	0.794*** (0.040)	0.798*** (0.043)	0.667*** (0.021)	0.676*** (0.017)	0.717*** (0.065)	0.674*** (0.037)	0.616*** (0.024)
$non\_pecuniary\_OSS_{it}$	-0.251* (0.141)	-0.009 (0.046)	0.194** (0.086)	0.091*** (0.028)	0.083*** (0.026)	0.107*** (0.042)	0.177** (0.071)	-0.004 (0.073)
ln (emp)	0.501*** (0.193)							
$non\_pecuniary\_OSS_{it}$ * ln (emp)	0.035** (0.015)							
Constant	-0.663 (0.913)	0.465** (0.211)	0.450*** (0.156)	0.453*** (0.151)	0.550*** (0.136)	0.448*** (0.143)	0.631*** (0.177)	0.453*** (0.086)
Year fixed effect?	Y	Y	Y	Y	Y	Y	Y	Y
Sample Restriction	-	Manuf.	Services	Excluding finance industries	Excluding agriculture and mining	IT-Producing Industries	IT-Using Industries	Non-IT Using or Producing Industries
SIC2 Instrument (for $non\_pecuniary\_OSS_{it}$)	Y	Y	Y	Y	Y	Y	Y	Y
County Instrument (for $non\_pecuniary\_OSS_{it}$)	Y	Y	Y	Y	Y	Y	Y	Y
Number of firm/year observations	12244	5880	1677	10555	11574	1168	4515	4714
Number of firms (groups)	1850	863	316	1644	1764	238	798	832
First Stage F-test	-	17.05	7.77	19.13	18.44	19.141	12.011	5.705
R^2	0.929	0.926	0.928	0.928	0.924	0.954	0.875	0.921

***p<.01, **p<.05, *p<.1. Standard errors are clustered at the firm level. All variables are the natural log of the underlying variable. All regressions are weighted with inverse-probability weightings based on the propensity of the firm to adopt non-pecuniary OSS. Columns 6, 7, and 8 use a one-year lag of OSS usage and instruments.

Table 11: **Robustness Checks**

DV: Value-Added (VA_{it})	1	2	3	4	5	6	7	8
Model	Pooled 2SLS	Pooled 2SLS	Pooled 2SLS	Pooled 2SLS	Pooled 2SLS	Pooled 2SLS	Pooled 2SLS	Pooled 2SLS
IT Capital (ITK_{it})			-0.016 (0.028)	-0.029 (0.020)	0.011 (0.030)	-0.036* (0.021)	-0.057* (0.034)	-0.025 (0.015)
IT Labor (ITL_{it})			0.059** (0.025)	0.058*** (0.019)	0.045* (0.027)	0.031** (0.013)	0.050*** (0.016)	0.052** (0.022)
Non-IT Capital (K_{it})	0.298*** (0.021)	0.297*** (0.022)	0.301*** (0.022)	0.316*** (0.014)	0.297*** (0.024)	0.313*** (0.012)	0.306*** (0.015)	0.301*** (0.022)
Non-IT Labor (L_{it})	0.643*** (0.018)		0.651*** (0.018)	0.607*** (0.017)	0.639*** (0.019)	0.653*** (0.015)	0.661*** (0.017)	0.653*** (0.018)
$non\_pecuniary\_OSS_{it}$	0.073*** (0.024)	0.068*** (0.025)		0.091*** (0.008)	0.061*** (0.017)	0.110*** (0.036)	0.093** (0.048)	0.062*** (0.015)
IT Capital and Labor combined	0.051*** (0.013)							
# of PCs and Servers		-0.013 (0.029)						
# of IT employees		0.040 (0.025)						
# of non-IT employees		0.668*** (0.019)						
Imputed cost for non-pecuniary OSS			0.040*** (0.015)					
Constant	0.485*** (0.122)	-1.666*** (0.136)	0.484*** (0.139)	0.431*** (0.070)	0.578*** (0.153)	0.515*** (0.051)	1.336*** (0.104)	0.458*** (0.146)
Year fixed effect?	Y	Y	Y	Y	Y	Y	Y	Y
County fixed effect?	N	N	N	N	N	N	Y	N
Robustness Check	Combined IT capital and 3x labor	Raw # for ITL, non-ITL, and ITK	Imputed price for OSS	6-year lag of OSS use	1-year lag of instruments	No IPW	County fixed-effect	All OSS variable
SIC2 Instrument (for $non\_pecuniary\_OSS_{it}$)	Y	Y	Y	Y	Y	Y	Y	Y
COUNTY Instrument (for $non\_pecuniary\_OSS_{it}$)	Y	Y	Y	Y	Y	Y	N	Y
Number of firm/year observations	12244	12244	12244	3670	10397	12244	12244	12244
Number of firms (groups)	1850	1850	1850	1182	1718	1850	1850	1850
First Stage F-test	22.38	19.52	20.72	165.68	25.08	36.74	40.88	19.27
R^2 (between)	0.920	0.922	0.920	0.934	0.928	0.900	0.931	0.920

***p<.01, **p<.05, *p<.1. Standard errors are clustered at the firm level. All variables are the natural log of the underlying variable. Regressions in columns 1-5 and 7-8 are weighted with inverse-probability weightings based on the propensity of the firm to adopt non-pecuniary OSS.

References

Abadie, A., & Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, *74*(1), 235-267.

Afuah, A. (2000). Do your co-opetitors' capabilities matter in the face of technological change? *Strategic Management Journal*, 21, pp. 387-404.

Altman, E., Nagle, F., & Tushman, M. (2014). Innovating without Information Constraints: Organizations, Communities, and Innovation When Information Costs Approach Zero. In Oxford Handbook of Creativity, Innovation, and Entrepreneurship, edited by Michael A. Hitt, Christina Shalley, and Jing Zhou. Oxford University Press.

Angrist, J.D., & Pischke, J.S. (2009). Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press, Princeton, NJ.

Aral, S., Brynjolfsson, E., & Wu, L. (2012). Three-way complementarities: Performance pay, human resource analytics, and information technology. *Management Science*, *58*(5), 913-931.

Aral, S., & Weill, P. (2007). IT assets, organizational capabilities, and firm performance: How resource allocations and organizational differences explain performance variation. *Organization Science*, *18*(5), 763-780.

Asay, M. (2013). Is Facebook The World's Largest Open Source Company? *ReadWrite.* Retrieved from http://readwrite.com/2013/10/17/is-facebook-the-worlds-largest-open-source-company on October 31, 2014.

Athey, S., & Ellison, G. (2014). Dynamics of Open Source Movements. *Journal of Economics & Management Strategy*, *23*(2), 294-316.

Athey, S., & Stern, S. (1998). *An empirical framework for testing theories about complementarity in organizational design*. National Bureau of Economic Research (NBER). Working Paper 6600.

Athey, S. & Stern, S. (2002) The impact of information technology on emergency health care outcomes. *RAND Journal of Economics*, 33(3), 399-432.

Baldwin, C. Y., & Clark, K. B. (2006). The architecture of participation: Does code architecture mitigate free riding in the open source development model?. *Management Science*, *52*(7), 1116-1127.

Baldwin, C., & Von Hippel, E. (2011). Modeling a Paradigm Shift: From Producer Innovation to User and Open Collaborative Innovation. *Organization Science*, *22*(6), 1399–1417.

Benkler, Y. (2002). Coase's Penguin, or, Linux and" The Nature of the Firm". *Yale Law Journal*, 369-446.

Black Duck Software. (2014). The Eighth Annual Future of Open Source Survey. Retrieved from https://www.blackducksoftware.com/future-of-open-source on Oct. 31, 2014.

Bloom, N. & J. Van Reenen. (2007). Measuring and explaining management practices across firms and countries. *Quarterly Journal of Economics*. **122**(4) 1351-1408.

Bloom, N., Sadun, R., & Van Reenen, J. (2012). Americans Do IT Better: US Multinationals and the Productivity Miracle. *American Economic Review*, 102(1), 167-201.

Brandenburger, A. M., & Nalebuff, B. J. (2011). *Co-opetition*. Random House LLC.

Bresnahan, T.F., E. Brynjolfsson, & L.M. Hitt. (2002). Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *Quarterly Journal of Economics*. **117**(1) 339-376.

Bridgman, B. (2013). *Home Productivity*. Bureau of Economic Analysis Working Paper 2013-03.

Brynjolfsson, E., & Hitt, L. (1996). Paradox lost? Firm-level evidence on the returns to information systems spending. *Management Science*, *42*(4), 541-558.

Brynjolfsson, E., & Hitt, L. M. (2003). Computing productivity: Firm-level evidence. *Review of economics and statistics*, *85*(4), 793-808.

Brynjolfsson, E., Hitt, L. M., & Yang, S. (2002). Intangible assets: Computers and organizational capital. *Brookings papers on economic activity*, *2002*(1), 137-198.

Brynjolfsson, E., & Milgrom, P. (2012). Complementarity in organizations. *The Handbook of Organizational Economics*, *11*.

Brynjolfsson, E. & Saunders, A. (2009). What the GDP Gets Wrong (Why Managers Should Care). *Sloan Management Review*, 51(1), 95-96.
Byrne, D., Oliner, S., & Sichel, D. (2013). Is the information technology revolution over? *Available at SSRN 2240961*.
Casadesus-Masanell, R., & Llanes, G. (2011). Mixed Source. *Management Science*, *57*(7), 1212–1230.
Ceccagnoli, M., Forman, C., Huang, P., & Wu, D. J. (2014). Digital platforms: when is participation valuable?. *Communications of the ACM*, *57*(2), 38-39.
Chatterji, A. K., & Fabrizio, K. R. (2013). Using users: When does external knowledge enhance corporate product innovation?. *Strategic Management Journal*.
Cole, S.R., & Hernan, M.A. (2008). Constructing Inverse Probability Weights for Marginal Structural Models. *American Journal of Epidemiology*, 168 (6), pp. 656-664.
Corbet, J., Kroah-Hartman, G. & McPherson, A. (2013). Linux Kernel Development: How Fast it is Going, Who is Doing It, What They are Doing, and Who is Sponsoring It (2013 Edition). Linux Foundation Whitepaper.
Corrado, C., & Hulten, R. (2013). Innovation Accounting. In Measuring Economic Sustainability and Progress, edited by Dale W. Jorgenson, J. Steven Landefeld, and Paul Schreyer. University of Chicago Press.
Corrado, C., Hulten, C., & Sichel, D. (2009). Intangible capital and US economic growth. *Review of Income and Wealth*, *55*(3), 661-685.
Dewan, S., & Min, C. K. (1997). The substitution of information technology for other factors of production: A firm level analysis. *Management Science*, *43*(12), 1660-1675.
Finley, K. (2013). Apple's Operating System Guru Goes Back to His Roots. *Wired*. Retrieved from www.wired.com/2013/08/jordan-hubbard/ on October 31, 2014.
Fitzgerald, B. (2006). The transformation of open source software. *MIS Quarterly*, 587-598.
Forman, C. (2005). The corporate digital divide: Determinants of Internet adoption. *Management Science*, *51*(4), 641-654.
Forman, C., Goldfarb, A., & Greenstein, S. (2005). How did location affect adoption of the commercial Internet? Global village vs. urban leadership. *Journal of Urban Economics*, *58*(3), 389-420.
Forman, C., Goldfarb, A., & Greenstein, S. (2008). Understanding the inputs into innovation: Do cities substitute for internal firm resources?. *Journal of Economics & Management Strategy*, *17*(2), 295-316.
Forrest, C. (2014). Salil Deshpande: Software Engineer, Venture Capitalist, Open Source Investor. *TechRepulic*. Retrieved from http://www.techrepublic.com/article/salil-deshpande-software-engineer-venture-capitalist-open-source-investor/ on October 31, 2014.
Fosfuri, A., Giarratana, M. S., & Luzzi, A. (2008). The penguin has entered the building: The commercialization of open source software products. *Organization Science*, *19*(2), 292-305.
FreeBSD Contributor List. https://www.freebsd.org/doc/en/articles/contributors/article.html, retrieved on October 27, 2014.
Furman, J. L., Porter, M. E., & Stern, S. (2002). The determinants of national innovative capacity. *Research Policy*, *31*(6), 899-933.
Gilder, G. (1995). The Coming Software Shift. *Forbes*, August 28, 1995.
Giera, J., & Brown, A. (2004). The Costs and Risks of Open Source – Debunking the Myths. Forrester Research Whitepaper.
Graham, R. (2014). 300k vulnerable to Heartbleed two months later. *Errata Security*. Retrieved from http://blog.erratasec.com/2014/06/300k-vulnerable-to-heartbleed-two.html on October 1, 2014.
Greenstein, S., & Nagle, F. (2014). Digital Dark Matter and the Economic Contribution of Apache. *Research Policy* 43, pp.623-631.
Hamilton, D. (2014). Mirantis Gains $100M in the Largest Series-B Investment Round in Open-Source Software History. Retrieved from http://www.thewhir.com/web-hosting-news/mirantis-gains-100m-largest-series-b-investment-round-open-source-software-history on October 31, 2014.
Han, K., Kauffman, R. J., & Nault, B. R. (2011). Returns to Information Technology Outsourcing. *Information Systems Research*, *22*(4), 824-840.

Han, K., & Mithas, S. (2013). Information technology outsourcing and non-IT operating costs: An empirical investigation. *MIS Quarterly*, *37*(1), 315-331.

Hann, I., Roberts, J., and Slaughter, S. (2013). All Are Not Equal: An Examination of the Economic Returns to Different Forms of Participation in Open Source Software Communities. *Information Systems Research* 24(3), pp. 520-538.

Hann, I., Roberts, J., Slaughter, S. and Fielding, R. (2002). Economic Incentives for Open Source Projects: Can Participation be Explained by Career Concerns? *Proceedings of the 22nd International Conference on Information Systems (ICIS)*, Barcelona, Spain, December 2002.

Harhoff, D., Henkel, J., & Von Hippel, E. (2003). Profiting from voluntary information spillovers: how users benefit by freely revealing their innovations. *Research Policy*, *32*(10), 1753-1769.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of Economic and Social Measurement, Volume 5, number 4* (pp. 475-492). NBER.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, 153-161.

Henkel, J. (2006). Selective revealing in open innovation processes: The case of embedded Linux. *Research Policy*, *35*(7), 953-969.

Hirano, K., Imbens, G. W., & Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, *71*(4), 1161-1189.

Hitt, L. M., & Brynjolfsson, E. (1996). Productivity, business profitability, and consumer surplus: three different measures of information technology value. *MIS Quarterly*, 121-142.

Hogan, J.W., & Lancaster, T. (2004). Instrumental variables and inverse probability weighting for causal inference from longitudinal observational studies. *Statistical Methods in Medical Research*, 13, pp 17-48.

Horvitz, D.G., & Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47, pp. 663-685.

Horovitz, B. (2013). Crowdsourcing rules for Super Bowl ads. *USA Today*. Retrieved from http://www.usatoday.com/story/money/business/2013/01/19/crowdsourcing-super-bowl-commercials-doritos-lincoln-pepsi/1842937/ on October 31, 2014.

Howe, J. 2008. *Crowdsourcing: Why the Power of the Crowd is Driving the Future of Business.* Crown Business, New York.

Huang, P., Ceccagnoli, M., Forman, C., & Wu, D. J. (2013). IT Knowledge Spillovers and Productivity: Evidence from Enterprise Software. *Available at SSRN 2243886*.

Huber, M. (2013). Identifying causal mechanisms (primarily) based on inverse probability weighting. *Journal of Applied Econometrics*, 2013.

Hulten, C. (2010). Decoding Microsoft: Intangible Capital as a Source of Company Growth. National Bureau of Economic Research (NBER) Working Paper 15799.

Imbens, G. W., & Kolesar, M. (2012). Robust Standard Errors in Small Samples: Some Practical Advice. *NBER Working Paper w18478*.

Jaffe, A. B., & Trajtenberg, M. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics*, *108*(3).

Jorgenson, D. W. (2001). Information technology and the US economy. *The American Economic Review*, *91*(1), 1-32.

Jorgenson, D. W., Ho, M. S., & Stiroh, K. J. (2005). Productivity, Volume 3: Information Technology and the American Growth Resurgence. *MIT Press Books*, *3*.

Kogut, B., & Metiu, A. (2001). Open-source software development and distributed innovation. *Oxford Review of Economic Policy*, *17*(2), 248-264.

Krishnamurthy, S. (2005). "An Analysis of Open Source Business Models," in Perspectives on Free and Open Source Software, J. Feller, B. Fitzgerald, S. Hissam, and K. Lakhani (eds.), MIT Press, Cambridge, MA, 2005, pp. 279-296.

Lakhani, K., & Von Hippel, E. (2003). How open source software works: “free” user-to-user assistance. *Research Policy*, *32*(6), 923–943.

Lakhani, K., Lifshitz-Assaf, H., & Tushman, M. (2012). Open innovation and organizational boundaries: the impact of task decomposition and knowledge distribution on the locus of innovation in *Handbook of Economic Organization: Integrating Economic and Organization Theory*, A. Grandori (ed.), Edward Elgar Publishing, Northampton, MA, pp. 355-382.

Lerner, J., Pathak, P. A., & Tirole, J. (2006). The dynamics of open-source contributors. *The American Economic Review*, 114-118.

Lerner, J., & Schaknerman, M. (2010). The comingled code: Open source and economic development. *MIT Press Books*.

Lerner, J., & Tirole, J. (2002). Some Simple Economics of Open Source. *The Journal of Industrial Economics*, *50*(2), 197–234.

Lunden, I. (2014). Alfresco Raises A Fresh $45M to Fuel Open-Source Enterprise Content Management. *TechCrunch.* Retrieved from http://techcrunch.com/2014/08/21/alfresco-raises-a-fresh-45m-to-fuel-open-source-enterprise-content-management/ on October 31, 2014.

MacCormack, A. (2003). Evaluating Total Cost of Ownership for Software Platforms: Comparing Apples, Oranges, and Cucumbers. AEI-Brookings Joint Center for Regulatory Studies Related Publication, April 2003.

MacCormack, A., Rusnak, J., & Baldwin, C. Y. (2006). Exploring the structure of complex software designs: An empirical study of open source and proprietary code. *Management Science*, *52*(7), 1015-1030.

McCue, T.J. (2013). For Motor Company Sees Open Source. *Forbes.* Retrieved from http://www.forbes.com/sites/tjmccue/2013/01/10/ford-motor-company-sees-open-source/ on October 31, 2014.

McElheran, K. S. (2014). Delegation in Multi-Establishment Firms: Adaptation vs. Coordination in I.T. Purchasing Authority. *Journal of Economics & Management Strategy*, 23 (2), 225-258.

O’Mahony, S. (2003). Guarding the commons: how community managed software projects protect their work. *Research Policy*, *32*(7), 1179–1198.

O'Mahony, S., & Ferraro, F. (2007). The emergence of governance in an open source community. *Academy of Management Journal*, *50*(5), 1079-1106.

Ostrom, E. (1990). Governing the commons: The evolution of institutions for collective action. Cambridge university press.

Phipps, S. (2014). Walmart’s investment in open source isn’t cheap. *InfoWorld*. Retrieved from http://www.infoworld.com/article/2608897/open-source-software/walmart-s-investment-in-open-source-isn-t-cheap.html on October 31, 2014.

Raymond, Eric. (1998). Goodbye, “free software”; hello, “open source”. Retrieved from http://www.catb.org/~esr/open-source.html on February 23, 2014.

Raymond, E. (1999). The cathedral and the bazaar. *Knowledge, Technology & Policy*, *12*(3), 23-49.

Russo, B., Braghin, B., Gasperi, P., Sillitti, A., and Succi, G. (2005). Defining TCO for the Transition to Open Source Systems. *Proceedings of the First International Conference on Open Source (OSS2005)*, pp. 108-112.

Saunders, A., & Brynjolfsson, E. (2013). Valuing IT-Related Intangible Assets. *Available at SSRN 2344949*.

Schofield, J. (2008). How many people make Windows 7? *The Guardian*. Retrieved from http://www.theguardian.com/technology/blog/2008/aug/19/howmanypeoplemakewindows7 on October 29, 2014.

Schumpeter, J. 1942. The Process of Creative Destruction. Chapter VII, pp. 81-86 in *Capitalism, Socialism, and Democracy.* Harper & Row, New York, NY.

Schwarz, M., & Takhteyev, Y. (2011). Half a Century of Public Software Institutions". *Journal of Public Economic Theory*, *12*(4), 609-639.

Shirky, C. (2008). Here Comes Everybody: The Power of Organizing Without Organizations. Penguin Press, New York.

Sinofsky, S. (2011). Introducing the team. *Microsoft Developer Network Blog*. Retrieved from http://blogs.msdn.com/b/b8/archive/2011/08/17/introducing-the-team.aspx on October 29, 2014.

Sorkin, A. & Peters, J. (2006). Google to Acquire YouTube for $1.65 Billion. *The New York Times.* Retrieved from http://www.nytimes.com/2006/10/09/business/09cnd-deal.html on October 31, 2014.

Syverson, C. (2011). What Determines Productivity? *Journal of Economic Literature*, 49(2), pp. 326-365.

Tambe, P., & Hitt, L. M. (2012). The Productivity of Information Technology Investments : New Evidence from IT Labor Data. *Information Systems Research*, *23*(3), 599–617.

Tambe, P., Hitt, L., & Brynjolfsson, E. (2011). The Price and Quantity of IT-Related Intangible Capital. Working paper.

Tambe, P., Hitt, L. M., & Brynjolfsson, E. (2012). The Extroverted Firm: How External Information Practices Affect Innovation and Productivity. *Management Science*, *58*(5), 843–859.

Varian, H. R., & Shapiro, C. (2003). Linux adoption in the public sector: An economic analysis. *Manuscript. University of California, Berkeley*.

Von Hayek, F. A. (1945). The use of knowledge in society. *The American Economic Review*, 519-530.

Von Hippel, E. (1986). Lead Users: A Source of Novel Product Concepts. *Management Science*, *32*(7), 791–805.

Von Hippel, E., & Von Krogh, G. (2003). Open source software and the “private-collective” innovation model: Issues for organization science. *Organization Science*, *14*(2), 209-223.

Von Krogh, G., Spaeth, S., & Lakhani, K. R. (2003). Community, joining, and specialization in open source software innovation: a case study. *Research Policy*, *32*(7), 1217–1241.

West, J., & Lakhani, K. R. (2008). Getting clear about communities in open innovation. *Industry and Innovation*, *15*(2), 223-231.

Wheeler, D. (2005). Why Open Source Software/Free Software (OSS/FS, FLOSS, or FOSS)? Look at the Numbers! available online at http://www.dwheeler.com/oss_fs_why.html.

Woods, D., & Guliani, G. (2005). Open Source for the Enterprise: Managing Risks, Reaping Rewards. O’Reilly Media.

Wooldridge, J. (2002). Inverse probability weighted M-estimators for sample selection, attrition, and stratification. *Portuguese Economic Journal*, 1, pp. 117-139.

Wooldridge, J. (2007). Inverse probability weighted estimation for general missing data problems. *Journal of Econometrics*, 141(2), 1281-1301.

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, *330*(6004), 686-688.

Woolley, A. W., & Fuchs, E. (2011). PERSPECTIVE-Collective Intelligence in the Organization of Science. *Organization Science*, *22*(5), 1359-1367.

Yarow, J. (2013). Microsoft’s Biggest Problem In One Chart. *Business Insider*. Retrieved from http://www.businessinsider.com/microsofts-biggest-problem-in-one-chart-2013-9 on October 31, 2014.

Yi, S. K. M., Steyvers, M., Lee, M. D., & Dry, M. J. (2012). The wisdom of the crowd in combinatorial problems. *Cognitive Science*, *36*(3), 452-470.

Young, R. & Johnson, D. R., (2009). A Comparison of Four Methods for Handling Missing Secondary Respondent Data. *Paper presented at the American Sociological Association Annual Meeting.*