Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Open Source Software and Firm Productivity

Frank Nagle

informs
http://pubsonline.informs.org/journal/mnsc/

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–25
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Open Source Software and Firm Productivity

Frank Nagle[a]

[a] Marshall School of Business, University of Southern California, Los Angeles, California 90089
Contact: naglef@marshall.usc.edu, http://orcid.org/0000-0002-3761-540X (FN)

Received: June 10, 2015
Revised: September 12, 2016; May 29, 2017; August 18, 2017; September 26, 2017
Accepted: October 3, 2017
Published Online in Articles in Advance: May 4, 2018

https://doi.org/10.1287/mnsc.2017.2977

Abstract. As open source software (OSS) is increasingly used as a key input by firms, understanding its impact on productivity becomes critical. This study measures the firm-level productivity impact of nonpecuniary (free) OSS and finds a positive and significant value-added return for firms that have an ecosystem of complementary capabilities. There is no such impact for firms without this ecosystem of complements. Dynamic panel analysis, instrumental variables, and a variety of robustness checks are used to address measurement error concerns and to add support for a more causal interpretation of the results. For firms with an ecosystem of complements, a 1% increase in the use of nonpecuniary OSS leads to an increase in value-added productivity of between 0.002% and 0.008%. This effect is smaller for larger firms, and the results indicate that prior research underestimates the amount of IT firms use.

History: Accepted by Chris Forman, information systems.
Supplemental Material: The online appendix is available at https://doi.org/10.1287/mnsc.2017.2977.

Keywords: productivity of IT • user innovation • open source software • economics of IS • technology strategy • IT policy and management

1. Introduction

In 2011, Marc Andreessen, the well-known venture capitalist, famously said, "Software is eating the world" (Andreessen 2011). Four years later, in 2015, the lesser-known venture capitalist John Vrionis responded, "It's actually open source software that's eating the world" (Vrionis 2015). Although this trend toward open source software (OSS) has been gaining more notoriety recently, OSS has long played an important role within firms and their IT infrastructure. This phenomenon has been widely discussed in the popular press as technology giants such as Apple, Google, and Facebook increase their reliance on OSS to grow their innovative and productive efforts (Sorkin and Peters 2006, Asay 2013, Finley 2013). Even Microsoft, whose former chief executive officer Steve Ballmer famously called one of the leading OSS projects, Linux, "a cancer" (Newbart 2001), has now fully embraced OSS and Linux, and Microsoft is now being called "an open-source company" (Vaughan-Nichols 2016). However, it is not only technology-focused companies that are relying on OSS: Ford, Walmart, and a host of other well-known non-IT brands use OSS to help drive innovation and production (McCue 2013, Phipps 2014). Additionally, these same technologies are allowing small start-ups to have a large impact, even when they are capital constrained, because of the nonpecuniary, or free, nature of these critical inputs. For example, WhatsApp, which was later acquired by Facebook for $19 billion, relied heavily on OSS from the earliest days of its inception (WhatsApp 2016). Furthermore, recent studies have shown that more than 50% of firms now use or contribute to OSS (Black Duck Software 2014), and billions of venture capital dollars are pouring into the OSS ecosystem (Black Duck Software 2014, Hamilton 2014). In addition, because of 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 OSS as an input into production, measuring the value it helps create can be difficult. In a classic Schumpeterian creative destruction process (Schumpeter 1942), OSS destroys old business models while creating new opportunities for growth. However, as with many digital goods, the value created by OSS 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 OSS, these measurement challenges have prevented researchers from analyzing how its impact varies across different firms and industries. Furthermore, it has been suggested that integrating OSS into the firm's production process can be more costly than comparable pecuniary inputs (Giera and Brown 2004), and consequently, its 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 nonpecuniary open source software 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. Research has shown that the increased use of unpriced goods of both a digital (Brynjolfsson and Saunders 2009, Greenstein and Nagle 2014) and nondigital (Bridgman 2013) nature may be an important factor in understanding recent trends in gross domestic product (GDP). Nonpecuniary digital goods, like OSS, 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 using OSS and other crowdsourced digital goods as key inputs into their productive and innovative processes. While some leading firms, such as Google and Facebook, have embraced OSS, others have shied away from relying on such inputs because of concerns about reliability, sharing with competitors, and the costs of restructuring business models to add externally developed digital goods directly into the innovation and production processes. More broadly, understanding the productivity implications of OSS 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.

This paper seeks to add insights to two important bodies of literature: the user innovation literature and the productivity returns to IT literature. The user innovation literature (e.g., von Hippel 1986)—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 et al. 2016) focuses almost exclusively on IT investments of a pecuniary nature, completely missing investments in nonpecuniary 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 nonpecuniary IT. Understanding the impact of such goods on firm productivity not only helps to contribute to the broad literature on the determinants of productivity[1] but also shows that user innovation is no longer a rare phenomenon and is becoming a key input into firm productivity. Additionally, the paper offers insights for practitioners that can be utilized to increase the profitability of the firm's operations and gain a competitive advantage by using open source goods as inputs. Finally, for policy makers, the results encourage policies that incentivize production of OSS and other public digital goods as a method for increasing firm and national productivity.

2. Free and Open Source Software as an Input into Production

As early as the 1980s, innovation and 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, often in the realm of OSS. However, most of the academic work on OSS has been focused on exploring supply-side mechanisms—why users contribute to OSS (Benkler 2002, Lerner and Tirole 2002, West and Lakhani 2008, Athey and Ellison 2014), how users join OSS projects (von Krogh et al. 2003), how users help each other contribute to OSS (Lakhani and von Hippel 2003), and how 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 et al. 2003, von Hippel and von Krogh 2003, Henkel 2006, Lerner et al. 2006, Fosfuri et al. 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[2]—who uses it, why they use it, and whether there are productivity benefits to using it remain unanswered questions. This is despite the fact that OSS—and, more broadly, nonpecuniary, 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 et al. 2012, Altman et al. 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), and productivity boosts from investments in IT are frequently mistaken for intangible firm-specific benefits (Brynjolfsson et al. 2002, Syverson 2011, Tambe et al. 2011, Saunders and Brynjolfsson 2016). Studies have also shown that IT-producing and using industries contributed a disproportionately large amount to the economic growth experienced in the United States, particularly from

1995 to 2004 (Jorgenson 2001, Stiroh 2002, Jorgenson et al. 2005). In addition to spending on IT capital, spending on IT labor has also been found to boost firm productivity (Tambe and Hitt 2012). Furthermore, participation in networks of practice adds IT-related knowledge spillovers that increase productivity (Huang et al. 2016). Relatedly, investments in IT outsourcing have been shown to have a positive impact on productivity (Han et al. 2011, Han and Mithas 2013), and investments in IT can impact public-sector value creation (Pang et al. 2014). 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 et al. 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,[3] it is not properly factored into such calculations. At a macro level, 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 United States alone (Greenstein and Nagle 2014), and the inclusion of intangibles[4] and nonpecuniary production have been shown to significantly alter GDP calculations (Corrado et al. 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 nonpecuniary OSS on firm productivity? After establishing a baseline answer to this question, the paper further considers the firm-level differences in extracting productive value from OSS, with a particular focus on the ecosystem of complements and capabilities necessary for productive use of OSS.

3. Theory Development

Because nonpecuniary OSS has no price, it cannot be measured by traditional means, and its impact on productivity has therefore gone unstudied in prior work. Importantly, compared with closed source and pecuniary open source software, using free and open source software can be risky, but it can also provide a number of additional benefits, so the net impact on productivity is unintuitive. This section discusses these risks and benefits to illuminate the empirical puzzle surrounding the productivity impact of using nonpecuniary OSS. It then discusses various characteristics of the firm that may moderate the main effect. In aggregate, it is not clear ex ante how the use of nonpecuniary OSS will impact firm productivity (if at all), motivating the empirical study that follows.

3.1. Risks and Costs of Using Nonpecuniary OSS

Compared with pecuniary OSS and closed source software, nonpecuniary OSS can be a risky and costly investment.[5] This section discusses the largest of these risks, including the fact that implementing 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. In aggregate, all of these risks and costs may lead to nonpecuniary OSS having a negative impact on productivity.

The allure of "free" software can be great for any capital-constrained firm considering implementing new software. 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 with 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 nonpecuniary OSS may actually be misleading and may lead to long-term costs that are 5%–20% higher than those of proprietary closed source software (Giera and Brown 2004). Such unseen costs are likely to increase the firm's overall costs, which would lead to a negative impact on productivity.

In addition to the direct monetary costs of supporting it, nonpecuniary OSS is often seen as riskier than pecuniary software for a number of reasons. To begin with, because a collective of users, rather than a central producer, creates nonpecuniary 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 users do not have a service agreement with any vendor (Woods and Guliani 2005). Relatedly, although larger OSS foundations, such as 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 on 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[6] 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.[7] Furthermore, vulnerabilities in open source software are often at a greater risk of exploitation (Ransbotham 2010). Perhaps the most concerning risk of all is the lack of a contractual relationship between a firm using nonpecuniary 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 nonpecuniary OSS, which means the use of such software is riskier than pecuniary OSS or closed source software where such agreements exist.

The lack of development and support, security concerns, and lack of contractual relationships can lead to users of nonpecuniary OSS spending unnecessary (and unproductive) time troubleshooting problems that arise with their software, rather than spending time contributing to activities that directly enhance the productivity of the firm. The view of nonpecuniary OSS as a risky and costly decision led to the commonly used phrase, "No one ever got fired for buying Microsoft."[8] 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. In aggregate, the various risks and costs laid out above could have a negative impact on the productivity of the firm.

3.2. Benefits of Using Nonpecuniary OSS

Despite all of the risks and costs discussed above, nonpecuniary OSS can also provide a number of benefits to firms willing and able to take on these risks and costs. These benefits include reduced up-front costs, collective intelligence of the crowd, and greater flexibility to alter and enhance the code base, all of which can lead to higher productivity.

The most salient benefit of using nonpecuniary OSS is the free nature of the software. Although, as discussed above, the actual cost of software is minimal compared with the costs of implementing it, the fact remains that firms using nonpecuniary 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 nonpecuniary 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 nonpecuniary 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 nonpecuniary 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.[9] This is the modern-day interpretation of earlier arguments by 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). Nonpecuniary OSS projects address this problem by allowing anyone to contribute to the development of the underlying code base. Therefore, the use of OSS allows a firm to harness the labor efforts of a wide collective of individuals. 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 and Fuchs 2011), including software development (von Hippel and von Krogh 2003). Furthermore, 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 nonpecuniary OSS has the added benefit of allowing firms to avoid holdup 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 holdup by the developer. However, if a firm relies on nonpecuniary OSS and needs a specific function, it can contribute to the code itself (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). Furthermore, the open nature of the software leads to a more modular architecture, which has been shown to allow for better integration (MacCormack et al. 2006) and can lead to higher productivity.

3.3. IT Capabilities and the Ecosystem of Complements

Since the productivity impact of nonpecuniary OSS is likely to be influenced by both the costs/risks and benefits discussed above, it is highly likely that some types of firms will benefit more from its use than others. In particular, those firms that already have an ecosystem of complements in place may benefit more from the use of nonpecuniary OSS. More specifically, nonpecuniary OSS (compared with both pecuniary OSS and closed software) requires a complement of technical capabilities for implementing and using such software. Although these capabilities are not necessarily of a higher quality than at other firms,[10] they are an important part of the ecosystem of complements that allows for the productive use of nonpecuniary OSS. Therefore, for a firm to successfully implement nonpecuniary OSS, it must already have a high degree of technical capabilities. These specialized capabilities and the complementary ecosystem surrounding nonpecuniary OSS are likely necessary for its use to have a positive productivity impact. These capabilities are likely to reside in firms that are either intensive IT users or are in an IT-producing industry (defined in Section 6.3). Such firms are likely to have a greater ability to implement nonpecuniary OSS and will be more apt to have the benefits outweigh the costs.

3.4. Additional Moderating Factors

Because of differences in capital constraints, it is likely that firm size will play a role in determining the productive impact of nonpecuniary OSS. For smaller firms, nonpecuniary OSS can play a critical role in allowing the IT capabilities of the firm to ramp up quickly, without expensive outlays for pecuniary software. However, as firms grow, it is likely that such cost savings will have less of an effect, and the costs of nonpecuniary OSS may outweigh the productivity benefits despite the fact that larger firms will have more resources to extract value from the use of nonpecuniary OSS.

4. Empirical Methodology

This section describes the empirical methodology employed to examine the puzzle developed above. First, it describes the estimation model, which is consistent with other models of the productivity of IT but accounts for nonpecuniary digital inputs. Then, it discusses identification concerns due to possible endogeneity as well as the methodologies employed to address these concerns.

4.1. Estimation Models

The data set will measure capital, labor, and various IT inputs at the firm. Before describing these 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 et al. 2012, Huang et al. 2016):

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

where Y_{it} is the production of firm i in time t, K_{it} is the amount of non-IT capital stock, and L_{it} is the amount of non-IT labor. Note that IT_{it} is the amount of IT capital stock, and A_{it} is a firm-specific efficiency multiplier that captures intangible assets such as management skill, institutional knowledge, and learning; IT_{it} captures both the value of IT hardware at the firm and three times the value of IT labor at the firm because of the importance of IT labor being used for internal software development efforts, the result of which is a capital good (Brynjolfsson and Hitt 1996, Hitt and Brynjolfsson 1996, Dewan and Min 1997, Huang et al. 2016).

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

$$\ln(VA_{it}) = \alpha \ln K_{it} + \beta \ln L_{it} + \gamma \ln IT_{it} + \ln A_{it} + \varepsilon_{it}. \tag{2}$$

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 1% 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 et al. 2012, Huang et al. 2016). 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 nonpecuniary OSS). Such an input is essentially uncounted 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 nonpecuniary open source software, $nonpecuniary\_OSS_{it}$, in a given period is added to the specification. Nonpecuniary OSS must be separated from pecuniary OSS because the latter is already measured by current productivity methods since it has a price.[11] The measurement of nonpecuniary 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:

$$\begin{aligned}\ln(VA_{it}) = {} & \alpha \ln K_{it} + \beta \ln L_{it} + \gamma_1 \ln IT_{it} \\ & + \gamma_2 \ln nonpecuniary\_OSS_{it} + \ln A_{it} + \varepsilon_{it}.\end{aligned} \tag{3}$$

Using Equation (3) as the preferred estimation equation, an estimate of the impact of nonpecuniary OSS usage can be obtained. It should be noted that in such a production function framework, when faced with an input with a price of zero (such as nonpecuniary OSS), a profit-maximizing firm should use an infinite amount of it. However, as mentioned in the discussion of TCO above, although the cost of the software may be zero, the cost of implementing it is not. At the very least, a pecuniary piece of hardware (server or PC) must be purchased for the software to run on. In particular, as mentioned above, since the empirical focus of this paper is on operating systems, one piece of hardware is required for each piece of nonpecuniary OSS. Therefore, since the total cost of implementation is greater than zero, demand for the nonpecuniary input will not be infinite.

As discussed above, it is likely that firms require an ecosystem of complements, including technical capabilities, to be able to fully gain productive value from the use of nonpecuniary OSS. As a proxy for the existence of such an ecosystem, a measure of the firm's IT intensity or of whether the firm is in an IT-producing industry is used. Although neither of these proxies directly measures the complete ecosystem of complements (including organizational practices or culture), they both are likely indicators that such an ecosystem may exist. Both measures will be used as an interaction term with the nonpecuniary OSS measure to understand the role the ecosystem of complements plays in obtaining productive value from nonpecuniary OSS. Furthermore, in such specifications the proxy for complementary capabilities (either IT intensity or IT producer status) will be interacted with the amount of IT capital at the firm to ensure that the coefficient on the interaction with nonpecuniary OSS is not capturing some additional level of benefit such firms obtain from all IT. As a robustness check for the primary specifications, the complementary capability proxy will be interacted with all variables in the regression to ensure this is also not occurring with any other inputs.

4.2. Identification Strategy

In an ideal experiment, one would randomly assign firms from the full population of U.S. firms to use or not use nonpecuniary OSS at varying levels of intensity. However, such an experiment is infeasible, and therefore observational data, discussed in the next section, are used. Like all studies of the impact of IT on productivity using observational data, this analysis is subject to endogeneity concerns. The primary endogeneity concern is the fact that firms endogenously decide whether or not to use nonpecuniary OSS and what amount to use. If firms that are, for example, better managed are both more likely to use nonpecuniary OSS and have higher levels of productivity, then the relationship between nonpecuniary OSS and productivity could not be interpreted as causal because of simultaneity bias. Furthermore, this could lead to an incorrect estimation of the size of the effect. An additional concern with using observational data is that there are measurement errors or assumptions that could impact the accuracy of the results. Furthermore, additional measurement error concerns arise since the observational data are based on self-reported surveys. All of these concerns would prevent a complete answer to the primary question that can be used to make recommendations to managers. Therefore, this paper employs a number of methods that help to address both of these concerns. These methods allow for the coefficient on use of nonpecuniary OSS to be interpreted in a more causal manner. Additionally, since not all micro characteristics of the firm are observed, the resulting coefficient can be interpreted as the impact of not only the nonpecuniary OSS itself but also the complementary assets 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 et al. 2002, Aral et al. 2012, Brynjolfsson and Milgrom 2012).

The length and size of the empirical data allow for a variety of methods to be used to address the primary endogeneity concerns discussed above. To start with, to control for unobserved time and industry trends, the models uses year fixed effects and industry fixed effects at the two-digit North American Industry Classification System (NAICS) level. The combination of these approaches helps eliminate unobserved time and industry effects that may bias the results. Although the two-digit NAICS-level industry definition is the traditional level of industry control used in productivity analysis, the more granular five-digit NAICS-level industry definition is also used as a robustness check. Since the data set is a panel, firm fixed effects models can be used in addition to standard ordinary least squares (OLS) regression. Such models take an additional step in the causal direction by controlling for unobserved firm characteristics and comparing a firm against itself, rather than against other firms. However, 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. However, because of the length of the panel (10 years), more advanced methods may be used to obtain even greater identification without sacrificing accuracy. In particular, the Arellano–Bond method (Arellano and Bond 1991) and the Blundell–Bond method (Blundell and Bond 1998) for dynamic panel analysis are used as the primary specifications.[12] Both of these methods were developed to better understand the causal nature of microproductivity estimations and have been used in studies of the productivity of IT (e.g., Tambe and Hitt 2012). Arellano–Bond (ABOND) uses lagged differences of production inputs as instruments to account for the endogeneity inherent in all inputs into productivity, including the amount of nonpecuniary OSS used at the firm. More specifically, ABOND uses a one-period lag of the dependent variable (value-added productivity) as a control and uses two-period, or longer, lags as instruments in a generalized method of moments (GMM) estimation. The Blundell–Bond (BBOND) method extends ABOND to create a system estimator that only requires a one-period lag for the instruments and reduces a small downward bias that occurs in ABOND when the actual value of a coefficient is high.

Although ABOND and BBOND help take a large step in the direction of interpreting the coefficients in a causal manner, additional specifications will be used as robustness checks against the main effect to add further support for a more causal interpretation and to further address measurement error concerns. These include the Levinsohn–Petrin GMM estimator (Levinsohn and Petrin 2003), inverse probability weighting (Horvitz and Thompson 1952; Wooldridge 2002, 2007), coarsened exact matching (Iacus et al. 2012), and an instrumental variable approach using two-stage least squares with two instruments. Finally, to further address the concern of simultaneity bias due to management quality, data from the World Management Survey (Bloom and Van Reenen 2007) are used to show that high-quality management is not correlated with, nor is it a predictor of, the use of nonpecuniary OSS for a subset of firms in the sample. Each of these methods are discussed further in the robustness check portion of the results section below.

5. Data

The data are split into two primary parts: OSS usage and financial statements, both of which are at the firm-year level. Data on which firms are using OSS come from the Harte Hanks IT Survey, a survey of IT usage at the establishment level for nearly 10,000 firms from 2000 to 2009. This database is used frequently in studies of the impact of IT on firm-level productivity (Brynjolfsson and Hitt 2003; Forman 2005; Forman et al. 2005, 2008; Bloom et al. 2012; Tambe et al. 2012; McElheran 2014; Huang et al. 2016). 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 are 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 are augmented with firm financial data. In particular, data are collected on firm expenditures on labor, capital, research and development (R&D), and advertising, as well as firm revenues and costs of materials. For public firms, this information is available via Standard & Poor's Compustat database. The firm's stock ticker symbol is used to match the Harte Hanks data to the Compustat data. Furthermore, an automated name-matching algorithm, combined with a by-hand analysis, ensured proper matching of firms and establishments in the Harte Hanks data to firms in the Compustat data, including subsidiaries of public firms. Therefore, sites within the Harte Hanks database that are owned by different firms in different years (e.g., via mergers or acquisitions) are associated with the correct parent firm and therefore the correct financial data. Although on average the Harte Hanks database contains information on 419,137 establishments and 9,957 firms per year, the final sample uses only public firms, as the model requires financial information filed in the firm's 10-K. This reduces the sample size to 1,663 firms and indicates that the results can best be applied to public firms. Furthermore, of these 1,663 firms, 97 firms do not report information on their operating system usage (discussed further below) and are dropped, which results in a final sample of 1,566 firms and 10,355 firm-year observations.[13] The sections below detail how these two data sets 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.

5.1. Variables

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 et al. 2016). First, deflated IT labor and non-IT labor (defined below) are both subtracted from the yearly operating costs (XOPR in Compustat), the result of which is deflated by the BLS Producer Price Index (PPI) by stage of processing for intermediate materials, supplies, and components at the two- or three-digit NAICS level.[14] The result is then subtracted from yearly sales (SALE in Compustat) deflated by the Bureau of Economic Association (BEA) Gross Domestic Product Price Index for gross output at the two- or three-digit NAICS level.[15]

IT Capital **(IT_{it}) *and IT Intensity*** **(*ITIntensity*$_{it}$).** The standard practice in prior literature in the field is to construct 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 as a result of the importance of IT labor being used for internal software development efforts, the result of which is a capital good (Brynjolfsson and Hitt 1996, Hitt and Brynjolfsson 1996, Dewan and Min 1997, Huang et al. 2016).[16] To calculate the value of IT hardware at the firm, the market value of the IT stock is estimated by multiplying the number of PCs and servers at the firm (from Harte Hanks[17]) by the average price of a PC or server that year from the Economist Intelligence Unit Telecommunications database. The BEA Price Index for nonresidential computers and peripherals is then used to deflate this value.[18] This method is consistent with prior work in this area (e.g., Brynjolfsson and Hitt 1996, Huang et al. 2016).

The value of IT labor is calculated by taking the number of IT workers at each firm (from Harte Hanks[19]) and multiplying by the mean annual wage for all computer and mathematical science occupations.[20] Although wages are the major portion of total compensation, this does not account for benefits, which can be significant in many industries. Therefore, the BLS Employer Cost for Employee Compensation table is used at the one-digit Standard Industrial Classification (SIC) level (for 2000–2003) and the two-digit NAICS level (2004–2009) to inflate the wages to a more accurate measure of total compensation.[21] Then, the BLS Employment Cost Index for wages and salaries for "White collar occupations—Management, professional, and related" at the industry level is used to deflate this value.[22] Consistent with prior literature, the IT labor cost is tripled and added to the IT hardware cost to obtain an aggregate value for IT capital. Additionally, because the costs of IT hardware and IT labor are being imputed, a robustness check using the raw number of PCs, servers, and IT employees will be run and shows that the results are consistent.

As discussed above, IT intensity will be used as a proxy for understanding whether or not a firm has the ecosystem of complements, including technical capabilities, necessary to extract productive value from its use of nonpecuniary OSS. The IT intensity variable is constructed by dividing the deflated value of the IT hardware at the firm in a given year by deflated sales in that year. This measure is consistent with other measures of IT intensity used in the literature (e.g., Chang and Gurbaxani 2012) and will be used as a continuous measure of the intensity of the firm's IT capital.

Non-IT Capital **(K_{it}).** Capital expenditure is reported in Compustat as yearly net total property, plant, and equipment (PPENT). However, this value relies on accounting measures of depreciation (rather than economic measures) and does not include any deflation for different vintages of capital equipment. Therefore, the methodology from Villalonga (2004), which builds on Lindenberg and Ross (1981), is used to develop a more accurate measure of capital stock.[23] Starting with PPENT data from 1987 or the year the firm became public, whichever occurs most recently, a value for capital stock based on the nominal investment from the current year is recursively constructed. The value of capital stock in year $t-1$ is first deflated using the appropriate gross domestic product (GDP) deflator for private nonresidential fixed investment from the National Income and Product Accounts, table 5.3.4. The resulting value is then depreciated by 5% (the standard rate of depreciation used in economic analysis; see, e.g., Villalonga 2004). Then, the nominal investment in capital stock for the current year t is calculated as the difference in PPENT from $t-1$ to t (since PPENT is itself reported in Compustat as a stock variable) plus the depreciation value (DP in Compustat). This calculation of the nominal investment in year t is added to the previously calculated value of capital stock in $t-1$. The resultant value represents the economically depreciated and deflated value of capital stock at the firm in year t. The value of IT hardware (defined above) is then subtracted from this value so that the hardware is not double-counted.

Non-IT Labor **(L_{it}).** Non-IT labor is constructed using a hybrid approach. When the firm reports its total labor expense (XLR in Compustat), this is used as the baseline labor expense. If the reporting footnote code indicates that this number is only for wages (and does not include benefits), the approach described above for IT labor is used to inflate the wages to a more accurate measure of total compensation. In both of these cases, IT labor is then subtracted from this total labor expense to yield the amount invested in non-IT labor.

In cases where XLR is not reported in Compustat, the labor cost must be imputed. This is done by 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 in that industry (at the three-digit SIC level for 2000–2001 and four-digit NAICS level for 2002–2009)[24] that year. For all methods of calculating non-IT labor expenses, the BLS Employment Cost Index for wages and salaries for private industry workers at the two-digit NAICS level is then used to deflate this result. This method of calculation is consistent with prior studies on IT productivity (Bresnahan et al. 2002, Brynjolfsson and Hitt 2003, Bloom and Van Reenen 2007). However, because the cost of labor is imputed, a robustness check with the raw number of non-IT employees shows the results are consistent.

Technology, R&D, and Brand Controls. Although the variables above capture expenditures on IT-related inputs, they do not capture how "cutting edge" the technology is at the firm. Therefore, the Harte Hanks data are used to construct measures of how widely spread such advanced technologies are at the firm. Each site interviewed reports whether it uses data warehousing software (DWS), database management software (DBMS), and/or enterprise resource planning software (ERP). For each of these technologies, a variable is constructed that represents the percentage of sites at the firm that have this technology (*DWS*%, *DBMS*%, and *ERP*%). This variable is equal to 0 if no sites at the firm use the technology and 1 if all sites at the firm use the technology.

Beyond the inclusion of high-level capital and labor variables, investments in R&D and brand (proxied by advertising) have been shown to have important effects on productivity (Brynjolfsson et al. 2002, Villalonga 2004, Hall et al. 2005, Corrado et al. 2009, Bardhan et al. 2013, and Saunders and Brynjolfsson 2016). Therefore, when these variables are reported in Compustat (XRD and XAD), they are used to create stock variables that are used as controls. If the variables are not reported, the assumption that the value is 0, rather than missing, is made. R&D stock and brand stock are calculated in a recursive manner, similar to the non-IT capital stock discussed above, following the methodology of Villalonga (2004). For R&D, the value reported in Compustat is a flow and represents the investments made in R&D in the current year. Similar to non-IT capital stock, the recursive construction of the variable starts in 1987 or the year the firm became public, whichever is more recent. The R&D stock in $t-1$ is first depreciated using the industry-specific rates based on the BEA/National Science Foundation data in Li and Hall (2016). For industries not appearing in this, a 15% depreciation rate is used, which is consistent with prior literature (Villalonga 2004, Saunders and Brynjolfsson 2016). The resulting value is then deflated by using the values for private businesses from table 4.1 of the BEA R&D Satellite Account from December 2010.[25] It is then added to the XRD value from Compustat for the current year to obtain an accurate measure of R&D stock. For brand (proxied by advertising stock), the same process is used, although a consistent 45% is used as the depreciation rate and PPI for advertising agencies is used to deflate values to current-year dollars.[26]

Nonpecuniary Open Source Software Usage. To measure nonpecuniary OSS usage at the firm, the number and type of operating systems used at the firm are measured. Although operating systems are certainly not the only nonpecuniary OSS used at the firm, they are important and frequently indicate the wider use of nonpecuniary OSS. Furthermore, the Harte Hanks survey asks firms what type of operating systems they use but does not always capture other types of nonpecuniary OSS. Because this only captures nonpecuniary OSS operating systems, the data set necessarily underestimates the total amount of nonpecuniary OSS used at the firm.

In addition to constructing a measure of nonpecuniary OSS operating systems, measures of pecuniary OSS and closed source operating systems are also constructed for use as controls and to better understand the software used at the firm. These three measures ($nonpecuniary\_OSS_{it}$, $pecuniary\_OSS_{it}$, 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 do not report the precise number of operating systems in use at a given firm. The data do, however, report the different types of operating systems used at each site at the firm. These operating systems are classified into three categories: nonpecuniary OSS, pecuniary OSS, or closed source. Table 1 shows the OSS operating systems in the data set.[27] 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 split over the total number of PCs at the site in a manner that is consistent with the overall market share for installed systems as reported by Gartner in its report "PC Installed Base by Operating System, 2000–2009" (Gartner 2010). However, before this split can occur, a correction must be applied for the fact that some sites and firms only use closed operating systems. For example, if a given site reports 100 PCs; one open source operating system, one pecuniary open source operating system, and one closed operating system; and the market share information for that year indicates that 90% of PCs were closed source, 7% were open source, and 3% were pecuniary open source, then a naïve approach to splitting the operating systems at the site would be to consider it to have 90 closed source PCs, 7 open source PCs, and 3 pecuniary open source PCs. However, this assumes that all sites have at least one of each of the three operating systems, which is not the case. Instead, the total number of PCs at all sites at all firms in the sample in a given year is calculated. Then the yearly split from Gartner is applied to the entire set of PCs in the sample to give a total number of PCs with each of the three types of operating systems. Then the number of PCs at sites with only closed operating systems is subtracted from the total number of PCs with closed operating systems. An updated percentage is then calculated to represent the likely breakdown of the three operating systems at the remaining firms (that report using an OSS operating system). This method allows for an accurate estimation of the total number of each type of PC operating system at each firm.[28] The same is done for servers.

This calculation yields an estimate of how many instances of a given type of operating system exist at the site in a manner that accounts for the average distribution of the three types. This is then aggregated to the firm level and divided by the number of surveyed

Table 1. Open Source Operating Systems

Pecuniary OSS operating systems	Nonpecuniary OSS operating systems	
Red Hat Linux	Berkeley Software Distribution (BSD)	Gentoo Linux
SUSE Linux	Conectiva	Mandrake Linux
SCO Linux	Debian	NetBSD
TurboLinux	Fedora	OpenBSD
	FreeBSD	Ubuntu

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 firmwide imputation of the number of each type of operating system that accounts for unsurveyed sites. This estimation will allow for a more granular interpretation of the primary effect.[29]

5.2. Summary Statistics

Table 2 shows the descriptive statistics of the firms in the data set. As mentioned above, there are 10,355 firm-year observations from 1,566 firms in the data set. 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, because of the Harte Hanks sampling methodology, larger firms are overrepresented in the sample, and very small firms (e.g., start-ups) 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 sized. For example, as shown in Table 2, the smallest company in the sample (Inter Tel Inc.) had sales of $4.67 million in its lowest-selling year (2001). Comparatively, the largest firm (Exxon Mobil Corp.) had sales of $425 billion in its highest-selling year (2008). Therefore, results should be interpreted as applying to medium and large firms. The firms in the data set 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 5,739.214, while the mean number of nonpecuniary OSS and pecuniary OSS operating systems are much lower at 67.496 and 130.578, respectively. Looking deeper into the data, there are 3,257 observations where firms use at least one nonpecuniary OSS operating system. For these 3,257 observations, the average number of nonpecuniary OSS operating systems is 214.571; 6,095 observations use no OSS (pecuniary or nonpecuniary) at all. Only 10 observations use exclusively OSS (pecuniary or nonpecuniary). These 10 observations comes from two companies, Novell Inc., a producer of nonpecuniary OSS, and Evans & Sutherland Computer Corp, a small producer of advanced computer graphics equipment. The results are robust to dropping both of these companies.

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 nonpecuniary OSS and the other two types of operating systems, pecuniary OSS and closed, are fairly low, while the correlation between pecuniary OSS and closed is comparatively high. Table 4 shows the breakdown of observations by industry using the two-digit NAICS classification. While 51% of the observations are from the manufacturing industry, there is also good representation from other key industries, such as trade, finance, services, and information. Furthermore, Table 4 shows the percentage of firms within the industry that use nonpecuniary OSS or any type of OSS operating system. The percentage of firms in an industry using nonpecuniary OSS varies between 16.33% (Retail Trade, NAICS 44–45) and 47.63% (Information, NAICS sector 51), with an overall average of 31.45%, and has no major outliers. The percentage of firms in an industry using any OSS varies between 21.70% (Retail Trade) and 66.67% (Agriculture, Forestry, Fishing and Hunting, NAICS sector 11). However, this maximum should

Table 2. Descriptive Statistics

Variable	Obs	Mean	Std. dev.	Min	Max
$sales_{it}$	10,355	6,423.017	19,607.140	4.671	425,071
VA_{it}	10,355	2,365.361	5,664.572	0.055	109,337
IT_{it}	10,355	464.896	1,982.940	0.434	83,801.52
K_{it}	10,355	4,178.616	14,252.850	0.350	211,839.5
L_{it}	10,355	1,309.277	3,517.399	0.221	80,913.23
$nonpecuniary\_OSS_{it}$	10,355	67.496	840.253	0	55,626.42
$pecuniary\_OSS_{it}$	10,355	130.578	3415.786	0	279,883.8
$closed_{it}$	10,355	5739.214	19,762.57	0	1,028,182
RD_{it}	10,355	592.212	2,836.813	0	55,535.63
AD_{it}	10,355	151.916	816.881	0	17,004.61
ERP%	10,355	0.114	0.198	0	1
DWS%	10,355	0.070	0.141	0	1
DBMS%	10,355	0.648	0.253	0	1
$ITIntensity_{it}$	10,355	0.003	0.006	0.0000015	0.248

Notes. Values for monetary variables are in millions of constant 2009 U.S. dollars; K_{it}, RD_{it}, and AD_{it} represent total stock value, rather than yearly flow value. Values for operating systems are in number of computers at the firm running operating systems in that category.

Table 3. Correlation Matrix

	VA_{it}	IT_{it}	K_{it}	L_{it}	$nonpecuniary\_OSS_{it}$	$pecuniary\_OSS_{it}$	$closed_{it}$	RD_{it}	AD_{it}	$ITIntensity_{it}$
VA_{it}	1.0000 (–)									
IT_{it}	0.5067 (0.000)	1.0000 (–)								
K_{it}	0.7572 (0.000)	0.3009 (0.000)	1.0000 (–)							
L_{it}	0.8616 (0.000)	0.6926 (0.000)	0.5995 (0.000)	1.0000 (–)						
$nonpecuniary\_OSS_{it}$	0.1000 (0.000)	0.0630 (0.000)	0.0865 (0.000)	0.1144 (0.000)	1.0000 (–)					
$pecuniary\_OSS_{it}$	0.1660 (0.000)	0.0495 (0.000)	0.0923 (0.000)	0.1533 (0.000)	0.0231 (0.019)	1.0000 (–)				
$closed_{it}$	0.4669 (0.000)	0.1463 (0.000)	0.3686 (0.000)	0.4126 (0.000)	0.1045 (0.000)	0.6268 (0.000)	1.0000 (–)			
RD_{it}	0.4361 (0.000)	0.0840 (0.000)	0.2955 (0.000)	0.2896 (0.000)	0.1114 (0.000)	0.1018 (0.000)	0.3143 (0.000)	1.0000 (–)		
AD_{it}	0.5534 (0.000)	0.2164 (0.000)	0.5656 (0.000)	0.4758 (0.000)	0.0869 (0.000)	0.0831 (0.000)	0.3335 (0.000)	0.4214 (0.000)	1.0000 (–)	
$ITIntensity_{it}$	−0.0643 (0.000)	−0.0381 (0.001)	−0.0494 (0.000)	−0.0523 (0.000)	0.0262 (0.008)	0.0921 (0.000)	0.2140 (0.000)	−0.0091 (0.353)	−0.0255 (0.009)	1.0000 (–)

Note. Values in parentheses are the significance level of the correlation.

be considered an outlier because NAICS 11 has a low number of observations, and most firms in the agriculture, forestry, fishing, and hunting industry are not publicly traded. Therefore, the more realistic range is between 21.70% and 60.88%, with an overall average of 41.14%.

6. Results and Discussion

This section presents the results of the empirical analysis and discusses the interpretation of these results. First, basic three-factor productivity results are presented and compared with those of other studies to confirm the consistency of the data and methods with prior research. Then, the baseline results from the inclusion of the continuous measurement of nonpecuniary OSS are presented. Following this, a detailed examination of the value-added productivity impact of nonpecuniary OSS for firms that are either intense IT users or are in IT-producing industries is given. Finally, an analysis of the relationship of the firm's management quality and decision to use nonpecuniary OSS is presented to help rule out concerns of simultaneity bias.

6.1. 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. Table 5 shows the results of the basic three-factor productivity analysis. Models 1–4 use OLS regression with increasingly restrictive fixed effects. Model 1 uses no fixed effects, model 2 introduces a year fixed effect, model 3 adds a NAICS-2 industry effect, and model 4 uses a NAICS-5 industry fixed effect. Model 5 uses panel regression with firm fixed effects, model 6 uses Arellano–Bond estimation, and model 7 uses Blundell–Bond estimation.[30] For models 1–5, the standard errors are robust and clustered by firm to account for any serial correlation in the error terms since the data set contains multiple observations of the same firm over different time periods (Angrist and Pischke 2009, Imbens and Kolesar 2016). Clustered standard errors are not possible in dynamic panel analysis, and therefore heteroskedasticity robust standard errors are used for models 6 and 7. Furthermore, as mentioned above, dynamic panel analysis relies on using lags (two periods for ABOND and one period for BBOND) for identification purposes, and therefore the sample size is reduced from 10,355 firm-year observations in the OLS models to 8,988 in the BBOND model and 7,650 in the ABOND model. Likewise, the number of firms (groups) in these models is reduced as well since some firms have fewer than three years of observations. As such, the GMM-type instruments for this analysis include the two-period lag (and prior) for ABOND and the one-period lag (and prior) for BBOND. Furthermore, first differences of all other right-hand-side variables are used as instruments for the differenced equations. Additionally, for the ABOND and BBOND models, the order 1 and order 2 autocorrelation test values are shown. For the results to be valid, we must reject the null of no autocorrelation for order 1 and not reject the null of no

Table 4. Industry Breakdown

Two-digit NAICS	Description	Frequency	% of all firms	% of firms using non-pecuniary OSS	% of firms using any OSS
11	Agriculture, Forestry, Fishing and Hunting	30	0.29	43.33	66.67
21	Mining, Quarrying, Oil and Gas Extraction	380	3.67	30.00	34.74
22	Utilities	624	6.03	29.81	38.30
23	Construction	172	1.66	35.47	38.95
31–33	Manufacturing	5,316	51.34	31.23	41.59
42	Wholesale Trade	500	4.83	29.40	41.40
44–45	Retail Trade	447	4.32	16.33	21.70
48–49	Transportation and Warehousing	303	2.93	25.08	33.66
51	Information	611	5.9	47.63	60.88
52	Finance and Insurance	567	5.48	22.75	30.16
53	Real Estate, Rental, and Leasing	107	1.03	26.17	29.91
54	Professional, Sci., and Tech. Services	410	3.96	42.44	53.90
56	Admin. and Support and Waste Management Services	323	3.12	32.20	43.34
61	Educational Services	56	0.54	37.50	42.86
62	Health Care and Social Assistance	273	2.64	39.93	50.92
71	Arts, Entertainment, and Recreation	48	0.46	22.92	35.42
72	Accommodation and Food Services	142	1.37	34.51	36.62
81	Other Services (except Public Admin.)	46	0.45	23.91	36.96
All		10,355	100	31.45	41.14

autocorrelation for order 2. The high R^2 values in models 1–4 are characteristic of such productivity studies.

The confidence interval of the coefficient on IT capital in model 5 overlaps with that of Huang et al. (2016), whose methodology this study most closely resembles. Furthermore, the model 3 coefficient on IT capital overlaps with the confidence interval of Brynjolfsson and Hitt (2003) in their one-year difference model with year and industry controls. The IT capital coefficient in model 3 is also very similar to the estimate of Tambe and Hitt (2012, table 9, column 1). These similarities help to add support to the validity of the data set used in this study.

6.2. Returns to Adoption of Nonpecuniary Open Source Software

Table 6 shows the results of including the continuous measure of nonpecuniary OSS adoption in the analysis. All models include a year fixed effect; an industry fixed effect (either NAICS-2 or NAICS-5); and controls for R&D stock, brand (proxied via advertising stock), and advanced technology intensity as defined above. Standard errors are clustered at the firm level for models 1–3 and are heteroskedasticity robust for models 4–8. The results show a negative and significant impact of nonpecuniary OSS when using the OLS model with a NAICS-2 industry control in model 1. However, when more restrictive methods are used in models 2–5, the coefficient is not distinguishable from zero, indicating a lack of support for a causal impact. Model 2 uses a NAICS-5 industry control, model 3 uses a firm fixed effect OLS, model 4 uses Arellano–Bond estimation, and model 5 uses Blundell–Bond estimation. Together, this analysis indicates that the adoption of nonpecuniary OSS has an impact on productivity that is not distinguishable from zero in the more restrictive models. After examining the direct effect, the preferred specification (the Arellano–Bond analysis) is used to calculate various interaction results related to firm size to further examine what types of firms may benefit from nonpecuniary OSS. As discussed above, it is likely that the cost benefit from nonpecuniary OSS matters more for smaller firms, and therefore we would expect smaller firms to have a more positive impact from using nonpecuniary OSS. Models 6–8 show the results of this analysis. The results are substantively similar for an analysis using the Blundell–Bond specification but are not shown because of space constraints. For models 7 and 8, the term "high" indicates that a firm is in the 75th percentile or above in the main sample. The interaction of nonpecuniary OSS with various measures of firm size—inclusion in the Fortune 1000, amount of sales, and number of employees—helps to better understand the relationship of firm size and the ability to gain productive value from the use of nonpecuniary OSS. All three indicate a negative and significant coefficient on the interaction term, indicating that larger firms receive less of a benefit from using nonpecuniary OSS than do smaller firms. However, although the baseline coefficient for nonpecuniary OSS is positive for smaller firms in all three models, it is only significant in model 6 (and even then, it is only significant at the 10% level).

The above analyses include all firms in the sample, regardless of adoption of OSS. Robustness checks

Table 5. Three-Factor Productivity Results

DV: *Value-Added* (VA_{it}) Model type:	1 OLS	2 OLS	3 OLS	4 OLS	5 OLS FE	6 ABOND	7 BBOND
IT Capital (IT_{it})	0.073***	0.066***	0.049***	0.039***	0.026***	0.023***	0.023***
	(0.008)	(0.009)	(0.008)	(0.007)	(0.006)	(0.005)	(0.005)
Non-IT Capital (K_{it})	0.305***	0.302***	0.257***	0.224***	0.059	−0.155	−0.081
	(0.013)	(0.013)	(0.014)	(0.019)	(0.048)	(0.102)	(0.095)
Non-IT Labor (L_{it})	0.669***	0.676***	0.728***	0.769***	0.789***	0.698***	0.680***
	(0.016)	(0.017)	(0.017)	(0.022)	(0.040)	(0.054)	(0.051)
Constant	0.216***	0.189***	0.193	0.311***	1.109***	0.021	−133.220
	(0.049)	(0.056)	(0.177)	(0.054)	(0.254)	(0.831)	(279.056)
Year fixed effect?	N	Y	Y	Y	Y	Y	Y
Industry fixed effect	—	—	NAICS-2	NAICS-5	—	NAICS-2	NAICS-2
Number of firm-year obs.	10,355	10,355	10,355	10,355	10,355	7,650	8,988
Number of firms (groups)	1,566	1,566	1,566	1,566	1,566	1,424	1,501
R^2 (within for panel)	0.910	0.911	0.930	0.952	0.444	—	—
Autocorrelation test order 1	—	—	—	—	—	−4.889	−5.545
z-score (p-value)						(0.000)	(0.000)
Autocorrelation test order 2	—	—	—	—	—	0.332	0.336
z-score (p-value)						(0.740)	(0.737)

Notes. Standard errors are clustered at the firm level for models 1–5 and are heteroskedasticity robust for models 6 and 7. All variables are the natural log of the underlying variable. DV, dependent variable; FE, fixed effects.
$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

Table 6. Continuous Adoption of Nonpecuniary OSS

DV: *Value-Added* (VA_{it}) Model type:	1 OLS	2 OLS	3 OLS FE	4 ABOND	5 BBOND	6 ABOND	7 ABOND	8 ABOND
IT Capital (IT_{it})	0.054***	0.040***	0.026***	0.023***	0.022***	0.022***	0.022***	0.022***
	(0.008)	(0.007)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Non-IT Capital (K_{it})	0.260***	0.226***	0.060	−0.154	−0.088	−0.157	−0.159	−0.160
	(0.014)	(0.019)	(0.049)	(0.102)	(0.087)	(0.102)	(0.102)	(0.102)
Non-IT Labor (L_{it})	0.725***	0.768***	0.790***	0.698***	0.673***	0.697***	0.696***	0.695***
	(0.017)	(0.022)	(0.040)	(0.054)	(0.050)	(0.054)	(0.054)	(0.054)
nonpecuniary_OSS$_{it}$	−0.004***	−0.001	−0.000	−0.000	0.000	0.002*	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)
nonpecuniary_OSS$_{it}$ × *Fortune 1000*						−0.003**		
						(0.001)		
nonpecuniary_OSS$_{it}$ × *High Sales*							−0.002**	
							(0.001)	
nonpecuniary_OSS$_{it}$ × *High Employee*								−0.002***
								(0.001)
Constant	0.134	0.292***	1.115***	0.014	0.184	−0.068	−0.023	−0.022
	(0.197)	(0.063)	(0.260)	(0.699)	(0.500)	(0.705)	(0.697)	(0.698)
Year fixed effect?	Y	Y	Y	Y	Y	Y	Y	Y
Industry fixed effect	NAICS-2	NAICS-5	—	NAICS-2	NAICS-2	NAICS-2	NAICS-2	NAICS-2
Number of firm-year obs.	10,355	10,355	10,355	7,650	8,988	7,650	7,650	7,650
Number of firms (groups)	1,566	1,566	1,566	1,424	1,501	1,424	1,424	1,424
R^2 (within for panel)	0.930	0.952	0.444					
Autocorrelation test order 1 z-score (p-value)	—	—	—	−4.903	−5.534	−4.924	−4.934	−4.919
				(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Autocorrelation test order 2 z-score (p-value)	—	—	—	0.320	0.328	0.342	0.327	0.332
				(0.749)	(0.743)	(0.732)	(0.744)	(0.740)

Notes. Standard errors are clustered at the firm level for models 1–3 and are heteroskedasticity robust for models 4–8. All regressions include controls for R&D stock, advertising stock, and advanced technology intensity (percentage of sites in the firm with DBMS, DWS, or ERP). All variables are the natural log of the underlying variable (plus a small number in the case of operating systems, R&D, and advertising). DV, dependent variable; FE, fixed effects.
$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

were run using only adopting firms. These robustness checks show a similar result, that nonpecuniary OSS adoption does not have a direct and statistically significant positive impact on adopting firms. These results are not shown because of space constraints, but they are available in Table A1 of the online appendix.

6.3. Returns to Adoption of Nonpecuniary OSS for IT-Intensive Firms and IT-Producing Firms

As discussed above, it is likely that not all firms will be impacted by nonpecuniary OSS in the same way. In particular, since nonpecuniary OSS is a highly specific input that may require an ecosystem of complements, including advanced technical capabilities, to implement, firms that are more likely to have these complements are more likely to obtain a positive productivity impact from implementing it. The ecosystem at firms that are either IT intensive or are in IT-producing industries are generally more likely to contain these technical capabilities than that at the average firm. Therefore, as discussed above, it is likely that such firms will benefit from the use of nonpecuniary OSS more than firms that are neither intense IT users nor in IT-producing industries. This analysis uses a definition of IT-producing industries based on that of Jorgenson et al. (2005), updated to use the NAICS classification scheme rather than SIC. These industries include NAICS codes 334 (Computer and Electronic Product Manufacturing), 511210 (Software Publishers), 518210 (Data Processing, Hosting, and Related Services), 541511 (Custom Computer Programming Services), and 541512 (Computer Systems Design Services). These industries are used to create a dummy variable, *IT-producer*, that is equal to 1 if the firm is in an IT-producing industry and 0 otherwise. Of the 10,355 firm-year observations in the analysis sample, 1,524 are from firms in IT-producing industries, representing 15% of the sample. Of these, 933 are nonadopters and 591 are adopters of nonpecuniary OSS. Adoption is slightly higher among IT-producing firms (40%) than all firms (31%), but the proportions are fairly similar to those in the overall sample. If, for example, 100% of IT-producing firms adopted nonpecuniary OSS, it would be difficult to obtain an accurate estimate of the coefficient. Since this is not the case, we can more causally interpret the resulting coefficient for IT-producing firms. The measure of IT intensity is continuous and is described above in Section 5. Importantly, since the IT-intensity term is crafted from a combination of a right-hand-side variable (*IT*) and a portion of the left-hand-side variable (*sales*), precisely interpreting this coefficient is difficult. Therefore, a binary measure of high IT intensity is also used for ease of interpretation. This dummy variable is equal to 1 if the firm is in the 95th percentile of IT intensity and 0 otherwise.[31]

The analyses in Table 7 examine the value-added productivity impact of nonpecuniary OSS for firms that are likely to have the necessary ecosystem of complements. Models 1–3 use the continuous IT intensity variable as a measure of this ecosystem, models 4–6 use the binary measure of IT intensity, and models 7–9 use the IT-producing industry variable. Models 1, 4, and 7 use an OLS estimation procedure with standard errors clustered at the firm level, and models 2, 5, and 8 use the Arellano–Bond estimation procedure with heteroskedasticity robust standard errors. Models 3, 6, and 9 use the Blundell–Bond estimation procedure with heteroskedastic robust standard errors. All models include controls for R&D stock, advertising stock, and advanced technology intensity (the percentage of sites in the firm with DBMS, DWS, or ERP). Models 1–3 also include a control for IT intensity. Models 1–3 include a control for the interaction of IT intensity with IT capital (not shown).[32] All variables are the natural log of the underlying variable (plus a small number in the case of operating systems, R&D, and advertising).

The coefficients on the interaction of nonpecuniary OSS and IT intensity in models 1–3 are positive and significant at least at the 5% level. The confidence intervals overlap, and therefore the coefficient from the Arellano–Bond analysis (model 2) is used for interpretation purposes. The coefficient on the interaction in this case is 0.409. If we consider two firms at the 95th percentile for IT intensity, but one uses OSS and the other does not, we can calculate the productivity boost the user obtains. The 95th percentile of IT intensity is 0.009. Therefore, a firm at this level of IT intensity has an aggregate coefficient on nonpecuniary OSS of $(0.009 \times 0.409) - 0.001 = 0.003$. This can be translated into dollars by considering the average nonpecuniary OSS using firm in the sample. Such a firm has 215 nonpecuniary operating systems and value added of \$3,464 million. Since the estimation procedure is log–log, this means a 1% increase in the amount of nonpecuniary OSS at the firm (or an additional 2.15 systems) leads to a $0.003 \times 0.01 \times \$3{,}464{,}000{,}000 = \$103{,}920$ increase in value added. Models 4–6 show similar results when using the dummy variable for high IT intensity, although the coefficients are slightly higher at $(0.009 - 0.004) = 0.005$ for the OLS specification, $(0.008 - 0.000) = 0.008$ for the Arellano–Bond specification, and $(0.004 - 0.000) = 0.004$ for the Blundell–Bond specification. Using the Arellano–Bond estimate (model 5) as an upper bound, the same 1% increase in nonpecuniary OSS at the firm leads to a $0.008 \times 0.01 \times \$3{,}464{,}000{,}000 = \$277{,}120$ increase in value added. It is important to note that further analysis of the linear combination of the coefficients shows that when considering the 95th percentile of IT intensity, the combination of the coefficient on OSS and the interaction coefficient is consistently

Table 7. Continuous Adoption of Nonpecuniary OSS with IT Intensity and IT Producer Interaction

DV: *Value-Added* (VA_{it}) Model type:	1 OLS	2 ABOND	3 BBOND	4 OLS	5 ABOND	6 BBOND	7 OLS	8 ABOND	9 BBOND
IT Capital (IT_{it})	0.026***	0.025***	0.025***	0.059***	0.023***	0.026***	0.067***	0.022***	0.020***
	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Non-IT Capital (K_{it})	0.052	−0.177*	−0.108	0.259***	−0.160	−0.099	0.255***	−0.157	−0.091**
	(0.049)	(0.100)	(0.086)	(0.007)	(0.100)	(0.085)	(0.007)	(0.102)	(0.025)
Non-IT Labor (L_{it})	0.786***	0.687***	0.663***	0.720***	0.695***	0.666***	0.722***	0.698***	0.674
	(0.042)	(0.054)	(0.049)	(0.009)	(0.054)	(0.049)	(0.009)	(0.054)	(0.018)
$nonpecuniary\_OSS_{it}$	−0.001*	−0.001	−0.001	−0.004***	−0.000	−0.000	−0.005***	−0.001	−0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
$nonpecuniary\_OSS_{it} \times ITIntensity_{it}$	0.317***	0.409**	0.370**						
	(0.115)	(0.187)	(0.164)						
$nonpecuniary\_OSS_{it} \times High\text{-}ITIntensity$				0.009***	0.008***	0.004**			
				(0.002)	(0.003)	(0.002)			
$nonpecuniary\_OSS_{it} \times IT\text{-}producer$							0.008***	0.004**	0.003**
							(0.002)	(0.002)	(0.001)
Constant	0.164	−0.001	0.333	0.155	−0.043	0.288	0.153	−0.025	0.192
	(0.197)	(0.698)	(0.505)	(0.101)	(0.683)	(0.500)	(0.098)	(0.705)	(0.216)
Year FE?	Y	Y	Y	Y	Y	Y	Y	Y	Y
Industry FE	NAICS-2	NAICS-2	NAICS-2	NAICS-2	NAICS-2	NAICS-2	NAICS-2	NAICS-2	NAICS-2
Number of firm-year obs.	10,355	7,650	8,988	10,355	7,650	8,988	10,355	7,650	8,988
Number of firms (groups)	1,566	1,424	1,501	1,566	1,424	1,501	1,566	1,424	1,501
R^2	0.931			0.931			0.933		
Autocorrelation test order 1 z-score (p-value)		−4.944	−5.439		−4.944	−5.498		−4.904	−5.418
		(0.000)	(0.000)		(0.000)	(0.000)		(0.000)	(0.000)
Autocorrelation test order 2 z-score (p-value)		0.327	0.319		0.370	0.322		0.365	0.321
		(0.744)	(0.749)		(0.712)	(0.748)		(0.715)	(0.742)
p-value for test of combined OSS and interaction coefficient				0.062	0.008	0.052	0.051	0.084	0.098

Notes. Standard errors are clustered at the firm level for models 1, 4, and 7, and they are heteroskedastic robust for all other models. All regressions include controls for R&D stock, advertising stock, advanced technology intensity (percent of sites in the firm with DBMS, DWS, or ERP). An additional control for IT intensity is also included in models 1–3. Models 1–3 include an interaction of IT intensity with IT capital (not shown). Similarly, models 4–6 include an interaction of the binary high IT intensity with IT capital. Likewise, models 7–9 include an interaction of IT producer with IT capital (not shown). All variables are the log of the underlying variable (plus a small number in the case of operating systems, R&D, and advertising). DV, dependent variable; FE, fixed effects.
***$p < 0.01$; **$p < 0.05$; *$p < 0.1$.

significant at the 10% level (or lower). The p-values for this test are shown in the final row of Table 7. Additional tests show that these results hold for firms that are as low as the 77th percentile of IT intensity. However, when considering the specification in model 5, using a cutoff of the 77th percentile results in a coefficient on the use of OSS of $(0.003 - 0.001) = 0.002$. In this case, a 1% increase in nonpecuniary OSS at the firm leads to a $0.002 \times 0.01 \times \$3{,}464{,}000{,}000 = \$69{,}280$ increase in value added, indicating a 75% drop in the coefficient (and the resulting increase in value added) when going from a 95th percentile cutoff to a 77th percentile cutoff.

Models 7–9 use whether or not a firm is in an IT-producing industry as an alternative measure of the ecosystem of complements at the firm. Models 7–9 consider all firms in the same regression and use an interaction term to assess the difference between IT-producing firms and others. Model 7 uses OLS, model 8 uses Arellano–Bond, and model 9 uses Blundell–Bond. All three include a control for the interaction of *IT-producer* with *IT Capital* (not shown). The coefficients for the variables of interest are positive and significant. After accounting for the baseline effect, the aggregate coefficients in these three models indicate that IT-producing firms have a coefficient of 0.003 on nonpecuniary OSS usage. This is very similar to the coefficient found for firms above the 95th percentile of IT intensity discussed above, and therefore the translation into dollars is similar. As with above, the combination of the coefficient on OSS and the interaction coefficient is consistently significant at the 10% level (or lower), and the p-values for this test are shown in the final row of Table 7.

Motivated by the results related to firm size in models 6–8 of Table 6, additional tests are run to examine the coefficient on a triple interaction of nonpecuniary OSS usage, the ecosystem of complements (IT intensity or IT producer status), and size of the firm (Fortune 1000 status, high sales, or high employee counts).

Table 8(a). Nonpecuniary OSS with IT Intensity Interactions with Firm Size

DV: *Value-Added* (VA_{it}) Model type:	1 OLS	2 ABOND	3 OLS	4 ABOND	5 OLS	6 ABOND
IT Capital (IT_{it})	0.061***	0.023***	0.061***	0.023***	0.062***	0.023***
	(0.008)	(0.005)	(0.008)	(0.005)	(0.008)	(0.005)
Non-IT Capital (K_{it})	0.257***	−0.177*	0.257***	−0.178*	0.258***	−0.179*
	(0.014)	(0.097)	(0.014)	(0.097)	(0.014)	(0.097)
Non-IT Labor (L_{it})	0.718***	0.700***	0.718***	0.700***	0.718***	0.700***
	(0.017)	(0.050)	(0.017)	(0.050)	(0.017)	(0.050)
$nonpecuniary\_OSS_{it}$	−0.004***	−0.001*	−0.004***	−0.001*	−0.005***	−0.001*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$nonpecuniary\_OSS_{it} \times ITIntensity_{it}$	0.424***	0.791***	0.412***	0.554**	0.394***	0.611**
	(0.149)	(0.299)	(0.143)	(0.278)	(0.136)	(0.253)
$nonpecuniary\_OSS_{it} \times ITIntensity_{it} \times Fortune\ 1000$	−0.231	−0.647**				
	(0.227)	(0.325)				
$nonpecuniary\_OSS_{it} \times ITIntensity_{it} \times High\ Sales$			−0.200	−0.224		
			(0.195)	(0.292)		
$nonpecuniary\_OSS_{it} \times ITIntensity_{it} \times High\ Employee$					−0.084	−0.442
					(0.226)	(0.305)
Constant	0.198***	−0.037	0.196***	0.028	0.192***	0.031
	(0.064)	(0.697)	(0.063)	(0.695)	(0.063)	(0.694)
Year FE?	Y	Y	Y	Y	Y	Y
Industry FE	NAICS-2	NAICS-2	NAICS-2	NAICS-2	NAICS-2	NAICS-2
Number of firm-year obs.	10,355	7,650	10,355	7,650	10,355	7,650
Number of firms (groups)	1,566	1,424	1,566	1,424	1,566	1,424

Notes. Standard errors are clustered at the firm level for models 1, 3, and 5 and are heteroskedastic robust for all other models. All regressions include controls for R&D stock, advertising stock, and advanced technology intensity (percentage of sites in the firm with DBMS, DWS, or ERP) (not shown). All models include IT intensity and an interaction of IT intensity with IT capital (not shown). All variables are the natural log of the underlying variable (plus a small number in the case of operating systems, R&D, and advertising). DV, dependent variable; FE, fixed effects.

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

As with the results that do not include measures of the ecosystem of complements (see Table 6), it is shown that larger firms obtain less benefit from using nonpecuniary OSS, although the results are considerably stronger when the IT-producer variable is used to proxy for the ecosystem of complements. As in Table 6, Fortune 1000 status, sales, or employees at or above the 75th percentile of the sample are used as indicators of firm size. Tables 8(a) and 8(b) show the results of this analysis. When using IT intensity as the proxy for the ecosystem of complements (see Table 8(a)), the coefficient on the triple interaction is negative, although it is only significant when using Fortune 1000 as an indicator of firm size and the Arellano–Bond specification (model 2). However, when using IT-producing industry as a proxy for the ecosystem of complements (see Table 8(b)), the coefficient on the triple interaction is negative and significant for all indicators of firm size and in both the OLS and Arellano–Bond specifications. In aggregate, this adds evidence to support the argument that even if they have an existing ecosystem of complements, larger firms do not benefit as much from the use of nonpecuniary OSS as do smaller firms.

Because of space constraints, additional robustness checks are performed only using the IT intensity proxy for the ecosystem of complementary capabilities. The results when using the IT producer proxy are consistent with the IT intensity proxy and are available in Table A3 in the online appendix. Table 9 shows the results of a variety of additional specifications designed to show the robustness of the primary results and help argue for a more causal interpretation of the coefficient on nonpecuniary OSS for firms that have the necessary ecosystem of complements. For all specifications, the results are consistent with the baseline estimations in Table 7. Models 1 and 2 in Table 9 repeat the analyses from models 1 and 3 of Table 7 but use an industry control at the more granular NAICS-5 level rather than at NAICS-2. For both specifications, the results are consistent with those from Table 7, and the confidence intervals of the key coefficients overlap. Model 3 shows the results of using the Levinsohn–Petrin GMM estimator (Levinsohn and Petrin 2003), which utilizes a two-stage process using the cost of materials as an exogenous productivity shock to the firm's production. Materials are calculated as sales minus operating expenses minus total labor costs. The result is then deflated by the BLS Producer Price Index by stage of processing for intermediate materials, supplies, and components at the two- or three-digit NAICS

Table 8(b). Nonpecuniary OSS with IT Producer Interactions with Firm Size

DV: *Value-Added* (VA_{it}) Model type:	1 OLS	2 ABOND	3 OLS	4 ABOND	5 OLS	6 ABOND
IT Capital (IT_{it})	0.068***	0.017***	0.071***	0.017***	0.070***	0.017***
	(0.007)	(0.005)	(0.007)	(0.005)	(0.007)	(0.005)
Non-IT Capital (K_{it})	0.251***	−0.151	0.250***	−0.152	0.252***	−0.151
	(0.014)	(0.104)	(0.014)	(0.103)	(0.014)	(0.104)
Non-IT Labor (L_{it})	0.721***	0.696***	0.719***	0.696***	0.719***	0.697***
	(0.017)	(0.056)	(0.017)	(0.056)	(0.017)	(0.056)
$nonpecuniary\_OSS_{it}$	−0.004***	−0.001	−0.004***	−0.001	−0.004***	−0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$nonpecuniary\_OSS_{it} \times IT\text{-}producer$	0.016***	0.006**	0.016***	0.005**	0.012***	0.005**
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)
$nonpecuniary\_OSS_{it} \times IT\text{-}producer \times Fortune\ 1000$	−0.026***	−0.006*				
	(0.005)	(0.003)				
$nonpecuniary\_OSS_{it} \times IT\text{-}producer \times High\ Sales$			−0.026***	−0.005**		
			(0.005)	(0.002)		
$nonpecuniary\_OSS_{it} \times IT\text{-}producer \times High\ Employee$					−0.020***	−0.006**
					(0.007)	(0.003)
Constant	0.213***	0.065	0.219***	0.085	0.199***	0.076
	(0.060)	(0.715)	(0.060)	(0.712)	(0.060)	(0.713)
Year FE?	Y	Y	Y	Y	Y	Y
Industry FE	NAICS-2	NAICS-2	NAICS-2	NAICS-2	NAICS-2	NAICS-2
Number of firm-year obs.	10,355	7,650	10,355	7,650	10,355	7,650
Number of firms (groups)	1,566	1,424	1,566	1,424	1,566	1,424

Notes. Standard errors are clustered at the firm level for models 1, 3, and 5 and are heteroskedastic robust for all other models. All regressions include controls for R&D stock, advertising stock, advanced technology intensity (percentage of sites in the firm with DBMS, DWS, or ERP) (not shown). All models include an interaction of IT producer with IT capital (not shown). All variables are the natural log of the underlying variable (plus a small number in the case of operating systems, R&D, and advertising). DV, dependent variable; FE, fixed effects.
$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

level. This methodology is consistent with that used in Tambe and Hitt (2012). Again, the coefficient is consistent with the coefficient in the primary specification (see Table 7, model 1). Model 4 shows the results of using inverse-probability weighting (IPW) in an OLS framework (Horvitz and Thompson 1952; Wooldridge 2002, 2007). The IPW procedure helps to control for endogenous selection by weighting firms based on their likelihood of adopting nonpecuniary OSS, improving the balance in the sample such that the coefficient may be interpreted in a more causal manner (Hirano et al. 2003, Huber 2013). This procedure and its application are detailed in Online Appendix B. The results of this estimation are in line with the other estimates, although the coefficient is only significant at the 10% level (however, it is significant at the 5% level when IT producer status is used as a measure of the ecosystem of complements; see Table A3 in the online appendix).

Model 5 shows the results of using coarsened exact matching (CEM), a matching methodology that allows for causal inference without requiring balance checking (Iacus et al. 2012). CEM matches treated (OSS adopters) and control (nonadopters) observations based on observables, and it drops observations without a good match. The matching is done with the same variables from the IPW matching (detailed in Online Appendix B), but without the operating system variables since those are considered the treatment variable in this setting. This leads to a smaller sample size but allows for a more causal interpretation of the results.[33] Again, the coefficient is positive and significant, although larger than in the prior specifications. However, the confidence intervals again overlap with the original estimate.

Models 6 and 7 of Table 9 use an instrumental variable approach in a further attempt to interpret the coefficients as more causal than correlational. These specifications each utilize two instruments based on the adoption of firms that face supply conditions similar to the focal firm and are therefore likely to be affected by similar shocks to supply. The first instrument is constructed based on a continuous measure of the nonpecuniary OSS usage habits of firms in the same two-digit NAICS-level industry as the focal firm. The second instrument is constructed based on a continuous measure of the nonpecuniary OSS usage habits of firms that are in the same county as the focal firm but whose adoption decision is exogenous to the firm itself. Specifically, nonpecuniary OSS requires specialized labor to implement and operate. It is likely that this labor is dispersed both geographically and by

Table 9. Model Robustness Checks for Nonpecuniary OSS with IT Intensity Interaction

DV: *Value-Added* (VA_{it})	1	2	3	4	5	6	7
Model	ABOND–NAICS-5	BBOND–NAICS-5	LevPet GMM	IPW OLS	CEM OLS	2SLS (emp)	2SLS (rev)
IT Capital (IT_{it})	0.024***	0.024***	0.092***	0.047***	0.083***	0.258***	0.262***
	(0.005)	(0.005)	(0.007)	(0.013)	(0.013)	(0.040)	(0.040)
Non-IT Capital (K_{it})	−0.170*	−0.170*	0.484***	0.230***	0.220***	0.304***	0.305***
	(0.103)	(0.103)	(0.099)	(0.023)	(0.018)	(0.017)	(0.017)
Non-IT Labor (L_{it})	0.680***	0.680***	0.716***	0.747***	0.736***	0.745***	0.745***
	(0.057)	(0.057)	(0.023)	(0.026)	(0.022)	(0.017)	(0.017)
$nonpecuniary\_OSS_{it}$	−0.002***	−0.002***	−0.008***	−0.002*	−0.004***	−0.130***	−0.133***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.024)	(0.024)
$nonpecuniary\_OSS_{it} \times ITIntensity_{it}$	0.604***	0.604***	0.575***	0.276*	1.084*	0.751**	0.790**
	(0.148)	(0.148)	(0.140)	(0.156)	(0.633)	(0.330)	(0.361)
Constant	0.121	−0.636		0.122	0.313	−1.736***	−1.775***
	(0.680)	(1.081)		(0.204)	(0.192)	(0.469)	(0.477)
Year FE?	Y	Y	Y	Y	Y	Y	Y
Industry FE	NAICS-5	NAICS-5	NAICS-2	NAICS-2	NAICS-2	NAICS-2	NAICS-2
Number of firm-year obs.	7,650	8,988	10,355	10,355	5,917	10,355	10,355
Number of firms (groups)	1,424	1,501	1,566	1,566	1,405	1,566	1,566
R^2	—	—	—	0.931	0.906	0.664	0.653
First-stage F-statistic	—	—	—	—	—	43.55	43.82
Cragg and Donald min. eigenvalue	—	—	—	—	—	9.304	9.351
Overidentification test results (p-value)	—	—	—	—	—	0.652	0.427

Notes. Standard errors are heteroskedasticity robust for models 1, 2, 6, and 7, bootstrapped for model 3, and clustered at the firm level for models 4 and 5. All regressions include controls for R&D stock, advertising stock, advanced technology intensity (percentage of sites in the firm with DBMS, DWS, or ERP), and an interaction of IT intensity with IT capital. All variables are the natural log of the underlying variable (plus a small number in the case of operating systems, R&D, and advertising). DV, dependent variable; FE, fixed effects; LevPet, Levinsohn–Petrin; emp, employee-weighted instrument; rev, revenue-weighted instrument.

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

industry, and therefore the number of other firms in a county or industry using nonpecuniary OSS is likely to exogenously shift the likelihood of a firm to also use nonpecuniary OSS. Both of these instruments are similar to instruments that have been used for other studies of the digital economy (e.g., Forman et al. 2005). Importantly, most firms in the sample were founded before OSS diffused widely. Therefore, the firm's decision to locate in a specific geography or industry is independent of OSS adoption patterns. The instruments are constructed as a measure of the mean nonpecuniary OSS usage by other establishments within a given firm's county or industry within the same year. However, since most firms operate at multiple sites that are geographically dispersed (but in only one primary industry), the geographic instrument for one firm-year observation is crafted by combining the nonpecuniary OSS usage in all counties where the firm has a site. Since most sites at a firm are not of the same size (e.g., a headquarters is generally larger than a remote site), this combination must be weighted. Two weighting mechanisms are used—percentage of the company's total employees that are at the site (model 6) and percentage of the company's total revenues that come from the site (model 7). Both methods result in a unique number for any given firm-year observation that indicates an exogenous shift to the likelihood that the firm will use nonpecuniary OSS. Adoption by other firms within a county and industry is calculated using the full Harte Hanks data set based on an average of 419,137 establishments per year, not just the establishments and firms used in the sample set, adding further exogeneity to the instrument. For interaction terms of nonpecuniary OSS and IT intensity or IT producer, additional instruments are created based on the interaction of the base IV and IT intensity or IT producer. This yields four instruments (industry usage, county usage, and the interaction of these two with IT intensity or IT producer) for two endogenous variables (the base variable and the interaction). The instruments are used in a two-stage least squares (2SLS) framework to first predict the amount of nonpecuniary OSS used by the firm, and then this result is used in the main regression estimating the impact of this use on productivity. The resulting coefficients are again consistent with those from the primary analysis in Table 7. The first-stage F-statistics are reasonably large (43.55 for the employee-weighted instrument and 43.82 for the revenue-weighted instrument), and the Cragg and Donald (1993) minimum eigenvalues are of a reasonable size (9.304 for the employee-weighted instrument and 9.351 for the revenue-weighted instrument). Furthermore, tests for overidentification (necessary since there are more instruments than endogenous regressors) show that overidentification is not a concern

Table 10. Sample and Variable Robustness for Nonpecuniary OSS with IT Intensity Interaction

DV: *Value-Added* (VA_{it})	1	2	3	4	5	6
Model type:	ABOND	ABOND	ABOND	ABOND	ABOND	ABOND
Robustness check:	Adopters only	Other OS controls	Raw nos. for *ITL*, *ITK*, and *L*	County fixed effect	Including pecuniary OSS	Exclude IT-producing, financial, and natural res.
IT Capital (IT_{it})	0.025**	0.026***		0.025***	0.024***	0.023***
	(0.012)	(0.005)		(0.005)	(0.005)	(0.004)
Non-IT Capital (K_{it})	0.037	−0.176*	−0.184*	−0.181*	−0.177*	0.045
	(0.110)	(0.101)	(0.103)	(0.100)	(0.100)	(0.061)
Non-IT Labor (L_{it})	0.496***	0.688***		0.687***	0.688***	0.672***
	(0.091)	(0.054)		(0.054)	(0.054)	(0.054)
$nonpecuniary\_OSS_{it}$	−0.004**	−0.001	−0.001	−0.001	−0.001	−0.001**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)
$nonpecuniary\_OSS_{it} \times ITIntensity_{it}$	1.727**	0.409**	0.439**	0.247*	0.393**	0.354***
	(0.701)	(0.187)	(0.180)	(0.150)	(0.188)	(0.094)
$pecuniary\_OSS_{it}$		−0.000			−0.000	
		(0.001)			(0.001)	
$Closed_{it}$		−0.003				
		(0.003)				
$pecuniary\_OSS_{it} \times ITIntensity_{it}$					−0.046	
					(0.140)	
Raw # of PCs and Servers			0.004			
			(0.011)			
Raw # IT Employees			0.025***			
			(0.007)			
Raw # non-IT Employees			0.707***			
			(0.053)			
Constant	1.197	0.008	−2.381***	0.023	−0.034	0.345
	(1.396)	(0.680)	(0.662)	(0.673)	(0.837)	(0.472)
Number of firm-year obs.	2,211	7,650	7,650	7,650	7,650	5,807
Number of firms (groups)	708	1,424	1,424	1,424	1,424	1,050

Notes. Standard errors are heteroskedasticity robust for all models. All regressions include controls for year, NAICS two-digit industry, R&D stock, advertising stock, advanced technology intensity (percentage of sites in the firm with DBMS, DWS, or ERP), IT intensity, and an interaction of IT intensity with IT capital. All variables are the natural log of the underlying variable (plus a small number in the case of operating systems, R&D, and advertising), except where noted. DV, dependent variable.
$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$.

(since the p-values are both much larger than 0.05). Therefore, the instrumental variable results add further weight for a causal interpretation and address some of the measurement error concerns related to the data set discussed above. The results in Table A3 in the online appendix examine the same robustness checks as those from Table 9 but use whether or not a firm is an IT producer as the proxy for the ecosystem of complements. The results in Table A3 show a similar consistency with the IT producer results from Table 7.

Table 10 shows the results of a variety of additional robustness checks related to sample and variable assumptions using the IT intensity proxy for the ecosystem of complementary capabilities. The results when using the IT producer proxy are consistent with the IT intensity proxy and are available in Table A4 in the online appendix. The results are consistent with the baseline estimations in Table 7. All models use only the Arellano–Bond specification because of space constraints. Model 1 limits the sample to adopters only. The coefficient on the interaction between nonpecuniary OSS and IT intensity remains positive and significant. Model 2 includes measures of the number of other types of operating systems (pecuniary OSS and closed) at the firm as additional controls, which does not change the primary coefficient of interest.

Since assumptions were made about the costs of labor and computers, model 3 uses the raw number of computers, IT employees, and non-IT employees to show these assumptions are not driving the results. Again, the primary coefficient of interest stays within the confidence interval of the coefficient from the original analysis in Table 7. There may 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 et al. 1993, Furman et al. 2002). Therefore, model 4 includes a county fixed effect, based on the firm's headquarters, in the analysis. The point estimate of the coefficient for the interaction term drops slightly but stays positive and significant at the 10% level. The analysis to this point has focused exclusively on nonpecuniary OSS, but it is possible that pecuniary OSS may have an impact as well, particularly for IT producers. Model 5 examines this possibility and shows coefficients on both the baseline use of pecuniary OSS and the interaction of pecuniary OSS with IT intensity that are not significant at the 10% level. This indicates that any benefit from using pecuniary OSS is not affected by the level of IT intensity at the firm, implying that the ecosystem of complements is not necessary for pecuniary OSS, which is consistent with the discussion in Section 3. Finally, model 6 shows the results of the primary specification when excluding financial, natural resources, and IT-producing firms. Such an exclusion is motivated by the differences in accounting rules and differences in IT input prices that affect these industries. Saunders and Brynjolfsson (2016) discuss these differences in depth. This exclusion is of particular importance in this study as classification as an IT-producing firm is used as an alternative means of determining whether a firm has the ecosystem of complements necessary to benefit from the use of nonpecuniary OSS. The results in model 6 are consistent with those in the primary analysis, which suggests that rather than being driven by an accounting or input price difference, the IT producer–related results are indeed driven by this ecosystem of complements.

6.4. Firm Management Quality

An endogeneity concern that has not yet been fully addressed is that the use of nonpecuniary OSS may be correlated with unobservable managerial practices that are likely to increase productivity. Although the primary data set does not allow completely ruling out such simultaneity bias, additional data from the World Management Survey (Bloom and Van Reenen 2007) are used to confirm that this is not driving the results.[34] The World Management Survey (WMS) asks a wide array of firms about their management practices every few years, starting in 2004. Of the 1,566 firms from the main data set for this paper, 183 appear at least once in the WMS data set. Although this is far from a complete overlap, it does represent nearly 10% of the firms in the data set. There are 217 firm-year observations that overlap between the two data sets. The firms that appear in both data sets are used to test the correlation between management practices and the use of OSS (both pecuniary and nonpecuniary). The results indicate that an increase in the quality of a firm's management practices is uncorrelated with the decision to use nonpecuniary or pecuniary OSS.[35] The results are consistent when examining the binary or continuous use of OSS and when controlling for the production inputs of the firm ($ITK_{it}, ITL_{it}, K_{it}$, and 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 at the 10% level. 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 as a result of management quality can be at least partially alleviated.

7. Conclusion

The results of this study show that the use of nonpecuniary OSS has a positive and measureable impact on value-added productivity for firms that already have an ecosystem of complementary capabilities as proxied by IT intensity or belonging to an IT-producing industry. The effect is consistently positive and well identified in a variety of specifications that account for various endogeneity and measurement concerns. Because the use of nonpecuniary OSS is only measured via operating systems, other firm investments in nonpecuniary OSS are not captured. Therefore, the true effect of all nonpecuniary OSS is possibly greater than the effect found in this study, but identifying that effect is left for future research.

Digging further into the main effect by exploring various split-sample analyses reveals that larger firms (based on sales, employees, and Fortune 1000 status) gain a smaller benefit from increased usage of nonpecuniary OSS. This result holds both with and without the consideration of the ecosystem of complements. However, because of the sample construction, even the smallest firms are still rather large. It is quite possible, even likely, that the use of nonpecuniary OSS has an even larger effect for firms that are very small and therefore capital constrained. However, because of data constraints, the effect of nonpecuniary OSS on small companies in general, and technology related start-ups in particular, is left for future research.

An additional limitation relates to understanding the quality of IT labor at the firm. As discussed above, a specific type of technical capability is needed to implement nonpecuniary OSS. Although there is currently no evidence to show that such labor is higher quality than other IT labor, it is possible that this may be the case. Therefore, the coefficient on nonpecuniary OSS may be partially capturing some benefit the firm is receiving from higher-quality labor. To properly disentangle these effects, detailed data on IT labor skills would be necessary. Therefore, it is not possible to test for this effect in the current setting. However, it is safe to say that the results for nonpecuniary OSS usage encompass the full benefits of the ecosystem of complements to OSS (including labor).

Although endogeneity is always a concern in productivity studies, this study takes steps to help rule out as many biases as possible to allow the results to be interpreted in a more causal manner. All of the regression results use year and industry fixed effects.

This helps to rule out alternative explanations due to time or industry trends. In the primary specifications, dynamic panel analysis (Arellano–Bond or Blundell–Bond) is used to rule out endogenous selection as a driving factor of the results. Additional specifications using inverse probability weighting, coarsened exact matching, and instrumental variables allow 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 more causal interpretation of the findings.

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 nonpecuniary OSS (and, more generally, nonpecuniary crowdsourced digital goods) as inputs into production. Furthermore, 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 nonpecuniary inputs, rather than misattributing them to firm intangible effects. This is especially important as information costs are increasingly approaching zero and the amount of nonpecuniary crowdsourced digital inputs firms use is likely to rise in the coming years. Indeed, it is quite possible that the results found in this study apply to a much wider range of goods, such as *Wikipedia*, free blueprints for 3D printed objects, open source hardware, crowdsourced innovation tournaments, and the digitization of consumers' opinions via online review sites and social media. 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 et al. 2012, Corrado and Hulten 2013, Altman et al. 2014), and they may have a similar unmeasured productivity impact to that of nonpecuniary OSS.

For practitioners, the results indicate that firms that already have the necessary ecosystem of complements may enhance their value-added productivity by increasing the amount of nonpecuniary OSS they employ in their production process. As more firms become more IT intensive, it is likely that a larger number of firms will have this ecosystem and can therefore benefit from the use of nonpecuniary OSS. Furthermore, smaller firms may benefit more than larger firms. 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 this paper, not all firms benefit to the same degree.

Acknowledgments

The author is grateful for helpful comments from Shane Greenstein, Carliss Baldwin, Yochai Benkler, Raj Choudhury, Anil Doshi, Lorin Hitt, Marco Iansiti, Ohchan Kwon, Karim Lakhani, Kristina McElheran, Hart Posen, Scott Stern, Neil Thompson, Mike Toffel, Joel West, Feng Zhu, two editors, and three anonymous referees. Excellent research assistance was provided by Jessica Bailey. Additional helpful comments were received from participants at Atlanta Competitive Advantage Conference 2014, American Economic Association 2015 Annual Meeting, Academy of Management (AOM) 2014 Annual Meeting, 2014 AOM BPS Dissertation Consortium, Consortium on Competitiveness and Cooperation 2014, Charles River Conference 2014, DRUID 2014, Harvard Business School Technology and Operations Management (HBS TOM) DBA Seminar 2014, HBS TOM Alumni Conference 2014, NBER Summer Institute 2016, NYU Engelberg Center Conference on Knowledge Commons 2014, 2014 Open and User Innovation Conference, Strategic Management Society Conference 2014, and ZEW Information and Communication Technologies 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, the University of California at Los Angeles, the University College London, the University of Maryland, the University of Pennsylvania, and the University of Southern California. The bulk of the work for this paper was done while the author was a doctoral student at Harvard Business School, where he was supported by the HBS Division of Research and Faculty Development. All mistakes remain the author's own.

Endnotes

[1] See Syverson (2011) for an overview of this literature.

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

[3] 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, only 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.

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

[5] The focus of this research is primarily on nonpecuniary OSS. The availability of pecuniary products, such as Red Hat Linux, which build on nonpecuniary OSS is important, but the risks associated with these products are lower as a result of the contractual relationship a customer has with the vendor, which greatly mitigates these risks.

[6] Linus's Law is attributed to Eric Raymond (1999, p. 30) but named after the founder of Linux, Linus Torvalds. It states, "Given enough eyeballs, all bugs are shallow," implying more people looking at the code is beneficial.

[7] The Heartbleed security bug was introduced into the OpenSSL cryptography library in December 2011 and was not noticed and fixed until April 2014 (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 and iOS operating systems.

[8] This phrase actually originated about IBM in the 1970s, long before OSS became prominent. However, it was ported to Microsoft in the 1990s as OSS started to gain traction in the marketplace.

[9] This statement is from a speech by Bill Joy, cofounder of Sun Microsystems, in 1990 (Gilder 1995).

[10] Hann et al. (2002, 2013) show that average participants in OSS do not receive higher wages at their jobs. However, they find those performing managerial tasks in the OSS community receive 18% higher wages.

[11] An important aspect of the OSS movement is the ability to build pecuniary software on top of nonpecuniary OSS. For example, Red Hat Enterprise Linux is built on the open source Linux kernel but is not free as a result of the additional functionality and support Red Hat provides. Conversely, a product such as Mandrake Linux is both open source and nonpecuniary. Therefore, pecuniary OSS is considered different from nonpecuniary OSS.

[12] Arellano–Bond is implemented via xtabond in Stata. Blundell–Bond is implemented using xtdpdsys in Stata.

[13] 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 time-series analysis and does not adversely affect the pooled analysis.

[14] Available at http://www.bea.gov/iTable/index_industry_gdpIndy.cfm (accessed September 25, 2017).

[15] Available at http://www.bea.gov/industry/xls/io-annual/GDPbyInd_VA_NAICS_1997-2014.xlsx (accessed September 25, 2017). The GOPI series (chain-type price indexes for gross output (2009 = 100)) is used for this deflator.

[16] The one notable exception to this is the study by Tambe and Hitt (2012), which focuses primarily on IT labor and therefore separates it from IT capital. Furthermore, software expenses 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 by the non-IT capital variable. Internal software development will be captured by the IT labor variable. This methodology is consistent with prior literature (e.g., Brynjolfsson and Hitt 1996, Huang et al. 2016).

[17] For most firms, Harte Hanks only surveys a sample of the sites within the firm. Therefore, the total number of PCs and servers must be imputed. In such cases, the average number of PCs/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/servers in the firm. The same procedure is used to calculate the number of IT employees and the number of each operating system at the firm. This leads to an imputed number of PCs/servers for the entire firm.

[18] Available at http://research.stlouisfed.org/fred2/data/B935RG3A086NBEA.txt (accessed September 25, 2017).

[19] 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.

[20] Obtained from the Bureau of Labor and Statistics: http://www.bls.gov/oes/2009/may/oes_nat.htm#15-0000 (accessed September 25, 2017).

[21] For example, the 2000 data can be found at http://www.bls.gov/news.release/history/ecec_06292000.txt (accessed September 25, 2017), and the 2004 data can be found at http://www.bls.gov/news.release/history/ecec_06242004.txt (accessed September 25, 2017).

[22] Deflator information for 2000–2005 is from table 3 of http://www.bls.gov/news.release/history/eci_01252001.txt (accessed September 25, 2017). Information for 2006–2009 is obtained from table 5 of http://www.bls.gov/news.release/history/eci_01312007.txt (accessed September 25, 2017).

[23] I thank a very helpful anonymous reviewer for detailed insights on implementing this methodology.

[24] Obtained from the Bureau of Labor and Statistics; for example, the data for 2009 can be found at http://www.bls.gov/oes/2009/may/oes_nat.htm#00-0000 (accessed September 25, 2017).

[25] Available at http://www.bea.gov/national/rd.htm (accessed September 25, 2017).

[26] Since the natural log of both R&D and brand stocks will be used in the regression, a small number (0.0000001) is added to each value to avoid the fact that the natural log of zero is undefined.

[27] Although some nonpecuniary OSS operating systems, such as Debian, are offered at a nominal pecuniary price by third-party vendors for the convenience of the distribution being preloaded on a CD or DVD, they are included in the nonpecuniary column because they are 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, similar to Red Hat Linux. Robustness checks were run against this assumption with no change to the primary results.

[28] I am grateful to an anonymous reviewer for pointing out the need to not take the naïve approach to estimating the number of each type of operating system at the firm, which would underestimate the amount of OSS.

[29] 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), a small number (0.0000001) 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 nonpecuniary 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 small number before taking the natural log should not significantly bias the results.

[30] When using methods that use a firm fixed effect for within firm identification, (models 5–7), the coefficient on non-IT capital becomes indistinguishable from zero. This is a result of the recursive calculation of non-IT capital stock discussed above, which limits the variation within firm from year to year. Therefore, the first differences do not provided sufficient variation from which to accurately measure the value of the coefficient. This can occur with the IT capital measure as well. Such results occur in other studies of IT productivity where the firm fixed effect biases the coefficient on capital toward zero (e.g., Huang et al. 2016, Tambe and Hitt 2012), often to the point where it is not distinguishable from zero (e.g., Acharya 2015). This phenomenon as it pertains to R&D has been discussed in the literature by Saunders and Brynjolfsson (2016) and Hall et al. (2005).

[31] Additional cutoff values are tested and discussed below.

[32] Although these additional control and interaction variables are not shown in Table 7, the complete regression table showing all coefficients for all variables and interactions can be found in Table A2 in the online appendix. This table shows additional specifications that include interactions of IT intensity with all variables including IT capital, non-IT capital, non-IT labor, R&D stock, advertising stock, and advanced technology intensity.

[33] While no procedure will guarantee a causal linkage in observational data, CEM has been shown to be an effective step in the causal direction, despite being based only on observables (Iacus et al. 2012).

[34] I am grateful to Nick Bloom, Raffaella Sadun, and John Van Reenen for access to the WMS data set.

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

References

Acharya RC (2015) ICT use and total factor productivity growth: Intangible capital or productive externalities? *Oxford Econom. Papers* 68(1):16–39.

Altman E, Nagle F, Tushman M (2014) Innovating without information constraints: Organizations, communities, and innovation when information costs approach zero. Hitt MA, Shalley C, Zhou J, eds. *Oxford Handbook of Creativity, Innovation, and Entrepreneurship* (Oxford University Press, New York), 353–378.

Andreessen M (2011) Why software is eating the world. *Wall Street Journal* (August 20), https://www.wsj.com/articles/SB10001424053111903480904576512250915629460.

Angrist JD, Pischke JS (2009) *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton University Press, Princeton, NJ).

Aral S, Weill P (2007) IT assets, organizational capabilities, and firm performance: How resource allocations and organizational explain performance variation. *Organ. Sci.* 18(5):763–780.

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

Arellano M, Bond S (1991) Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Rev. Econom. Stud.* 58(2):277–297.

Asay M (2013) Is Facebook the world's largest open source company? *ReadWrite* (October 17), http://readwrite.com/2013/10/17/is-facebook-the-worlds-largest-open-source-company.

Athey S, Ellison G (2014) Dynamics of open source movements. *J. Econom. Management Strategy* 23(2):294–316.

Baldwin CY, Clark KB (2006) The architecture of participation: Does code architecture mitigate free riding in the open source development model? *Management Sci.* 52(7):1116–1127.

Baldwin C, von Hippel E (2011) Modeling a paradigm shift: From producer innovation to user and open collaborative innovation. *Organ. Sci.* 22(6):1399–1417.

Bardhan I, Krishnan V, Lin S (2013) Business value of information technology: Testing the interaction effect of IT and R&D on Tobin's q. *Inform. Systems Res.* 24(4):1147–1161.

Benkler Y (2002) Coase's penguin, or, Linux and the nature of the firm. *Yale Law J.* 112(3):369–446.

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

Bloom N, Van Reenen J (2007) Measuring and explaining management practices across firms and countries. *Quart. J. Econom.* 122(4):1351–1408.

Bloom N, Sadun R, Van Reenen J (2012) Americans do IT better: US multinationals and the productivity miracle. *Amer. Econom. Rev.* 102(1):167–201.

Blundell R, Bond S (1998) Initial conditions and moment restrictions in dynamic panel data models. *J. Econometrics* 87(1):115–143.

Bresnahan TF, Brynjolfsson E, Hitt LM (2002) Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *Quart. J. Econom.* 117(1):339–376.

Bridgman B (2013) Home productivity. Bureau of Economic Analysis Working Paper 2013-03, U.S. Bureau of Economic Analysis, Washington, DC.

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

Brynjolfsson E, Hitt LM (2003) Computing productivity: Firm-level evidence. *Rev. Econom. Statist.* 85(4):793–808.

Brynjolfsson E, Milgrom P (2012) Complementarity in organizations. Gibbons R, Roberts J, eds. *The Handbook of Organizational Economics* (Princeton University Press, Princeton NJ), 11–55.

Brynjolfsson E, Saunders A (2009) What the GDP gets wrong (why managers should care). *Sloan Management Rev.* 51(1):95–96.

Brynjolfsson E, Hitt LM, Yang S (2002) Intangible assets: Computers and organizational capital. *Brookings Papers Econom. Activity* 2002(1):137–198.

Byrne D, Oliner S, Sichel D (2013) Is the information technology revolution over? *Internat. Productivity Monitor* 25(Spring), 20–36.

Casadesus-Masanell R, Llanes G (2011) Mixed source. *Management Sci.* 57(7):1212–1230.

Chang YB, Gurbaxani V (2012) Information technology outsourcing, knowledge transfer, and firm productivity: An empirical analysis. *MIS Quart.* 36(4):1043–1053.

Corrado C, Hulten R (2013) Innovation accounting. Jorgenson DW, Landefeld JS, Schreyer P, eds. *Measuring Economic Sustainability and Progress*, Chap. 18 (University of Chicago Press, Chicago).

Corrado C, Hulten C, Sichel D (2009) Intangible capital and US economic growth. *Rev. Income Wealth* 55(3):661–685.

Cragg JG, Donald SG (1993) Testing identifiability and specification in instrumental variable models. *Econometric Theory* 9(2): 222–240.

Dewan S, Min CK (1997) The substitution of information technology for other factors of production: A firm level analysis. *Management Sci.* 43(12):1660–1675.

Finley K (2013) Apple's operating system guru goes back to his roots. *Wired* (August 8), www.wired.com/2013/08/jordan-hubbard/.

Fitzgerald B (2006) The transformation of open source software. *MIS Quart.* 30(3):587–598.

Forman C (2005) The corporate digital divide: Determinants of Internet adoption. *Management Sci.* 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. *J. Urban Econom.* 58(3):389–420.

Forman C, Goldfarb A, Greenstein S (2008) Understanding the inputs into innovation: Do cities substitute for internal firm resources? *J. Econom. Management Strategy* 17(2):295–316.

Fosfuri A, Giarratana MS, Luzzi A (2008) The penguin has entered the building: The commercialization of open source software products. *Organ. Sci.* 19(2):292–305.

Furman JL, Porter ME, Stern S (2002) The determinants of national innovative capacity. *Res. Policy* 31(6):899–933.

Gartner (2010) Forecast: PC market by operating system, worldwide, 2001–2009. Report, Gartner, Stamford, CT.

Giera J, Brown A (2004) The costs and risks of open source—Debunking the myths. White paper, Forrester Research, Cambridge, MA.

Gilder G (1995) The coming software shift. *Forbes* (August 28), 154.

Graham R (2014) 300k vulnerable to Heartbleed two months later. *Errata Security* (blog) (June 21), http://blog.erratasec.com/2014/06/300k-vulnerable-to-heartbleed-two.html.

Greenstein S, Nagle F (2014) Digital dark matter and the economic contribution of Apache. *Res. Policy* 43(4):623–631.

Hall BH, Jaffe A, Trajtenberg M (2005) Market value and patent citations. *RAND J. Econom.* 36(1):16–38.

Hamilton D (2014) Mirantis gains $100M in the largest series-B investment round in open-source software history. *Whir* (October 21), http://www.thewhir.com/web-hosting-news/mirantis-gains-100m-largest-series-b-investment-round-open-source-software-history.

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

Han K, Kauffman RJ, Nault BR (2011) Returns to information technology outsourcing. *Inform. Systems Res.* 22(4):824–840.

Hann I, Roberts J, Slaughter S (2013) All are not equal: An examination of the economic returns to different forms of participation in open source software communities. *Inform. Systems Res.* 24(3):520–538.

Hann I, Roberts J, Slaughter S, Fielding R (2002) Economic incentives for open source projects: Can participation be explained by career concerns? *Proc. 22nd Internat. Conf. Inform. Systems (ICIS), Barcelona, Spain*, 365–372.

Harhoff D, Henkel J, von Hippel E (2003) Profiting from voluntary information spillovers: How users benefit by freely revealing their innovations. *Res. Policy* 32(10):1753–1769.

Hayek FA (1945) The use of knowledge in society. *Amer. Econom. Rev.* 35(4):519–530.

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

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

Hitt LM, Brynjolfsson E (1996) Productivity, business profitability, and consumer surplus: Three different measures of information technology value. *MIS Quart.* 20(2):121–142.

Horvitz DG, Thompson DJ (1952) A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* 47(260):663–685.

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 DJ (2016) IT knowledge spillovers, absorptive capacity, and productivity: Evidence from enterprise software. Working paper, University of Maryland, College Park.

Huber M (2013) Identifying causal mechanisms (primarily) based on inverse probability weighting. *J. Appl. Econometrics* 29(6): 920–943.

Hulten C (2010) Decoding Microsoft: Intangible capital as a source of company growth. NBER Working Paper 15799, National Bureau of Economic Research, Cambridge, MA.

Iacus SM, King G, Porro G (2012) Causal inference without balance checking: Coarsened exact matching. *Political Anal.* 20(1):1–24.

Imbens GW, Kolesar M (2016) Robust standard errors in small samples: Some practical advice. *Rev. Econom. Statist.* 98(4):701–712.

Jaffe AB, Trajtenberg M, Henderson R (1993) Geographic localization of knowledge spillovers as evidenced by patent citations. *Quart. J. Econom.* 108(3):577–598.

Jorgenson DW (2001) Information technology and the US economy. *Amer. Econom. Rev.* 91(1):1–32.

Jorgenson DW, Ho MS, Stiroh KJ (2005) *Productivity, Information Technology and the American Growth Resurgence*, Vol. 3 (MIT Press, Cambridge, MA).

Kogut B, Metiu A (2001) Open-source software development and distributed innovation. *Oxford Rev. Econom. Policy* 17(2):248–264.

Krishnamurthy S (2005) An analysis of open source business models. Feller J, Fitzgerald B, Hissam S, Lakhani K, eds. *Perspectives on Free and Open Source Software* (MIT Press, Cambridge, MA), 279–296.

Lakhani K, von Hippel E (2003) How open source software works: "Free" user-to-user assistance. *Res. Policy* 32(6):923–943.

Lakhani K, Lifshitz-Assaf H, Tushman M (2012) Open innovation and organizational boundaries. Grandori A, ed. *Handbook of Economic Organization: Integrating Economic and Organization Theory* (Edward Elgar Publishing, Northampton, MA), 355–382.

Lerner J, Schankerman M (2010) *The Comingled Code: Open Source and Economic Development* (MIT Press, Cambridge, MA).

Lerner J, Tirole J (2002) Some simple economics of open source. *J. Indust. Econom.* 50(2):197–234.

Lerner J, Pathak PA, Tirole J (2006) The dynamics of open-source contributors. *Amer. Econom. Rev.* 96(2):114–118.

Levinsohn J, Petrin A (2003) Estimating production functions using inputs to control for unobservables. *Rev. Econom. Stud.* 70(2): 317–341.

Li WCY, Hall BH (2016) Depreciation of business R&D capital. NBER Working Paper 22473, National Bureau of Economic Research, Cambridge, MA.

Lindenberg EB, Ross SA (1981) Tobin's *q* and industrial organization. *J. Bus.* 54(1):1–32.

MacCormack A (2003) Evaluating total cost of ownership for software platforms: Comparing apples, oranges, and cucumbers. AEI-Brookings Center for Regulatory Studies Related Publication, AEI, Washington, DC.

MacCormack A, Rusnak J, Baldwin CY (2006) Exploring the structure of complex software designs: An empirical study of open source code. *Management Sci.* 52(7):1015–1030.

McCue TJ (2013) Ford Motor Company sees open source. *Forbes* (January 10), http://www.forbes.com/sites/tjmccue/2013/01/10/ford-motor-company-sees-open-source/.

McElheran KS (2014) Delegation in multi-establishment firms: Adaptation vs. coordination in I.T. purchasing authority. *J. Econom. Management Strategy* 23(2):225–258.

Newbart D (2001) Microsoft CEO takes launch break with the *Sun-Times*. *Chicago Sun-Times* (June 1), https://web.archive.org/web/20011115003306/http://www.suntimes.com/output/tech/cst-fin-micro01.html.

O'Mahony S (2003) Guarding the commons: How community managed software projects protect their work. *Res. Policy* 32(7): 1179–1198.

O'Mahony S, Ferraro F (2007) The emergence of governance in an open source community. *Acad. Management J.* 50(5):1079–1106.

Ostrom E (1990) *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge University Press, Cambridge, MA).

Pang MS, Lee G, DeLone WH (2014) IT resources, organizational capabilities, and value creation in public-sector organizations. *J. Inform. Tech.* 29(3):187–205.

Phipps S (2014) Walmart's investment in open source isn't cheap. *InfoWorld* (August 22), http://www.infoworld.com/article/2608897/open-source-software/walmart-s-investment-in-open-source-isn-t-cheap.html.

Ransbotham S (2010) An empirical analysis of exploitation attempts based on vulnerabilities in open source software. *Proc. 9th Workshop Econom. Inform. Security, Cambridge, MA*.

Raymond E (1999) The cathedral and the bazaar. *Knowledge, Tech. Policy* 12(3):23–49.

Russo B, Braghin B, Gasperi P, Sillitti A, Succi G (2005) Defining TCO for the transition to open source systems. *Proc. First Internat. Conf. Open Source Systems*, 108–112.

Saunders A, Brynjolfsson E (2016) Valuing IT-related intangible assets. *MIS Quart.* 40(1):83–110.

Schumpeter JA (1942) The process of creative destruction. *Capitalism, Socialism and Democracy* (Harper & Brothers, New York), 81–86.

Schwarz M, Takhteyev Y (2011) Half a century of public software institutions. *J. Public Econom. Theory* 12(4):609–639.

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

Sorkin A, Peters J (2006) Google to acquire YouTube for $1.65 billion. *New York Times* (3, Part 1), http://www.nytimes.com/2006/10/09/business/09cnd-deal.html.

Stiroh KJ (2002) Information technology and the US productivity revival: What do the industry data say? *Amer. Econom. Rev.* 92(5):1559–1576.

Syverson C (2011) What determines productivity? *J. Econom. Literature* 49(2):326–365.

Tambe P, Hitt LM (2012) The productivity of information technology investments: New evidence from IT labor data. *Inform. Systems Res.* 23(3):599–617.

Tambe P, Hitt L, Brynjolfsson E (2011) The price and quantity of IT-related intangible capital. *ICIS 2011 Proceed.* 16. https://aisel.aisnet.org/icis2011/proceedings/economicvalueIS/16.

Tambe P, Hitt LM, Brynjolfsson E (2012) The extroverted firm: How external information practices affect innovation and productivity. *Management Sci.* 58(5):843–859.

Varian HR, Shapiro C (2003) Linux adoption in the public sector: An economic analysis. Working paper, University of California, Berkeley, Berkeley.

Vaughan-Nichols SJ (2016) Why Microsoft is turning into an open-source company. *ZDNet* (June 9), http://www.zdnet.com/article/why-microsoft-is-turning-into-an-open-source-company/

Villalonga B (2004) Intangible resources, Tobin's *q*, and sustainability of performance differences. *J. Econom. Behav. Organ.* 54(4): 205–230.

von Hippel E (1986) Lead users: A source of novel product concepts. *Management Sci.* 32(7):791–805.

von Hippel E, von Krogh G (2003) Open source software and the "private-collective" innovation model: Issues for organization science. *Organ. Sci.* 14(2):209–223.

von Krogh G, Spaeth S, Lakhani KR (2003) Community, joining, and specialization in open source software innovation: A case study. *Res. Policy* 32(7):1217–1241.

Vrionis J (2015, December 6) It's actually open source software that's eating the world. *VentureBeat* (December 6), http://venturebeat.com/2015/12/06/its-actually-open-source-software-thats-eating-the-world/.

West J, Lakhani KR (2008) Getting clear about communities in open innovation. *Indust. Innovation* 15(2):223–231.

WhatsApp (2016) WhatsApp and Open Source. Accessed March 23, 2016, https://web.archive.org/web/20160304091651/whatsapp.com/opensource.

Wheeler D (2005) Why open source software/free software (OSS/FS, FLOSS, or FOSS)? Look at the numbers! Accessed September 25, 2017, 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, Sebastopol, CA).

Wooldridge JM (2002) Inverse probability weighted M-estimators for sample selection, attrition, and stratification. *Portuguese Econom. J.* 1(2):117–139.

Wooldridge J (2007) Inverse probability weighted estimation for general missing data problems. *J. Econometrics* 141(2):1281–1301.

Woolley AW, Fuchs E (2011) Collective intelligence in the organization of science. *Organ. Sci.* 22(5):1359–1367.

Yarow J (2013) Microsoft's biggest problem in one chart. *Bus. Insider* (September 13), http://www.businessinsider.com/microsofts-biggest-problem-in-one-chart-2013-9.