IT Knowledge Spillovers, Absorptive Capacity, and Productivity: Evidence from Enterprise Software

Peng Huang
R.H. Smith School of Business
University of Maryland
College Park, MD 20742
huang@umd.edu

Marco Ceccagnoli
Scheller College of Business
Georgia Institute of Technology
800 West Peachtree Street, NW
Atlanta, GA 30308
marco.ceccagnoli@scheller.gatech.edu

Chris Forman
Dyson School of Applied Economics and Management
Cornell University
Ithaca, NY 14853
chris.forman@cornell.edu

D. J. Wu
Scheller College of Business
Georgia Institute of Technology
800 West Peachtree Street, NW
Atlanta, GA 30308
dj.wu@scheller.gatech.edu

PRELIMINARY, PLEASE DO NOT CIRCULATE WITHOUT PERMISSION

Abstract

We examine the productivity implications of external knowledge flows obtained through an internet-enabled discussion forum in which IT professionals help each other resolve problems related to the implementation and use of enterprise software. Consistent with the absorptive capacity framework (Cohen and Levinthal, 1989, 1990), we show that IT spillovers are more likely to accrue to firms with prior related investments in advanced enterprise software, and less likely to accrue when the external knowledge is more difficult to learn, such as for business functional knowledge or knowledge originating from relatively newer and emerging discussion forums. We also show that when knowledge is difficult to learn, the ability to derive the value of IT spillovers is more dependent on prior related IT investments. Similar to R&D investments, the dual role of IT investments has important implications for estimating their "true" rate of return.

1. Introduction

While information technology (IT) systems have been shown to create significant value for the firms who adopt them, the returns often appear with a delay (Brynjolfsson and Hitt 2003) and the outcomes may vary widely (Aral and Weill 2007, Bloom et al. 2012, Bresnahan et al. 2002). Firms investing in new IT systems must often undertake complementary innovation, sometimes termed co-invention, to adapt general purpose IT systems to the idiosyncratic needs of organizations (Bresnahan and Greenstein 1996). While sometimes these innovations are related to technical adaptations to IT hardware and software systems, they also frequently involve changes to organizational elements such as business processes (Bartel et al. 2007, Bresnahan et al. 2002, Dranove et al. 2014).

The necessary human capital to deploy these systems is scarce and unequally distributed.[1] However, there are a range of formal and informal means that firms can use to acquire the necessary knowledge to build human capital.[2] The available mechanisms range from hiring workers who have acquired the necessary expertise by working on similar projects at other firms (Tambe and Hitt 2014a), exchange of knowledge with third party consultants who have been hired by the firm (e.g., Chang and Gurbaxani 2012b, Ko et al. 2005), to knowledge exchange that is mediated by communication with industry or supply chain participants (Caselli and Coleman 2001, Chang and Gurbaxani 2012a). However, despite the widespread use of these mechanisms, it remains the case that the benefits to the deployment of business IT systems vary greatly (Aral and Weill 2007), emerge only over time (Brynjolfsson and Hitt 2003, Dranove et al. 2014), and arrive more quickly for firms who already have access to related skills (e.g., Bresnahan and Greenstein 1996, Dranove et al. 2014). That is, firms are unable to perfectly contract for the required human capital to deploy IT systems.

While there is widespread agreement that firms face difficulty in acquiring the necessary human capital to deploy IT systems, because of data constraints we have limited understanding why this is the case. This is particularly true for human capital accumulation augmented by external knowledge flows, which typically leave no paper trail. In this paper we seek to address this gap in the literature by investigating the conditions under which firms realize benefits from efforts to accumulate human capital related to deploying IT systems. Our particular interest is in understanding how prior related investments related to IT systems facilitate the process of accumulating human capital through external knowledge

[1] For a review of the literature see, for example, Brynjolfsson and Milgrom (2012). For a recent example, see Tambe and Hitt (2014a).

[2] Investments in education, training, health and values that cannot easily be separated from people are regarded as human capital (Becker 2008). As will be discussed in further detail below, in our setting knowledge acquisition builds human capital through a variety of formal and informal means. We will therefore use the terms “acquire knowledge” and “accumulate human capital” interchangeably.

flows. We introduce and test a model where IT investment serves dual purposes: IT not only creates value for the investing firm directly through the use of the systems, but also indirectly contributes to productivity by deepening the firm's knowledge stock that helps the firm to absorb external knowledge flows through its absorptive capacity.[3]

We follow recent work that has used records of digital interactions as proxies for aggregate human capital investments within the firm (e.g., Tambe and Hitt 2014a). Much of the prior work in this area has aggregated the human capital of individual workers as reported in resume databases or through social media platforms like LinkedIn (Tambe 2014, Tambe and Hitt 2014a). However, the deployment of IT systems requires the development of human capital that is at least partially firm-specific, combining both internal and external knowledge of how to implement specific technologies within the context of specific situations. To address this concern, we develop a new measure where we observe human capital accumulation related to a firm's IT systems by directly studying the knowledge acquisition activity of IT workers on an online discussion board. Our setting enables us to measure characteristics about the knowledge acquired, the extent to which it is related to a firm's existing IT systems, and how the combination of these two shapes the ability of workers to successfully absorb new knowledge acquired from the discussion board.

Specifically, we study knowledge acquisition activities related to implementation of enterprise software systems from one of the largest vendors, SAP. We measure knowledge flows by observing activity on a question and answer forum sponsored by the SAP Community Network (SCN). Our approach is to estimate productivity models that are augmented by a factor of production that includes the firm's knowledge stock, where the stock of knowledge is computed based upon both prior investments in related domains, and knowledge spillovers as a result of the activities of firm employees on the forum. We estimate this model over a longitudinal sample of 275 firms among the Fortune 1000 that are SAP adopters.

We first estimate a benchmark model that assumes firms are homogenous in their ability to absorb and exploit external knowledge. We show that SAP user firms that experience an increase in knowledge inflows through their employees' participation in the SAP Community Network also experience an increase in productivity. For an average firm in our sample, one additional answered question translates into an additional $13.30 million in added value for the firm. Similar to the use of patent citations as a measure of knowledge flows in the R&D literature, we acknowledge that our measure of knowledge

[3] In this paper we defined absorptive capacity as a firm's ability to assimilate, transform, and apply external knowledge (Cohen and Levinthal 1989, Cohen and Levinthal 1990).

flows may capture other efforts aimed at augmenting human capital within the firm.[4] Thus, the size of the coefficient may partially reflect the returns to these other knowledge acquisition activities. We probe the robustness of our results to different assumptions regarding the measurement of knowledge flows and the omission of unobserved variables.

We next turn to an examination of the conditions under which knowledge flows are more likely to be absorbed by the firm. Using data from the forum and from an external data set related to IT investments, we measure key factors that may affect a firm's absorptive capacity. Building on prior work (Cohen and Levinthal 1989), we estimate a model that extends our production functional model with the determinants of absorptive capacity. We focus on two factors that may influence the absorption of external knowledge: the presence of prior related knowledge at the firm, and characteristics of the knowledge that may influence the ease with which it will be transferred and absorbed. Consistent with prior studies that have focused on absorption of knowledge related to research and development (e.g., Cohen and Levinthal 1989), we find that external knowledge is more readily absorbed when firms have made investments in prior related knowledge, that absorption is less likely when the knowledge is difficult to learn, and that the importance of prior related knowledge to knowledge absorption is greater when the transferred knowledge is more difficult to learn.

We contribute to recent work that has sought to understand the productivity benefits of IT investment and related business process innovation in several ways. First, as noted above, we advance recent work seeking to understand the productivity benefits of external knowledge acquisition or "IT spillovers" (e.g., Chang and Gurbaxani 2012a,b, Tambe and Hitt 2014a) by demonstrating the conditions under which external knowledge acquisition leads to higher productivity. For example, external knowledge that can be easily codified and is less context-dependent is more readily absorbed and transferred, leading to greater contribution to productivity.

Further, while prior information systems research has emphasized the importance of related IT investments to achieving benefits from IT-enabled business process innovation, our research advances the frontier in several ways.[5] In particular, our measurement strategy allows us to capture differences in the nature of knowledge that is being acquired. This allows us to test an implication of the absorptive capacity model that has previously been difficult to study in the context of IT research: namely, that the importance of prior related enterprise IT investments to absorbing new external knowledge related to business process innovation will be greatest in those environments where the knowledge is more context-

[4] For a discussion of the measurement error from using patent citations as proxy for knowledge flows in the R&D literature, see Roach and Cohen (2013).

[5] For a review of prior research on absorptive capacity in information systems research, see Roberts et al. (2012).

dependent and has not yet been written down in abstract, codified terms. Understanding these differences is important; focusing on aggregate knowledge flows can give rise to an aggregation bias that may create faulty policy prescriptions.

More broadly, our research approach provides a unique strategy to measure human capital acquisition that accompanies business process innovation. In the past, due to the inherent measurement difficulties, efforts to measure this type of human capital acquisition have mostly created proxies using survey measures (e.g., Aral and Weill 2007, Bresnahan et al. 2002, Sambamurthy et al. 2003), investments in related technologies (Greenstein and Nagle 2014, Nagle 2015), or, more recently, using data from resumes and social network profiles (e.g., Tambe and Hitt 2014a). One challenge faced by many of these approaches is that they are often both costly to implement and are limited in their ability to measure firm-specific human capital investments. This research approach provides insights for other researchers on how to use alternative, archival data to study questions in this research area.

2. Theory and Hypotheses

2.1 External Knowledge Input, Absorptive Capacity, and Productivity

The effective implementation of IT within organizations has emphasized the view of IT as an enabler of bsiness process innovation. Business process innovation requires both a range of investments in computing hardware and software, as well as changes to process flows, human capital, and other organizational practices. For example, it may involve changes in employee discretion, or the way that information flows within an organization (Bresnahan et al. 2002). Prior research has emphasized that business process innovation involves co-invention, the post-adoption invention of complementary business processes and adaptations that make investments in IT useful (Bresnahan and Greenstein 1996). In the context of enterprise software, adopters of ERP systems must change configuration tables so that they align with business processes (Ko et al. 2005). Further, local business rules must be incorporated into ERP software through a process of configuration and customization; for example, a new sales order transaction will generate a change in inventory values, and such changes are determined by specific inventory valuation methods.

The knowledge and expertise of using IT to enable business process innovation is typically embodied within IT workers (e.g., Tambe and Hitt 2014). This expertise can be accumulated through a process of on-the-job skill acquisition that has been shown to have an important impact on labor productivity in a variety of contexts (Benkard 2000, Lieberman 1984, Thornton and Thompson 2001), or facilitated by accessing external knowledge sources. Researchers have long understood how such knowledge flows from external sources can augment the human capital of scientists (Furman et al. 2005, Griliches 1992) and increasing attention has recently been focused on examining their impact on the human capital of IT workers (Chang and Gurbaxani 2012b). Another channel through which knowledge can be transferred between firms is the direct acquisition of human capital through employment contracts. For example, through the acquisition of experienced IT workers, firms can obtain access to knowledge gained by these workers through their training at their previous employer (Tambe and Hitt 2014b).

In addition to formal contractual relationships, knowledge of how to use IT to enable business process innovation can also be transferred through informal interactions between firms. These are often labeled knowledge spillovers, since external knowledge flows affect the productivity of the receiving firm (Griliches 1979). In the context of enterprise software, there are many channels through which these types of informal interactions take place; we provide several examples in the specific context of enterprise resource planning such as SAP, for more examples see Iansiti and Lakhani (2009). The Americas' SAP Users' Group (ASUG) hosts face-to-face meetings where users can share experiences of implementing SAP software and benchmarking best practices both through formal presentations and through informal interactions between individual participants.[6] TechEd, the premier technical education conference sponsored by SAP, offers various hands-on workshops, demo-driven lectures, showcase exhibitions, and Q&A sessions on the latest developments in enterprise software by SAP and its solution partners.[7] SAP also provides opportunities for knowledge transfer and human capital development via online channels. As described in further detail below, the SAP Developer Network (SDN) provides an opportunity for SAP users, partners, and SAP employees to provide user-to-user support through the use of web collaboration tools (Lakhani and von Hippel 2003, Van Alstyne and Benbya 2011).

Recent evidence from other avenues for knowledge exchange such as open source software development and standards-setting processes suggests that participation in such environments, whether virtual or physical, can augment the human capital of participants (Lakhani and von Hippel 2003, Waguespack and Fleming 2009). In the context of developing and implementing IT systems, this type of human capital augmentation will make IT workers more productive when engaging in valuable business process innovation, which will in turn have a positive impact on firm productivity (Tambe and Hitt 2014). In line with this thinking, we investigate whether exposure to external knowledge inputs through informal channels like online communities will be positively associated with firm total factor productivity (TFP).

More importantly, firms may differ in their abilities of assessing the value of the external IT knowledge and applying it for productive use. It is well understood in the context of R&D that outside sources of knowledge are an important input into the innovation process (Cohen and Levinthal 1990), and

[6] In 2014, more than 75,000 users attended such meetings (https://www.asug.com/about/what-we-do).

[7] http://events.sap.com/teched-global/en/home

a firm's investment in R&D serves dual purposes – it not only generates new information, but also enhances the firm's ability to identify, assimilate and exploit knowledge from its external environment (Cohen and Levinthal 1989). This ability, termed "absorptive capacity", is shown to be path-dependent and is a function of prior knowledge accumulation (Cohen and Levinthal 1990). Applying this logic to a firm's IT investment (Roberts et al. 2012), we argue that it also serves dual purposes – it not only creates value through the use of the systems, but also equips the firm with the ability to appreciate the value of external IT knowledge, assimilate it, and apply it for productive use. Therefore, a firm's IT-specific absorptive capacity is a function of its prior accumulation of related knowledge stock, which contributes to the firm's productivity indirectly when opportunities to absorb external knowledge inputs arise.

2.2 Research Model of the Drivers of Absorptive Capacity

We assume that a firm's knowledge stock related to the use of enterprise information systems is one of the drivers of its productivity and performance. Following the work of Cohen and Levinthal (1989), henceforth C&L, we model the knowledge stock as

$$Z_{it} = M_{it} {S_{it}}^{\gamma_{it}}, \tag{1}$$

where M_{it} represents a firm's cumulated investment in enterprise systems, which include investments in a combination of hardware, software, and human capital. S_{it} represents flows of external knowledge available in the public domain that can be accessed through various channels, as discussed above.[8] γ_{it} measures the extent to which the focal firm is able to recognize the value of new, external information, assimilate it, and effectively utilize it in a business setting. It therefore represents the firm's absorptive capacity. Equation (1), similar to C&L, but within an IT investment setting, embodies the idea that the effect of IT knowledge flows on productivity is critically conditioned by a firm's absorptive capacity. Following C&L, we also assume that

$$\gamma_{it} = f(M_{it}, D_{it}), \tag{2}$$

i.e., absorptive capacity is a function of the firm's cumulated investments in enterprise systems and the characteristics of outside knowledge that make learning more difficult (or difficulty of learning, D_{it}). Given the above framework, and in line with the theory in C&L, we present below a set of hypotheses related to the drivers of absorptive capacity.

---

[8] For simplicity, we ignore the role of extra-industry knowledge here. In C&L, extra-industry knowledge includes "output of government or university laboratories." Our model and data analysis assumes that knowledge stocks are formed based on the cumulated enterprise IT investments and flows of external knowledge. We have experimented with an alternative model and data analysis in which we use stocks instead of flows of external knowledge and our key predictions continue to hold.

2.2.1 Prior Related IT Investments

Firms engaging in new business process innovation and who wish to augment their human capital through the acquisition of knowledge from external relationships will experience greater benefits when they have already made inroads through internal knowledge accumulation (Cohen and Levinthal 1989, Ko et al. 2005). As noted earlier, within the context of business process innovation, general purpose IT innovations that have been developed outside of the firm will need to be adapted to specific firm needs (Bresnahan and Greenstein 1996).

The impact of prior knowledge in facilitating the absorption and adaptation of external knowledge to the focal firm's idiosyncratic needs will be salient when some portion of that prior knowledge stock is related to that acquired externally (Cohen and Levinthal 1990). In our setting, prior experience with deploying business applications software will be more valuable in absorbing knowledge about ERP systems than that with deploying IT infrastructure such as PCs and servers. We investigate the salience of this traditional hypothesis within the absorptive capacity literature by investigating the extent to which IT expertise accumulated within an organization as a result of cumulated investments in enterprise systems increases the productivity benefits of external knowledge acquisition. Referencing equations (1) and (2) of our research model, we formulate the following hypothesis:

Hypotheses 1: $\frac{\partial \gamma_{it}}{\partial M_{it}} > 0$. Firms are better able to exploit external knowledge related to enterprise software when they have made prior investments in enterprise systems.

2.2.2 Characteristics of External Knowledge

The costs of transferring knowledge across firm boundaries will depend upon the nature of the knowledge (Teece 1977). C&L highlight that the value of external knowledge to a firm will be greater when that knowledge is easier to assimilate and exploit. Knowledge may be difficult to transfer because it is tacit (Kogut and Zander 1992) or costly to adapt and use in a new environment (von Hippel 1994). One view has held that knowledge will be less complex and easier to transfer across firm boundaries when it can be written down in terms of a limited number of essential elements (Arora and Gambardella 1994).

In our context, business process innovation requires both the implementation and installation of technical systems as well as complementary changes to processes, decision rights, and the organization of economic activity within firms.[9] These latter types of changes are commonly believed to be more costly

[9] There is by now an active literature that has investigated the implications of these latter changes for productivity and firm value, non-convexities in their payoffs, and the circumstances under which they are most costly to perform. See, for example, Attwell (1992), Bresnahan and Greenstein (1996), Bresnahan et al. (2002), Hubbard (2003), Forman et al. (2005), and Bloom et al. (2012).

and difficult to perform than the former (Brynjolfsson et al. 2006), because they are often context-dependent. For example, decisions related to physical and human capital, processes, and organization may interact in unexpected ways and generate significant non-convexities in payoffs (Milgrom and Roberts 1990). As a result, these latter activities require identifying the correct system of activities within the context of the firm, and implementing them successfully. In contrast, the implementation and installation of technical systems are often less context-dependent. For example, programmers will often reuse software written in other applications (Barns and Bollinger 1991, King and Lakhani 2013). Even when they do not involve direct copying, lessons learned by programmers and other technical IT workers are more likely to generalize to other settings. For example, within the context of the open source Apache project, Lakhani and von Hippel (2003) report that users who provide information to a help forum spend 98% of their time on the website reading the answers to other posted questions. This effort helps them to learn about problems that other Apache users are experiencing that in turn aid them in managing their own Apache installations.

The discussion above suggests that knowledge related to the implementation of technical systems will be more abstract and less context-dependent than that related to complementary organizational changes required in business process innovation, and therefore is easier to transfer across firm boundaries. As a shorthand, we label the first type of knowledge 'technical knowledge' and the latter 'business functional knowledge.' If the cost of knowledge transfer is higher – or the knowledge is difficult to learn – the knowledge flow may have a lower likelihood of being successfully absorbed and applied to productive use.

The ease of transfer may also depend upon the novelty of external knowledge. When knowledge is new, there may be less information on how to use it properly (Cohen and Levinthal 1989, von Hippel 1994). For example, avoidance of problems when using a new process machine may require a great deal of information about the setting where it is to be applied (von Hippel and Tyre 1995). Further, in environments undergoing rapid technological change, absorption of new knowledge may require keeping track of new developments within the field (Cohen and Levinthal 1989). With reference to IT spillovers, knowledge that is related to new IT developments such as new emerging fields will be more difficult to transfer.

To summarize the above discussion, we expect that, all else equal, transfers of knowledge that are difficult to learn will have smaller productivity effects. Referencing our key relationships formalized by equations (1) and (2), we therefore formulate the following hypothesis:

Hypotheses 2: $\frac{\partial \gamma_{it}}{\partial D_{it}} < 0$. Firms are less able to exploit external knowledge related to enterprise software when the external knowledge is more difficult to learn.

2.2.3 The Interaction of Prior Related IT Investments and Characteristics of External Knowledge

The stock of related IT investments and the nature of knowledge acquired externally are also characterized by important interdependencies. In particular, the role of prior IT investments in the process of knowledge assimilation depends upon the nature of knowledge the firm is seeking to acquire. A related point has been demonstrated within the R&D literature (Cohen and Levinthal 1989), which has shown that related internal R&D becomes more important to the acquisition of external knowledge when that external knowledge is more complex and less targeted at the needs of the firm. In these circumstances, the firm's prior related investment becomes vital to the acquisition and assimilation of external knowledge.

As discussed above, we focus on two factors making external knowledge more difficult to learn, which in turn might influence the extent to which prior related IT investments are critical drivers of a firm's absorptive capacity. First, we expect that the prior IT investments play a greater role in absorbing knowledge acquired from external sources when a large fraction of external knowledge involves business functional knowledge. In other words, firms with prior related IT investments will receive greater benefits (i.e., higher productivity) from external knowledge flows related to business functional knowledge – knowledge that is difficult to be transferred and absorbed. This is because it is more challenging for the receiving firm to translate insights gained from business functional knowledge acquisition into a valuable set of actions related to processes, decision rights, and organization. However, firms with prior related investments on how to implement business application software will be able to derive value from inflows of business functional knowledge—in this setting, firms are able to put insights learned into productive use. However, prior IT investments and associated cumulated IT human capital will have little influence on the value achieved from flows of technical knowledge, because transferring such knowledge requires little adaptation and customization.

Second, prior IT investments will be more critical to knowledge absorption when the external knowledge is more novel. In addition to the findings C&L mentioned above, a related point has been made in the R&D literature by Escribano et al. (2009), who found that absorptive capacity plays a more critical role in environments where knowledge changes rapidly. In our context, we expect prior IT investments and experience to help the receiving firms better understand, evaluate, and more effectively utilize new knowledge in emerging IT domains.

To summarize, referencing by equations (1) and (2), we formulate the following hypothesis:

Hypotheses 3: $\frac{\partial^2 \gamma_{it}}{\partial M_{it} \partial D_{it}} > 0$. Prior investments in enterprise systems are more important to the firms' ability to exploit external knowledge when the external knowledge is more difficult to learn.

3. Research Context

Our research questions require a robust measure of inter-firm knowledge flows related to the use of IT. We use the online community network created by SAP AG, the largest enterprise software vendor by revenue during our sample period. SAP established in 2003 an internet-based network of practice, the SAP Developer Network (SDN). The SDN was later expanded to include a community for business process experts, and was expanded still further over time to incorporate other communities that interface with SAP's products. Given this increase in breadth, the SDN is now known as the SAP Community Network.

The SCN serves as a resource repository and a platform for SAP users, developers, architects, consultants and integrators to collaborate and exchange knowledge on the adoption, implementation and customization of SAP solutions. It hosts forums, expert blogs, a technical library, article downloads, a code sharing gallery, e-learning catalogs, wikis and other facilities through which users contribute their knowledge. As of 2008 there were over 199,000 active users from 224 different countries in the community.

We choose enterprise software as the background for measuring IT knowledge flows for several reasons. First, investment in enterprise software and its implementation accounts for a significant portion of total business-related IT spending (Brynjolfsson et al. 2002) – in some cases accounting for as much as 75% of corporate IT investment (McAfee and Brynjolfsson 2008), and adoption of enterprise software has been shown to be associated with significant improvement in firm financial and operational performance (Hitt et al. 2002). However, implementing enterprise software is complex and requires complementary business process innovation: because of these challenges, projects frequently take longer than expected and benefits take a long time to achieve (McAfee 2002). Last, knowledge of how to implement ERP systems is widely distributed among users (Yusuf et al. 2004) and, because of the heterogeneous environments in which systems are implemented, expertise on how to deploy them is not easily contracted upon and internal human capital accumulation occurs as users learn how to deploy software functionality in their organizations through a series of projects (Walker 2012). Thus, our environment offers a useful first test case for understanding the interrelationships between the internal stock of IT knowledge human capital and efforts to acquire additional human capital through external interactions.

A unique feature of the SAP Community Network is that the knowledge flows from one user to another can be observed and quantified. The community has a contributor recognition program (CRP) that awards points to community users for contributing technical articles, code samples, videos, wiki entries, forum posts, and weblogs. For example, when users participate in a forum discussion, they can receive points for posting solutions to existing discussion threads marked as questions, if their answers are deemed helpful by the person who asks the question. SAP publicly recognizes its most active contributors. For example, on the "Top Contributors" page, the top 50 contributors measured by total reward points are listed. On each discussion forum page, the top three contributors to that forum are listed with their total points received. In addition, SAP identifies and provides special status to exceptional and high-value contributors by granting them the honorary title of "SAP Mentor".

Participation in the community network is voluntary and anyone can register as a user by providing basic personal information. In particular, users can optionally list the firm that employs them. Using this piece of information, it is possible to aggregate the knowledge flows to firms whose employees actively participate in SCN (details on the measurement of the knowledge flows will be introduced in the next section). The user's profile also lists the country that the user comes from, her relationship to SAP, email address, phone number, expertise, and LinkedIn profile page. Figure 1 presents a sample user profile.

[Insert Figure 1 Here]

To track knowledge flows between the users of SCN, we focus on user interactions through the most frequently used communication format: the discussion forums. Although users of SCN may access knowledge through other formats such as wikis, blogs, and articles, the number of active participants in these other formats is smaller in comparison to those in the discussion forums, and knowledge flows arising from the use of these other channels are unfortunately not measurable. The primary purpose of the discussion forums is to provide an avenue for conversations between the community users to help one another solve problems encountered during the implementation, deployment and use of SAP software (Fahey et al. 2007). The forums are organized according to domains of knowledge or expertise, each of which usually corresponds to a technical domain (such as database or operating system), a particular SAP software module, or the application of SAP to a particular industry.[10]

Conversations in each forum are organized by discussion thread. Each *thread* is initiated by a knowledge seeker, who posts a specific question in a topic forum of her choice. Knowledge contributors, on the other hand, post responses that are attempts at answering the question. A discussion thread is comprised of a list of *messages*, and each message (either a question or an answer attempt) contains the

[10] Examples of forums include SAP on SQL server, data transfers, ERP manufacturing, product life cycle management, CRM-interaction center, and SAP for automotive solutions.

information about the member who posts the message, the body of the message, and a time stamp. Once a correct answer (determined by the knowledge seeker) is received, the discussion thread is closed. Figure 2 presents a sample discussion thread in the forum of SAP ERP Manufacturing - Production Planning, with a question, a correct answer, and a helpful answer.

[Insert Figure 2 Here]

We developed a web scripting tool and obtained the complete history of SAP forum discussions from 2004 to 2008. The dataset includes about 1.1 million discussion threads with 5.0 million messages posted in 209 topic-specific forums. Table 1 presents some summary statistics of the evolution of the SCN over our sample period, including numbers of registered users, topic forums and the discussion threads posted in these forums. Overall, we find that the online community has experienced rapid growth since its establishment: by the end of our sample roughly one quarter of all questions raised are solved by the collective effort of the community users, and the average time to obtain a correct solution is less than five days.

[Insert Table 1 Here]

4. Data and Methods

4.1 Estimation models

We adopt a production function approach and extend it by introducing the role of knowledge stock related to enterprise systems. A typical production function relates firm output to factors of input (Hulten 2001). For example, a simple form of a three-factor Cobb-Douglas production function has been widely used in prior studies on IT productivity (Brynjolfsson and Hitt 1996, Dewan and Min 1997, Mittal and Nault 2009):

$$Y = AK^{\alpha}L^{\beta}C^{\eta} \tag{3}$$

Where Y is the quantity of production output, K is the stock of non-IT capital, L is the stock of labor, C is the stock of IT capital, and A denotes the total factor productivity (TFP). TFP is defined as the output contribution that is not explained by the factor inputs and is often interpreted as technological progress. In this case, the output elasticity of IT capital, $\eta = \partial lnY/\partial lnC$, represents the percentage increase in output due to a one percent increase in IT capital. To incorporate the role of external IT spillovers, we follow the literature on R&D spillovers by adding to (3) a factor that captures the knowledge stock related to enterprise software – Z_{it} as in equation (1) – which is a function of cumulated investments in enterprise systems, M_{it}, and knowledge obtained from outside the boundary of the firm, S_{it}

(e.g., Chang and Gurbaxani 2012a, Griffith et al. 2006). Assuming the output elasticity of Z_{it} is φ, we have:

$$Y_{it} = AK_{it}^{\alpha}L_{it}^{\beta}C_{it}^{\eta}{M_{it}}^{\varphi}{S_{it}}^{\varphi\gamma_{it}} \tag{4}$$

As in equations (1) and (2), γ_{it} reflects a firm's absorptive capacity, assumed to be a function of the firm's prior investments in enterprise systems (M_{it}), the degree to which the external knowledge is more difficult to assimilate (D_{it}), and their interaction, i.e., using a translog specification,

$$\gamma_{it} = \gamma_0 + \gamma_1 lnM_{it} + \gamma_2 lnD_{it} + \gamma_3 lnM_{it}lnD_{it} \tag{5}$$

The interaction term highlights how the importance of prior IT investments to knowledge assimilation will depend upon the difficulty of learning, a key feature of the model of C&L. Note that in C&L $\gamma_0 = 0$. We add it here to be more general and it allows absorptive capacity to have an independent effect on the production function.

We can now rewrite the production function as

$$Y_{it} = AK_{it}^{\alpha}L_{it}^{\beta}C_{it}^{\eta}{M_{it}}^{\varphi}S_{it}^{\varphi(\gamma_0+\gamma_1 lnM_{it}+\gamma_2 lnD_{it}+\gamma_3 lnM_{it}lnD_{it})} \tag{6}$$

Or in log form

$$y_{it} = a + \alpha k_{it} + \beta l_{it} + \eta c_{it} + \varphi m_{it} + {\gamma_0}' s_{it} + {\gamma_1}' m_{it}s_{it} + {\gamma_2}' d_{it}s_{it} + {\gamma_3}' m_{it}d_{it}s_{it} \tag{7}$$

Where we use standard notation and denote logged terms using lowercase letters. Further, $\gamma_0' = \varphi\gamma_0$ and we use similar notation for the other knowledge absorption terms. After introducing firm and year fixed effects and the idiosyncratic error, the model becomes

$$\begin{aligned} y_{it} = a &+ \alpha k_{it} + \beta l_{it} + \eta c_{it} + \varphi m_{it} + {\gamma_0}' s_{it} + {\gamma_1}' m_{it}s_{it} + {\gamma_2}' d_{it}s_{it} + {\gamma_3}' m_{it}d_{it}s_{it} \\ &+ \mu_i + \delta_t + \varepsilon_{it} \end{aligned} \tag{8}$$

We use equation (8) to test our hypotheses. As explained in section 2.2, H1 asserts that the stock of internal IT knowledge is associated with greater success of assimilating external knowledge for productive use ($\frac{\partial\gamma_{it}}{\partial M_{it}} > 0$). From equation (5) and (6) we have the output elasticity of S, or $\eta_S = \frac{\partial y_{it}}{\partial s_{it}} = \varphi\gamma_{it}$. Therefore, testing H1 is equivalent to testing $\frac{1}{\varphi} * \frac{\partial\eta_S}{\partial M_{it}} > 0$, or $\frac{1}{\varphi M_{it}} * \frac{\partial^2 y_{it}}{\partial s_{it}\partial m_{it}} > 0$. If the output elasticity of the knowledge stock is positive (φ>0), a standard assumption in the productivity literature, then testing H1 is equivalent to testing $\frac{\partial\eta_S}{\partial m_{it}} > 0$, or $(\gamma_1' + {\gamma_3}' d_{it}) > 0$. That is, the test of H1 will depend upon the d_{it}, or difficulty of knowledge flows, at which the derivative is evaluated. In keeping with standard practice, we will provide test statistics at mean values but will also characterize the derivative at other points in the distribution. A test of H2 ($\frac{\partial\gamma_{it}}{\partial D_{it}} < 0$), is similarly equivalent to testing $\frac{\partial\eta_S}{\partial d_{it}} < 0$, or $(\gamma_2' + \gamma_3{}' m_{it}) < 0$. In addition, we expect the stock of internal IT knowledge is more critical to absorbing external knowledge flows when they are more difficult to learn ($\frac{\partial^2\gamma_{it}}{\partial M_{it}\partial D_{it}} > 0$), which is equivalent to testing $\gamma_3' > 0$.

4.2 Data

We construct a dataset of publicly traded firms that are SAP adopters. Our data come from a variety of sources. Our primary measure of knowledge transfer comes from user activities in the discussion forums on the SAP Community Network. To identify SAP adopters, we obtained a detailed list of all installations of SAP product modules in the United States prior to the end of year 2004 from SAP AG. We use the Harte Hanks Computer Intelligence (CI) Technology database to collect firm-level IT investment data. The CI database records detailed information about IT infrastructure for most of the Fortune 1000 firms, including data on the quantity of mainframes, peripherals, minicomputers, servers and PC systems, as well as other IT hardware stocks. The CI database has been widely used by prior studies to investigate issues related to IT productivity (e.g., Brynjolfsson and Hitt 2003, Chwelos et al. 2010, Dewan et al. 2007). The CI data were matched with Standard and Poor's Compustat database to obtain financial data that we use to construct measures of production output, non-IT capital stock and labor expenses.

4.2.1 Sample

Our sample is constructed in several steps. It begins in 2004 with the start of SCN and ends in 2008, which is the last year that we have IT data.[11] To obtain the data for our sample, we first retrieve the set of firms that were on the Fortune 1000 list at least once during 2004-2008 and match them to Compustat data. We then match these firms with the CI database. Because we are interested in knowledge spillovers related to the implementation and the related business process innovation in SAP products, we further restrict our sample to those firms that had installed at least one SAP module prior to the end of 2004. The final sample is an unbalanced panel of 275 firms with 1240 observations over a 5-year period.

---

[11] While SCN originated in 2003, it was re-launched in 2004 with a new reward system so we use 2004 as the starting point for our analysis.

4.2.2 Variables

Here we describe the variables used in our analysis. We first document the variables that measure productivity, labor, and IT and non-IT capital, and then describe the variables measuring knowledge flows and those related to absorptive capacity.

Production Function Inputs and Outputs

IT Capital

Our measure of IT capital is derived from the CI database. The information in the database covers major categories of IT hardware investments made by firms, such as personal computing, systems and servers, networking, software, storage and managed services. Historically the CI database has provided direct measures of IT capital stock, but this measure is not available over the years of our sample. As a proxy, we adopt the method used by Brynjofsson and Hitt (1995), Hitt and Brynjofsson (1996), Dewan and Min (1997), and Gu et al. (2008) and measure the IT capital stock using an estimate of the market value of the IT hardware systems plus three times the current year IT labor expenses. Inclusion of IT labor expense in the calculation of IT capital is justified by the fact that a large fraction of IT labor expenses is dedicated to the development of computer software, which is a capital good. The assumption that underlies this method is that the current IT labor spending is a good proxy for the IT labor expenses in the recent past, and IS staff "stock" depreciates fully in three years (Brynjolfsson and Hitt 1995).

The first component of this variable is equal to the market value of total PCs and servers currently owned by the firm, converted to constant 2005 US dollars. To be specific, we collect market prices of PCs and servers in the United States from two report series produced by the *Gartner Dataquest Market Statistics* database—Gartner Worldwide Server Forecast and Gartner Worldwide PC Forecast—from 2004 to 2008. These two report series present detailed statistics on the number of shipments, prices, vendor revenues and other related information about PC and Servers, broken down to the level of each geographic region and market segment.[12] Our market prices for PCs and servers are calculated as the average user price across their respective market segments within the United States. These prices are then multiplied by the quantities of PCs and servers owned by the firm to derive the market value of the IT computer assets for each firm. Finally, we deflate the market value by the Bureau of Economic Analysis (BEA) price index for computers and peripheral equipment.

[12] Gartner Dataquest defines PC market segments as: desk-based, mobile, professional, and home. Server market segments are defined by CPU types, which include x86, IA64, RISC, and other. The database covers the global regions of Asia/Pacific, Eastern Europe, Latin America, Middle East & Africa, and Western Europe. Several country level statistics are also available, including the United States, Canada, and Japan.

The second component of IT capital stock is IT-related labor expenses. The CI database provides the number of IT employees of the sample firms at the site level (which falls into one of the following ranges: 1-4, 5-9, 10-24, 25-49, 50-99, 100-249, 250-499, and 500 or more), where a site represents a particular firm-location, much like the concept of establishment in Census data. We aggregate the site-level employee numbers to the firm level to derive the total number of IT-related employees hired by the firm. For each range, we take the middle value of the range as the number of IT employees. IT labor prices are obtained from the Occupational Employment and Wage Estimates series from the Bureau of Labor Statistics (BLS) Occupational Employment Statistics (OES), and we use the mean annual wage of computer and mathematical occupations as the average labor price for IT employees. As the wage reported by the OES series does not reflect benefits, we multiply the wage number by the ratio of total compensation to salary, which is obtained from BLS Employer Costs for Employee Compensation (ECEC) series. The IT labor expense is then deflated by the BLS Employment Cost Index (ECI) for private industry workers.

Production Output

We follow prior literature (Brynjolfsson and Hitt 2003, Dewan and Min 1997) and use added value as the measure of production output, which equals deflated sales less deflated materials. Compared to sales, added value is said to be less noisy and more comparable across industry sectors (Dewan and Min 1997). Annual sales numbers are retrieved from Compustat, and we deflate them using industry-specific (at 2-digit NAICS sector) price deflators from BEA *Gross Output and Related Series by Industry*. Materials are calculated by subtracting undeflated labor and related expenses (Compustat data item XLR) from undeflated total operating expenses (Compustat data item XOPR), and deflating by the BLS Producer Price Index (PPI) for intermediate materials, supplies, and components.

Non-IT Capital

The calculation of total capital stock is similar to that in Brynjolfsson and Hitt (2003) for ordinary capital. Specifically, the gross book value of capital stock (property, plant and equipment (Total-Gross), Compustat data item PPEGT) is deflated by an industry-specific capital investment deflator reported in BLS 1987-2010 Detailed Capital Measures.[13] In order to apply the deflators, the average age of capital stock is calculated as the ratio of total accumulated depreciation (Compustat data item DPACT) to current depreciation (DP). We then subtract the deflated computer capital from deflated total capital to get the value of non-IT capital.

Non-IT Labor

---

[13] Retrieved from http://www.bls.gov/mfp/mprdload.htm.

Consistent with prior studies on IT productivity (Bresnahan et al. 2002, Brynjolfsson and Hitt 2003), total labor expense is either obtained directly from Compustat labor and related expenses (data item XLR), or calculated as the product of a firm's reported number of employees (Compustat data item EMP) and industry-average labor cost per employee, and deflated by the BLS Employment Cost Index (ECI) for private industry workers. Average labor cost per employee is obtained from *National Sector NAICS Industry-Specific estimates* series of BLS OES. To account for the fraction of benefits in total compensation, we multiply the wage number by the ratio of total compensation to salary, which is obtained from BLS *Employer Costs for Employee Compensation* (ECEC) series. Non-IT labor is defined as the difference between deflated total labor expense and IT labor expense.

Variables Measuring Knowledge Flows and Absorptive Capacity

In this section we describe the set of variables that measure the flows of external knowledge (S_{it}) and the factors influencing the absorption of knowledge, D_{it} and M_{it}. Except as noted otherwise, we take logs of these variables when they are entered into our empirical model.

S_{it}: *External Knowledge Flows*

We measure flows of IT-related knowledge acquired externally, S_{it}, from forum conversations that took place on the SAP Community Network. For each question that is posted, the rules of the SAP reward program specify that the knowledge seeker can use her discretion to judge the quality of answers posted by knowledge contributors and distribute reward points as follows: 10 reward points for correct answers (at most 1 answer can be evaluated as correct), 6 points for very helpful answers (at most 2 answers), and 2 points for helpful answers (no limit on number). We define a knowledge inflow as an incident where a knowledge seeker gives reward points to knowledge contributors in recognition of their quality responses. As noted above, we use a crawler program to identify user information such as location and firm. Next, we select all the users that reside in the United States, and match them to firms in our sample by examining their employer affiliations and domains of their email addresses.

For each user a who is an employee of firm i, we retrieve all the discussion threads that were initiated by a in year t, and examine the history of the answers posted by other forum users. If a received any correct, very helpful, or helpful answers in year t, the total number of reward points she gave to the knowledge contributors are used as a proxy for inward IT spillovers to a. The reward points are then aggregated across all the threads posted by a in year t to derive an individual-level knowledge inflow, S_{at}. The firm level spillover variable is defined as the sum of knowledge inflows of all the individuals who are employees of the firm:

$$S_{it} = \sum_{a \in F_i} S_{at}$$

where F_i is the set of users who are employees of firm i. We exclude in this measure within-firm knowledge flows, i.e., knowledge flows in which both the source and the recipient are employed by the focal firm i.

Our measure of knowledge inflow is likely to suffer from measurement error due to missing data on the knowledge seekers who did not report their employers.[14] However, we observe no systematic differences in knowledge inflow between questions asked by knowledge seekers who reported their employers and those asked by seekers who didn't reveal their employers: the average inflow per question per year is 2.91 for non-reporting seekers and 3.08 for reporting seekers, and the difference is not statistically significant (p=0.35). If firms strategically promote employee activity in SCN and other communities (Mehra et al. 2011), then S_{it} may serve as a proxy for the broader receptivity of the firm to external inflows. We consider this possibility further in sections 5.1 and 5.2.

D_{it}: *Difficulty of Learning*

We create two measures of difficulty of learning based upon the knowledge flows in S_{it}. First, prior research has emphasized two distinct dimensions of IT knowledge that are particularly relevant in the process of adopting an information system: technical knowledge and business functional knowledge. As noted above, technical knowledge is primarily related to enabling technologies such as operating systems, programming languages, database management systems, networks, and telecommunications. In contrast, business functional knowledge is related to the adaptation of business processes and functions in the implementation of information systems and the use of IT to achieve business goals (Ko et al. 2005, Lee et al. 1995).

We distinguish between these two sources of knowledge by examining the forum in which a knowledge seeker's question is raised. We define a forum as technical-oriented if the forum is dedicated to topics related to low-level, enabling technologies of an enterprise system such as programming languages, database technologies, data transfer issues, and reporting and formatting tools. In contrast, we define a forum as business-oriented if the discussion topics in the forum focus on the configuration of the enterprise system to implement a particular business function or process, such as monitoring employee

[14] For example, among all the discussion threads that are initiated by US knowledge seekers during our sample period, only 48% (23,973 out of 49,977) of them have a seeker who reported her employer.

performance, coordination of supply chains, consolidating procurement processes, or managing projects.[15,16]

Preliminary statistics reveal some interesting contrasts between the two categories of knowledge that the firms receive from SCN. First, we notice that there is an asymmetry in the demand for the two types of knowledge: there are fewer forums that are dedicated to topics related to business functions (77) than those dedicated to technical questions (116). In addition, community users raise more technical questions (658,574) than business functional questions (453,271) during our sample period. Second, on the supply side of the knowledge exchanges, we observe that business-related questions in general receive fewer responses from knowledge contributors (4.55 vs. 4.81 responses per question asked, $p<0.01$), they are more difficult to solve (with the probability of obtaining a correct answer to a question equal to 0.226 vs. 0.262, $p<0.01$), and therefore the knowledge seekers of business-related questions in general receive fewer inward knowledge flows per question (reward points per question equal to 5.33 vs. 5.72, $p<0.01$). As a result of the interaction of these demand-side and supply-side factors, firms in our sample on average receive more technical knowledge inflows (mean=6.75 reward points for each firm-year in our sample) than business functional knowledge inflows (mean=3.75 reward points).

Our first measure for the Difficulty of Learning D_{it} uses the share of the inflows that are related to business functional knowledge. Specifically, we define D_{it} as (1 + business functional knowledge flows)/ (1+ total knowledge flows), where 1 is added to both the numerator and denominator to avoid taking the log of 0 when the variable d_{it} is entered into the regression.

We use as an alternative measure for the Difficulty of Learning a variable capturing the novelty of external knowledge and the rate at which it is changing. We measure this using differences in the timing of introduction of discussion forums on SCN. We construct D_{it} as the percentage of knowledge flows that are derived from forums less than one year old at the time the knowledge flow takes place. The underlying assumption is that new forums relate to recently introduced products and services which are in the early stage of their life cycle. Such products have not been widely deployed, and generalized best

---

[15] Some examples of technical-oriented forums are: Java Programming, Form Printing, SAP on SQL Server, Service-Oriented Architecture, and Data Transfers. Some examples of business-oriented forums are: Logistic Materials Management, Sales and Distribution General, ERP Operations – Quality Management, Knowledge Management & Collaboration, and Product Lifecycle Management.

[16] The classification process is done by independent coding from two of the authors who have conducted prior research related to SAP enterprise software, and two research assistants who have experiences using SAP software products. The rare inconsistencies in the classification results across the coders (6 out of 209) were resolved by checking the content of the forums and discussions among them. We remove forums for which we do not have sufficient information to make a classification, or those for which the classification scheme does not apply (for example, forums such as career center, suggestions and comments, etc.). In total, 16 forums out of 209 are excluded for these reasons.

practices have not been well-established. As a result, it may be more difficult to translate solutions that have been deployed successfully at one firm within the context of another. The variable, D_{it}, is defined as (1+knowledge flows from new forums)/(1+total knowledge flows).

M_{it}: Prior Investments in Enterprise Systems

Unlike our variables S_{it} and D_{it} which are variables for which we have direct measures of knowledge flows and their characteristics, we do not directly observe prior investments in enterprise systems. As is well-known, measures of direct (i.e., software license fees) and indirect (i.e., human capital investments) spending related to enterprise system cannot generally be observed except through survey measures such as those employed by Brynjolfsson et al. (2007). In the absence of direct measures of M_{it} we compute two separate proxies based on data from inside and outside the forum. Because these data are approximations to the underlying variable of interest, we create indicator variables to employ a flexible functional form.

We first measure M_{it} using the SAP installation data we obtained from SAP AG. The implementation of SAP enterprise software is highly modular, and a typical SAP system consists of a series of technical and functional modules.[17] On top of these technical and functional modules, the user firms can optionally install and deploy a series of advanced business suite products such as CRM (customer relationship management), SCM (supply chain management), SEM (strategic enhanced management), APO (advanced planner optimizer), EP (enterprise portal), KM (knowledge management), or SRM (supplier relationship management). To measure prior related investments with SAP products, we created a binary indicator variable that is set to 1 if any of these advanced business suites is installed by our focal firm prior to year 2004 (the first year of our sample), and 0 otherwise.

Because prior related knowledge depreciates over time and our benchmark measure does not vary during our sample period, we create an alternative measure of prior related knowledge based on the participation of the firm's employees in the SCN community. Specifically, we compute the cumulative contributions to SCN forums (measured by reward points earned) made by all the employees of firm i prior to year t and create a binary variable whose value is set to 1 if cumulative contribution made by a firm's employees is greater than the sample mean. We view these variables as related, and use them together to triangulate our understanding of the behavior of same (ultimately unobserved) variable. An

[17] Typical SAP technical modules are ABAP (Advanced Business Application Programming) and BASIS (Business Application Software Integrated Solution). Typical SAP functional modules are FICO (Finance & Controlling), HR (Human Resource), PP (Production Planning), MM (Material Management), SD (Sales & Distribution), PM (Plant Management), PS (Project System) and QM (Quality Management).

analysis of the data supports this assertion: the mean knowledge outflows for firms in which M_{it}=1 is more than twice as large as those for which M_{it}=0.

Table 2 reports the summary statistics of the variables. The average firm in the sample has sales of $16.65 billion, added value of $5.49 billion, and 41,167 employees, consistent with our sample being large publicly-traded, Fortune 1,000 SAP adopters. In addition, firms in our sample invest heavily in IT capital, which has a mean level of $97.49 million and maximum of $1.18 billion. Table 3 provides the correlation matrix among the key variables.

[Insert Table 2 and Table 3 Here]

Table 4 presents a breakdown of the sample firms by vertical industries, which is based on 2-digit NACIS sectors. It is notable that firms in manufacturing industry account for the majority (66%) of the sample, followed by utilities firms (8%).

[Insert Table 4 Here]

4.3 Identification

As is well known, one challenge to estimating production functions such as regression equation (8) is the endogeneity of factor inputs. Knowledge terms s_{it} , m_{it}, and d_{it} are all the results of current or prior decisions by the firm, as are factor inputs k_{it}, l_{it}, and c_{it}. In this section we discuss the processes that give rise to the realizations of each of these variables, how they influence our efforts to identify the causal relationships of interest, and the implications for our estimation strategy.

The term S_{it} reflects the outcome of knowledge acquisition efforts by the firm. As discussed above, we measure these based upon knowledge inflows from SCN. Specifically, we will use a measure of the value of inflows from the SCN forums. We view these as a proxy for the firm's total knowledge acquisition efforts, which may include other knowledge acquisition activities such as participation in conferences or activity on other parts of SCN.[18] In this way, S_{it} follows the example of prior work in the literature on IT spillovers which has focused on specific types of knowledge acquisition such as that arising from employee movements (Tambe and Hitt 2014), from knowledge exchange with consulting firms (Chang and Gurbaxani 2012b), or the literature on R&D spillovers which has used patent citations (Furman et al. 2005) to proxy for knowledge flows.[19]

---

[18] This could include efforts to acquire knowledge from reading other posts on the SCN forum. See Lakhani and von Hippel (2003) for another example of such knowledge acquisition efforts within the context of open source software.

[19] However, our approach differs from approaches that use "spillover pools" that weight investments in R&D or IT by proximity to the firm to arrive at a proxy for the potential for spillovers (see, for example, Cheng and Nault 2007, Jaffe 1986).

Knowledge transfer to the firm will be influenced both by efforts by the firm to acquire new knowledge and by the collective response from the community to the knowledge acquisition efforts.[20] Efforts at knowledge acquisition will depend upon the firm's efforts to deploy and enhance enterprise software systems. In our context, this corresponds to the number of questions asked by the firm on the SCN forums. The number of questions asked on the forum may be higher when the firm is deploying a new system, for example. The deployment of new systems will increase both the need for complementary knowledge to deploy the system, and employee incentives to participate in the forum. Moreover, at the firm level, firms may endogenously decide whether to allow their employees to participate in communal activities like open source or community forums based on the marginal product of these activities for the firm (Mehra et al. 2011). This can influence both employee's incentives to participate and to report their employer in our data. In our regressions we use investment in IT hardware and software to control for firm efforts to deploy enterprise software, however we note that the number of questions may pick up a broad range of efforts related to human capital acquisition and business process innovation related to the deployment of enterprise software.[21] To further understand our results, in many regressions we include a control for the number of questions as well as the number of employees that participate in SCN, and observe how changes in knowledge inflows influence our results conditional on controlling for these. This allows us to control for a broad range of activities that may be correlated with business process innovation taking place at the firm.

Knowledge transfer to the firm will also depend upon the quality of responses that firm employees receive from their queries. Ultimately this is a human capital matching problem that will depend, among other things, upon the supply of related human capital and the decisions of others to allocate attention to the questions raised (Haas et al. 2015); this effort allocation may depend both upon the quality of the question asked (i.e., its clarity), the difficulty of the question, the distribution of expertise related to the question among the community members, as well as the reputation of the firm's employees and how often they have contributed to the community (e.g., Forman et al. 2008, Haas et al. 2015). To tease out the influence of potential variations in the knowledge supply decisions of others, we instrument for knowledge inflows using predicted values of knowledge inflows based on activity within the forums that the firm participates in.

The variable M_{it} denotes the firm's prior investments in enterprise software that will enhance the firm's knowledge stock, itself a critical driver of absorptive capacity. Our baseline measure of M_{it} is

[20] More broadly, knowledge transfer will be influenced by the channels available through which knowledge travels. In our setting, these will be fixed and determined by the forum that we study.

[21] The challenges of measuring activity related to business process innovation are well known. For further examples and discussion, see Bresnahan et al. (2002), Brynjolfsson et al. (2002), and Dranove et al. (2014).

based on investments made in SAP software prior to the start of our sample period. Our use of firm fixed effects will difference out unobserved time-invariant elements that may be correlated with M_{it} and will influence productivity, and will also mean that the main effect of M_{it} will not be separable from the firm fixed effects in estimation. Because the value of prior related investments may depreciate over time, we explore the robustness of our results to alternative measures that are time-varying, including a measure based on the firm's prior contributions to the community.

The difficulty of learning external knowledge, D_{it}, depends on the type of knowledge transferred to the firm. Like S_{it}, it is an outcome of the firm's knowledge acquisition strategy; firms with a high value of D_{it} are those who have asked and who have received answers to a large number of complex or idiosyncratic questions. Because high values of D_{it} may pick up heterogeneity in the types of projects pursued by the firm, we explore the robustness of our results to our use of a measure of D_{it} based only on the types of questions asked by the firm. This measure will capture heterogeneity in the types of projects pursued by the firm, but will not directly capture the incidence of knowledge inflows.

As in most papers exploring firm productivity in general and the productivity of IT inputs in particular, we face the challenge that k_{it}, l_{it}, and c_{it} are the outcome of firm-level input decisions, and so may be correlated with unobserved productivity shocks. As we describe in further detail below, we explore the robustness of our results to the use of dynamic panel data estimation.

More broadly, our econometric approach is affected by two challenges that are common to studies seeking to measure the influence of external knowledge flows and human capital accumulation (Adams 2006, Cohen 2010, Griliches 1992). First, knowledge flows, by their very nature, are difficult to observe and measure. Second, effort to seek and acquire new knowledge is a choice variable and so may be influenced by unobserved factors. We do not possess an exogenous shock that will influence the propensity to observe or acquire new knowledge. Our estimation strategy will be to first estimate a log-linear version of equation (4) that assumes $\gamma_{it} = \gamma$ for all firms (that is, no role for absorptive capacity) and explore the robustness of results to a range of alternative measures, omitted variables, falsification exercises, and instrumental variable strategies. Next, we estimate equation (8) and explore whether our results are consistent with a model of absorptive capacity. By simultaneously testing hypotheses 1-3 within the context of the absorptive capacity framework, we circumscribe the nature of unobserved heterogeneity that may be influencing our results; that is, if our results reflect omitted variables, they must be omitted variables that give rise to similar results for γ_1', γ_2', and γ_3' as predicted by our absorptive capacity model.

5. Results

In this section we first establish a link between inward knowledge flows and productivity by estimating a log-linear version of equation (4) that assumes $\gamma_{it} = \gamma$ for all firms (that is, firms are homogenous in their ability to absorb external knowledge). This allows us to establish a set of baseline results, and to examine the robustness of our results to different assumptions about measurement and omitted variables in a less complicated (log-linear) model. We then test our full model (equation (8)) that relaxes the assumption of constant absorptive capacity.

5.1 Analyses with Homogenous Absorptive Capacity

In this section we estimate a model where we assume that firms are homogeneous in their absorptive capacity (γ). Specifically, assuming prior investments in enterprise systems (m) remain constant over time, regression equation (8) reduces to the following model

$$y_{it} = a + \alpha k_{it} + \beta l_{it} + \eta c_{it} + \gamma s_{it} + \mu_i + \delta_t + \epsilon_{it} \ , \tag{9}$$

where γ reflects the average level of absorptive capacity of the sample firms. Since equation (8) and (9) are nested, we will present results from a nested specification test comparing these two models in section 5.3.

In Columns 1 and 2 of Table 5 we report the result from estimating equation (9) using fixed effects and random effects panel data models, respectively. The coefficients for the knowledge inflow terms are significant in both models ($p<0.05$), indicating that firms with greater inward knowledge flows produce more output, given the same amount of investment in capital, labor and IT. The result from a Hausman test comparing the fixed effects and random effects estimates rejects the orthogonality of the random effects and the regressors (p-value <0.01). Based on the result of the Hausman test we estimate the remainder of our models using fixed effects models.

[Insert Table 5 Here]

The results from our baseline fixed effects model in column 1 imply that a one percent increase in the amount of inward knowledge flows is associated with 0.009 percent increase in the added value produced by a firm. Considering that the added value of an average firm in our sample is $5.491 billion, this translates into a $0.48 million increase in production output. For the average firm in our sample, obtaining external knowledge corresponding to one correctly answered question (i.e., knowledge inflow moving from sample mean, 10.71 points, to 20.71 points) increases added value from $2,654.47 million to $2,667.77 million, a $13.30 million increase. For comparison, using data from an earlier sample period, Hitt et al. (2002) show that firms who adopt SAP ERP systems experience an increase of between 2.7

percent and 1.7 percent in productivity. During a later period (1998-2005) that partially overlaps with our sample, Aral et al. (2006) find that ERP adoption was associated with a 6.9% increase in productivity. The average firm in their sample had sales of \$8.466 billion and cost of goods sold of \$5.774 billion, or a value added of \$2.692 billion. Thus, ERP adoption translated into an increase in value-added of \$185.7 million.

It is instructive to reflect on what these calculations mean. It is commonly argued that investments in enterprise systems reflect both the value of the systems and that of the complementary business process innovation required to make the system useful (e.g., Bresnahan et al. 2002, Bresnahan and Greenstein 1996). The results of these prior studies on the productivity benefits of ERP systems reflect both the value added to the system itself and that of the complementary efforts, as measured by the incidence of ERP investments.

In contrast, our estimates reflect the value of complementary business process innovation (and possibly some incremental investments in new (unmeasured) enterprise system modules), as measured by the acquisition of external knowledge among firms who have already invested in ERP systems. Because they reflect incremental innovation that extends the value of existing ERP systems, the value-added in our study is lower than that of prior work. However, because they may be capturing business process innovation that extends beyond the effects of the activity on the forum, the value-added estimates may be larger than that experienced from the answer to any one particular question. For example, the sample discussion thread in Figure 2 involves a query about material exclusion in the sales order process. This could represent part of a broader effort to reconfigure the sales order process within SAP. Because it is an extension or change to an existing implementation of SAP, the productivity benefits of this extension are likely to be smaller than that of an incidence of an SAP implementation. However, because it may capture a broader effort at improving the sales order process, in our analyses the value to the firm may appear larger than the narrow benefits of having that particular question answered.

We also note that out estimate of the output elasticity of IT capital (0.010), is comparable to that in Tambe and Hitt (2014b) (0.027, p.64), or in Tambe and Hitt (2012) (0.032, p.609), and the differences may be explained by the different sample we use, which consists of ERP adopters with the majority in the manufacturing industry.

5.2 Robustness of Baseline Analyses

As noted above, the estimates associated with the effect of our IT spillover variable may reflect the effects of a bundle of human capital accumulation efforts related to adoption and installation of ERP systems. This conjecture is consistent with other work researching IT and productivity, where IT investments and organizational practices pick up a range of activities around business process innovation (Bloom et al. 2012, Bresnahan et al. 2002). To study how this type of activity might be influencing our results, in columns 3-8 of Table 5 we examine a range of robustness checks aimed at controlling for SAP-related activity within the firm.

We first control for heterogeneity in SAP activity by interacting the number of pre-sample SAP modules of the focal firm, obtained from the SAP installation data described above, with a set of time dummies in column 3. This variable will pick up the intensity of SAP activity during our sample if unobserved investments in SAP during the sample period are correlated with prior SAP investments. The results in column 3 of Table 5 show that the estimated effect of knowledge flows is little changed in comparison to the baseline results in column 1.

As an additional effort to proxy for other kinds of firm activity, in columns 4-6 we add controls for firm activity on the forum – these activities on the forum will be correlated with SAP-related activity, however because they do not indicate answers to questions they will not specifically pick up human capital deepening within the firm. In column 4 we show results that add a control for the cumulative number of registered users who are the focal firm's employees in the SCN, and in Column 5 we present model results that explicitly include the total number of questions that are raised by a firm's employees (recall that a question is usually the first message that initiates a discussion thread) in a year – regardless of whether the question was solved or received helpful answers – as a control. Interestingly, the number of questions is negatively correlated with productivity; this may reflect unobserved human capital differences within the firm – specifically, firms with lower level of human capital may ask more questions (Di Maggio and Van Alstyne 2013). In Column 6 we present a model that incorporates both the number of registered users and number of questions as controls. In all cases, the coefficient of our inward spillover variable remains positive and economically and statistically significant. Thus, even when controlling for other types of activity in the network, it appears that the IT spillovers are positively correlated with productivity. In sum, this set of results suggests that spillovers are capturing something beyond the value of new investments in SAP modules and systems.

To increase the confidence in our interpretation of the regression results, we also conduct a falsification test where we examine whether the timing of the effects of the spillovers are appropriate. If the benefits of spillovers appear before they should – in other words, prior to a firm actually receiving knowledge flows – then this suggests our results are affected by omitted variable bias. In Column 7 of Table 5 we report a model in which we add a variable that is equal to the spillovers in the next period. If our interpretation is correct, we expect that the future value of spillovers should have no effect on the productivity of the current year. This is indeed the result that we find. To rule out the possibility that the insignificant future value of spillovers is driven by serial correlation in the spillover measure, in column 8 of Table 5 we also present a model where we only include the future spillover as the explanatory variable, without adding the current period spillovers.[22] Again we find future value of spillovers does not have a significant effect on productivity. While any conclusion that we can draw from this model is tentative due to the short length of our panel, these results provide an additional piece of evidence in support of our interpretation of the results.

In our section on identification we discussed how the size of knowledge inflows were based upon the number of questions asked and the likelihood of answering a question. Our regressions that include the number of questions controls for omitted factors that could be correlated both with the propensity to ask questions as well as productivity, however it remains possible that our estimates of knowledge inflows may be influenced by omitted factors that influence the likelihood of responses and that may be correlated with productivity. For example, workers with greater IT skills may have more reputation in the community and may be better able to describe the questions, leading to a higher likelihood that their questions be answered. To address this, we instrument for knowledge inflows using activity within the forums that the firm participates in. Specifically, for each question asked by the focal firm we compute the predicted number of responses and the predicted number of correct responses based on regressions where the predictors are forum fixed effects, year fixed effects, the number of questions posted in the focal forum and year, the number of users in the focal forum and year, the average number of replies per question in the focal forum and year, the average number of views in the focal forum and year, and the average solution rate in the focal forum and year. All of these forum-year variables exclude the focal question. We then sum these predicted values across all of the questions asked by the firm in the year, and use this as an instrument for knowledge inflows. Table 6 presents the regression results using each of these instruments separately; we do not include them together because the instruments are highly correlated with one another. In all cases, the Angrist-Pischke first stage F-tests of exclusion restrictions reject the null, suggesting the instruments are not weak. The Stock-Yogo critical values, reported for 10% maximal IV size, further confirm the validity of the instruments. The second stage results are qualitatively similar and in the same direction as in our baseline estimates but larger; the larger estimates may reflect the instruments removing measurement error or a local average treatment effect.

[Insert Table 6 Here]

In Appendix 1 we examine the results of additional robustness tests. We study the robustness of our findings to different ways of computing spillovers from the SCN forum. Prior research has studied the

[22] We thank an anonymous reviewer for suggesting this analysis.

effects of spillovers based upon pools of IT investment from firms in the same industry or through supply chain interactions (e.g., Cheng and Nault 2007, 2012). We study whether our results reflect spillovers from firms in the same industry by including a spillover pool in our regressions. To study whether our results are driven by a small number of "superstar" individuals who are very active in the forum, we rerun our regressions excluding from the spillover term knowledge inflows from questions asked by these individuals. Some prior research has emphasized how answering questions can lead to human capital deepening (e.g., Di Maggio and Van Alstyne 2013), so we further examine the robustness of our results to controlling for a measure of same-period knowledge outflows from firm employees. Our results are robust to all of these changes, though the magnitude of IT spillovers effect is reduced in some cases. Overall, these results provide further support for the idea that our IT spillover measure reflects, to some extent, a firm's effort to deepen its human capital. To study the robustness of our results to different assumptions about omitted variables, in Appendix 2 we discuss the results of GMM-based dynamic panel data models that treat knowledge inflows, IT capital, and lagged value-added as endogenous variables.

5.3 Test of Absorptive Capacity Model and Hypotheses

In this section we show how the value of IT spillovers will depend on the firm's absorptive capacity. We present the results from the full model (equation 8) in Table 7. We use these regressions to show that, consistent with the theory of absorptive capacity, the value of external knowledge flows depends upon prior enterprise software investments and the difficulty of learning.

[Insert Table 7 about here]

Results in column 1, using the Advanced SAP module variable as the measure of prior investments, form our baseline results. We use this measure both because it signals prior use of enterprise systems and because this measure is fixed at pre-sample values and so will not be influenced by changes in unobserved firm-level variables during our sample period. As noted above, we use two variables to measure of difficulty of learning. In panel A of Table 7 (columns 1-2) we use the percentage of business-related knowledge flows as the measure of difficulty of learning, while in panel B (columns 3-4) difficulty of learning is measured by the percentage of knowledge flows from newly introduced forums.

First, we run a specification test comparing the absorptive capacity model as specified in equation (8) to the baseline model with homogeneous absorptive capacity (equation 9). This amounts to a test of the null hypothesis that

$$\gamma_1' = 0, \gamma_2' = 0, \gamma_3' = 0.$$

We obtain an F-statistic equal to 12.12, rejecting the null at the 1% significance level. This result suggests that the ACAP model overall is the preferred specification.

Because regression equation (8) involves 3-way interactions, the testing of H1 and H2 cannot be directly performed by examining the coefficient estimates of $\gamma_1{}'$ and $\gamma_2{}'$. As we explained earlier, testing of H1 is equivalent to $(\gamma_1' + \gamma_3{}'d_{it}) > 0$ when d_{it} is at the sample mean. We calculate this linear combination which has a value of (0.0153 + 0.0041*(-0.124)) = 0.0148 (p<0.01), which is supportive of H1.

Consistent with the theory of absorptive capacity, we observe that prior related investment in enterprise software serves dual purposes: beyond its direct contribution to productivity, it also contributes to productivity indirectly by enhancing a firm's IT-related absorptive capacity, thereby allowing the firm to identify and exploit external knowledge. To quantify the economic implications of absorptive capacity for the value of knowledge inflows, we compute the output elasticities when the firm has made large prior SAP investments versus those when the firm has not made such investments, while holding the difficulty of learning at the sample mean. These are reported in Table 8(a). In the former case ($m_{it} = 1$), a one percent increase in the amount of inward knowledge flows is associated with a 0.007 percent increase in the added value produced by the firm (p<0.1). In contrast, firms with low prior enterprise software investments ($m_{it} = 0$) actually see negative value from IT spillovers (elasticity=-0.007, p<0.01). The difference between the two output elasticities is 0.014 and highly statistically significant (p<0.01), which suggests that the indirect contribution of enterprise software investment cannot be ignored when opportunities of inter-firm spillover are present. Failure to consider the dual effects of enterprise IT investments will lead to underestimation of their true returns.

[Insert Table 8 about here]

We next examine the effects of the characteristics of knowledge. A test of H2 is equivalent to testing $(\gamma_2' + \gamma_3{}'m_{it}) < 0$ when m_{it} is at the sample mean. We calculate this linear combination which has a value of -0.0055+0.0041*0.814=-0.0022 (p=0.13). H2, which states that the effects of inward knowledge flows on productivity will be less when those knowledge flows are more difficult to learn, is consistent with our data, but slightly short of significance at conventional levels. Because of the role of prior enterprise software investments (m_{it}) in the expression, the effects of difficulty will depend upon the presence of prior enterprise software investments. In particular, among firms with low prior enterprise software investments ($m_{it} = 0$), the test of H2 will degenerate to $\gamma_2' = -0.006$ ($p < 0.01$).

To quantify how difficulty of knowledge influences the economic impact of knowledge inflows, we compare the elasticity of inward knowledge flows when the knowledge flow is mainly business-related (d_{it} =3rd quartile) to that when it is mainly technical-related (d_{it} =1st quartile),[23] while holding the prior SAP investments at the sample mean. We report the results in Table 9(a). In the former case, inward knowledge flows have a combined effect (coefficient) of 0.005 (n.s.) while in the latter case inward knowledge flows have a positive marginal effect of 0.011 ($p<0.01$). The difference between the elasticities under the two cases is negative (-0.006, p=0.13), which suggests that the return of external knowledge flows is decreasing in the difficulty of learning external knowledge. However, the difference is not statistically significant, again owing to the moderating effects of prior knowledge. When evaluated at the point where firms have low enterprise software investments ($m_{it} = 0$), there is a -0.018 difference in the output elasticity of knowledge flows between knowledge flows that are mainly business-related and those that are mainly technical-related ($p<0.01$).

[Insert Table 9 about here]

Lastly, we assess the interaction effect of prior investments in enterprise software and difficulty of learning on the returns of knowledge inflows. The test of H3 can be performed directly by examining $\gamma_3' > 0$ in regression (8). We observe a positive and significant coefficient estimate of the three-way interaction s_{it}*d_{it}*m_{it} (γ_3'=0.004, $p<0.05$), supporting H3. That is, prior IT investments play a greater role when external knowledge is difficult to learn.

Table 10a provides further documentation of the economic implications of prior enterprise software investments and difficulty of knowledge learning by calculating the output elasticities w.r.t. knowledge inflows under 4 scenarios: {high prior IT investments, low prior IT investments} X {high difficulty of learning, low difficulty of learning}. For difficulty of learning, we compute the first and third quartiles of d_{it} among the observations that have nonzero knowledge inflows. When prior enterprise software investments are low and most knowledge is business functional, the output elasticity of external knowledge inflows are actually negative and significant (-0.008, $p<0.01$). In contrast, when prior investments are high and most inflows are technical, then knowledge inflows have a positive output elasticity (0.012, $p<0.05$). The significant column difference of the row differences, presented in the bottom-right cell (0.013, $p<0.05$), further confirms that the marginal effect of prior IT investment on the value of spillover is higher when external knowledge is difficulty to learn (0.015, or first row difference) than when external knowledge is easy to learn (0.002, or second row difference).

[Insert Table 10 about here]

[23] For the sample firms that have nonzero knowledge inflows, the 1st quartile corresponds to a scenario where 4% of knowledge flows are business related, while the 3rd quartile has 93% business related knowledge.

Our second measure of prior cumulated investments in enterprise software is a binary indicator that is set to 1 if the firm's cumulative contribution to SCN prior to year t is greater than sample mean. In column 2 of Table 7 we present results of estimation equation (8) using this variable, which appear very similar to those using prior investments in SAP applications. With regard to H1, the linear combination $(\gamma_1' + \gamma_3' d_{it})$ when d_{it} is at the sample mean is 0.013 ($p<0.1$), a result which is supportive of H1 and consistent with what we found when using SAP modules as the measure of prior IT investments. In table 8(b) we present computations of output elasticities evaluated at different values of m_{it} while holding d_{it} at the mean levels, to show the contrast between a firm that has made prior IT investments versus one that hasn't.

We find support for H2. Holding m_{it} at the sample mean, the linear combination $(\gamma_2' + \gamma_3' m_{it})$ is negative and significant (-0.005, $p<0.05$). The calculations of marginal effects, presented in Table 9(b), compare a firm that received difficult-to-learn knowledge spillovers versus one that received easy-to-learn knowledge spillovers, and provide further evidence of the economic implications of difficulty on knowledge absorption.

Finally, the three-way interaction $s_{it}*d_{it}*m_{it}$ is positive and highly significant (γ_3'=0.006, $p<0.01$), supporting H3 that prior related investments are more critical when the knowledge obtained from external sources is difficult to learn. Calculations of the marginal effects of knowledge flows, presented in Table 10(b), provide further evidence that absorptive capacity shapes how knowledge inflows contribute to firm productivity.

We repeat these tests using an alternative measure of the difficulty of learning – the percentage of knowledge flows obtained from SCN forums that are newly introduced, and present the results in Panel B of Table 7 (columns 3-4). The direction and significant of the results is similar to that in columns 1-2, with the exception of the 3-way interaction $s_{it}*d_{it}*m_{it}$ in column 3, which is slightly short of significance at conventional levels (γ_3'=0.002, p=0.15). Overall, our empirical tests of the full model lend support to our hypotheses as predicted by the theory of absorptive capacity as applied to IT investments (Cohen and Levinthal 1989).

5.4 Robustness of the Absorptive Capacity Models

As we discussed earlier, one of the major identification challenge we face is that our knowledge flows measure may pick up unobserved incremental IT investments such as deployment of new enterprise software modules or investments in on-job training that are correlated with the propensity of acquiring external knowledge. However, if these unobserved factors are responsible for the results in Table 7, then they must behave in such a way that they give rise to similar interaction patterns with prior related knowledge and the difficulty of learning as predicted by our absorptive capacity model. In Table 11 we present a set of falsification tests that provide additional evidence that supports the interpretation that our results reflect the interplay between prior IT investments, external knowledge flows, and productivity.

Particularly, if a firm makes incremental investments related to enterprise systems such as implementation of new modules or enhancing human capital through internal training, these investments will be reflected in the IT capital variable (recall that our definition of IT capital includes both IT hardware and IT labor). In column 2 of Table 10 we present a model where we add the interaction between IT capital and the variables related to absorptive capacity – prior related knowledge and the difficulty of learning. For comparison, we reproduce the result from column 1 of Table 7 and put it in column 1 of Table 10. We find that the interaction patterns involving IT capital in column 2 are significantly different from those in column 1. In fact, the direction of the interactions $c*d$ and $c*d*m$ are completely opposite to $s*d$ and $s*d*m$; these results suggest that the effect of unobserved incremental IT investments do not behave in conformance with the predictions of the absorptive capacity framework. In column 3 we present a model where we include the sets of interaction terms that involve both knowledge flows and IT capital, and again the results show that only external knowledge acquisition behave in ways that are consistent with the predictions of the absorptive capacity theory. In sum, these results do not support the alternative hypothesis that our results are caused by unobserved enterprise software investments that are correlated with knowledge inflows.

[Insert Table 11 about here]

Another alternative interpretation of our results regarding the effect of IT spillovers, s, is that they reflect the implications of a broader set of human capital acquisition activities at the firm level. Consistent with this interpretation, our absorptive capacity model can be modified in a way that we no longer attempt to separately identify the effects of M_{it} and S_{it} and instead combine these two terms to reflect the aggregate knowledge stock related to human IT capital investments (Z_{it} in our earlier model). In this case, Z_{it} is a function of both internal IT human capital investments, M_{it}, and knowledge obtained from outside the boundary of the firm, S_{it} (e.g., Chang and Gurbaxani 2012a, Griffith et al. 2006), with our spillovers measure representing a proxy for Z_{it}. Under this alternative model and interpretation, our earlier estimation strategy continues to hold, with a slight change in interpretation of the results. Namely, prior IT investments are associated with greater success in accumulating both internal human capital and assimilating external knowledge for productive use ($\frac{\partial \gamma_{it}}{\partial M_{it}} > 0$). Further, we expect that human capital accumulation has a larger impact on productivity if a large fraction of the knowledge is difficult to learn

($\frac{\partial \gamma_{it}}{\partial D_{it}} < 0$). In addition, prior IT investments are more valuable to absorbing human capital when they are more difficult to learn ($\frac{\partial^2 \gamma_{it}}{\partial M_{it} \partial D_{it}} > 0$).[24]

We performed a series of additional robustness tests of the absorptive capacity model, and present these results in Appendix 3. First, we test the same model specifications while controlling for the number of questions raised by a firm's employees and the number of employees that participate in SCN, and our findings are consistent with those presented in Table 7. Second, we also evaluate the robustness of our findings to the alternative measures of knowledge inflows, to the exclusion of knowledge inflows from superstar employees, and to controlling for the effect of spillover pools.

6. Conclusions

This paper shows that knowledge flows related to the implementation and use of enterprise software are associated with a substantial rise in firm productivity. However, our work suggest that firms are heterogeneous in their abilities of exploiting external knowledge for productive use - the effects of access to external knowledge are critically conditioned by a firm's prior IT investments and by the nature of external IT knowledge flows. These results are consistent with a well-established R&D literature based on the concept of absorptive capacity.

We adopt a novel measurement strategy that allows us to exploit a rich dataset based on activity taking place in an online discussion forum, an increasingly used channel that firms use to augment the human capital necessary to deploy IT systems. By combining a novel data source with an established theoretical framework, we show that the effect of external knowledge flows is stronger for firms with prior investments in enterprise software, especially those who deployed advanced software applications related to business processes. We find that absorptive capacity is lower when external IT knowledge is difficult to learn, such as for business functional knowledge or knowledge originating from relatively newer and emerging discussion forums. However, it is precisely in these environments in which prior investments in enterprise software have their most significant impact on facilitating the absorption of knowledge.

We contribute to an existing literature on IT spillovers by highlighting the importance of the dual roles of IT investments for a firm's productivity. While IT hardware, software and labor have a direct effect on productivity, prior investment in enterprise software also contributes indirectly to productivity by enhancing a firm's IT-related absorptive capacity, which in turn increases the value of external

[24] A formal write-up of this model is available upon request.

knowledge flows when opportunities for spillover are present. Such indirect effects of IT investments on productivity are more pronounced when external knowledge is more difficult to learn.

These results have several implications for firms. For one, firms that fail to account for these indirect effects will underestimate the productivity implications of their IT investments. Further, there are broader implications for firms given the shift away from on-premises computing to cloud computing. Historically, investments in applications software came bundled with investments in how to deploy the systems. That is, firms deploying enterprise software were required to make complementary investments in business process innovation. As firms increasingly deploy service-based application software that may require smaller investments to deploy, this may influence their ability to respond to new enterprise IT-based opportunities in the future. This shares some similarities with earlier concerns about whether 'offshoring' software development would lead to a hollowing out of the labor force in the US (e.g., Levy and Murnane 2005).

Further, our results provide insights for why firms participate in online communities such as SCN. An existing body of work has argued that workers contribute to open source projects to develop their skills (e.g., Lakhani and von Hippel 2003, Lakhani and Wolf 2005) and more recent work has argued that firms may provide incentives for workers to contribute to such projects to develop the accumulated human capital within the firm (Mehra et al. 2011). However, our work provides evidence of the benefits of participation not only through inflows but also through contributions because of the implications for absorptive capacity.

While our research contributes to the literature on IT spillovers, it is worthwhile noting that the process by which 'spillovers' are generated in our context vary significantly. In contrast to much work on the IT spillovers (e.g., Cheng and Nault 2007, 2012; Tambe and Hitt 2014a, 2014b) and the traditional R&D literature on absorptive capacity (e.g., Cohen and Levinthal 1989, 1990), knowledge flows in our setting are not externalities arising from investments in product or business process innovation from firms in the same industry, supply chain, or network. Instead, they arise from deliberate decisions on the part of firms to ask and answer questions. In that way, they bear some similarity to the nature of knowledge flows arising from the transactional relationship between an IT services provider and its clients (Chang and Gurbaxani (2012a, b). Given that the nature of the knowledge transferred is likely already customized to the firm's needs, the continued importance of absorptive capacity is striking.

Our work also contributes to prior research that has sought to understand the interrelationships between IT investment, business process innovation, and productivity.[25] The literature on business

[25] See, for example, Bloom et al. (2009), Bresnahan and Greenstein (1996), Bresnahan, Brynjolfsson, and Hitt (2002), Dranove et al. (2014), Forman et al. (2005), Ichniowski and Shaw (2003), and Bartel et al. (2007).

process innovation has been hampered by a number of challenges, namely the difficulty of measuring the inputs and outputs of the innovation process. While measurement of innovation is always problematic (Cohen 2010, Mortensen and Bloch 2005), measurement of business process innovation is particularly difficult because it leaves behind no tangible 'footprints' such as patents. Our work provides further insights into the role of external knowledge flows in augmenting internal human capital through a unique measurement strategy that uses behavior online to capture inputs into business process innovation that could not be previously measured directly. While we acknowledge that our measures may not capture all such human capital accumulation, they are in the spirit of the literature on business process innovation that uses proxies for hard-to-measure inputs and outputs and acknowledges output elasticities may capture variance related to some unmeasured activity (e.g., Anderson et al. 2003, Bresnahan et al. 2002). More broadly, our research also adds to an existing body of research in the IS literature that has sought to understand the role of absorptive capacity in appropriating returns from IT investments (Roberts et al. 2012). While much of this literature emphasizes the role of path dependence, our measurement strategy for knowledge flows allows us an unusual opportunity to measure the interrelated effects of knowledge spillovers, prior related investments, and the type of knowledge.

As in any study seeking to measure the productivity implications of IT investments, business process innovation, and human capital acquisition, our study is subject to limitations. One advantage of our study over prior work is our ability to measure knowledge flows directly. However, as has been noted elsewhere in the paper, our estimation strategy and robustness are shaped by the unique data-generating process of knowledge acquisition in our setting which involves endogenous choices to raise and answer questions. Further, as in any study such as ours, we must be cognizant of whether organizational investments such as external knowledge acquisition are correlated with other unobserved variables. To address these concerns, we examined the robustness of our results to falsification exercises, instrumental variable estimates, and a range of measurement strategies. While our results are robust to these efforts, we leave it to future work to study the robustness of our results to other contexts.

Our research highlights opportunities for new research. One interesting possibility would be to investigate at a more disaggregate level how related human capital investments influence the benefits that individual workers experience from participation in related communities (Huang and Zhang). This could be accomplished by tying community activity to databases of worker skills, using data from sites such as LinkedIn. Another possibility would be to measure the firm-level benefits of worker participation in the production of collective action goods such as open source software. We hope our research spurs additional work in these important areas.

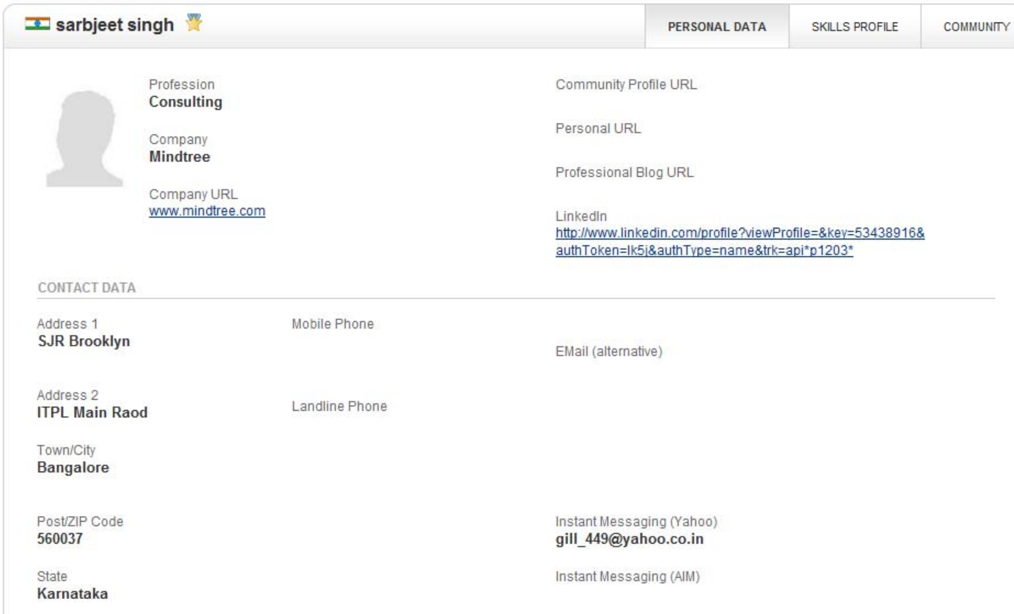

Figure 1: Sample User Profile

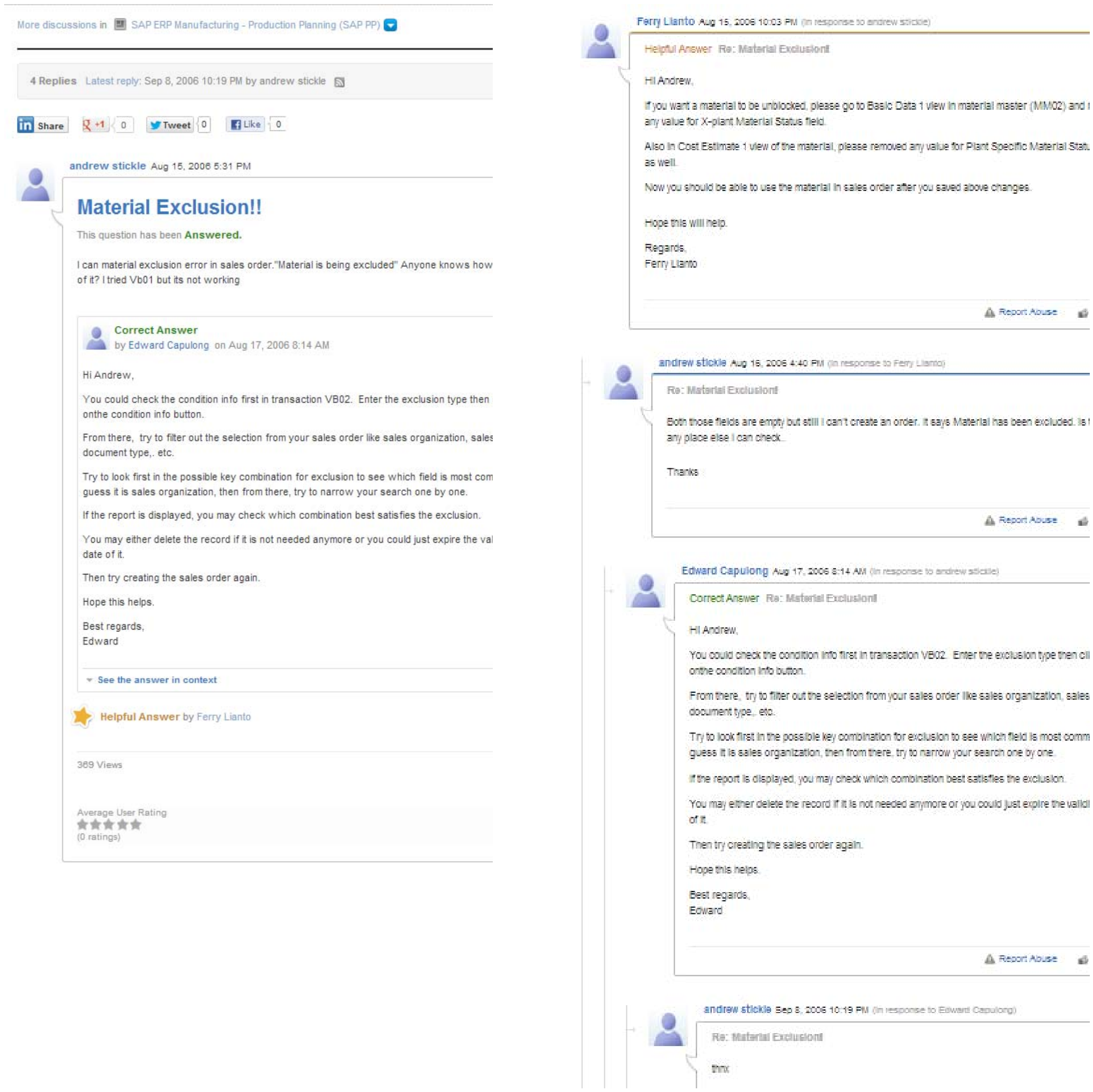

Figure 2: Sample Discussion Thread

Table 1: Evolution of SAP Community Network

Year	Number of registered users	Number of active forums	Number of new threads initiated in the year	Average number of replies for threads initiated in the year	Fraction of questions solved	Fraction received helpful answers	Fraction received very helpful answers	Number of days until correct answer
2004	19,289	57	16,296	4.679	0.107	0.073	0.098	13.378
2005	43,226	83	67,225	5.394	0.244	0.271	0.295	4.735
2006	80,981	141	176,422	5.160	0.242	0.293	0.314	3.359
2007	137,552	179	394,183	4.731	0.227	0.260	0.287	4.219
2008	198,975	209	463,740	4.625	0.252	0.255	0.256	4.512

Table 2: Summary Statistics

Variable	Mean	Std. dev	Min	Max
Annual sales (million $)	16649.51	33024.88	298.91	364392.40
Added value (million $)	5491.17	8811.34	118.11	73242.29
Non-IT capital (million $)	12526.01	29661.25	48.44	321772.70
IT capital (million $)	97.49	138.29	0.00	1181.67
Non-IT labor (million $)	2781.93	4405.87	28.75	40586.13
No. of employees (thousands)	41.17	59.59	0.66	428
Knowledge inflows (reward points)	10.71	93.82	0	2190
Difficulty of learning (pct. of knowledge related to business)*	0.43	0.40	0.00	1
Difficulty of learning (pct. of knowledge from new forums)*	0.11	0.21	0.00	1
Advanced SAP Business Process Suites (binary)	0.81	0.39	0	1
High Related Knowledge in Human Capital (binary)	0.10	0.30	0	1
Number of SCN users	2.94	6.52	0	97
Number of questions	3.58	12.45	0	220

Number of observations: 1240. Number of firms: 275.
*: Summary stats for Difficulty are based on observations with nonzero knowledge flows.

Table 3: Pearson Correlation Matrix of Selected Variables

	Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1	Annual sales	1.00												
		(-)												
2	Added value	0.86	1.00											
		(0.00)	(-)											
3	Non-IT capital	0.83	0.77	1.00										
		(0.00)	(0.00)	(-)										
4	IT capital	0.34	0.48	0.33	1.00									
		(0.00)	(0.00)	(0.00)	(-)									
5	Non-IT labor	0.55	0.81	0.37	0.58	1.00								
		(0.00)	(0.00)	(0.00)	(0.00)	(-)								
6	No. of employees	0.52	0.76	0.37	0.56	0.93	1.00							
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(-)							
7	IT spillover	0.02	0.04	0.00	0.00	0.06	0.03	1.00						
		(0.56)	(0.12)	(0.96)	(0.94)	(0.04)	(0.22)	(-)						
8	Difficulty (business)	-0.10	-0.16	-0.04	-0.06	-0.18	-0.12	-0.37	1.00					
		(0.00)	(0.00)	(0.14)	(0.03)	(0.00)	(0.00)	(0.00)	(-)					
9	Difficulty (new forums)	-0.09	-0.15	-0.04	-0.05	-0.17	-0.11	-0.41	0.80	1.00				
		(0.00)	(0.00)	(0.17)	(0.11)	(0.00)	(0.00)	(0.00)	(0.00)	(-)				
10	Adv. SAP Suites	0.07	0.08	0.00	0.05	0.07	0.07	0.04	-0.09	-0.10	1.00			
		(0.02)	(0.00)	(0.88)	(0.08)	(0.02)	(0.02)	(0.12)	(0.00)	(0.00)	(-)			
11	High Human Capital	0.10	0.13	-0.01	0.02	0.16	0.13	0.20	-0.31	-0.34	0.08	1.00		
		(0.00)	(0.00)	(0.75)	(0.43)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(-)		
12	Users	0.20	0.32	0.10	0.12	0.37	0.28	0.18	-0.30	-0.40	0.15	0.43	1.00	
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(-)	
13	Questions	0.11	0.21	0.03	0.06	0.26	0.19	0.81	-0.52	-0.59	0.10	0.39	0.56	1.00
		(0.00)	(0.00)	(0.26)	(0.04)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(-)

Significance levels are shown in the parentheses.

Table 4: Industry Segments of the Sample

2-digit NAICS	Description	Freq.	%
11	Agriculture, Forestry, Fishing and Hunting	5	0.4
21	Mining, Quarrying, and Oil and Gas Extraction	25	2.02
22	Utilities	104	8.39
23	Construction	8	0.65
31-33	Manufacturing	824	66.45
42	Wholesale Trade	51	4.11
44-45	Retail Trade	48	3.87
48-49	Transportation and Warehousing	17	1.37
51	Information	71	5.73
52	Finance and Insurance	12	0.97
53	Real Estate and Rental and Leasing	10	0.81
54	Professional, Scientific, and Technical Services	31	2.5
56	Administrative and Support and Waste Management and Remediation Services	10	0.81
62	Health Care and Social Assistance	9	0.73
72	Accommodation and Food Services	15	1.21
Total		1,240	100

Table 5: Baseline Spillover Models

Variables	(1) Fixed effects	(2) Random effects	(3) Include (# of SAP modules) X (year dummies)	(4) Includes # of users	(5) Includes # of questions	(6) Includes # of users and # of questions	(7) Spillover timing	(8)
Non-IT capital	0.109**	0.241***	0.107**	0.110**	0.110**	0.110**	0.083*	0.082*
	(0.043)	(0.017)	(0.043)	(0.043)	(0.043)	(0.043)	(0.048)	(0.048)
IT capital	0.010*	0.009*	0.010*	0.010*	0.011*	0.010*	0.009	0.009
	(0.006)	(0.015)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Non-IT labor	0.726***	0.683***	0.725***	0.727***	0.729***	0.728***	0.745***	0.746***
	(0.059)	(0.018)	(0.059)	(0.059)	(0.059)	(0.059)	(0.061)	(0.061)
Knowledge flows	0.009**	0.011**	0.009**	0.011***	0.015***	0.014**	0.008**	
	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)	(0.006)	(0.003)	
Knowledge flows (t+1)							0.004	0.003
							(0.003)	(0.003)
Log(registered users)				-0.020		-0.011		
				(0.013)		(0.017)		
Log(questions)					-0.015*	-0.009		
					(0.009)	(0.010)		
Constant	1.726***		1.749***	1.683***	1.659***	1.670***	1.762***	1.766***
	(0.452)		(0.455)	(0.456)	(0.457)	(0.456)	(0.477)	(0.476)
Observations	1,240	1,240	1,240	1,240	1,240	1,240	1,206	1,206
Number of firms	275	275	275	275	275	275	268	268
R-squared	0.575	0.569	0.576	0.577	0.577	0.577	0.580	0.578

The dependent variable is the natural logarithm of Value Added. All models except Column (2) use firm-level fixed effects and year dummies. Column (2) estimates include 2-digit NAICS dummies. Robust standard errors (clustered by firm) in parentheses except for Column 2. All R-square values are "within" estimates that do not include the explanatory power of the fixed effects.
*** p<0.01, ** p<0.05, * p<0.1

Table 6: Instrumenting for Knowledge Inflows

	Predicted Number of Correct Answers		Predicted Number of Replies		Predicted Number of Points	
	(1)	(2)	(3)	(4)	(5)	(6)
Variables	First Stage	Second Stage	First Stage	Second Stage	First Stage	Second Stage
Non-IT capital	0.051	0.111***	0.045	0.110**	0.052	0.110**
	(0.124)	(0.037)	(0.123)	(0.037)	(0.123)	(0.037)
IT capital	-0.073**	0.012**	-0.072*	0.011**	-0.070**	0.011**
	(0.034)	(0.005)	(0.034)	(0.005)	(0.034)	(0.005)
Non-IT labor	-0.168	0.731***	-0.155	0.731***	-0.163	0.730***
	(0.124)	(0.050)	(0.123)	(0.050)	(0.124)	(0.050)
Knowledge flows		0.031**		0.028**		0.027**
		(0.013)		(0.012)		(0.012)
Log(users)	-0.543***	0.002	-0.518***	0.0004	-0.524***	-0.0005
	(0.100)	(0.018)	(0.099)	(0.981)	(0.099)	(0.017)
Log(questions)	0.761***	-0.029	0.704***	-0.026	0.717***	-0.025
	(0.113)	(0.018)	(0.109)	(0.017)	(0.110)	(0.017)
Excluded instrument	0.138***		0.008***		0.007***	
	(0.033)		(0.001)		(0.001)	
Observations	1,227	1,227	1,227	1,227	1,227	1,227
Number of firms	262	262	262	262	262	262
F-statistic	17.37		30.78		25.16	
Stock and Yogo critical values	16.38		16.38		16.38	
R-squared	0.491	0.573	0.498	0.574	0.500	0.575

The dependent variable is the natural logarithm of Value Added. All models use firm-level fixed effects and year dummies. Robust standard errors (clustered by firm) in parentheses. All R-square values are “within” estimates that do not include the explanatory power of the fixed effects.
13 observations were dropped in the IV regressions due to singletons.
*** p<0.01, ** p<0.05, * p<0.1. Stock and Yogo critical values are reported for 10% maximal IV size.

Table 7: Test of Absorptive Capacity Hypotheses

Variables	Panel A: d_{it} = Percentage of Business Knowledge Inflows		Panel B: d_{it}=Percentage of Knowledge Inflows from New Forums	
	(1)	(2)	(3)	(4)
	m_{it}=advanced SAP module	m_{it}= high human capital	m_{it}=advanced SAP module	m_{it}= high human capital
Non-IT capital	0.110**	0.104**	0.110**	0.106**
	(0.043)	(0.044)	(0.043)	(0.043)
IT capital	0.010*	0.010*	0.010*	0.009
	(0.006)	(0.006)	(0.006)	(0.006)
Non-IT labor	0.726***	0.729***	0.725***	0.726***
	(0.059)	(0.056)	(0.059)	(0.056)
Knowledge flows (s_{it})	-0.008***	-0.004	-0.011**	-0.008
	(0.003)	(0.004)	(0.004)	(0.006)
Prior related investment (m_{it})	--	0.055*	--	0.054*
		(0.028)		(0.028)
$s_{it}*m_{it}$	0.015***	0.014*	0.017**	0.019**
	(0.005)	(0.008)	(0.007)	(0.008)
$s_{it}*d_{it}$ (difficulty of learning)	-0.006***	-0.005**	-0.003***	-0.003**
	(0.001)	(0.002)	(0.001)	(0.002)
$s_{it} * d_{it} * m_{it}$	0.004**	0.006***	0.002	0.004*
	(0.002)	(0.002)	(0.002)	(0.002)
Constant	1.715***	1.748***	1.720***	1.752***
	(0.454)	(0.443)	(0.453)	(0.442)
Observations	1,240	1,240	1,240	1,240
Number of firms	275	275	275	275
R-squared	0.576	0.580	0.576	0.579

The dependent variable is the natural logarithm of Value Added. All models use firm-level fixed effects and year dummies. Robust standard errors in parentheses. All R-square values are “within” estimates that do not include the explanatory power of the fixed effects.
*** p<0.01, ** p<0.05, * p<0.10

Table 8: The Impact of Prior IT Investments on the Relationship between External Knowledge and Productivity

(a) m_{it}=advanced SAP modules

	High SAP investments (m_{it}=1)	Low SAP investments (m_{it}=0)	Difference
(d_{it} =mean)	0.007*	-0.007***	0.014***

(b) m_{it}=high human capital

	High human capital investments (m_{it}=1)	Low human capital investments (m_{it}=0)	Difference
(d_{it} =mean)	0.010	-0.003	0.013*

Notes: Output elasticity of knowledge inflow is reported. *** p<0.01, ** p<0.05, * p<0.10.

Table 9: The Impact of Difficulty of Learning on the Relationship between External Knowledge and Productivity

(a) m_{it}=advanced SAP modules

	High difficulty (d_{it}=3rd quartile)	Low difficulty (d_{it}=1st quartile)	Difference
(m_{it} =mean)	0.005	0.011***	-0.006 (p=0.13)

(b) m_{it}=high human capital

	High difficulty (d_{it}=3rd quartile)	Low difficulty (d_{it}=1st quartile)	Difference
(m_{it} =mean)	-0.003	0.013**	-0.016**

Notes: Output elasticity of knowledge inflow is reported. *** p<0.01, ** p<0.05, * p<0.10.

Table 10: Interaction of Prior IT Investments and Difficulty of Learning

(a) m_{it}=advanced SAP module

	High SAP investments (m_{it}=1)	**Low SAP investments (m_{it}=0)**	**Row diff.**
Difficulty=high (d_{it} =3rd quartile)	0.007	-0.008***	0.015***
Difficulty=low (d_{it} =1st quartile)	0.012**	0.010***	0.002
Column diff.	-0.005	-0.018***	0.013**

(b) m_{it}=high human capital

	High human capital investments (m_{it}=1)	**Low human capital investments (m_{it}=0)**	**Row diff.**
Difficulty=high (d_{it} =3rd quartile)	0.010	-0.004	0.014*
Difficulty=low (d_{it} =1st quartile)	0.007	0.013**	-0.006
Column diff.	0.003	-0.017**	0.020***

Notes: Output elasticity of knowledge inflow is reported. *** p<0.01, ** p<0.05, * p<0.10.

Table 11: Horse Race between Internal IT Adoption and External Knowledge Flows

	(1) Knowledge Flows	(2) IT Capital	(3) IT Capital and Knowledge Flows
Non-IT capital	0.110**	0.109**	0.110**
	(0.043)	(0.043)	(0.043)
IT capital (c_{it})	0.010*	0.005	0.004
	(0.006)	(0.007)	(0.008)
Non-IT labor	0.726***	0.726***	0.726***
	(0.059)	(0.059)	(0.059)
Knowledge flows (s_{it})	-0.008***	0.006	-0.009***
	(0.003)	(0.004)	(0.003)
Prior related investment (m_{it})	--	--	--
s_{it} * m_{it}	0.015***		0.016***
	(0.005)		(0.005)
s_{it} *d_{it} (difficulty of learning)	-0.006***		-0.005***
	(0.001)		(0.001)
s_{it} * d_{it} * m_{it}	0.004**		0.005***
	(0.002)		(0.002)
c_{it} * m_{it}		0.006	0.008
		(0.010)	(0.010)
c_{it} * d_{it}		0.018*	0.002
		(0.009)	(0.009)
c_{it} * d_{it} * m_{it}		-0.020**	-0.004
		(0.009)	(0.009)
Constant	1.715***	1.724***	1.719***
	(0.454)	(0.452)	(0.454)
Observations	1,240	1,240	1,240
Number of firms	275	275	275
R-squared	0.576	0.576	0.576

The dependent variable is the natural logarithm of Value Added. All models use firm-level fixed effects and year dummies. Robust standard errors in parentheses. All R-square values are “within” estimates that do not include the explanatory power of the fixed effects.

*** $p<0.01$, ** $p<0.05$, * $p<0.10$

References

Adams, J.D. 2006. Learning, internal research, and spillovers. *Economics of Innovation and New Technology*. **15**(1) 5-36.

Anderson, M.C., R.D. Banker, S. Ravindran. 2003. The new productivity paradox. *Communications of the ACM*. **46**(3) 91-94.

Aral, S., E. Brynjolfsson, D.J. Wu. 2006. Which Came First, IT or Productivity? Virtuous Cycle of Investment and Use in Enterprise Systems.

Aral, S., P. Weill. 2007. IT Assets, Organizational Capabilities, and Firm Performance: How Resource Allocations and Organizational Differences Explain Performance Variation. *ORGANIZATION SCIENCE*. **18**(5) 763-780.

Arora, A., A. Gambardella. 1994. The changing technology of technological change: general and abstract knowledge and the division of innovative labour. *Research policy*. **23**(5) 523-532.

Barns, B., T.B. Bollinger. 1991. Making reuse cost-effective. *IEEE software*. **8**(1) 13-24.

Bartel, A., C. Ichniowski, K. Shaw. 2007. How Does Information Technology Affect Productivity? Plant-Level Comparisons of Product Innovation, Process Improvement, and Worker Skills. *The Quarterly Journal of Economics*. **122**(4) 1721-1758.

Becker, G.S. 2008. Human capital. The concise encyclopedia of economics. The Library of Economics and Liberty, http://www.econlib.org/library/Enc/HumanCapital.html.

Benkard, C.L. 2000. Learning and Forgetting: The Dynamics of Aircraft Production. *The American Economic Review*. **90**(4) 1034-1054.

Bloom, N., R. Sadun, J. Van Reenen. 2012. Americans do IT better: US multinationals and the productivity miracle. *The American Economic Review*. **102**(1) 167-201.

Bresnahan, T., S. Greenstein. 1996. Technical progress and co-invention in computing and in the uses of computers. *Brookings Papers on Economic Activity. Microeconomics*. **1996** 1-83.

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

Brynjolfsson, E., D. Fitoussi, L. Hitt. 2007. The IT Iceberg: Measuring the Tangible and Intangible Computing Assets, MIT working paper.

Brynjolfsson, E., L. Hitt. 1995. Information technology as a factor of production: The role of differences among firms. *Economics of Innovation and New Technology*. **3**(3&4) 183 - 200.

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

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

Brynjolfsson, E., L.M. Hitt, S. Yang. 2002. Intangible assets: Computers and organizational capital. *Brookings Papers on Economic Activity*. **2002**(1) 137-181.

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

Caselli, F., W.J. Coleman. 2001. Cross-Country Technology Diffusion: The Case of Computers. *American Economic Review*. **91**(2) 328-335.

Chang, Y.B., V. Gurbaxani. 2012a. The impact of IT-related spillovers on long-run productivity: An empirical analysis. *Information Systems Research*. **23**(3) 868-886.

Chang, Y.B., V. Gurbaxani. 2012b. Information technology outsourcing, knowledge transfer, and firm productivity: An empirical analysis. *MIS Quarterly*. **36**(4) 1043-1063.

Cheng, Z., B.R. Nault. 2007. Industry level supplier-driven IT spillovers. *Management Science*. **53**(8) 1199-1216.

Cheng, Z., B.R. Nault. 2012. Relative industry concentration and customer-driven IT spillovers. *Information Systems Research*. **23**(2) 340-355.

Chwelos, P., R. Ramirez, K.L. Kraemer, N.P. Melville. 2010. Does technological progress alter the nature of information technology as a production input? New evidence and new results. *Information Systems Research*. **21**(2) 392-408.

Cohen, W.M. 2010. Fifty years of empirical studies of innovative activity and performance. *Handbook of the Economics of Innovation*. **1** 129-213.

Cohen, W.M., D.A. Levinthal. 1989. Innovation and learning: the two faces of R & D. *The economic journal*. **99**(397) 569-596.

Cohen, W.M., D.A. Levinthal. 1990. Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly*. **35**(1) 128-152.

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

Dewan, S., C. Shi, V. Gurbaxani. 2007. Investigating the risk return relationship of information technology investment: Firm-level empirical analysis. *Management Science*. **53**(12) 1829-1842.

Di Maggio, M., M. Van Alstyne. 2013. Information Sharing, Social Norms and Performance, MIT working paper, available at SSRN: http://ssrn.com/abstract=1893164

Dranove, D., C. Forman, A. Goldfarb, S. Greenstein. 2014. The trillion dollar conundrum: Complementarities and health information technology. *American Economic Journal: Economic Policy*. **6**(4) 239-270.

Escribano, A., A. Fosfuri, J.A. Tribó. 2009. Managing external knowledge flows: The moderating role of absorptive capacity. *Research policy*. **38**(1) 96-105.

Fahey, R., A.C. Vasconcelos, D. Ellis. 2007. The impact of rewards within communities of practice: a study of the SAP online global community. *Knowledge Management Research & Practice*. **5** 186-198.

Forman, C., A. Ghose, B. Wiesenfeld. 2008. Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*. **19**(3) 291-313.

Furman, J.L., M.K. Kyle, I.M. Cockburn, R. Henderson. 2005. Public & Private Spillovers: Location and the Productivity of Pharmaceutical Research. *Annals of Economics and Statistics*. **2005**(79/80) 165-188.

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

Griffith, R., R. Harrison, J. Van Reenen. 2006. How special is the special relationship? Using the impact of US R&D spillovers on UK firms as a test of technology sourcing. *The American Economic Review*. **96**(5) 1859-1875.

Griliches, Z. 1979. Issues in assessing the contribution of research and development to productivity growth. *Bell Journal of Economics*. **10**(1) 92-116.

Griliches, Z. 1992. The search for R&D spillovers. *Scandinavian Journal of Economics*. **94**(Supplement) S29-S47.

Gu, B., L. Xue, G.N. Ray. 2008. IT Governance and IT Investment Performance: An Empirical Analysis.

Haas, M.R., P. Criscuolo, G. George. 2015. Which problems to solve? Online knowledge sharing and attention allocation in organizations. *Academy of Management Journal*. **58**(3) 680-711.

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

Hitt, L.M., D.J. Wu, X. Zhou. 2002. Investment in enterprise resource planning: Business impact and productivity measures. *Journal of Management Information Systems*. **19**(1) 71-98.

Huang, P., Z. Zhang. Participation in Open Knowledge Communities and Job-Hopping: Evidence from Enterprise Software. *MIS Quarterly*. **40**(3) 785-806.

Hubbard, T.N. 2003. Information, decisions, and productivity: On-board computers and capacity utilization in trucking. *The American Economic Review*. **93**(4) 1328-1353.

Hulten, C.R. 2001. Total factor productivity: a short biography. C.R. Hulten, E.R. Dean, M.J. Harper, eds. *New Developments in Productivity Analysis*, University of Chicago Press, 1-54.

Iansiti, M., K.R. Lakhani. 2009. SAP AG: Orchestrating the Ecosystem. Harvard Business School Case 9-609-069.

Jaffe, A.B. 1986. Technological opportunity and spillovers of R & D: Evidence from firms' patents, profits, and market value. *The American Economic Review*. **76**(5) 984-1001.

King, A., K.R. Lakhani. 2013. Using open innovation to identify the best ideas. *MIT Sloan management review*. **55**(1) 41.

Ko, D.G., L.J. Kirsch, W.R. King. 2005. Antecedents of knowledge transfer from consultants to clients in enterprise system implementations. *MIS Quarterly*. **29**(1) 59-85.

Kogut, B., U. Zander. 1992. Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization science*. **3**(3) 383-397.

Lakhani, K.R., E. von Hippel. 2003. How open source software works:"free" user-to-user assistance. *Research policy*. **32**(6) 923-943.

Lakhani, K.R., R. Wolf. 2005. Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects. J. Feller, B. Fitzgerald, S. Hissam, K.R. Lakhani, eds. *Perspectives on Free and Open Source Software*, MIT Press, Cambridge.

Lee, D.M.S., E.M. Trauth, D. Farwell. 1995. Critical skills and knowledge requirements of IS professionals: A joint academic/Industry investigation. *MIS Quarterly*. **19**(3) 313-340.

Levy, F., R.J. Murnane 2005. *The new division of labor: How computers are creating the next job market*, Princeton University Press.

Lieberman, M.B. 1984. The learning curve and pricing in the chemical processing industries. *The RAND Journal of Economics*. **15**(2) 213-228.

McAfee, A. 2002. The impact of enterprise information technology adoption on operational performance: An empirical investigation. *Production & Operations Management*. **11**(1) 33-53.

McAfee, A., E. Brynjolfsson. 2008. Investing in the IT that makes a competitive difference. *Harvard Business Review*. **86**(7/8) 98-107.

Mehra, A., R. Dewan, M. Freimer. 2011. Firms as incubators of open-source software. *Information Systems Research*. **22**(1) 22-38.

Milgrom, P., J. Roberts. 1990. The economics of modern manufacturing: Technology, strategy, and organization. *The American Economic Review* 511-528.

Mittal, N., B.R. Nault. 2009. Investments in information technology: Indirect effects and information technology intensity. *Information Systems Research*. **20**(1) 140-154.

Mortensen, P.S., C.W. Bloch 2005. *Oslo Manual-Guidelines for collecting and interpreting innovation data*, Organisation for Economic Cooperation and Development, OECD.

Nagle, F. 2015. Crowdsourced Digital Goods and Firm Productivity: Evidence from Open Source Software. *Harvard Business School Research Paper*(15-062).

Roberts, N., P.S. Galluch, M. Dinger, V. Grover. 2012. Absorptive Capacity and Information Systems Research: Review, Synthesis, and Directions for Future Research. *MIS Quarterly*. **36**(2) 625-648.

Sambamurthy, V., A. Bharadwaj, V. Grover. 2003. Shaping agility through digital options: Reconceptualizing the role of information technology in contemporary firms. *MIS quarterly* 237-263.

Tambe, P. 2014. Big data investment, skills, and firm value. *Management Science*. **60**(6) 1452-1469.

Tambe, P., L.M. Hitt. 2012. The productivity of information technology investments: New evidence from IT labor data. *Information Systems Research*. **23**(3) 599-617.

Tambe, P., L.M. Hitt. 2014a. Job hopping, information technology spillovers, and productivity growth. *Management science*. **60**(2) 338-355.

Tambe, P., L.M. Hitt. 2014b. Measuring spillovers from information technology investments. *Information Systems Research*. **25**(1) 53–71.

Teece, D.J. 1977. Technology transfer by multinational firms: The resource cost of transferring technological know-how. *The economic journal*. **87**(346) 242-261.

Thornton, R.A., P. Thompson. 2001. Learning from experience and learning from others: An exploration of learning and spillovers in wartime shipbuilding. *American Economic Review*. **91**(5) 1350-1368.

Van Alstyne, M., H. Benbya. 2011. How to Find Answers Within Your Company. *MIT Sloan Management Review*. **52**(2) 65.

von Hippel, E. 1994. Sticky Information and the Locus of Problem Solving: Implications for Innovation. *Management Science*. **40**(4) 429-439.

von Hippel, E., M.J. Tyre. 1995. How learning by doing is done: problem identification in novel process equipment. *Research Policy*. **24**(1) 1-12.

Waguespack, D.M., L. Fleming. 2009. Scanning the commons? Evidence on the benefits to startups participating in open standards development. *Management Science*. **55**(2) 210-223.

Walker, B.A.S.B.B. 2012. Successfully Navigating the Turbulent Skies of a Large-Scale ERP Implementation, HEC Montreal Centre for Case Studies.

Yusuf, Y., A. Gunasekaran, M.S. Abthorpe. 2004. Enterprise information systems project implementation:: A case study of ERP in Rolls-Royce. *International Journal of Production Economics*. **87**(3) 251-266.

IT Knowledge Spillovers, Absorptive Capacity, and Productivity: Evidence from Enterprise Software

Appendix 1: Endogeneity and Identification with Homogeneous Absorptive Capacity

While we use fixed effects to control for time-invariant firm-level heterogeneity, there might be time-varying, unobserved factors that are correlated with both our measurement of knowledge spillovers and firm productivity. In addition, the presence of measurement errors in our spillover variable may lead to bias in estimation. Although it's impossible to control for all the unobservable factors, in the following we present a systematic discussion of the various endogeneity issues, and the measures we take to address each one of them. We focus on 5 sources of potential endogeneity in this discussion: (1) mismeasurement of our spillover variable, (2) unobserved spillovers through searching existing body of knowledge on SAP discussion forums, (3) whether the effect we observe is driven by a few super star IT employees, (4) more generally, how the results are affected by unobserved changes in IT human capital, and (5) IT spillovers through other channels that may be correlated with our measure.

First, endogeneity concerns may arise because of the way of measuring our IT spillovers variable. For example, one potential issue with our baseline spillover measure is that knowledge seekers may not pay enough attention to the answers posted, or simply lack the expertise to judge the quality of the answers, leading to mismeasurement of the spillover variable. Moreover, we may mismeasure the magnitude of spillovers if a knowledge seeker rewards too many knowledge contributors by marking their posts as helpful (since, unlike correct answers and very helpful answers, there is no limit on the number of helpful answers per thread).

Such measurement errors, if they exist, would most likely result in an attenuation bias in the estimates. However, to assess how measurement error might influence our results, we perform two separate analyses. First, we calculate the spillover variable using reward points from only correct and very helpful answers, therefore preventing the large number of helpful answers from inflating the spillover variable. We present the result of this analysis in Column 1 of Table A1. Second, we construct the spillover variable by simply counting the number of questions that are resolved (questions that received either a correct answer or at least a very helpful answer) without using reward points as weight. The result is presented in Column 2 of Table A1. We find that the findings are robust to different ways of measuring the spillover variable.

[Insert Table A1 about here]

Second, one possible source of omitted variable bias in our estimation is due to our limited ability to observe other related activities on SCN. In particular, our measurement of spillovers is based on forum Q&A discussions, and for such spillovers to occur the knowledge seeker must explicitly ask a question in the forum in the first place. However, we note that many knowledge seekers may obtain knowledge spillovers without explicitly asking questions, especially when similar problems have been resolved by other community members already – they can perform a keyword search on SCN forums and find existing solutions to their problems, instead of initiating a new Q&A discussion thread. This alternative channel of knowledge spillovers is especially important for more mature SAP products and modules, because it is likely that solutions to many issues are readily available through accumulation of existing knowledge pool. To address this issue, we construct a new variable – the size of existing knowledge pool –that captures the effect of possible spillovers through search. Particularly, we make use of the SAP installation data and take advantage of the variations in the different product modules installed by the firms in our data. The basic assumption behind this exercise is that if a large knowledge pool exists for the particular SAP modules a company uses, the firm is more likely to benefit from knowledge spillovers through searching the knowledge pool.

To construct this variable, we observe that the online forums are organized so that each forum is dedicated to a specific topic, which often corresponds to an SAP product module or an industry solution. In addition, SAP introduced the forums gradually over time, with only 57 forums in 2004 and grow to 209 forums in 2008 (see details in Table 1). To identify which forums are most useful for a firm, we create a mapping table that associates each SAP product module with the most relevant forum. Using the modules installed at each firm in 2004, we establish a "installed SAP modules--SCN topic forums" correspondence, and for each firm-year observation, we count the number of existing, resolved cases (questions that received correct answers) by the end of that year in all the forums that are relevant to the firm in question. In column 3 and 4 of Table A1 we present the results of the models that control for the effect of existing knowledge pool, using two different types of operationalization, respectively: the first is a simple count of solved cases, and the second is the number of quality (reward points) weighted solved cases. As a robustness check, we also explore an alternative way of identifying the most useful forums to a firm, based on the firm's historical usage of the SAP forums: for firm *i* and year *t*, we find a list of forums that firm *i* has been historically active (i.e. at least one employee of the firm has posted messages in the forum) prior to or during year *t*. Once the list of relevant forums is identified, the existing knowledge pool for each firm-year observation is similarly constructed using the collection of solved cases in these forums. In column 5 of Table A1 we present the result of the model that incorporates this alternative measure of knowledge pool. We find that our results are robust to the effect of searching existing knowledge pool.

A third endogeneity concern is related to an alternative explanation of our results: rather than capturing the value of knowledge spillovers, our main independent variable of interest is mainly driven by a few highly productive, super-star employees of the recipient firm who are SAP enterprise software experts and are very active on the SAP Community Network, and therefore its effect is confounded with firm-level human capital endowment.[26] If this is true, we expect the actual marginal effect of knowledge spillovers would be less than suggested by its estimated coefficient.

We explicitly test this alternative hypothesis by constructing a measure of the number of superstar employees in the sample firms, using a gamification feature introduced by SAP on SCN. Specifically, around year 2009 SAP introduced a reputation system that awards platinum, gold, silver and bronze medals to active contributors to SAP online community according to their levels of lifetime contribution. SAP made the medal winners highly visible: the medal badges are displayed in the business card of the contributors, in the discussion forums and blogs whenever the individuals make a contribution, and next to the individuals' name on the list of top contributors. This feature allows us to identify the superstar members in the SAP community network. We created a variable, medalist, which is the sum of the number of employees that have earned such elite titles. The longitudinal variation of this variable comes from timing of the medalists' entries into SCN and their tenures at their employers. We present two separate analyses to disentangle the effect of having superstar employees from that of general knowledge spillovers. In column 6 of Table A1 we run a regression where we add the new variable, medalists, in addition to the knowledge spillover measurement. We find that although the magnitude of the spillover effect is reduced once we control for the superstar employee effect, its contribution to firm productivity is still positive and significant. In column 7 of Table A1 we run an alternative specification where we construct a measure of spillover that excludes the knowledge inflows acquired by medalists/superstar employees of the firm. The result indicates that our finding is not completely driven by superstars: there remains a positive spillover effect, although we do observe that the marginal effect is smaller once we remove the impact from superstars.

Fourth, related to the above-mentioned issue, on a broader level, our spillover measure is likely to be correlated with firm-level general human capital deepening in enterprise software, or the intensity of use of the SAP Community Network. It is possible that firm investment in human capital is driving both enhanced productivity and higher level of knowledge inflows from SCN, therefore our observed effect of spillover on productivity is in fact a spurious relationship due to simultaneity. While we acknowledge it is difficult to obtain data on firm-level investment in human capital or its IT employees, we try to control its effect by one manifestation of such investment: the knowledge contribution made by a firm's employees,

[26] We thank an anonymous reviewer for suggesting this point.

or knowledge outflows, on SCN. The underlying assumption is that a firm with more capable IT employees –or a higher level of human capital – is more likely to be able to help other members who seek knowledge by answering their questions, resulting in a higher level of knowledge outflow. Similar to the way in which we calculate knowledge inflow, for each firm-year we calculate the sum of the reward points received by its employees (which occur when they correctly answer other knowledge seekers' questions), and use it to proxy for firm-level changes in human capital in enterprise software. We present the result of a model with this control variable in Column 8 of Table A1. In column 9 of Table A1, we control for both super star employees and human capital in enterprise software. Our result shows that the marginal effect of knowledge spillover is reduced when we control for firm-level human capital, but there remains a positive contribution of knowledge spillovers.

Finally, endogeneity may cause bias in our coefficient estimate of knowledge spillover if our measure of spillover is correlated with those occurred via other spillover channels, such as those mediated through external knowledge pools. In prior literature, such external knowledge pools are usually modeled as weighted sum of IT investments of other firms in the same industry or facing the same technological opportunities (Tambe and Hitt 2014b). We present two sets of results where we explicitly account for other spillover channels using a pooled approach, similar to the one adopted by Hitt and Tambe (2014b). First, we construct the spillover pool by industry proximity and define the pool as the sum of IT investments made by all other companies (among fortune 1000) in the same 3-digit NAICS industry, and present the results in column 10 of Table A1. Second, we also construct the spillover pool by both industry proximity and technological proximity. Specifically, we define the pool as the sum of the IT investments by all other firms (among fortune 1000) that 1) fall into the same 3-digit NAICS industry as the focal firm, and 2) have installed SAP enterprise software before 2004. The results of incorporating the spillover pool are presented in column 11 of Table A1. We find that our knowledge spillover measure is only moderately correlated with industry spillover pool (ρ=0.026) and industry-technology spillover pool (ρ=0.026), and the coefficient estimate remain robust after we add the pooled measures of IT spillovers.[27]

Appendix 2: Dynamic Panel Estimates with Homogeneous Absorptive Capacity

To further address the remaining endogeneity issues of IT spillover, we relax the assumption that IT-related spillover is strictly exogenous and use the generalized method of moments (GMM)-based dynamic panel data models estimator (Arellano and Bond 1991, Blundell and Bond 1998). In particular,

[27] We note that unlike Hitt and Tambe (2014b), the spillover pool is not significant in column 10 of Table A1. However, this difference is likely due to the smaller sample we use here – we only consider the 275 SAP user firms among Fortune 1000 in this study. Indeed, when we run the same model using Fortune 1000 sample we get positive and significant estimate of spillover pool, consistent with Hitt and Tambe (2014b).

consistent with Arellano and Bond (1991), we construct internal instruments within the data that is provided by dynamic panel data model. More specifically, we use difference GMM, employing lag term of our endogenous variables, IT spillovers and further lags of added value, and all difference of other exogenous variables including year dummies as our instrument variables for differenced equation. We checked the validity of the moment conditions required by difference GMM using the Hansen test, which does not reject the assumption that our instruments are exogenous (Arellano and Bond 1991, Roodman 2009). We also test the validity of generalized method of moments (GMM) in our dynamic panel data model. The test results indicate that our model specification is based on no serial correlation in the first-differenced disturbances.

[Insert Table A2 about here]

We report the results from the baseline Arellano-Bond estimator of dynamic panel data model with lagged dependent variable in column 1 of Table A2, where IT spillover is treated as exogenous. In column 2 we treat IT-related spillovers as endogenous and use internal instrument variables to instrument for both lagged dependent variable and IT spillovers. The coefficient estimate of IT spillover in column 2 remains supportive of a positive contribution of IT spillover to productivity. In column 3 we present the dynamic panel model where we assume IT capital as endogenous, a similar specification used by Tambe and Hitt (Tambe and Hitt 2014a). We find the endogeneity of IT capital actually produces downward bias on its coefficient – the marginal effect of IT capital is slightly higher after controlling for its endogeneity. In column 4 we present a dynamic panel model where we treat both IT capital and IT spillover as endogenous, and the estimates are little changed. Overall, the Arellano-Bond estimates provide support for that our main finding is robust to the endogeneity of IT spillover.

Appendix 3: Additional Robustness Tests of the Absorptive Capacity Model

First, to account for related activities on SCN that are correlated with knowledge inflows, we estimate the model specification in equation (8) that incorporates the nature of knowledge flows (e.g., difficulty of learning) and prior related investments while controlling for the number of SCN users and the number of questions raised by a firm's employees, and present the results in Table A3. We find very similar results.

[Insert Table A3 about here]

In addition, we evaluate whether our findings are robust to alternative measures of knowledge inflows, to the exclusion of knowledge inflows from superstar employees, and to the effect of spillover pools, using the approaches similar to those presented in Appendix 1. Particularly, we calculate the spillover variable using reward points from only correct and very helpful answers, and present the result of this analysis in Column 1 of Table A4. Column 2 reports the results where we construct the spillover variable by simply counting the number of questions that are resolved (questions that received either a correct answer or at least a very helpful answer) without using reward points as weight. In column 3 we run a model specification where we use a measure of spillover that excludes the knowledge inflows acquired by medalists/superstar employees of the firm. We also control for the effects of the spillover pools, using both the industry spillover pool variables and the industry-technology spillover pool variable, as defined in Appendix 1, and present the results in column 4 and column 5. All the findings are consistently supported.

[Insert Table A4 about here]

References

Arellano, M., S. Bond. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*. **58**(2) 277-297.

Blundell, R., S. Bond. 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of econometrics*. **87**(1) 115-143.

Roodman, D. 2009. A note on the theme of too many instruments. *Oxford Bulletin of Economics and statistics*. **71**(1) 135-158.

Tambe, P., L.M. Hitt. 2014a. Job hopping, information technology spillovers, and productivity growth. *Management science*. **60**(2) 338-355.

Tambe, P., L.M. Hitt. 2014b. Measuring spillovers from information technology investments. *Information Systems Research*. **25**(1) 53–71.

Table A1: Endogeneity and Unobserved Heterogeneities

Variables	(1) Alternative measures of spillovers	(2)	(3) Knowledge pool	(4)	(5)	(6) Superstar employee	(7)	(8) Human resource	(9) Superstar +Human resource	(10) Spillover pools	(11)
Non-IT capital	0.109**	0.109**	0.109**	0.109**	0.109**	0.111**	0.108**	0.102**	0.103**	0.110**	0.106**
	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)	(0.044)
IT capital	0.010*	0.010*	0.010*	0.010*	0.010*	0.010*	0.010*	0.009*	0.009*	0.010*	0.010*
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Non-IT labor	0.726***	0.725***	0.725***	0.725***	0.725***	0.725***	0.727***	0.724***	0.724***	0.726***	0.728***
	(0.059)	(0.059)	(0.059)	(0.059)	(0.058)	(0.059)	(0.059)	(0.057)	(0.057)	(0.059)	(0.059)
IT spillover	0.009**	0.016**	0.009**	0.009**	0.010***	0.008**		0.007*	0.006†	0.009**	0.009**
	(0.004)	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)		(0.004)	(0.004)	(0.004)	(0.004)
IT spillover (excluding superstar)							0.007*				
							(0.004)				
knowledge pool			-0.001								
			(0.005)								
knowledge pool (weighted)				-0.002	-0.002						
				(0.005)	(0.002)						
medalists						0.041			0.033		
						(0.029)			(0.040)		
knowledge outflow (cumulative)								0.010*	0.010*		
								(0.006)	(0.006)		
Spillover pool										0.019	0.010
										(0.014)	(0.011)
Constant	1.726***	1.727***	1.691***	1.701***	1.731***	1.722***	1.727***	1.794***	1.790***	1.587***	1.679***
	(0.452)	(0.452)	(0.464)	(0.466)	(0.447)	(0.452)	(0.452)	(0.446)	(0.447)	(0.446)	(0.453)
Observations	1,240	1,240	1,240	1,240	1,240	1,240	1,240	1,240	1,240	1,240	1,240
R-squared	0.575	0.575	0.575	0.575	0.576	0.575	0.575	0.578	0.578	0.576	0.575
Number of firms	275	275	275	275	275	275	275	275	275	275	275

The dependent variable is the natural logarithm of Value Added. In column 1 the spillover measure includes only reward points from correct and very helpful answers. In column 2 the spillover measure is defined as the number of resolved questions without using reward points as weight. Searchable knowledge base is defined as # of solved questions in related forums in column 3, and reward point weighted knowledge pool in related forums in column 4 and 5. In column 4 a forum is defined as related if a firm had installed related the SAP product. In column 5 a forum is defined as related if a firm is historically active in that forum. IT spillover in column 7 excludes spillovers from superstar employees.
All models use firm-level fixed effects and year dummies.
Robust standard errors (clustered by firm) in parentheses. Within R-squares are reported for panel data models.
Columns (3) - (10) compute IT spillovers using reward points from correct, very helpful, and helpful answers, Column (1) uses only reward points from only correct and very helpful answers, and Column (2) computes IT spillovers by counting the number of correct answers to questions.
*** p<0.01, ** p<0.05, * p<0.10, † p=0.10

Table A2 Arellano-Bond Dynamic Panel Estimate

Variables	(1) With lagged DV	(2) With lagged DV, spillover as endogenous	(3) With lagged DV, IT capital as endogenous	With lagged DV, both spillover and IT capital as endogenous
Lag of Added Value	0.489**	0.561**	0.263*	0.339**
	(0.237)	(0.220)	(0.148)	(0.146)
Non-IT capital	-0.001	-0.009	0.011	0.003
	(0.057)	(0.059)	(0.054)	(0.056)
IT capital	0.014***	0.014***	0.015***	0.015***
	(0.005)	(0.005)	(0.005)	(0.005)
Non-IT labor	0.544***	0.534***	0.595***	0.585***
	(0.077)	(0.073)	(0.060)	(0.060)
IT spillover	0.007*	0.007*	0.007*	0.007†
	(0.004)	(0.004)	(0.004)	(0.004)
Log(registered users)	-0.009	-0.008	-0.009	-0.007
	(0.016)	(0.017)	(0.015)	(0.015)
Log(questions)	-0.005	-0.006	-0.004	-0.006
	(0.009)	(0.009)	(0.008)	(0.008)
Constant	0.112	-0.318	1.421	0.963
	(1.404)	(1.303)	(0.917)	(0.894)
Model fit	Wald chi2(13) =376.74 Prob > chi2 =0.0000	Wald chi2(13) = 365.81 Prob > chi2 = 0.0000	Wald chi2(13) =380.26 Prob > chi2 = 0.0000	Wald chi2(13) = 385.79 Prob > chi2 =0.0000
Instruments for differenced equation	GMM-type: L(2/.).log(avalue) Standard: first difference of all explanatory variables	GMM-type: L(2/.).log(avalue) L(2/.).log(spillover) Standard: first difference of all explanatory variables	GMM-type: L(2/.).log(avalue) L(2/.).log(IT) Standard: first difference of all explanatory variables	GMM-type: L(2/.).log(avalue) L(2/.).log(spillover) L(2/.).log(IT) Standard: first difference of all explanatory variables
Arellano-Bond test for zero autocorrelation in first-differenced errors	First order: z = -1.31 Pr > z = 0.191 Second order: z = -0.99 Pr > z = 0.321	First order: z = -1.63 Pr > z = 0.104 Second order: z = -0.98 Pr > z = 0.325	First order: z = -0.96 Pr > z = 0.335 Second order: z = -1.01 Pr > z = 0.315	First order: z = -1.39 Pr > z = 0.165 Second order: z = -1.01 Pr > z = 0.315
Hansen test of overid. restrictions	chi2(2) = 3.75 Prob > chi2 = 0.154	chi2(8) = 7.99 Prob > chi2 = 0.435	chi2(8) = 7.26 Prob > chi2 = 0.509	chi2(14) = 14.73 Prob > chi2 = 0.397
Observations	706	706	706	706
Number of firms	249	249	249	249

The dependent variable is the natural logarithm of Value Added. Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.10, † p=0.10

Table A3: Test of Absorptive Capacity Hypotheses, Controlling for Questions and Users

Variables	Panel A D = Percentage of Business Knowledge Inflows		Panel B D=Percentage of Knowledge Inflows from New Forums	
	(1) m=advanced SAP module	(2) m= high human capital	(3) m=advanced SAP module	(4) m= high human capital
Non-IT capital	0.111**	0.103**	0.111**	0.105**
	(0.043)	(0.044)	(0.043)	(0.044)
IT capital	0.010*	0.010*	0.010*	0.010*
	(0.006)	(0.006)	(0.006)	(0.006)
Non-IT labor	0.729***	0.734***	0.728***	0.731***
	(0.059)	(0.055)	(0.059)	(0.055)
Log(registered users)	-0.011	-0.012	-0.011	-0.012
	(0.017)	(0.017)	(0.017)	(0.016)
Log(questions)	-0.009	-0.014	-0.009	-0.014
	(0.010)	(0.010)	(0.010)	(0.010)
Knowledge flows (s)	-0.004	0.003	-0.007	-0.002
	(0.004)	(0.005)	(0.005)	(0.007)
Prior related investment (m)	--	0.070**	--	0.069**
		(0.029)		(0.029)
s*m	0.016***	0.011	0.017***	0.017**
	(0.005)	(0.008)	(0.007)	(0.008)
s*d (difficulty of learning)	-0.005***	-0.005**	-0.003***	-0.004**
	(0.001)	(0.002)	(0.001)	(0.002)
s* d*m	0.004**	0.006***	0.002	0.004**
	(0.002)	(0.002)	(0.002)	(0.002)
Constant	1.659***	1.691***	1.663***	1.691***
	(0.459)	(0.442)	(0.458)	(0.441)
Observations	1,240	1,240	1,240	1,240
Number of firms	275	275	275	275
R-squared	0.578	0.583	0.578	0.583

The dependent variable is the natural logarithm of Value Added. All models use firm-level fixed effects and year dummies. Robust standard errors (clustered by firm) in parentheses. All R-square values are "within" estimates that do not include the explanatory power of the fixed effects.
*** p<0.01, ** p<0.05, * p<0.10

Table A4: Robustness Tests of Absorptive Capacity Model

Variables	(1)	(2)	(3)	(4)	(5)
	Alternative measures of spillovers		Spillovers that exclude superstar	Spillover Pool	
Non-IT capital	0.111**	0.111**	0.109**	0.111**	0.107**
	(0.043)	(0.043)	(0.044)	(0.043)	(0.044)
IT capital	0.010*	0.010*	0.010*	0.010	0.010*
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Non-IT labor	0.729***	0.729***	0.729***	0.729***	0.731***
	(0.059)	(0.059)	(0.059)	(0.059)	(0.059)
Log(registered users)	-0.011	-0.009	-0.005	-0.011	-0.011
	(0.017)	(0.017)	(0.005)	(0.017)	(0.017)
Log(questions)	-0.009	-0.011	-0.015	-0.010	-0.009
	(0.010)	(0.010)	(0.017)	(0.010)	(0.010)
Knowledge flows (s)	-0.004	-0.006	-0.004	-0.003	-0.003
	(0.005)	(0.008)	(0.011)	(0.004)	(0.004)
Spillover Pool				0.018	0.011
				(0.014)	(0.011)
s*m (prior related investments)	0.017***	0.031***	0.014***	0.016***	0.016***
	(0.005)	(0.009)	(0.005)	(0.005)	(0.005)
s*d (difficulty of learning)	-0.006***	-0.010***	-0.005***	-0.005***	-0.005***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
s* d*m	0.005**	0.009***	0.004*	0.004**	0.004**
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
Constant	1.660***	1.660***	1.671***	1.524***	1.609***
	(0.459)	(0.459)	(0.459)	(0.453)	(0.459)
Observations	1,240	1,240	1,240	1,240	1,240
Number of firms	275	275	275	275	275
R-squared	0.578	0.577	0.576	0.578	0.578

The dependent variable is the natural logarithm of Value Added. All models use firm-level fixed effects and year dummies. Robust standard errors (clustered by firm) in parentheses. All R-square values are "within" estimates that do not include the explanatory power of the fixed effects.

In column 1 the spillover measure includes only reward points from correct and very helpful answers. In column 2 the spillover measure is defined as the number of resolved questions without using reward points as weight. IT spillover in column 3 excludes spillovers from superstar employees. Difficult of learning is measured by the percentage of business knowledge inflows; prior related investments is measured by deployment of advanced SAP modules.

*** p<0.01, ** p<0.05, * p<0.10