Three Essays in the Economics of Information Technology

by

Heekyung Hellen Kim

Ph.D. Civil and Environmental Engineering
Massachusetts Institute of Technology (1995)

SUBMITTED TO THE SLOAN SCHOOL OF MANAGEMENT
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

AT THE

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2013

ARCHIVES
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
JUL 01 2013
LIBRARIES

© 2013 Heekyung Hellen Kim. All rights reserved.

The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part in any medium now known or hereafter created.

Signature of Author ______________________________

Sloan School of Management
June 1, 2013

Certified by ______________________________

Erik Brynjolfsson
Schussel Family Professor
Professor of Information Technology, Technological Innovation, Entrepreneurship and Strategic Management
Thesis Supervisor

Accepted by ______________________________

Ezra Zuckerman Sivan
Nanyang Technological University Professor
Professor of Technological Innovation, Entrepreneurship and Strategic Management
Chair, MIT Sloan Ph.D. Program

Three Essays in the Economics of Information Technology

by

Heekyung Hellen Kim

Submitted to the Sloan School of Management
On May 3, 2013 in Partial Fulfillment of the
Requirements for the Degree of Doctoral Philosophy in
Management Science

ABSTRACT

The first chapter is to investigate the impact of a free on-line repository of research articles on the diffusion of their ideas measured by the citation counts. The key questions that this chapter answers are as following: 1) does a free on-line repository of research articles increase the diffusion of their scholarly ideas measured by their citations?; 2) who benefits from the free access? By using a dataset from the Social Science Research Network (SSRN), an open repository of research articles, and employing a natural experiment that allows the effect of free access separate from other confounding factors, this study identifies the causal effect of free access on the citation counts as well as shows a heterogeneous effect of free access on both supply and demand side.

The second chapter is to study the correlation between CEO pay and information technology. The hypothesis is that IT increases "effective size" of the firm that a top manager controls and thus her marginal productivity. In turn, in an efficient market, the firms with a higher degree of information technology will reward their CEOs with a higher compensation.

The third chapter is to examine whether firms that emphasize decision making based on data and business analytics ("data driven decision making" or DDD) show higher performance. Using detailed survey data on the business practices and information technology investments of 179 large publicly traded firms, this study finds that firms that adopt DDD have output and productivity that is 5-6% higher than what would be expected given their other investments and information technology usage. Furthermore, the relationship between DDD and performance also appears in other performance measures such as asset utilization, return on equity and market value. Using instrumental variables methods, this study finds evidence that the effect of DDD on the productivity do not appear to be due to reverse causality. These results provide some of the first large scale data on the direct connection between data-driven decision making and firm performance.

Thesis Supervisor: Erik Brynjolfsson

Title: Schussel Family Professor of Management Science, Professor of Information Technology, Director, The MIT Center for Digital Business

Acknowledgement

First of all, I thank my committee members. I thank Erik Brynjolfsson for his generous support, patient guidance, and encouragement. Without his support, I could not have even started this PhD. I thank Scott Stern for his advice on the first chapter. I feel very lucky that he joined MIT and I had an opportunity to get his guidance while I was at MIT. I thank Ajay Agrawal, who was willing to be in my committee member at a late stage of my PhD and provided me emotional as well as intellectual support when I needed it badly. I thank Lorin Hitt for his thorough comments and feedback on the second chapter.

I thank Ezra Zuckerman for his guidance when I was in search of what I wanted to research. He always made time for me whenever I needed it.

I thank Sharon Cayley and Hillary Ross who made sure that I followed all the administrative procedures to complete the PhD program.

I thank Greg Gordon and Ralph Dandrea at SSRN for providing their data for the first chapter.

I thank my fellow graduate students and colleagues for being a friend, a critique, and a discussion partner.

Last but not least, I thank my mother for being my role model. Her courage and resilience have inspired me throughout my life as well as this PhD journey.

Contents

Chapter 1. The Effect of a Free On-line Repository on the Diffusion of Scholarly Ideas

Abstract

This study investigates the impact of a free on-line repository of research articles on the diffusion of their ideas measured by the citation counts. The key questions that this paper answers are that: 1) does a free on-line repository of research articles increase the diffusion of their scholarly ideas?; 2) who benefits from the free access? By using a dataset from the Social Science Research Network (SSRN), an open repository of research articles, and employing a natural experiment that allows the effect of free access separate from other confounding factors such as quality differentials, early exposure, low search cost, and promotion effect, this study identifies the causal effect of free access on the citation counts. The natural experiment in this study is that a select group of articles is posted on SSRN at a time chosen by their authors' affiliated organizations or SSRN, not by their authors. Using a difference-in-difference method and comparing the citation profiles of the articles before and after the posting time on SSRN against a group of control articles published in the same journal and issue, I estimated that the free access increases the citation approximately by 10%. This effect is not, however, homogenous across all the articles. In the supply side, the boost in the citation is driven by the articles published in low-tiered journals before the online posting. The research articles authored by the scholars affiliated with top institutes tend to receive a boost from the online posting. In the demand side, the scholars in the developing countries appear to get more benefits from the free access to the posted articles than those in the developed countries do because the fraction of the citing authors from developing countries to the number of all authors increases after the cited articles are posted on SSRN.

Keywords: Open access, knowledge management, value of free access, diffusion of knowledge

Introduction

Scientific knowledge progresses as science challenges and builds upon prior beliefs and findings. The ability of a society to make scientific progress, therefore, depends on how well the society can validate, organize, maintain, and offer researchers access to the prior knowledge (Rosenberg 1963; Rosenberg 1979; Heller and Eisenberg 1998; David 2005; Mokyr 2002). Many researchers have reported the roles of various institutions in effectively diffusing prior knowledge. For example, Furman and Stern (2011) showed that biological resource centers (BRCs), "living libraries" where biological materials are deposited for researchers, improved the diffusion of knowledge by both certifying and offering access to the deposited materials. Zuckerman and Merton (1971) defined scholarly journals as an institution: the scholarly journals certify the content of the research articles they publish while offering access to the articles for other researchers. These institutions validate the prior knowledge as well as offer access to the knowledge for other researchers and scholarly ideas are distributed through these institutions after being validated.

The advent of the Internet and digital publishing has, however, changed the way that scholarly ideas are disseminated. The validation and distribution of scholarly ideas are no longer provided by the same institution. Scholarly ideas are disseminated through working papers, blogs, Twitter, and many forms other than refereed journals, regardless of their contents being validated. As of 2010 Blogpulse tracked over 166 million blogs, and estimated that 70,000 blog posts were created per day. Research articles are often available for free access in open repositories in the Internet. For example, arXiv.org provides an online repository for researchers in physics, mathematics, computer science, and quantitative biology to post unpublished and published manuscripts for public viewing. As of May 2011, arXiv.org has posted approximately 6,000 new articles per month and total of 677,000 research articles since its inception in 1991. Social Science Research Network (SSRN), an open online repository of research articles in the field of social science, has archived 300,000 research articles as of 2010 since its inception in 1993 and provides free access to those unpublished and published articles. This paper examines how such an online non-refereed repository of research articles has changed the diffusion of scholarly ideas.

The open online repositories of research articles provide two functions that most of traditional journals do not, free access[1] and early exposure, at the expense of no referees or quality control. Anyone who has the access to the Internet can download the full text of articles from such an on-line repository for free while most of traditional journals allow access to their articles only for their (paid) subscribers. Research articles submitted to the online repository is exposed to readers immediately upon submission

[1] Refereed journals increasingly offer free readership. The first type is not subscription-based; the fees are charged to authors for their submission of articles, not readers, and the journals are available for free readership. The journals of the Public Library of Science such as *PLoS Biology* and *PLoS Medicine* are the most notable examples of such type. The second type is that subscription-based journals offer authors to buy open access to their articles. These journals have, therefore, both open access articles to the readers and articles with pay-wall in the same issue.

because there is no refereeing process. Because it often takes a few years from submission to publication (in a refereed journal), many authors post their unpublished manuscripts in an online repository or their own website for free readership before their articles are published in a journal or book (refereed or not). The free access should, in theory, accelerate the diffusion of ideas. However, previous research findings (e.g., Lawrence 2001a; Gaule and Maystre 2011; Davis et al. 2008; Eysenbach 2006) have not reached a consensus on the issue. Moreover, to the best of the author's knowledge, no study has identified whether the effect is homogeneous across different authors and readers. This paper investigates the differential effect of the free on-line repository on the readers as well as the authors while reconciling the differences among the scholars on the effect of free access by separating it from confounding factors in a natural experiment.

Ever since Lawrence (2001a) reported that research articles published in online free computer science proceedings received over an 300% of citations of non-free articles based on a cross-sectional study, many researchers have challenged the findings using a more rigorous identification strategy such as a randomized experiment and an instrumental variable method (e.g., Davis et al. 2008; Gaule and Maystre 2011). The difficulty to measure the effect of free access arises because two other factors are often confounded with free access. The first is selection bias. Often authors select their better articles to post in the Internet for free readership. An apparent difference in citations between free and non-free articles may be, therefore, due to the inherent quality difference. The second is early viewership. Most research articles are available as a working paper for readership well before they are published. Depending on a journal, it takes a few years for a research article to be published. The open repositories of research articles such as arXiv.org and SSRN (Social Science Research Network) allow authors to post any research article. Therefore, a difference in citations between free and non-free articles may arise from that the free articles had a longer time to receive a citation than the non-free articles (Moed 2007). A cross-sectional study to compare citations between free and non-free articles fails to separate the confounding factors and what appears to be the effect of free access can be the effect of other factors.

To identify the effect of free access, this study employs a natural experiment allowed by a unique feature of data from an online repository of research articles, Social Science Research Network (SSRN). SSRN was established in 1993 to enhance the dissemination of research ideas in social science; it covers 24 research disciplines, emphasizing law, economics, and finance. It had archived approximately 300, 000 research articles as of 2010, hosting over 300 working paper series. I exploit three aspects of SSRN to identify the effect of free online access. First, when an organization joins SSRN and starts a research paper series for the organization, a large number of research papers, often over 100 papers, is submitted to SSRN and posted at once, the timing of posting being exogenous to the authors or the quality of the papers. The articles may be chosen by the authors, but the timing of posting those articles is not decided by authors. I do not assume that there is no inherent difference in quality between the articles selected for posting on SSRN and the articles not selected. The assumption here is that the inherent difference in quality between the selected articles for posting on SSRN and unselected articles up to the timing of posting should not change after the timing of posting. Secondly, some portion of research papers posted on SSRN has already been published in journals prior to being posted on SSRN, having a citation profile over time before being posted on SSRN. The citation profile over time before posting on SSRN allows the estimation of quality difference between the articles selected for posting on SSRN and the unselected articles from the same journal and issue. The third relevant trait of SSRN arises from its having been established over 15 years, with many articles available on SSRN for more than 5 years; this relative longevity enables me to compile their citation trajectories over 5 years after they were freely available on SSRN. Because receiving citations from other published articles takes time, allowing enough time for the free access articles to accrue citations is important.

The natural experiment in this study is that some articles, which had been published in refereed journals at least 4 years earlier and thus had a citation profile over time prior to being posted on SSRN, were posted on SSRN at an exogenously chosen time ("treated articles") and other articles, which had been published in the same journal, volume, and issue as their counterpart treated articles, were never posted on SSRN ("control articles"). The treatment in this natural experiment is, therefore, the posting on SSRN. The identification of free access comes from the fact that the timing of posting the treated articles was not decided upon by their authors. By comparing the citation profiles of the treated articles before and after the posting time on SSRN with those of control articles with similar characteristics using a difference-in-difference method, I estimated the effect of free access on citations, not confounded with the quality of the treated articles. First, the results show that SSRN articles receive 60-80 % higher citations than their matched control articles even before being posted on SSRN, indicating that the articles are of higher quality. They receive an additional 10-20% of citations after being posted on SSRN, which is likely to be driven by the free access that SSRN provides, much smaller than what was reported by previous studies (i.e. Lawrence, 2001a).

The next is to investigate the differential effect across the authors (supply) and the readers or citing scholars (demand). First, the cited articles were divided into two groups in various criteria: 1) high vs. low journal impact factor, 2) high vs. low profile institutes, and 3) high vs. low profile authors. Secondly, the readers or the citing authors were also divided into two groups: 1) high vs. low income countries, 2) across vs. within the discipline between the cited and citing authors, and 3) distant vs. close in geographic distance between the authors and the readers. In the supply side, the observed increase appears to be driven by the research articles previously published in low-tiered journals or journals with low journal impact factor. It is consistent with the fact that top-tiered journals tend to be subscribed more widely and be accessible before the SSRN-posting while the low-tiered journals may not be widely distributed and the SSRN-posting may provide an additional outlet for those articles to the readers. The articles authored by at least one affiliated with a high-profile institute seem to have much larger effect than those with a low-profile institute. Especially, the effect becomes more pronounced with the articles published in a low-tiered journal and authored by a scholar affiliated with a top institute. When the authors were divided into high vs. low profile researchers, the articles authored by high profile researchers tend to be cited more than other researchers even before their articles are posted on SSRN, not particularly after the posting. In the demand side of the effect, the citing authors tend to be more from developing countries than developed countries after the posting on SSRN. When the distance between the citing authors and the cited authors was measured, the treated articles or the articles eventually posted on SSRN tend to be closer in knowledge distance to the citing articles even before the posting on SSRN but the posting on SSRN doesn't appear to change the distance. The geographic distance between the cited authors and the citing authors is also unaffected by the posting on SSRN.

The contribution of this paper is that: 1) to identify a causal relationship between free access and the diffusion of scholarly ideas; 2) to separate the quality differential from the free access and quantify those effects separately; and 3) most importantly, to demonstrate a heterogeneous effect of the free on-line repository on the authors and the readers of the articles, separately. The result of this study reconciles the difference among scholars on the open access advantage, showing that the effect of free access is not as large as what was reported by some scholars (i.e. Lawrence, 2001a) but still larger than what was reported as unobservable by other scholars (i.e. Davis et al., 2008). This study shows that the apparent increase in the citations for the open access articles mostly come from the selection bias, much more than from the free access. More importantly, this study is the first to show the heterogeneous effect of the open access.

Open Access Debate

The first theoretical proposition that I test is that scholarly ideas with free access should diffuse more widely and, as a result, their citation counts, a proxy for diffusion, should increase. However, the empirical findings on this relationship have been inconsistent, and researchers are still debating whether there is a causal relationship between free access and citations. The debate on the existence of an "open access advantage" for scholarly communications started when Lawrence (2001b) reported that freely-accessible online computer science proceedings received more than three times the average number of citations of papers as their counterpart paper journals. The open access advantage refers to the fact that free and unrestricted access to research papers gives them an advantage in receiving citations. Many researchers have reported that freely available papers in the Web receive more citations in a variety of disciplines (Norris 2008), such as computer science (Lawrence 2001b), astrophysics (Schwarz and Kennicutt 2004; Metcalfe 2005), physics (Harnad and Brody 2004), mathematics (Antelman 2004; Davis and Fromerth 2007), philosophy (Antelman 2004), political science (Antelman 2004), engineering (Antelman 2004), law (Donovan and Watson 2011), and multi-disciplinary sciences (Eysenbach 2006). Other researchers, however, argue that what appeared to be an open access advantage may be attributable to either early viewing or self-selection or both. For example, Moed (2007) reports that arXiv.org accelerates citation due to the fact that it makes papers available *earlier* rather than by making them *freely* available. On the other hand, Kurtz et al. (Kurtz et al. 2005; Kurtz and Henneken 2007) report that authors tend to make more citable papers such as those published in journals with higher impact factors freely available, suggesting that self-selection, not unrestricted accessibility, causes the increased citation of open access papers. Conducting an experiment that randomly assigned certain articles for free access at publishers' websites, Davis et al. (2008) reported that there is no evidence of an open access advantage for citation counts in the 2 years subsequent to publication. In summary, the open access advantage ranges, researchers argue, from zero to 300% of citations of non-free research articles; early exposure and quality difference have been identified as the potential confounding factors for the overestimation of the effect of free access. Without an identification strategy capable of separating the confounding factors as well as allowing a reasonably long time for articles to receive citations after publication, an unbiased estimate on the value of free access cannot be made.

Moreover, the heterogeneous effect of the open access, if any, has not been identified by the previous studies that reported the open access advantage. The beneficiaries of the open access can be categorized in two groups in terms of supply and demand: the supply side is the authors of the open research articles and the demand side is the readers or other scholars to cite those articles. It is an empirical question whether the open access helps the articles written by well-known authors or published in well-known journals more than those with less-known authors or journals, in other words, the open access promotes Superstar effect. Alternatively, the open access can promote the readership of articles written by less-known authors or published in less-known journals, providing an additional outlet for those articles. With respect to the demand side, the readers in the organizations and countries without access to a wide range of journals may get the benefit from the open access. While there is a report that the scholars in the developing country such as India may not have access to research articles as widely as those in the developed countries have (GaulÃ© 2009), no study has empirically shown that the open access helps the readers in the developing countries more than those in the developed countries. This study reports the heterogeneous effect of the open access, empirically showing that the research articles published in lower-tiered journals tend to get a boost in citations after free access and the readers in the developing countries tend to cite the research articles more after their being posted on SSRN.

Identification Strategy

Previous studies have identified three factors may cause the increases in citations for research articles with open access: 1) free access; 2) early exposure; and 3) quality difference. These three factors are often confounded, causing a biased estimate of the effect of each factor. I employ an identification strategy to separate the effect of free access from other potential confounding factors with a longitudinal dataset. In this study, I focus on a setting where two requirements are met. The first is that an exogenous shock exists to make the research articles available for free access. In this setting, authors do not choose the time to post their articles for free access; instead, the organizations that the authors are affiliated with or the websites that host those articles choose the time to post articles for free access even if the authors choose which of their articles are posted. The second requirement is that these articles were already published for some time before being posted for free access. This requirement served two purposes: 1) the effect of early exposure is removed and, 2) more importantly, these articles have an observable citation trajectory over time before the posting, allowing the comparison of the citation trajectory before and after the posting. The research articles I chose were posted at a time decided upon not by their authors but by the authors' affiliated organizations or their hosting website and had been already published at least 4 years before being posted for free access, meeting the two requirements.

An inference on the effect of any event based on a comparison before and after the event should address a time trend that may concur with the effect of the event. A standard approach to address the time trend is to include time dummy variables in empirical equations. Merely including time dummy variables, however, is not enough to address the time trend when the dependent variable is citation counts. This inadequacy of merely including time dummy variables to account for the time trend exists because the citation profile over time is often specific for each research article, depending on when the articles were first published, when their citing articles were published, the interaction between the publication time of cited articles and citing articles, and the quality of the article. To separate the time and age effects, I used a difference-in-difference estimator by including a set of control articles with characteristics similar to their counterpart treated articles. I chose the control articles based on the following criteria: that they were published in the same journal, volume, and issue; and that they have their own observable citation trajectory over time, as their counterpart treated samples do. The difference-in-difference method that I used in this study was illustrated as in Figure 1. For example, a research article posted on SSRN in 2000 at an exogenous timing is selected for the study. Because the journal and volume published the article are known, the citation profiles for all the articles in the volume before and after the year of posting on SSRN are constructed. The selection bias is determined from the difference in the citations between the articles posted on SSRN, as indicated as SSRN paper, and the articles published in the same volume but not posted on SSRN, as indicated as control paper. The counterfactual citations, as denoted in a dotted line, that SSRN paper would have received after being posted on SSRN was constructed from the citations that the control papers received, on the assumption that their citation trend would be similar to that before the posting event. The difference in citations between the observed citations and the counterfactual citations is interpreted as SSRN effect or the effect of free access.

The limitation of the difference-in-difference estimator is that the counterfactual trajectory of the treated articles is accounted for by the control articles and the quality of match between the treated and control articles is critical. Therefore, in the next analysis for a tighter match between control and treated articles, I used the coarse exact matching method (CEM) to choose a subset of treated articles that can be matched to their control articles with respect to citation profiles over time and total citation counts up to the year when their matching treated articles were posted on SSRN. The citation profile of the control articles provide the counterfactual citation profiles over time that the treated articles would have without being posted for free access. It is, however, possible that the inherent difference between treated and control articles results in the different trajectory after the posting event. Because a traditional fixed effect estimator without using any control unit does not rely on the quality of match between treated and control units, it can be an alternative to the difference-in-difference estimator. Using a traditional fixed effect estimator and some common functional forms that other researchers have used for citation profiles, I also estimated the effect of free access.

For the most statistical analysis when the dependent variable was citation counts, I used a conditional fixed effect negative binomial model and a conditional fixed effect Poisson model. While some studies have successfully used the conditional negative binomial model for panel estimation of overdispersed count data (e.g., Hausman et al. 1984; Furman and Stern 2011), it has been reported that the conditional fixed-effects negative binomial model is not a true fixed-effects model because it fails to control for all of the predictors that are fixed over time (Allison and Waterman 2002; Guimaraes 2008; Hilbe 2007). An alternative is to use a conditional fixed effect Poisson model but handle overdispersion of data by bootstrapping the sample without assuming any distribution of data or using a quasi-maximum-likelihood estimator to estimate a robust standard error (Hilbe 2007). As there are trade-offs in using one over the other model specification, I present the result using all of them in the first analysis. For the following analyses, I present the results from using only the conditional fixed effect Poisson model with robust standard error. For some analyses on the heterogeneous effect when the fraction was the dependent variable, I used a time-series ordinary least square (OLS) model.

Data

Data Construction and Source

The data source I relied on for this study is the SSRN, complemented with the Web of Science. The SSRN was established in December of 1993 by Social Science Electronic Publishing Inc. to facilitate worldwide dissemination of social science research. Since then, the number of archived papers and delivered downloads has increased exponentially (Figures 2 and 3). For the year from May 2010 to May 2011, SSRN received 56,000 papers and delivered 8.6 million downloads. As of August 2010, SSRN had archived 298,243 research articles, of which 189,625 articles had full texts free of charge. Downloading and posting a research article on SSRN is free and open to anyone. However, a research organization is charged when SSRN hosts a research paper series for the organization. In addition, certain user services are charge-based: for example, an email alert or delivery service for research articles on certain topics or written by certain authors, suited to users' preferences, is provided to users at a charge. SSRN records posting and revision date of posted articles, tracks citations and number of downloads even before citing or cited articles are published in traditional scholarly journals, and identifies whether some papers in multiple versions are in fact the same paper, removing any erroneous counts of citation or posting of the same paper.

SSRN does not report whether their posted articles are published in refereed journals unless the authors or the organizations indicate it. In order to identify the publication status of SSRN articles, I matched the title and the names of authors with those in Web of Science and collected the information on the publication status. If they were identified as published, I collected data on the total citations, the publication source, the names of authors of citing papers, and the publication source of the citing papers. The matching method I used is hardly perfect: some research articles from SSRN may have been published with slightly different titles and erroneously identified as unpublished. The imperfect matching error can lead to two cases: 1) some published SSRN articles may be excluded from the study erroneously or 2) they may be categorized as control articles erroneously if they happened to be published in the same journal, volume, and issue as the other SSRN articles identified as published. In the first case, the exclusion of those published articles should not affect the estimate in one way or the other as their exclusion from the study is random. The second case can lead to a biased estimate on the effect of the free access. The estimate to which this error leads is, however, an underestimate, not an overestimate, of the effect of free access. The difference in citations between the control and treated articles that I observe may be smaller than the true difference without the error because the control articles are contaminated by the treated articles. What I report from the analysis is, therefore, going to be a downward bias, if any, due to this imperfect matching.

The identification strategy exploits a unique feature of SSRN's practice of posting articles. While authors can post their papers at any time of their choice, there is a general trend of a large number of papers submitted to SSRN for posting at once by organizations, especially when SSRN starts a new paper series for the organizations. As Figure 4 shows, the number of newly submitted papers to SSRN per month spikes whenever an organization starts a new research working paper series or submits a large number of papers for the series. This spike in submissions of papers from organizations is also illustrated in Figure 5. For example, an organization, A, submitted 445 papers in one day, May 4, 2000, when it started a new research paper series. The timing of posting these papers is decided by the organization or SSRN, not by the authors of these papers. In other words, the timing of posting is exogenous to the quality of papers. Therefore, the increase in citations after posting on SSRN can be attributed to either the time trend or free accessibility available on SSRN. The time trend is accounted for by matching the SSRN articles with non-SSRN articles that were published in the same journal, volume, and issue as described in more detail in the next paragraph. I identified 13,000 articles that were posted in a large number at once, at least more than 100 articles from the same organization in the same month. I confirmed with SSRN that these articles were posted either by SSRN or the organization, not by the authors, typically at the start of a new research paper series for the organization. Among those articles, I chose 385 articles that had been published at least 4 years prior to the posting year on SSRN.

When SSRN articles were identified as published in refereed journals and subsequently chosen to be included in this study, their publication source such as journal name, volume, issue, and publication date was also identified from Web of Science. Once the publication source was identified, I collected the title, the total citations, and the authors of the articles, other than SSRN articles, published in *the same publication source* as the SSRN articles. I chose those articles that were published in the same journal, volume, and issue as SSRN articles were published as "control articles," counterparts to the SSRN articles which are thought to be the "treated" articles. The treatment is whether the articles are posted on SSRN after having been published in refereed and non-free journals for at least 4 years: the treated articles are posted on SSRN and the control articles are not, while they both were published in the same journal, volume, and issue at the same publication time. I compiled the list of the control and treated articles with their titles, authors, and publication sources. From each article in the complied list, I collected data from Web of Science on citing papers such as name of authors and publication source if each article in the complied list received any citation from other articles published in journals that Web of Science tracked. The final data set I compiled from both Web of Science and SSRN included counts of self-citations over time for both treated and control articles, counts of non-self citations over time for both treated and control articles, posting dates on SSRN for treated articles, posting organizations for treated articles, publication sources for both treated and control articles, and names of authors for both treated and control articles.

Supply Side Analysis

For the analysis of the heterogeneous effect on the supply side, I divided the sample into two groups in various criteria. The first criterion was the Journal Impact Factor (JIF) published by Web of Science as of 2006. Some articles were published in the journals where the Journal Impact Factor was not known in 2006. Excluding those articles, I compiled 283 articles and they were grouped into the two: one is published in the journals above the median JIF and the other published in the journals below the median JIF. The two different groups were analyzed separately or with an interaction term in order to investigate the heterogeneous effect of the free on-line repository on the citations, which may differ depending on the prestige and thus the degree of the distribution of the journals where the articles were originally published. The second criterion was the quality of the institute with which the authors were affiliated. From the list used by Kim et al.(Kim et al. 2009), the top 25 institutes in the field of finance were identified. A binary variable was set to be 1 for the article published in the field of finance (and economics) and authored by at least one scholar affiliated with the top institute. The last criterion was the quality of the authors. Using the ranking of economists published by Repec as of 2011, I assigned a binary variable to be 1 for an economist if she is in the top 10% economists and 0 otherwise. I also selected the articles only published in the economics journals listed under the subject of economics by Web of Science.

Demand Side Analysis

The readers or the citing scholars were also grouped in various ways. The first was the countries to which the citing authors' affiliations belonged. I collected the address of the authors of all the articles citing the treated and control articles. This analysis is to test a hypothesis that the free access will give the benefit to the readers in developing countries more than those in developed countries. Using the definition of the developed country by Wikipedia, I marked the country in the citing author's address as either developing or developed country. The fraction of the citing authors in all citing authors was estimated in two ways: one was to count the number of citing articles authored by scholars all from the developing countries and get the fraction of the number of those articles among the number of all the citing articles in each year; the second was to count the number of authors from the developing countries in each citing article, get the fraction of those authors in each citing article, and average the fraction over each year. The fractions in both estimates were dependent variables in this analysis, unlike the previous analyses, and a time-series OLS was used for the estimation of the SSRN effect.

The second criterion was the distance between the cited author and citing author. The research question was whether the SSRN-posting makes the distance closer or further apart. The distance was measured both in knowledge base and in geography. For the analysis on the knowledge distance, the articles published in journals of the three fields only, which are economics, finance, and law. The list of the journals under each field was obtained from Web of Science. Economics and finance were considered as one field as there are many journals cross-listed under the two subjects and law was considers as the other field. Only when the cited article was published in the journal in the fields of either economics/finance or law, the article was selected for the analysis. Then the source of the citing articles for the selected articles was parsed and checked as either "within" or "across". The citing article was marked as "within" if it was published in the journal in the same field where the cited article was published and as "across" otherwise.

The geographic distance between the cited and the citing author was estimated only between the reprint authors or the first authors if the reprint author is not available. The articles were selected only when those authors' addresses were in the USA. The distance, averaged over each year, was regressed on the SSRN variables. A time-series OLS was used for this analysis, because the dependent variable was not a count variable as citations in the previous analysis.

Descriptive Statistics

The descriptive statistics on both treated and control are shown in Tables 1, 2, and 3. The numbers of SSRN articles and of their matched control samples are 385 and 3,820, respectively (Table 1). The total citations that SSRN articles received, 47.2, was twice as high as those of their matching control articles, 24.4. The publication year of the SSRN articles ranged from 1970 to 2006. The average number of years for which the SSRN articles had been published when they were posted on SSRN was 10.3. The numbers of the journals and issues in which these SSRN articles were published were 165 and 337, respectively.

In order to show the differences in the characteristics of the SSRN and their control articles before posting on SSRN, I tabulated the descriptive statistics on those before and after the posting (Table 2 and 3) separately. The SSRN articles received more citations (2.0 on average, Table 2) than their control articles (1.1 on average, Table 2), even prior to the posting year. The difference in cumulative citations, which are non-self citation counts that the articles received up to the posting year of the treated articles, is more pronounced: 12.5 for the SSRN articles and 6.8 for the control articles. These differences between the SSRN articles and their control articles even prior to the posting suggest that the SSRN articles are higher quality than their control articles. After posting on SSRN, the differences in both citations per year and cumulative citations between the SSRN articles and their control articles seem to become greater, suggesting that the posting may cause the increased gap in citation counts between the SSRN articles and their control articles (Table 3). The effect of the posting on the citation counts is quantified by the empirical equations described in the next section.

Results and Discussion

Average Effect

The difference in citations between SSRN-articles and non-SSRN articles before and after posting year is quite clear, as shown in Figure 5. The first graph (Figure 5a) shows all the samples, where the posting year of all of the SSRN articles is set to be zero. At the year -20, the citation of the SSRN articles seems to increase with time more than their matching control samples (Figure 5a). Therefore, excluding the articles which were published longer than 20 years before being posted on SSRN, I graphed the citation changes over time of SSRN articles and their matching control articles (Figure 5b). Even prior to being posted on SSRN, the SSRN articles showed a higher number of citations than their matching control articles. This finding is consistent with the reports of numerous studies that articles with free access tend to be of higher quality (e.g., Davis and Fromerth 2007; Kurtz et al. 2005). In this setting, the authors did not choose the timing of posting but the authors may have chosen which of their articles would be posted on SSRN. Even if it was the authors' affiliated organization that chose which articles to post, it is likely that they chose better articles for posting. Many researchers reported that the selection bias may explain the observed difference in citations between open access articles and other articles (e.g., Schwarz and Kennicutt 2004; Davis and Fromerth 2007; Kurtz et al. 2005; Moed 2007; Metcalfe 2005). The first difference in citation between the treated and the control articles before posting on

SSRN can, therefore, indicate the inherent quality difference between the treated and the control. The second difference between pre-posting and after-posting, however, can be attributed to the posting on SSRN after the natural citation trend with aging is accounted for by the matching control articles and the quality difference between the treated and control articles is measured by the first difference, because the timing of posting was not chosen by the authors.

For a statistical analysis, I used a difference-in-difference method for panel data (Wooldridge 2007) as in the following empirical equation, similar to what was used by Furman and Stern (2011):

$$Cite_{igt} = f(\varepsilon_{igt};\ \alpha_g + \lambda_t + \beta_1 SSRN_i + \beta_2 (SSRN \times After\_Posting)_{it}) \qquad \text{---- (1)}$$

where $Cite_{igt}$ is citation counts that an article, *i*, received at a year, *t*, when it was published in a journal, volume, and issue, *g*. The subscripts *i, g,* and *t* indicate article, group, and time, respectively. Each group means the same journal, volume, and issue. The α_g and λ_t indicate a fixed effect for the group and citation year, respectively. *SSRN* is a binary variable, 1 if posted on SSRN at some point and 0 if not posted on SSRN. This variable is time-invariant and for all time periods it is either 1 or 0. *After_Posting* is a time-variant binary variable, equal to 1 only for years after the treated article is posted on SSRN and 0 otherwise. In this specification, I am interested in not only the effect of posting on SSRN, which is captured by the interaction term, *SSRN x After_Posting*, but also inherent differences between SSRN and non-SSRN articles, captured by the term, *SSRN*, alone. In order to show the average difference in citation counts between SSRN articles and non-SSRN articles even prior to being posted on SSRN in this specification, I included as control articles all of the research articles published in the same journal and issue as a SSRN article was published. The coefficient β_1, for the binary variable, *SSRN*, captures the possible differences between the SSRN articles and non-SSRN articles prior to being posted on SSRN.

In the conditional fixed-effect negative binomial model (4-1 in Table 4), SSRN articles appear to receive 164.5% of citations of their matching control articles, even prior to being posted on SSRN, consistent with the earlier figure. The coefficient for (*SSRN x After_Posting*), 0.158 or 1.171 as the exponentiated value, tells that the SSRN articles gained an additional 17% citation counts after being posted on SSRN compared to their counterpart control samples that were never posted on SSRN. In the model 4-2 and 4-3 where a conditional fixed effect Poisson model was used, the estimate on the coefficient for (*SSRN x After_Posting*) was 0.099 (Model 4-2 and 4-3). The standard error for the model 4-3 becomes large because the model accommodates distribution of data other than Poisson. Nonetheless, the posting on SSRN seems to increase the citation counts over 10% across all of the models at a statistical significance level of $p<0.10$. I attribute this gain to free access offered by SSRN. Among the three potential factors to increase citations for articles with open access identified by previous researchers, which are free access, early exposure, and quality difference, I excluded the early exposure factor because all of these articles were already published before posted on SSRN. Conditional on the assumption that the quality of articles is not correlated with the timing of the posting, the quality difference should be accounted for by the coefficient for *SSRN* but not by the coefficient for *SSRN x After_Posting*. The control articles may be available as well for free access somewhere other than SSRN. If this is the case, what is estimated by the *SSRN* coefficient in this model is an underestimate, not an overestimate of the effect of free access. It is, however, possible that what SSRN provides is not a passive free access to a research article but an active promotion. Knowing that there is no barrier to access to posted articles, the authors or the organizations that the authors are affiliated with may cite their own articles more than they would otherwise and put a link to their articles on SSRN whenever they cite these articles. In addition, SSRN provides some services to users to draw attention to popular papers or papers to suit users' specific interests. This kind of service may give additional readership for the articles posted on SSRN and increase citations. However, the promotion is not a cause but a consequence of free access.

While the above specification, (1), provides an estimate of the difference between SSRN articles and non-SSRN articles, the potential for substantial heterogeneity among articles (even though they are published in the same journal, volume, and issue) may lead to a biased estimate of the impact of SSRN posting on subsequent citation. Therefore, the article-specific fixed effect (c_i) is included as in the following specification:

$$Cite_{igt} = f(\varepsilon_{igt};\ c_i + \lambda_t + \delta_{t-pubyear} + \beta_3(SSRN\ \times After\_Posting)_{it}) \qquad \text{----(2)}$$

This specification tests for the impact of posting on SSRN by estimating the changes in citations after an article is posted on SSRN. The age and time effect which may affect the citation counts are accounted for by including the year and age fixed effect, λ_t and $\delta_{t-pubyear}$, along with the control articles with similar characteristics. In this specification, only one control article, among the non-SSRN articles published in the same journal and issue as the SSRN article, was selected to match one SSRN article. Two other criteria for the selection of a control article, in addition to being published in the same journal and issue as the SSRN article, were used: 1) the control article should have a similar citation-year profile for 4 years prior to the posting year of its matching SSRN article and 2) the control article should have total citation counts close to its matching SSRN article up to the posting year. If no article meeting the criteria is found to match a SSRN article, the SSRN article was excluded from the analysis. As a result, the number of SSRN articles included in this analysis was smaller than in the earlier analysis. The resulting articles consist of 145 SSRN articles and 145 control articles.

This specification was also tested with both a conditional fixed effect Poisson and negative binomial models. They were qualitatively similar and only the result from the conditional fixed effect Poisson model with robust standard error was presented in Table 5. The coefficient for *SSRN x After_Posting* was 0.122 or 112.9% (5-1). In other words, these articles gain approximately 13% in citation counts after being posted on SSRN. The magnitude is similar to what was obtained with the group fixed effect in the earlier specification (10% in Model 4-2 and 4-3). This interpretation, however, depends on the assumption that the SSRN and their control articles have the same aging profile. It is possible that SSRN articles may have longer-lived citation profiles, which would result in an upward bias on the estimate of *SSRN x After_Posting*. To address this possibility, I include a separate linear time trend term for SSRN articles, *SSRN x Age,* in (5-2) while all the other dummy variables are included as in (5-1). The coefficient for *SSRN x Age* is insignificant while the coefficient for *SSRN x After_Posting* increases, suggesting that the differences in citation profiles between SSRN articles and control articles do not cause an upward bias on the estimate of the posting effect.

In the next two models, 5-3 and 5-4, I estimate the posting effect only with SSRN articles, excluding the control articles. In the panel analysis, it is common not to include control samples. As the time-invariant fixed effect of an article is differenced out from the estimating equation, the citation change with time can be attributed to the posting on SSRN. To exclude the control articles, however, one should assume an underlying citation-age profile common to all articles. For example, McCabe and Snyder (2011) assumed citation counts to be a concave function of age, and Furman and Stern (2011) specified one of their models with a concave function of age and a polynomial expansion of calendar year. Following the functional forms in these previous studies, I included publication age and its square term in the model (5-3) along with calendar year dummy variables. The coefficient for *SSRN x After_Posting* increases in this model as the coefficients for both age and age-squared term are negative. In the next model, (5-4), a polynomial expansion of year variable was included in place of calendar year dummy variables. In both models, the coefficient for *(SSRN x After_Posting)* was significant at $p<0.05$. It seems that the estimate on the coefficient for *(SSRN x After_Posting)* seems robust to different model specifications, suggesting that the effect of free access on the diffusion of scholarly ideas is statistically significant as predicted by the theory.

Heterogeneous Effect on Supply Side

The benefit of the open access may differ across the authors of the articles posted on SSRN. Because the articles of this study were all published in a journal before being posted on SSRN, the prestige of the journals where the articles were published was known. The Journal Impact Factor changes with year and JIF in the year of 2006 was used for this analysis. The median of the Journal Impact Factor of the journals where the sample articles were published was 1.92. Approximately 100 articles were published in a journal, of which Journal Impact Factor (JIF) was not available in the year of 2006. Excluding those articles, the articles were grouped into two: one is published in the journals above 1.92 of Journal Impact Factor and the other published in the journals below 1.92 of Journal Impact Factor. As shown in Table 6, the overall effect of the free access on the citation becomes statistically insignificant due to a smaller size of samples (6-1). The effect of free access on the articles published in the below-median JIF journals is, however, much stronger and statistically significant (6-2 and 6-3). To illustrate the difference in the effect between the low-tiered journal and the high-tiered journal, an interaction term was included in the model 6-4 and the SSRN-posting effect was much smaller in the articles published in the journals above median JIFs. The result suggests that the free access provides an additional outlet for the articles published in less distributed journals rather than further promotes the articles that were already well-known.

It can be still the case, however, that the articles published in low-tiered journals by high-profile authors receive a disproportionate boost from the SSRN-posting. This question is answered in the table 7, 8, and 9. Among the articles published in the field of finance, the authors affiliated with top institutes seem to have a much larger boost in the citations from the SSRN-posting (7-2). When the articles were selected when they were published in a combined list of journals in the economics and finance, the effect of top institute becomes statistically insignificant (8-2). The SSRN tends to host working paper series from the top institutes and the model 8-2 shows that the SSRN effect becomes insignificant when the top institute variable was included, suggesting that the quality differentials between SSRN and non-SSRN articles can be captured by that the author is affiliated with top institutes or not. A similar effect is observed for the high profile researchers. Instead of categorizing the institutes as in Table 7 and 8, I ranked researchers themselves and divided them into a group of top 10% as ranked by Repec as of 2011 and the rest. The top 10% economists do not seem to have a boost from the SSRN-posting (9-2 and 9-3). They receive a higher citation count regardless of the SSRN. There is, however, a limitation of this analysis: the ranking of the economists change with time. I used the ranking from the year 2011, while the publication years and posting years articles in this analysis range from 1990s to 2006.

Heterogeneous Effect on Demand Side

On the demand or the readers, the benefit of the open access may not be the same. While it is known that the scholars in the developing countries have a limited access to the academic journals compared to those in the developed countries (GaulÃ© 2009), the differential effect of the open access advantage across countries of different income has not been reported yet. I collected the addresses of the citing authors and identified the countries where they are located. If all authors of a citing article are affiliated with organizations in the developing countries, the citing article is marked as one from the developing countries. If at least one author of a citing article is affiliated with organizations in the developed countries, the citing article is marked as one from the developed countries. After marking all of the citing articles either as one from the developing countries or as one from the developed countries, the ratio of those from the developing countries to all the citing articles each year was estimated. Similarly to the regression of (1), the ratio was regressed on *SSRN* and *SSRN x After_Posting*. In this analysis, the dependent variable is no longer a count variable but a fraction and a time-series OLS was used. As shown in Table 10, the ratio increases upon the SSRN-posting. The prestige of the journals does not appear to make a difference (10-2). In the next analysis, the fraction of the authors from the developing countries in all the authors in each citing article was estimated. For example, a citing article was authored by two scholars; one from developing country and the other from developed country. Then the fraction is 0.5. This fraction was obtained for each citing article and averaged over each year. The result of the regression, when the fraction was the dependent variable, shows that, upon SSRN-posting, the fraction increases (11-1). This increase is mostly driven by the low-tiered journals (11-2, 11-3, and 11-4), because when the articles are divided into two groups, one published in journals with the above-mean JIF and the other published in journals with the below-mean JIF, the ratio only in the articles published in the below-mean JIF increases upon the SSRN-posting (11-2). The result suggests that the scholars or citing authors that have a limited access to the articles published in journals distributed less widely get an access to those articles upon the SSRN-posting.

The open access may promote the citation across the field more than the citation within the field. The scholars may get access to the journals in their own field whether they are available in their affiliated organizations or not through other channels such as contacting the authors and their peers. If this is the case, the open access may promote the citation across the field. To see this effect, the samples were first reduced only to those published in the journals in the field of economics, finance, and law. The reduced samples were regressed as before and compared to the all samples (12-1 for all samples and 12-2 for reduced samples). The SSRN-posting effect is insignificant in both samples. The next analysis was to make the citations only by the articles in the same field per year as the dependent variable (12-3).

The result does not show the SSRN-effect. The next model (12-4) uses a time-series OLS with the fraction of within-field citing to all citing articles being the dependent variable. Again, the SSRN-effect does not appear to affect the fraction.

The spillover of knowledge over geographic distance has been studied by many researchers (i.e., Griffith et al. 2011;Abramovsky and Simpson). I investigate whether the open access changes the distance between the citing and the cited authors. As shown in Table 13, the geographic distance between the citing and cited authors does not change upon the SSRN-posting. By the time that these articles are posted on SSRN, they were published already for 4 years by the design of this study. In addition, the samples included in this analysis were only those with both citing and cited authors being located in the USA. It is not, therefore, surprising that the distance is not affected by the SSRN-posting.

Other Potential Factors

The results shown both in Table 4 and 5 are the increased citation upon posting on SSRN or SSRN-effect. Although I attribute the SSRN-effect to free access, there are other potential effects associated with posting on SSRN, except the early exposure and the selection bias that this study controlled for. The first is a low search cost associated with SSRN. SSRN is a repository, providing a database of research articles and allowing an easy search for a research article. Even if a research article is freely accessible at other sites such as its author's personal webpage, the free article may not be easily searchable and thus not be cited as it would be if posted on SSRN. This effect is not due to free access *per se*, but due to low search cost. This argument would be applicable to unpublished SSRN articles that are not available in other widely used database. The SSRN articles in this study are, however, already published at least for four years and easily searchable in the Web of Science, a more commonly used and much more exhaustive database of published research articles. The way with that I identified the publication source of a research article posted on SSRN was to match the title of the SSRN article and its authors to the Web of Science database. Therefore, by design, the SSRN article included in this study had to be searchable by the Web of Science. For a citing author to locate an old published research article only because it is available in SSRN although it is also searchable in the Web of Science, she must have an access to SSRN but not to the Web of Science. The difference between the two databases in this context is not a difference in the search cost but a difference in the access cost. SSRN is free to any user while the Web of Science is only available to subscribing individuals or the users affiliated with subscribing organizations. The second is a marketing or promotion of the research articles by the SSRN. SSRN not only provides a passive outlet to post research articles but send a personalized email to its users for newly posted articles. The increased citation may be due to this new promoting effect by SSRN not by the free access. As shown in Table 6, the boost in the citation is observed only in the articles published in the low-tiered journals. If the promotion had increased the citation, the effect would be even more pronounced for the articles published in the high-tiered journals. Therefore, the wider access to the articles is a likely cause for the boost.

Conclusion

The main contribution of this study is to report the heterogeneous effect of the open access as well as the causal relationship between free access and citations. In theory, free access to ideas should help their diffusion, and research articles with free access should receive more citations, a proxy for diffusion. However, previous empirical studies have not been able to separate the effect of free access from selection bias and have reported inconsistent findings from no or negative effect to an over-300% increase vs. citations of non-free articles. By using a natural experiment that estimates the effect of free access separate from that of confounding factors, this study identifies the effect of free access to research articles on citation counts. The effect is not as large as some previous studies have reported. It is less than 20%. When the sample size was reduced, the effect became statistically insignificant. The selection bias or the quality differential explained the most of the increased citations of the open access articles.

However, the effect was not heterogeneous across the authors and readers. The articles previously published in low-tiered journals drove the effect while the articles published in high-tiered journals did not get a boost in citations from the free access. The articles published in the journals in the field of finance and authored by scholars affiliated with top institute tend to get a boost from the SSRN-posting. After the free posting of the articles, their citing authors tend to become more from developing countries than they were before the free posting. The distance in both knowledge and geography between the citing and cited authors doesn't seem to change with the free posting of the articles.

References

Abramovsky, L., and Simpson, H. 2011. "Geographic Proximity and Firm-University Innovation Linkages: Evidence from Great Britain," *Journal of Economic Geography* (11:6), pp. 949-977.

Allison, P.D., and Waterman, R.P. 2002. "Fixed-Effects Negative Binomial Regression Models," *Sociological Methodology* (32:1), pp. 247-265.

Antelman, K. 2004. "Do Open-Access Articles Have a Greater Research Impact?," *College & research libraries* (65:5), p. 372.

David, P.A. 2005. "From Keeping Nature's Secrets to the Institutionalization of Open Science," *Code: Collaborative ownership and the digital economy*), pp. 85-108.

Davis, P.M., and Fromerth, M.J. 2007. "Does the Arxiv Lead to Higher Citations and Reduced Publisher Downloads for Mathematics Articles?," *Scientometrics* (71:2), pp. 203-215.

Davis, P.M., Lewenstein, B.V., Simon, D.H., Booth, J.G., and Connolly, M.J.L. 2008. "Open Access Publishing, Article Downloads, and Citations: Randomised Controlled Trial," *BMJ: British Medical Journal* (337).

Donovan, J.M., and Watson, C.A. 2011. "Citation Advantage of Open Access Legal Scholarship," *SSRN eLibrary*).

Eysenbach, G. 2006. "Citation Advantage of Open Access Articles," *PLoS biology* (4:5), p. e157.

Furman, J., and Stern, S. 2011. "Climbing Atop the Shoulders of Giants: The Impact of Institutions on Cumulative Knowledge Production," *American Economic Review* (101:5), pp. 1933-1963.

GaulÃ©, P. 2009. "Access to Scientific Literature in India," *Journal of the American Society for Information Science & Technology* (60:12), pp. 2548-2553.

Gaule, P., and Maystre, N. 2011. "Getting Cited: Does Open Access Help?," *Research Policy*).

Griffith, R., Lee, S., and Van Reenen, J. 2011. "Is Distance Dying at Last? Falling Home Bias in Fixed-Effects Models of Patent Citations," *Quantitative economics* (2:2), pp. 211-249.

Guimaraes, P. 2008. "The Fixed Effects Negative Binomial Model Revisited," *Economics Letters* (99:1), pp. 63-66.

Harnad, S., and Brody, T. 2004. "Comparing the Impact of Open Access (Oa) Vs. Non-Oa Articles in the Same Journals," *D-lib Magazine* (10:6).

Hausman, J., Hall, B.H., and Grtltches, Z. 1984. "Econometric Models for Count Data with an Application to the Patents-R&D Relationship," *Econometrica* (52:4), pp. 909-938.

Heller, M.A., and Eisenberg, R.S. 1998. "Can Patents Deter Innovation? The Anticommons in Biomedical Research," *science* (280:5364), p. 698.

Hilbe, J.M. 2007. *Negative Binomial Regression*. Cambridge: Cambridge University Press.
Kim, E.H., Morse, A., and Zingales, L. 2009. "Are Elite Universities Losing Their Competitive Edge?," *Journal of Financial Economics* (93:3), pp. 353-381.
Kurtz, M.J., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M., Henneken, E., and Murray, S.S. 2005. "The Effect of Use and Access on Citations," *Information Processing & Management* (41:6), pp. 1395-1402.
Kurtz, M.J., and Henneken, E.A. 2007. "Open Access Does Not Increase Citations for Research Articles from the Astrophysical Journal," *Arxiv preprint arXiv:0709.0896*).
Lawrence, S. 2001a. "Free Online Availability Substantially Increases a Paper's Impact," *Nature* (411:6837), p. 521.
Lawrence, S. 2001b. "Online or Invisible," *Nature* (411:6837), p. 521.
McCabe, M.J., and Snyder, C.M. 2011. "Did Online Access to Journals Change the Economics Literature?," *Social Science Research Network (SSRN), January* (23).
Metcalfe, T.S. 2005. "The Rise and Citation Impact of Astro-Ph in Major Journals," *Arxiv preprint astro-ph/0503519*).
Moed, H.F. 2007. "The Effect of "Open Access" on Citation Impact: An Analysis of Arxiv's Condensed Matter Section," *Journal of the American Society for Information Science and Technology* (58:13), pp. 2047-2054.
Mokyr, J. 2002. *The Gifts of Athena: Historical Origins of the Knowledge Economy*. Princeton Univ Pr.
Norris, M. 2008. "The Citation Advantage of Open Access Articles,").
Rosenberg, N. 1963. "Technological Change in the Machine Tool Industry, 1840â€"1910," *The Journal of Economic History* (23:04), pp. 414-443.
Rosenberg, N. 1979. "Technological Interdependence in the American Economy," *Technology and Culture* (20:1), pp. 25-50.
Schwarz, G.J., and Kennicutt, R.C. 2004. "Demographic and Citation Trends in Astrophysical Journal Papers and Preprints," *Arxiv preprint astro-ph/0411275*).
Wooldridge, J. 2007. "What's New in Econometrics? Imbens/Wooldridge Lecture Notes; Summer Institute 2007, Lecture 10: Difference-in-Differences Estimation," *NBER. http://www. nber. org/minicourse3. html, last accessed at March* (19), p. 2009.
Zuckerman, H., and Merton, R.K. 1971. "Patterns of Evaluation in Science: Institutionalisation, Structure and Functions of the Referee System," *Minerva* (9:1), pp. 66-100.

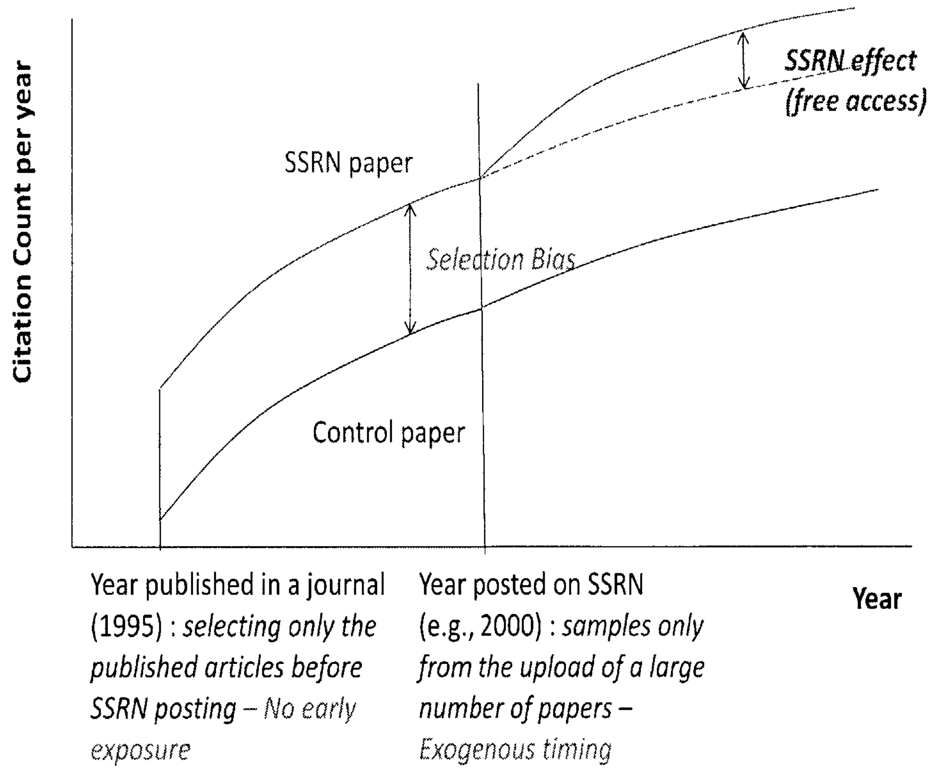

Figure 1. Identification Strategy

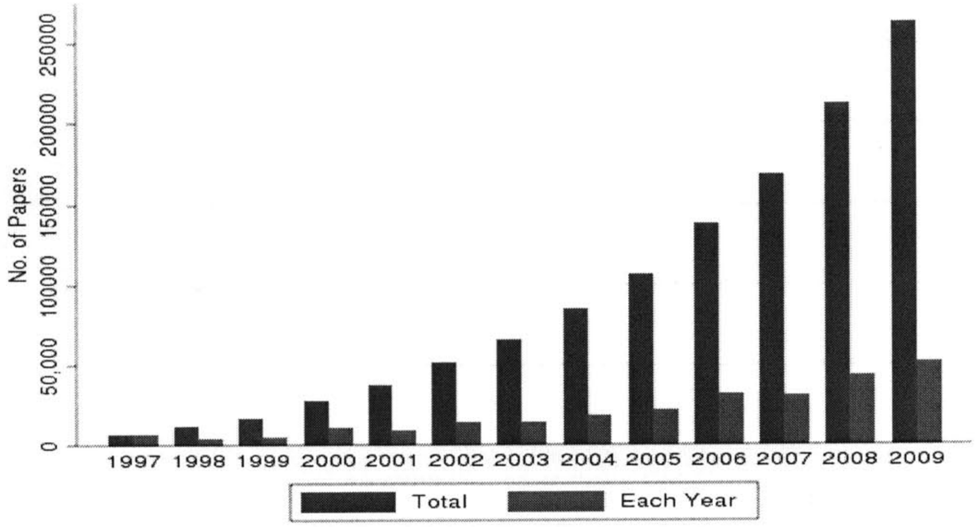

Figure 2. Number of papers posted on SSRN.

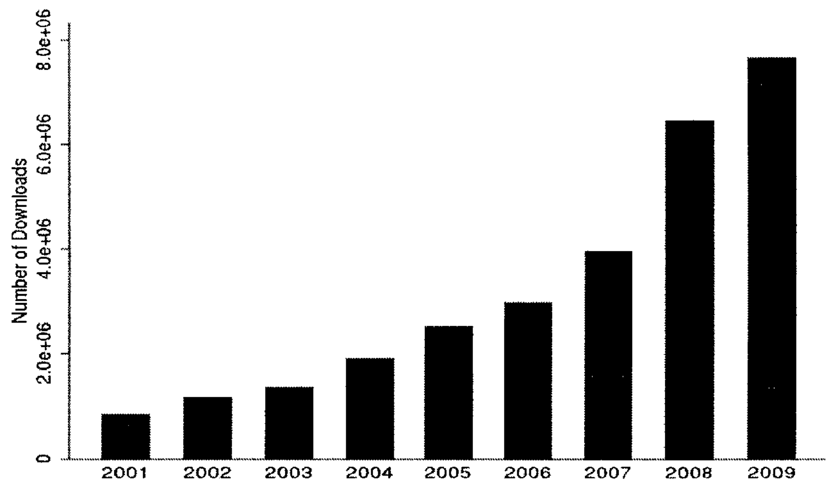

Figure 3. Number of downloads.

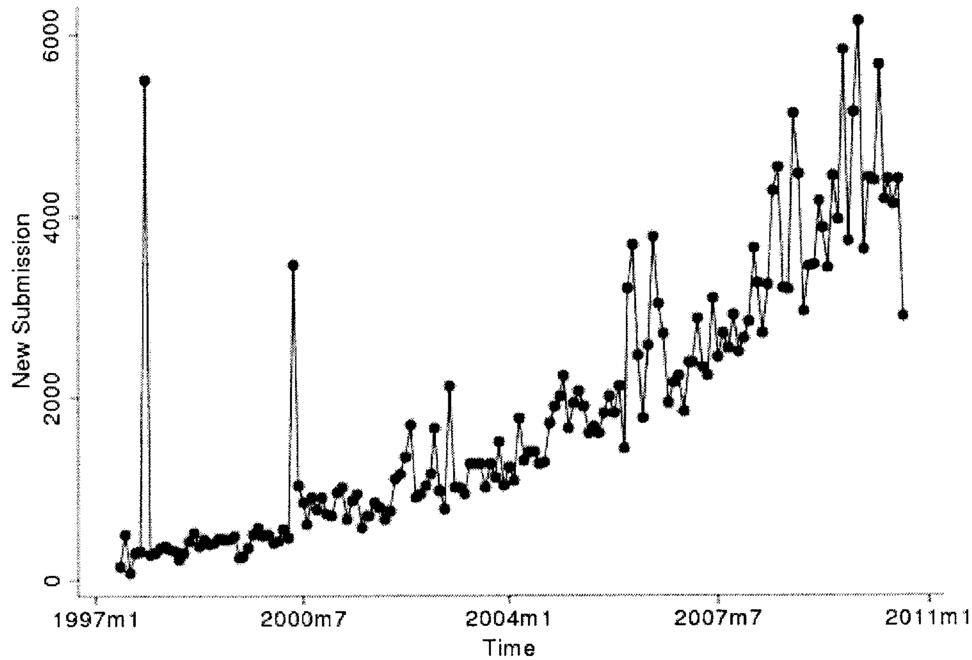

Figure 4. Number of new papers posted on SSRN per month.

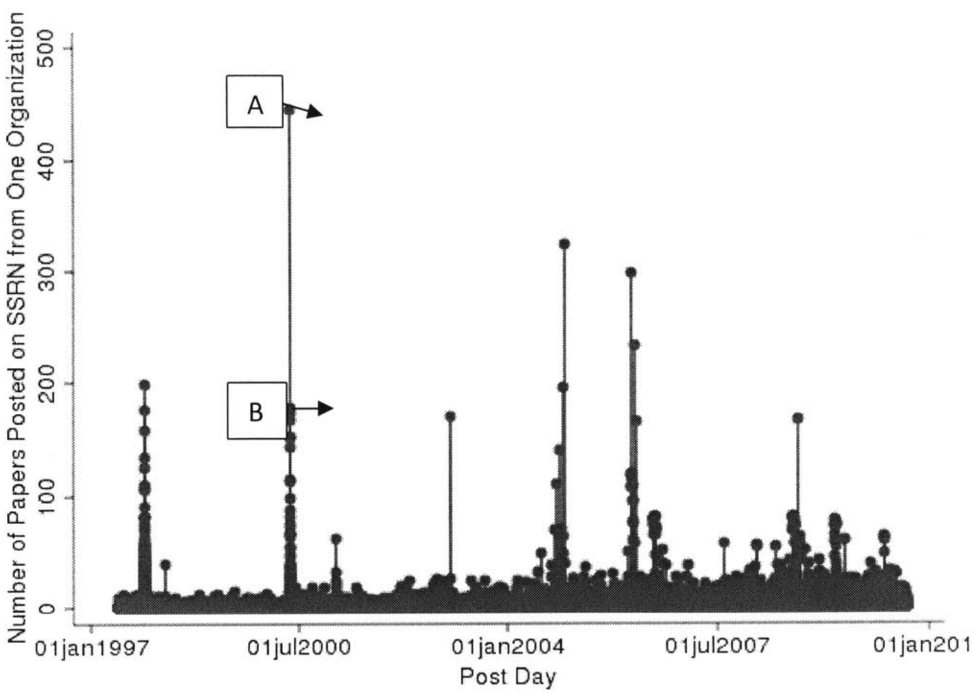

Figure 5. Number of new papers submitted from one organization in one day. Each circle represents the number of papers submitted from one unique organization on that day. For example, an organization, A, posted 445 papers in May 4, 2000, and an organization, B, posted 178 papers in the same day.

Table 1. Article characteristics for samples used in the longitudinal study. Control articles were drawn from the same journal, volume, and issue where SSRN-articles were published.

Treated Samples (SSRN articles, n=385)	Mean	Std.	Min.	Max.
Total citations up to 2010 since published	47.2	126.9	0	1898
Publication year	1994.5	7.3	1970	2006
Year posted on SSRN	2004.8	3.7	2000	2010
Years since publication when posted on SSRN	10.3	6.6	4	35
Number of journals where sample articles were published	165			
Number of Journal/Vol/Issue where sample articles were published	337			
Observations	385			
Control Samples (non-SSRN articles, n=3820)	Mean	Std.	Min.	Max.
Total citations up to 2010 since published	24.4	62.9	0	1387
Publication year	1992.9	7.4	1970	2006
Year posted on SSRN	Not Applicable			
Years since publication when posted on SSRN	Not Applicable			
Number of journals where sample articles were published	165			
Number of Journal/Vol/Issue where sample articles were published	337			
Observations	3820			

Table 2. Article-year characteristics BEFORE posting on SSRN for samples used in the longitudinal study. Control articles were drawn from the same journal, volume, and issue where SSRN articles were published.

	Treated Samples (SSRN articles)			
	Mean	Std.	Min.	Max.
Citations per year*	2.0	5.2	0	117
Cumulative citations	12.5	35.7	0	740
Year	1997.2	6.8	1971	2009
Years since publication	6.8	6.2	0	34
Years since posting on SSRN	-7.8	6.2	-35	-1
Observations	3979			
	Control Samples (Non-SSRN articles)			
	Mean	Std.	Min.	Max.
Citations per year	1.1	3.0	0	102
Cumulative citations	6.8	19.1	0	804
Year	1996.1	6.8	1971	2009
Years since published	6.9	6.1	0	34
Years since posting on SSRN	Not Applicable			
Observations	42053			

Table 3. Article-year characteristics AFTER posting on SSRN for samples used in the longitudinal study. Control articles were drawn from the same journal, volume, and issue where SSRN articles were published.

	Treated Samples (SSRN articles)			
	Mean	Std.	Min.	Max.
Citations per year	5.0	14.4	0	289
Cumulative citations	48.9	105.6	0	1898
Year	2006.6	2.8	2001	2010
Years since publication	14.4	6.4	5	40
Years since posting on SSRN	4.4	2.8	1	10
Observations	1998			
	Control Samples (Non-SSRN articles)			
	Mean	Std.	Min.	Max.
Citations per year	2.1	5.8	0	149
Cumulative citations	22.2	53.3	0	1387
Year	2006.3	2.8	2001	2010
Years since publication	15.6	6.5	5	40
Years since posting on SSRN	Not Applicable			
Observations	23235			

* Citation in all tables is non-self citation.

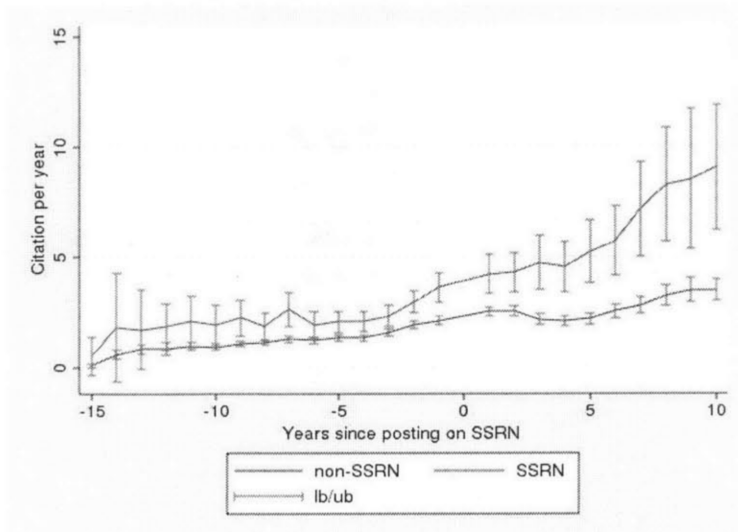

Figure 6. Citation-age profile (a) all samples; (b) a subset of SSRN and non-SSRN articles that were published less than 20 years before SSRN articles were posted on SSRN. The error bar is one standard deviation.

Table 4. Value of free access: SSRN effect for longitudinal samples,

	Conditional Fixed Effect Negative Binomial (4-1)	Conditional Fixed Effect Poisson (4-2)	Quasi-ML Poisson (4-3)
SSRN	0.498***	0.585***	0.585***
	(0.0620)	(0.0122)	(0.1207)
	[1.645]	[1.795]	[1.795]
SSRN x After_Posting	0.158***	0.099***	0.099*
	(0.0548)	(0.0162)	(0.0552)
	[1.171]	[1.104]	[1.104]
Constant	-0.771*		
	(0.3341)		
	[0.462]		
Group Fixed Effect	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes
N of article-years	71265	71265	71265
N of articles	4205	4205	4205
N of SSRN article-year	5977	5977	5977
N of SSRN articles	385	385	385
N of Journal/Vol/Is	337	337	337
N of Journal	165	165	165
Log-Likelihood	-95067	-141962	-141962

Exponentiated forms of coefficients (or Incidence-Rate Ratios) are reported in brackets.

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

Table 5. Value of Free Online Access with Article-Fixed Effect.

	Baseline diff-in-diffs specification (5-1)	Interacting SSRN articles with age (5-2)	Identification based only on variation within SSRN articles with age functions (5-3)	Identification based only on variation within SSRN articles with age and year functions (5-4)
SSRN x After_Posting	0.122* (0.0701) [1.129]	0.179* (0.0925) [1.196]	0.307** (0.1361) [1.360]	0.176** (0.0885) [1.192]
SSRN x Age		-0.008 (0.0129) [0.992]		
Age			-2.155*** (0.2610) [0.116]	-2.115*** (0.2473) [0.121]
Age-squared			-0.003*** (0.0010) [0.997]	-0.003*** (0.0010) [0.997]
Year				1.954*** (0.2475) [7.056]
Year-squared				0.004*** (0.0008) [1.004]
Age Fixed Effect	Yes	Yes	No	No
Calendar Year Fixed Effect	Yes	Yes	Yes	No
Article Fixed Effect	Yes	Yes	Yes	Yes
N of article-years	4425	4425	2153	2153
N of articles	290	290	145	145
Log-Likelihood	-3947	-3947	-2243	-2281

Exponentiated forms of coefficients (or Incidence-Rate Ratios) are reported in brackets.

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

Table 6. The heterogeneous effect on the supply I: Journal Impact Factor of the journals where the treated and control sample articles were published before posting year

DV=Citation per year	All articles	Articles published only in the journals with JIF <= 1.9	Articles published only in the journals with JIF > 1.9	All articles
	(6-1)	(6-2)	(6-3)	(6-4)
Citation per year				
SSRN	0.393**	0.571*	0.217	0.599*
	(0.1485)	(0.2464)	(0.1278)	(0.2397)
	[1.482]	[1.770]	[1.243]	[1.820]
SSRN x Post_SSRN	0.047	0.238*	-0.181	0.184
	(0.0828)	(0.1056)	(0.1093)	(0.1074)
	[1.049]	[1.269]	[0.835]	[1.202]
SSRN x High_JIF				-0.412
				(0.2718)
				[0.663]
SSRN x Post_SSRN x High_JIF				-0.315*
				(0.1452)
				[0.730]
N of article-years	27617	13359	14208	28266
N of articles	2908	1620	1282	3018
N of SSRN article-year	3235	1973	1262	3235
N of SSRN articles	283	195	88	283
N of Journal/Vol/Is	264	187	77	283
N of Journal	132	112	21	146
Log-Likelihood	-77083	-30986	-44950	-77610

Exponentiated forms of coefficients (or Incidence-Rate Ratios) are reported in bracket
*** p<0.001, ** p<0.01, * p<0.05

Table 7. The heterogeneous effect on the supply II: Citation of SSRN-articles authored by at least one scholar from top institutes (the field of Finance only)

	(7-1)	(7-2)
Citation per year		
SSRN	0.208	0.822*
	(0.1137)	(0.3469)
	[1.232]	[2.275]
SSRN x Post_SSRN	0.297*	-0.392*
	(0.1216)	(0.1830)
	[1.346]	[0.676]
Top Institute		0.590*
		(0.2480)
		[1.804]
SSRN x Top Institute		-0.772*
		(0.3686)
		[0.462]
SSRN x Post_SSRN x Top Institute		0.755***
		(0.2261)
		[2.128]
N of article-years	3284	3284
N of articles	299	299
N of SSRN article-year	438	438
N of SSRN articles	33	33
N of Journal/Vol/Is	31	31
N of Journal	9	9
Log-Likelihood	-9911	-9628

Exponentiated forms of coefficients (or Incidence-Rate Ratios) are reported in bracket
*** p<0.001, ** p<0.01, * p<0.05

Table 8. The heterogeneous effect on the supply III: Citation of SSRN-articles authored by at least one scholar from top institutes (the field of Economics and Finance)

DV= Citation per year	(8-1)	(8-2)
Citation per year		
SSRN	0.317**	0.213
	(0.1157)	(0.1903)
	[1.373]	[1.237]
SSRN x Post_SSRN	-0.041	-0.024
	(0.0901)	(0.1424)
	[0.960]	[0.977]
Top Institute		0.403***
		(0.0775)
		[1.497]
SSRN x Top Institute		0.024
		(0.2304)
		[1.025]
SSRN x Post_SSRN x Top Institute		-0.023
		(0.1648)
		[0.977]
N of article-years	22496	22496
N of articles	2255	2255
N of SSRN article-year	2409	2409
N of SSRN articles	194	194
N of Journal/Vol/Is	175	175
N of Journal	68	68
Log-Likelihood	-63889	-63007

Exponentiated forms of coefficients (or Incidence-Rate Ratios) are reported in bracket
*** $p<0.001$, ** $p<0.01$, * $p<0.05$

Table 9. The heterogeneous effect on the supply IV: High-Profile Researchers

	(9-1)	(9-2)	(9-3)
Citation per year			
SSRN	0.339**	0.130	0.125
	(0.1304)	(0.1364)	(0.1363)
	[1.404]	[1.138]	[1.133]
SSRN x Post_SSRN	-0.111	-0.189	-0.189
	(0.0955)	(0.1261)	(0.1261)
	[0.895]	[0.828]	[0.828]
High-Profile Researcher		0.343***	0.335***
		(0.0947)	(0.0954)
		[1.409]	[1.398]
High-Profile Researcher x SSRN		0.160	0.395
		(0.2022)	(0.2572)
		[1.174]	[1.485]
High-Profile Researcher x SSRN x Post_SSRN		0.072	0.118
		(0.1449)	(0.1737)
		[1.075]	[1.125]
High-Profile Researcher x SSRN x High-JIF			-0.386
			(0.2783)
			[0.680]
High-Profile Researcher x SSRN x Post_SSRN x High-JIF			-0.117
			(0.1815)
			[0.890]
N of article-years	20133	20133	20133
N of articles	2036	2036	2036
N of SSRN article-year	2158	2158	2158
N of SSRN articles	173	173	173
N of Journal/Vol/Is	155	155	155
N of Journal	64	62	64
Log-Likelihood	-55967	-55092	-54907

Exponentiated forms of coefficients (or Incidence-Rate Ratios) are reported in bracket

*** p<0.001, ** p<0.01, * p<0.05

Table 10. The Heterogeneous Effect on the Demand Side I: the fraction of the citing articles authored by scholars all from the developing countries each year

DV=The fraction of citing articles authored by scholars all from developing countries per year	(10-1)	(10-2)
SSRN	-0.009	-0.006
	(0.0054)	(0.0079)
	[0.991]	[0.994]
SSRN x Post_SSRN	0.016*	0.025*
	(0.0076)	(0.0105)
	[1.016]	[1.025]
SSRN x High_JIF		-0.007
		(0.0103)
		[0.993]
SSRN x Post_SSRN x High_JIF		
		-0.023
		(0.0134)
		[0.977]
Constant	0.652	0.651
	(0.3527)	(0.3524)
	[1.920]	[1.918]
N of article-years	27617	27617
N of articles	2908	2908
N of SSRN article-year	3235	3235
N of SSRN articles	283	283
N of Journal/Vol/Is	264	264
N of Journal	131	133
Log-Likelihood	6648	6652

*** $p<0.001$, ** $p<0.01$, * $p<0.05$

Table 11. The heterogeneous effect on the demand side of citation II: the fraction of the citing authors from developing countries in each citing article, averaged over all citing articles per year.

DV= Fraction of the developing countries of the citing authors' affiliations	(11-1) All articles	(11-2) All articles published in journals of JIF <= 1.92	(11-3) All articles published in journals of JIF > 1.92	(11-4) All articles
SSRN	-0.009 (0.0050) [0.991]	-0.012 (0.0074) [0.988]	-0.003 (0.0057) [0.997]	-0.015* (0.0074) [0.985]
SSRN x Post_SSRN	0.017** (0.0066) [1.017]	0.025** (0.0097) [1.026]	0.002 (0.0072) [1.002]	0.029** (0.0094) [1.030]
SSRN x High_JIF				0.014 (0.0093) [1.014]
SSRN x Post_SSRN x High_JIF				-0.031** (0.0116) [0.969]
Constant	-0.171 (0.1858) [0.843]	-0.367 (0.3324) [0.693]	0.093 (0.1119) [1.098]	-0.169 (0.1861) [0.844]
N of article-years	27073	12949	14075	27073
N of articles	2798	1550	1243	2798
N of SSRN article-year	3200	1951	1249	3200
N of SSRN articles	272	185	87	272
N of Journal/Vol/Is	247	174	73	244
N of Journal	126	102	18	123
Log-Likelihood	8421	2802	6002	8424

*** p<0.001, ** p<0.01, * p<0.05

Table 12. The heterogeneous effect on the demand side of citation III: Citations within or across disciplines for the cited articles published in the field of law, economics, or finance. Economics and finance are considered one field; law is the other field. If the cited article was published in the journals of economics/finance or law and the citing article was published in the journals in the same field, respectively, the citation was defined as a citation within the field.

	All citations per year (All articles) (12-1)	All citations per year (Articles only in the field of economics, finance, or law) (12-2)	Citations only by the articles in the same field per year (12-3)	Fraction of within-field citing (12-4)
SSRN	0.393**	0.307**	0.355**	0.017
	(0.1485)	(0.1119)	(0.1138)	(0.0151)
SSRN x Post_SSRN	0.047	-0.033	-0.002	0.020
	(0.0828)	(0.0878)	(0.0918)	(0.0132)
Constant				2.154***
				(0.3843)
N of article-years	27617	23316	23316	23316
N of articles	2908	2430	2430	2430
N of SSRN article-year	3235	2596	2596	2596
N of SSRN articles	283	225	225	225
N of Journal/Vol/Is	264	206	206	206
N of Journal	131	90	90	91
Log-Likelihood	-77083	-65029	-54945	-6925

*** p<0.001, ** p<0.01, * p<0.05

Table 13. The heterogeneous impact on the demand side of citation IV: Geographic distance between the reprint or the first author of cited articles and the reprint or the first author of their citing articles, averaged over each year

	Geographic distance between cited and citing author (All articles) (13-1)	Geographic distance between cited and citing author (Articles published in journals with JIF<=1.92) (13-2)	Geographic distance between cited and citing author (Articles published in journals with JIF>1.92) (13-3)	Geographic distance between cited and citing author (All articles) (13-4)
SSRN	-234.5*	-359.8*	-58.6	-404.2**
	(99.2)	(140.2)	(137.4)	(138.4)
SSRN x Post_SSRN	105.5	116.6	15.2	210.5
	(119.3)	(176.1)	(156.7)	(172.0)
SSRN x High_JIF				366.8
				(194.0)
SSRN x Post_SSRN x High_JIF				-238.5
				(230.6)
Constant	3973.8	4081.8	1892.7	3871.4
	(4126.2)	(7002.9)	(4110.8)	(4075.6)
N of article-years	17253	7078	10175	17497
N of articles	1681	759	922	1725
N of SSRN article-year	2451	1429	1022	2451
N of SSRN articles	200	125	75	200
N of Journal/Vol/Is	153	89	59	149
N of Journal	70	50	16	74
Log-Likelihood	-161239	-66413	-94768	-163529

Chapter 2. Strength in Numbers: Does Data-Driven Decisionmaking Affect Firm Performance?

(with Erik Brynjolfsson and Lorin Hitt)

Abstract

We examine whether firms that emphasize decision making based on data and business analytics ("data driven decision making" or DDD) show higher performance. Using detailed survey data on the business practices and information technology investments of 179 large publicly traded firms, we find that firms that adopt DDD have output and productivity that is 5-6% higher than what would be expected given their other investments and information technology usage. Furthermore, the relationship between DDD and performance also appears in other performance measures such as asset utilization, return on equity and market value. Using instrumental variables methods, we find evidence that the effect of DDD on the productivity do not appear to be due to reverse causality. Our results provide some of the first large scale data on the direct connection between data-driven decision making and firm performance.

Keywords: Business Analytics, Decisionmaking, Productivity, Profitability, Market Value

Acknowledgements: We thank Andrew McAfee, Roger Robert, Johnson Sikes and participants at the Workshop for Information Systems and Economics and participants at the 9th Annual Industrial Organization Conference for useful comments and the MIT Center for Digital Business for generous financial support.

Introduction

We have been witnessing a data revolution; firms gather extremely detailed data from and propagate knowledge to their consumers, suppliers, alliance partners, and competitors. Part of this trend is due to the widespread diffusion of enterprise information technology such as Enterprise Resource Planning (ERP), Supply Chain Management (SCM), and Customer Relationship Management (CRM) systems (Aral et al. 2006; McAfee 2002), which capture and process vast quantities of data as part of their regular operations. Increasingly these systems are imbued with analytical capabilities, and these capabilities are further extended by Business Intelligence (BI) systems that enable a broader array of data analytic tools to be applied to operational data. Moreover, the opportunities for data collection outside of operational systems have increased substantially. Mobile phones, vehicles, factory automation systems, and other devices are routinely instrumented to generate streams of data on their activities, making possible an emerging field of "reality mining" (Pentland and Pentland 2008). Manufacturers and retailers use RFID tags to track individual items as they pass through the supply chain, and they use the data they provide optimize and reinvent their business processes. Similarly, clickstream data and keyword searches collected from websites generate a plethora of data, making customer behavior and customer-firm interactions visible without having to resort to costly or ad-hoc focus groups or customer behavior studies.

Furthermore, leading-edge firms have moved beyond passively collecting data to actively conducting customer experiments to develop and test new products. For instance, Capital One Financial pioneered a strategy of "test and learn" in the credit card industry where large number of potential card offers were field-tested using randomized trials to determine customer acceptance and customer profitability (Clemons and Thatcher 1998). While these trials were quite expensive, they were driven by the insight that existing data can have limited relevance for understanding customer behavior in products that do not yet exist; some of the successful trials created led to products such as "balance transfer cards," which revolutionized the credit card industry. Online firms such as Amazon, eBay, and Google also rely heavily on field experiments as part of a system of rapid innovation, utilizing the high visibility and high volume of online customer interaction to validate and improve new product or pricing strategies. Increasingly, the culture of experimentation has diffused to other information-intensive industries such as retail financial services (Toronto-Dominion Bank, Wells Fargo, PNC), retail (Food Lion, Sears, Famous Footwear), and services (CKE Restaurants, Subway) (see Davenport 2009).

In this paper, we develop a measure of the use of "data-driven decision making" (DDD) that captures a shift in the decision-making process. This shift is from a practice, where managers make a decision first and collect data to support the decision, to a new practice where data come first and managers make a decision based on the data rather than their instinct or "gut feelings". This is possible because the cost of collecting and analyzing data has become very low and the types of data to collect and analyze do not have to be decided *a priori*. While there have been many studies on the relationship between various IT-enabled capabilities and firm performances (see {Mithas, #833}for review), there have been very few studies to measure this shift itself[2] and a causal relationship between the shift and firm performance. The shift from intuition-driven to data-driven in managerial decision-making process is enabled only when firms have IT capability to collect and analyze data, but the capability itself is not sufficient for the shift. The firm should have a culture and leadership to embrace the new decision-making process, allowing the replacement of the top managers' intuition with data. We show that IT investment doesn't necessarily lead to DDD in all firms and there is a fair amount of variation in DDD with a given amount of IT investment. DDD appears to be different from IT investment, the presence of ERP system, IT infrastructure, and IT usage, while it is correlated with those conventional IT measures. We examine what drives the data-driven decision-making process in the firms: firm age, consistency of business practices, and adjustment cost. We use these drivers as instruments to address the potential endogeneity issue in the causality between DDD and firm performance. Combining measures of DDD captured in a survey of 179 publicly traded firms in the US with public financial information and private data on overall information technology investments, we examine the relationships between DDD and productivity, financial performance and market value. We find that DDD is associated with a 5-6% increase in their output and productivity, beyond what can be explained by traditional inputs and IT usage. Supplemental analysis of these data using instrumental variables methods and alternative models suggest that this is a causal effect, and not driven by the possibility that productive firms may have a greater propensity to invest in DDD practices even in the absence of real benefits.

Theoretical Framework

IT and Firm Performance

Since the mid-1990s, it has been recognized that information technology is a significant driver of productivity at the business unit (Barua et al. 1995), firm (e.g., Brynjolfsson and Hitt 1996; Bresnahan et al. 2002; {Mithas, 2012 #831}; Kohli and Devaraj 2003), industry (e.g., Jorgenson and Stiroh 2000; Melville et al. 2007) and economy level (Oliner and Sichel 2000; Jorgenson and Stiroh 1999). The early studies about IT and firm performances have used IT investment as a proxy for IT-capability (Oliner and Sichel 2000; Jorgenson and Stiroh 1999). Since then scholars have directly constructed IT-related measures (see {Mithas, 2011 #833} for review) and identified the mechanism on how IT increases the firm performances. To name a few, IT-related measures such as IT-resources (e.g., {Melville, 2004 #710};{Wade, 2004 #834}), IT-capabilities (e.g., {Bharadwaj, 2000 #835};{Santhanam, 2003 #836};{Mithas, 2011 #833}), IT-infrastructure (e.g., {Lewis, 2003 #837}; {Weill, 2002 #838};{Weill, 2002 #839}), introduction of ERP systems (Hitt et al. 2002; Anderson et al. 2003;Aral et al. 2006; McAfee 2002; Aral et al. 2009; McAfee and Brynjolfsson 2008), or an actual IT usage (Devaraj and Kohli 2003) have been shown as a driver for firm performance. In other studies, organizational structure (e.g. Bresnahan

[2] There is a growing volume of case evidence that this relationship is indeed true, at least in specific situations (e.g., Davenport and Harris 2007; Ayres 2008; Loveman 2003). However, there is little independent, large sample empirical evidence on the value or performance implications of adopting these technologies, especially in academic literature.

et al. 2002; Francalanci and Galal 1998; Tambe et al. 2009) complementary to IT investment have been examined as a mechanism for IT to increase firm performance.

While the previous studies have surveyed many IT-enabled processes, very few studies have examined the data-driven decision-making process as a variable of interest. It is partly because DDD is a paradigm change from intuition to data as a basis for managerial decision-making, only feasible after a fair amount of practices of the IT-enabled processes. After a couple of decades since the introduction of MIS (Management Information System), such a change has become possible and we are beginning to witness the change. While the previous studies have not differentiated whether IT helps making a decision or implementing the decision, our study is particularly concerned about the importance of IT in decision-making rather than implementation process of a decision. We differentiate the role of IT whether data support an implementation of a decision or lead to the decision itself. In other words, we are concerned about the relative importance of data and intuition in the decision-making step.

MIS has typically supported strategic decisions *after* managers make the decisions. DDD is opposite to such a process: data comes first without *a priori* decision and then managers make decisions based on the data[3]. For example, Amazon.com continually experiments with their websites for the best appearance. "We have been implementing changes on <online site> based on opinion, gut feeling or perceived belief. It was clear that this was no way to run a successful business...Now we can release modifications to the page based on purely on statistical data" (Kohavi, Longbotham et al., 2009). Kohavi et al. (2009) illustrates how data can overrule the traditional approach of relying on intuition and "the highest paid person's opinion" (HiPPO) for corporate decision-making. "Greg Linden at Amazon created a prototype to show personalized recommendations based on items in the shopping cart. Linden notes that while the prototype looked promising, a marketing senior vice-president was dead set against it, claiming it will distract people from checking out. Greg was forbidden to work on this any further. Nonetheless, Greg ran a controlled experiment" to prove that the new feature would bring in a lot more sales at a statistically significant level and the HiPPO was wrong. The top manager's intuition on a new service used to determine whether the new service should be developed or not. But data and experiments have replaced the manager's intuition with hard data: data comes first and then decision follows. The way that MIS is helping managerial decision in this context is very different from the traditional ways in the past: data is playing a leading, not a supporting, role in decision-making process.

---

[3] Hal Varian, the chief economist at Google, mentioned that economists at Google had a new approach for an economic study, based on data: instead of proposing a hypothesis first and collecting and analyzing data to support or reject the hypothesis, they examine the data first and approach a phenomenon without a priori hypothesis. This is possible at Google because they can collect fine-grained data real-time at a very low cost. The shift from intuition-driven to data-driven in the managerial decision-making is analogous to the shift in the scientific method in Biology as well. The scientific method employed in Biology has changed from theory-driven to data-driven since the invention of microscope that allowed collection of data on cells.

Our study captures this process as DDD and relates it to firm performance. Moreover, using an instrumental variable method, we demonstrate a causal relationship between DDD and firm performance.

DDD and Firm Performance

While the use of "big data" and "data analytics" has been spotlighted in the popular business press as a key driver for firm performance in recent years, there have been very few large-scale econometric studies that relate DDD to firm performances and most studies have been case studies and surveys on DDD. For example, Loveman (2003), the CEO of Caesar's Entertainment, states that use of databases and decision-science-based analytical tools was the key to his firm's success. Davenport and Harris (2007) have listed many firms in a variety of industries that gained competitive advantage through use of data and analytical tools for decision making such as Proctor and Gamble and JC Penney. They also show a correlation between higher levels of analytics use and 5-year compound annual growth rate from their survey of 32 organizations. A more recent study (Lavalle et al. 2010) has reported that organizations using business information and analytics to differentiate themselves within their industry are twice as likely to be top performers as lower performers. Our study advances the understanding about the relationship between DDD and firm performance by applying a standard econometric method to survey and financial data on publicly traded large 179 firms.

Measuring the Impact of Information Technology Investments

Productivity

The literature on IT value has used a number of different approaches for measuring the marginal contribution of IT investment accounting for the use of other firm inputs and controlling for other firm, industry or temporal factors that affect performance (see a summary of these in Hitt and Brynjolfsson 1996). Our focus will be on determining the marginal contribution of DDD on firm performance, after controlling for IT investment and IT usage. As we will describe later, DDD will be captured by an index variable (standardized to mean zero and variance one) that captures a firm's position on this construct relative to other firms we observed, and can be incorporated directly into various performance measurement regressions.

The key regression model in this paper is a productivity regression shown in the equation (1). The most commonly used measure of performance in this literature is multifactor productivity, which is computed by relating a measure of firm output such as Sales or Value-Added, to firm inputs such as capital (K), labor (L), and information technology capital or labor (IT). Different production relationships can be modeled with different functional forms, but the most common functional form assumption is the Cobb-Douglas production function which provides the simplest relationship between inputs and outputs that is consistent with economic production theory. The model is typically estimated in firm-level panel data using controls for industry and year, and inputs are usually measured in natural logarithms. The residuals of this equation can be interpreted as firm productivity after accounting for the contributions of all inputs (sometimes called "multifactor productivity" or the "Solow residual"). Including additional firm factors additively into this equation can then be interpreted as factors that "explain" multifactor productivity and have a direct interpretation as the marginal effect of the factor on firm productivity. This results in the following estimating equation:

$$\ln(sales)_{it} = \beta_0 + \beta_1 \ln(m)_{it} + \beta_2 \ln(k)_{it} + \beta_3 \ln(ITE)_{it} + \beta_4 \ln(NonIT\ Employee)_{it} + \beta_4 (\mathrm{DDD})_{it} + controls + \varepsilon \quad \text{-- (1)}$$

where m is materials, k is physical capital, ITE is the number of IT employees, Non-IT Employee is the number of Non-IT employees, and DDD is our data-driven decision-making variable. The controls include industry and year. To help rule out some alternative explanations for our results we also include the firm's explorative tendency and the firm's human capital such as importance of typical employee's education and average worker's wage as control variables. Our performance analysis is based on a five year panel (2005-2009) including a single cross-section of DDD data observed in 2008 match to all years in our panel.[4]

Profitability

An alternative method of measuring firm performance is to relate an accounting measure of profitability to the construct of interest and other control variables. This approach is particularly popular in the management literature, and has been employed in many studies that have examined the performance impact of ERP (e.g., Hitt et al. 2002; Aral et al. 2006). However, it has the disadvantage that it is less theoretically grounded than other performance measurement methods, but has a significant advantage that it allows a diversity of interpretations of performance, and is closely related to how managers and securities analysts actually compare the performance of firms. The general form of this estimating equation is:

$$Log(Perform.Numerator)_{it} = \beta_0 + \beta_1 log(IT)_{it} + \beta_2 (DDD)_{it} + \beta_3 log(Perform.Denominator)_{it} + control + \varepsilon \quad \text{---(2)}$$

The performance numerators and denominators for the profitability ratio we tested are summarized in Table 1. Note that the IT in this equation is also the number of IT employees and used as a proxy for IT-related capability as in the equation (1).

Table 1. Performance numerator and denominator in the profitability analysis

Profitability Ratio	Performance Numerator	Performance Denominator
Return on Assets	Pretax Income	Assets
Return on Equity	Pretax Income	Equity
Asset Utilization	Sales	Assets

[4] This assumes that our measure of DDD in 2008 is correlated with the true value of DDD in other years. We test whether our results are sensitive to this assumption and find no evidence that the relationship between measured DDD and productivity varied over the sample period.

Market Value

The final performance metric we examined is the total market value of the firm. Accounting measures such as return on assets, return on equity, and return on sales have some weaknesses in capturing firm performance: 1) they typically only reflect past information and are not forward looking; 2) they are not adjusted for risk; 3) they are distorted by temporary disequilibrium effects, tax laws, and accounting conventions; 4) they do not capture the value of intangible assets; 5) they are insensitive to time lags necessary for realizing the potential of organizational change. Financial market-based measures can be a useful alternative to these accounting measures. In particular, variants on Tobin's q ratio, defined as the ratio of the stock market valuation of a firm to its measured book value, has been used as measure of business performance (Chen and Lee 1995), intangible assets (Hall 1993; Hirschey 1982), technological assets (Griliches 1981), and brand equity (Simon and Sullivan 1993).

In the context of IT-investments, market value has been used to estimate the value of intangible assets such as organizational capital associated with IT assets (e.g. Brynjolfsson et al. 2002; Saunders and Brynjolfsson 2010; Brynjolfsson et al. 2011). The underlying principle is that the total value of financial claims on the firm should be equal to the sum of the firm's assets (Baily et al. 1981; Hall et al. 2000; Hall 2001). Therefore, the value of intangible assets can be estimated by subtracting the value of other tangible inputs from the sum of financial claims. Other researchers used Tobin's q to examine the effects of information technology on firm performance (Bharadwaj et al. 1999). Related work found that e-commerce announcements (Subramani and Walden 2001) and Internet channel addition (Geyskens et al. 2002) were correlated with changes in market value.

We build on the intangible assets literature and model the value of financial claims against the firm, MV, as the sum of each of its n assets, A.

$$MV = \sum_{i=1}^{n} A_i \qquad \text{---------------- (3)}$$

What the above model formulates is that the market value of a firm is simply equal to the current stock of its capital assets when all assets can be documented and no adjustment costs are incurred in making them fully productive. However, in practice firm value can deviate significantly from tangible book value. For instance, at the time of writing Google is valued at approximately $190 billion but the company lists $40 billion in total assets on its balance sheet. The difference, $150 billion, can be interpreted as the sum of its intangible assets.

Following the emerging literature on IT and intangible assets, we consider three classes of intangibles – those related to information technology and its associated organizational complements (captured as IT employees), brands (captured as advertising), and technology (captured as R&D investment). We also consider the possibility that the value of some types of assets increase with the presence of DDD (similar to the treatment of organizational assets in Brynjolfsson et al. 2002). This yields the following equation:

$$(MV)_{it} = \beta_0 + \beta_1 K_{it} + \beta_2 (OA)_{it} + \beta_3 (IT)_{it} + \beta_4 (DDD)_{it} \times A_{it} + controls + \varepsilon_{it} \qquad \text{----------------- (4)}$$

where MV is the market value of the firm, K is the capital, OA is other assets, IT is either IT capital or the number of IT-employees, DDD is our data-driven decision-making variable, A is an asset (capital, other assets, or non-IT-employee) and controls include industry, year, the ratio of R&D expense to sales, and the ratio of advertising expense to sales. Unlike the productivity equation using Cobb-Douglas formulation, the market value equation does not use the log form because the market value is believed to be the sum of the value of firm's each asset. This formulation does not usually include the number of employees or other human capital as firms cannot own the employees. Unlike other assets, employees can leave their firms at their will. It can be, however, argued that the human capital is an important asset for a firm as the firm can have management practices to attract and keep their key employees. We, therefore, test our regression with or without the number of employees. This also provides a more natural relationship since one would generally expect that firms of different sizes would have a different marginal effect on market value as DDD (measured as a standardized index) varies.

Endogeneity of DDD

All of the performance methods above must either be interpreted as conditional correlations rather than causal relationships or rely on an assumption that DDD is exogenous with respect to firm performance. For the purposes of this study, neither is an attractive approach since the former limits the managerial relevance of this analysis, and the latter is unlikely to be true (although a number of recent studies have suggested that the bias on at least IT investment due to endogeneity is not large – see Tambe and Hitt 2011).

The literature on IT value has generally used two types of approaches for directly addressing endogeneity concerns. First, researchers can make arguments of temporal precedence either by including lagged values of other input variables (e.g. Brynjolfsson and Hitt 1996;Dewan and Kraemer 2000), or by looking at differences in performance before and after a system becomes live rather than when the investment is made (Aral et al. 2006; Hitt and Frei 2002). Second, econometric methods that rely on internal instruments in panel data (such as the Arellano and Bond, or Levinsohn and Petrin estimators) can be used to control for endogeneity under the assumption that changes in past investment levels are uncorrelated with current performance. However, both of these approaches rely on significant temporal variation in the variables of interest, and cannot be readily applied to our context since we have a single cross-sectional observation of DDD. However, we are able to pursue the more traditional instrumental variables approaches, where researchers specify a set of factors (instruments) that drive the demand for the endogenous factor but are not correlated with the unobserved component of performance.

In prior work, researchers have used measures of the composition of IT (relative proportion of mainframes versus PCs) and the overall age of capital within an organization (Brynjolfsson and Hitt 2003) under the assumption that these factors determine the ability of a firm to adapt their IT infrastructure to changing business needs. Recent work by Brynjolfsson, Tambe and Hitt (Tambe and Hitt 2011) attempts to more directly measure the IT-related adjustment costs or organizational inertia (see e.g. Hannan and Freeman 1984; Nelson and Winter 1982) by developing a scale capturing the factors that facilitate or inhibit IT investment such as senior management support or organizational culture, and used this scale as an additional instrument.

To these existing instruments, we add additional instruments that may be especially useful in explaining cross-sectional variation in DDD. Prior work has specifically linked organization experience, operationalized as *firm age*, to organizational inertia (Henderson and Clark 1990; Henderson 1993; Bresnahan et al. 2009; Balasubramanian and Lee 2008; Tushman and Anderson 1986). By this argument, younger firms are more likely able to adopt new innovations such as business analytics or other technologies underlying DDD, thus leading to a negative correlation between DDD and firm age (which is observed in our data). To reduce the possibility that our instrument would be invalidated by a correlation between innovation-driven productivity and firm age (see Huergo and Jaumandreu 2004), we include controls for innovation activity when this instrument is used. It is also possible that firm age has a correlation with productivity due to learning by doing (e.g. Cohen and Levinthal 1989;Argote et al. 2003;Levitt and March 1988;Nass 1994) but since this would yield positive correlation between firm age and productivity, any bias from using this instrument would likely reduce our observed effect of DDD, making the results more conservative.

Another potential demand driver for DDD is the degree of consistency in business practices. Brynjolfsson and McAfee (2008) argue that one way in which firms are able to capture the value of IT-related innovation, including discoveries facilitated by DDD, is that they can replicate good ideas across the organization. This is motivated by the observation that information (e.g. Shapiro and Varian 1999) or specific information about innovative practices (e.g., Jones 1999) is non-rival and therefore more valuable with scale. Thus, firms that have demonstrated the ability to deploy common business practices across large numbers of organization units are likely to be more effective users of DDD, and therefore more likely to have invested in developing DDD capabilities than firms that have disparate business practices.

Thus, our set of instruments includes constructs employed in prior literature for capital age (Brynjolfsson and Hitt 2003) barriers to IT adoption (Brynjolfsson et al. 2011) as well as new measures of firm age, and consistence of business practices. As we will show later, these constructs pass the normal empirical instrument validity tests, and when utilized, demonstrate that our observation relationships between DDD and performance are robust to concerns about reverse causality.

All of the dependent and independent variables used in the regression equations are summarized in Table 2.

Table 2. Summary of Independent and Dependent Variables in the Regression Analysis

(*The list of the performance numerators and denominators are listed in Table 1.)

Equation Number	Dependent Variable	Independent Variable	Control Variable
1	Log(Sales)	Log(Material); Log(PPE); Log(IT-employees); Log(Non-IT Employees); DDD	Industry; year; explorative tendency; importance of typical employee's education; average worker's wage
	DDD	Firm Age; Consistency of Business Practices	Industry; year
2	Log(Performance Numerator)*	Log (IT-Employee); DDD; Log(Performance Denominator)*	Industry; year
4	Market Value	PPE; Other Asset; IT-Employee; DDD	Industry; year

Data and Measures

Business Practice

Our business practice and information system measures are estimated from a survey administered to senior human resource (HR) managers and chief information officers (CIO) from large publicly traded firms in 2008. The survey was conducted in conjunction with McKinsey and Company and we received responses from 330 firms. The survey asks approximately 80 questions about business practices as well as organization of the information systems function and usage of information systems. The questions extend a previous wave of surveys on IT usage and workplace organization administered in 1995-1996 and 2001 (Hitt and Brynjolfsson 1997; Brynjolfsson et al. 2011), but adds additional questions on innovative activities, the usage of information for decision making, and the consistency of their business practices. To explore the effect of DDD, we used the survey response to construct measures of firms' organizational practices. We combine these measures with publicly available financial data. This yielded 179 firms with complete data for an analysis of firm productivity covering all major industry segments ver the period from 2005 to 2009. The exact wording of the survey questions appears in Appendix.

Data-Driven Decision Making (DDD). We constructed our key independent variable, data-driven decision making (DDD), from three questions of the survey: 1) the usage of data for the creation of a new product or service, 2) the usage of data for business decision making in the entire company, and 3) the existence of data for decision making in the entire company (see Appendix I for the exact wording for the questions). The difference between the first and second question was the context when the data was used: the first asked the context of creation of a new service and product, and the second question was about the decision-making in the entire company beyond the creation of a new product/service such as HR management. These three questions were a part of a survey which consists of approximately 80 questions on organizational structures, IT usage, and general practices on human resources management. We extracted those three questions from the survey based on the relevance of the questions to our variable of interest, DDD, rather than the general IT use and practices.

We created DDD by first standardizing (STD) each factor with mean of zero and standard deviation of 1 and then standardizing the sum of each factor:

DDD = STD(STD(use of data for creation of a new product or service) + STD(use of data for business decisions in the entire company) + STD(existence of data for such a decision))

Adjustment Cost. A measure for the adjustment cost was constructed from 6 survey questions. Respondents were asked to describe the degree to which the following 6 factors facilitate organizational changes: financial resources, skill mix of existing staff, employment contracts, work rules, organizational culture, customer relationships, and senior management involvement. Similarly to DDD, we created the composite index by first standardizing each factor with mean of zero and standard deviation of 1 and then and then standardizing the sum of the scale components.

Consistency of Business Practices. Consistency of business practices ("Consistency") is constructed as a composite of responses to six survey questions on consistency of business practices across operating units, within business units, across functions, and across geographies (4 questions); the effectiveness of IT for supporting consistent practices; and consistency of prioritization of projects. Similarly to DDD, the consistency measure was created by first standardizing each factor with mean of zero and standard deviation of 1 and then standardizing the sum of the scale components.

Exploration (EXPR). Firm's tendency to explore a new market or technology and to engage in radical innovation was used as a control variable because firm age, one of our instruments, may be correlated with a firm's innovative activity which, in turn, can affect productivity and other performance measures. It was a composite index of 8 survey questions regarding the firm's tendency to explore new markets or technologies. This index was also standardized in the same manner as the consistency and DDD measures.

General Human Capital. The importance of typical employee's education and the average worker's wage were used as a proxy for the firm's human capital.

Other IT-related Measures

Information Technology Staff. The survey included the questions about IT budgets, outsourcing, change of IT budgets from 2008 to 2009, and full time IT employment. The number of full-time IT employees for the year 2008 was asked in the survey, but for the year 2009 it was estimated from the questions on IT budget. Using the change of IT budget from 2008 to 2009, the percentage of outsourcing, and IT FTE for 2008, we were able to estimate the IT FTE for the year 2009. The year from 2005 and 2006, we used data collected in a previous study (Tambe and Hitt 2011). For the year 2007, a value interpolated from 2005, 2006, 2008 and 2009 was used. The number of non-IT employees is equal to the number of employees reported on Compustat less our computed IT employment measure.

While the construction of the IT input series is less than ideal, we do not believe that this introduces any biases in the analysis, and enables us to extend existing IT input datasets almost through the current period. Tambe and Hitt (2011) showed that IT employees appear to be a good proxy of overall IT input, at least for conducting productivity analyses (results using IT capital and IT employees are essentially the same, with the IT employee data showing less error variance). To reduce the impact of using different sources over time, we include year dummy variables that will control for any scaling differences. The remaining variance in these measures is likely noise which may tend to bias our results toward zero, making them more conservative.

IT-usage. The percentage of the employees using IT-related technology was used as a proxy for the actual IT usage.

IT-governance. The way in which IT projects are governed and prioritized was also surveyed. It was a composite index of 6 survey questions (Table 2), of which the exact wordings are shown in Appendix II. This index was also standardized in the same manner as the DDD measure.

IT-infrastructure. A measure for the management practices regarding IT infrastructure was constructed by combining three survey questions (Table 2). The exact wordings of the questions are shown in Appendix II. This index was standardized in the same manner as the DDD measure.

Other Data

Production Inputs and Performance. Measures of physical assets, employees, sales and operating income were taken directly from the Compustat Industrial Annual file from 2005 to 2009. Materials were estimated by subtracting operating income before tax and labor expense from sales. In the case that labor expense was not available, it was estimated from number of employees and the industry average wage for the most disaggregated industry data available that matched the primary industry of the firm.

Following prior work (Brynjolfsson et al. 2002), we calculated market value as the value of common stock at the end of the fiscal year plus the value of preferred stock plus total debt. The R&D ratio and the advertising expense ratio were constructed from R&D expenses and advertising expense divided by sales, respectively. The missing values were filled in two ways; 1) using the averages for the same NAICS code industry and 2) creating a dummy variable for missing values and including the dummy variable in the regression. The results were essentially the same for our variables of interest.

Firm Age. Firm age was collected from a semi-structured data site (http://www.answers.com) where available, and supplemented with additional data from firm websites and the Orbis database. Firm age was the founding year subtracted from the year of the observation. In case that multiple firms were merged, we used the founding year of the firm which kept its name. For mergers where the new entity did not retain either prior firm name, we used the founding year of the oldest firm engaged in the merger.

Table 2. Construction of Measure of Organizational Practices

	Range of scale	Mean	Std. Dev.	Cronbach's Alpha
Measure 1: Data-Driven Decision-making (DDD)				0.58
The typical basis for the creation of a new product or service (HR survey q13a)	1-5[5]	2.97	1.13	
The degree of using data for a decision-making for the ***entire company*** (HR survey q16j)	1-5	3.85	0.85	
The existence of data needed for a decision-making (HR survey q16p)	1-5	3.43	0.87	
Measure 2: Adjustment cost				0.69
Please rate whether the following factors at your company facilitate or inhibit the ability to make organizational changes: (1:inhibit significantly, 5:facilitate significantly) (HR survey q11)				
a) Skill mix of existing staff	1-5	3.22	1.19	
b) Employment contracts	1-5	2.89	0.65	
c) Work rules	1-5	2.98	0.83	
d) Organizational culture	1-5	3.31	1.27	
e) Customer relationships	1-5	3.69	1.02	
f) Senior management involvement	1-5	4.11	0.98	
Measure 3: Consistency				0.77
Looking across your entire company, please rate the level of consistency in behaviors and business processes across operating units (HR survey q1)	1-5	3.02	0.75	
Regarding the first core activity of your company, the consistency within business unit (HR survey q9a)	1-5	3.79	0.93	
Regarding the first core activity of your company, the consistency across functions (e.g., sales, finance, etc) (HR survey 9b)	1-5	3.38	0.99	
Regarding the first core activity of your company, the consistency across geographies (HR survey q9c)	1-5	3.53	0.99	
Effectiveness of IT in building consistent systems	1-5	3.50	0.85	

[5] Scale ranges from 1 to 5, with 5=greatest reliance on data

and processes for each operating unit (IT survey q13b)				
Measure 4: Exploration (EXPR)				0.58
IT facilitates to create new products (IT survey 11a)	1-5	3.78	1.22	
IT facilitates to enter new markets (IT survey 11b)	1-5	3.68	1.15	
IT supports growth ambitions by delivering services or products that set us apart from competitors (IT survey 12c/HR survey 15c)	1-4	2.52; 2.56	1.08; 1.01	
IT plays a leading role in transforming our business (IT survey 12d/HR survey 15d)	1-4	2.90; 3.01	1.13; 1.12	
IT partnering with the business to develop new business capabilities supported by technology (IT survey 13f/HR survey 14e)	1-5	3.33; 0.96	3.01; 1.09	
Strong ability to make substantial/disruptive changes to business processes (HR survey 16l)	1-5	2.90	1.05	
Measure 5: General human capital				
EDUCATION: The importance of educational background in making hiring decisions for the "typical" job (HR survey q4)	1-5	3.34	1.00	
% of employees using PC/terminals/workstations (HR survey q7a)	%	77.0	27.1	
% of employees using e-mails (HR survey q7b)	%	73.0	29.1	
Measure 6: IT governance				0.63
The consistency of IT project and approval processes (IT survey 15a)	1-3	2.48	0.73	
The involvement of business with IT projects (IT survey 15b)	1-3	2.32	0.61	
The management of IT demand (IT survey 15c)	1-3	2.52	0.53	
The sharing of IT cost to the business (IT survey 15d)	1-3	1.80	0.77	
The tracking/managing IT-project benefits (IT survey 15e)	1-3	1.61	0.68	
The measurement and reporting of IT-projects performance (IT survey 16a)	1-5	2.86	1.01	
Measure 7: IT infrastructure management				0.41
The management of service level (IT survey 19a)	1-3	1.72	0.70	
The management of standards/products in infrastructure (IT survey 19b)	1-3	2.31	0.64	
The management of IT capacity (IT survey 19c)	1-3	2.26	0.77	

Results and Discussion

Productivity Tests

The descriptive statistics for our variables are tabulated in Table 2 and Table 3. Most of the business practice measures were captured on 5-point Likert scales with a mean on the order of 3-4 and a standard deviation of approximately 1. When formed into scales, the control variables for adjustment costs and consistency of business practices appear to be fairly internally consistent with Cronbach's alpha of .69 and .77 respectively. The DDD measure shows a Cronbach's alpha of 0.58, which is consistent with the fact that firms can pursue some aspects of DDD (such as using data to develop new products) independently of the others. The same appears true for the exploration measure. The distributions of DDD is somewhat positively-skewed; the mode in the histogram of DDD is greater than its mean (Figure 1). There is an industry difference in the distribution of DDD (Figure 2). Not surprisingly the industry of information, professional services, finance, and insurance (NAICS 1-digit code 5 industries) shows a more positively skewed distribution of DDD than other industries (Figure 2). Manufacturing industry appears to have a positively skewed distribution of DDD as well (Figure 2). Recognizing the industry differences in the DDD distribution, we added industry dummy variables to control for the industry difference in our analysis. The average firm in our sample is large, with a geometric mean of approximately $2.3 billion in sales, 6000 non-IT employees and 172 IT employees.

Table 3. Production Function Variables (N=111, Year 2008 cross section)

Variable	Mean	Std.Dev.
Log(Sales)	7.76	0.90
Log(Material)	7.18	1.02
Log(Capital)	6.26	1.64
Log(Non-IT Employee)	8.70	1.05
Log(IT-Employee)	5.15	1.22
Log(Avg. Workers' Wage)	11.1	0.63

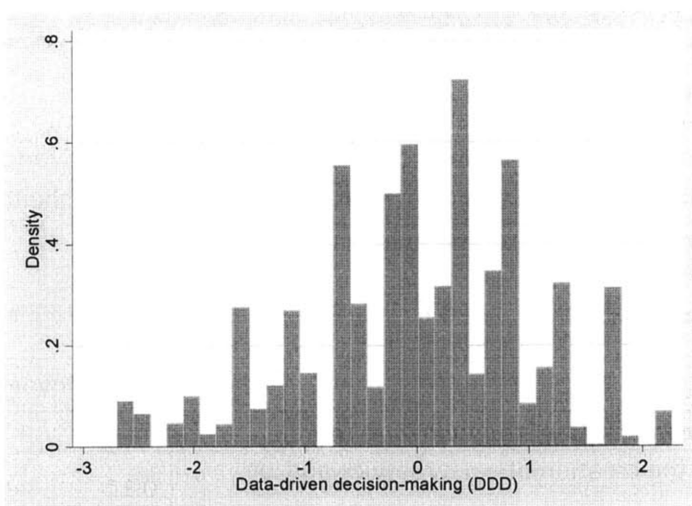

Figure 1. Distribution of DDD

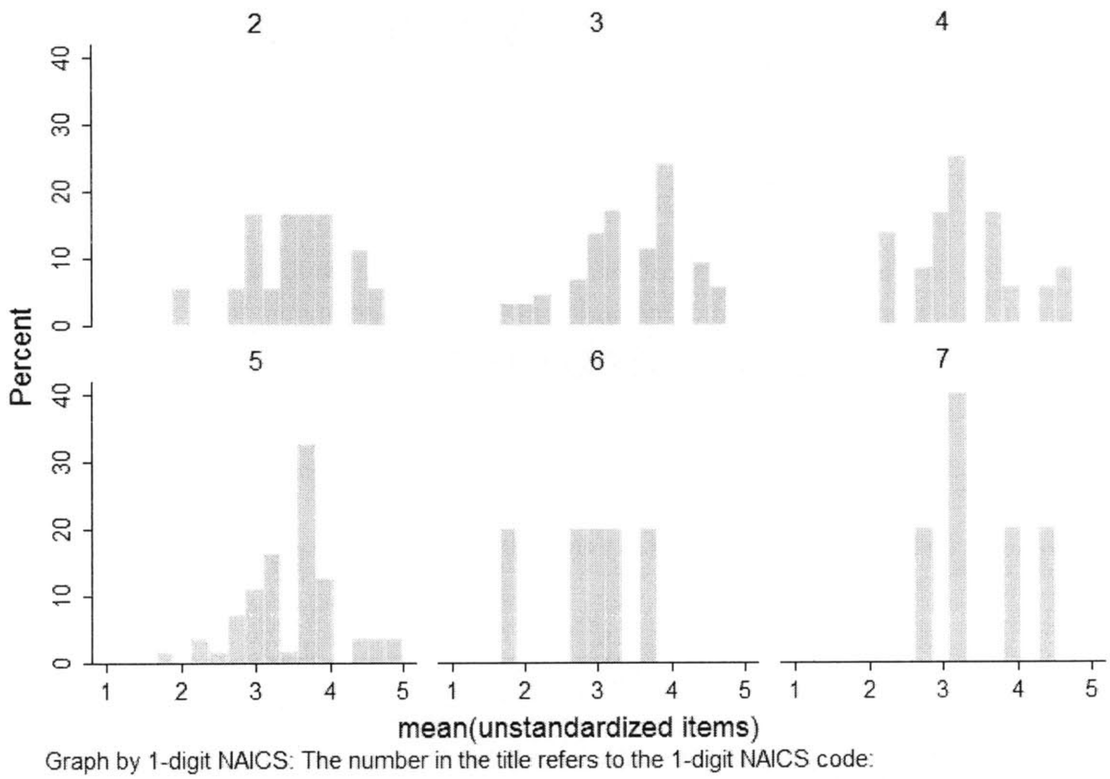

Figure 2. Distribution of DDD in different industries (2: Utilities and Construction; 3: Manufacturing; 4: Wholesale and Retail Trade; 5: Information, Professional Services, and Finance/Insurance; 6: Health care and social assistance; 7: Arts, Entertainment, Recreation, Accommodation, and Food Services)

Table 4 reports the conditional correlation of our key construct, data-driven decision-making (DDD), with the number of IT employees, after controlling for the industry and the total number of employees and shows that the correlation is 0.103 between IT staff and DDD.

Table 4. Correlations between DDD and the ratio of IT-employee to total employee

	IT Employee per total employee
DDD composite (average of the following three)	***0.18****
1. Use data for the creation of a new service and/or product (q13a)	0.11
2. Have the data we need to make decisions in the entire company (q16p)	0.15*
3. Depend on data to support our decision making (q16j)	0.17*

(Partial correlation for each pair, after controlling for industry. ***p<0.01, **p<0.05, *p<0.1)

Interestingly, this correlation is slightly lower than correlations between IT and other organizational complements which tend to be on the order of 20%. This may be because, as a new practice, DDD may be in the process of diffusing across firms. Firms that were historically high in IT may or may not have made investments in DDD. This will tend to lower estimates of correlations, but strengthen the power of tests for performance. In fact, if the correlation between DDD and IT investment were perfect, it would be impossible to distinguish the performance effects of the two.

The primary results regarding the relationship between DDD and productivity are shown in Table 5. All results are from pooled OLS regressions, and errors are robust and clustered by firm to provide consistent estimates of the standard errors under repeated sampling of the same firms over time. To rule out an alternative explanation, we included average worker's wage as a measure of human capital in all models. The first column (1) shows a baseline estimate of the contribution of IT to productivity during our panel from 2005 to 2009. The coefficient estimate on IT measure (the number of IT-employees) is about 0.056 ($t=2.8$, $p<0.01$), which is broadly consistent with the results from previous studies (e.g. Tambe and Hitt 2011). In column (2), we include our variable of interest, DDD and the coefficient estimate on DDD is 0.046 (s.e.=0.02, $p<0.01$) while the coefficient estimate on IT remains the same. This suggests that firms with one standard deviation higher score on our DDD measure are, on average, about 4.6% more productive than their competitors. It should be noted that this result is *after* controlling IT use; that is, the additional variation in productivity can be explained by the variation in DDD among the firms with the same amount of IT use.

To check the robustness of our assumption that the effects of DDD did not vary over the test period (2005-2009), we subdivide our sample into smaller periods and repeat our main productivity analysis. We find that when the sample is restricted to periods around our survey (2008-2009) the results are similar to the full sample (see Table 5) suggesting that we are not biasing our results by extending the data to prior periods. We can also compare the results of different subsamples over time in fully balanced panel of 72 firms. While the precision of the estimates is significanly reduced, the coefficients on DDD are virtually identical whether we consider the full sample, the pre-survey subsample (2005-2007) or the survey period (2008-2009) (see Table 6). We confirmed this observation with a Chow test which showed no significant variation in the DDD coefficient between subperiods. This suggests that our results are not biased by extending the panel in the time dimension.

While our preferred interpretation of the OLS results is that DDD is causing higher performance, there are at least two plausible endogeneity problems that could lead to this estimate having a positive bias. First, it is possible that high performing firms have slack resources enabling them to invest in a number of innovative activities including DDD, which would lead to a reverse causal relationship between performance and DDD. Second, there may be omitted variables such as management quality or greater firm-specific human capital that could be associated with both higher performance and the use of DDD, also creating upward bias. To address these problems, we treat DDD as endogenous and use three instruments: adjustment costs, firm age, and consistency of business practices. In addition, we extend the base specification to include a measure of innovation (EXPR) to remove any potential omitted variables bias related to the innovative activity in our sample firms, as well as measures of firm human capital.

Table 5. OLS Regressions of DDD on Productivity Measures

DV=Log(Sales)	(1) 2005-2009	(2) 2005-2009	(3) 2008-2009
DDD		0.046***(0.02)	0.043**(0.02)
Log(Material)	0.54***(0.04)	0.53***(0.04)	0.51***(0.04)
Log(Capital)	0.095***(0.02)	0.096***(0.02)	0.10***(0.03)
Log(IT-Employee)	0.056***(0.02)	0.057***(0.02)	0.12***(0.03)
Log(Non-IT Employee)	0.25***(0.03)	0.25***(0.03)	0.24***(0.04)
Constant	-1.48***(0.40)	-1.44***(0.37)	-1.10**(0.46)
Number of firms	179	179	113
Observations	681	681	211
R-squared	0.94	0.94	0.94
Other Controls	Industry; Year; Log(average worker's wage)		

(Robust standard errors in parenthesis. *** p<0.01, ** p<0.05, *p<0.1)

Table 6. Regression analysis of balanced panel when the sample period was divided into two periods.

DV=Log(Sales)	(1)2005-2009	(2) 2005-2007	(3) 2008-2009
DDD	0.058**(0.02)	0.054**(0.03)	0.052*(0.03)
Log(Material)	0.50***(0.05)	0.52***(0.08)	0.48***(0.04)
Log(Capital)	0.14***(0.03)	0.15***(0.03)	0.13***(0.04)
Log(IT-Employee)	0.039(0.03)	0.005(0.03)	0.11***(0.04)
Log(Non-IT Employee)	0.24***(0.05)	0.22***(0.03)	0.26***(0.05)
Constant	-1.43***(0.44)	-1.44***(0.45)	-1.43***(0.55)
Number of firms	72	72	72
Observations	360	216	144
R-squared	0.95	0.95	0.96
Other Controls	Industry; Year; Log(Average Worker's Wage)		

(Robust standard errors in parenthesis. *** p<0.01, ** p<0.05, *p<0.1)

First, we run OLS regression including these additional control variables. The OLS result for the coefficient estimate on DDD with these controls (column (1) in Table 7), 0.045 (t=2.7, p<0.01), was statistically the same as that without the additional control variables (0.046 with s.e.=0.02, the column (2) in table 5). While our preferred interpretation of the OLS results is that DDD is causing higher performance, there are at least two plausible endogeneity problems that could lead to this estimate having a positive bias. First, it is possible that high performing firms have slack resources enabling them to invest in a number of innovative activities including DDD, which would lead to a reverse causal relationship between performance and DDD. Second, there may be omitted variables such as management quality or greater firm-specific human capital that could be associated with both higher performance and the use of DDD, also creating upward bias. To address these problems, we treat DDD as endogenous and use three instruments: adjustment costs, firm age, and consistency of business practices. In addition, we extend the base specification to include a measure of innovation (EXPR) to remove any potential omitted variables bias related to the innovative activity in our sample firms, as well as measures of firm human capital.

Table 5We then conduct an instrumental variables regression using 2SLS and find that the coefficient on DDD is slightly higher than the prior OLS estimates (0.059, p<0.10) but is less precisely estimated due to the use of IV (see column 2 in Table 7). Nonetheless, our instrument set does pass the usual tests for weak instruments (the F-statistic on the excluded instruments in the 1st stage is 20 – see Staiger and Stock 1997) for a justification of this test). In addition, Hausman test fails to reject the null hypothesis that the OLS and IV coefficients are the same, thus suggesting that any biases due to endogeneity are small. Finally, because we have three instruments but only a single endogenous variable, we can conduct tests of over identification restrictions (the Sargan Test) and find that the coefficient on DDD is unaffected by the choice of instruments within our instrument set. Overall, these tests suggest that our original tests are unbiased, and firms that are one standard deviation above the means of our DDD scale have received a 5-6% productivity increase over the average firm.

Table 7. IV-Regressions of DDD on Productivity Measures

	(1) OLS	**(2) IV**
Variable	**DV=Log(Sales)**	**DV = Log(Sales)**
DDD	0.045*** (0.017)	0.059* (0.031)
Log(Material)	0.53*** (0.040)	0.53***(0.040)
Log(Capital)	0.097*** (0.024)	0.096***(0.024)
Log(Non-IT Employee)	0.26*** (0.031)	0.26***(0.031)
Log(IT Employee)	0.054*** (0.020)	0.055**(0.020)
Importance of Employee Education	0.018(0.02)	0.016(0.02)
Log(Avg. Workers' Wage)	0.20***(0.031)	0.20***(0.029)
Exploration	-0.009(0.02)	-0.012(0.023)
Controls	Industry, Year	The same as in (1)
Observations	681	681
Number of Firms	179	179
(Adj.) R-square	0.94	0.94
Overid Test (Sargan Test)		0.75
Hausman Test		0.67

(Standard errors clustered around firms are in parentheses, *p<0.10, **p <0.05, ***p < 0.01. DV means dependent variable. The overide test tests the null hypothesis that the estimates using each one instrument are the same. The Hausman Test tests the null hypothesis that OLS is consistent. The numbers for the Sargan and Hausman test indicate p-value. The industry control is at 2-digit NAICS level for manufacturing industries and 1-digit NAICS level for other industries. The years are from 2005 to 2009.)

In this construct, we do not and cannot differentiate whether the firm relies on IT capabilities in-house or outsourced IT developments for DDD. Our interpretation of the responses to the survey questions is that the respondents would rate the degree of DDD based on their firms' implementation of DDD consistently across firms regardless of the origin of the IT capabilities. It is, however, possible that the respondents may be less accurate in judging the degree of DDD if most of the IT capabilities are outsourced. The bias due to this inaccuracy may be both upward and downward: some HR managers may underestimate their DDD due to the less presence of IT personnel or overestimate their DDD due to the automated system housed somewhere else. The measurement error of this kind is the limitation of our study, but not different from any other survey-based study.

Business Profitability Test

We estimated the impact of DDD on three performance measures; return on assets (ROA), return on equity (ROE), and asset utilization (sales/assets) (see the equation 2). Because the data for some firms lacked values necessary to these regressions, we used 174 firms, a subset of the 179 firms used in the productivity estimation. IT appears to be significantly correlated with two profit measures in the expected direction (ROA and Asset Utilization) but not ROE (Table 8). DDD appears to be correlated with ROE and Asset Utilization at $p<0.05$. The point measure on the estimates for the coefficient of DDD ranges from 6 to 8% although these differences are not statistically significant across regressions. It should also be noted that the coefficients on the denominators are all significantly less than one, which would be the value expected if a pure ratio best fit the data.

Overall, these results are consistent in direction and magnitude with prior work using similar methods (Hitt et al. 2002; Aral et al. 2006) which showed the installation of ERP systems was correlated increases in some profitability measures. When this analysis is repeated with extended controls using instrumental variables regressions (seeTable 9), we find that the results are reinforced for return on assets, but are too imprecisely estimated for the other factors to make any conclusions. For the most part, the results are neither statistically different from the OLS results or from zero. Furthermore, the Sargan test statistic for the asset utilization regression is borderline significant, raising questions as to whether our instrument set can be used for this analysis. Thus, we are unable to make any inferences on whether the profit relationship is causal, most likely due to the reduced power (relative to productivity models) of this profit ratio analysis.

Table 8. Regressions of DDD on different performance measures

Interpretation	Return on Asset		Return on Equity		Asset Utilization	
Dependent Variable=	Log(Pretax Income)		Log(Pretax Income)		Log(Sales)	
DDD		0.063 (0.05)		0.067** (0.03)		0.076** (0.04)
Log(IT-Employee)	0.10* (0.06)	0.11* (0.06)	-0.045 (0.04)	-0.043 (0.04)	0.065* (0.04)	0.067* (0.04)
Log(Asset)	0.65*** (0.06)	0.64*** (0.06)			0.39*** (0.04)	0.37*** (0.04)
Log(Equity)			0.90*** (0.04)	0.89*** (0.04)		
Log(Non-IT Employee)	0.16** (0.07)	0.16** (0.07)	0.14*** (0.05)	0.14*** (0.04)	0.36*** (0.05)	0.36*** (0.05)
Constant	-1.77* (0.10)	-1.79* (1.01)	-2.26*** (0.6)	-2.29*** (0.57)	1.52** (0.62)	1.50** (0.61)
Number of	174	174	174	174	174	174

firms						
Number of observations	564	564	564	564	564	564
R-squared	0.67	0.78	0.84	0.84	0.82	0.83
Other controls	Industry; Year; Log(avg. worker's wage); Importance of education of typical employees					

(Standard errors clustered around firms are in parentheses, *p<0.10, **p <0.05, ***p < 0.01.)

Table 9. Profitability Regressions with Extended Firm-Specific Control Variables

Interpretation	Return on Asset		Return on Equity		Asset Utilization	
DV=	Log(Pretax Income)		Log(Pretax Income)		Log(Sales)	
	OLS	2SLS	OLS	2SLS	OLS	2SLS
DDD	0.068 (0.049)	0.19* (0.11)	0.059** (0.029)	0.092 (0.063)	0.066* (0.034)	0.034 (0.062)
Log(IT-Employee)	0.069 (0.054)	0.070 (0.053)	-0.041 (0.037)	-0.041 (0.036)	0.051 (0.035)	0.049 (0.035)
Log(Total Asset)	0.69*** (0.07)	0.64*** (0.08)			0.42*** (0.05)	0.43*** (0.06)
Log(Equity)			0.90*** (0.04)	0.89*** (0.04)		
Number of Firms	174	174	174	174	179	179
Number of Observations	568	568	565	565	682	682
R-square	0.76	0.76	0.85	0.85	0.84	0.84
Overid Test		0.77		0.27		0.04
Hausman Test		0.23		0.54		0.54
Other Control variables	Industry, Year, Log(R&D expense), Log(Advertising expense), Log(Capital), Log(Total number of employees), Log(Market share), Importance of employees' education					

(Standard errors clustered around firms are in parentheses, *p<0.10, **p <0.05, ***p < 0.01.)

Market Value Test

We also examined the relationship between DDD and the market value of firms. This regression relates market value to the three types of assets; PP&E, other assets, and IT. We repeat this analysis using two proxies for IT assets. First, we estimate IT budgets over time using the actual observation in 2008 and the ratio of IT employees to budgets to estimate the values in all other years. Second, we used the IT employees estimate directly (the same measure as in the productivity analysis)

The results measuring IT using budgets is presented in Table 10. Examining the control variables we find that the coefficient for property, plants and equipment (PP&E) is larger than the theoretical value of $1 (closer to $2 per dollar of PP&E) while the coefficient on other assets is substantially less (close to $0.20). The high value on PP&E may indicate short-run adjustment costs, or correlations with omitted assets; the low value on other assets perhaps suggests that stockholders do not believe that they will receive the full value of these assets, on average. Each dollar of IT budget is associated with $26 of market value, which after converting from a flow to a stock measure (see Saunders and Brynjolfsson 2010 – they find approximately a 2:1 ratio between IT spending and IT capital stock), suggests that each dollar of IT capital is associated with about $13. This is slightly higher than estimates reported in prior work (which are on the order of $10), but the differences are not statistically significant. We also find that the coefficient on IT is higher when combined with DDD. The interaction term implies, after a stock to flow adjustment, that IT capital is correlated with $6.5 of additional market value in firms that are one standard deviation higher in DDD. The fact that the IT budget coefficient drops slightly when the interaction is included is consistent with DDD being somewhat related to IT. Interestingly, the interactions are not significant (economically or significantly) for PP&E or other assets, highlighting the special role of IT in enabling DDD. Some researchers reported that IT capital investments make a larger contribution to overall firm risk than non-IT capital investments and about 30% of the gross return on IT investment corresponds to the risk premium associated with IT risk (Dewan et al. 2007). The valuation we estimated may, therefore, include the risk premium. Even if we only take 70% of the market value associated with IT as an actual value of IT asset, it is still significantly more than a theoretical $1, suggesting an intangible asset value associated with IT.

Results are similar when proxy IT capital with IT employees. The results suggest that each employee is associated with $8.2K of IT capital stock, and that firms that invest in DDD have an addition $3.1K of value per employee for each standard deviation of DDD above the mean. The results are otherwise similar to the prior analysis except there appears to be a small positive interaction between DDD and other assets (we are reluctant to interpret this because of the heterogeneity of other assets, and the low direct coefficient). Altogether, these analysis suggest that firms that adopt DDD have a higher market value, and that this value is most closely related to their level of IT capital.

Table 10. Regressions of DDD on market value, using IT budget as a proxy for IT capital

DV=Market Value	(1)	(2)	(3)	(4)
Property, Plant and Equipment (PPE)	2.00*** (0.71)	1.95*** (0.64)	1.91*** (0.59)	1.98*** (0.64)
IT Budget (ITB)	25.7*** (8.55)	19.1*** (5.34)	24.4*** (7.3)	21.5*** (6.3)
Other Assets (OA)	0.18*** (0.04)	0.19*** (0.03)	0.18*** (0.04)	0.21*** (0.035)
DDDxITB		13.6* (8.1)		
DDDxPPE			0.44 (0.52)	
DDDxOA				0.32 (0.18)
Constant		-747.5 (3,329)	-2,747 (4,238)	-992.3 (2,007)
Number of observations	481	481	481	481
Number of firms or observations	110	110	108	
R-squared	0.69	0.714	0.70	0.71

(Robust standard errors in parentheses, *p<0.10, **p <0.05, ***p < 0.01.)

Table 11. Regressions of DDD on market value

DV=Market Value	(1)	(2)	(3)	(4)
Property, Plant and Equipment (PPE)	1.77*** (0.50)	1.75*** (0.45)	1.72*** (0.43)	1.75*** (0.46)
IT-Employee (ITE)	8262*** (2003)	6348*** (1649)	7983*** (1864)	7505*** (1714)
Other Assets (OA)	0.19*** (0.03)	0.20*** (0.03)	0.19*** (0.03)	0.21*** (0.03)
DDDxITE		3097** (1267)		
DDDxPPE			0.30 (0.38)	
DDDxOA				0.24* (0.13)
Constant	-5494 (3360)	-4487 (2799)	-5953 (3396)	-5332 (3066)
Number of firms	179	179	179	179
Number of observations	676	676	676	676
R-squared	0.75	0.77	0.76	0.77
Other Controls	Industry, Year, R&D per sales, Advertising expense per sales			

(1: Information Technology was the number of IT-employees, used to proxy for IT asset. Standard errors clustered around firms are in parentheses, *p<0.10, **p <0.05, ***p < 0.01.)

Conclusion

Case literature and economic theory suggest a potential connection between data driven decision making and productivity. By analyzing a large sample of firms, we find that DDD is indeed associated with higher productivity and market value, and that there some evidence that DDD is associated with certain measures of profitability (ROE, asset utilization). Our results are consistent with different measures of our IT variable and changes in the time period of the panel. In the productivity estimation, it appears to be robust to the use of instrumental variables methods to control for reverse causality or other forms of endogeneity. Collectively, our results suggest that DDD capabilities can be modeled as intangible assets which are valued by investors and which increase output and profitability.

References

Anderson, M.C., Banker, R.D., and Ravindran, S. 2003. "The New Productivity Paradox," *Communications of the ACM* (46:3), pp. 91-94.

Aral, Brynjolfsson, and Wu. 2009. "Testing Three Way Complementarities: Performance Pay, Monitoring and Information Technology," *ICIS*.

Aral, S., Brynjolfsson, E., and Wu, D. 2006. "Which Came First, It or Productivity? The Virtuous Cycle of Investment and Use in Enterprise Systems," *International Conference on Information Systems*, pp. 1-22.

Argote, L., McEvily, B., and Reagans, R. 2003. "Managing Knowledge in Organizations: An Integrative Framework and Review of Emerging Themes," *Management Science* (49:4), pp. 571-582.

Attewell, P., and Rule, J. 1984. "Computing and Organizations: What We Know and What We Don't Know," *Communications of the ACM* (27:12), pp. 1184-1192.

Ayres, I. 2008. *Super Crunchers: Why Thinking-by-Numbers Is the New Way to Be Smart*. Bantam.

Baily, M.N., Gordon, R.J., and Solow, R.M. 1981. "Productivity and the Services of Capital and Labor," *Brookings Papers on Economic Activity* (1981:1), pp. 1-65.

Balasubramanian, N., and Lee, J. 2008. "Firm Age and Innovation," *Industrial and Corporate Change* (17:5), p. 1019.

Barua, A., Kriebel, C.H., and Mukhopadhyay, T. 1995. "Information Technologies and Business Value: An Analytic and Empirical Investigation," *Information Systems Research* (6:1), p. 3.

Bharadwaj, A.S., Bharadwaj, S.G., and Konsynski, B.R. 1999. "Information Technology Effects on Firm Performance as Measured by Tobin's Q," *Management Science*), pp. 1008-1024.

Blackwell, D. 1953. "Equivalent Comparison of Experiments," *Ann. Math. Statist.* (24), pp. 265-272.

Bresnahan, T., Brynjolfsson, E., and Hitt, L. 2002. "Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence*," *Quarterly Journal of Economics* (117:1), pp. 339-376.

Bresnahan, T., Greenstein, S., and Henderson, R. 2009. "Schumpeterian Competition and Diseconomies of Scope: Illustrations from Leading Historical Firms in Computing."

Brynjolfsson, E., Hitt, L., and Yang, S. 2002. "Intangible Assets: How the Interaction of Computers and Organizational Structure Affects Stock Market Valuations," *Brookings Papers on Economic Activity: Macroeconomics* (1), pp. 137-199.

Brynjolfsson, E., and Hitt, L.M. 1996. "Paradox Lost? Firm-Level Evidence of High Returns to the Information Systems Spending," *Management Science* (42:4).

Brynjolfsson, E., and Hitt, L.M. 2003. "Computing Productivity: Firm-Level Evidence," *Review of economics and statistics* (85:4), pp. 793-808.

Brynjolfsson, E., Tambe, P., and Hitt, L. 2011. "The Extroverted Firm: How External Information Practices Affect Productivity," *Management Science* (Forthcoming).

Chen, K.C.W., and Lee, C.W.J. 1995. "Accounting Measures of Business Performance and Tobin's Q Theory," *Journal of Accounting Auditing and Finance* (10), pp. 587-587.

Clemons, E.K., and Thatcher, M.E. 1998. "Capital One: Exploiting an Information-Based Strategy," IEEE, pp. 311-320 vol. 316.

Cohen, W.M., and Levinthal, D.A. 1989. "Innovation and Learning: The Two Faces of R & D," *The Economic Journal* (99:397), pp. 569-596.

Davenport, T. 2009. "How to Design Smart Business Experiments," *Harvard Business Review*), February.

Davenport, T.H., and Harris, J.G. 2007. *Competing on Analytics: The New Science of Winning*. Harvard Business Press.

Devaraj, S., and Kohli, R. 2003. "Performance Impacts of Information Technology: Is Actual Usage the Missing Link?," *Management Science*), pp. 273-289.

Dewan, S., and Kraemer, K.L. 2000. "Information Technology and Productivity: Evidence from Country-Level Data," *Management Science*), pp. 548-562.

Dewan, S., Shi, C., and Gurbaxani, V. 2007. "Investigating the Risk-Return Relationship of Information Technology Investment: Firm-Level Empirical Analysis," *Management Science* (53:12), Dec, pp. 1829-1842.

Francalanci, C., and Galal, H. 1998. "Information Technology and Worker Composition: Determinants of Productivity in the Life Insurance Industry," *MIS Quarterly*), pp. 227-241.

Galbraith, J.R. 1974. "Organization Design: An Information Processing View," *Interfaces* (4:3), pp. 28-36.

Geyskens, I., Gielens, K., and Dekimpe, M.G. 2002. "The Market Valuation of Internet Channel Additions," *Journal of Marketing* (66:2), pp. 102-119.

Griliches, Z. 1981. "Market Value, R&D, and Patents," *Economics Letters* (7:2), pp. 183-187.

Hall, B.H. 1993. "The Stock Market's Valuation of R&D Investment During the 1980's," *The American Economic Review* (83:2), pp. 259-264.

Hall, R.E. 2001. "The Stock Market and Capital Accumulation," *American Economic Review* (91:5), pp. 1185-1202.

Hall, R.E., Cummins, J.G., and Lamont, O.A. 2000. "E-Capital: The Link between the Stock Market and the Labor Market in the 1990s," *Brookings Papers on Economic Activity* (2000:2), pp. 73-118.

Hannan, M., and Freeman, J. 1984. "Structural Inertia and Organizational Change," *American sociological review* (49:2), pp. 149-164.

Henderson, R. 1993. "Underinvestment and Incompetence as Responses to Radical Innovation: Evidence from the Photolithographic Alignment Equipment Industry," *The RAND Journal of Economics* (24:2), pp. 248-270.

Henderson, R.M., and Clark, K.B. 1990. "Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms," *Administrative Science Quarterly* (35:1), pp. 9-30.

Hirschey, M. 1982. "Intangible Capital Aspects of Advertising and R & D Expenditures," *The Journal of industrial economics* (30:4), pp. 375-390.

Hitt, L., and Brynjolfsson, E. 1997. "Information Technology and Internal Firm Organization: An Exploratory Analysis," *Journal of Management Information Systems* (14:2), p. 101.

Hitt, L.M., and Brynjolfsson, E. 1996. "Productivity, Business Profitability, and Consumer Surplus: Three Different Measures of Information Technology Value," *Management Information Systems Quarterly* (20:2).

Hitt, L.M., and Frei, F.X. 2002. "Do Better Customers Utilize Electronic Distribution Channels? The Case of Pc Banking," *Management Science*), pp. 732-748.

Hitt, L.M., Wu, D., and Zhou, X. 2002. "Investment in Enterprise Resource Planning: Business Impact and Productivity Measures," *Journal of Management Information Systems* (19:1), pp. 71-98.

Holmstrom, B., and Milgrom, P. 1991. "Multitask Principal–Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, and Organization* (7:special issue), p. 24.

Huergo, E., and Jaumandreu, J. 2004. "Firms' Age, Process Innovation and Productivity Growth," *International Journal of Industrial Organization* (22:4), pp. 541-559.

Jones, C.I. 1999. "Growth: With or without Scale Effects?," *The American Economic Review* (89:2), pp. 139-144.

Jorgenson, D.W., and Stiroh, K.J. 1999. "Information Technology and Growth," *American Economic Review, Papers and Precedings* (86:2).

Jorgenson, D.W., and Stiroh, K.J. 2000. "U.S. Economic Growth at the Industry Level," *American Economic Review Papers and Proceedings* (90:2).

Kohli, R., and Devaraj, S. 2003. "Measuring Information Technology Payoff: A Meta-Analysis of Structural Variables in Firm-Level Empirical Research," *Information Systems Research* (14:2), Jun, pp. 127-145.

Lavalle, S., Hopkins, M.S., Lesser, E., Shockley, R., and Kruschwitz, N. 2010. "Analytics: The New Path to Value," *MIT Sloan Management Review* (Fall).

Levitt, B., and March, J.G. 1988. "Organizational Learning," *Annual Review of Sociology* (14), pp. 319-340.

Loveman, G. 2003. "Diamonds in the Data Mine," *Harvard Business Review* (81:5), pp. 109-113.

McAfee, A. 2002. "The Impact of Enterprise Information Technology Adoption on Operational Performance: An Empirical Investigation," *Production and Operations Management* (11:1), pp. 33-53.

McAfee, A., and Brynjolfsson, E. 2008. "Investing in the It That Makes a Competitive Difference," *Harvard Business Review* (86:7/8), p. 98.

Melville, N., Gurbaxani, V., and Kraemer, K. 2007. "The Productivity Impact of Information Technology across Competitive Regimes: The Role of Industry Concentration and Dynamism," *Decision Support Systems* (43:1), Feb, pp. 229-242.

Melville, N., Kraemer, K., and Gurbaxani, V. 2004. "Review: Information Technology and Organizational Performance: An Integrative Model of It Business Value," *MIS Quarterly* (28:2), pp. 283-322.

Nass, C. 1994. "Knowledge or Skills: Which Do Administrators Learn from Experience?," *Organization Science* (5:1), pp. 38-50.

Nelson, R.R., and Winter, S.G. 1982. *An Evolutionary Theory of Economic Change*. Belknap press.

Oliner, S.D., and Sichel, D.E. 2000. "The Resurgence of Growth in the Late 1990s: Is Information Technology the Story?," *Federal Reserve Board Working Paper*).

Pentland, A., and Pentland, S. 2008. *Honest Signals: How They Shape Our World*. The MIT Press.

Radner, R. 1993. "The Organization of Decentralized Information Processing," *Econometrica: Journal of the Econometric Society*), pp. 1109-1146.

Saunders, A., and Brynjolfsson, E. 2010. "Valuing It-Related Intangible Capital," Working Paper, MIT.

Shapiro, C., and Varian, H.R. 1999. *Information Rules: A Strategic Guide to the Network Economy*. Harvard Business Press.

Simon, C.J., and Sullivan, M.W. 1993. "The Measurement and Determinants of Brand Equity: A Financial Approach," *Marketing science* (12:1), pp. 28-52.

Staiger, D., and Stock, J.H. 1997. "Instrumental Variables Regression with Weak Instruments," *Econometrica* (65:3), pp. 557-586.

Subramani, M., and Walden, E. 2001. "The Impact of E-Commerce Announcements on the Market Value of Firms," *Information Systems Research* (12:2), pp. 135-154.

Tambe, P., Hitt, L., and Brynjolfsson, E. 2009. "The Extroverted Firm: How External Information Practices Affect Productivity," *working paper*).

Tambe, P., and Hitt, L.M. 2011. "The Productivity of Information Technology Investments: New Evidence from It Labor Data," *SSRN eLibrary*).

Tushman, M.L., and Anderson, P. 1986. "Technological Discontinuities and Organizational Environments," *Administrative Science Quarterly* (31:3), pp. 439-465.

Appendix I. Business Practices or HR Survey

A. Basic information

1. Looking across your entire company, please rate the level of consistency in behaviors and business processes across operating units.

1. Not at all consistent	2. Slightly consistent	3. Somewhat consistent	4. Very consistent	5. Extremely consistent
☐	☐	☐	☐	☐

During the course of this survey, we will ask you about a number of business practices and behaviors. To help ensure that we collect accurate information, please respond only for the operating unit (e.g., BU or geography) that you know the best.

2. How would you rate the number of products you offer in your primary business?

	1	2	3	4	5
a. In absolute terms	Single product ☐	☐	Very broad product line ☐	☐	Multiple broad product lines ☐
b. Compared to competitors	Significantly fewer ☐	☐	About the same ☐	☐	Significantly more ☐

B. Employees

Questions in this section refer to "typical employees," defined as non-managerial, non-supervisory personnel directly involved in producing your company's main product or delivering its main service.

3. Please write in the title of the "typical employee" that is most important to creating value in your primary business unit: ______________________

4. How important is educational background in making hiring decisions for this "typical" job?

1. Not at all	2. Slightly	3. Somewhat	4. Very	5. Extremely
☐	☐	☐	☐	☐

5. To what extent are the following work practices used for this "typical" job? [Please remember to answer only for the operating unit you know the best.]

	1. Not at all	2. Slightly	3. Somewhat	4. Very	5. Extremely
a.Self-managing teams	☐	☐	☐	☐	☐
b.Cross-training	☐	☐	☐	☐	☐
c.Team building and group cohesion activities	☐	☐	☐	☐	☐
d.Employee involvement in important work processes	☐	☐	☐	☐	☐
e.Promotion based on teamwork	☐	☐	☐	☐	☐

6. For this "typical" job, who is generally responsible for the following tasks?

	1. Exclusively Typical Employees	2. Mostly Typical Employees	3. Equally	4. Mostly Managers	5. Exclusively Managers
a.Setting the pace of work	☐	☐	☐	☐	☐
b.Deciding how tasks should be performed	☐	☐	☐	☐	☐

c.Making hiring decisions	☐	☐	☐	☐	☐
d.Allocating staff or funding to projects	☐	☐	☐	☐	☐

7. What percentage of all employees use the following technologies in their regular work?

   a. General purpose computers (e.g. PC, workstation, or terminal): _____%

   b. Email: _____%

   c. Mobile email via Blackberry, Windows Mobile, etc: _____%

   d. Instant messaging: _____%

C. Activities

Questions in this section refer to "Core activities," defined as business processes that help your company outperform or differentiate you from competitors (e.g., Product development, Supply chain management). "Non-core activities" don't provide any such differentiation for your company and are not directly related to producing or bringing your main product or service to market.

8. Core activities

   a. Please list the 3 most important core activities for your primary business unit in order of ***importance***.
   1: ____________ 2: ____________ 3: ____________

   b. Would these have been different 5 years ago? Yes / No

9. For the first activity you listed above, how consistent is this activity across the following dimensions?

	1. Not at all consistent	2. Slightly consistent	3. Somewhat consistent	4. Very consistent	5. Extremely consistent
a. Consistency within business unit	☐	☐	☐	☐	☐

b. Consistency across functions (e.g., sales, finance, etc)	☐	☐	☐	☐	☐
c. Consistency across geographies	☐	☐	☐	☐	☐

10. On average, how long does it take your organization to perform the following activities?

	<1 week	1 month	3 months	6 months	1 year	1-2 years	>2 years	N/A
a. Design, create and introduce a new product or service after approval	☐	☐	☐	☐	☐	☐	☐	☐
b. Hire a new typical employee once need is identified	☐	☐	☐	☐	☐	☐	☐	☐
c. Respond to a competitor's price change	☐	☐	☐	☐	☐	☐	☐	☐
d. Introduce a new production technology or method	☐	☐	☐	☐	☐	☐	☐	☐
e. Respond to a competitor's new product or service	☐	☐	☐	☐	☐	☐	☐	☐

D. Organizational changes and innovation

11. Please rate whether the following factors **at your company** facilitate or inhibit the ability to make organizational changes.

	1. Inhibits significantly	2. Inhibits somewhat	3. No effect	4. Facilitates somewhat	5. Facilitates significantly
a. Financial resources	☐	☐	☐	☐	☐

b. Skill mix of existing staff	☐	☐	☐	☐	☐
c. Employment contracts	☐	☐	☐	☐	☐
d. Work rules	☐	☐	☐	☐	☐
e. Organizational culture	☐	☐	☐	☐	☐
f. Customer relationships	☐	☐	☐	☐	☐
g. IT systems	☐	☐	☐	☐	☐
h. Senior management involvement	☐	☐	☐	☐	☐

12. Where do you most often get ideas for innovation in your products or services? [Choose all that apply.]

    a. Product managers
    b. Line employees
    c. Service employees
    d. Sales and marketing staff
    e. R&D staff
    f. Executive management
    g. Other employees
    h. Customers
    i. Suppliers
    j. Distributors/retailers

13. How are decisions made for the creation of a new product or service?

	1	2	3	4	5
a.What is the typical basis for this decision?	Data ☐	☐	☐	☐	Experience and expertise ☐
b.Who is empowered to make this decision?	An individual ☐	☐	☐	☐	Group via discussion and consensus ☐
c.Where does the authority to make this decision reside?	Centrally (top of organization) ☐	☐	☐	☐	Locally ☐

E. Information technology

14. Currently, how effective is your IT organization in each of the following areas?

	1. Not at all	2. Slightly	3. Somewhat	4. Very	5. Extremely
a. Providing basic services cost-effectively (e.g., e-mail, laptops/desktops)	☐	☐	☐	☐	☐
b. Delivering new projects or enhancements on time and within budget	☐	☐	☐	☐	☐
c. Reactively responding to business needs to improve existing systems or functions	☐	☐	☐	☐	☐
d. Proactively engaging with business leaders to **refine existing** processes and systems	☐	☐	☐	☐	☐
e. Partnering with the business to **develop new business capabilities** supported by technology	☐	☐	☐	☐	☐

15. Please rank the following IT capabilities in terms of how important they are for your company now.

	Rank Order
a. IT delivers basic technology services to the business at the lowest cost	____
b. IT improves the efficiency and cost of our business operations	____
c. IT supports our growth ambitions by delivering services or products that set us apart from competitors	____
d. IT plays a leading role in transforming our business	____

F. Work practices

16. To what extent do the following statements describe the work practices and environment of your ***entire company***.

	1. Describes not at all	2. Describes slightly	3. Describes somewhat	4. Describes considerably	5. Completely describes
a. Executives devote a significant part of their time to recruiting	☐	☐	☐	☐	☐
b. Competitive benchmarks are regularly used in corporate strategic planning meetings	☐	☐	☐	☐	☐
c. Project teams often include employees from customers, suppliers or business partners	☐	☐	☐	☐	☐
d. Our company embeds many processes in technology	☐	☐	☐	☐	☐
e. We are usually the leading-edge adopter of new technologies in our industry	☐	☐	☐	☐	☐
f. Our company stresses operational excellence (efficient execution) over innovation	☐	☐	☐	☐	☐
g. We invest heavily in promoting our corporate culture	☐	☐	☐	☐	☐

h. Project teams often include employees from both business and IT working as peers	☐	☐	☐	☐	☐
i. We rely on extensive collaboration across functions/groups	☐	☐	☐	☐	☐
j. We depend on data to support our decision making	☐	☐	☐	☐	☐
k. We have a strong ability to make incremental changes or improvements to our business processes	☐	☐	☐	☐	☐
l. We have a strong ability to make substantial/ disruptive changes to our business processes	☐	☐	☐	☐	☐
m. We have a strong ability to disseminate changes to our business processes	☐	☐	☐	☐	☐
n. We rigorously track and ensure capture of the benefits or returns specified in business cases	☐	☐	☐	☐	☐
o. We monitor performance of activities through a set of well-defined metrics	☐	☐	☐	☐	☐
p. We have the data we need to make decisions	☐	☐	☐	☐	☐
q. Decisions about new IT investments are clearly rooted in business needs	☐	☐	☐	☐	☐
r. Our business executives understand IT	☐	☐	☐	☐	☐
s. Our IT executives understand the business	☐	☐	☐	☐	☐
t. Analysis and improvement are part of everyone's job	☐	☐	☐	☐	☐
u. We encourage experimentation to continually improve our offerings	☐	☐	☐	☐	☐
v. Our business processes have become more consistent over the past 3 years	☐	☐	☐	☐	☐
w. We have a modular business organization that promotes flexibility	☐	☐	☐	☐	☐
x. We have a cross-functional process	☐	☐	☐	☐	☐

orientation					

17. Do your employees have an effective way to do each of the following activities?

	1. Not at all	2. Slightly	3. Somewhat	4. Very	5. Extremely	6. N/A
a. Broadcast/show what they're working on within the company	☐	☐	☐	☐	☐	☐
b. Describe their experience and expertise	☐	☐	☐	☐	☐	☐
c. Collaboratively produce documents (e.g., memos, spreadsheets, blueprints)	☐	☐	☐	☐	☐	☐
d. Find a colleague with specific expertise	☐	☐	☐	☐	☐	☐
e. Find information/ knowledge about a particular topic	☐	☐	☐	☐	☐	☐

Appendix II. Technology Survey

A. Basic information

1. Does your group support the entire company (versus a specific operating unit or division)? Yes / No

2. What is the ending month and year for your company's most recent reporting year?
   Month ________ Year ________

3. IT Spending profile:
   a. Total Official IT Budget for Fiscal Year 2009: $_______ MM

   b. Please estimate the breakdown of this budget across the following dimensions [Please ensure these values add to 100%]:
      i. Application Development: _______ %
      ii. Application Maintenance: _______ %
      iii. Infrastructure : _______ %
      iv. Overhead/other: _______ %

   c. How did the IT budget change from the last fiscal year? [Choose one]
      i. Decreased by more than 10 percent
      ii. Decreased by 6 to 10 percent
      iii. Decreased by 1 to 5 percent
      iv. No change
      v. Increased by 1 to 5 percent
      vi. Increased by 6 to 10 percent
      vii. Increased by more than 10 percent

   d. Approximately what percentage of the IT budget is spent on outsourced services? _____%

   e. Please estimate the total IT-related spend outside the IT budget (also known as 'shadow IT'), as a percentage of the official IT Budget for FY 2009. [Choose one]
      i. 0-5%
      ii. 6-10%
      iii. 11-15%
      iv. 16-20%
      v. Greater than 20%

4. How many IT FTEs does your company have (including employees and contractors/external staff, but not including outsourced staff)? ________

5. Please describe the level of centralization or decentralization for your major IT resources.

a. Application development & maintenance	Primarily consolidated into one organization ☐	Mix of centralized and decentralized resources ☐	Primarily decentralized ☐
b. Infrastructure	Primarily consolidated into one organization ☐	Mix of centralized and decentralized resources ☐	Primarily decentralized ☐

6. Where does the executive in charge of IT report (if the role has matrixed reporting, please specify the most important relationship in practice)? [Choose one]
   a. CEO
   b. COO
   c. CFO
   d. BU executive(s)
   e. Other ___________

7. CIO Profile: [Select one answer for each of the following questions]
   a. How old is the CIO?
      Under 40 / 41-45 / 46-50 / 51-55 / 56-60 / Over 60 / Don't Know

   b. What is the highest degree that the CIO holds?
      BA/BS / MBA / MA/MS / PhD / None / Don't Know

   c. What was the CIO's undergraduate major?
      Economics / Computer Science / Humanities / Other / Don't Know

8. How many instances of the following systems are you running globally? [Select one] For each system, are there multiple (competing) systems in each geography/business unit?

	None or N/A	One (1)	Few (2-3)	Many (4+)		Are there multiple systems in a geography or BU? [Check if yes]
a. ERP/SCM	☐	☐	☐	☐		☐
b. CRM	☐	☐	☐	☐		☐
c. HR	☐	☐	☐	☐		☐
d. Financial systems	☐	☐	☐	☐		☐

B. Use of technology

9. How would you generally characterize your company's approach to use technology in support of business capabilities? (Choose one)
   a. Leading edge adopter
   b. Fast follower
   c. Mature adopter
   d. Late adopter

10. How is IT strategy primarily developed? [Choose one]
    a. Business and IT strategy are tightly integrated and influence each other.
    b. Business strategy is developed with some input from IT.
    c. Business strategy is developed first and used to guide IT strategy.
    d. Business and IT strategy are not linked.

11. Please rate whether information technology at your company currently facilitates or inhibits your ability to do the following:

	1. Inhibits Significantly	2. Inhibits somewhat	3. No effect	4. Facilitates somewhat	5. Facilitates Significantly	6. N/A or Not important
a.Create new products	☐	☐	☐	☐	☐	☐
b.Enter new markets	☐	☐	☐	☐	☐	☐
c.Share	☐	☐	☐	☐	☐	☐

knowledge						
d.Deliver year-over-year productivity gains	☐	☐	☐	☐	☐	☐
e.Track customer- or segment-level profitability	☐	☐	☐	☐	☐	☐

12. Please rank the following IT capabilities in terms of how important they are for your company now. [Rank order these statements from 1-4]

	Rank Order
a.IT delivers basic technology services to the business at the lowest cost	____
b.IT improves the efficiency and cost of our business operations	____
c.IT supports our growth ambitions by delivering services or products that set us apart from competitors	____
d.IT plays a leading role in transforming our business	____

13. Currently, how effective is your IT organization in each of the following areas?

	1. Not at all	2. Slightly	3. Somewhat	4. Very	5. Extremely
a.Providing basic services cost-effectively (e.g., e-mail, laptops/desktops)	☐	☐	☐	☐	☐
b.Building consistent systems and processes for each operating unit	☐	☐	☐	☐	☐
c.Delivering new projects or enhancements on time and within budget	☐	☐	☐	☐	☐
d.Reactively responding to	☐	☐	☐	☐	☐

business needs to improve existing systems or functions					
e.Proactively engaging with business leaders to **refine existing** processes and systems	☐	☐	☐	☐	☐
f.Partnering with the business to **develop new business capabilities** supported by technology	☐	☐	☐	☐	☐

14. What factors encourage or inhibit new technology investments for your company?

	1. Inhibits Significantly	2. Inhibits somewhat	3. No effect	4. Facilitates somewhat	5. Facilitates Significantly
a.Technical skills of existing IT staff	☐	☐	☐	☐	☐
b.Business domain knowledge of existing IT staff	☐	☐	☐	☐	☐
c.Ability to hire new IT staff	☐	☐	☐	☐	☐
d.Access to financial resources	☐	☐	☐	☐	☐
e.Business unit support for IT projects	☐	☐	☐	☐	☐
f.Senior management support	☐	☐	☐	☐	☐
g.Existing systems or data	☐	☐	☐	☐	☐
h.Existing IT infrastructure	☐	☐	☐	☐	☐
i.Ability to quantify project benefits	☐	☐	☐	☐	☐
j.Organizational culture	☐	☐	☐	☐	☐
k.IT project approval process	☐	☐	☐	☐	☐
l.Difficulties in system migration	☐	☐	☐	☐	☐

C. Technology practices

15. For the following questions about **governance and project prioritization practices,** please select the option that best describes your company's actual practices. (If there are differences between the "official policy" and actual practice, please indicate the actual practice.)

	1	**2**	**3**
A. How consistent are your project prioritization and approval processes?	Ad hoc process across operating units ☐	Standard processes within operating units, but no cross-unit prioritization ☐	Standardized, cross-enterprise process ☐
B. How involved is the business with IT projects?	Steering Committee only for the largest projects ☐	Varied engagement for medium/large projects ☐	Detailed, ongoing involvement for all projects ☐
C. How do you manage demand?	No real demand management – "first come, first served" basis ☐	Informal, qualitative prioritization ☐	All requests prioritized on quantified business value and effort ☐
D. How are IT costs shared/charged to the business?	Not charged to business units ("Cost center") ☐	Block allocations to business units ("IT Tax") ☐	Chargeback based on usage ("Service catalog") ☐
E. How do you track and manage the capture of project benefits?	Benefits are not incorporated into budgets ☐	Rigorously track and ensure capture of cost savings ☐	Rigorously track and ensure capture for both cost and revenue ☐

16. How do you measure and report performance for IT projects? [Choose one]
    a. No formal mechanism
    b. Ad hoc reporting
    c. Retrospective sKey Performance Indicators (KPIs) and reporting
    d. Real-time KPIs with only management-level reporting
    e. Real-time KPIs with cascading dashboards used by all levels of the organization

- *If you selected C, D, or E*: Do these KPIs or metrics include business-specific measures or outcomes?
  Yes / No

- *For all answers*: Do you use the same reporting approach for ongoing or 'business as usual' activities?
  Yes / No

17. For the following questions about **application development practices,** please select the option that best describes your company's actual practices. (If there are differences between the "official policy" and actual practice, please indicate the actual practice.)

	1	**2**	**3**
A. How do you track developer productivity?	No standard productivity measures ☐	Progress based on person months/weeks ☐	Output/functionality-based measures (e.g., Use Case or Function Points) ☐
B. How do you manage resources/skills?	Primarily dedicate staff to domains, applications or business-units ☐	Matrixed approach ☐	Company-wide pool of resources assigned as needed for specific project phases ☐
C. How much feedback do you get from users?	Users provide input at the beginning of a project ☐	Users provide input at the beginning and end of a project ☐	Users provide daily or weekly feedback on builds ☐
D. Do developers have uninterrupted periods of time scheduled to	Varies by developer or group ☐	Team members are encouraged to create time-blocks to focus on	Scheduled interruption-free times during the week are

complete tasks?		core tasks ☐	enforced ☐

18. Please describe your preferred approach to software development/acquisition. (For this question, "Core activities," are defined as business processes that help your company outperform or differentiate you from competitors (e.g., Product development, Supply chain management). "Non-core activities" don't provide any such differentiation for your company and are not directly related to producing or bringing your main product or service to market.)

a. What is your preference for buying vs. building applications?

a.Systems supporting core activities	Use packaged software "as-is" ☐	Customize packaged software ☐	"Build your own" ☐
b.Systems supporting non-core activities	Use packaged software "as-is" ☐	Customize packaged software ☐	"Build your own" ☐

b. Where do you most often run these applications?

a.Systems supporting core activities	Install and run on-premise ☐	Host at a 3rd party site ☐
b.Systems supporting non-core activities	Install and run on-premise ☐	Host at a 3rd party site ☐

c. What is your approach to sourcing applications?

a.Systems supporting core activities	Assemble best-of-breed ☐	Standardize on one or two vendors ☐
b.Systems supporting non-core activities	Assemble best-of-breed ☐	Standardize on one or two vendors ☐

19. For the following questions about **infrastructure management practices,** please select the option that best describes your company's actual practices. (If there are differences between the "official policy" and actual practice, please indicate the actual practice.)

	1	**2**	**3**
A. How do you manage service levels?	Few formal SLAs, expect "best effort" ☐	Most SLAs are defined/codified ☐	Complete set of SLAs, used in service catalog ☐
B. How do you manage standards/ products in infrastructure?	Infrastructure is custom built to project requirements ☐	Two or three standardized options at each level of the stack (e.g., Win/Unix or Oracle/DB2) ☐	Limit to a few standard bundles based on combinations of hardware, software and service levels ☐
C. How do you manage capacity (e.g., storage or computing power)?	Ad hoc, new capacity added as needed ☐	Annual process to define capacity requirements, based on historical trends ☐	Ongoing process to evaluate demand using a number of indicators ☐

20. How do you typically manage shadow IT activities? [Choose one]
    a. Try to prevent any IT activities unsanctioned by IT Department
    b. Tolerate, as long as shadow projects follow corporate standards
    c. Encourage, as long as shadow projects follow corporate standards
    d. Encourage, and use as a basis for experimentation, regardless of corporate standards
    e. Ignore

Chapter 3. CEO Pay and Information Technology

(with Erik Brynjolfsson)

Abstract

Compensation for CEO's and other top executives has increased dramatically in recent decades, drawing increasing scrutiny from policy-makers, researchers, and the broader public. This finds that information technology (IT) intensity strongly predicts compensation of top executives. The examination of panel data from 2507 publicly traded firms over 15 years controls for other types of capital, number of employees, market capitalization, median worker wages, industry turbulence, firm or industry fixed effects, and other factors. In order to examine the role of IT more closely, this study gathers detailed information on the business practices and information technology investments of 165 large publicly traded firms. The data-driven decision-making (DDD) is correlated with a significant increase in CEO pay even after controlling for the average worker's wage, suggesting that the data-driven decision-making may further widen the gap between top managers and average workers. The interpretation of this finding builds on earlier work which found a correlation between CEO pay, firm size, and general skill. First, IT increases the information available to the top executives for decision-making, magnifies their ability to propagate instructions throughout the firm, and improves the monitoring and enforcement of those instructions. When a top executive's instructions are implemented with higher fidelity, the fortunes of the firm will more closely mirror her performance. From the perspective of the top executive, this increases "effective size" of the firm that she controls and thus her marginal productivity. In turn, in an efficient market, this will increase overall executive compensation. Second, IT makes managerial decision-making skill more general and transferrable to different firms and industries. This generality of managerial skill allows the top talent to have more outside options, increasing the CEO pay.

Keywords: executive pay, IT impacts, decision-making, centralization, income inequality, executive compensation

Introduction

This paper examines the relationship between information technology (IT) and top executives' pay. A substantial rise in top executives' pay for the last three decades has been well-documented (i.e., Hall and Liebman 1998; Hall and Murphy 2003; Bebchuk and Fried 2005; Bebchuk and Grinstein 2005). For instance, the ratio of CEO pay to average worker pay has increased from 70 in 1990 to 300 in 2005, widening the income gap between the top earners and median workers.

We examine panel data from 2507 publicly traded firms over 15 years and find that IT intensity is significantly correlated with CEO pay, playing a key role in the recent rise in CEO compensation. The relationship is robust to the inclusion of controls for other types of capital, number of employees, market capitalization, median worker wages, industry turbulence, firm or industry fixed effects, and other factors. Our interpretation of this finding builds on earlier work that ties the increase in CEO pay to the increase in the average size of firms. Specifically, we hypothesize that information technology (IT) has amplified the ability of CEO's decisions to affect the fortunes of whole firm. This increases the *effective size* of the firms that CEOs can influence, resulting in an increase in the marginal productivity of CEOs and thus their pay.

Many researchers have reported that CEO pay is highly correlated with firm size (e.g. Roberts 1956; Barro and Barro 1990; Kostiuk 1990; Rosen 1992). Gabaix and Landier (2008) proposed a model in which the best CEO manages the largest firm at competitive equilibrium as this maximizes the CEO impact. In other words, the CEO of the largest firm has the highest marginal productivity and thus receives the highest pay. Our model extends their work by considering how the *effective size* of the firm varies with both the nominal size and IT intensity. We define the concept of effective firm size as the size of firm that CEO can effectively influence and control. The ability of CEO to manage large firms depends on technology for communicating instructions, replicating processes and monitoring employees. The less perfect the communication between top managers and employees, the less effective the CEO's impact becomes. For instance, if a CEO's instructions are accurately propagated throughout only a part of the firm, then the effective size is not as great as it would be if the instructions were accurately and precisely propagated to every part of the firm. Similarly, the ability of the CEO to use IT to better monitor compliance with instructions will also increase the effective size of the firm. The measure of firm size relevant to the marginal productivity of CEO is not the nominal firm size, measured by the number of employees or assets or market capitalization as used in previous literature, but the effective size of firm, measured by the resources that the CEO can actually influence and control. The benefits of having a "superstar" CEO will be greater the larger the effective size of the firm.

We argue that IT can increase the effective size of firm through a variety of mechanisms, and thus we hypothesize that IT-intensive firms pay their CEOs more than those of the same nominal size but with less IT. Examining over 2500 firms in 53 industries from 1992 to 2006, we find that the increase in IT intensity, as defined as the ratio of the IT capital stock to total capital stock, not only explains the difference in CEO pay across industries but is also strongly correlated with the increase in CEO pay over time.

The Role of IT

Effective Firm Size

There has been a radical increase in the use of IT in the US economy in the last few decades. The improvements in the power of IT and the sharp decrease in the price of computers have increased the computing power of American companies by over 33 times during the last two decades. Even in nominal terms, IT stock has grown. Between 1987 and 2004, IT stock per full-time employee (FTE) increased from $1600 to over $5100 while the percentage of IT of investment in total tangible wealth from nearly doubled from 12% to over 23% during this period.

The increases in the quality and quantity of IT have radically affected the ability of top executives to keep informed about activities throughout the organization, and to respond quickly and precisely with instructions. The increase in IT capital and investment in the last two decades has changed and penetrated every step of businesses. Enterprise Resource Planning (ERP) is one example; by now, companies in virtually every industry have adopted ERP despite that it is relatively recent. According to one estimate, spending on these enterprise IT platforms already accounted for over 50% of all U.S. corporate IT investment in 2001 (McAfee 2002). The enterprise IT platforms have enabled a rapid and reliable implementation of a new business practice throughout firms. The retail pharmacy firm CVS provides an illustrative example (McAfee 2005; Brynjolfsson et al. 2007). CVS had a problem of declining customer satisfaction, especially in their core prescription drug fulfillment business. The key problem was identified: there was often a delay in the insurance verification because of incomplete or inaccurate information about customers. Therefore, the management team changed the process. Overall customer satisfaction increased from 86% to 91% using the new system, which represented an enormous improvement.

Most importantly, once CVS had developed an improved prescription drug ordering process, they embedded it in their enterprise IT, which made it possible to rapidly replicate the process to over 4,000 retail stores throughout their entire firm within one year. The IT system assured that each pharmacist would implement the new process exactly as it was designed by management. This rapid and high fidelity firm-wide business process change across thousands of retail stores, enabled by their enterprise IT, led to significant increases in customer retention, increasing the firm value. Contrast this with a pre-IT approach which might have involved sending memos to store managers urging them to adopt the new process and perhaps training sessions exhorting the benefits of changing the procedure and explaining how to do it. In the old system, less than 100% of the store managers would be likely to comply with such a process change even after several years. Thus, the IT-enabled approach to business process replication amplified the impact of management's decision to change a business process. As a result, the firm's market value increased significantly reflecting the rapid and consistent firm-wide implementation of management's decision.

This example illustrates how the <u>effective</u> size of the firm affected by management decisions can be increased as IT facilitates the rapid and reliable implementation of new processes. In particular, the impact of CEO decisions can be magnified by IT as instructions and innovations are propagated, implemented and monitored. In turn, *ceteris paribus*, equilibrium CEO compensation can be expected to increase when the CEO's marginal revenue product is greater.

Generality of Managerial Skill

Numerous researchers have noted that some information is not easily transferable because it "tacit" (i.e., Polanyi 1958; Rosenberg 1982) or "specific" (i.e., Jensen and Meckling 1976) or "sticky"(i.e., Von Hippel 1994). Advancement in information technology has, however, brought in a wave of codification of knowledge. Balconi 2002) presented an illustrative example of the changes in the progressive codification of technological knowledge occurred over the last 30 years formerly embodied in skilled workers in the steel industry. Until the end of the 1960s instruments to measure the temperature and chemical composition of liquid steel were not widely available so measurements were carried out based on some physical characteristics observable by sight. For example, "in order to know the temperature of liquid steel, a sample was taken out of the furnace, poured upon an iron plate and the temperature was deduced by observing the forming of the spot, its shape and the way it solidified and attached itself to the plate." The ability to recognize the temperature by sight was acquired with 5 years of experience or more. By the 1970s the content and the temperature of liquid steel was routinely conducted by automated measurement tools but the process was still controlled by workers and line managers. By the end of 1980s the whole processing cycle was automated. Since the mid-90s, workers are principally involved in monitoring automated data collection and processing equipment, and approving process changes recommended by the processing algorithms. Over time, the knowledge once embodied in experienced workers has been supplanted by machine monitoring and largely automated decision making. A consequence is that the specialized knowledge only possessed by the most skilled individuals can now be embedded in computer instruments and available to distant operators in the operating rooms throughout the entire company. More broadly, firms now gather and propogate knowledge not only from production workers, but from their consumers, suppliers, alliance partners, and competitors much faster and more cheaply with the aid of information technology, make information available for distant decision makers.

Knowledge generation does not stop at passive analysis of data. New and broadly available software has enabled managers to conduct active experiments with their new business ideas and base their decisions on scientifically valid data. From banks such as PNC, Toronto-Dominion, and Wells Fargo to retailers such as CKE Restaurants, Famous Footwear, Food Lion, Sears, and Subway to online firms such as Amazon, eBay, and Google, firms test many business ideas through a randomized test before launch, called as "information-based strategy" (Davenport 2009). This information-based strategy alienates the high-level manager's tacit knowledge – the knowledge of what was or was not likely to be a successful business innovation. Just as the process of controlling steel processes was transformed by improved measurement of temperature, the innovative process in an online business is now being transformed by the improved measurement of customer preferences. However, this process automation cannot substitute for strategic decision making, although it does allow strategic decisions to be moved from managers at the level of a task or a business unit, to a manager with responsibility for the entire firm. At the same time, the skill to make such a strategic decision becomes a skill to interpret numeric values generated by the computers, much more transferrable to other firms and industries. Consequently, the managerial skills become more general with the advent of information technology and a top talent can be sought after with high pay.

Related Research

Our study is related to three streams of literature; one is the effect of IT on centralization and decentralization of decision-making, the second is the effect of IT on income inequality, and the third is the rise of CEO compensation.

A large stream of literature studies effects of IT on command, control, coordination, and organization of firms. Rather than reviewing the voluminous literature here, we refer the reader to previous studies (i.e., Leavitt and Whisler 1958;Rule and Attewell 1989;Gurbaxani and Whang 1991;Malone et al. 1987;Brynjolfsson and Mendelson 1993;Brynjolfsson et al. 1994;Brynjolfsson and Hitt 2000) and the literature reviews cited therein. In theory, IT could shift power either toward the center or away from it, leading to centralization or decentralization of a firm. In the former case, IT makes local knowledge available to top managers and management can be more centralized. In this case, one might expect higher CEO relative pay. On the other hand, in the latter case, IT makes local knowledge of one department available to employees (as well as to top managers) in other departments and employees can coordinate tasks among themselves more easily with the need for CEO involvement. In this case, one might expect less CEO relative pay. Ultimately, the net effect is an empirical question. Our study addresses this question by combining data from CPS, Compustat, and BEA and examining the correlation of IT with CEO pay.

Although not directly related to the scope of this study, changes in firm size induced by IT have been also studied based on the economic theories of Coase (1937) and Williamson (1973; 1981). While the effect of IT on firm size is a highly valuable research question, our study does not attempt to answer the question. We take firm size exogenous as numerous studies to research CEO compensations including Gabaix and Landier's paper (2008) have treated. It is conceivable that IT changes firm size and the correlation of CEO compensation and IT is mediated by the change of the nominal firm size itself. In our study we concentrate on the change of the effective firm size by IT while the nominal firm size is assumed exogenous. The interpretation of our results with and without this assumption is discussed in the result section.

The second stream of literature to which our study is related is the role of IT in leading to the wage inequality in the whole economy (i.e., Autor et al. 1998). Our view on the role of IT in increasing CEO pay is close to those of Garicano (Garicano and Rossi-Hansberg 2006;Garicano 2000). They explain that knowledge has hierarchy; some knowledge attached to lower-ranked employee is used to solve a routine problem while other knowledge attached to top managers to process a convoluted non-routine problem. As the cost of communication among agents decreases, the complicated problems that lower-ranked employees cannot solve can be easily passed to their superiors. Once solved, the solution can be disseminated easily with the lowered cost of communication. This can lead to the dependency of the problem-solving on a few "superstars" (Rosen 1981) and thus a higher wage for the superstars. Their explanation is consistent with our view that IT increases the effective size of a firm by easily passing non-routine problems to a few problem solvers and then disseminating the solution to the entire firm. Therefore, the dependency of the IT-intense firm on the few problem solvers and thus the sensitivity of the IT-intense firm to those few superstars become greater, leading to a higher compensation for those superstars in IT-intense firms. The CEOs may or may not be the problem solvers but they are a key to recognize, place, and reward the superstars in their firm. Our model does not, however, imply that a greater wage inequality within the same firm, when the firm is IT-intense, would be observed. Some firms or industries can concentrate many problem-solvers or superstars for the whole economy. The wage inequality within such firms may not be as great as in the whole economy. Our study focus is the differences of CEO pay across time and industry, not the wage inequality within the same firm, as a result of the increased effective size of the firm due to the increased IT intensity. To the best of our knowledge, no previous study has reported the determining power of IT in CEO compensation.

The third stream of literature relevant to our study is studies on CEO compensation. There have been many discussions, theories and explanations for the rise of CEO pay in academia as well as mass media. Following Gabaix and Landier's (2008) summary of the literature on the topic, we recapitulate the literature in four views.

The first is an agency view (Jensen and Murphy 1990;Dow and Raposo 2005;Holmstrom and Kaplan 2001). In order to give incentives or reward to CEOs to cope with more volatile and uncertain business environments and bring more innovative idea in the business strategy, firms have to pay their CEOs more. The second is a malfunctioning market view (Yermack 1997;Bertrand and Mullainathan 2001;Bebchuk and Fried 2005;Bebchuk et al. 2002;Hall and Murphy 2003). This view is that an efficient market mechanism doesn't work in CEO pay, explaining the rise of CEO pay by an increase in management entrenchment, a loosening of social norms against excessive pay, or simply inability to estimate the true cost of a common type of compensation such as stock-option compensation. The third is a view on the change of CEO job itself (Frydman 2005). CEO's skills have become more general and they can have more outside options, putting upward pressure on pay. The fourth view is a market equilibrium view (Gabaix and Landier 2008;Tervio 2008). This view is to explain the rise of CEO pay as the result of the rise in firm size. At market equilibrium, the most talented CEO is matched to the largest firm and paid the highest as the marginal productivity of the most talented CEO is the largest when she is matched to the largest firm and thus the efficiency increases.

Our paper is most closely linked to the market equilibrium view; the increased marginal productivity of CEOs resulted from the increase of firm size as well as from the shift to the generality of managerial skills is a major determinant of the recent rise of CEO pay. We articulate that what is relevant to the marginal productivity of CEO is an <u>effective</u> size of firm, an increasing function of IT intensity of the firm as well as the nominal firm size. The more IT-intense environment a firm has, the lower the cost of communication and access of knowledge as articulated by Garicano and Rossi-Hansberg (2006) and thus the larger the effective size of the firm becomes. As a result, the impact of CEOs becomes more significant, that is, the marginal productivity of CEO increases. The rise of CEO pay is, therefore, a manifestation of the increased marginal productivity of CEO. At the same time, the managerial skill to interpret data and make a strategic decision based on the data, even if the data itself may be firm-specific, becomes more general and the top talent can have more outside options. Therefore, the IT-intensive firms should pay more for their CEOs.

Our argument is also related to the study by Baker and Hall (Baker and Hall 2004), who explained the CEO pay based on the increased marginal productivity. CEOs perform two kinds of tasks; one is a task to vary with firm size and the other invariant to firm size. For example, a managerial action to buy a corporate jet may affect value of a firm to the same degree regardless of the firm size. A managerial decision to implement new processes or systems throughout a firm, however, will affect the value of a large firm more than that of a small firm. Some managerial decisions affect the value in an absolute amount and others in a percentage. CEOs need to be given an appropriate incentive to take her best efforts to both tasks. Especially when a firm is large the task to affect the firm value in percentage becomes more important and an appropriate incentive to elicit her best effort on the task should be given to the manger. We argue that when a firm is IT-intense as well as large, the impact of CEO decision to affect the firm value in percentage becomes even more important. When a firm is large, implementing a new business process or strategy across multiple departments and branch offices, even if underscored by the top managers, can be slow and prone to uneven and unreliable application of the new process. IT, however, can help a reliable and speedy implementation of the new process in the way anticipated by the top manager upon deciding on the adoption. In other words, the effective size of firm that its CEO can control becomes larger when IT helps her top managerial decision reach the entire firm faster and more reliably.

To examine the robustness of our model, we included other control variables in the analysis. The first was the median worker wage of the industry. Workers' wage difference caused by industry difference has long been recognized (i.e, Katz and Summers 1989;Krueger and Summers 1987;Krueger and Summers 1988;Gibbons and Katz 1992). For example, the median workers' wage in the industry of computer systems and related services, $65,000, was over 3-times higher than that in the industry of food services, $18,000, in 2005. The presence of the industry difference in wages has been often explained by factors such as differences in unmeasured abilities of workers across industry, compensating differences, rent sharing, and efficiency wages. Therefore, CEO pay in an industry where median worker wage is high may be high for those reasons. Since we are capturing the industry difference only by IT intensity, there may be an upward omitted variable bias for IT intensity. Our paper does not attempt to identify the cause for the median worker wage difference across industry but takes the cause as a given unknown factor. Therefore, we examined the impact of IT intensity on CEO pay by including the median worker wage as a control variable. Our hypothesis is that IT investment will affect the marginal productivity of top managers more than median workers, resulting in that other industry differences causing median workers' salary will affect CEO pay less once IT investment variable is separated.

The second control variable we included was the industry turbulence. Some researchers have reported that IT-intensive industries tend to be more turbulent than others (Brynjolfsson et al. 2007). It may be that firms in turbulent industries face more competitive business environments and benefit disproportionately from hiring more talented and thus more expensive CEOs. The increased CEO pay due to IT intensity may be a result of the more competitive business environment that the firm faces in her industry not the increased effective size of her firm. The industry turbulence, as defined as the average of rank changes of firms from year to year over industry, was included as a control variable to explore this possibility.

Model

Our model extends the model of Gabaix and Landier (2008), that can be summarized as the following:

1) CEOs have different levels and managerial talent and are matched to firms competitively;

2) in equilibrium, the best and thus the highest paid CEO manages the largest firms, as this maximizes their impact and economic efficiency; and

3) the CEO pay also increases with the average size of firm in the economy.

In other words, this theory states that, if there are two firms of different sizes and two managers of different talent, a competitive equilibrium exists in the way that the larger firm hires the more talented manager at a higher pay than the smaller firm does. Our contribution is to extend the concept of firm size in their theory to an "effective" size of firm, defined as the firm size that top managers can control and reach as information technology has integrated firm data and enabled a replicable, speedy, and firm-wide business process aided by an enterprise IT system, and test the theory with empirical analysis.

We briefly walk through Gabaix and Landier's model here. Consider the problem of hiring a CEO with talent, *T*, faced by a particular firm. The firm has "baseline" earnings of a_0. At t=0, it hires a manager of talent *T* for one period. The manager's talent *T* increases the firm's earnings according to

$$a_1 = a_0(1 + C \times T) = a_0 + a_0 \times C \times T \tag{1}$$

for some $C > 0$, which quantifies the effect of talent on earnings. Consider one extreme case that the CEO's actions at date 0 impact earnings only in period 1. The firm's earnings are $(a_1, a_0, a_0, \ldots.)$. The other extreme case is that the CEO's actions at date 0 impact earnings permanently. Then the earnings become $(a_1, a_1, a_1, \ldots)$. In both cases, the firm's problem can be written as the following:

$$\max_T S \times (1 + C \times T) - W(T) = \max_T S + S \times C \times T - W(T) \tag{2}$$

where $S = \frac{a_0}{1+r}$ for the former (where CEO's talent impacts the firm's earnings only the first period) or $\frac{a_0}{r}$ for the latter (where CEO's talent impacts the firm's earnings permanently), *r* is the discount rate, and $W(T)$ is the wage of CEO with talent, T. Eqn (1) can be generalized as $a_1 = a_0 + C{a_0}^{\gamma} + independent\ factors$, for a non-negative γ. The maximization problem of (2) becomes that:

$$\max_T (S + S^{\gamma} \times C \times T - W(T)) \tag{3}$$

Let's call $w(m)$ the equilibrium compensation of each CEO with index *m* which can be thought of the CEO's ranking or quantile in talent. The problem of (3) can be rewritten as that:

$\max_m (CS(n)^{\gamma} T(m) - w(m))$ (4)

A competitive equilibrium consists of:

i. a compensation function $W(T)$, which specifies the market pay of a CEO of talent T, and

ii. an assignment function $M(n)$, which specifies the index $m = M(n)$ of the CEO heading firm n in equilibrium,

such that

iii. each firm chooses its CEO optimally: $M(n) \in arg \max_m (CS(n)^{\gamma} T(m) - W(T(m))$, and

iv. the CEO market clears, that is each firm gets a CEO.

As in equilibrium there is associative matching: $m = n$,

$w(n) = \int_N^n CS(u)^{\gamma} T'(u) du + w(N)$ (5)

Assuming a specific functional form for $T'(u)$ with the use of the extreme value theory, Gabaix and Landier provided the solution for a CEO pay in terms of the size of a reference firm as well as the CEO's firm. The most relevant equation for our study is expressed in terms of the effective size of firm as following[6]:

$w = D_* \widehat{S_*}^{\alpha} \hat{S}^{\beta}$ (6)

where w is CEO pay, $\hat{S}$ and $\widehat{S_*}$ are the effective size of the CEO's firm and a reference firm, respectively, and α and β are positive constants. D_* is a function of the marginal talent of CEO, $T'(n_*)$ of the reference firm and the size of the reference firm. In all equations, the subscript * indicates attributes for a reference firm.

The effective size of a firm in Gabaix and Landier's model is a function of the sensitivity of the firm to CEO talent and the nominal size of the firm. We extend this concept further; CEO may be able to reach a greater portion of her firm if her firm is more integrated through its IT system. In other words, we hypothesize that the effective size (that CEO can affect) increases as a firm is more integrated through an information technology (IT) system. This hypothesis reflects that IT reduces, to name a few, the cost of communication between top managers and employees, the cost of implementing new business processes, and the cost of monitoring employee performances. For example, a large retailer such as Wal-Mart has adopted an enterprise IT system and the inventories and sales data from approximately 4,000 retail stores in the USA alone can be accessed and analyzed in its headquarter on real-time. Therefore, the effective size that top managers in its headquarter can reach has been widened with the centralized IT system allowing the global access to the data only local managers were accessible to in the

[6] This is the equation (25) of Proposition 3 on p.85 by Gabaix and Landier.

past without such a centralized IT system. Therefore, we assume that the effective size is an increasing function of both IT and nominal size.

$$\hat{S} = cI^{\delta}S \tag{7}$$

where c is a constant, I is the IT intensity, δ is a constant, and S is the nominal size of the firm.

The nominal size can be of various measures; they usually include employee numbers, sales, market capitalization, and assets. Following Gabaix and Landier's agnostic approaches, we used employee numbers, physical capital assets, and market capitalization as the nominal size and empirically chose the best proxy. As shown in the result section, the market capitalization turned out to be the best proxy for the nominal size. This is also what Gabaix and Landier found. The equations (6) and (7) yield that:

$$w = D_* c_*^{\alpha} I_*^{\alpha\delta} S_*^{\alpha} c^{\beta} I^{\beta\delta} S^{\beta} = A_*^{\mu} I_*^{\varepsilon} S_*^{\alpha} c^{\beta} I^{\rho} S^{\beta} \tag{8}$$

where $A_* = D_* c_*$, and μ, ε, and ρ are constants.

The resulting empirically testable equation is the following:

$$\ln(w_{i,t}) = \beta_1 + \beta_2 \ln(A_{*t-1}) + \beta_3 \ln(I_{*t-1}) + \beta_4 \ln(S_{*t-1}) + \beta_5 \ln(I_{i,t-1}) + \beta_6 \ln(S_{i,t-1}) \tag{9}$$

where *i* and *t* indicate an index for firm and time, respectively. This equation implies that the compensation for a CEO this year ($w_{i,t}$) is determined by the effective size of a reference company (I_{*t-1} and S_{*t-1}) as well as the CEO's firm ($I_{i,t-1}$ and $S_{i,t-1}$) in the previous year along with other characteristics associated with the reference company (A_{*t-1}). This time lag lessens the potential simultaneity problem between CEO pay and the effective firm size.

A_* is a variable relevant to a reference company such as the marginal talent of CEO of the reference company and the sensitivity of the reference firm to its CEO talent (not captured by the effective size). This measure captures business environments that the CEO's firm of interest is under, on the assumption that the CEO's firm would face a similar environment as its reference firm. For example, a firm in a highly-competitive industry may reward its CEO's talent more because it takes more talent to win over the competition; a firm in the industry where more educated workers belong may also reward its CEO because it needs a CEO to understand the complexities of the tasks that the workers of his firm may have to deal with. Therefore, we used two measures for A_* which can capture the business environment to some degree: industry turbulence and industry-median worker income. To compare our results with Gabaix and Landier's results, we conducted our analysis with and without the variable, A_* as Gabaix and Landier assumed that A_* may be the same for all firms and dropped it in their empirical analysis. In both cases, IT intensity was statistically significant.

In our test, we used IT intensity, defined as the ratio of IT capital stock to the total capital stock of plant, equipment and IT capital, as the IT variable. I_{*t-1} is IT intensity for a reference company in year t-1. Whether we choose a median-sized firm or the 250th-ranked firm[7] in year t-1 as a reference firm, I_{*t-1}

[7] The 250th-ranked firm was a reference firm in Gabaix and Landier's empirical analysis.

depends mostly on the year variable as IT intensity has substantially grown as time. The variation in IT intensity over time for a reference firm was not very different from the variation in IT intensity over time for the whole economy. Therefore, we used the IT intensity for the whole economy in year t-1 for I_{*t-1}except the models with fixed year effect. In the models with fixed year effect, the IT intensity for the whole economy was not included because there is no variation within a fixed year. The model with fixed year effect resolves the issue not whether the difference in the firm-level IT intensity explains the differences in CEO pay within an industry at a given year, but whether the differences in CEO pay for the same sized firm across industry are explained by the differences in the industry-level IT intensity at a given year.

We used industry-level IT as our measure of $I_{i,t-1}$ in the equation (9). This stems from two reasons, one practical and the other theoretical. The practical reason is that high quality firm-level IT data does not exist for the full period we seek to consider. Accordingly, industry-level IT intensity can serve as a noisy measure of firm-level IT intensity. The second reason for this approach is that using the industry-level IT intensity can mitigate a potential endogeneity problem; suppose that for some reason highly paid CEOs tend to spend more on IT investments in their own firms, leading to a positive correlation between CEO pay and firm-level IT intensity. By taking the IT intensity at the industry level and by also using the previous year's data as a covariate in the regression to determine the current year's CEO pay at firm level, the endogeneity problems are reduced.

The variables, S_{*t-1} and $S_{i,t-1}$, are the nominal size of a reference company and the CEO's firm of interest in the previous year, respectively. For firm-level analysis, we chose the 250th-ranked firm in the previous year as the reference firm following Gabaix and Landier's model. For industry-level analysis, there is no difference between a reference firm and a firm of interest and we treat those as one industry size variable. We used the total labor cost and the total net value of physical capital in each industry in the previous year for the industry size variable (S_{*t-1} and $S_{i,t-1}$).

Two models, one using employee number and net value of physical capital and the other using market capitalization, were compared to select a better proxy for firm size in the context of CEO pay. As shown in Table 1, the market capitalization captured the firm size better, consistent with Gabaix and Landier's result. For all other models at firm level, we used the market capitalization as the firm size variable.

It should be also noted that we excluded the IT-producing industries; computer software and hardware industries. Although the impact of CEO decisions and skills on the firm value in these industries is likely to be amplified due to IT as well, IT in these industries may affect many other aspects of their businesses in a different way. Therefore, the correlation of CEO pay with IT intensity in these computer industries may be very different from that in other industries merely using IT. To focus on our main points of the influence of IT usage on CEO pay, we excluded the IT-producing industries.[8]

[8] However, the inclusion of IT-producing industries does not change our main results.

Data Sources and Variables

IT intensity at industry level

We followed a method similar to one described in a previous study (Brynjolfsson et al. 2007) to estimate IT intensity at industry level. In summary, IT intensity was defined to be IT capital stock divided by the sum of Plant, Equipment and IT capital stock. The capital stock data for IT, Plant and Equipment are available from the Bureau of Economic Analysis's (BEA) "Tangible Wealth Survey" for 63 industry sectors at approximately three-digit NAICS level from 1947 to 2006. We used two variables for IT intensity; one is the IT intensity in the whole economy each year and the other is the IT intensity of each industry each year. As IT intensity is either at industry-level or at the whole economy, our firm-level analysis uses the industry-level IT intensity data.

CEO compensation and firm-level company data

We used two Compustat databases, *Industrial* and *Executives*, for the period from 1992 to 2006. Compustat provides commercially available databases for public companies. The Industrial database provides firm characteristics such as physical assets, employee numbers, common stock and sales, while the Executives database provides data on compensation of the top executives up to 13 from each company. The Executives database is compiled from proxy statements filed by the companies in compliance with Securities and Exchange Commission (SEC) regulations and covers S&P 1500 companies starting in 1992.

The CEO compensation, w, was the variable, *tdc1*, from Compustat Executives data set. The *tdc1* includes salary, bonus, other annual, restricted stock grants, LITP payouts, all other, and value of option grants. We selected companies with at least three executives included in the database.

For industry-level analysis, the compensation (tdc1) of CEO or three top executives from each firm was summed over each industry. As a size variable, we used two variables; one is the labor cost and the other is the physical capital. The labor cost of each industry was estimated by the total employee number of the selected companies in each industry multiplied by the median worker wage of the industry. The median worker wage of each industry was obtained from the Current Population Survey (CPS) as explained later.

For firm-level analysis, the firm size, s, was represented in two ways; one is the employee number and physical capital, the other is the market capitalization following Gabaix and Landier (2008). They reported that the market capitalization was the best proxy for firm size. It was calculated by the equation, *data199* x *abs(data25) + data6 – data60 – data74*, where *data199* is the share price of closing at fiscal year, *data25* is Common Shares Outstanding, *data6* is Total Assets, *data60* is Total Common Equity, and *data74* is Deferred Taxes. All nominal quantities were converted into 2000 dollars using the GDP deflator from the BEA.

Firm-level business practices

Our business practice and information system measures are estimated from a survey administered to senior human resource (HR) managers and chief information officers (CIO) from large publicly traded firms in 2008. We received matched responses (both HR and CIO) from 127 firms, HR only responses from 122 firms, and only CIOs-only responses from 81 firms. The survey asks over 80 questions on business practices as well as organization and usage of information systems. The questions extend a previous wave of surveys on IT usage and workplace organization administered in 1995-1996 and 2001 (Hitt and Brynjolfsson 1997 Brynjolfsson et al.), but adds additional questions on organizational structure, the usage of information for decision making, and the consistency of their organizational practices. To test our hypothesis, we used the survey response to construct measures of firms' organizational practices. We combine these measures with publicly financial data and CEO compensation. This yielded 164 firms with complete data for an analysis of firm productivity, and 165 firms during the same period with the variables needed to analyze CEO compensation. Our sample spans manufacturing, retail/wholesale trade, information, and finance/insurance industries over the period from 2005 to 2009.

We constructed our key independent variable, data-driven decision-making (DDD), from three questions of the survey: 1) the usage of data for the creation of a new product or service, 2) the usage of data for business decision-making in the entire company, and 3) the existence of data for decision-making in the entire company. Two other measures from the survey were also constructed to indicate the general employee's human capital; 1) the importance of typical employee's education and 2) the average of % of employees using e-mail and % of employees using PCs, workstation, or terminal. All measures except the ones based on percentages were capture in 5-point Likert scales.

We created DDD by first standardizing each factor with mean of zero and standard deviation of 1 and then standardizing the sum of each factor for each composite measure:

DDD = STD(STD(use of data for creation of a new product/service) + STD(use of data for business decision in the entire company) + STD(existence of data for such a decision))

Industry-level median worker's income

We used data from the Current Population Survey (CPS) to estimate the median worker's income for the period from 1992 to 2006. CPS is a monthly survey of about 50,000 households conducted by the Bureau of the Census for the Bureau of Labor Statistics. The survey provides data related to employment such as wage, industry and occupation and a variety of demographic characteristics such as age, marital status and education attainment. We used their wage data for fulltime workers, *incwage*, to estimate median workers' wage income in each industry each year. We also used their total income data, to be more comparable to CEO pay data which include bonuses and stock options as well as salaries, and compared its result with that using *incwage*. The results from both variables were very close and we report the result using *incwage*.

Industry Classification

A hybrid industry classification was created to merge the IT capital data from BEA and the median worker's wage income data from CPS. The resulting industry classification has 58 sectors as listed in the appendix, fewer than 63 sectors from BEA. The industry classification that CEO compensation and company data are based on is SIC and NAICS. The SIC and NAICS industry codes were, therefore, converted to our hybrid industry classification for the analysis. It should be noted that the industry classification used by Gabaix and Landiers (2008) shown in Table 2 is Fama-French industry classification (48 sectors), different from our industry classification (58 sectors). Excluding the observations with missing variables and firms in IT-producing industries, our data universe consisted of 2,507 firms, 20,087 observations, and 53 industry sectors over 15 years.

Results

The number of employees in the firm and firm's physical capital (net property, plant, and equipment) were used as size variables in the model 1-1A. The IT variables were statistically and economically significant. Following Gabaix and Landier (2008), we then used market capitalization as a size variable as in the model 1-1B.

Inclusion of IT variables in the model changed the coefficient associated with the reference firm size (that is, the size of the 250^{th}-ranked firm). We were able to replicate the findings of Gabaix and Landier when we used their model (the result was not shown). However, when we added IT to the model, the coefficient associated with the size of the 250^{th}-ranked firm became a half of the value in the earlier model (model 1-1B in Table 1). The effect of the size of the reference firm seems to be captured by the IT intensity in the whole economy. Moreover, the overall IT intensity of the whole economy is as important as the firm size in determining the CEO pay as shown in each of the models of Table 1.

In the models with fixed industry effect or fixed firm effect (1-2 and 1-3), IT intensity of the firm's industry remains positive, but is not statistically significant. This is understandable, since fixed effects undoubtedly over-control for the effects of IT that are constant within a firm or industry. For any given year, however, the CEO pay difference is explained by the industry difference in IT intensity as shown in model 1-4; each 1% change in IT intensity is correlated with a bit more than 0.1% change in CEO pay across firms.

Robustness Check of Firm-Level Analysis

We further checked the robustness of the model by including two potential variables which may have the same trend with IT intensity. This is to relax Gabaix and Landier's assumption that A_* in Eqn (9) is the same across firms. We argue that A_* may be the same across firms in their industry but different from firms in other industries. Therefore, we used industry-level measures for two variables. The first was median worker wage. IT-intense industries may be industries to use more highly educated employees and pay their employees more. Accordingly, CEOs in those industries may be paid more. The second was industry turbulence. Throughout all models in Table 2, the IT intensity variables positive and is usually statistically significant.

Table 1. Firm-Level Analysis of CEO Pay I[9]					
	Log(CEO Compensation)$_{i,t}$				
	(1-1A)	(1-1B)	(1-2)	(1-3)	(1-4)
Log(Employee Number, $S_{i,t}$)	0.15*** (0.01)				
Log(Physical Capital, $S_{i,t}$)	0.18*** (0.01)				
Log(Market Capitalization, $S_{i,t-1}$)		0.42*** (0.01)	0.44*** (0.004)	0.35*** (0.009)	0.42*** (0.004)
Log(250th-ranked Firm Market Capitalization in the whole economy, S_{*t-1})		0.28*** (0.05)	0.25*** (0.05)	0.37*** (0.04)	
Log(IT intensity of the whole economy, I_{*t-1})	1.21*** (0.05)	0.40*** (0.10)	0.45*** (0.10)	0.37*** (0.09)	
Log(IT intensity of their industry, $I_{i,t-1}$)	0.25*** (0.02)	0.12*** (0.01)	0.04 (0.03)	0.03 (0.03)	0.12*** (0.005)
Industry Fixed Effect	No	No	Yes	No	No
Firm Fixed Effect	No	No	No	Yes	No
Year Fixed Effect	No	No	No	No	Yes
Number of Firms	2507	2507	2507	2507	2507
Number of Industries	53	53	53	53	53
Observations	20087	20087	20087	20087	20087
Adjusted R-squared	0.32	0.46	0.50	0.68	0.46

Median worker wage

An unexpected result was that, when IT and the other controls are included, the median worker wage was actually negatively correlated with CEO pay. This result is not inconsistent with our hypothesis that problem solving is concentrated to a few superstars as IT becomes more prevalent, leading to a higher pay to the superstars and perhaps less responsibility and pay for the median worker in firms that allocate decision-making in this way. This result should be, however, taken with caution. Our data of median worker income is at industry level including small firms while CEO pay is at firm level from S&P 1500 firms. Therefore, our result doesn't mean that CEO compensation is negatively correlated with median worker compensation in the same firm or a firm of similar size. It is well-known that large firms tend to pay their employees higher salary than smaller firms. However, using the income data of median workers only working for the subset of relatively large firms do not change our results of IT being significantly correlated with CEO pay (results are not shown), suggesting that IT increases the pay for CEO or a few superstars much more highly than median workers even among large firms.

Industry Turbulence

Industry turbulence was also a significant variable to determine CEO pay in these specifications, in similar magnitude to the industry-level IT intensity (Table 2). This result is consistent with the explanation that the highly competitive business environment in recent decades may have increased CEO pay Dow and Raposo 2005. As with firm size itself, the increase in industry turbulence, however,

[9] ***, **, *, and ~ indicate 0.1%, 1%, 5%, and 10% level of significance, respectively.

may also be caused in part by high IT intensity. In fact, this has been specifically argued by some researchers Brynjolfsson et al. 2007: IT-intense industries may be highly turbulent industries and IT intensity may allow the industry turbulence. More turbulent industries may benefit more from highly talented CEOs and thus pay them more. Thus, our coefficients may actually underestimate the role of IT. The full effect of IT may be larger than is implied by the coefficient of IT intensity in these models, since a part of IT's effect may work through changes in industry turbulence and nominal firm size.

Table 2. Firm-Level Analysis of CEO Pay II

	Log(CEO Compensation)$_{i,t}$			
	(2-1)	(2-2)	(2-3)	(2-4)
Log(Market Capitalization, $S_{i,t-1}$)	0.43*** (0.007)	0.44*** (0.004)	0.35*** (0.009)	0.43*** (0.003)
Log(250th-ranked Firm Market Capitalization, S_{*t-1})	0.39*** (0.05)	0.35*** (0.05)	0.44*** (0.04)	
Log(IT intensity of the whole economy, I_{*t-1})	0.26~ (0.10)	0.29** (0.11)	0.26** (0.09)	
Log(IT intensity of their industry, $I_{i,t-1}$)	0.10*** (0.01)	0.03 (0.03)	0.02 (0.03)	0.10*** (0.006)
Log(Industry turbulence, A_{*t-1})	0.08*** (0.01)	0.09*** (0.02)	0.07*** (0.01)	0.08*** (0.006)
Log(median worker wage in their industry, A_{*t-1})	-0.36*** (0.04)	-0.12~ (0.07)	-0.08 (0.06)	-0.36*** (0.02)
Industry Fixed Effect	No	Yes	No	No
Firm Fixed Effect	No	No	Yes	No
Year Fixed Effect	No	No	No	Yes
Number of Firms	2507	2507	2507	2507
Number of Industries	53	53	53	53
Observations	20087	20087	20087	20087
Adjusted R-squared	0.48	0.51	0.68	0.48

Nominal Firm Size and IT intensity

Our model views that IT widens top managers' reach inside their firms of a given size. However, IT can also play an important role in changing firm size. Taking theories of firms on transaction or coordination costs (Coase, 1937; Williamson, 1973, 1981), many researchers have proposed that IT advancement would change firm size. IT can reduce transaction cost between firms, favoring the market over firms and leading to smaller firms (i.e., Brynjolfsson et al. 1994;Malone et al. 1987). On the other hand, IT can reduce coordination cost within a firm and lead to a larger firm (i.e., Zhu 2004; Armour and Teece 1978). Some researchers also reported that IT may lead to smaller firms in manufacturing and larger firms in retail and service sectors (Wood et al. 2008). Other research results on the effect of IT and firm size showed that large firms in IT-intensive industries tend to become larger, suggesting that IT may increase firm size of large firms (Saunders 2008). The role of IT as well as other factors in determining firm size is beyond the scope of this paper. It is, however, worth mentioning how the role of IT in changing firm size can change our result. In case that IT leads to larger firm, IT has an indirect impact on CEO pay increase by increasing firm size while in our model IT has a direct impact on CEO impact by increasing CEO reach within her firm. Our estimate only captures the direct impact and may be underestimated. In case that IT leads to smaller firms, IT may decrease CEO pay by decreasing the firm size and as a result the CEO pay increase may be lowered. However, if IT in large firms leads to the increase in their nominal firm size for most sectors, as in Saunders' study (2008), the nominal size of the firms in our study, S&P 1500 firms, may have increased due to IT. Then IT may increase the nominal firm size as well as the effective firm size and our study may have underestimated the IT effect on CEO pay.

The basics of our model are to take the nominal or effective firm size as exogenous from CEO pay. However, it is possible that the growth of the firm size and CEO pay may reflect a positive feedback cycle. For example, our model can be read that the 2nd-ranked CEO is match to run the 2nd largest firm as measured by *effective* size (reflecting both IT and the nominal size), in a competitive equilibrium at t=1. If the CEO grows the firm to become the 1st-ranked largest firm in effective size by the end of t=1; then she will be managing the largest firm at t=2. Thus, the CEO who managed the 2nd-ranked largest firm in effective size at t=1 become most highly paid at t=2.

Interestingly, while CEO pay has been found to be correlated with nominal firm size after 1970, no correlation was found for earlier years (Frydman and Saks 2007). Our result is consistent with this fact as well. The integration of data within firm has become possible with IT advancement only for the last few decades. A case of Johnson and Johnson (Ross 1995) illustrates an example of increase in the effective firm size upon the introduction of IT system. Johnson & Johnson, founded in 1886, had a long history of managing its operating companies as independent business, embracing the decentralized management approach as a path to increased flexibility, accountability and creativity. Each operating company had its own marketing and sales units and its executives were compensated based on the performance of their company, not the corporation as a whole. Problems with the decentralized management approach surfaced when customers tried to limit their vendor interactions; in certain cases one customer got calls from up to 18 representatives from different J&J companies. Over time, Johnson and Johnson started introducing various structures to increase inter-company cooperation and a more coherent management practice. A business strategy to achieve the coherent management practice over the whole corporation was to build and introduce a corporate IT system to share data across business units and practice cross-company cooperation among 160 operating companies. In other words, the effective size of Johnson and Johnson that the corporate headquarters reached before the introduction of IT system could be much smaller than that after the introduction of IT system while their nominal size may seem unchanged. Therefore, CEO pay may seem poorly correlated with the nominal firm size when the nominal firm size does not accurately reflect its effective size.

Conclusion

This paper examined the importance of information technology (IT) in CEO compensation. IT intensity was generally the most significant variable explaining CEO pay. We reason that IT intensity allows a more integration of a firm and increases the effective size of the firm that a top manager like CEO can reach and control. Therefore, the firm's CEO pay, the marginal productivity of the firm's CEO, would be based on the effective size of the firm, a function of not only its nominal size but also its IT intensity. We argue that, in equilibrium, the CEO of the largest firm in terms of effective size will tend to be paid the most. As effective firm size grows, reflecting improvements in IT, so will the average pay of CEOs and other top executives. Our model was robust with inclusion of other potential variables such as median worker wage and industry turbulence.

While the correlations we find are fairly robust, it is also possible that IT intensity affects CEO pay through other mechanisms. For instance, IT may have standardized skill sets and IT-intense industries may require less firm-specific skills than other industries. This is consistent with the result by Frydman (2005) that CEO skills have become less firm-specific. Or perhaps, IT is a proxy for the information intensity of a firm, and information-intensive industries pay their CEOs more. However, in any event, the finding of such a strong and large correlation between relative CEO pay and IT intensity is both novel and economically significant.

Our findings have important implications for how we think about the growth in CEO pay and inequality. If the underlying drivers relate to changes in the way technology propagates instructions and innovations within the firm, then firms may wish to adjust their recruiting, hiring and retention policies as well as their IT investments and architectures. Similarly, policy-makers will want to craft tax and labor policies which can mitigate inequality while preserving efficiencies in talent matching and executive decision-making as much as possible. To the extent that we expect continuing advances in Moore's law and the power of IT to amplify and propagate decisions, the relationships we uncover in this paper can likewise be expected to continue and even grow. This could portend a world with increased productivity growth as innovations are ever replicated with ever more fidelity and speed throughout the firm, but also even more "superstar" pay for top executive and an even more skewed distribution of income.

References

Ante, S. E. "Giving the Boss the Big Picture" *Business Week*, February 13, 2006.

Armour, H., and Teece, D. 1978. "Organizational Structure and Economic Performance: A Test of the Multidivisional Hypothesis," *The Bell Journal of Economics*), pp. 106-122.

Autor, D.H., Katz, L.F., and Krueger, A.B. 1998. "Computing Inequality: Have Computers Changed the Labor Market?," *Quarterly Journal of Economics* (113:4), pp. 1169-1213.

Baker, G.P., and Hall, B.J. 2004. "Ceo Incentives and Firm Size," *Journal of Labor Economics* (22:4), pp. 767-798.

Balconi, M. 2002. "Tacitness, Codification of Technological Knowledge and the Organisation of Industry* 1," *Research Policy* (31:3), pp. 357-379.

Barro, J.R., and Barro, R.J. 1990. "Pay, Performance, and Turnover of Bank Ceos," *Journal of Labor Economics* (8:4), pp. 448-481.

Bebchuk, L., and Fried, J. 2005. "Pay without Performance: Overview of the Issues," *Journal of Applied Corporate Finance* (17:4), pp. 8-23.

Bebchuk, L., Fried, J., and Walker, D. 2002. "Managerial Power and Rent Extraction in the Design of Executive Compensation," *University of Chicago Law Review* (69), pp. 751-846.

Bebchuk, L., and Grinstein, Y. 2005. "The Growth of Executive Pay," *Oxford Review of Economic Policy. Winter* (21:2), pp. 283-303.

Bertrand, M., and Mullainathan, S. 2001. "Are Ceos Rewarded for Luck? The Ones without Principals Are*," *Quarterly Journal of Economics* (116:3), pp. 901-932.

Brynjolfsson, E., and Hitt, L. 2000. "Beyond Computation: Information Technology, Organizational Transformation and Business Performance," *The Journal of Economic Perspectives*), pp. 23-48.

Brynjolfsson, E., Malone, T., Gurbaxani, V., and Kambil, A. 1994. "Does Information Technology Lead to Smaller Firms?," *Management Science*), pp. 1628-1644.

Brynjolfsson, E., McAfee, A., Sorell, M., and Zhu, F. 2007. "Scale without Mass: Business Process Replication and Industry Dynamics," MIT and HBS, Cambridge.

Brynjolfsson, E., and Mendelson, H. 1993. "Information Systems and the Organization of Modern Enterprise," *Journal of Organizational Computing* (3:3), pp. 245-255.

Brynjolfsson, E., Tambe, P., and Hitt, L. 2011. "The Extroverted Firm: How External Information Practices Affect Productivity," *Management Science* (Forthcoming).

Davenport, T. 2009. "How to Design Smart Business Experiments," *Harvard Business Review*), February.

Dow, J., and Raposo, C. 2005. "Ceo Compensation, Change, and Corporate Strategy," *Journal of Finance*), pp. 2701-2727.

Frydman, C. 2005. "Rising through the Ranks: The Evolution of the Market for Corporate Executives, 1936-2003," Harvard University.

Frydman, C., and Saks, R. 2007. "Historical Trends in Executive Compensation, 1936-2005," MIT Sloan School of Management

Federal Reserve Board of Governors, Cambridge.

Gabaix, X., and Landier, A. 2008. "Why Has Ceo Pay Increased So Much?," *Quarterly Journal of Economics*), February, pp. 49-99.

Garicano, L. 2000. "Hierarchies and the Organization of Knowledge in Production," *Journal of Political Economy* (108:5), pp. 874-904.

Garicano, L., and Rossi-Hansberg, E. 2006. "Organization and Inequality in a Knowledge Economy*," *The Quarterly Journal of Economics* (121:4), pp. 1383-1435.

Gibbons, R., and Katz, L. 1992. "Does Unmeasured Ability Explain Inter-Industry Wage Differentials?," *Review of Economic Studies* (59:July), pp. 515-535.

Gurbaxani, V., and Whang, S. 1991. "The Impact of Information Systems on Organizations and Markets," *Communications of the ACM* (34:1), pp. 59-73.

Hall, B.J., and Liebman, J.B. 1998. "Are Ceos Really Paid Like Bureaucrats?," *Quarterly Journal of Economics* (113), pp. 653-691.

Hall, B.J., and Murphy, K.J. 2003. "The Trouble with Stock Options," *Journal of Economic Perspectives* (17:3), pp. 49-70.

Hitt, L., and Brynjolfsson, E. 1997. "Information Technology and Internal Firm Organization: An Exploratory Analysis," *Journal of Management Information Systems* (14:2), p. 101.

Holmstrom, B., and Kaplan, S. 2001. "Corporate Governance and Merger Activity in the United States: Making Sense of the 1980s and 1990s (Digest Summary)," *Journal of Economic Perspectives* (15:2), pp. 121-144.

Jensen, M., and Murphy, K. 1990. "Performance Pay and Top-Management Incentives," *Journal of Political Economy* (98:2), p. 225.

Jensen, M.C., and Meckling, W.H. 1976. "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure," *Journal of Financial Economics* (3:4), pp. 305-360.

Katz, L., and Summers, L. 1989. "Industry Rents: Evidence and Implications," *Brookings Papers on Economic Activity: Microeconomics*), pp. 209-275.

Kostiuk, P.F. 1990. "Firm Size and Executive Compensation," *Journal of Human Resources* (25:1), pp. 91-105.

Krueger, A., and Summers, L. 1987. "Reflections on the Inter-Industry Wage Structure," in *Unemployment and the Structure of Labor Markets,* K. Lang and J. Leonard (eds.). Oxford: Basil Blackwell.

Krueger, A., and Summers, L. 1988. "Efficiency Wages and the Inter-Industry Wage Structure," *Econometrica* (56:March), pp. 259-293.

Leavitt, H., and Whisler, T. 1958. "Management in the 1980's," *Harvard Business Review* (November-December), pp. 41-48.

Malone, T., Yates, J., and Benjamin, R. 1987. "Electronic Markets and Electronic Hierarchies," *Communications of the ACM* (30:6), p. 455.

McAfee, A. 2002. "The Impact of Enterprise Information Technology Adoption on Operational Performance: An Empirical Investigation," *Production and Operations Management* (11:1), pp. 33-53.

McAfee, A. 2005. "Pharmacy Service Improvement at Cvs (a)," Case Study 606-015, Harvard Business School.

Polanyi, M. 1958. *Personal Knowledge.* Routledge London.

Roberts, D.R. 1956. "A General Theory of Executive Compensation Based on Statistically Tested Propositions," *Quarterly Journal of Economics* (70:2), pp. 270-294.

Rosen, S. 1981. "The Economics of Superstars," *American Economic Review* (71:5), pp. 845-857.

Rosen, S. 1992. "Contracts and the Market for Executives," in *Contract Economics,* L. Werin and H. Wijkander (eds.). Cambridge, MA
Oxford: Blackwell, pp. 181-211.
Rosenberg, N. 1982. *Inside the Black Box: Technology and Economics.* Cambridge Univ Pr.
Ross, J. 1995. *Johnson & Johnson: Building an Infrastructure to Support Global Operations.* Center for Information Systems Research, Sloan School of Management, Massachusetts Institute of Technology.
Rule, J., and Attewell, P. 1989. "What Do Computers Do?," *Social Problems*), pp. 225-241.
Tervio, M. 2008. "The Difference That Ceos Make: An Assignment Model Approach," *American Economic Review* (98:3), pp. 642-668.
Von Hippel, E. 1994. "" Sticky Information" And the Locus of Problem Solving: Implications for Innovation," *Management Science* (40:4), pp. 429-439.
Wood, C., Shin, H., and Devaraj, S. 2008. "Technology's Effect on Firm Structure: Manufacturing Vs. Service," *ICIS*, Paris, France.
Yermack, D. 1997. "Good Timing: Ceo Stock Option Awards and Company News Announcements," *Journal of Finance*), pp. 449-476.
Zhu, K. 2004. "The Complementarity of Information Technology Infrastructure and E-Commerce Capability: A Resource-Based Assessment of Their Business Value," *Journal of Management Information Systems* (21:1), pp. 167-202.