THE COST OF RESEARCH TOOLS AND THE DIRECTION OF INNOVATION: EVIDENCE FROM COMPUTER SCIENCE AND ELECTRICAL ENGINEERING*

Jeff Furman
Boston University & NBER
(furman@bu.edu)

Florenta Teodoridis
University of Southern California
(florenta.teodoridis@marshall.usc.edu)

Abstract

We examine how a change in the cost of access to knowledge influences the direction of inventive activity. To do this, we leverage an unanticipated and substantial reduction in the cost of motion-sensing research technology that occurred with the introduction and subsequent hacking of the Microsoft Kinect system. To estimate whether this shock induces changes in scientists' research trajectories, we employ novel measures based on machine learning (topic modeling) techniques as well as traditional measures based on bibliometric indicators of knowledge accumulation. Our analysis demonstrates that the Kinect shock increases the diversity of research of both incumbents and entrants in motion-sensing and that the effect is greater for entrants than for incumbents. Importantly, the increase in diversification of entrants extends to projects that do not directly use motion-sensing, suggesting that the reduction in cost acts as a conduit towards unexplored research trajectories.

Keywords: direction of innovation, research costs, (JEL I23, O31, O33, O34)

* Authors in alphabetical order; both authors contributed equally. We thank Ajay Agrawal, Kevin Bryan, Avi Goldfarb, Danielle Li, Olav Sorenson and Scott Stern for substantive discussions of the analysis, and seminar participants at numerous conferences and seminars for constructive feedback. We are grateful for funding from the National Science Foundation's Program on the Science of Science & Innovation Policy (SciSIP), grant SES-1564368 & SES-1564381. We thank Aparna Anand, Avik Basu, Robyn Johnson and Lindsay Raymond for excellent research assistance. All errors are our own.

I. Introduction

In 2011, Peter Thiel, the high-profile founder of PayPal and Palantir turned technology investor, famously published an investment manifesto with the subtitle “We wanted flying cars, instead we got 140 characters.” This provocative lament on the direction of contemporary technological evolution highlights a first-order question regarding why innovation takes the direction that it does. Since the NBER’s seminal publication of the volume *The Rate and Direction of Inventive Activity* in 1962 (NBER, 1962; Lerner & Stern, 2012), researchers have made more progress in studying drivers of the *rate* than of the *direction* of inventive activity (Acemoglu, 2012). In this paper, we contribute to research addressing this deficit by investigating the impact of research technology, an increasingly salient yet understudied input in knowledge production (Stephan, 2012), on the direction of knowledge creation in ideas space.

In particular, we ask two questions related to the diversity of research lines: First, to what extent does a substantial reduction in the cost of using a research tool affect the diversity of scientists’ research trajectories? Building on models developed by Aghion, Dewatripont, and Stein (2008), and Murray et al. (2016), we argue that the impact on research diversity of changes in the cost of research tools will depend on the location of researchers in knowledge space, i.e., either inside or outside the area affected by the change in research costs. Thus our second research question: Do reductions in the cost of research tools differentially affect researchers as a function of their location in knowledge space?

Identifying a relationship between changes in the cost of research technology and the direction of inventive effort is difficult because numerous unobservable factors are likely associated with both changes in the cost of such tools and the evolution of research lines. To generate causal insights regarding the consequences of changes in research technology costs on idea recombination, we need an instrument that is correlated with such changes but not with factors affecting research trajectories other than through their effect on the cost of research equipment. Building on Teodoridis (2017), we propose that the introduction of the Microsoft Kinect gaming system provides just such a shock to the cost of motion-sensing research tools.

Kinect was launched in November 2010 as an add-on device for Microsoft’s Xbox 360 gaming console. Whereas prior gaming systems required the use of handheld controllers to interact with games, Kinect employed motion sensors to track gamers’ physical movements (e.g., punching or kicking motions) to enable them to interact with the Xbox using only their bodies to control gameplay. Soon after Kinect’s launch, hackers developed and released through the open-source community a driver that enabled devices other than the Xbox to interact with Kinect, thus making it possible for scholars in electrical engineering, computer science, and electronics to harness Kinect’s motion-sensing data for use in research applications. The launch and hacking of Kinect reduced the cost of motion-sensing technology by an order of magnitude, from multiple thousands of dollars to approximately $150 (Teodoridis, 2017).

In this paper, we consider the introduction of the Kinect system to be a plausibly exogenous shock, since its value to scientific research was neither a consideration nor a motivation for Microsoft's development of the tool, nor was it anticipated by the research community. Indeed, Microsoft was surprised by the research response to Kinect and initially worked actively to prohibit the introduction of an open-source Kinect driver. We utilize the Kinect shock in difference-in-differences estimations to ascertain the impact of its introduction on measures of researcher-level project choice in the domains of computer science and electrical and electronics engineering research.

To develop a dataset to enable our analysis, we begin with the Institute of Electrical and Electronics Engineers (IEEE) *Xplore* database, a curated database that includes published and conference papers in electrical engineering, computer science, and electronics. We compile individual researcher histories using author identifiers assigned by the IEEE and checked for accuracy by the authors. We distinguish between researchers who had been active in motion-sensing research prior to the shock and those who engage with motion-sensing after the shock, and attempt to get an understanding of the type of innovation these researchers undertake because of the change in cost of research technology.

Measuring such effects is challenging. Characterizing the direction of innovation with empirical variables requires capturing changes in an ever-evolving, multidimensional space of ideas. To estimate changes in research trajectories, one first needs to define boundaries of research trajectories. This creates a paradox, however, as boundaries of research trajectories are part of the core unknown to be estimated. To address this challenge, we propose novel measures of units of knowledge in ideas space based on topic modeling analysis, using unassisted machine learning techniques that represent the latent categorization of academic publications without relying on an ex-ante, pre-imposed, structure (such as author-defined keywords or institutionally defined research fields). Our analysis also considers more traditional measures of knowledge trajectories based on observable characteristics of academic publications: a measure of diversification based on a stable taxonomy maintained by the IEEE, a count of new authors, and a count of new publication outlets. Each measure involves advantages and disadvantages that we discuss in our analysis. Although each measure captures somewhat different attributes of diversification, we believe that, taken together, they provide an informative window into the effects of the cost reduction on the direction of innovation.

The results indicate an average increase in diversification of research output at the individual level across all our measures. The estimates indicate a disproportionate increase of 0.60 to 0.74 topics (out of approximately 50) in the publication portfolio of researchers who had been active in motion-sensing research before and after the Kinect shock. These researchers also experience a 34% disproportionate increase in the number of yearly new co-authors and a 38% increase in the number of yearly new publication outlets. Next, following theoretical insights informed by Aghion, Dewatripont, and Stein (2008) and Murray at el. (2016), we separately consider the effect of Kinect on within- and outside-area researchers. We define within-area researchers as incumbents in motion-sensing research (before the launch of Kinect). We find that outside-area researchers exhibit a higher increase in diversity. Specifically, outside-area researchers experience an increase of about one additional category (out of approximately 50) in the diversification of their research portfolios, a 47% to 49% increase in research portfolio diversification, while within-area researchers experience a 6% to 7% increase. Furthermore, outside-area researchers increase the number of new yearly co-authors by 48% (1.11 additional new co-authors) and the number of new yearly publication outlets by 49% (1 in 3 new publication outlets), while within-area researchers increase their number of new co-authors by 8% and that of new publication outlets by 10%. Important, the increase in diversification of outside-area researchers is not fully explained by diversification into motion-sensing. Specifically, while the reduction in cost of research technology increases the possibility for outside-area researchers to engage with projects in the knowledge area of the cost reduction, it also increases the probability of generating novel combinations that could open new lines of inquiry. Indeed, we find that the effect on diversification persists, with only small reductions in magnitude, when eliminating post-Kinect motion-sensing publications from the portfolio of outside-area researchers.

We interpret these results as evidence that a large reduction in the cost of research technology acts as a conduit to increased diversity in research trajectories, which Acemoglu (2002) suggests as a desirable outcome for economic growth. For example, an outside-area researcher in our sample who, prior to Kinect, was focused on research involving sound waves, extended his research to the study of infant seizures by developing detection techniques combining audio and video inputs. Another researcher who was involved in motion-sensing prior to Kinect took advantage of the opportunities opened by the cost reduction to adapt computer vision detection and visualization algorithms to developing malaria diagnostic and tumor identification techniques.

Our analysis underscores the central role played by research tools in knowledge production and sheds light on the processes by which their cost influences how research trajectories evolve. Understanding these processes is important for research in the economics and sociology of science and for the study of the strategic management of science-based organizations. In addition to direct welfare implications, factors influencing the evolution of research trajectories are important because of the complex interplay of financial incentives, career incentives, institutional features, and researcher preferences for the construction and maintenance of research communities and the development and evolution of invisible colleges (Crane, 1969; David and Dasgupta, 1994, Stephan, 1994; Stern, 2004). Furthermore, factors influencing the direction of research are particularly salient for the study of strategic behavior of science-based organizations, including research laboratories in both private firms and universities. Just as project selection and the identification of niches for specialization are important for private firms competing for profits, investments in resources that affect the direction of knowledge creation should be important for academic institutions and research laboratories competing for reputation and opportunities for knowledge creation. Furthermore, the magnitude of the impact of costs of research tools draws attention to the role of market power in technology retailing and its influence on the direction of innovation.

II. Direction of Innovation

Interest in understanding the factors that affect the rate and direction of technical change has been sustained for more than a half-century because of the central importance of knowledge for economic growth and the interest of public policy in supporting the accretion of ideas-driven growth (Solow, 1956; Nelson, 1962; Romer, 1990; Lerner and Stern, 2012). The difficulties associated with obtaining socially optimal levels of innovation are long understood to arise from difficulties with appropriability and indivisibility, which yield externalities and fixed costs that negatively affect initial inventors relative to subsequent inventors and result in underinvestment in innovation (Arrow, 1962; Scotchmer, 1991).

More recently, scholars have clarified that these distortions affect the rate of technical advancement in part by affecting its direction. Acemoglu (2012) demonstrates that market incentives, crafted around intellectual property protection policies, yield underinvestment in research diversity relative to the social optimum. In turn, this leads to excessive concentration of innovation in the types of technologies that offer rents in the near term at the expense of alternative approaches that may yield higher returns in the long term. Bryan and Lemus (2017) underscore this point, noting that suboptimal investment in diversity arises even in the absence of intellectual property policy interventions. This work paints a somewhat pessimistic picture of the relationship between research diversity and economic growth. On the more optimistic side, Acemoglu (2012) suggests that the underprovision of commercial incentives for diversity in invention may be mitigated, though not fully, through diversity in basic science research, which serves as an input into commercial innovation and for which the incentives for diversity are not dampened by commercial imperatives. Indeed, Aghion, Dewatripont, and Stein (2008) argue that "the private sector's ownership of a given idea will not yield as diverse an array of useful next-generation ideas as would be generated in academia" (p. 630), since the incentives of firms for commercial success impose constraints on their researchers' choices whereas the commitment of academia to freedom in project selection enables its researchers to choose across broad sets of scientifically feasible research lines.

Project selection in academia is not, however, entirely free of constraints. First, the incentive system in academia gives rise to numerous concerns about freedom of project selection, including reputation, personal satisfaction, financial remuneration, and a number of other, more eclectic considerations (Merton, 1957; Stephan, 1996, 2012). For example, scholars target their research for certain journals in response to monetary awards (Franzoni et al., 2011) and reputation benefits associated with such publications, in an attempt to secure grants as urged by universities (Stephan, 2012) or to meet tenure requirements. Second, competition in ideas space and the expansion of the knowledge frontier leads researchers to specialize in increasingly narrower niches (Jones, 2009; Agrawal, Goldfarb, and Teodoridis, 2016) and to move away from crowded research areas (Borjas and Doran, 2013). Third, external factors such as science and innovation policies, firm and government investments, market demand-pull factors, and features of local research environments, including geographic endowments and firm policies, can also affect researchers' choice of projects (Dasgupta and David, 1994; Cockburn and Henderson, 1994, 1996). For example, massive research commitments, such as the Manhattan Project, the War on Cancer, and the Large Hadron Collider at CERN, accelerate the process of knowledge generation, although their impact on research direction is less clear. Such projects may increase research diversity by bringing together scholars from different areas and enabling broad exploration to solve specific problems, but a focus on achieving particular research and solving specific questions may limit inquiry to narrower research trajectories.

An important but not fully understood factor that influences freedom of project selection in academia is the cost of research materials and tools (Stephan, 2012). Mokyr (2002) details the salient influence of technology in general and research technology in particular in affecting the cost of access to knowledge. Mokyr (2002) argues that the cost of tools is important for research progress and project selection, since it affects the ability to access and leverage knowledge, the existence of which does not guarantee that such knowledge can be accessed and put to productive use. At the extreme, some research questions cannot be addressed at all in the absence of sufficiently well-developed research tools, since the cost of their use is effectively infinite prior to their development. For example, inquiry in astronomy and multiple areas of physics was deeply constrained before the invention of the telescope, in the early 1600s (Van Helden, 1975); and the invention of the microscope and its improvements by Hooke, in the 1660s, and van Leeuwenhoek, in the 1670s, famously initiated the field of microbiology (Gest, 2004). More recently, Furman and Stern (2011) demonstrate that changes in the institutional regime that facilitated access to biomedical materials at lower cost led to an increase in researchers' propensity to use knowledge associated with those materials. Similarly, Murray et al. (2016) document that the NIH-led elimination of IP restrictions on sharing genetically engineered mice led to an increase in the number of academic publications that draw on knowledge associated with those mice. In addition to demonstrating an increase in the rate of knowledge production once the costs of access to research technology decline, each paper provides suggestive evidence relating increased access to research tools and diversity in research.

In this paper, we directly focus on understating implications of reductions in cost of research technology on individual researchers' diversity in portfolio trajectories. We focus on the individual researcher because individuals are at the root of the knowledge creation process. Thus, understanding the impact of cost conditions on researchers' choice of portfolio trajectories provides a solid step towards understanding the impact of cost conditions on aggregate diversity in research direction. Our analysis draws from a series of theoretical insights to hypothesize on implications of cost reductions on researchers' portfolio diversification.

Murray et al. (2016) build on Aghion, Dewatripont, and Stein (2008), who emphasize control-rights considerations, and argue that the fundamental trade-off between academia and the private sector is one of creative control versus focus, to consider the role of cost of access to knowledge within academia. Their model is centered around the role of a manager (of research lines). The more early-stage the research on each line, the lower the probability that the manager will have vision to see where the line is going and, hence, direct researchers to pursue the line further. The lower that probability, the higher the freedom of researchers to move from one research line to another. While the authors provide empirical support by focusing on the role of openness in academia and implications for aggregate diversity as observed through academic publications, their theoretical model reveals two additional important insights.

First, the impact on diversity is determined by individual-level mechanisms influencing the project selection freedom of individual researchers. This insight underscores the importance of focusing our analysis at the individual researcher level. Second, the impact on diversity depends on the interplay between the magnitude of cost reduction and the location of researchers in knowledge space either inside or outside the area of the cost reduction. Specifically, large reductions in the cost of research technology were found to be democratizing (Teodoridis, 2017), i.e., to compensate for the fixed costs of within-area expertise (Cohen and Klepper, 1992, 1996). Conversely, lower reductions in the cost of research technology increase returns to specialization (Teodoridis, 2017), i.e., result in increased freedom of project choice only for within-area researchers.

Drawing on these insights, we hypothesize that a large reduction in the cost of research technology will lead to an increase in the diversity of research at the individual researcher level, as a result of outside-area researchers exploiting the opportunities opened by the reduction in cost. For example, Teodoridis (2017) shows that a large reduction in the cost of research technology alters the optimal team composition in knowledge production to include outside-area specialists. To the extent that changes in collaboration reflect changes in researchers' project choices, this is consistent with an increase in diversity of outside-area researchers. Similarly, Murray et al. (2016) show that openness, the elimination of restrictions on sharing genetically engineered mice, led to an increase in the arrival of new researchers on academic publications citing the use of mice. This suggests that the portfolio of outside-area researchers changed to include academic publications citing the use of mice. Furthermore, we argue that the resultant increase in the diversity of outside-area researchers would move beyond direct engagement with the knowledge area of the research technology, since access to a new knowledge area increases the possibility of novel recombinations (Uzzi et al., 2013) and, thus, new lines of inquiry. This is important, as it extends the benefits of cost reduction beyond the diversity effects discussed in Murray et al. (2016) to considering the role of costs of research tools as a conduit towards enabling new lines of inquiry. These effects would not occur with lower reductions in the cost of research technology.

The effect on diversity of within-area researchers is less clear. On the one hand, a large reduction in the cost of research technology does not create the same opportunities for recombination for these researchers since the cost reduction arises within their domain of knowledge. Therefore, we would not expect large cost reductions to induce diversification among within-area researchers. On the other hand, Aghion, Dewatripont, and Stein (2008) note a direct effect of openness in academia: that of greater mobility of researchers across projects. In other words, a large reduction in cost increases openness even among within-area researchers, facilitating greater mobility between projects. This effect could lead to an increase in diversity among these researchers to the extent that greater mobility leads to some degree of engagement in different projects. For example, Teodoridis (2017) discusses a replacement effect, where large reductions in costs of research technology substitute for within-area specialists in teams engaged with the knowledge captured in the research tool, thus freeing these within-area researchers to pursue other projects. In the presence of lower reductions in cost, we anticipate similar effects insofar as increasing returns to within-area expertise manifest through involvement in more diverse projects.

All in all, these considerations suggest that low reductions in costs of research technology would have an impact on diversity to the extent that the increased returns to within-area expertise result in engagement with more diverse projects. Conversely, large reductions in costs of research technology lead to larger effects on diversity because, in addition to the possibility of within-area researchers becoming engaged with more diverse projects, such reductions in cost enable outside-area researchers to engage with the knowledge area of the tool as a conduit towards enabling new lines of inquiry.

III. Kinect

The shock we examine in this paper involves a large reduction in costs of research tools. The setting appears unusual at first glance: It is the result of the unexpected impact of Microsoft's successful launch of a controller-free video game system designed to compete with rival products launched by Nintendo and Sony. In the two months following Kinect's launch on November 4, 2010, Microsoft sold more than 8 million units (>130,000 Kinect units per day), outpacing the iPhone and the iPad to become the Guinness World Records' all-time fastest-selling consumer electronic device (Bilton, 2011). The surprise that makes Kinect valuable for our research context is not, however, its commercial success but its wide-ranging and near-immediate impact on scholarship in motion-sensing, computer vision, and related areas of inquiry. In the sections that follow, we trace the history of Kinect and outline its applicability to various research communities.

II.1. Microsoft's Introduction of the Kinect System

In November 2010, Microsoft introduced the Kinect system for its Xbox video game console with the aim of competing with handheld gesture-recognition remotes introduced previously by Nintendo (Wii) and Sony (PlayStation). With the Kinect, Microsoft attempted to leapfrog its video console rivals by creating the first hands-free controller for electronic devices, a game controller system that responded to the natural movements of the player. The Kinect system was composed of a color (RGB) camera, a depth sensor, and a multi-array microphone. These physical features, along with pattern recognition software, enabled Kinect to recognize in three dimensions the movements and facial expressions of multiple individuals and to acknowledge and distinguish their voice commands (Zhang, 2012). The Kinect system translated this motion-sensing information into actions enabling players to control game play.

In addition to being a treat for video gamers, Kinect also offered a feast for hackers, who descended upon the system with the aim of developing code that would enable users to access the vast data obtained by Kinect's sensors and link it to other devices, especially Windows-based computers. These efforts received a twofold infusion of interest on Kinect's launch day. First, Adafruit Industries, a manufacturer of do-it-yourself electronics kits operated by recent alumni of MIT's Media Lab, Limor Fried and Phillip Torrone, offered a $1,000 prize for the first individual or organization to post an open-source Kinect driver to GitHub (Carmody, 2010a). The second spur of interest arose as a result of Microsoft's actively (and quixotic) effort to thwart the hackers. In a same-day response to Adafruit's prize offer, Microsoft released a statement to CNet: "Microsoft does not condone the modification of its products… [and will] …work closely with law enforcement and product-safety groups to keep Kinect tamper-resistant" (Terdiman, 2010). Adafruit responded immediately by doubling its Kinect driver bounty to $2,000, further intensifying the race for the driver.

Only two days after launch, a member of the NUI (Natural User Interface) Group, named *AlexP*, demonstrated his ability to control Kinect using Windows 7. His refusal to share his code and Microsoft's denial that Kinect had been hacked led Adafruit to increase its prize yet again, this time to $3,000. By November 10, Adafruit had acquired a piece of equipment called a USB analyzer that enabled it to collect and share data from Kinect's sensors. Using these data, multiple teams raced to decode and reverse-engineer the Kinect driver. The race was won the next morning, November 11, by a Spanish computer science undergraduate student, Héctor Martín, who did not own an Xbox but who had purchased a Kinect that morning when it went on sale in Europe (Giles, 2010). Within days of the driver's release, researchers and hobbyists had adapted Kinect for numerous uses, including the creation of 3-D computer holograms and a modified iRobot Roomba that could responded to human hand and voice commands and could create visual maps of the rooms it had visited (Wortham, 2010).

During the week that hackers had raced to create an open-source driver to harvest Kinect's data, consumers purchased nearly a million Kinect units. In the wake of Martín's success, Microsoft initially continued to resist working with the hacker community and even refused to acknowledge that its system had been hacked; the firm's attitude towards their efforts soon softened and its public statements became less negative.[1] Within ten days of the release of Martín's open-source driver, the potentially wide-ranging technological and scientific applications of Kinect came into greater focus and Microsoft's approach changed completely. Microsoft had become convinced of the value of embracing the experimentation of the hobbyist and scientific communities and announced that it would not pursue legal remedies against those who adapted the Kinect system for such purposes and that it would, indeed, work with those communities to enable Kinect's broader applications (Carmody, 2010b). By June 2011, Microsoft offered the unofficial adaptations of Kinect an official embrace through its release of a software developers' kit (SDK) that support noncommercial applications.[2]

II.2. Motion-sensing Research & the Introduction of Microsoft Kinect

In addition to the particularly strong interest in Kinect among the hobbyist and hacker communities, the technology has had a substantial influence on academic research, by reducing by an order of magnitude the cost of employing motion-sensing as a tool in the process of scientific research. Prior to Microsoft's development of Kinect, low-quality depth-sensing motion-sensing technologies were available for academic research at a cost of thousands of dollars, through tools such as "time of flight" cameras. The introduction of Kinect improved motion-sensing quality while reducing the cost to $150. The tool's perceived impact among academics is evident in a series of researcher quotes reported in Teodoridis (2017). One particularly effusive researcher states:

---

1 "What has happened is someone has created drivers that allow other devices to interface with the Kinect for Xbox 360... The creation of these drivers, and the use of Kinect for Xbox 360 with other devices, is unsupported... We strongly encourage customers to use Kinect for Xbox 360 with their Xbox 360 to get the best experience possible" (BBC News, November 12, 2010).

2 An interesting coda to the Adafruit-Kinect saga is that it was partially financed and inspired by a Microsoft insider, Johnny Lee, who was frustrated with the firm's intransigent approach towards enabling the system to be used for purposes other than Xbox gaming. In February 2011, after Microsoft announced its support for Windows-based drivers of the Kinect system, Lee, who had left Microsoft to join Google in the interim, expressed his enthusiasm for the release and his frustration with Microsoft's reluctance to do so at the time of the system's release: "Yay! This makes me happy. Microsoft officially announces support for Windows Drivers for the Kinect Camera as a free download in the Spring. This was something I was pushing really hard on in the last few months before my departure …It's unfortunate this couldn't have happened closer to launch day. But, perhaps it took all the enthusiasm of the independent developer community to convince the division to do this. It certainly would have been nice if all this neat work was done on Microsoft software platforms. I actually have a secret to share on this topic. When my internal efforts for a driver stalled, I decided to approach AdaFruit to put on the Open Kinect contest. For obvious reasons, I couldn't run the contest myself. Besides, Phil and Limor did a phenomenal job, much better than I could have done. Without a doubt, the contest had a significant impact in raising awareness about the potential for Kinect beyond Xbox gaming both inside and outside the company. Best $3000 I ever spent" (Lee, 2011). Adding credence to Lee's account, Adafruit quickly corroborated the report on its own blog (Adafruit, 2011).

> It's incredibly powerful to have a cheap way to obtain even a noisy estimate of a user's pose. $200 is zero dollars for a research lab, while $2,000 and $20,000 [the costs of the alternative technologies] are $2,000 and $20,000 – a lot of money! Rehabilitation researchers and robotics researchers love the Kinect because of all the new types of applications it enables. It just works! That is great. (p. 46)

Before the launch of Kinect, most work using motion-sensing tools and algorithms was conducted by a subset of researchers in the broad field of computer vision (motion-sensing researchers for short). Computer vision is an interdisciplinary field in computer science and electrical engineering that studies how computers can be enabled to perceive and interact with the physical world based on information gleaned from captured and processed images.[3] Topics in computer vision include 3D model building; automotive assistance, safety, and traffic management; biometrics and fingerprint recognition; eye and head tracking, gesture recognition; image recognition; machine inspection; medical imaging; motion capture; object description and recognition; optical character recognition (OCR); photography; robotics; safety monitoring; sports analysis; and surveillance.[4] The subgroup of motion-sensing researchers were focused on projects that involved the detection, interpretation, and application of visual and other sensory information associated with moving objects. Similar to fields like statistics and cryptography, motion-sensing research includes scholars who develop the tools, techniques, and algorithms that are central to the field's development, as well as researchers in a number of application areas who employ the field's tools and contribute to their improvement.

After the launch of Kinect, the use of motion-sensing tools expanded within the computer vision community and expanded to other areas of computer science and electrical engineering. First, as the story of its launch and hacking suggests, the arrival of the Kinect system was perceived as a profound shock in costs and capabilities for the computer vision research community: Han et al. (2013) describe the impact simply and clearly: "With the invention of the low-cost Microsoft Kinect sensor, high-resolution depth and visual (RGB) sensing has become available for widespread use." *Wired* magazine described the potential for Kinect to impact robotics more evocatively:

---

[3] Interest in using computers to represent the visual world began in the era of computer punch cards in the 1950s (e.g., Kirsch et al., 1957). A series of publications and conference volumes of research on pattern recognition, spatial representation, and signal detection began in the 1960s (e.g., Lugt, 1964, Sutherland, 1968; Haralick et al. 1973) and yielded textbooks in the 1970s and early 1980s (e.g., Winston, 1975; Hanson and Riseman, 1978; Brady and Barrow, 1981).

[4] Computer vision groups are most often located in computer science departments (e.g., MIT's Computer Vision group within CSAIL, the Computer Science and Artificial Intelligence Laboratory) and/or engineering schools (e.g., the Computer Vision Group at UC-Berkeley, and the CAVE Computer Vision Laboratory at Columbia). Master's programs in computer vision are increasing in number (e.g., Carnegie Mellon's, MSCV, "Master of Science – Computer Vision"), as are undergraduate programs (e.g., Boston University's Undergraduate Concentration in Computer Vision). The University of Central Florida recently celebrated its 30th year of NSF-funded training for undergraduate research assistants in computer vision (UCF, 2017). Computer vision research groups also exist at major technology firms, including Microsoft.

> "For 25 years, the field of robotics has been bedeviled by a fundamental problem: If a robot is to move through the world, it needs to be able to create a map of its environment and understand its place within it. Roboticists have developed tools to accomplish this task, known as simultaneous localization and mapping, or SLAM. But the sensors required to build that map have traditionally been either expensive and bulky or cheap and inaccurate. Laser arrays cost a few thousand dollars and weigh several pounds, and the images they capture are only two-dimensional. Stereo cameras are less expensive, lighter, and can construct 3-D maps, but they require a massive amount of computing power. Until a reasonably priced, easier method could be designed, autonomous robots were trapped in the lab. … When something is that cheap, it opens up all sorts of possibilities," says Ken Conley of Willow Garage, which sells a $500 open source robotics kit that incorporates the Kinect. (The previous non-Kinect version cost $28 0,000.) (Tanz, 2011)

Similarly, soon after Kinect became available, Chang et al. (2012) compared the potential of Kinect with that of OptiTrack, a high-precision motion-capture system historically used in rehabilitation tracking and costing substantially more than Kinect (approximately $3,000-$6,000 in 2017). Other work using Kinect includes studies on improving the physical performance of youths with motor disabilities (e.g., Chang et al. 2011), measuring human gait (Stone and Skubic, 2013; Preis et al., 2015), and improving facial recognition (Nanni et al., 2014).

In addition to its impact on areas that were, historically, the focus of motion-sensing research, Kinect was adopted by researchers outside the traditional motion-sensing research domains. For example, Richards-Riessetto et al. (2012) describe the value of Kinect for work in archaeology, and Rafibakhsh et al. (2012) describe its value for construction engineering and management. More broadly, after the launch of Kinect, motion-sensing appears to have found its way into an increasing variety of research projects with applications in a wide set of domains, from artificial intelligence and virtual reality to education, healthcare, music, cinematography, market research and advertising. For example, faculty and graduate students at MIT's CSAIL laboratory have designed a motion-sensing system, called Emerald, which tracks individuals' movements within their homes, can alert medical personnel in the event of a medical catastrophe or fall, and can even be used to predict fall events.

Overall, the launch of Microsoft Kinect appears to have changed the opportunity set for research in computer science and electrical engineering through the introduction of a low-cost, relatively high-quality motion-sensing tool. The cost reduction appears to be plausibly exogenous and appears to have been a surprise to the incumbent research community, the community of potential users that had been working outside traditional motion-sensing topics, and even to Microsoft itself. We thus employ the shock in our analysis as a plausibly exogenous reduction in the cost of research that is not correlated with researchers' characteristics or with their anticipated research choices, except indirectly through its impact of the cost reduction on researchers' project choices.

IV. Data and Empirical Strategy

Our analysis is focused on estimating the impact of a large reduction in the cost of research technology on researchers' propensity to pursue new research directions, or, in other words, to produce output with a higher degree of diversity. We face two main challenges. First, identifying a relationship between changes in the cost of research technology and the direction of inventive effort is complicated by the prospect that multiple unobservable factors may be associated with changes in both the cost and the evolution of research lines. Second, it is difficult to measure diversity in research lines since researchers may manipulate flexible measures of research direction. Further, fixed measures of research direction, such as bibliographic taxonomies, may be too rigid to adapt as new research areas emerge over time. To address the first challenge, we exploit the quasi-natural experiment provided by the launch of Kinect using a difference-in-differences analysis. To address the second challenge, we employ novel measures of research direction based on advances in machine learning, i.e., topic modeling, meant to uncover latent categorizations in large bodies of text.

Our data draw on the population of publications, early-access publications, and conference proceeding papers included in the IEEE *Xplore* database, which covers nearly 200 computer science and electrical engineering journals and more than 1,800 conference proceedings, between 2001 and 2014. We conduct our estimation on the subset of papers published in the four years before and four years after the launch of Kinect (2007-2014), and we use the remainder of the data to obtain better estimates of researchers' pre-Kinect research behavior and trends.

III.1. Empirical Strategy

Our differences-in-differences analysis compares research trajectories before and after the launch of Kinect. Formally, we estimate for researcher *i* and year *t*:

$$Diversification_{it} = \beta(TreatedResearcher_i * AfterKinect_t) + Age_i^2 + \delta_i + \gamma_t + \epsilon_{it} \quad (1)$$

where $TreatedResearcher_i$ is a dummy variable equal to 1 if research *i* is a treated unit, and 0 otherwise. We define treated researchers as individuals who published at least one motion-sensing paper in either the period before Kinect's launch, the period after the launch, or both. We explore a number of different definitions of treated researchers. Specifically, we distinguish between (1) an aggregate average effect of the reduction in the cost of research technology on all researchers' project diversification, (2) an average effect on individuals close in knowledge space to the area of the cost reduction (within-area researchers), and (3) an average effect on individuals further away in knowledge space from the area of the cost reduction (outside-area researchers). To identify motion-sensing publications, we search the full text and metadata of publications in the IEEE *Xplore* database using carefully identified keywords, through interviews with subject matter experts and cross-referenced with IEEE's taxonomy.[5]

$AfterKinect_t$ is a dummy variable equal to 1 if the observation year is between 2011 and 2014, namely after Kinect's launch, and 0 otherwise. Age_i^2 represents the squared age of researcher *i*. Age is calculated as the number of years since the occurrence of the first publication in our large dataset, starting in 2001. The age term itself is captured in δ_i , which represents individual fixed effects, which control for time-invariant individual attributes. The term γ_t captures year-specific fixed effects that account for changes in publication trends over time. As a consequence of including individual and time fixed effects, the terms $TreatedResearcher_i$ and $AfterKinect_t$ drop out of the estimating equation. $Diversification_{it}$ represents different measures of diversification of research lines at the individual level *i* at time *t*, as detailed in the next section.

Our main coefficient of interest is β. We interpret a positive value of β as indicating a higher increase in diversification at the individual level for treated researchers after the launch of Kinect, when compared with the change in diversification of matched researchers from before to after the launch of Kinect. In other words, a positive value of β indicates a positive effect of the reduction in cost of research technology on researchers' propensity to diversify their research lines.

We construct a plausible counterfactual using coarsened exact matching (CEM; Iacus et al., 2011, 2012) based on individual researcher characteristics in the before period (2007-2010). Specifically, we match on (1) yearly productivity, (2) the number of co-authors in each of the four years before Kinect's launch, and (3) a measure of diversification across knowledge topics between 2007 and 2010. We measure productivity as a count of publications weighted by citations and diversification as 1 minus the Euclidian distance in the space of IEEE-defined research categories.[6] We conduct three matching procedures, one for each of our three definitions of treated researchers: (1) individuals who published in motion-sensing either before or after the launch of Kinect, (2) individuals who published in motion-sensing before the launch of Kinect (incumbents or within-area researchers), and (3) individuals who published in motion-sensing only after the launch of Kinect (entrants or outside-area researchers). We follow this approach to distinguish between an aggregate average effect on all researchers who undertake motion-sensing and the differential

[5] These terms are available upon request and are described in greater detail in Teodoridis (2017).

[6] IEEE assigns each publication to one of the 51 main categories listed in its taxonomy. We calculate the Euclidian distance based on the percentage of keywords from each category that a researcher collects in her publication portfolio between 2007 and 2010. Formally, we calculate:

$$DiversificationIndex_i = 1 - \sqrt{\sum_{k=1}^{51} CategoryPercentage_{ik}^2} \quad (2)$$

where *i* is the individual researcher and $CategoryPercentage_{ik}^2$ represents the squared percentage of keywords assigned to researcher *i*'s publications in each of the *k* main 51 categories of the IEEE taxonomy. Note that, by construction, the measure is less than or equal to 1 and never 0, and it increases with higher levels of keywords spread across IEEE categories.

contribution to this effect of outside-area and within-area researchers, while ensuring appropriate counterfactuals for each subgroup – we employ CEM with weights rather than one-on-one matching to use as much of the available data as possible.

We argue that our matching procedure provides a plausible counterfactual for the treated researchers, i.e., those who engage with motion-sensing. We match on productivity in the before period to ensure that our results on changes in diversity at the individual level are not confounded by researchers at the right tail of the productivity distribution. This is important since the probability of having publications across multiple categories increases with publication output. We match on the number of co-authors in the before period to ensure that our results are not driven by researchers' abilities or preferences for collaborating more intensely or more broadly. This matters for the analysis, since higher levels of collaboration could be correlated with more diverse output as each new collaborator increases the potential pool of expertise and perspectives. We also match on the level of diversification in the pre-Kinect period to ensure that our results are not driven by individuals with a taste for diversity that may manifest regardless of changes in costs of research technology. Our results remain robust to using the full set of untreated researchers as counterfactual in fixed-effects estimations, under the assumption that all researchers publishing in IEEE outlets are at risk of engaging with new technological developments in their research.[7]

III.2. Measuring Research Diversification

To estimate changes in research trajectories, one first must define the boundaries of research trajectories. This creates a paradox, as the boundaries of research trajectories are part of the core unknown to be estimated. Unlike physical space, which consists of a well-known number of dimensions and distances between locations, ideas space consists of an unknown number of dimensions, the distances between which cannot be uniquely measured and which evolve in unanticipated ways over time (Doran, 2017).[8]

Aghion, Dewatripont, and Stein (2008) model the development of ideas along research lines. In some cases, such as in mathematics, fields are relatively well defined, and stable field distinctions can form the basis for inquiry about location and movement in ideas space (Borjas and Doran, 2012, 2015; Agrawal, Goldfarb, and Teodoridis, 2016). In most fields, however, it is difficult to measure such lines or to identify where they branch. To overcome these challenges, work on the direction of research activity often focuses on measures of research breadth (Grupp, 1990; Rafols and Meyer, 2010), indicator variables, or other measures of topic overlap to reflect whether researchers are roughly in the same field/domain (Boudreau et

[7] Available from the authors. Not included due to space constrains.

[8] Doran (2017) describes ideas space as consisting of an "arbitrary n-dimensional space" and notes that there is neither a "unique way to decide how 'close' two n-vectors are" nor how "to decide how quickly an n-vector is moving to a different position."

al., 2017), or considers the development of ideas based on references (citations) to one or more particular papers in a stream of research (e.g., Furman and Stern, 2011; Azoulay et al., 2015).

These analyses typically rely on observable indicators of innovation output, such as keywords and taxonomies, and employ measures such as citation maps, publication outlets, and collaboration metrics. While helpful in providing some evidence of the evolution of research trajectories, these approaches face many of the challenges described above. For example, research that categorizes research direction based on curated taxonomies has the advantage of consistency but faces the trade-off of either being stable and thus comparable over time or being dynamic and thus evolving with the changing research landscape at the cost of consistency and classification standardization. Author-assigned keywords, or any other set of keywords not extracted from a defined vocabulary, fare better in capturing new knowledge directions but lack structure and may be more subject to gaming. This limits the interpretability of hierarchical connections between keywords and changes in such relationships over time. Similarly, while the citation revolution, as a method for tracing knowledge linkages (DeSolla Price, 1970; Griliches, 1990), significantly helped advance our understanding of factors influencing the process of knowledge creation, it is subject to the same concerns, since measuring diversity in citation maps requires some form of categorization. Like author-assigned keywords, the selection of backwards and forwards citations is subject to social processes and strategic behaviors that complicate their interpretation for understanding changes in research trajectories.

The research context we examine, motion-sensing research and related work in computer science and electrical engineering, is not characterized by a relatively stable set of keywords and research topics. Indeed, the past two decades have seen the emergence of many new domains of research and associated new keywords enabled by ever-advancing computing power, materials science, and methods within these fields. To measure research direction in these fields, we leverage advances in machine learning and computing power to develop measures that make use of more complete information in academic publications. We propose measures based on two topic modeling algorithms, which we have adapted for inference in our context. Our empirical analysis also includes a set of more traditional measures of research diversification based on observable characteristics of academic publications, including (a) a measure of diversification based on the stable taxonomy maintained by the IEEE, (b) a count of new authors, and (c) a count of new publication outlets.

The main advantage and, hence, contribution of measures based on machine learning analysis is the ability to identify similarities between bodies of text without predefined assumptions about their structure. Machine learning algorithms evolved from the study of pattern recognition in computer science but have increasingly found applications in a variety of fields, including genetics, medical imaging, computational biology and bioinformatics, image recognition, social network analysis, and economics and public policy (Athey and Imbens, 2015).[9] Currently, there are a large number of algorithms customized for various tasks. While limitations remain, their complexity and accuracy is rapidly evolving. Note that the intended purpose of these algorithms is prediction, not inference. Their success rests with their ability to reveal the latent structure of a corpus of texts in order to predict with high accuracy where a new text would fit in the structure. We are interested in identifying the latent categorization of ideas in motion-sensing and related fields in ways that (a) are less subject to the strategic behavior of researchers and (b) are sensitive to the fact that motion-sensing research and related fields evolve over time.

In the service of these objectives, we employ two types of algorithms, which we adapt for our purposes: Latent Dirichlet Allocation (LDA) and Hierarchical Dirichlet Process (HDP) (Blei et al., 2003). The algorithms fall into the topic modeling category of unassisted machine learning. LDA and HDP are probabilistic models that employ a hierarchical Bayesian analysis of text (see, e.g., Hofmann, 1999; Blei et al., 2003; Buntine and Jakulin, 2004). The intuition is that of a generative process, in which the data are assumed to be characterized by a set of observed variables (words in the document or vocabulary) that develop from a set of hidden variables (the topic structure) (Blei et al., 2003). Both algorithms generate collections of words (topics) that are found to appear together in the input text with a certain probability. In other words, the input text is 'assigned' to topics with a certain probability.

We conduct our analysis using the abstract of the academic publications in our dataset as input text into LDA and HDP. We run each algorithm per year for the full set of publications available in our dataset. We modify the algorithms to output the set of words describing each topic and to list the publication IDs of each abstract used to identify those topics. Each publication ID is assigned a score, which can be thought of as a probability of 'belonging' to a topic. We ignore scores below 1% probability of belonging to a topic. Despite substantial advances in computing power, each process is computationally intensive. Each run of LDA or HDP using our data requires days of computing time. As result, we have run the current analysis allocating papers to topics using article abstracts but not full text. We hope to be able to use publication full text in future work.

Each algorithm involves trade-offs and limitations. First, LDA does not automatically calculate the number of topics that represents the optimal latent categorization of the input text. Researchers using LDA must therefore specify the number of topics into which they would like the algorithm to group the text. Choosing a number of categories that is too large could yield category definitions that are overly narrow, while choosing a number that is too small could yield category definitions that are too broad. HDP relaxes this constraint, as it has a built-in algorithm to identify the optimal number of topics per group of input text.

[9] Indeed, we note that a curious feature of our paper is that we are diversifying our own work by applying a research tool, machine learning, now available at lower cost in a field (the economics of innovation and science) in which it has not been regularly applied in the past.

Second, both algorithms treat the input text as a one-time group for which the latent categorization needs to be revealed. In other words, the algorithms cannot automatically track the evolution of topics over time by updating the set of keywords in each category over time. Third, the algorithms operate under a bag-of-words assumption, in which each word is treated independently. The optimal algorithm for our analysis would be one in which the number of topics and the set of words that characterize each topic were allowed to change over time to reflect the changing nomenclature and structure of scientific fields. To our knowledge, however, the field of machine learning has not yet developed any such reliable algorithms.

We take a few steps to mitigate some of the limitations of the existing algorithms. First, we present results from both LDA and HDP. We also run a sensitivity algorithm on multiple instances of LDA, consistent with state-of-the-art practices in computer science, to identify the number of optimal topics with the highest probability of accuracy.[10] We identify the optimal number of topics to be between 50 and 60, in line with the optimal number of topics calculated by the HDP.[11] In our results, we show estimates using results of an LDA with 50 topics and HDP. All our results remain robust to LDA with 40, 60, 90, and 120 topics. Note that the higher number of topics would, by construction, result in fragmentation of the optimal number of topics. We also run an alternative version of HDP that relaxes some of the bag-of-words assumption to confirm no significant changes in the optimal number of topics. Furthermore, we use the LDA and HDP output in a regression with time fixed effects; hence our results are not hindered by the fact that the algorithms are executed on a per-year basis and thus reveal the latent categorization of topics for each year in our dataset.[12]

We calculate diversification as a yearly weighted measure of spread across categories, as identified by the LDA and HDP algorithms. Specifically, we first sum the number of topics where the focal researcher has their papers assigned by LDA and HDP in the focal year. Next, we sum the probabilities of assignment to said categories and multiply that number by the number of topics. The probabilities can be thought of as the percentages of assignment to categories. Hence, the product of number of topics and sum of probabilities gives a weighted measure of the number of topics in which the focal researcher publishes in the focal year. For ease of interpretability of regression coefficients, we take the natural logarithm of this number.

[10] We would be happy to provide details on these processes upon request.

[11] We consulted with experts in computer science, electrical engineering, and electronics to ensure that the LDA and HDP topics reflect credible categorizations in this line of research.

[12] Computer scientists are working on a variety of extensions of these algorithms. We chose to use algorithms that are considered robust among computer scientists rather than current experimental ones aimed at advancing the frontier in topic modeling. We did so not only to ensure reliability of our results but also to ensure the most optimal available processing times (e.g., one instance of LDA takes 7 to 10 days to run on our dataset, when using an 8-core processor with 64 GB of RAM) and the realistic possibility of reverse-engineering the algorithms to obtain the set of probabilities for topics assignment at the publication level. This is important, as many algorithms with faster processing times or relaxed assumptions are doable precisely because their goal is to predict rather than explain the latent categorization. For example, versions of these algorithms encode words into numbers to speed up processing, but, in the process, they cut the link back to the actual input text, thereby making it impossible to extract the topic assignment probabilities for each publication ID.

Consistent with the recommendation of Doran (2017) that work measuring the direction of research trajectories should employ multiple measures, we also present results using more traditional measures of diversification, which are based on publication attributes. The first measure is a yearly reversed Euclidian distance in the space of 51 IEEE categories in computer science, electrical engineering, and electronics. To calculate this measure, we apply equation (2) to yearly publication data over the period of interest (2007-2014), four years before and four years after the launch of Kinect. The second measure counts the number of new co-authors that the focal author has in the observation year relative to previous years. To count the number of new co-authors, we take advantage of our full dataset going back to 2001. We do so since, by definition, the count of new co-authors requires a few years of reference data to get closer to reflecting the true number of new co-authors and to not be upward-biased due to left-side data truncation. The third measure reflects the number of new publication outlets in which a researcher publishes each year, relative to each prior year, going back to 2001.

We chose these three measures since they capture different attributes associated with changes in research lines. Specifically, the Euclidian-based measure follows previous literature in employing the data captured in the IEEE bibliometric categorization to create an index of concentration (or diversification) within ideas space, as defined by the taxonomy. The advantage of this approach rests with the tractability and stability over time of the categorization structure. The disadvantage stems from the same attributes, as we discussed above. In particular, the fixed taxonomy fails to capture changes in the categorization structure over time that would otherwise indicate changes in research lines. The other two measures include the count of new co-authors and the count of new publication outlets. The first of these indicates changes in research lines to the extent that collaboration patterns reflect changes in the bases of expertise associated with a researcher's project choices. The second measure indicates changes in research direction to the extent that different journals address different audiences and cover different areas of ideas space. The downside of these two measures is that they are focused on outcomes that indirectly reflect the content or intellectual focus of academic research. It is possible to change co-authors and publication outlets while continuing to work on the same knowledge trajectory, and it is also possible to continue working with old co-authors and publish in the same journals while shifting one's research focus. As a result, we prefer our more novel measures, which we think are more likely to reflect changes in research direction because they are more tied to the content of researchers' published works.

Our analysis presents results from five main measures of diversification: two measures based on topic modeling algorithms (HDP and LDA with 50 topics) and three measures based on traditional observable characteristics of academic publications, namely a Euclidian distance in the ideas space as mapped by IEEE's taxonomy, a count of new authors per period, and a count of new publication outlets per period. Each measure captures different attributes of diversification, each with its own limitations and merits. Taken together, we argue, they paint an informative picture of the evolution of research lines.

III.3. Sample Construction and Descriptives

We collect data on every publication, early-access publication, and conference proceeding paper available through IEEE *Xplore* between 2001 and 2014. These data include 2,492,451 publications and 1,670,888 unique author names in the fields of computer science, electrical engineering, and electronics. Because of the importance of publications in conference proceedings in computer science and related fields, the IEEE possesses advantages relative to other libraries of publications, including the Web of Science and Scopus.

We focus our analysis on the four years leading up to and the four years following the launch of Kinect at the end of 2010, i.e., 2007-2010 and 2011-2014. We do so to ensure comparable timeframes and to allow for some publication data for controls and other measures that require a longer-run observation of publication trends, such as author age and changes in number of yearly new co-authors and new publication outlets. The 2007-2014 dataset consists of 1,776,125 publications authored by 1,391,313 individuals as identified by IEEE. Within this subset, we further distinguish between researchers who published both before and after Kinect's launch (430,779), only in the before period (442,395), and only in the after period (518,139). We conduct our main analysis on the first subset, that of researchers who published both before and after Kinect's launch. We do so (1) to ensure that our main results are not driven by zeroes due to exits from or entry into our observation period and (2) to allow for an observable period before Kinect's launch from which we can identify trends and construct plausible counterfactuals. We further eliminate outliers from our main dataset, namely individuals with fewer than three and more than 50 publications before Kinect's launch (2007-2010). We eliminate researchers with fewer than 3 publications since some of our diversification measures rely on individuals' breadth of publications and low productivity mechanically translates into low diversification. Note that this set of researchers also includes authors who occasionally publish in outlets tracked by IEEE. We eliminate authors with more than 50 publications in the before period to account for potential disambiguation effects in the IEEE algorithm assigning unique author identifiers. Such author IDs would appear as very productive and potentially diversified individuals, risking an upward bias to our estimations. We identify a total of 3,200 researchers with over 50 publications in the four-year period before Kinect's launch, less than 1% of our main sample.

All our results remain robust to altering these cutoff decisions and to including the full set of data available to us. Specifically, our results remain robust (1) to considering the full dataset, 2001-2014; (2) to considering other cutoff points for the minimum and maximum number of publications; (3) to eliminating cutoffs for the minimum and maximum number of publications and utilizing the full set of 2007-2014 authors; and (4) to including authors who exit after 2010. Since our estimation strategy precludes including authors who published only in the post-Kinect period, i.e., we do not observe a before period in order to identify an appropriate counterfactual, we provide descriptive statistics for this group that demonstrate alignment with our main findings and argument.

In Table 1.a, we present average values of our three definitions of treated researchers for each of our five measures of diversification. In Table 1.b, we show the equivalent values for our matched samples. First, we draw attention to imbalance between our treated and control group that necessitates employing a CEM procedure. Second, the descriptive statistics foreshadow our main findings: (1) an overall increase in diversification, (2) a positive impact on diversification for within-area researchers, and (3) a higher and more pronounced impact on diversification on outside-area researchers compared with within-area researchers. Further, we observe differences across our diversification measures relative to baseline trends. Some of the effects come from mitigating the overall decrease in diversification, while others come from an accentuated baseline trend of increasing diversification. Specifically, our Euclidian measure of diversification, which captures the information contained in IEEE's taxonomy relative to changes in ideas space, shows an overall trend of decreasing diversification, with the reduction in cost of research technology dampening that effect. The same trends can be observed in our measure of diversification based on counts of new publication outlets. The measure based on counts of new co-authors shows similar trends in the full sample but reverses course in the matched sample, while the two topic-modeling-based measures show an overall trend of increase in diversification (for both the full and matched samples) that is accentuated by the reduction in cost of research technology. This is important for at least two reasons. First, it underscores our point regarding the difficulty of crafting an all-encompassing measure of diversification. Second, it highlights the differences in the various aspects of the knowledge creation process captured by each measure. For example, the overall trend of decrease in diversification as captured by the Euclidian measure appears to contrast with the overall trend of increase in diversification captured by our topic modeling measures (LDA and HDP). This apparent inconsistency may, however, result from an increase in the fragmentation of research lines, an effect aligned with the "knowledge burden" effect (Jones 2009, 2010). More broadly, the differences suggest a need for approaching studies on the direction of research trajectories using empirical strategies that triangulate across multiple measures.

******* Please insert Tables 1.a and 1.b about here *******

Since our empirical strategy is limited to evaluating the effect of the cost reduction on the group of researchers for which we have access to observables in the pre-Kinect period, we describe the group of individuals for which we do not have such data in Table 2. Specifically, Table 2 captures the number of new unique author identifiers, as listed in the IEEE database, that appear in the dataset in each of the eight years of interest. We distinguish between researchers entering our sample with at least one motion-sensing publication in their first year (Column 1) and researchers entering with other types of publications (Column

2). In Column 3 we calculate the ratio of new entries in motion-sensing research and observe that before the launch of Kinect about 1% of entry was occurring through publishing in motion-sensing. However, the percentage increases by 31% immediately following the Kinect launch and continues to increase, up to 2.5 times in 2014, relative to the period before Kinect. We interpret these data as additional evidence of increasing research diversification at the individual level. This finding is in line with Murray et al. (2016), who document that the elimination of IP restrictions on sharing genetically engineered mice led to an increase in the number of new authors on academic publications drawing on the knowledge embedded in the mice. The group of new authors presumably includes both new researchers that fall under the category captured in our Table 2 and those we capture in the outside-area group of treated researchers.

******* Please insert Table 2 about here *******

We present descriptive statistics that show the balance in our CEM procedure across our three definitions of treated researchers in Appendix A tables A1 to A3. We show average values for all our covariates used in the matching procedure, in both the full sample (Columns 1-3) and the matched sample (Columns 4-6). Across all definitions, treated researchers are more productive. They have a higher number of co-authors and publish in more IEEE categories than the rest of the population. However, this fact results from a more skewed distribution in the full sample. The CEM procedure balances these observables for each of our three definitions of treated researchers. Our sample size is reduced by the matching procedure and by our eliminating outliers, as described above.

V. Empirical Analysis: The Kinect Shock and Research Diversity

In our empirical analysis, we estimate the impact of the substantial reduction in the cost of motion-sensing research technology resulting from the introduction of Kinect on the direction of research activity in related areas. We show results using five measures of diversification, two based on novel machine learning approaches and three based on traditional approaches based on observable characteristics of researchers' publication portfolios.

In Table 3, we present the results of our main difference-in-differences estimation in equation (1), considering researchers as treated if they published in motion-sensing either before or after the launch of Kinect. The first two columns report results using the measures of diversification based on topic modeling, while the following three columns report diversification measures based on direct bibliometric measures. First, we observe that all measures show evidence of an increase in diversification of innovative output triggered by the reduction in cost of research technology. Second, we draw attention to the close estimated magnitudes of the two measures based on topic modeling techniques. In robustness checks, all other LDA-based measures of diversification remain close to these estimated coefficients. This is important, as it demonstrates an internal consistency between the topic modeling algorithms. This is an interesting finding, considering that each algorithm was run under some different assumptions, with the aim of identifying biases associated with those assumptions.

******* Please insert Table 3 about here *******

Columns 1 and 2 show evidence of a 27% increase in the diversification of yearly publications of treated researchers after the launch of Kinect relative to the control group. The analysis based on HDP indicates that the research cost reduction led to an increase of 0.6 topics out of approximately 50 topics covering the complete set of research in computer science, electrical engineering, and electronics. The LDA results imply that the cost shock induced an increase of 0.74 topics out of 50 total topics. These findings are consistent with those obtained using the traditional diversification measures. The coefficient in the remaining columns imply an 8% increase in Euclidian-based diversification, a 34% increase in the number of yearly new co-authors, and a 38% increase in the number of yearly new publication outlets.

We test the timing of these effects by replacing our $AfterKinect_t$ with a set of dummy variables for each sample year. We plot the estimated difference in yearly level of diversification between our treated and control researchers in Figure 1. All values are computed relative to 2010. In each case, we find an increase in average researcher diversification following the launch of Kinect that begins immediately after the launch and persists across time.

******* Please insert Figure 1 about here *******

Next, we estimate whether the effect on diversification has a differential impact on researchers outside and within motion-sensing as informed by the discussion in Section II. Table 4 repeats our analysis of equation (1) focusing on the group of outside-area researchers. The results of each analysis in this table suggest that the reduction in the cost of motion-sensing research technology induced diversification among treated outside-area researchers. The HDP-based diversification measure shows an estimated 47% increase in diversification, and the LDA-based measure shows a 49% increase in diversification. This amounts to about one additional category out of a total of 50 categories mapping the entire set of publications in computer science, electrical engineering, and electronics. The Euclidean-based measure indicates an increase of 12%, while the number of new co-authors shows an increase of 48% (1.11 additional new co-authors) and the number of new publication outlets shows an increase of 49% (1 in 3 additional new publication outlets). As before, we test the timing of these effects and display the yearly estimated differences in diversification measures between treated and control units in Figure 2.

******* Please insert Table 4 and Figure 2 about here *******

In Table 5, we verify whether the effects on diversification within the group of outside-area researchers persist outside the set of papers directly engaging with motion-sensing. In line with our theoretical arguments, we find evidence of such effects. Specifically, the topic-modeling-based measures evidence a 34% increase in diversification directly attributable to the reduction in cost (3 additional topics for every 4), the Euclidian-based measures show a 6% increase, the number of new co-authors increases by 27% (more than 1 in 2 new co-authors), and the number of new publication outlets increases by 49% (1 in 3 additional new publication outlets). As before, we test the timing of these effects and display the yearly estimated coefficients in Figure 3.

******* Please insert Table 5 and Figure 3 about here *******

In Table 6 and Figure 4, we report the results of our analysis focusing on the set of researchers who published motion-sensing research prior to the introduction of Kinect, i.e., within-area researchers. The LDA measure indicates a 6% increase in diversification directly attributable to the reduction in cost of research technology. The HDP measure indicates similar effects in which post-Kinect diversification increases by approximately 7%. The Euclidian-based diversification increases by 3%, the number of yearly new co-authors increases by 8%, and the number of yearly publication outlets increases by 10%.

******* Please insert Table 6 and Figure 4 about here *******

Our theoretical discussion suggests that changes in diversification of within-area researchers might be influenced by the level of within-area expertise. We test this possibility by replacing our dummy treatment variable with a continuous variable capturing the degree of involvement in motion-sensing before the launch of Kinect, calculated as the percentage of motion-sensing publications in one's yearly research portfolio (Table 7). We find that the effect on diversification indeed increases with higher levels of involvement in motion-sensing before the launch of Kinect.

******* Please insert Table 7 about here *******

IV.3. Complementary Descriptive Analyses Based on Scopus Data

The results of our regression analyses document that the reduction in the cost of motion-sensing technology induces an increase in the diversity of researchers' project portfolios. Within-area researchers contribute less to this effect, though their contribution is positive and statistically and economically significant. Among this group, researchers closest to motion-sensing in the pre-Kinect period increase their diversification the most. The greatest increase in research diversity accrues to the group that did not publish motion-sensing research prior to the Kinect shock but who entered in its wake.

Ideally, we would also like to gain insight into the type of knowledge these individuals create as part of the identified effect of diversification. However, the difficulty of capturing implications for the direction of innovation also poses challenges to our ability to speak directly on the type of knowledge these researchers produce as part of the observed increase in diversification. To shed some light on this, we provide some descriptive examples. Specifically, we select a random group of researchers from our treated and control samples and look up their publication portfolio in Scopus, a comprehensive bibliographical database tracking academic publication across all areas of science. We take advantage of a particular feature of this database, namely its "analyze" function. Scopus allows users to select a group of papers and to analyze its properties using a proprietary Scopus algorithm. The analysis displays, among other things, pie charts with the distribution of publications across knowledge areas, as defined by the Scopus taxonomy. Since the data we use in our analysis is collected from a different database (IEEE *Xplore*), which processes and classifies papers without any connection to Scopus, focusing on the Scopus analysis algorithm also provides an additional piece of evidence to our findings. We include our examples of randomly selected treated and control researchers and the pie-chart analysis of their pre- and post-Kinect publication portfolios in Appendix B. These charts suggest an increase in diversification following the launch of Kinect that is consistent with our econometric analysis and provide a glimpse into the type of knowledge these researchers create. We observe two broad avenues through which researchers increase their diversification following Kinect: (1) an increase in the number of areas where they publish and (2) an increase in the percentage of their publications across knowledge areas in their portfolio.

VI. Discussion and Conclusion

Understanding the direction of inventive activity is important for research in the economics and sociology of science and the study of the strategic management of science-based organizations. The direction of research exploration is of central concern in economics because of its relationship to the rate of economic growth (Acemoglu, 2002, 2010; Bryan and Lemus, 2016). In addition to its direct impact on welfare, issues associated with research direction are of interest to economics because of the complex interplay of financial incentives, career incentives, institutional features, and researcher preferences that shape this area of economic activity (David and Dasgupta, 1994; Stephan, 1994; Stern, 2004). For the sociology of science, the evolution of research activity is important for its impact on the construction and maintenance of research communities, the development and evolution of invisible colleges, and for changing norms in research. Issues associated with research direction are particularly salient for the study of strategic behavior of science-based organizations, including academic research labs and universities. Whereas private firms compete for profits, research laboratories and academic institutions compete for resources and reputation. As is the case with private firms, project selection and the identification of niches for specialization is one of the approaches that PIs can apply in their efforts to compete. By making such differentiating investments, science-based organizations affect the direction and breadth of basic research and shape opportunities for follow-on innovation.

Relative to research on the rate of inventive activity, the direction of inventive activity is less fully studied, in part because it faces such conceptual and empirical challenges. As Doran (2017) notes, ideas exist in a multidimensional space in which neither the number of dimensions nor the distance between the dimensions can be well known or characterized, particularly since these dimensions evolve in ways that are difficult to anticipate and can take time to recognize. It is possible to make progress through qualitative work that provides rich detail on the evolution of ideas or the trajectories of individual researchers' work. However, the problems inherent in measuring a project's location in ideas space have made it difficult to conduct this work on a larger scale or to draw generalizable conclusions. Azoulay et al. (2015) conceive of research fields based on the set of articles judged to be sufficiently "related" to a particular work, based on a National Library of Medicine algorithm for measuring article similarity (the PUBMED Related Algorithm). Other projects measure research diversity using self-selected keywords, co-authorship ties, and measures based on citation networks, such as cosine similarity measures. Just as many of these measures were enabled by advances in data storage and computational speeds, further advances in computing power have enabled the analysis of large bodies of text. We leverage these advances to develop measures of research direction that reflect the probabilistic assignment of papers to categories of research topics. Using unassisted machine learning tools, we construct one set of topic categories based on a fixed number of categories and another in which the number of categories is chosen by algorithm. These measures of research diversity have the advantage of being less subject to strategic behavior than citation networks or keywords, as they employ the full text of paper abstracts.

In this paper, we use these measures, complemented by more traditional measures, to investigate how a first-order reduction in the cost of motion-sensing technology affects researchers' project portfolios. Based on our analysis of the nearly 2.5 million publications indexed by the IEEE *Xplore* database between 2007 and 2014, we find that motion-sensing incumbents and entrants substantially increase their research diversity on all of our metrics relative to researchers matched on observable pre-Kinect characteristics. Among motion-sensing incumbents, we find that the effect on diversification is most pronounced for researchers with high levels of involvement in motion-sensing. We also find that outside-area researchers exhibit a greater increase in research diversity than inside-area researchers. Further, we find that motion-sensing entrants increase the diversity of their research portfolios even outside of motion-sensing, suggesting that the research technology acts as a conduit to opportunities for knowledge creation beyond the area of the cost reduction. These findings are aligned with theoretical arguments suggesting that the effect on diversification is large only in the presence of substantial reductions in the cost of research technology.

Together, our results paint a picture in which a reduction in the cost of research tools increases the breadth of research activity, expanding the trajectories of scholars both inside and outside the affected area. Our analysis underscores the important role of research tools and highlights the processes by which scientific research trajectories evolve. We hope that further work will build upon the topic modeling tools that we have introduced and explore new areas in which to apply unassisted machine learning to track the evolution of knowledge in ideas space.

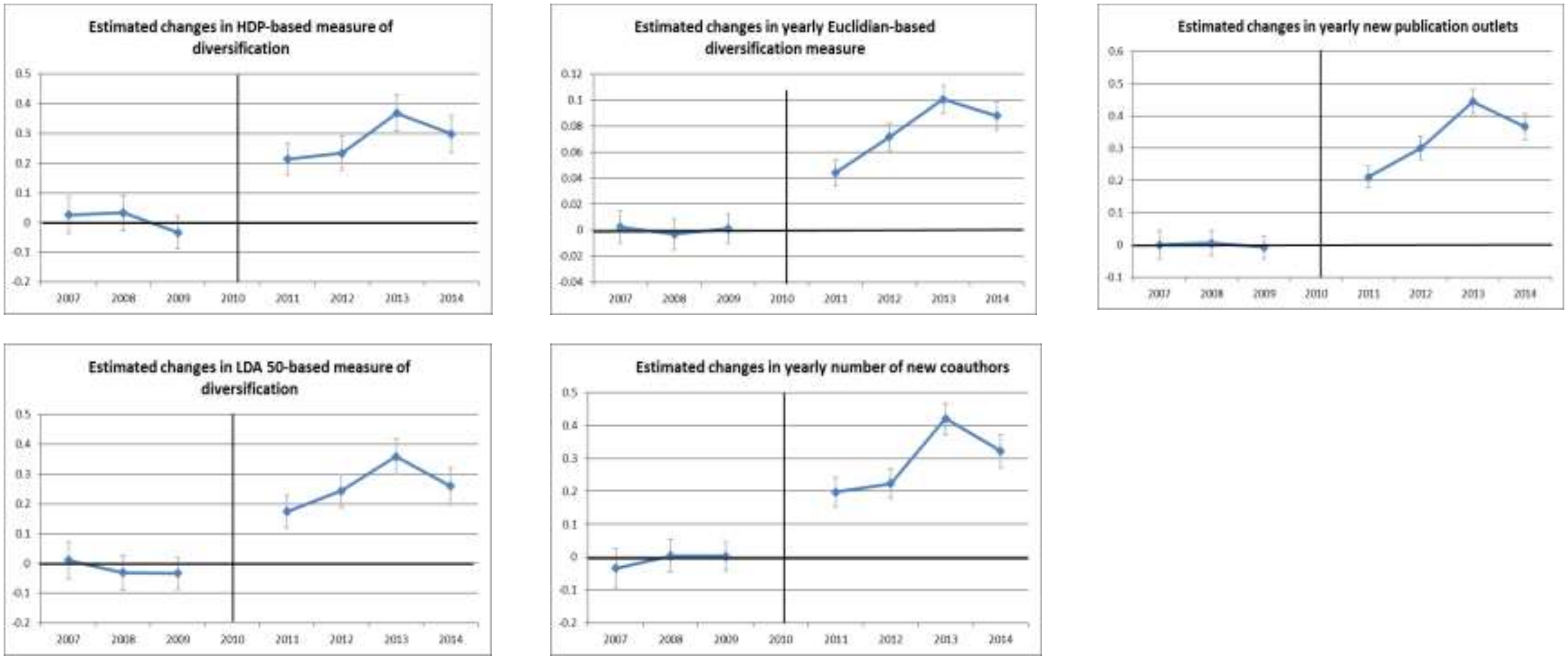
REFERENCES

1. Acemoglu, Daron (2002) “Directed Technical Change,” *Review of Economic Studies*, 69(4), 781-809.
2. ACEMOGLU, Daron (2012) “Diversity and Technological Progress” in The Rate and Direction of Inventive Activity Revisited, Scott Stern and Josh Lerner, ed. Chicago: IL: University of Chicago Press, 319-356.
3. Adafruit (2011) “Johhny Lee was behind the Kinect Bounty Contest -with a little help from us :),” Adafruit blog, https://blog.adafruit.com/2011/02/21/johhny-lee-was-behind-the-kinect-bounty-contest-with-a-little-help-from-us/, posted February 21, 2011 at 7:41 pm, accessed, July 24, 2017.
4. Aghion, Philippe, Mathias Dewatripont, and Jeremy C. Stein (2008) “Academic freedom, private-sector focus, and the process of innovation,” *Rand Journal of Economics*, 39(3), 617–635.
5. Agrawal, A., Goldfarb. A., Teodoridis, F. (2016). “Understanding the Changing Structure of Scientific Inquiry,” *American Economic Journal: Applied Economics* 8(1): 100-128.
6. Arrow, Kenneth J. (1962) “The Economic Implications of Learning by Doing,” *Review of Economic Studies*, 29(3), 155-173.
7. Athey, Susan and Guido Imbens (2015) “Lectures on Machine Learning,” NBER Summer Institute.
8. Azoulay, Pierre, Joshua S. Graff Zivin, and Jialan Wang (2010) “Superstar Extinction,” *Quarterly Journal of Economics*, 125 (2): 549-589.
9. Azoulay, Pierre, Jeffrey L. Furman, Joshua L. Krieger and Fiona Murray (2015) “Retractions”, *Review of Economics and Statistics*, 97(5): 1118-1136.
10. BBC News (2010) “Kinect Hacked Days After Release,” BBC News, http://www.bbc.co.uk/news/technology-11742236, posted November 12, 2010; accessed 24 July 2017.
11. Bilton, Nick (2011) “Microsoft Sells 10 Million Kinects, Breaking Record,” New York Times Blogs, https://bits.blogs.nytimes.com/2011/03/09/microsoft-sells-10-million-kinects-breaking-record/, posted March 9, 2011 6:08pm; accessed 24 July 2017.
12. Borjas, George and Kirk Doran (2012) “The Collapse of the Soviet Union and the Productivity of American Mathematicians,” *Quarterly Journal of Economics*, 127(3), 1143-1203.
13. Borjas, George and Kirk Doran (2015) “Cognitive Mobility: Native Responses to Supply Shocks in the Space of Ideas,” *Journal of Labor Economics*, 33(S1), S109-S145.
14. Boudreau, Kevin J., Eva C. Guinan, Karim R. Lakhani, Christoph Riedl (2017) “Looking Across and Looking Beyond the Knowledge Frontier: Intellectual Distance, Novelty, and Resource Allocation in Science,” *Management Science*, 62(10), 2765-2783.
15. Brady, Michael and H.G. Barrow (1981) Computer Vision, Amsterdam, NL: North-Holland.
16. Buntine, Wray and Aleks Jakulin (2004) “Applying discrete PCA in data analysis,” In Proceedings of the 20th conference on Uncertainty in artificial intelligence, 59-66.
17. Bryan, Kevin and Jorge Lemus (2017) “The Direction of Innovation”, working paper.
18. Carmody, Tim (2010a) “Adafruit Offers $1000 Bounty for Open-Source Kinect Drivers,” *Wired Magazine*, https://www.wired.com/2010/11/adafruit-offers-1000-bounty-for-open-source-kinect-drivers/, posted November 04, 2010, 1.28pm; accessed 24 July 2017.
19. Carmody, Tim (2010b) “How Microsoft Learned to Stop Worrying and Love Open Kinect,” *Wired Magazine*, https://www.wired.com/2010/11/open-kinect/, posted November 22, 2010, 2.25pm; accessed July 24,2017.
20. Chang, Yao-Jen, Shu-Fang Chen, and Jun-Da Huang (2011) “A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities,” *Research in Developmental Disabilities*, 32(6), 2566-2570.
21. Chang, Chien-Yen, Belinda Langel, Mi Zhang, Sebastian Koenig, Phil Requejo, Noom Somboon, Alexander A. Sawchuk, and Albert A. Rizzo (2012) “Towards Pervasive Physical Rehabilitation Using Microsoft Kinect,” 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops, 159-162.
22. Cockburn, Iain and Rebecca Henderson (1994) “Measuring Competence? Exploring Firm Effects in Pharmaceutical Research,” *Strategic Management Journal*, 15(S1), 63-84.
23. Cohen, M. Wesley and Steven Klepper (1992) “The Tradeoff between Firm Size and Diversity in the Pursuit of Technological Progress,” *Small Business Economics*, 4(1), 1-14.

24. Cohen, M. Wesley and Steven Klepper (1996) "A Reprise of Size and R&D," *The Economic Journal*, 106(437), 925-951.
25. Crane, Diana (1969) "Social Structure in a Group of Scientists: A Test of the "Invisible College" Hypothesis," *American Sociological Review*, 34(4), 335-352.
26. Dasgupta, Partha and Paul A. David (1994) "Toward a New Economics of Science," *Research Policy*, 23(5), 487-521.
27. de Solla Price, Derek (1976) "A General Theory of Bibliometric and Other Cumulative Advantage Processes," *Journal of the American Society for Information Science*, 27(5): 292–306.
28. Doran, Kirk (2017) "The Direction of Science: Natural Experiments," presentation at the June 2017 DRUID Conference, New York, NY.
29. Franzoni, C., Scelatto, G., Stephan. P. (2011) "Changing Incentives to Publish," *Science*, 333:702-703.
30. Furman, J., Stern, S. (2011) "Climbing Atop the Shoulders of Giants: The Impact of Institutions on Cumulative Knowledge Production," *American Economic Review*, 101(5): 1933-1963.
31. Gest, Howard (2004) "The Discovery of Microorganisms by Robert Hooke and Antoni van Leeuwenhoek, Fellows of the Royal Society," *Notes and Records of the Royal Society of London*, 58(2), 187-201.
32. Giles, Jim (2010) "Inside the race to hack Microsoft's Kinect," *New Scientist*, 208(2789), 1.
33. Griliches, Zvi. (1990) "Patent Statistics as Economic Indicators: A Survey," *Journal of Economic Literature*, 28(4): 1661–707.
34. Grupp, Hariolf (1990) "The Concept of Entropy in Scientometrics and Innovation Research," *Scientometrics*, 18(3-4), 219-239.
35. Han, Jungong, Ling Shao, Dong Xu, and Jamie Shotton (2013) "Enhanced Computer Vision with Microsoft Kinect Sensor: A Review," *IEEE Transactions on Cybernetics*, 43(5), 1318-1334.
36. Hanson, Allen R. and Edward M Riseman (1978) Computer Vision Systems: Papers from the Workshop on Computer Vision Systems, held at the University of Massachusetts, Amherst, Massachusetts, June 1-3, 1977. New York, NY: Academic Press.
37. Haralick, Robert M., K. Shanmugam, and Its'hak Dinstein (1973) "Textural Features for Image Classification," *IEEE Transactions on Systems, Man, And Cybernetics*, 3(6), 610-621.
38. Henderson, Rebecca and Iain Cockburn (1996) "Scale, Scope, and Spillovers: The Determinants of Research Productivity in Drug Discovery," *RAND Journal of Economics*, 27(1), 32-59.
39. Hofmann, Thomas (1999) "Probabilistic latent semantic analysis," in Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence, 289-296.
40. Iacus, S. M. and King G. Porro (2011) "Multivariate matching methods that are monotonic imbalance bounding", *Journal of American Statistics Association*, 106(493): 345-361.
41. Iacus, S.M. and King G. Porro (2012) "Causal inference without balance checking: Coarsened exact matching", *Political Analysis* 20(1):1-24.
42. Jones, B. (2009) "The Burden of Knowledge and the Death of the Renaissance Man: Is Innovation Getting Harder?" *Review of Economic Studies*, 76(1): 253-281.
43. Jones, B. (2010) "As Science Evolves, How Can Science Policy?" *NBER Innovation Policy and the Economy*, 11: 103-131.
44. Kirsch Russell. A., Lee Cahn, C. Ray, and Genevieve H. Urban (1957) "Experiments in processing pictorial information with a digital computer," IRE-ACM-AIEE '57 (Eastern): Papers and discussions presented at the December 9-13, 1957, Eastern Joint Computer Conference, 221-229.
45. Lee, Johnny Chung (2011) "Windows Drivers for Kinect, Finally!" procrastineering blog, http://procrastineering.blogspot.com/2011/02/windows-drivers-for-kinect.html, posted February, 21, 2011, accessed July 24, 2017.
46. Lugt, A. Vander (1964) "Signal Dectection by Complex Spatial Filtering," *IEEE Transactions on Information Theory*, 10(2), 139-145.
47. Mokyr, Joel (2002) "The Gifts of Athena: Historical Origins of the Knowledge Economy," Princeton University Press: Princeton, NJ.
48. Murray, F., Aghion, P., Dewatripont, M., Kolev, J., & Stern, S. (2016) "Of Mice and Academics: Examining the Effect of Openness on Innovation," *American Economic Journal: Economic Policy*, 8(1), 212-52.

49. Lerner, Josh and Scott Stern, eds. (2012) The Rate and Direction of Inventive Activity Revisited, Chicago, IL: University of Chicago Press.
50. National Bureau of Economic Research (1962) The Rate and Direction of Inventive Activity: Economic and Social Factors, Princeton, NJ: Princeton University Press.
51. Merton, Robert K. (1957) "Priorities in Scientific Discovery: A Chapter in the Sociology of Science," *American Sociological Review*, 22(6), 635-659.
52. Nanni, Loris, Alessandra Lumini, Fabio Dominio, and Pietro Zanuttigh (2014) "Effective and precise face detection based on color and depth data," *Applied Computing and Informatics*, 10(1), 1-13.
53. Nelson, R. R. (1982) "The Role of Knowledge in R&D Efficiency," *Quarterly Journal of Economics*, 97(3): 453-470.
54. Preis, Johannes, Moritz Kessel, Martin Werner, and Claudia Linnhoff-Popien (2012) "Gait recognition with Kinect," In 1st international workshop on Kinect in pervasive computing, P1-P4.
55. Rafibakhsh, Nima, Jie Gong, Mohsin K. Siddiqui, Chris Gordon, and H. Felix Lee (2012) "Analysis of Xbox Kinect sensor data for use on construction sites: Depth accuracy and sensor interference assessment," Construction Research Congress 2012: Construction Challenges in a Flat World, 848-857.
56. Rafols, I., & Meyer, M. (2010) "Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience," *Scientometrics*, 82(2), 263–287.
57. Richards-Rissetto, H., Remondino, F., Agugiaro, G., von Schwerin, J., Robertsson, J. and Girardi, G., (2012) "Kinect and 3D GIS in archaeology," IEEE, 18th International Conference on Virtual Systems and Multimedia, 331-337.
58. Romer, P. M. (1990) "Endogenous Technological Change," *Journal of Political Economy*, 98(5):71-102.
59. Solow, R. M. (1956) "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics*, 70: 65-94.
60. Stephan, Paula E. (1996) "The Economics of Science," *Journal of Economic Literature*, 34 (3), 1199-1235.
61. Stephan, Paula E. (2012) "How Economics Shape Science", Harvard University Press, Cambridge, Massachusetts.
62. Stern, Scott (2004) "Do Scientists Pay to be Scientists?" *Management Science*, 50(6), 835-853.
63. Scotchmer, S. (1996) "Standing on the shoulders of giants: Cumulative research and the patent law," *The Journal of Economic Perspectives*, 5(1), 29-41.
64. Stone, Erik and Marjorie Skubic (2013) "Unobtrusive, Continuous, In-Home Gait Measurement Using the Microsoft Kinect," *IEEE Transactions on Biomedical Engineering*, 60(10) 2925 – 2932.
65. Sutherland, Ivan E. (1968) "A head-mounted three dimensional display," AFIPS '68: Proceedings of the December 9-11, 1968, Fall Joint Computer Conference, Part I, 757-764.
66. Tanz, Jason (2011) "Kinect hackers are changing the future of robotics," Wired Magazine, posted, June 28, 2011; accessed July 24, 2017.
67. Teodoridis, Florenta (2017) "Understanding Team Knowledge Production: The Interrelated Roles of Technology and Expertise," *Management Science*, forthcoming.
68. Terdiman, Daniel (2010) "Bounty Offered for Open-Source Kinect Driver," CNET News, https://www.cnet.com/news/bounty-offered-for-open-source-kinect-driver/, posted November 4, 2010, 12.50pm, updated 4.03pm and 5.07pm; accessed 24 July 2017.
69. Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013) "Atypical combinations and scientific impact" *Science*, 342(6157), 468-472.
70. Van Helden, Albert (1975), "The Historical Problem of the Invention of the Telescope," 13(4), History of Science, 251-263.
71. Winston, Patrick Henry, ed. (1975) The Psychology of Computer Vision, New York, NY: McGraw-Hill.
72. Wortham, Jenna (2010) "Off the Shelf and into the garage," New York Times, November 22, page B1.
73. Zhang, Zhengyou (2012) "Microsoft Kinect Sensor and Its Effect," *IEEE Multimedia*, 19(2), 4–10.
74. Zucker, Lynne G., Michael R. Darby, and Marilynn B. Brewer (1998) "Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises," *American Economic Review*, 88, 290-306.

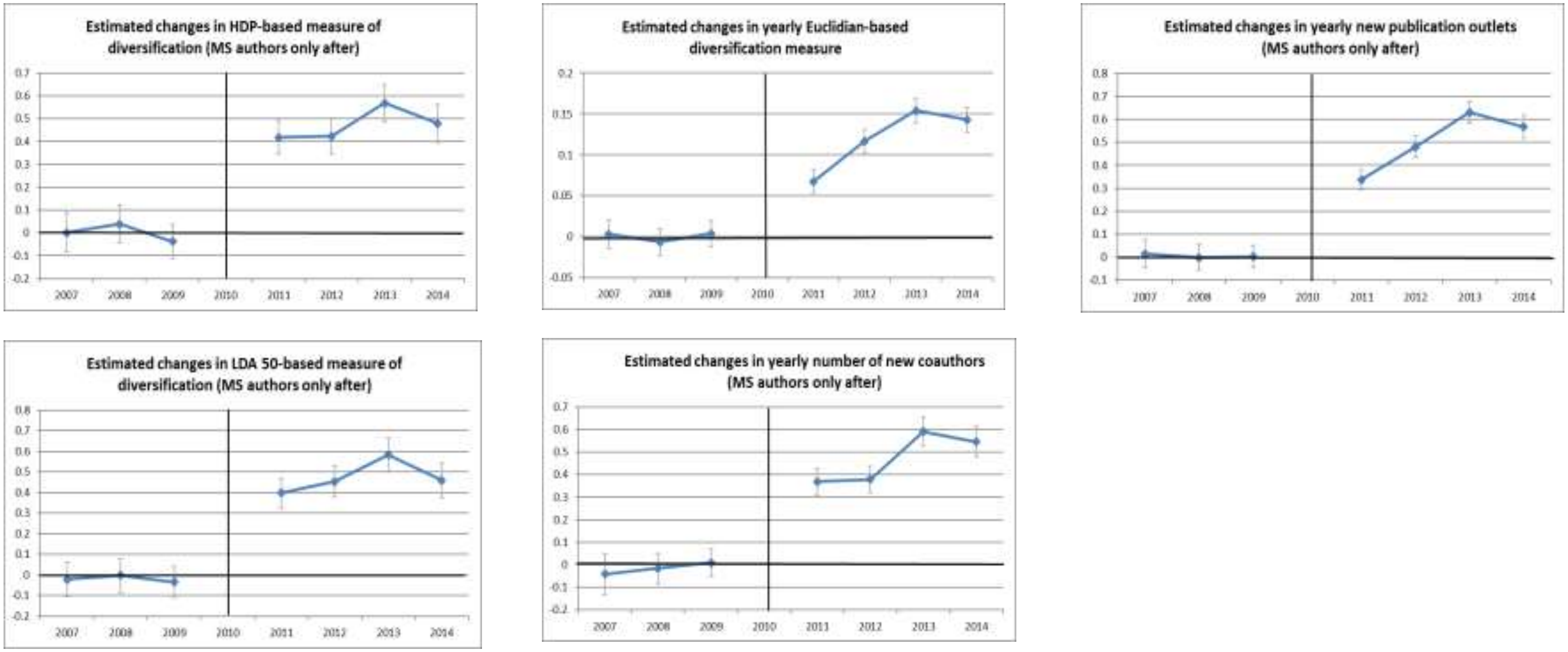
Figure 1
Change in individual-level diversification for researchers who published in motion-sensing either before or after the launch of Kinect



Notes: We base this figure on our 2007-2014 dataset. Each point on the graph represents the coefficient value on the covariate *TreatedResearcher x Year* and thus describes the relative difference in diversification between treated and control authors in that year. The bars surrounding each point represent the 95 percent confidence interval. All values are relative to the base year of 2010.

Figure 2

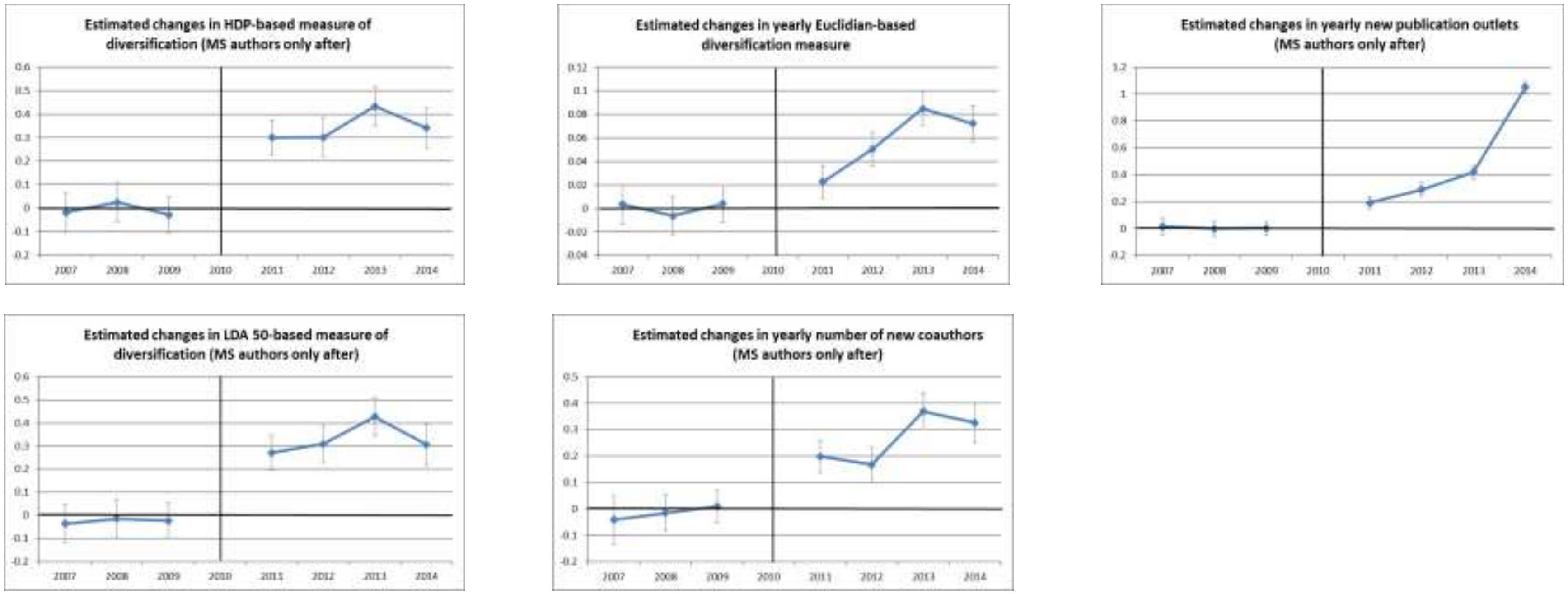
Change in individual-level diversification for researchers who published in motion-sensing after the launch of Kinect



Notes: We base this figure on our 2007-2014 dataset. Each point on the graph represents the coefficient value on the covariate *TreatedResearcher x Year* and thus describes the relative difference in diversification between treated and control authors in that year. The bars surrounding each point represent the 95 percent confidence interval. All values are relative to the base year of 2010.

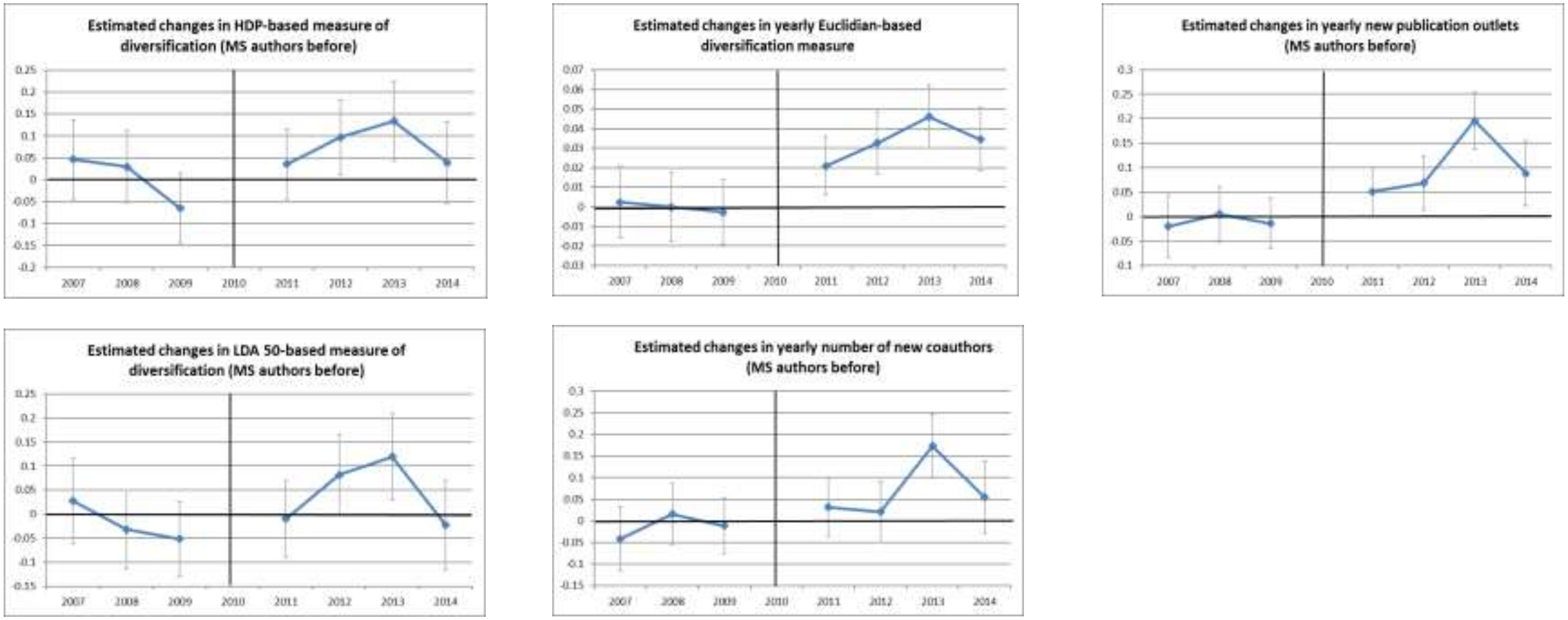
Figure 3

Change in individual-level diversification for researchers who published in motion-sensing after the launch of Kinect, when excluding their motion-sensing publications



Notes: We base this figure on our 2007-2014 dataset. Each point on the graph represents the coefficient value on the covariate *TreatedResearcher x Year* and thus describes the relative difference in diversification between treated and control authors in that year. The bars surrounding each point represent the 95 percent confidence interval. All values are relative to the base year of 2010.

Figure 4
Change in individual-level diversification for researchers who published in motion-sensing before the launch of Kinect



Notes: We base this figure on our 2007-2014 dataset. Each point on the graph represents the coefficient value on the covariate *TreatedResearcher x Year* and thus describes the relative difference in diversification between treated and control authors in that year. The bars surrounding each point represent the 95 percent confidence interval. All values are relative to the base year of 2010.

Table 1.a: Descriptive statistics for our main sample of researchers and their publications (2007-2014)

Descriptive statistics											
		(1) Topic modeling				(2) Classic metrics					
Treated as		HDP topics weighted by fit probability (log)		LDA 50 topics weighted by fit probability (log)		Yearly Euclidian-based diversification		New Collaborators		New Publication Outlets	
		Treated	Controls	Treated	Controls	Treated	Controls	Treated	Controls	Treated	Controls
Researchers who published in motion-sensing either before or after or both	Before	3.269	2.695	3.820	3.159	0.489	0.387	6.123	4.031	1.755	1.153
		(1.593)	(1.406)	(1.575)	(1.407)	(0.269)	(0.288)	(9.299)	(6.939)	(1.987)	(1.407)
	After	3.521	2.773	3.981	3.150	0.461	0.321	6.764	3.558	1.751	0.927
		(1.695)	(1.494)	(1.681)	(1.499)	(0.272)	(0.289)	(10.457)	(6.532)	(2.176)	(1.368)
Observations	Before	71,658	451,148	71,658	451,148	88,556	643,380	88,556	643,380	88,556	643,380
	After	69,110	393,897	69,110	393,897	88,556	643,380	88,556	643,380	88,556	643,380
Researchers who published in motion-sensing before	Before	3.349	2.695	3.908	3.159	0.504	0.387	6.357	4.031	1.779	1.153
		(1.601)	(1.406)	(1.580)	(1.407)	(0.264)	(0.288)	(9.432)	(6.939)	(1.992)	(1.407)
	After	3.449	2.773	3.906	3.150	0.438	0.321	6.182	3.558	1.580	0.927
		(1.702)	(1.494)	(1.693)	(1.499)	(0.281)	(0.289)	(9.987)	(6.532)	(2.084)	(1.368)
Observations	Before	41,415	451,148	41,415	451,148	50,196	643,380	50,196	643,380	50,196	643,380
	After	37,512	393,897	37,512	393,897	50,196	643,380	50,196	643,380	50,196	643,380
Researchers who published in motion-sensing after	Before	3.158	2.695	3.700	3.159	0.470	0.588	5.815	4.031	1.724	1.153
		(1.574)	(1.406)	(1.559)	(1.407)	(0.275)	(0.095)	(9.113)	(6.939)	(1.979)	(1.407)
	After	3.606	2.773	4.071	3.150	0.491	0.560	7.526	3.558	1.975	0.927
		(1.684)	(1.494)	(1.663)	(1.499)	(0.257)	(0.109)	(10.996)	(6.532)	(2.271)	(1.368)
Observations	Before	30,243	451,148	30,243	451,148	38,360	643,380	38,360	643,380	38,360	643,380
	After	31,598	393,897	31,598	393,897	38.360	643,380	38.360	643,380	38.360	643,380

Table 1.b: Descriptive statistics for our main matched (CEM) sample of researchers and their publications (2007-2014)

Descriptive statistics											
		(1) Topic modeling				(2) Classic metrics					
Treated as		HDP topics weighted by fit probability (log)		LDA 50 topics weighted by fit probability (log)		Yearly Euclidian-based diversification		New Collaborators		New Publication Outlets	
		Treated	Controls	Treated	Controls	Treated	Controls	Treated	Controls	Treated	Controls
Researchers who published in motion-sensing either before or after or both	Before	2.133	2.154	2.733	2.679	0.357	0.355	2.307	2.286	0.860	0.892
		(1.075)	(1.098)	(1.076)	(1.116)	(0.295)	(0.294)	(3.355)	(3.287)	(0.990)	(1.028)
	After	2.599	2.344	3.085	2.761	0.356	0.282	3.186	2.332	0.980	0.730
		(1.328)	(1.235)	(1.318)	(1.237)	(0.287)	(0.284)	(5.055)	(4.377)	(1.247)	(1.041)
Observations	Before	17,909	109,274	17,909	109,274	28,192	187,680	28,192	187,680	28,192	187,680
	After	18,264	95,812	18,264	95,812	28,192	187,680	28,192	187,680	28,192	187,680
Researchers who published in motion-sensing before	Before	2.118	2.141	2.725	2.668	0.360	0.358	2.242	2.258	0.839	0.883
		(1.066)	(1.090)	(1.057)	(1.108)	(0.296)	(0.294)	(3.086)	(3.074)	(0.964)	(1.014)
	After	2.418	2.350	2.890	2.766	0.313	0.283	2.565	2.359	0.789	0.739
		(1.256)	(1.251)	(1.252)	(1.253)	(0.288)	(0.285)	(4.555)	(4.387)	(1.111)	(1.055)
Observations	Before	8,409	65,544	8,409	65,544	13,168	111,936	13,168	111,936	13,168	111,936
	After	7,643	56,500	7,643	56,500	13,168	111,936	13,168	111,936	13,168	111,936
Researchers who published in motion-sensing after	Before	2.123	2.141	2.714	2.673	0.350	0.350	2.311	2.279	0.872	0.879
		(1.068)	(1.088)	(1.078)	(1.105)	(0.294)	(0.294)	(3.514)	(3.433)	(1.001)	(1.021)
	After	2.762	2.326	3.254	2.748	0.398	0.280	3.836	2.336	1.172	0.720
		(1.366)	(1.242)	(1.348)	(1.239)	(0.281)	(0.284)	(5.724)	(4.401)	(1.360)	(1.042)
Observations	Before	8,647	85,095	8,647	85,095	13,808	147,160	13,808	147,160	13,808	147,160
	After	9,817	74,390	9,817	74,390	13,808	147,160	13,808	147,160	13,808	147,160

Table 2: New researcher entries in electrical engineering, computer science, and electronics, per year, as observed through publications logged in the IEEE Xplore database (2007-2014)

	Number of authors entering IEEE Xplore academic publication		
	Number of authors entering with at least one motion-sensing publication	Number of authors entering with publications in other areas	Percentage of motion-sensing entry
2007	1,103	122,322	0.90%
2008	1,307	124,670	1.04%
2009	1,188	120,554	0.99%
2010	1,502	152,237	0.99%
2011	1,794	137,233	1.31%
2012	2,449	135,514	1.81%
2013	2,805	116,184	2.41%
2014	2,985	119,265	2.50%

Table 3: Researchers experience a change in the level of diversification of their publication portfolio after the launch of Kinect

Treated as researchers who published motion-sensing papers either in the before or after period or both Controls determined through Coarsened Exact Matching (CEM)					
	(1) Topic modeling		(2) Classic metrics		
	HDP topics weighted by fit probability (log)	LDA 50 topics weighted by fit probability (log)	Yearly Euclidian-based diversification	New Collaborators	New Publication Outlets
Treated x AfterKinect	0.271***	0.270***	0.076***	0.295***	0.322***
	(0.017)	(0.017)	(0.003)	(0.013)	(0.011)
				[1.343***]	[1.380***]
Age - sq	0.004***	0.004***	-0.001***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
				[1.002***]	[1.002***]
Year FE	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes
LL				-975,119.90	-382,210.95
R-sq	0.020	0.018	0.024		
Observations	241,259	241,259	431,744	430,424	430,592

Notes: The data is a panel at the author level based on publication data between 2007 and 2014. All models have robust standard errors clustered at the individual level. +significant at 15%, *significant at 10%, **significant at 5%, ***significant at 1%

Table 4: Evidence of change in diversification of publication portfolios for researchers who published in motion-sensing only after the launch of Kinect

	Treated as researchers who published motion-sensing papers in the after period Controls determined through Coarsened Exact Matching (CEM)				
	(1) Topic modeling		(2) Classic metrics		
	HDP topics weighted by fit probability (log)	LDA 50 topics weighted by fit probability (log)	Yearly Euclidian-based diversification	New Collaborators	New Publication Outlets
Treated x AfterKinect	0.472***	0.488***	0.120***	0.478***	0.490***
	(0.023)	(0.023)	(0.004)	(0.018)	(0.014)
				[1.613***]	[1.632***]
Age - sq	0.004***	0.004***	-0.001***	0.002***	0.002***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
				[1.002***]	[1.002***]
Year FE	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes
LL				-727,391.20	-281,979.16
R-sq	0.029	0.029	0.034		
Observations	177,949	177,949	321,936	320,856	321,152

Notes: The data is a panel at the author level based on publication data between 2007 and 2014. All models have robust standard errors clustered at the individual level. +significant at 15%, *significant at 10%, **significant at 5%, ***significant at 1%

Table 5: The result on diversification for researchers who published in motion-sensing only after the launch of Kinect not mechanically driven by engagement with motion-sensing

	Treated as reserachers who published motion-sensing papers in the after period Controls determined through Coarsened Exact Matching (CEM) All DVs exclude motion-sensing papers				
	(1) Topic modeling		(2) Classic metrics		
	HDP topics weighted by fit probability (log); exclude MS papers	LDA 50 topics weighted by fit probability (log); exclude MS papers	Yearly Euclidian-based diversification; exclude MS papers	New Collaborators; exclude MS papers	New Publication Outlets; exclude MS papers
Treated x AfterKinect	0.348***	0.344***	0.057***	0.272***	0.490***
	(0.024)	(0.024)	(0.004)	(0.021)	(0.014)
				[1.312***]	[1.632***]
Age - sq	0.004***	0.004***	-0.001***	0.002***	0.002***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
				[1.002***]	[1.002***]
Year FE	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes
LL				-723,748.85	-282,404.98
R-sq	0.022	0.022	0.030		
Observations	176,538	176,538	321,936	320,824	321,152

Notes: The data is a panel at the author level based on publication data between 2007 and 2014. All models have robust standard errors clustered at the individual level. +significant at 15%, *significant at 10%, **significant at 5%, ***significant at 1%

Table 6: Evidence of change in diversification of publication portfolios for researchers who published in motion-sensing before the launch of Kinect

	Treated as researchers who published motion-sensing papers in the before period Controls determined through Coarsened Exact Matching (CEM)				
	(1) Topic modeling		(2) Classic metrics		
	HDP topics weighted by fit probability (log)	LDA 50 topics weighted by fit probability (log)	Yearly Euclidian-based diversification	New Collaborators	New Publication Outlets
Treated x AfterKinect	0.076***	0.058**	0.034***	0.078***	0.104***
	(0.025)	(0.025)	(0.004)	(0.020)	(0.017)
				[1.081***]	[1.110***]
Age - sq	0.004***	0.004***	-0.001***	0.002***	0.002***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
				[1.002***]	[1.002***]
Year FE	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes
LL				-559,122.94	-220,327.75
R-sq	0.013	0.009	0.026		
Observations	138,096	138,096	250,208	249,296	249,568

Notes: The data is a panel at the author level based on publication data between 2007 and 2014. All models have robust standard errors clustered at the individual level. $^{+}$significant at 15%, *significant at 10%, **significant at 5%, ***significant at 1%

Table 7: The result on diversification for researchers who published in motion-sensing before Kinect is more pronounced for individuals with higher involvement in motion-sensing before Kinect

	Treated as researchers who published motion-sensing papers in the before period Controls determined through Coarsened Exact Matching (CEM)				
	(1) Topic modeling		(2) Classic metrics		
	HDP topics weighted by fit probability (log)	LDA 50 topics weighted by fit probability (log)	Yearly Euclidian-based diversification	New Collaborators	New Publication Outlets
Fraction of MS x	0.211***	0.165***	0.084***	0.241***	0.306***
AfterKinect	(0.054)	(0.053)	(0.009)	(0.049)	(0.038)
				[1.273***]	[1.358***]
Age - sq	0.005***	0.004***	-0.001***	0.002***	0.002***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
				[1.002***]	[1.002***]
Year FE	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes
LL				-559,102.20	-220,319.76
R-sq	0.012	0.008	0.026		
Observations	138,096	138,096	250,208	249,296	249,568

Notes: The data is a panel at the author level based on publication data between 2007 and 2014. All models have robust standard errors clustered at the individual level. $^{+}$significant at 15%, *significant at 10%, **significant at 5%, ***significant at 1%

APPENDIX A

Table A1: CEM balance, where treated researchers are defined as scientists who published in motion-sensing either before or after Kinect

CEM balance						
	Full Sample			Matched Sample (CEM)		
	Treated	Controls	t-stat	Treated	Controls	t-stat
Citation-weighted publication count 2007	14.055	7.397	88.48	2.340	2.317	0.85
Citation-weighted publication count 2008	13.385	7.802	79.28	2.644	2.633	0.32
Citation-weighted publication count 2009	8.065	4.729	98.34	2.017	1.991	1.58
Citation-weighted publication count 2010	7.038	4.301	101.60	2.367	2.339	1.70
Author count 2007	13.957	9.052	71.31	3.678	3.643	0.76
Author count 2008	14.292	9.890	74.76	4.157	4.168	0.30
Author count 2009	14.491	9.517	98.25	4.766	4.752	0.42
Author count 2010	15.837	10.528	99.55	6.282	6.241	1.03
Diversification index 2007-2010	0.683	0.640	162.42	0.657	0.657	
Observations	177,112	1,286,760		56,384	375,360	

Table A2: CEM balance, where treated researchers are defined as scientists who published in motion-sensing before Kinect

CEM balance						
	Full Sample			Matched Sample (CEM)		
	Treated	Controls	t-stat	Treated	Controls	t-stat
Citation-weighted publication count 2007	15.868	7.397	87.83	2.323	2.300	0.61
Citation-weighted publication count 2008	14.609	7.802	75.44	2.664	2.615	1.36
Citation-weighted publication count 2009	8.687	4.729	91.31	1.957	1.918	1.73
Citation-weighted publication count 2010	7.287	4.301	87.20	2.220	2.190	1.32
Author count 2007	15.212	9.052	70.02	3.609	3.575	0.73
Author count 2008	15.253	9.890	70.80	4.318	4.329	0.24
Author count 2009	15.290	9.517	89.36	4.585	4.602	0.37
Author count 2010	16.362	10.528	85.91	6.028	5.989	0.67
Diversification index 2007-2010	0.687	0.640	137.17	0.661	0.661	0.42
Observations	100,392	1,286,760		26,336	223,872	

Table A3: CEM balance, where treated researchers are defined as scientists who published in motion-sensing after Kinect

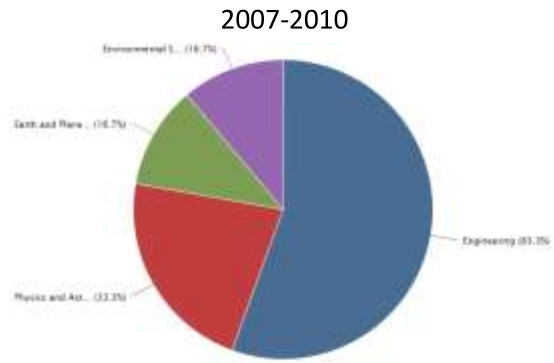
CEM balance						
	Full Sample			Matched Sample (CEM)		
	Treated	Controls	t-stat	Treated	Controls	t-stat
Citation-weighted publication count 2007	11.683	7.397	41.18	2.136	2.110	0.70
Citation-weighted publication count 2008	11.784	7.802	39.81	2.330	2.378	1.10
Citation-weighted publication count 2009	7.252	4.729	52.70	1.946	1.934	0.62
Citation-weighted publication count 2010	6.712	4.301	62.89	2.433	2.405	1.20
Author count 2007	12.314	9.052	32.99	3.430	3.451	0.34
Author count 2008	13.035	9.890	36.89	3.946	3.966	0.45
Author count 2009	13.445	9.517	54.08	4.777	4.738	0.85
Author count 2010	15.150	10.528	60.50	6.474	6.431	0.77
Diversification index 2007-2010	0.677	0.640	94.98	0.653	0.653	0.02
Observations	76,720	1,286,760		27,616	294,320	

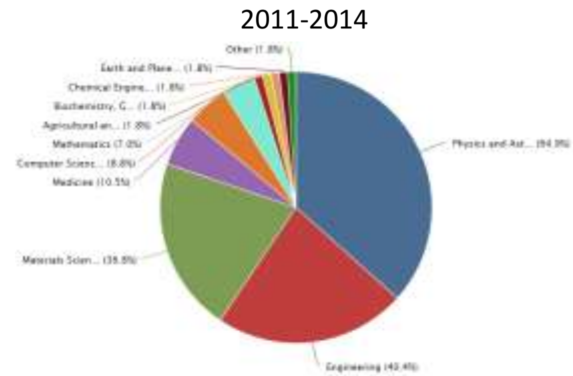
APPENDIX B

Examples of changes in publication portfolio distribution across research topics, as defined by Scopus

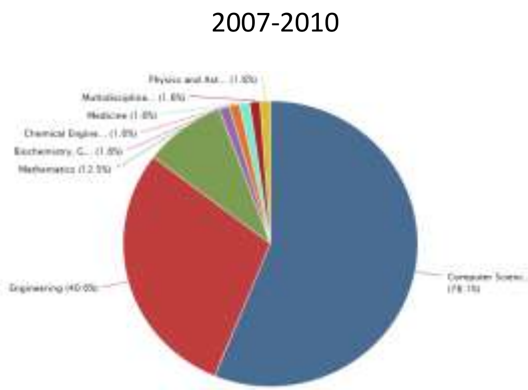
(names not shown due to privacy concerns)

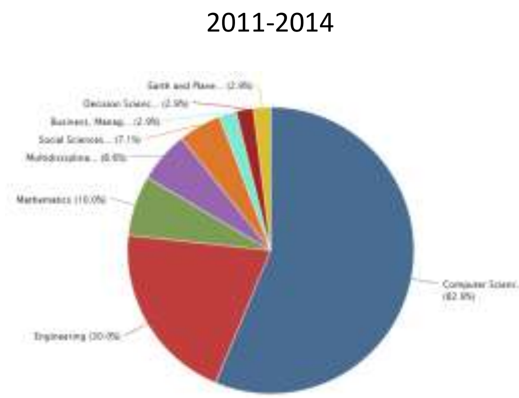
Example 1. Treated researcher





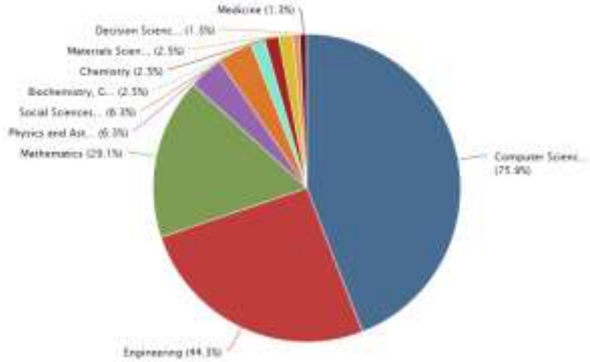
Example 2. Treated researcher



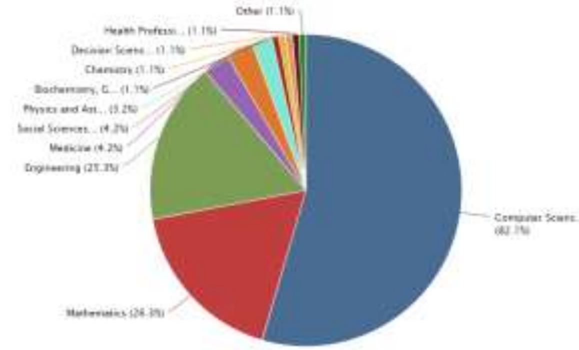


Example 3. Treated researcher

2007-2010

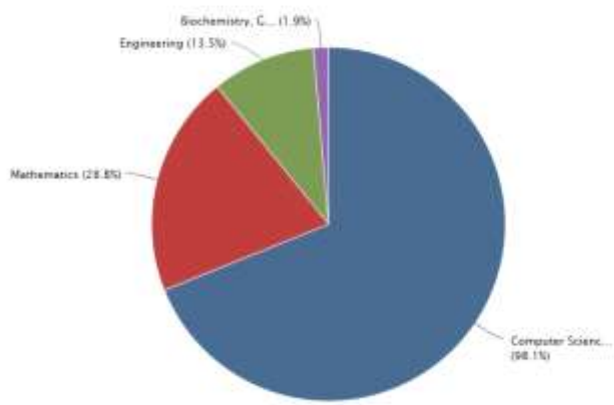


2011-2014

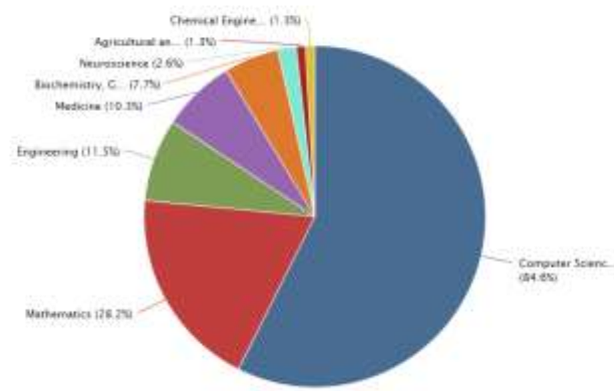


Example 4. Treated researcher

2007-2010

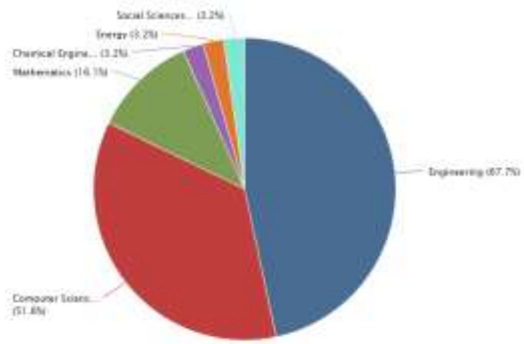


2011-2014

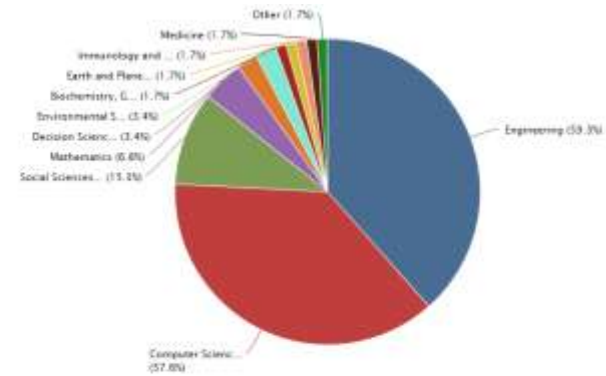


Example 5. Treated researcher

2007-2010

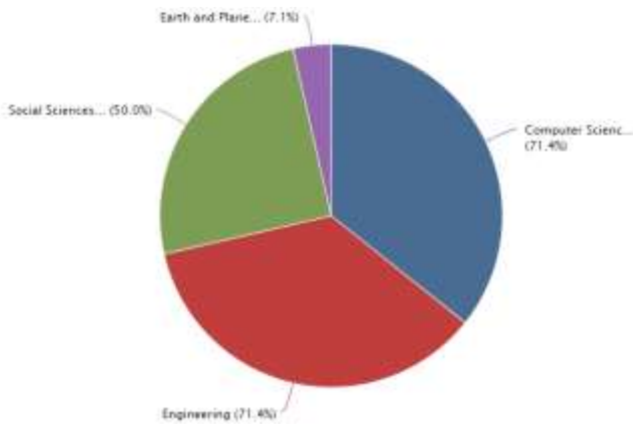


2011-2014

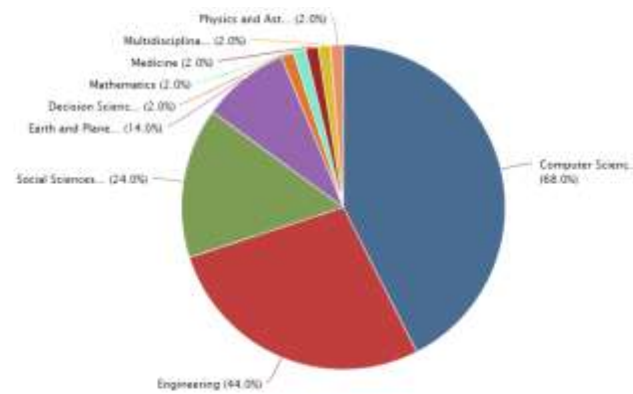


Example 6. Treated researcher

2007-2010

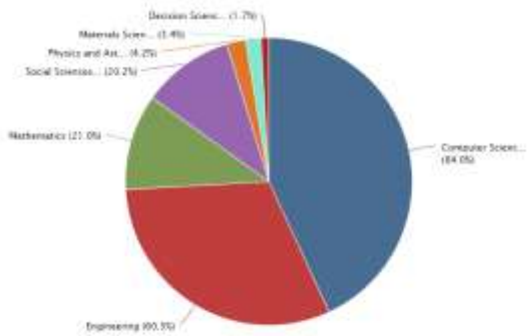


2011-2014

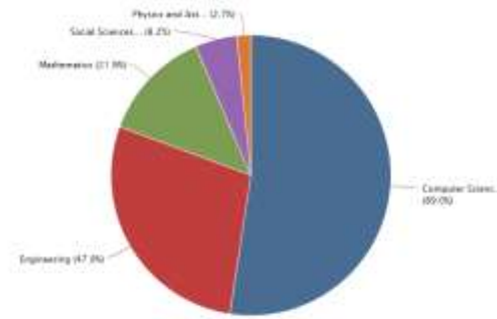


Example 7. Control researcher

2007-2010

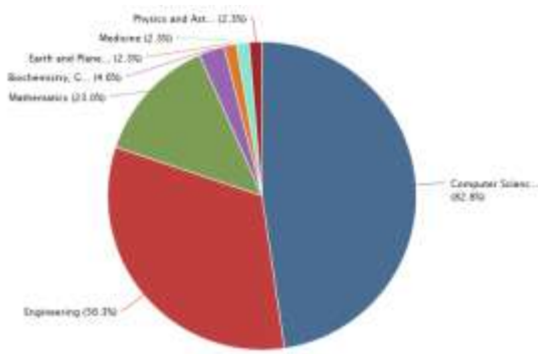


2011-2014

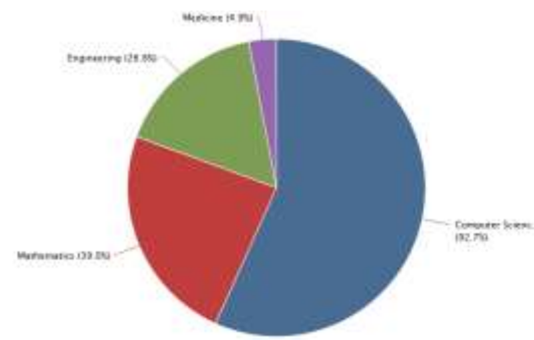


Example 8. Control researcher

2007-2010

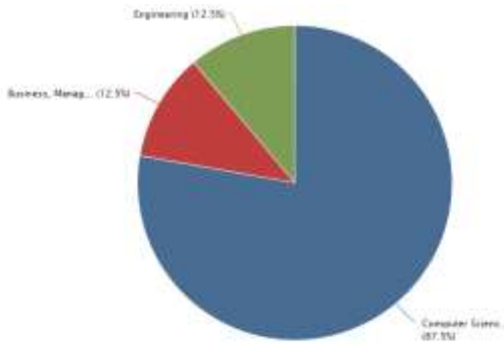


2011-2014

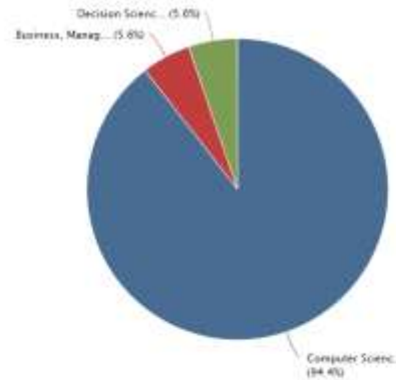


Example 9. Control researcher

2007-2010

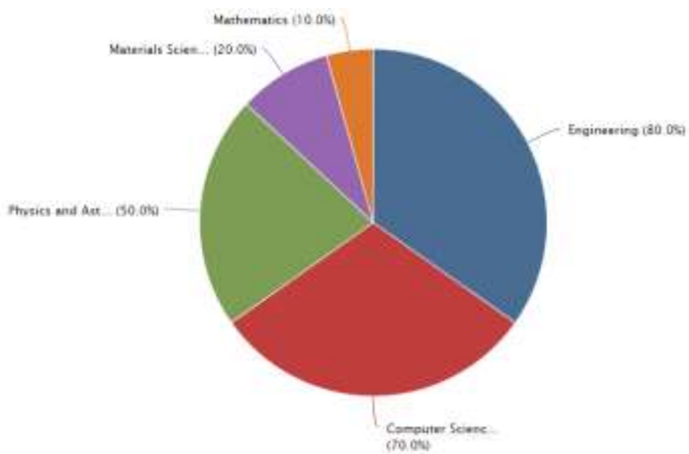


2011-2014

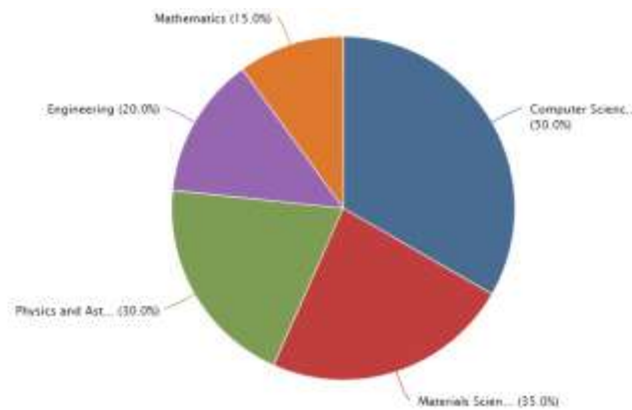


Example 10. Control researcher

2007-2010

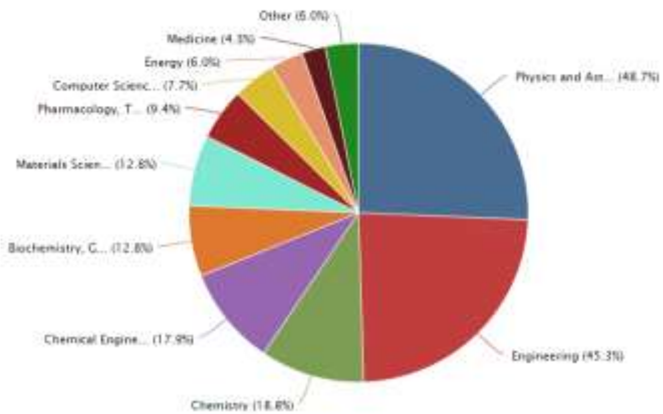


2011-2014

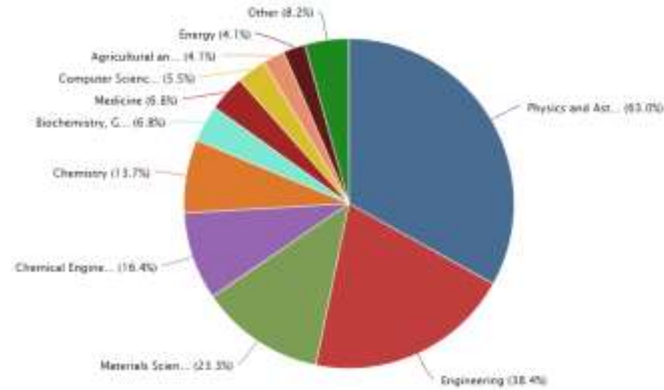


Example 11. Control researcher

2007-2010

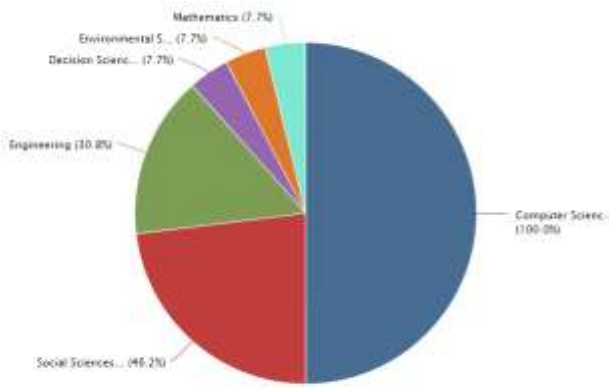


2011-2014



Example 12. Control researcher

2007-2010



2011-2014