THE ROLES OF KNOWLEDGE COMPLEXITY AND LOCATION COMPLEXITY FOR THE STRUCTURE OF KNOWLEDGE BUILDING IN A GLOBAL ENVIRONMENT

by

JESSICA RAE SALMON

A Dissertation submitted to the

Graduate School – Newark

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Management

written under the direction of

John A. Cantwell

and approved by

Newark, New Jersey

October 2017

ABSTRACT OF THE DISSERTATION

THE ROLES OF KNOWLEDGE COMPLEXITY AND LOCATION COMPLEXITY FOR THE STRUCTURE OF KNOWLEDGE BUILDING IN A GLOBAL ENVIRONMENT

By JESSICA RAE SALMON

Dissertation Director:

John A. Cantwell

Scholars have long studied the complexity of knowledge in innovation. More recently, research has begun to focus attention on the role of knowledge recombination as a way to understand knowledge complexity, knowledge growth, and evolutionary search. Yet little is known about knowledge complexity in the broad context of globalization. We build on knowledge recombination patterns in global innovation activities to develop our theory of the relationship through which earlier contributions to knowledge become inputs to subsequent knowledge building that generates more or less complex knowledge artifacts. We propose that knowledge complexity rises when recombined elements are sourced across two dimensions of distance, characterized by combining sources taken from disparate knowledge fields and distinct geographical locations. The study draws upon and compares three alternative ways of measuring the complexity of technological knowledge through patent data.

This dissertation establishes two new methods for measuring complexity and adapts a third measure for wider applicability in research. Study 1 results show there no clear relationship between technological distance and complexity as measured through either co-classification or cross-classification data. We establish the growth of the ICT era has also facilitated increases in knowledge complexity while the turbulence from ICT is indirectly increasing knowledge complexity. We end this study with a direct comparison of two measures for knowledge complexity to establish which aspects of complexity each best reflects. In study 2, we find increasing knowledge complexity also increases locational complexity. Digging deeper we see there are divergent effects from the use of both knowledge complexity measures when investigating locational complexity which further establishes the uniqueness of each knowledge complexity measure. We further assess the representative distinguishing characteristics of each complexity measure. We also establish here that ICT is contributing to increasing locational complexity as ICT is a connector of both technology fields and geographic locations. In study 3 we examine the outliers of the relationships examined for each complexity measure.

Acknowledgements

I first wish to thank my adviser, John Cantwell, for his support, assistance, mentorship, and friendship with whom I spent many a fine afternoon drinking tea and laughing our way through our meetings. I am also thankful to my dissertation committee members Farok Contractor, Lucia Piscitello, and Sinéad Monaghan for graciously joining me in this process, their feedback provided, and their gentle guidance over the last few years. To Farok for supporting this research direction and expanding my research network through allied projects. To Lucia for her timely assistance with methodology and developing my methods skill base. To Sinéad for being that friend and work associate who always knew the right thing to say even when I didn't want to hear it.

I gratefully acknowledge and am immensely thankful for the financial support provided by the Rutgers Business School – Newark for years of assistantships through the Rutgers PhD program, the RBS and GS-N Dean's Office Summer Research Scholarship, and the Prudential Chair in Business Research Award which enabled me to finish my dissertation in the expected timeframe. My many thanks to Goncalo Filipe, Monnique DeSilva, and Dawn Gist who helped me over and through various administrative obstacles.

I also extend my special thanks to my colleagues Pallavi, Beyza, Seho, Alex, Andres, Salma, Yuan-Yuan, Jeongho, Sarah, Kyungjoong, Nuruzzaman, Suli; Grace, Sebastian, Fernando, Shoshana, Kelichi, Andika, Vincent, Wen, Nilofar, Emine, Lutisha, Hafiz, Setiadi, and Steve for joining me on this incredible intellectual adventure. I wish nothing but the best for all of you.

I wish to thank my parents Joe and Darcy, along with my sister Tara – thanks for reminding that the world is still turning outside my dissertation.

With deep gratitude I wish to thank my husband Justin who provided the final year of financial support and made dinner every night during those final crucial months. I'm so lucky.

Last, I'd like to thank my cats Biddy and Wednesday who would make themselves available to chase crumpled bits of lousy dissertation writing…my how I provided you with hours of fun at the beginning.

TABLE OF CONTENTS

LIST OF TABLES, GRAPHS, AND ILLUSTRATIONS: IN TEXT

Chapter 3: Data Overview

Chapter 7: Empirical Results

Study 1

Study 2

LIST OF TABLES, GRAPHS AND ILLUSTRATIONS: IN APPENDICES

Appendix A0 – Data Overview

Appendix A1 – Study 1

Appendix A2 – Study 2

Appendix A3 – Study 3

CHAPTER 1: INTRODUCTION

During the mechanization era lasting from 1770-1870 (Anderson, 2001; Vertova, 1998; 2002), considered the first technological innovation paradigm, Karl Benz built and patented his Benz Patent Motorwagen. Having built three copies[1] of the same model petroleum-powered automobile[2] in 1885-6, this is widely considered to be the birth of mass-production in the automobile industry. An innovation paradigm serves as "carrier branch" for innovations to occur along (Anderson, 2001; Kuhn, 1962). By definition, the three core or primary systems an automobile needs is an engine, a drivetrain, and a steering mechanism anything else is considered a secondary system. A thick streak of perfectionism and a penchant for depression made Karl believe his invention should not be promoted until it was tinkered to perfection. Karl's moneyed wife Bertha Benz was of a different opinion as revealed when she slipped away unannounced one morning leaving a note on the kitchen counter[3] informing her husband she took the three-wheeled car (not road tested!) and her two teenage sons to visit her mother[4] some 60 miles away but will be back in a few days (Lienhard, n.d.)[5]. This trip served two other purposes – to reassure Karl his invention worked for distances greater than one kilometer, and as a live marketing stunt to attract further investors and buyers. Automobile historians point to this event as the beginnings of the automobile industry. Being the first road trip in a personal automobile, there was nothing resembling modern highways and zero gas stations along the way. Bertha had to stop along the drive at various pharmacies to pick

---

[1] One of these three automobiles is still running today.
[2] Prior to this, vehicles and extended transport were designed for multiple passengers in a steam-powered bus.
[3] She also sent a telegram upon reaching her mother's house the same day announcing their safe arrival.
[4] Bertha's family provided funding for Karl's research and prototypes.
[5] As a point of reference, a horse-driven carriage on good roads could cover 50 miles in 6-7 hours in the early 1800's (Austen & Shapard, 2004).

up some “fuel[6]” from the chemist. An innovator herself she had to use her hat pin to de-clog a fuel line in the carburetor along the way, used her garter to reinsulate a faulty wire cover[7], improved the braking system by having a cobbler apply leather to the brakes[8], and had a blacksmith repair the drive chain.

The automobile represents a complex system because it contains a series of artifacts or parts working together as a whole. Complexity can be measured in several different ways. First it can be measured in terms of the output characteristics (e.g. steering wheel, seating, engine, brakes), second in terms of where Karl sourced the various antecedent or contributing characteristics from (e.g. horse-drawn carriages, bicycles, and likely train technologies). Because knowledge has the tendency to “stick” in an area, some geographic locations become known for specific types of specialized knowledge. Therefore a third way in which complexity can be measured is in terms of the geographic location(s) from where those characteristics (e.g. bicycle manufacturers) were drawn upon or sourced from.

Continuing with our example of the automobile, further improvements were made to it during the following Chemical Engineering innovation paradigm lasting from 1870-1970 (Anderson, 2001), these additional chemical innovations were layered onto the existing mechanical engineering platform. Originally car tires were made from iron bars bent into a circle (this is the case in the Benz Patent Motorwagen). Looking to add comfort to the passenger’s ride, self-taught chemist Robert William Thomson stepped

[6] The fuel was a form of distilled petroleum called ligroin; also commonly used as a cleaning agent at the time.
[7] At the time, electrical wires were insulated with fabric instead of today’s zero-conductivity rubber.
[8] This is considered to be the first prototype brake pads.

into the picture and developed pneumatic rubber ties to replace the unforgiving iron bars[9] (Johnson, 2008). This tire was re-invented 40 years later for the same reason[10] by John Lloyd Dunlop ("History of the Passenger Tire," n.d.). These first tires were white as that is the natural color of rubber. Drivers found the tires did not have much durability, were prone to bursting, and looked dirty very quickly. Chemists continued to work on this and eventually determined that adding the inert chemical element carbon[11] would increase the distance traversed 100 fold, increased tensile[12] strength 1008%, and had the added bonus of changing the color of the tires to black thus hiding the dirt and grime of travel (Hiskey, 2011)[13]. This represents an increase in the complexity of automobile because it now includes mechanical engineering elements and chemical engineering elements that have been recombined into a single output.

Carrying our example of the automobile even further, during the current era of Electronic Engineering the automobiles are becoming wired and filled with sensors that report back various statistics to the on-board electronic control module – thus electronic innovations are layered onto the existing chemical and mechanical engineering technologies. Sticking with our tire example, during this era electronic engineers integrated sensors into the wheels which feedback information on the tire pressure, automatic braking system (ABS), and cruise control. This represents a further increase in complexity as the automobile now contains elements of mechanical engineering,

[9] This was not a commercial success; some believe this tire was ahead of its time.
[10] Dunlop invented the air-filled tire to ease his son's headaches from bicycling and claimed no knowledge of Robert William Thomson earlier invention.
[11] The company Binney & Smith approached the Goodrich Tire Company with this solution in the early 1900's (Hiskey, 2011). Binney & Smith is now known as Crayola Crayons.
[12] The amount of force needed to make an object burst or break.
[13] Modern vehicle tires with "white walls" are a vestige from the time when tires were all white. In the modern era the effect of white walls is superficial and purely cosmetic thus not a potential durability concern.

chemical engineering, and electrical engineering. This dissertation seeks to examine what is driving this increase in complexity and what aspects of complexity each of the three methods of measurement (output characteristics, contributing parent characteristics, and the location sourcing characteristics) best reveals.

Scholars have long studied innovation and the complexity thereof in knowledge building as a key source of competitive advantage and value creation (Fleming & Sorenson, 2001; Frenken, 2006; Nightingale, 1998; Trajtenberg, Henderson, and Jaffe, 1997; Celo, Nebus, & Wang, 2015). Innovation is a socially intensive process of recursive problem solving whereby functional answers are sought piecemeal through knowledge's amorphous and recipe-like nature from core and supporting technologies (Arthur, 2007; Nahapiet & Ghoshal, 1998; Nelson & Winter, 1982; Rugman & Verbeke, 2001). We define complex knowledge as that which relies upon rich interactions and interdependences and in which the configuration is of great importance (Baumann & Siggelkow, 2013; Ganco, 2015; Kauffman, 1993; Simon, 1962). The extant literature suggests the act of knowledge recombination is a mechanism through which novel knowledge may be created. In this process, knowledge grows in part through the blending of antecedent knowledge streams in novel forms through trial and error processes which may result in an artifact with greater complexity (Arthur, 2007; Hagarden, 1998; Olsson & Frey, 2002; Weitzman, 1998). The recombination literature therefore provides a useful lens for examining complex knowledge building and the structure of it as this framework accounts for both the characteristics and historical development of knowledge building (Fleming, 2001; Olsson & Frey, 2002; Weitzman, 1996; 1998). Extant research emphasizes global value creation by connecting distant knowledge sources for innovation (Antonelli, Krafft, & Quatraro, 2010; Cano-Kollmann, Cantwell, Hannigan, Mudambi, & Song, 2016; Cantwell & Noonan, 2004; Fleming & Sorenson, 2001; Trajtenberg et al., 1997; Vagnani, 2012; Yayavasram & Chen, 2015). Knowledge is a complex system (Ganco, 2013, 2015; Simon, 1962), we focus on the mechanisms forging distant connections acting on that system as a means of value creation for global firms

Traditionally, the complexity of distant knowledge recombinations has been studied within a single industry or knowledge field (Fleming & Sorenson, 2001; Ganco, 2013; Kaplan & Vakili, 2015; Vagnani, 2012). Studies find technologically distant knowledge recombinations can produce complex, value-creating innovations (Fleming & Sorenson, 2001; Kaplan & Vakili, 2015). However, research shows specialized fields of knowledge and industries cluster in specific geographic locations (Marshall, 1920; Saxenian, 1994; Tallman, Jenkins, Henry, & Pinch, 2004). For example, San Francisco is known as Silicon Valley in the USA, London is known for its financial center, and the Port wine cluster has a long-established presence in Portugal. Congregating thusly often brings these firms locational advantages, access to knowledge spillovers, a targeted labor pool, with complementary services and suppliers nearby (Krugman, 1991; Porter, 1990). Taking a practical approach, reaching across technology fields to recombine knowledge suggests a simultaneous reaching to distinct geographic clusters and therefore a systemic increase in the complexity of the technological knowledge system. This is why we must examine both technology field and geographic location together.

In the present context of globalization, the singular use of a technological knowledge field or industry to measure knowledge complexity in the knowledge recombination literature does not directly address the additional geographic complexity element contributed from traversing physical distances to achieve said recombination. It is important to jointly understand the contributions of these two knowledge building inputs through which global innovation connections are made to encourage the facilitation of value creation and capture likelihood by firms.

This dissertation sets out to outline and assess the properties of complex knowledge through the joint consideration of technology fields and geographic locations as two knowledge building inputs during periods of globalization. To do so, we propose to shift the approach so as to examine the complexity of novel knowledge artifacts when they are recombined across both these conditions: technology field and geographic location. A goal of this research is to establish the uniqueness of the three measures of complexity developed here – knowledge artifact complexity, knowledge sourcing complexity, and location sourcing complexity – all of which are built using the information on a given focal patent with the intention of revealing that firms access different types of knowledge expertise and hence different types of distributed knowledge systems. In this approach we expand the parameters to envelop all possible technological knowledge fields without industry constraints; likewise we do not limit the potential geographic locations. It is reasonable to expect bridging technology field disciplines normally implies that a physical distance must also have been traversed in some way, given that locations are specialized in their activity. The implications for this work extends to firms seeking to appropriate rents from innovative activities – particularly breakthrough innovations, public policy makers to encourage firms to locate within a relevant cluster, and influential managers deciding how to facilitate and support the process of novel knowledge development.

The dissertation is organized as follows. In constructing our position, we first suggest that which facilitates knowledge complexity in global innovation and examine two proposed determinants of knowledge complexity. Study one examines how technological knowledge is changing in complexity across technology fields while comparing and contrasting two alternative measures of this phenomenon with one being novel to the literature. Study two then builds on this base by layering location complexity onto the model and is another novel measure contributed to the literature. Study three is a case study of the joint knowledge complexity and locational complexity relationship from the prior study, exploring when this joint consideration results in outlying behavior.

CHAPTER 2: LITERATURE REVIEW

2.0 Fragmentation of Global Value Chains

Fragmentation of production systems suggests both labor and geography are fragmented between firms and across geographic space (Gereffi, Humphrey, Sturgeon, 2005; Ietto-Gillies, 2014). The fragmentation of production systems has shown a steady increase in the trade of components and services and away from the trade of final goods (Grossman & Rossi-Hansberg, 2008; Johnson & Noguera, 2012; Schmitt, & Van Biesebroeck, 2013; Sturgeon, Van Biesebroeck, & Gereffi, 2008) highlighting changes in the global value chains. Global value chains (GVCs) (Gereffi, 1994) and Global Production Networks (GPNs) (Yeung & Coe, 2014) are composed of a global lead firm which capitalizes on the high rent activities (e.g. innovation, R&D, marketing, branding) while outsourcing and offshoring the low return functions (e.g. manufacturing, repair) (Azmeh & Nadvi, 2014; Gereffi, 1999; Gereffi & Korzeniewicz, 1994). In doing so, these global leader firms are becoming more assembly-oriented when expanding the niches for labor-intensive activities. This international fragmentation of labor, a prominent trend beginning in the 1970's, permits producers in different countries and likely with diverse ownership structures to form systems of production for subparts and components (Arndt & Kierzkowski, 2001; Gereffi, 2005). Filling that niche is considered a success by local producers but is demanding as these suppliers must meet the established demands of the lead firms in terms of price, delivery times, and compliance with labor, environment, and quality standards (Kaplinsky, 2005; Nadvi, 2008). While the global economy is becoming more integrated, production systems have become more disintegrated

(Feenstra, 1998) which implies and has led to the international trade in components and services to grow in proportion (Yeats, 2001; Hummels, Rapoport, & Yi, 1998).

This fragmentation in production systems implies products are becoming more complex. We can think of fragmentation (or simplification) and complexity as being the two extremes of a spectrum. Movement along this spectrum with regards to the flow for a related collection of knowledge artifacts may be mediated by the degree of modularization. A complex system may exhibit modularity if the components can be designed and innovated upon independently but are compatible and can work together in support of a unified whole (Baldwin & Clark, 2000; 2006). Modularity is built upon the premise of a hierarchy of primary and supporting technologies and the near-decomposability such that modular product architecture is one where the product taken as a whole can be decomposed into unique and stable subparts (Baldwin & Clark, 2000; Ethiraj & Levinthal, 2008; Henderson & Clark, 1990; Parnas, 1972; Simon, 1962). The basic design principle of modularity is to encapsulate interdependencies into stand-alone units (thus fragmenting it) and to minimize any reciprocal interdependencies between the modularized units of a whole (thus illustrating the hierarchical nature). As the size of the system increases, the nonlinearity and nonadditive nature of complexity is revealed as the interdependencies disrupt the design process growing in scale faster than proportionate. Modularization bounds the design complexity of the whole product while encouraging incremental and localized innovation. Modularity permits an architectural structure to achieve economies of scale in a global market by drawing on production capabilities external to the traditional boundaries of the firm (Langlois 2007, Sturgeon 2002).

Advances in ICT, an increasing sophistication in logistic capabilities, and greater flows through global economic openness is enabling this modularization or “fine-slicing” of various firm activities which also encourages firms to relocate aspects of the value chain to explore diverse knowledge streams, exploit more fruitful foreign markets, and access efficient lower cost locations (Andersson & Pedersen, 2010; Buckley, 2009, 2011; Contractor, Kumar, Kundu, & Pederson, 2010; Dunning, 1993). At a global scale, fine-slicing is a function of innovation which disaggregates value chain activities into subsystems (Andersson & Pedersen, 2010). These constituent pieces can be allocated “offshore” for geographic fine-slicing and “outsourced” for organizational fine-slicing (Contractor, et al., 2010). Firms can appropriate innovation rents by allowing external cooperating and competing actor’s access to their technology particularly in globalized and more spatially dispersed environments; in doing so modularization is encouraging both a vertical and horizontal disaggregation of the production system (Langlois & Robertson, 1992; Rezk, Srai, & Williamson, 2016; Sturgeon, 2002; Contractor, et al., 2010) and in the process we expect a gradual dispersion of the knowledge base.

Kodama illustrates how innovation in many high-tech industries comes from a process of technology fusion (1986; 1995) or convergence (2014) where hybrid technologies are developed by recombining existing disparate technologies. Thus, technological diversification is increasingly stemming from the growing interrelatedness of formerly unconnected technologies (Cantwell & Piscitello, 2000). The uptake of information and communications technology (ICT) as an innovation paradigm has proven to be a prime example of this type of fusing technology which is connecting the historically unconnected technologies. At the same time, ICT has also lowered the cost of communication and coordination of activities in disaggregated and dispersed R&D settings thus also connecting historically difficult to connect geographic locations (Andersson & Pedersen, 2010; Gooris & Peeters, 2016; Larsen & Pedersen, 2009; Roberts, 2000). This two-pronged connecting property of ICT has encouraged advances in ICT to enable firms to fragment their various systems to decouple, disperse, and shift towards a globalized production configuration (Andersson, Forsgren, & Holm, 2007; Azmeh & Nadvi, 2014; Rezk, Srai, & Williamson, 2016; Yamin, & Sinkovics, 2010). Multi-technology systems may be brought together in part by ICT where these firms may choose to locate in centers of excellence for industries other than their own for the development of fields outside their primary interest in the ICT sectors. The firm's attraction to different and locationally distinct potential inputs may be driven by the complexity of recombination necessary for high-tech and cutting-edge technology development which is associated with risk, uncertainty, high R&D costs, the need to cooperate with other firms in the area, all the while increasing the flow of knowledge within MNEs across political boundaries (Cantwell & Santangelo, 2002).

The multi-technology firm (Granstrand & Oskarsson, 1994; Granstrand, Patel, & Pavitt, 1997; Patel & Pavitt, 1997) thus has developed out of technology characteristics becoming increasingly complex in nature as the characteristics exhibit increasing interrelatedness and a degree of fusion from technologies that were historically separate (Kodama, 1992). This model permits the firm to adapt to changing techno-socioeconomic conditions characterized by innovation opportunities stemming from technological diversification. As such the range of technologies the firm must be familiar with is wider than its product offerings, reflecting the conditions of the current innovation paradigm where the development of a wide range of technological capabilities is essential to exploit prospective technology recombinations thus facilitating fruitful output. The core reason behind this is the increasing number of technologies drawn upon during the production of a single product.

While the environment is rapidly changing exhibiting increasing technological relatedness and growing knowledge complexity, firms increasingly draw on a more diversified knowledge base relative to their respective extent of product or market diversification by engaging in technologically-motivated inter-firm alliances and locating in clusters of complementary skills (Cantwell, 2008; Madhok & Phene, 2003). Essentially multi-technology firms can be thought of as those taking advantage of the increasing technological relatedness.

When MNEs locate some part of their R&D or production in a complementary or alternative center of innovation from its industry the firm is able to gain access to novel and potentially useful avenues for development which, given time, it may integrate into its exiting activities and lines. Modern multi-technology firms inherently require a broader diversity of technological expertise (than those historically c.f. Cantwell & Fai, 1999) in order to produce its given product line, this provides a strong incentive to capitalize on centers of innovation for the different fields of expertise in which they are located (Cantwell, 2008; Brusoni, Prencipe, Pavitt, 2001). Bayer provides an excellent example of a multi-technology firm. Its primary output field is in chemicals but it is also actively sourcing in mechanical engineering, information technology, instruments, and so forth from locations revealing those respective innovation strengths.

2.1 Facilitators of Knowledge Complexity

Globalization has enabled firms to access and source distant knowledge with comparatively greater ease than yesteryear. Specialized forms of knowledge congregate in specific areas (Marshall, 1920; Porter, 1990). Knowledge has a degree of tacitness to it that helps it to stick in an area (Nightingale, 1998; Searle, 1995). Firms initially search for new knowledge locally, then search more distantly forging connections when the sought knowledge is not found in the local area (Ethiraj & Levinthal, 2004; Fleming, 2001; Gavetti & Levinthal, 2000). To innovate using this specialized knowledge, a firm must travel to and interact with the target location. Firms may interact by recombining its extant knowledge with that of the specialized location to form knowledge artifacts or patents. This act of recombination is twofold, disparate technology fields are being joined as is the tacit knowledge from each location, with the knowledge artifact likely becoming more complex. To explain this further, we first focus on the factors facilitating knowledge complexity – globalization, the clustering of knowledge, innovation, and recombination; then we expand upon that which we propose to develop through distant recombinations – technology field and geographic location.

2.1.1 Globalization

Through globalization, the world has become more interdependent, interconnected, and yet spread out (Berry, Guillén, Hendi, 2014; Fernandes, 2011; Goerzen, Asmussen, Nielsen, 2013). In doing so, globalization has been playing a critical role in the development of a more complex structure for technological knowledge building. Above and beyond individual firms engaging in knowledge sourcing behaviors, there is a systematic trend towards globalization as both general business and the world is becoming interconnected on a daily basis making knowledge based connections are more common.

In searching for novel innovation solutions, it becomes more likely that firms explore outside their geographic area and so firms extend outside the immediate vicinity. International businesses are particularly well positioned to be sensitive to the distinct locations of specific knowledge as they have established international knowledge-sourcing patterns in place. MNC units are more likely to seek advice with partner units elsewhere. Partner units can be in contact with other multinational firms, diaspora, international research centers, and university programs (Lorenzen & Mudambi, 2013; Thomas, 2016; Yusuf, 2008).

2.1.2 Innovation

Innovation outputs are sought as a source of future returns for firms around the globe. Innovation is a social practice of recursive problem-solving as useful solutions are sought through piece-meal knowledge recombinations from a hierarchy of core and supporting fields (Arthur, 2007; Henderson & Clark, 1990; Nahapiet & Ghoshal, 1998; Schumpeter, 1950; Simon, 1962). Innovative solutions can be sought and applied along the length and breadth of the value chain. Innovating firms engage in complex problem solving, often requiring novel combinations of knowledge. Because of its amorphous nature (Rugman & Verbeke, 2001), knowledge can be partially used and incorporated in knowledge building activities in addition to existing knowledge becoming relevant to the innovation activities and processes undertaken by the firm. Often times it is a variety of subproblems in the supporting technologies, more than the focal problem of the primary core technologies where most of the work needs to be done for the envisioned idea to function (Arthur, 2007; Simon, 1962). Solving this hierarchy of problems suggests the firm has to tap into various forms of specialized knowledge within and beyond its own traditional borders in the system (Rothaermel & Alexandre, 2009; Weigelt & Miller, 2013) and potentially from innovation patterns in different knowledge fields.

Firms operating in a single industry may co-locate in a geographically proximate area, which may over time development into a cluster reflecting the local knowledge specialization (Krugman, 1991; Porter, 1998; Shaver, 1998). Agglomeration economies may emerge when related firms specialize in various factor inputs of intermediate product or process supply (Marshall, 1920; Porter, 1990). Learning the uncodified aspects of knowledge occurs through informal ties effectually encouraging the confinement of knowledge to the area (Miller, Zhao, & Calantone, 2006). Operating in these zones may augment competitive advantages of the clustering firms. However partially because of this local specialization, no one geographic area can have the entire range of expertise that comes to be needed.

Commonly search is localized until a point at which it is determined a functional answer is not found, then search extends more distantly to find potential solutions for problem solving (Vernon, 1966; Nelson & Winter, 1982; March 1991). Knowledge applications that are derived from incremental and localized search are easier to find and therefore associated with more competition, and represent less of a competitive advantage (Ethiraj & Levinthal, 2004; Fleming, 2001; Gavetti & Levinthal, 2000). They are in and of themselves less likely to be radical because it is an incremental extension.

Successfully recombined radical knowledge may be harder to develop because search occurs over a diverse assortment of distant fields. This novelty however represents a greater competitive advantage and the breakthrough knowledge is associated with less immediate competition.

2.1.3 Recombination

Knowledge recombination, the process of trial and error whereby antecedent knowledge is combined with other antecedent knowledge, can lead to new knowledge generation as a form of innovation (Arthur, 2007; Olsson, 2000; Olsson & Frey, 2002; Weitzman, 1996, 1998). Some literature on knowledge complexity (c.f. Arthur, 2007; Fleming, 2001; Frenken, 2006; Sorenson, Rivkin, & Fleming, 2006, Ganco, 2015) suggests that more complex knowledge development is supported by an even larger knowledge base (that is eventually used), which is developed through a series of trial attempts. This suggests that recombining selected components of knowledge in some novel fashion tends to be associated with an increase in the complexity of that knowledge. It is suggested that while a wide variety of forms of recombinant knowledge may potentially be envisaged (Weitzman, 1996; 1998), what becomes critical is the process through which specific combinations can be realized and brought to fruition within a reasonable time horizon. This combinatorial process is generally designed to address problem-solving concerns in ways that are achievable and deliverable. Workable solutions may often depend upon having access to a very large body of supporting or complementary functional knowledge. It has been argued that when particularly distant articles of knowledge are brought together, the outcome may be both numerous qualified recombinatorial failures and a select few novel innovations (Arthur, 2007; Fleming, 2001; Weitzman, 1998).

As time progresses, natural limitations to recombination emerge. The knowledge recombination process may become increasingly expensive exhibiting increasing costs, diminishing returns to creativity (Olsson, 2000), decreasing growth rates, technical imbalances (Rosenberg, 1976), intra-context friction (Weitzman, 1998), and reverse salients (Hughes, 1987) may emerge. The scope of innovation can become confined which encourages the discovery and development of alternative routes to achieving functional answers to problem-solving. As innovation patterns becomes naturally hindered, at some point less expensive alternatives emerge and become an emergent new innovation pattern (Dosi, 1982; Freeman, 1991; Kuhn, 1962; Olsson, 2000; Von Tunzelmann, Malerba, Nightingale, & Metcalfe, 2008).

These shifts are identified when the distance between knowledge fields are brought closer together over time through mutual learning via problem-solving. For example, the intellectual distance between the chemistry field and the biology field was shortened by the establishment of the intermediary field of chemical biology which share antecedent knowledge of both chemistry and biology (Schreiber & Nicolaou, 1994a, 1994b). In developing such intermediary fields, overtime there is less distinction and separation between the fields of study. This suggests an increase in complexity as the knowledge field crossovers are jointly creating recombined knowledge and are exhibiting greater cross-field associations.

Even though reliability decreases and uncertainty increases during recombination in periods of innovation pattern shifting (Grant, 1996; Katila & Ahuja, 2002; Takeishi, 2002), more distant combinations may also be associated with a greater scope for novelty (Ahuja & Lampert, 2001; Guilford, 1967). Here, technological change arises from the distinctive knowledge recombination approaches inherently available to each firm. During periods of innovation pattern change, new routines are sought through the act of problem-solving and knowledge field borders are crossed. Prior innovation patterns are commonly partially integrated into the following predominant pattern when alternatives emerge from various problem-solving recombination efforts.

In the current innovation era guided by ICT (Andersen, 2001; Kodama, 1992; Kuhn, 1962), it itself makes connections between areas of knowledge that were previously quite separate from one another and is a classic example of an effect of globalization. ICT has encouraged the rapid codification of knowledge and facilitated its diffusion across knowledge fields and across geographic space. Global value chains have encouraged the development of ICT applications relevant to the specific specialization of activity in each location (Kumar, 2001; Chopra & Meindl, 2012).

2.1.4 Paradigm Change

A paradigm, conceptualized as a historical epoch, represents an established path of knowledge accumulation that has become familiar to firms in a given industry or context (Anderson, 2001; Dosi, 1982; Kuhn, 1962). If there is an established paradigm for knowledge search and recombination in a given context or industry, such that innovation becomes more path-dependent, then new knowledge recombinations can then be developed through a relatively greater reliance upon established sourcing methods.

It is suggested that while a wide variety of forms of recombinant knowledge may potentially be envisaged (Weitzman, 1998), what becomes critical is the process through which specific combinations can be realized and brought to fruition within a reasonable time horizon. This combinatorial process is generally designed to address problem-solving concerns in ways that are achievable and deliverable. Workable solutions may often depend upon having access to a very large body of supporting or complementary functional knowledge (Brusoni, Prencipe, Pavitt, 2001). This leads us to suggest that for an enterprise to access complementary knowledge outside its core paradigm, a search is conducted over a wider array of established paradigms than those in which the firm has some significant prior experience. Even though reliability decreases and uncertainty increases during recombination across paradigms, more distant combinations may also be associated with a greater scope for novelty.

A paradigmatic shift is identified when the distance between either clusters or closely related fields are brought closer together over time through mutual learning via problem-solving (Anderson, 2001; Kuhn, 1962). If one were to take a cross-section of a paradigm before and after a shift, there would be less distinction and separation between the fields of study afterwards. This suggests an increase in complexity as the clusters crossovers are jointly creating recombined knowledge and are exhibiting greater cross-field associations. During periods of paradigmatic change, new innovation routines are sought through the act of problem-solving and paradigm borders are crossed. The search for problem-solving knowledge in a different field or cluster than a focus firm's core area(s) may be sought by increasing the firm's cross-field associations. These associations may initially be made peripherally. These cross-field associations, we suggest can begin more openly, particularly during nebulous paradigmatic change. Attempting to borrow from and blend two disparate paradigms suggests an unstable linking of both knowledge and the network, therefore suggesting an increase in the complexity of both relative to pre-paradigmatic change contexts. As a new predominant paradigm is attempting to establish, formally locking into an unknown and untested direction for search and development is inappropriate. During these periods of unstable paradigmatic shifting the broadening of cross-field associations in an attempt to link distinct technology fields may be done via the use of informal and indirect ties to complement the continued existence of formal ties for intra-paradigm development. Compared to a formal tie indicated when firms have contractually outlined obligations to one another (see Lincoln, 1982, a review), an informal tie is identified when firms exchange reciprocally and trust emerges over time without legally-based obligations (Arthur, 2007; Baldwin & von Hippel, 2011; Olsson, 2000; Pavitt, 2002). Firms can also source knowledge indirectly from non-immediate and community-based pools of which the firm is an attendant.

It has been suggested that these informal and indirect networks critically enhance the paradigm transition process (Arthur, 2007) by serving several purposes. They help identify good ideas by providing early access to a spectrum of potentially useful knowledge fragments (Winter, 1984), offer diverse and contradictory knowledge (Burt, 2004), and alerting the actor to previous efforts (Arthur, 2007). Rather the mutual exchange of knowledge between two parties informally can assist in trial and error search for functional answers to problems. Opportunities engendered by regular informal and indirect exchanges could develop into a growing openness in the collaborative knowledge-seeking networks of firms that are engaged in this process (Brusoni, Prencipe, & Pavitt, 2001; Grigoriou & Rothaermel, 2014; Hobday, Davies, & Prencipe, 2005; Langlois, 2003). This broadening of the firm's network alters the composition of the knowledge network and its configuration while new routines and patterns for innovation are being developed and complexity is increasing.

2.2 Determinants of Knowledge Complexity

A complex system, one such as knowledge, is one that cannot be easily broken down into the contributing building blocks because each piece is expected to interact in a nonadditive and nonlinear manner (Ganco, 2015; Simon, 1962). As such, we focus on one aspect of the contributors to the complexity of knowledge – that of distance. We expect the recombinations of distant knowledge to lead to more complex knowledge artifacts. We begin by assessing distance in terms of theoretical space, then in terms of geographic space.

The expected complexity level of local and distant search contains a debate. Some authors indicate local search can foster incremental complexity growth (Taylor & Greve, 2006; Weisberg, 1999) while others expect high complexity levels to emerge in distant search (Ahuja & Lampert, 2001; Guilford, 1967; Hargadon & Sutton, 1997). We follow with the expectation that local search produces more incremental development and distant search expecting to produce more breakthroughs. We assess knowledge complexity by investigating the underlying structure of knowledge building inputs – i.e. the pattern of knowledge artifact development in theoretical and physical space. We do so by investigating the architecture of each knowledge artifact – by that we mean the pattern of knowledge domains used and the originality of that pattern, for both the theoretical space attributes and physical space contributions of each knowledge artifact in the entire structure.

2.2.1 Theoretical Space: Technology Field Distance

Novel knowledge may be considered distant when the antecedent knowledge drawn upon stems from distinct and unconnected technology fields (Antonelli, Krafft, & Quatraro, 2010; Cano-Kollmann, Cantwell, Hannigan, Mudambi, & Song, 2016; Fleming & Sorenson, 2001; Kodama, 1992; Trajtenberg et al., 1997). Here knowledge complexity has risen from innovators recombining the available core knowledge field with supporting, peripheral, or unconnected knowledge fields. Knowledge may become more complex through the act of innovative problem-solving when its recombined antecedents were dispersed across distinct technology knowledge fields. With this progression of integration, knowledge has experienced a growing level of complexity. An example of this would be the knowledge recombination of photography equipment with medical equipment resulting in endoscopic cameras.

2.2.2 Physical Space: Geographic Distance

The systematic trend towards globalization and its epistemic communities are making knowledge based connections are more common. The world is becoming more interconnected with knowledge based connections becoming more common as people travel and move more. For example, people move more and bring their indigenous knowledge with them to the new geographic location and typically exhibit greater knowledge recombinatorial abilities in innovation (Bäker, 2015; Franzoni, Scellato, & Stephan, 2014; Scellato, Franzoni, & Stephan, 2015).

Accessing a different knowledge domain from the firm's core knowledge field(s), suggests the intentional establishment of cross-field associations. Put simply, increasing complexity may occur when antecedent knowledge is sourced from distinct geographic locations for recombination. This in turn suggests geographically, and likely internationally, dispersed connections are inherently necessary for cross-field associations to be purposefully established for recombining available knowledge to a greater degree of complexity in the search for solving problems in innovation along the value chain.

CHAPTER 3: DATA

The primary research question places several demands on the data. Initially, the data must cover a broad range of technologies. Secondly, it must track both the antecedent contributions in recombination as well as the final characteristics of the artifact. Thirdly, the data must have a long time horizon so as to be able to distinguish original from common recombination patterns. In order to satisfy these demands, I analyze every knowledge artifact in selected fields of origin in the global population of granted USPTO (United States Patent and Trademark Office) patents between 1976 to 2014 (n = 1,340,799). A patent secures exclusive rights to the inventor(s) from unauthorized usage of the knowledge artifact for a given length of time. By USPTO definition, a granted patent is an original contribution whereby it must be novel and nonobvious.

Patents are rich with standardized information resulting in an attractive data source for researchers. Any given patent has a title, brief abstract, a complete description of the knowledge artifact so as to provide an individual record of the knowledge frontier, along with the application and grant dates. The legal owner of the patent, termed assignee, and contributing inventor(s) are listed by name along with the city and country of residence. This information may be used as a basis for investigations into geographic considerations. Each patent is assigned a minimum of one technology class, most have multiple technology class assignments, used to categorically indicate the genre of the knowledge artifact. Adding to the allure of patent data research, all prior granted patents are reclassified when the USPTO office determines a new technology class is warranted, ensuring a historically consistent classification scheme back to the first patent in 1790.

These patent technology classes indicate the characteristics of the knowledge artifact. Patents include citations to antecedent patents and appropriate scientific publications facilitating a method to trace the parental roots of the resulting technological knowledge development. All patents cite at least one antecedent patent and many cite several antecedent patents. Computerized access to this data is publically available from 1963 to the present and from 1975 onward for patent citations. This database contains systemized and detailed information on long-term innovation patterns, consequently it is able to support micro-, mezo-, and macro-level analysis.

Despite being flush with information, so long as there has been research using patents there has been a tandem debate regarding the appropriateness from large-scale economic patent data researchers (c.f. Scherer, 1965; Schmookler, 1966). Patent data may be an imperfect source when commonly voiced concerns acknowledge not all innovations are patented or patentable, patenting propensity varies across nations, industries, and firms and the relationship between firm size and innovativeness has been questioned (Levin, Klevorick, Nelson, & Winter, 1987; Pavitt, 1988; Griliches, 1988; Kleinknect & Reinders, 2012). Those recommending the use of Research and Development (R&D) instead are presented with Pavitt's 1988 work which denotes how this too is a biased innovation measure for similar reasons of the relative importance of measured R&D, variations across technologies and sectors, and unaccounted informal R&D occurring outside the established purview. Mansfield's 1986 survey showed when patenting is unimportant a firm still applies for 66% of all their patentable inventions; where patenting is important that ratio rises to 84% (Cantwell, 2006). Assuaging one limitation, there are several established methods for handling proposed patenting propensity data limitations. A common method, known for its power, is the construction and normalization of indices for patent propensity variations. This method is represented as a revealed technology advantage (RTA) index (Soete, 1981), the index of internationalization (Cantwell, 1995), or the corporate technological competitiveness index (Cantwell & Sanna-Randaccio, 1993). Studies have found large firms are more inclined to patent regardless of immediate usability over smaller firms (Acs & Audretsch, 1989; Cantwell, 2006; Pavitt, 1988). Patents have also been shown to correlate with additional measures of technological knowledge activity and innovation performance (Pavitt, Robson, & Townsend, 1987). Altogether with an awareness of propensities and multiple methods for managing debates, patents serve as a robust proxy for innovation, while better so for large over small firms (Cantwell, 2006; Griliches, 1990; Basberg, 1987; Acs & Audretsch, 1989).

3.1 Data Descriptives

The data covers the period 1976-2014, it has been divided into three equal time periods of 13 years which has been used to demonstrate its ability to answer the dissertation questions. The first time period is 1976 – 1988 and contains 158,426 unique patents, the second time period is 1989 – 2001 and contains 383,308 patents, the third time period is 2002 – 2014 and contains 799,720 patents. For the purposes of demonstrating the ability of the data to answer the dissertation questions in the proposal, the patents from eight Tech56 fields, with two from each CEMTO industry field, were chosen for interesting cross-field variation. The CEMTO industry fields are a high level of patent data aggregation representing Chemical, Electrical, Mechanical, Transport, and

Other. The Other sector represents Tech56 fields that do not fit easily into one of the main four categories; only four Tech56 fields fall into this category. If a pairing were to link two of the CEMT fields (e.g. C and E), then this represents crossing broad macro area bounds and thus represents bridging a large and unusual distance. The primary subclass of the patent indicates which technology field or industry sector it is categorized into. Tech56 fields 8 and 12 were selected to indicate that which pertains to chemicals and pharmaceuticals (Chemical), Tech56 fields 16 and 29 indicate mechanical engineering (Mechanical), Tech56 fields 40 and 41 indicate information and communications technologies (Electrical), and Tech56 fields 42 and 43 represent transportation equipment (Transport). These eight fields total 1,340,799 patents of the more than 6 million patents of the final database, or approximately 20%. To be clear, complexity is calculated for every patent without limitation on what field or sector it is identified as, the aggregating schemes are only used to make the results more understandable by collecting all related patents into a common framework.

There are many other methods that can be used to show a change in knowledge complexity or location complexity. The basic model of the NK complexity measure is how it benchmarks against the observed likelihood of coincidence of a pair of subclasses. With regards to calculating knowledge complexity, the relatedness of activities can be used to show if a pairing is within or across technology fields; with regards to location complexity, the relatedness of locations and the distance in miles between the capital cities of referencing and referenced locations are commonly used. As study 1 and study 2 are exploratory in nature, it is important to include here how an established model may be used to depict complexity in addition to the adapted NK models used here. The relatedness between any two technologies or locations can be calculated by both simple and sophisticated means. In the simple calculation (Technological Diversification measure) the objective is to determine if the paired technologies indicate an intra-CEMTO or inter-CEMTO pairing (c.f. Cantwell & Zhang, 2011). In the sophisticated calculation (Technological Diversification measure) the pattern of patenting activity is used to determine the perception of complementarity between pairs of technologies (c.f. Cantwell & Noonan, 2004). As the NK method of complexity used here is uncommon in the international business literature, I have also included the simple method that clearly indicates if a pairing crosses the broad macro CEMTO areas or Continents as a way to corroborate the expected increase in complexity as measured by the NK model.

3.1.1 Cross-Classifications or Knowledge Artifact Complexity (KAC)

Complexity as measured by cross-classifications subclasses is the first of three methods we will investigate to examine the changing levels of complexity over time. Some patents only list a single subclass and while it is believed the knowledge artifact was developed through a process of recombination the measure cannot accommodate this therefore these patents have been controlled for and represent 10% of the data (Table A1). Overall the number of unique primary and secondary subclass codes has increased over time for both the primary and secondary subclasses, this casually suggests complexity may be increasing.

Table A2 shows a simple model of complexity where the pairs of primary to secondary subclasses as aggregated to CEMTO levels. The pairs are restricted to that of the primary CEMT field. Overall there appears to be an increase in intraCEMTO pairings which suggests a decrease in complexity. However when we look closer, CC and MM both decrease in intraCEMTO pairings thus suggesting an increase in complexity, EE and TT both increase in intra CEMTO pairings thus suggesting a decrease in complexity.

The calculation for complexity by subclasses is made by dividing the number of subclasses on a given patent by a weight. The number of subclasses is a simple count of all the subclasses on a given patent. The weight is calculated by dividing the count of subclasses previously recombined with primary subclass *i* by the total count of patents that reference subclass *i*. The weight is cumulative over the three periods thus encompassing the entire database by the final time period. More explicitly, for a given primary subclass, step 2 counts the number of unique subclasses it has been paired with and step 3 counts all of the patents that have a listing for subclass *i*. Table 1 (below) shows the three stages to the complexity calculation.

As an example of how to read the chart, in time period 1976-1988, any given citing patent in Tech56 field 8 has an average of 5.890 subclasses listed. For any given subclass in Tech56 field 8, the given subclass has been also been observed as paired with an average of 49.452 other unique subclasses. Next we count all of the unique patents in the prior and given time period(s) (in the case of first period – we only have the given period) that reference the given subclass, for Tech56 file 8 that is an average of 37.458 patents. If we were to then compute (5.890 / (49.452 / 37.458)), we would arrive at the average complexity value for any given Tech56 field 8 patent, 4.500.

Table 1: The three stages of the subclass complexity calculation, and results

Average	1976-1988 tech56	average	1989-2001 tech56	average	2002-2014 tech56	average
Step 1: Number of subclasses on a given patent	8	5.889881	8	6.108823	8	7.065013
	12	5.720942	12	6.802399	12	5.724588
	16	4.605391	16	5.2521	16	5.743612
	29	3.789493	29	3.824852	29	4.003815
	40	4.516428	40	4.486499	40	4.908166
	41	3.379128	41	3.966817	41	4.081229
	42	3.344211	42	2.873669	42	3.178002
	43	3.455925	43	3.520608	43	3.339278
Step 2: Count of subclasses previously recombined with primary subclass *i*	8	49.5625	8	231.289215	8	609.408896
	12	230.141525	12	622.911390	12	1104.00403
	16	75.2690887	16	193.619635	16	355.199530
	29	93.2691955	29	203.555056	29	330.455760
	40	105.21595	40	153.320791	40	474.900312
	41	117.475074	41	277.110248	41	533.980389
	42	113.214683	42	236.456167	42	512.586979
	43	76.154899	43	189.788783	43	420.631572
Step 3: Total count of patents that reference subclass *i*.	8	37.4583333	8	222.404902	8	574.542343
	12	320.881655	12	1709.60797	12	4202.95545
	16	79.2915040	16	272.894853	16	570.842153
	29	144.013793	29	345.435757	29	608.674384
	40	290.735189	40	382.645704	40	1427.95905
	41	245.949002	41	898.410794	41	2548.17969
	42	302.029700	42	768.775829	42	2082.81248
	43	81.0744330	43	284.873838	43	966.644240

Subclass Complexity Results

	1976-1988 tech56	complexity(avg)	1989-2001 tech56	complexity(avg)	2002-2014 tech56	complexity(avg)
Subclass Complexity	8	4.500946	8	5.495304	8	6.538764
	12	6.623147	12	15.0358	12	22.9347
	16	4.333273	16	6.468287	16	7.274913
	29	4.757893	29	5.335217	29	5.881689
	40	10.89805	40	9.398081	40	13.68053
	41	6.25737	41	10.62334	41	15.42614
	42	7.880091	42	7.373808	42	9.804443
	43	3.570292	43	5.138704	43	6.972797

Co-classification Results and Conclusions:

When calculating complexity via subclasses, we see complexity is increasing, Table 1. Tech56 fields 40 and 42 both decrease slightly but all eight fields rise to level above their respective starting points in the final period. Overall the table shows how all the Tech56 fields are rising in complexity using the subclass measure.

When we examine the slow overall trend of increasing intra-CEMTO pairings and decreasing inter-CEMTO pairings (Table A2) we may anticipate a slow change such that complexity is decreasing over time according to the subclass model. Or more specifically, we expect CC and MM to show increasing complexity by not EE and TT. However upon examining subclass complexity over the time period (Table A3), we see that in fact all Tech56 fields are increasing in complexity. Why does this divergence occur? It may be that there is spurious diversity in the subclass classification scheme because there are so many subclasses to choose from and may be similar in nature. Thus, the simple model may be showing the unsatisfactory movement of subclasses and thus a lot of spurious noise. Relatedly, it may the CEMTO level is too broad for the purpose of examining complexity at this micro level, as was also found to be the case for subclasses in Cantwell & Zhang (2011). It may also be that inter-Tech56 field pairings are occurring within the CEMTO fields but this level of aggregation is too high to adequately demonstrate.

3.1.2 Cross-classification or Knowledge Sourcing Complexity (KSC)

Complexity as measured by patent co-classification is the second of three methods we will investigate to examine the changing levels of complexity over time. Overall the number of unique citing (the focal patent) and cited (the antecedent patents of the focal patent) citations is increasing over time (Table A3), this again casually suggests complexity may be increasing.

Table A5 outlines the percentage of citing and cited fields by Tech56 field and by year. In the 1976-1988 timeframe, Tech56 field 8 cites a total of 35 unique citing subclasses. Tech56 field 8 also references 403 cited subclasses from across the 56 Tech56 fields. Overall, the eight Tech56 fields cite all 56 fields.

Table A6 shows the simple measure of complexity where the percentage of pairs as a citing patent is paired with a cited patent is aggregated to the CEMTO level. Of note is the overall 5.67 percentage point increase of interCEMTO pairings suggesting knowledge is growing in complexity over time. This is quite different from Appendix 0 Table 2 where in cross-classifications aggregated to CEMTO levels there was a 2.38 percentage point decrease of interCEMTO pairings. Here fields CC, MM, and TT all show decreasing intraCEMTO pairs thus suggesting an increase in complexity. Paired field EE shows an increase then plateauing of intraCEMTO pairs, but overall suggests a decrease in complexity.

The calculation for complexity by citations is made by dividing the number of cited patents on a given citing patent by a weight. The weight is calculated by dividing the count of citations previously recombined with citation *i* by the total count of patents that reference citation *i*. The weight is cumulative over the three periods thus encompassing the entire database by the final time period. More explicitly, for a given primary citation, step 2 counts the number of unique citations it has been paired with, while step 3 counts all of the patents that have a listing for citation *i*. Table 2 (below) shows the three stages to the complexity calculation.

As an example of how to read the chart, in time period 1976-1988, any given citing patent in Tech56 field 8 has 4.968 cited patents listed. For any given citation in Tech56 field 8, the given citation has been also been observed as paired with 31.518 other unique citations. Then we count all of the unique patents in the prior and given (in the case of first period – we only have the given period) time period(s) that reference the given citation. If we were to then calculate (4.968 / (31.518 / 20.792)), we would arrive at the average complexity for any given Tech56 field 8 patent, 3.106.

Table 2: The three stages of the citation complexity calculation, and results

	tech56	1976-1988 average	1989-2001 average	2002-2014 average
Step 1: Number of cited patents on a given citing patent	8	4.968128	10.33502	20.65722
	12	3.939234	7.429368	15.59686
	16	3.947664	8.699477	18.33418
	29	3.94604	7.267243	12.75643
	40	4.459687	8.217237	15.89453
	41	5.131822	9.604885	17.65533
	42	4.666504	8.037944	11.17674
	43	3.617509	7.707064	14.67904
Step 2: Count of citations previously recombined with primary citation *i*	8	31.5179290	297.486842	1224.42869
	12	86.7718658	547.601314	1724.34069
	16	40.6412696	248.805533	779.381761
	29	60.1949958	238.966408	650.41017
	40	85.378219	265.342227	1219.60479
	41	137.473602	611.680449	1715.72985
	42	99.7463989	366.465218	1102.04127
	43	39.8360748	232.442314	1041.96769
Step 3: Total count of patents that reference citation *i*	8	20.7928286	217.875506	832.270618
	12	192.263193	1372.49782	5445.79922
	16	42.9928253	264.976898	760.992541
	29	109.347228	352.212954	837.548950
	40	188.308197	533.833553	1977.41275
	41	263.096892	1851.69503	4994.72632
	42	265.909601	864.013458	2745.89297
	43	55.9245005	302.98824	1334.73072

Citation Complexity Results

	tech56	1976-1988 complexity(avg)	1989-2001 complexity(avg)	2002-2014 complexity(avg)
Citation Complexity	8	3.106263	7.315406	13.31857
	12	6.115171	12.96697	37.48836
	16	3.657164	7.574245	13.455
	29	5.68054	8.931501	13.39928
	40	8.179387	10.43254	23.59896
	41	7.914577	17.43078	38.29422
	42	9.911439	15.21351	21.18359
	43	4.791656	9.626425	18.32828

Citation Results and Conclusions:

When calculating complexity via citations, we see complexity is increasing, Table 2. All eight Tech56 fields rise to level above their respective starting points in the final period. Thus the table shows how all the Tech56 fields are rising in complexity using the citation measure. Both the simple model of CEMTO paired technologies and the citation complexity calculation indicate all Tech56 fields are increasing in complexity.

Consistent with observations comparing KAC and KSC, modularity may play a role in mediating fragmentation and complexity. In this case, modularity may imply KAC does not rise or even decreases while KSC values increases substantially. In such a situation, complexity is bore by the unified production system but not by the contributing individual knowledge artifacts per se. With regards to the decrease in fields 40 and 42 in KAC period 2 (See above Table 1), this may be because of a field-wide increase in modularity from the first period to the second period. The KSC values for those two fields (See above Table 2) increase which adds suggestive confirmation what modularity is playing a part in complexity. This will be examined further in future research.

3.1.3 Location Connectivity or Location Sourcing Complexity (LSC)

Complexity as measured by locations is the last of three methods we will investigate to examine the changing levels of complexity over time. Important to note is the relatively low count of citing and cited countries as compared to either subclasses or citations. In addition 95% of the data is made up by a small number of countries (n = 9 – 13) and reference countries (n = 10 – 19). There is a wider diversity of meaningful categories in subclass and citation data than in location data. Because of this it was necessary to aggregate the country-level data to continent-level, which also maintains alignment with the aforementioned aggregating scheme such that continents are approximating the CEMTO aggregation level.

In Appendix A0 Table 8, I show the simple model of complexity where the pairs of primary to secondary location are aggregated to the continent level. There is comparatively greater activity across continents than there is across CEMT sectors. We also see there is an overall decrease in intracontinent pairing thus suggesting an increase in overall complexity. When self-paired Africa, Europe, and South America all show a decrease of intracontinent pairings, thus suggesting an increase in complexity. Two continents show mixed results where Asia shows a slight increase in intracontinent pairings and Australia[14] a slight decrease. North America shows an increase in intracontinent pairing and thus suggests a decrease in complexity.

This complexity measure was originally framed to handle tens of thousands of subclass codes but because of the relatively small count of countries (approximately 200) to possible subclass codes (approximately 80,000) this measure had to be modified. It is these intracontinent and intercontinent weights which were used to calculate locational complexity in order to produce values that represent the intention of the original complexity measure. The calculation for complexity by locations is made by dividing the number of cited locations patents on a given citing location patent by a weight (Table 11). Secondarily, this also suggests this original complexity measure has limitations making it specific to certain contexts. The weight was calculated by determining the percentage of number of times the citing continent was paired with a cited continent over the entire time period where the weight is cumulative over the three periods (Table 10).

[14] Australia also includes patents from Oceania.

To exemplify, in the 1976-1988 time period, Tech56 field 8 has an average of 5.068 pairwise locations per patent. Each pairwise combination is then divided by the weight of step 2 in which the percentage of the citing continent of a given patent in Tech56 field 8 has been paired with 0.217 (21.7%) other cited continents, averaged across all pairwise combinations. The average complexity for a patent given of each location in Tech56 field 8 is 29.097. This final complexity value is averaged for all the pairwise combinations in step 2, Note: A simple division of step 1 to step 2 results in Simpson's Paradox (Simpson, 1951) for the final complexity value; i.e. this will calculate an average of averages without accounting for differences in the various frequencies associated with each contributing average. To prevent this inaccurate descriptive statistic, the final location complexity value shown here is weighted by the number of pairwise continent:rcontinent combinations to prevent confounding and counter-intuitive results.

Table 3: Steps to Location Complexity

Average	1976-1988		1989-2001		2002-2014	
Step 1: Number of citing and cited locations	tech56	average	tech56	average	tech56	average
	8	5.067857	8	10.92727	8	21.69914
	12	4.174273	12	7.950888	12	16.34115
	16	4.067032	16	9.157815	16	19.12959
	29	4.06529	29	7.614947	29	13.38361
	40	4.620563	40	8.526991	40	16.30152
	41	5.701122	41	10.83852	41	18.99685
	42	4.849125	42	8.674921	42	12.12163
	43	3.7723	43	8.200159	43	15.46583
Step 2: Weight: Frequency of citing continent: cited continent pairings, averaged by n time period patents	tech56	average	tech56	average	tech56	average
	8	.217447	8	.196671	8	.194254
	12	.176784	12	.180875	12	.184525
	16	.190071	16	.168237	16	.1626725
	29	.197547	29	.162592	29	.1529048
	40	.17434	40	.155046	40	.1460445
	41	.172384	41	.17734	41	.1914234
	42	.132071	42	.130042	42	.1323335
	43	.175608	43	.163286	43	.1553333

Location Complexity Results

	1976-1988		1989-2001		2002-2014	
Location Complexity	tech56	complexity(avg)	tech56	complexity(avg)	tech56	complexity(avg)
	8	29.09668	8	209.1063	8	252.8576
	12	67.70354	12	119.392	12	240.6965
	16	93.59246	16	196.7498	16	370.0665
	29	83.98908	29	158.927	29	239.8846
	40	47.26549	40	66.59346	40	131.7363
	41	54.10012	41	95.24223	41	201.7026
	42	89.44292	42	186.0097	42	195.5729
	43	62.61966	43	115.3215	43	175.8323

It is possible to calculate the location complexity for a single country. By normalizing the citing country, I can stabilize the Tech56 fields to the citing country. This controls for any effect being specific to the country of origin and controls for variations in the proportion of citing patents associated with a given country. In order, the top three patenting countries are United States, Japan, Germany and they hold this order across all three periods, see Appendix A0 Table 9. These three countries may also be used to represent the three main patenting continents of North America, Asia, and Europe, see Appendix A0 Table 10. Beyond the top three countries small number problems begin to again influence the usability of the results; although small number problems begin to appear in Germany's calculations.

Location Conclusions:

Both the simple model of CEMTO paired continent locations and the location complexity calculation indicate all Tech56 fields are increasing in complexity from the first period.

General Conclusions

Table 4: Summary table of complexity output by measure

	1976-1988		1989-2001		2002-2014	
Subclass Complexity	tech56	complexity(avg)	tech56	complexity(avg)	tech56	complexity(avg)
	8	4.500946	8	5.495304	8	6.538764
	12	6.623147	12	15.0358	12	22.9347
	16	4.333273	16	6.468287	16	7.274913
	29	4.757893	29	5.335217	29	5.881689
	40	10.89805	40	9.398081	40	13.68053
	41	6.25737	41	10.62334	41	15.42614
	42	7.880091	42	7.373808	42	9.804443
	43	3.570292	43	5.138704	43	6.972797

Citation Complexity	tech56 | complexity(avg)	tech56 | complexity(avg)	tech56 | complexity(avg)
	8 | 3.106263	8 | 7.315406	8 | 13.31857
	12 | 6.115171	12 | 12.96697	12 | 37.48836
	16 | 3.657164	16 | 7.574245	16 | 13.455
	29 | 5.68054	29 | 8.931501	29 | 13.39928
	40 | 8.179387	40 | 10.43254	40 | 23.59896
	41 | 7.914577	41 | 17.43078	41 | 38.29422
	42 | 9.911439	42 | 15.21351	42 | 21.18359
	43 | 4.791656	43 | 9.626425	43 | 18.32828
Location Complexity	tech56 | complexity(avg)	tech56 | complexity(avg)	tech56 | complexity(avg)
	8 | 29.09668	8 | 209.1063	8 | 252.8576
	12 | 67.70354	12 | 119.392	12 | 240.6965
	16 | 93.59246	16 | 196.7498	16 | 370.0665
	29 | 83.98908	29 | 158.927	29 | 239.8846
	40 | 47.26549	40 | 66.59346	40 | 131.7363
	41 | 54.10012	41 | 95.24223	41 | 201.7026
	42 | 89.44292	42 | 186.0097	42 | 195.5729
	43 | 62.61966	43 | 115.3215	43 | 175.8323

From the abridged dataset sampling all major fields, our basic assumption that knowledge complexity and location complexity are rising overtime is confirmed. The use of these three methods (subclasses, citations, and locations) are interesting because they do in fact reflect different aspects of knowledge complexity and locational complexity, as they grow in different rates, show different patterns of development, and how this can be used to inform methodological choices in future studies.

Comparing subclasses and citations we can see how the measures are positively but not strongly related. The complexity values vary in growth rate by period and vary in the outcome value. Cross-classifications showed mixed trends when comparing the number of codes, the simple complexity calculation, and the NK method for complexity, whereas Co-classification data was uniformly supportive. This may suggest cross-classification data is more random or chaotic in nature than co-classification data which is why the cross-classification results seem to be dampened compared to co-classification results. Although location complexity is not directly comparable, we can see it shows the greatest rate of increase.

CHAPTER 4: STUDY 1

4.1 Introduction

It is commonly understood how knowledge building can be achieved through recombination (trial and error) whereby novel knowledge can emerge (Antonelli, 2009; Fleming, 2001; Sorenson, Rivkin, & Fleming, 2006; Ganco, 2015). From this body of work, we know a lot about the complexity of knowledge via industry or classification recombination (Fleming & Sorenson, 2001; Ganco, 2015; Kaplan & Vakili, 2015; Vagnani, 2012) but relatively little about the architecture of individual knowledge artifacts composing the underlying structure. In other words we know less about the pattern of the technological knowledge domains utilized and the originality of that recombination pattern in global innovation. Complexity through knowledge recombination has frequently been studied within a single industry or knowledge field (Fleming & Sorenson, 2001; Ganco, 2013; Kaplan & Vakili, 2015; Vagnani, 2012). As a result of this body of work, we know a lot about industry/classification recombination, complexity of knowledge, and historic innovation pattern but relatively little about the global pattern of the technological knowledge domains of expertise utilized and the originality of that recombination pattern, i.e. the knowledge inputs of the underlying structure. In this context, the architecture of knowledge building artifacts which contribute to the structure of global knowledge building.

The recombination framework provides a useful lens for examining the individual artifacts of novel knowledge building and their composition (Celo, Nebus, & Wang, 2015; Fleming & Sorenson, 2001; Ganco, 2015). This literature accounts for the characteristics and development of knowledge building for knowledge complexity. We argue that the more dispersed the underlying structure is for the knowledge search domain, more complex knowledge will be produced regardless of the measure used. This research therefore has a nested contribution: (1) to determine if the structure of knowledge building is associated with a rise in technological knowledge complexity, (2) to compare the trends revealed by the common knowledge complexity measurement approaches, and (3) illustrate the different types of distributed knowledge systems firms draw upon to build a knowledge artifact.

Technological knowledge complexity is commonly calculated through the use of patents, briefly because of the demands of the research question(s) and the stability of the patent coding system over time. By recognizing two key approaches[15] for calculating technological knowledge complexity (patent characteristic / co-classification and citation / cross-classification data, respectively reflecting an outcome and pathway measure), we can produce an independent validity test for each measure and establish the trends each measure is more inclined to indicate at a finer level of analysis. Having identified an outcome measure (Fleming and Sorenson, 2001) that calculates the complexity of novel technological knowledge admitting for the spread of knowledge domain and the originality of the recombination pattern, we then build an equivalent methodological construction for a pathway measure. Both measures of complex technological knowledge used here calculate the complexity of knowledge but each represents different aspects they potentially therefore may be measuring different traits.

[15] A third measure is that of the co-occurrence of key words in the patent texts (c.f. Engelsman and van Raan, 1991; Kaplan & Vakili, 2015; Bhattacharya & Basu, 1998). Key words are known to exhibit polysemy and to evolve in meaning (Chang et al., 2009). These factors lead to a definitional instability (Mei, Shen, & Zhai, 2007) and present unreliability for a measure applied over a longer time horizon, therefore an assessment of them is outside the scope of this research but presents fallow ground for future research.

Co-classification data has been shown to reflect product relatedness as represented by the technology characteristics of the artifact (Cantwell & Piscitello, 2000, 2004; Piscitello 2004). Thus we define product (or business) relatedness as the increase in a corporation's technological base by a varied assortment of technological competencies (Piscitello, 2004; Cantwell et al, 2004). We propose complexity measured in this method represents Knowledge Artifact Complexity (KAC) where the higher the KAC value, the more likely the Tech56 field is associated with interrelated applications.

Cross-classification data has been shown to reflect how generalizable (specialized) a technology field is (Hall, Jaffe, Trajtenberg, 2001; Hall & Trajtenberg, 2004; Trajtenberg, Jaffe, Henderson, 1997). We propose complexity measured in this second methods represents Knowledge Sourcing Complexity (KSC) where the higher the KSC value, the more likely the Tech56 field is associated with generalizable applications.

This work promotes new hypotheses regarding the underlying structure of innovation and direct methodological comparisons assessed on a global population data set. The results are expected to show how both measures can reflect the underlying structure of knowledge building and the increase in technological knowledge complexity and the distinct methodological contributions of each measure.

4.2 Hypothesis Development

Innovating firms are commonly in search of novel combinations of knowledge for problem-solving. Knowledge can grow incrementally as well as in leaps and bounds (Baumann & Siggelkow, 2013; Kauffman 1993). While envisioning a wide variety of innovative artifacts is possible (Weitzman 1996; 1998), what becomes a critical factor is the process of blending specific knowledge artifacts for a potentially fruitful contribution in a reasonable time horizon. Workable and potential solutions may therefore depend on the ability to tap into a larger source of complementary and or supporting functional knowledge with varying forms of specialized knowledge (Brusoni, Prencipe, Pavitt, 2001; Kapoor & Adner, 2012). These alternative routes emerge or are sought when potential solutions cannot be reasonably discovered within the current constraints and context (Hughes, 1987; Olsson & Frey, 2002; Weitzman, 1998).

The scope for novelty is greater in recombining previously unconnected or more technologically distant antecedent knowledge streams, even if there is greater uncertainty and lower reliability present (Grant, 1996; Katila & Ahuja, 2002). It has been argued when notably distant prior knowledge artifacts are recombined the result may be both numerous qualified failures and a select few novel innovations (Arthur, 2007; Fleming, 2001; Weitzman, 1998). Stated alternatively, the merging of two distinct technology fields represents a comparatively more complex knowledge stream than if each were to continue unaffected by the other.

When the architecture of the knowledge artifact draws on particularly novel or distant knowledge domains, the result is expected to register as more complex that its antecedents. Measures for complex technological knowledge have a dual mandate in assessing the architecture of each knowledge artifact; they must to calculate the difficulty (ease) of recombining knowledge to build said artifact, while accounting for the originality (commonality) of the pattern of the recombined knowledge streams relied upon to arrive at the knowledge artifact. Unbounded by a single empiric definition of technological knowledge, two methods emerged as powerful indicators of novel knowledge that we can apply to measuring the complexity of knowledge – that of patent subclasses (Antonelli, Krafft, & Quatraro, 2010; Breschi, Lissoni, & Malerba, 2004; Fleming & Sorenson, 2001; Vagnani, 2012; Yayavasram & Chen, 2015) and patent citations (Trajtenberg, Henderson, & Jaffe, 1997; Cantwell & Noonan, 2004; Engelsman & van Raan, 1991; Engelsman & van Raan, 1994; Zhang, Jiang, Cantwell, 2015).

The first method, KAC, follows the perspective of patent classification data when each individual patent is associated with multiple USPTO classes (Fleming & Sorenson, 2001; Yayavasram & Chen, 2015; Vagnani, 2012; Antontelli, Krafft, & Quatraro, 2010) which reflect a description of the endpoint characteristics of an artifact. Using this method reflects a description of the outcome characteristics of the knowledge artifact.

As recombination is an act of construction from prior knowledge, complex technological knowledge can also been calculated using patent citation data (Cantwell & Noonan, 2004; Engelsman & van Raan, 1991; Trajtenberg et al, 1997) which reflects some key aspects of the flow of knowledge that was directly antecedent to the development of an artifact. Using this method, KSC, reflects the pathway of development along which portions of antecedent knowledge were utilized to build the final knowledge artifact iteration. This second method was developed for this dissertation to illustrate and measure a second type of knowledge expertise utilized by innovating firms.

Additionally, because these measures are similar but distinct it is likely each will pick up different trends within that underlying structure. Both measures of complex technological knowledge used here calculate the complexity of knowledge but each represents different aspects they potentially therefore may be measuring different traits.

This presents an interesting opportunity novel to the literature, for what is the result of a direct and equivocal comparison of characteristics to pathway data for the measurement of knowledge complexity? By using the same data set to investigate whether the structure of knowledge building is associated with a rise in complexity, I compare the trends revealed by each measure matched to each patent. These two methods are likely positively correlated but may not have a straightforward or direct correlation. A second objective therefore is to more clearly ascertain the properties of these measures and the aspects of complexity each best reveal.

What are the drivers of complexity? Complexity literature indicates that more interconnections present in the system, the greater is the complexity (Kauffman, 1993; Fleming & Sorenson, 2001; Ganco, 2015). When technologies become interconnected by linking distant and previously unsuccessfully recombined technologies in novel ways, we can expect complexity to increase because the more distance technologies are being recombined into a single knowledge artifact. With regards to product relatedness, if a range of products are produced by common firms across international space, then we can supposed there is some commonality underlying the knowledge artifact characteristics which is then reflected in the technologies they represent. Technology relatedness here is here defined by linkages between the technology classes and thus frequently co-occurring (Bell & Pavitt, 1993; Patel & Pavitt, 1994; Piscitello, 2004). Thus we expect that as the degree of technology distance rises, so too do both of the complexity measures.

Hypothesis 1: As the degree of technology distance between technology fields increase, knowledge complexity will increase.

Secondly, knowledge complexity may also be facilitated by paradigm change. An existing paradigm represents an established path of knowledge accumulation that can become familiar to firms in a given context (Anderson, 2001; Kuhn, 1962). Paradigms also reflect the way value is represented – in the Information Age value is more intangible because it is based on the complex knowledge system (Langlois, 2003). Commonly through combinatorial efforts the previous paradigm becomes partially integrated into the following predominant paradigm. General purpose technologies (GPTs) provide a conduit along with formerly unconnected technologies may be recombined such that rising technological interrelatedness enables actors to recombine knowledge in a useful manner (Dosi, 1984; Perez, 1985; Freeman, 1987; Freeman & Perez, 1988; Freeman & Louça, 2001). Information and Communication Technologies (ICTs) are recognized as a type of GPT and currently are the leading innovation paradigm (Hagedoorn & Schakenraad, 1992; Santangelo, 2002). The ICT era developed opportunities to fuse together technologies previously unrelated (Kodama, 1992) and thus we expect an increase in knowledge complexity.

If there is continuity in the knowledge building pattern then actors are continuing in a paradigm such that the pattern of problem solving is predictable and well-established. However once a disruption occurs in the pattern of knowledge recombination, we expect different knowledge connections to be made regardless of an affiliation with ICT. We call this turbulence - the degree of change or flux. Having controlled for ICT (i.e. the direct effect of ICT on complexity), what are the effects of ICT-based turbulence on complexity? In other words, is ICT also indirectly affecting complexity through the turbulence it causes across the technology fields? As ICT is the conduit for system-wide change, ICT-based turbulence is expected to increase complexity.

Hypothesis 2a: As the spread of ICT increases, knowledge complexity will increase.

Hypothesis 2b: As ICT-based turbulence increases, knowledge complexity will increase.

4.3 Research data and design

4.3.1 Data

I analyze every knowledge artifact in selected fields of origin in the global population of granted USPTO patents from 1976 – 2014 (n = 1,340,799). The majority of the dataset was derived from the US patent database of John Cantwell at Rutgers University - Newark, currently being updated and extended with the help of Salma Zaman. Primarily I use the patent number, year of grant, citations, and classifications of each patent.

Complexity data is expected to reveal non-normal results. Thus we are encouraged to normalize the data by taking the logarithm of KAC and KSC complexity values. Plotting the data confirms this supposition.

It is likely that KAC and KSC reflect parts of the intellectual structure such that they are complementary measures. Co-classifications and cross-classifications have a consistent meaning across the entire database which brings some stability to the definitional meaning. As each measure pertains to a single unique patent there is likely to be some confluence of knowledge themes although fundamentally each measure is independent. Thus we expect the two measures, KAC and KSC, to be positively but not strongly correlated. A simple correlation matrix inclusive of all three time periods reveals the correlation of KAC and KSC to be 0.1919 and the normalized correlation of KAC and KSC to be 0.4771. Thus in either instance we find a pattern of confirmation that the designed measures support our expectations.

Firms know more than they make (Brusoni, Prencipe, Pavitt, 2001), thus we can expect the sourcing pattern will be greater than the characteristics present. In other words, KSC will have greater complexity values than KAC. Sourcing patterns of innovation show greater breadth because problem solving requires innovation in core and supporting fields with much of the development focused on the supporting fields (Arthur, 2007). This may be because citation data can reflect a bias towards the social networks and relationships of the inventor(s) when they disclose the patents that influenced the development of the focal knowledge artifact (Alcacer & Gittelman, 2006; Engelsman & van Raan, 1991). Comparatively, classification data may represent narrower themes of the knowledge field and the identifying characteristics of the knowledge artifact.

4.3.2 Variables

The innovativeness of each knowledge artifact is commonly measured by its degree of technological knowledge complexity via the interdependence of the contributing antecedent knowledge (Fleming & Sorenson, 2001). This measure has been used repeatedly in the literature for the purpose of calculating the complexity of knowledge embodied in a patent using the historical ease (difficulty) of recombining the constituting elements (c.f. Antonelli, Krafft, & Quatraro, 2010; Breschi, Lissoni, & Malerba, 2004; Fleming & Sorenson, 2001; Vagnani, 2012; Yayavasram & Chen, 2015). Using Fleming and Sorenson's (2001) measure we can calculate the difficulty of recombination for the architecture of each knowledge artifact. KAC reflects the extent to which knowledge characteristics are spreading across subclasses. We can measure this by observing both the spread of technology domains and the commonality or originality of the knowledge building pattern used to construct the knowledge artifact. The intuition behind the metric is fairly straightforward: if an artifact has knowledge attributes that are commonly and easily recombined across different technology domains then the complexity of that artifact is low. But when the knowledge attributes have either hardly or never been successfully recombined then the level of complexity is high. The logic is, if an artifact embodies attributes that have been recombined with a wide variety of technology domains, then the artifact components all together are not particularly complex. Conversely, when the artifact attributes are only capable of being recombined with a small and select set or are completely novel in their recombination, then we can presume those attributes are highly complex in arrangement.

4.3.2.1 Measure 1: Knowledge Artifact Complexity

Subclass references are used by the USPTO to indicate the technology characteristics of the artifact. All patents are assigned at least one technology subclass with many having several subclasses. In the denominator equation, we first assess the commonality of each contributing subclass. The denominator focuses on the commonality (inverse of originality) of technology domains, in this measure via subclasses i in patent l. We begin by identifying every iteration of subclass *i* in every patent of the dataset. The numerator of the denominator equation is arrived at by tallying the different subclasses which appear alongside subclass *i* on all previous patents. These are repeated for as many times as there are subclasses on the focal patent. The full expression of the denominator in the lower equation is achieved by summing the results of all previous uses of that subclass. This equation captures the ease of recombination and is indicated with an increased value when a particular subclass is recombined with a variety of technology domains, while controlling for all the applications in that subclass. To measure the spread of knowledge characteristics used for the entire artifact, we average the count of subclasses of that particular patent by the sum of the contributing subclasses commonality.

Dependent Variable 1: Knowledge Artifact Complexity =

$$\frac{\text{Count of subclasses on patent } l}{\sum_{l \in i} \left(\dfrac{\text{Count of subclasses previously recombined with primary subclass } i}{\text{Total count of patents that reference subclass } I\text{, accumulating by periods}} \right)}$$

The use of co-classifications to calculate the complexity of a knowledge artifact is a very specific measure which does not capture all of what we think of with regards to complexity. To illuminate our understanding we turn to the use of patent citations which have also frequently been used in the literature for the purpose of calculating the complexity of the knowledge artifact – although not in this fashion. Cross-classifications are used to reflect the extent to which knowledge characteristics are spreading across technology fields. This measure is based on the path of knowledge building according to the inventors assessment.

Measure 2: Knowledge Sourcing Complexity

As this measure is intended to be as direct and as equivocal as possible we follow the precedent established in Fleming and Sorenson (2001). The number of cited patent primary classifications on a patent is the measure for the number of pathway attributes of that knowledge artifact. Citations are used by the inventor(s) to indicate which antecedent knowledge was used to build the focal artifact. This second method focuses on the commonality of technology domains utilized, in this measure via cited patent primary classifications in patent *l*. We identify the primary classification of cited patent *i* on citing patents from 1976-2014. The numerator of the denominator equation is arrived at by counting the different cited patent primary classifications which were used to build the focal patent on all previous patents. The denominator of the denominator equation is achieved by summing the previous uses of that primary classification cited patent. To measure the technological knowledge spread of the entire artifact, we average the count of citations on that particular patent over the sum of the commonality values for each contributing classification.

Dependent Variable 2: Knowledge Sourcing Complexity =

$$\frac{\text{Count of primary classification citations on patent } l}{\sum_{l \in i} \left(\dfrac{\text{Count of citations previously recombined with primary citation } i}{\text{Total count of patents that reference citation } i\text{, accumulating by period}} \right)}$$

4.3.2.2 Independent Variables

Measures of interconnectedness reflecting distance must incorporate artifact relatedness and knowledge sourcing diversity. Interconnectedness is operationalized by two variables.

Technological Diversification

The first method is done using technology diversification via the proportion of technologies per knowledge artifact (Cantwell, 2004; Cantwell & Zhang, 2011; Zhang, 2010). This variable captures the proportion of citations at the patent level that are within or across Tech56 fields. The diversification ratio is defined in the following manner:

$$DIV_i = \mu_i / \sigma_i$$

Such that DIV_i for KSC is the ratio of the knowledge cited by the focal patent *i*; μ_i indicates the mean shares of cited patents of citing patent *i* from all Tech56 fields, and σ_i denotes the standard deviation of shares of the cited patents again across all Tech56 fields for each citing patent *i*.

Technological Distinctiveness

The second method is technology relatedness (Breshci et al., 1998; Noonan, 2002; Teece et al., 1994) which we invert to technology distinctiveness (Cantwell, Noonan, & Zhang, 2008). This variable captures the relatedness of technology within each paired Tech56 field by computing:

$$R_{ij} = (n_{ij} - \mu_{ij}) / \sigma_{ij}$$

In which is n_{ij} the actual number of linkages between technology *i* and *j*; is the anticipated number of linkages between technology *i* and *j*; and σ_{ij} is the standard deviation of the expectation.

To invert technology relatedness into technology distinctiveness,

$$D_{ij} = \max(R) - R_{ij}$$

Where Dij is the technological distinctiveness for a pair of citing and cited knowledge artifacts in fields *i* and j, Max(R) is the maximum value of R for any possible combination of fields (12 in this case), and R_{ij} is the relatedness for any combination of technology fields represented as *i* paired with *j*.

Change in share of ICT

The amount of ICT is operationalized as the percentage of ICT classes in a given Tech56 field. We follow precedent and use the six Tech56 fields that are commonly considered to be ICT-oriented (Santangelo, 2001). This will alert us to the uptake of the ICT innovation paradigm.

Turbulence

Turbulence is measured as the correlation in technology field profiles from the first period to the third and last period.

4.3.2.3 Controls

Single class dummy

On average, 10% of the patents in this database are composed of only one subclass. Although it is likely these artifacts are developed through a process of recombination, these measures cannot adequately observe the process. Thus, we include a dummy to control for outcomes that may interfere with the analysis.

Time dummy control

We also hold with the received trend – and common sense – that as time goes by, complexity will increase. For this reason, we also include a dummy variable for each time frame.

Number of subclasses

Historically complexity has been measured as the number of components in a knowledge artifact (c.f. Ghosh, Martin, Pennings, & Wezel, 2013; Marengo, Pasquali, Valente, & Dosi, 2012). Commonly results from this measure are mixed. In the attempt to build a more accurate representation of complexity, some measures include a second calculation for the number of components as distinct from the interactions between the components. We include this value here as a crude measure of complexity.

Number of citations

We measure the number of citations to parent patents each focal patent lists. This also serves as a crude measure of complexity.

Number of Unique Tech56 Classes via Citations

We measure the number of unique Tech56 classes are preset on a patent. We do so because it is expected that the greater the number of unique Tech56 classes, the greater the anticipated complexity.

Number of trials

We measure the number of times knowledge artifacts utilize the exact same set of subclasses to have a count of the commonality of certain patterns. This helps to control for the effect of exhausting local search opportunities (Fleming, 2001; Olsson, 2000). This can occur when an inventor finds a useful set of components and recombines them repeatedly; at the same time the potential to find additional functional recombinations declines.

Degree of Technology Field Connectivity

We also include the inverse of the probability that a given Tech56 field cites another given Tech56 field through the use of subclasses. This helps to capture any residual information on the commonality of a pairwise set of combinations.

4.4 Discussion

This study proposes to explore the properties of each knowledge complexity measure and unravel the aspects of complexity each best reveals. The results of this study are expected to contribute to complexity theory by addressing the various emerging underlying structures of knowledge building. The results of this study are also expected to assist in explaining how globalization has facilitated the increased complexity of knowledge, and in particular the complexity and diversity of knowledge recombinations.

CHAPTER 5: STUDY 2

5.1. Introduction

With innovation being a social practice of recursive problem-solving where useful solutions are sought through piece-meal knowledge contributions from a hierarchy of core and supporting fields (Arthur, 2007; Nahapiet & Ghoshal, 1998; Simon, 1962), innovation is likely to draw on multiple areas of expertise. Some of these areas of expertise may be core to the firm while others may be supporting or periphery. Industries aggregate in specific geographic regions (e.g. Boston's Route 128, Port wine in Portugal, Finance in London, Silicon Valley in San Francisco), which suggests to access this knowledge a firm must traverse a given geographic distance. Commonly search for innovative solutions begins near the firm's home base but when a satisfying solution is not found nearby the search organically extends further outside its industry cluster and therefore likely stretches outside its immediate geographic area particularly for supporting knowledge solutions (Ethiraj & Levinthal, 2004; Fleming, 2001; Gavetti & Levinthal, 2000). International businesses are better positioned to search outside their immediate vicinity because they already have subsidiary units, partners, and access to diaspora-based relationships elsewhere (Lorenzen & Mudambi, 2013; Thomas, 2016; Yusuf, 2008).

This importance of geography for innovation have been touted repeatedly in the innovation literature spanning topics such as clusters (Marshall, 1920; Porter, 1990), knowledge spillovers (Griliches, 1992, 1998; Mansfield, 1988, 1991), knowledge tacitness (Nightingale, 1998; Searle, 1995), and economic geography (Krugman, 1991; Lorenzen & Mudambi, 2013; Storper & Walker, 1989). Commonly in the context of technological knowledge complexity, distant knowledge recombination has been considered implicitly if not explicitly to be defined solely through the lens of distant technology knowledge fields (c.f. Fleming & Sorenson, 2001; Trajtenberg et al, 1997, Weitzman, 1998). In the present context of globalization where connectivity between areas is systematically rising, this approach does not directly address the additional locational complexity resulting from traversing geographic distances in achieving knowledge recombination – in other words, less attention has been paid to the changes in knowledge complexity unfettered by a single definition of distance. This presents the question, how does the relationship between the knowledge building components of technology field and geographic location affect the knowledge building structural complexity of the knowledge artifact?

Commonly international business studies incorporate distance between locations as a measure for spatial variation (c.f. Dunning & Lundan, 2008; Rugman, 1981, 2005) – e.g. administration, culture, economic, institution, language, and religious distances (c.f. Berry, Guillén, Zhou, 2010; Zaheer, Schomaker, & Nachum, 2012) location decisions (Berry, et al., 2010), exports (Beugelsdijk, Hennart, & Slangen, 2011) – and yet there has been little done to incorporate geographic distance into our understanding of the complexity of any given recombined knowledge artifact and the underlying structure to which it comprises. As location data represents the degree of international connectivity of the places (Cantwell & Iammarino, 1998; Cantwell & Iammarino, 2000; Cantwell, Iammarino, & Noonan, 2001), we expect this represents the third and final form of complexity examined here, that of Location Sourcing Complexity (LSC). When a Tech56 field is more reliant on cross-locational linkages, this implies it is the type of knowledge that relies more on international knowledge-based communities. Whereas Tech56 fields that exhibit less locational complexity, are those for which the relevant knowledge communities tend to be more geographically independent, specialized, and internally more self-sufficient. To exemplify, on one end of the spectrum is a globally mobile community in which every cluster depends on its relationships with others. On the other end of the spectrum are low complexity locations reflecting Marshallian clusters (1920) in which activities depend on local connections. Here we investigate the geographic origins of every knowledge artifact antecedent(s) in order to trace the degree of international connectivity of the locations with the intent of discovering how each artifact adds to our understanding of the underlying structure and drivers of complex technological knowledge building.

To be clear, locational complexity is distinct from geographic distance (Allen, 1977; Funk, 2014; Whittington, Owen-Smith, & Powell, 2009) or locational diversity (Amin & Cohendet, 2004; Dunning, 1970). Although we expect geographic distance or proximity to influence the pattern of knowledge building, rather we measure the extent and likelihood of knowledge interactions between any given pair of locations. This is likely to be related to the pure physical distance between those locations and indeed many studies in international business interpret geographic distance as a physical distance commonly measured in miles or travel time. Thus we include measures of geographic distance and locational diversity as controls. And indeed in this literature and the economic geography literature there is a greater tendency to think of geography in wider terms likely encompassing barriers or constraints to cultural and institutional distances (c.f. Beugelsdijk, McCann, & Mudambi, 2010; Song, 2014).

5.2 Hypothesis Development

The knowledge base is becoming more interconnected as particularly complex technological knowledge is built in which it draws on multiple domains of expertise. Following similar logic as recombining technological knowledge fields, the scope for novelty in recombining knowledge from different locations is greater even if it is less reliable and more uncertain (Grant, 1996; Katila & Ahuja, 2002; Takeishi, 2002). Because knowledge is socially constructed, it has a degree of tacitness that embeds and ties it to the geographic region in which it was developed or taught (Nightingale, 1998; Searle, 1995). Blending geographically dispersed knowledge is expected to result in a more complex knowledge artifact. The links forged between previously difficult to connect locations have been facilitated by decreases in the cost of connecting and decreases in the cost of transferring information; as such globalization has encouraged the world to become more spread out but interconnected. Although complex knowledge is difficult to transfer across great distances ties to distant collaborators can still facilitate the development of complex knowledge artifacts (Bell & Zaheer, 2007; Whittington, Owen-Smith, & Powell, 2009). Overall, the merging of distinct innovation patterns may result in a greater degree of locational complexity. Location Sourcing Complexity (LSC) reflects the extent to which a technology field relies on separate autonomous knowledge communities or internationally connected knowledge communities, revealing cross-country complexity. Thus the higher the LSC, the more likely the field relies upon internationally connected knowledge communities.

We now look at the extent to which the specialization of fields and locations align historically, to the extent they do not is indicative of complexity. In this situation, the underlying structure of knowledge building is becoming more dispersed across technology fields but also likely across geographic locations. Taking this into account we anticipate along with a change in the knowledge complexity comparing the antecedents to the artifact, we also anticipate greater geographic distance will produce an artifact with greater locational complexity. A wider knowledge base may indicate a wider geographic base and even though correlated, they are not the same. We expect a correlation between the recombination of technological field dispersion and geographic dispersion such that data points representing the joint consideration of knowledge complexity and geographic complexity are expected to fall along the diagonal. We expect a high but not perfect correlation of an artifact's locational complexity as it related to its technological field complexity. In study 1 we expect to establish that KAC and KSC represent different aspects of complexity and here we need to assess their relationship with LSC jointly. Therefore we make various assumptions about the relationship among LSC, KAC, and KSC. First that all three are generally positively related, and secondly that LSC is more related to KSC than KAC because they both formed from cross-classification data.

Hypothesis 3a: As KAC and KSC rise, LSC will rise.

Hypothesis 3b: As LSC and KSC rise, KAC will rise.

Hypothesis 3c: As LSC and KAC rise, KSC will rise.

Location complexity may also be facilitated by paradigm change. Information and Communication Technologies (ICTs) are recognized as a type of GPT and currently are the leading innovation paradigm (Hagedoorn & Schakenraad, 1992; Santangelo, 2002).

The ICT era developed opportunities to fuse together technologies previously unrelated (Kodama, 1992) and locations previously unconnected. Because ICT inherent connects both technology fields and locations, we anticipate location complexity will rise.

If there is continuity in the knowledge building pattern then actors are continuing in a paradigm such that the pattern of problem solving is predictable and well-established. However once a disruption occurs in the pattern of knowledge recombination, we expect different knowledge connections to be made regardless of an affiliation with ICT. We call this turbulence - the degree of change or flux. Having controlled for ICT, what are the effects of turbulence on complexity? As ICT is the conduit for system-wide change, turbulence is expected to increase complexity but not as rapidly as ICT. Lastly, there may be a modifying effect of ICT and KC combined that are causing an increase in complexity. We also test this.

Hypothesis 4a: As the spread of ICT increases, LSC will increase.

Hypothesis 4b: As the spread of ICT increases, KSC will increase.

Hypothesis 4c: As the spread of ICT increases, KAC will increase.

Hypothesis 5a: ICT interacted with KAC and KSC cause an increase in complexity.

Hypothesis 5b: ICT interacted with LSC and KAC cause an increase in complexity.

Hypothesis 5c: ICT interacted with LSC and KSC cause an increase in complexity.

5.3 Research Data and Design

5.3.1. Data

The research question exerts several demands the data must satisfy. The data must cover a broad range of technologies. It must track antecedent and finalized knowledge artifact information. It must have a long time horizon to establish trends. Lastly, it must also reveal the geographic locations of the various antecedent artifacts inventor(s). To satisfy these demands, I analyze every knowledge artifact in selected fields of origin in the global population of granted USPTO patents from 1976 – 2014 (n = 1,340,799). The geographic location of the primary inventor is expected to represent the place where the innovation occurred.

Again, complexity data is expected to reveal non-normal results. Thus we are encouraged to normalize the data by taking the logarithm of LSC complexity values. Plotting the data confirms this supposition.

It is reasonable to suppose LSC will become higher the longer the time horizon. Knowledge stickiness is expected to decline as time passes because the actors become more familiar with a wider set of sources, thus becoming more apparently connected (Markusen, 1996). Indeed, in Table 11 (Chapter 3) we see the longer the time horizon, the higher the value of LSC.

5.3.2 Variables

Study 1 establishes if knowledge artifact co-classification and knowledge artifact cross-classifications are distinctly useful in measuring different but related aspects of technological knowledge complexity. The results of that study dictate which method or methods to use in future studies concerned with technological knowledge complexity – including this second study. If both methods are distinctly useful, then both methods will be used; if the methods yield similar results, then I will choose one method.

In order to maintain parallelism between calculating the degree of technological knowledge complexity and the degree of locational complexity, we built an equivalent measure for the geographic distance between recombined antecedent knowledge sources. In other words, here we are calculating the dispersion of locations for knowledge building instead of dispersion of technology fields for knowledge building. We must move beyond a simple count measure of utilized locations because they do not reflect all that a complex system is – “rich interactions and interdependencies such that the configuration is of great importance” (Chapter 1 – Introduction). In other words, in keeping with the definition of complexity, it is not just a count of the number of links involved but also the structure of the pattern of those linkages.

5.3.2.1 Measure 3 – Locational Sourcing Complexity

This measure continues to reflect the precedent established in Fleming and Sorenson (2001) for technological knowledge complexity. All patents cite at least one antecedent patent and many cite several antecedent patents. From each knowledge artifact we extract the primary classes and the geographic location of each first named inventor from each cited patent (i.e. the key contributing field and location from each contributing antecedent patent). This is expected to reflect the overall sum and density of the knowledge building links with more distant linkages weighed more heavily. This calculation focuses on the nearness of geographic locations used, in this measure via the location of the first named inventor of every citing patent listed on patent *l*. Again to add clarity, we only use the location of each first named inventor. We identify the location of the primary inventor of cited patent *i* on citing patents from 1976 – 2014. The numerator of the denominator equation is arrived at by counting the different cited patent primary inventor locations which was used to build the focal patent on all previous patents. The denominator of the denominator equation is achieved by summing the previous uses of that first named inventor location. To measure the locational spread of the entire artifact, we average the count of locations of the cited patents over the sum of the nearness values for each contributing first named inventor location.

Dependent Variable 3: Locational Sourcing Complexity =

$$\frac{\text{Count of locations on the cited patents}}{\sum_{l \in i} \left(\begin{array}{l} \text{Frequency of citing continent: cited continent pairings, averaged by n} \\ \quad \text{time period patents, accumulating by period} \end{array} \right)}$$

5.3.2.2 Independent Variables =

KAC, KSC

Having shown in Study 1 that both KAC and KSC are distinctly useful for measuring knowledge complexity, here we include both as possible drivers for increasing locational complexity.

Change in share of ICT

The amount of ICT is operationalized as the percentage of ICT classes in a Tech56 field.

5.3.2.3 Controls

Country Distance

Geographic distance has been shown to be a factor in innovation practices, thus we control for it (Berry, 2014; Cantwell, Iammarino, Noonan, 2001; Cantwell & Vertova, 2004). Distance between locations is measured in miles and calculated via great circle distance between those two points – more commonly known as “As the crow flies.” We measure this in two ways – the average of all pairwise combinations from the primary location to secondary locations and the sum of the pairwise combinations. This distance is taken in miles between the capital city of each country and subsequently logged.

Degree of Country Connectivity

We also include the probability that a given country cites another given country. This helps to capture any residual information on the commonality of a pairwise set of combinations.

Number of location citations

To maintain as many parallels are possible we continue the trend from study 1 to include the number of location citations present as a simple measure of complexity.

Number of subclasses

Historically complexity has been measured as the number of components in a knowledge artifact. Commonly results from this measure are mixed. In the attempt to build a more accurate representation of complexity, some measures include a second calculation for the number of components as distinct from the interactions between the components. We include this value here as a simple measure of complexity.

5.4 Discussion

This study attempts to examine the changes in dispersion of location(s) used for new knowledge creation over time. The results of this study are expected to contribute to complexity theory by expanding the definition of distance to that of the joint consideration of technology field and geography. The results of this study are expected to add to the literature on the role of location in international innovation.

CHAPTER 6: STUDY 3

6.1 Introduction

To add greater depth to our understanding of recombination in this context, I next assess the data points that are uncorrelated – particularly those data points that are outliers. Outliers are expected to occur when the recombination of technology fields *or* geographic locations occurs. In other words, I focus on the conditions under which technology field complexity and locational complexity are *not* related. In doing so, I expect to demonstrate some boundaries of the anticipated relationship; i.e. that which hinders or constrains the joint complexity of knowledge and geography.

6.2 Hypothesis Development & Methods

Although stemming from the statistics, the patents collected here are examined in a case study model. The primary mode of identifying these outliers is by examining the residuals of the full KAC regression and the full KSC regression from Study 2. We expect they will be collected in terms of two characteristics – that which have high knowledge complexity values in either KAC or KSC and low locational complexity; and those with high locational complexity and low knowledge complexity. Within those two general groups the outliers have been further collected into cases typified by distinguishing characteristics.

Hypothesis 6: *Two sets of complexity outliers are expected – one with recombinations of positive residuals and one set in negative residuals.*

6.3.1. Data

The research question exerts several demands the data must satisfy. The data must cover a broad range of technologies. It must track antecedent and finalized knowledge artifact information. It must have a long time horizon to establish trends. Lastly, it must also reveal the geographic locations of the various antecedent artifacts inventor(s). To satisfy these demands, I analyze every knowledge artifact in selected fields of origin in the global population of granted USPTO patents from 1976 – 2014 (n = 1,340,799). The geographic location of the primary inventor is expected to represent the place where the innovation occurred.

As the purpose here is to assess the outliers, it is only those patents which fall outside the relationship between knowledge complexity and locational complexity that are examined here.

6.4 Discussion

The results of this study are also expected to elucidate some of the limitations on the joint consideration of technology field and geography.

CHAPTER 7: EMPIRICAL RESULTS

7.1 Hypotheses Testing – Study 1[16]

By examining KAC and KSC in tandem we are able to examine the relationship between them and where key distinguishing differences emerge as we wish to establish first that these measures are distinct and second which characteristics of complexity each best reveals. At the same time we examine several drivers of complexity.

7.1.1 Knowledge Artifact Complexity

Below, the results of the regression for Knowledge Artifact Complexity for time period 1 (1976-1988) are reported in model 1; time period 2 (1989-2001) in model 2; and time period 3 (2002-2014) in model 3 on A1 Table 10 using the Technological Diversification measure.[17] A1 Tables 4-6[18] present the descriptive statistics for periods 1, 2, and 3 for Knowledge Artifact Complexity (KAC). A1 Tables 7-9 present the two-tailed correlations for periods 1, 2, and 3 of the relevant variables for Knowledge Artifact Complexity using the Technological Diversification measure.[19] There are no problematic correlations observed among the variables.[20] Both the regressions and models are statistically significant. The dependent variable and the relevant controls are all significantly different from zero. All three independent variables Technological

[16] Supporting Tables, Graphs, and Figures for this study are located in Appendix A1.
[17] The results of the regression for Knowledge Artifact Complexity in time period 1, time period 2, and time period 3 in models 1, 2, and 3 respectively are located on A1 Table 14 using the Technological Distinctiveness measure.
[18] All descriptive statistics and correlation tables are located in the Appendix.
[19] A1 Tables 11 to 13 present the two-tailed correlations for periods 1, 2, and 3 of the relevant variables for Knowledge Artifact Complexity using the Technological Distinctiveness measure.
[20] There was some concern that the Technological Diversification and Technological Distinctiveness measures may be endogenous to both complexity measures KAC and KSC as they are conceptually similar both reflecting artifact relatedness and sourcing diversity albeit in different ways. Correlation results show endogeneity is not a concern, A1 Tables 26-27, as the relevant correlations range from 0.1200-0.1544.

Diversification measure, the squared term of the Technological Diversification, and ICT_SharePerField are all significantly different from zero.

The core purpose here is to establish increasing technological field distance is a driver of knowledge complexity. Hypothesis 1 receives mixed support and no clear determination can be made. This may be because of the double edged nature of recombination such that once a distance is linked it automatically becomes less of a distance for the following occasions in which it is used because it has become more familiar. The regressions show the relationship to be nonlinear such that it is generally inverse-U shaped. Using the Technological Diversification weight, we see in periods 1 and 2 an inverse-U shape to the relationship of technological distance and KAC[21]. We observe a change in period 3 where the relationship between KAC and both of the weights invert becoming U shaped.[22]

7.1.2 Knowledge Sourcing Complexity

A1 Tables 15-17 present the descriptive statistics for periods 1, 2, and 3 for Knowledge Sourcing Complexity (KSC). Tables A1-18 to 20 present the two-tailed correlations for periods 1, 2, and 3 of the relevant variables for Knowledge Artifact Complexity using the Technological Diversification measure[23]. There are no problematic

[21] Results for periods 1-3 are mirrored using the Technological Distinctiveness measure and thus this second measure serves to both confirm and function as a robustness check.
[22] This unexpected inversion in period 3 prompted running the Akaike's Information Criterion (AIC) test and the Schwarz's Bayesian Information Criteria (BIC) test in order to compare the linear and nonlinear model to determine which is more accurate. The results of both tests determined the nonlinear form to be more accurate, understood by the model with smaller AIC and BIC values, however the difference was small. As there is no clear difference between the two models in addition to there being no clear reason why the model would invert from period 2 to 3, the results were graphed. When graphed (see A1 Graph 28-30), the results show the predominant effect to be a slightly negative linear trend for all three periods.
[23] A1 Tables 22-24 present the two-tailed correlations for periods 1, 2, and 3 of the relevant variables for Knowledge Sourcing Complexity using the Technological Distinctiveness measure.

correlations observed among the variables. The results of the regression for Knowledge Sourcing Complexity using the Technological Diversification measure in time period 1, 2, and 3 are reported in A1 Table 21 in models 1, 2, and 3 respectively[24]. Both the regressions and models are statistically significant. The dependent variable and the relevant controls are all significantly different from zero[25]. The three independent variables Technological Diversification measures in linear and squared formats, and ICT_SharePerField all of which are significantly different from zero.

As with before, the main goal here is to establish increasing technological field distance is a driver of knowledge complexity. Once again the result for Hypothesis 1 is mixed and unclear. During period 1, the relationship between complexity and distance is U shaped but in periods 2 and 3 for the Technological Diversification[26] measure, results again show an inverted U-shape.[27]

---

[24] The results of the regression for Knowledge Sourcing Complexity using the Technological Distinctiveness weight in time period 1, 2, and 3 are reported in A1 Table 25 in models 1, 2, and 3 respectively.

[25] In addressing the use of count models (e.g. the number of subclasses or the number of citations) for indicating complexity, we can use any given KAC regression to see the best number of subclasses coefficient ranges between 0.08 to 0.15 while for any given KSC regression the coefficient for citations ranges begin around 0.16 and drop to 0.01 by period 3. Said differently, the use of count models do a poor job of indicating the knowledge complexity of a given patent. This is further confirmed in assessing the count model output from the specialty repressions on Tables 3-4, these results also indicates pure count models are not good measures for complexity, nor are count models able to distinguish KAC from KSC as each as they shows various degrees of self-inflation and unreliability over time.

[26] Also true for the Technological Distinctiveness weight.

[27] The Akaike's Information Criterion (AIC) test and the Schwarz's Bayesian Information Criteria (BIC) test were run to compare the pure linear form of the model with the parabolic model. These tests compare the linear and nonlinear model to determine which is more accurate. The results of both tests determined the nonlinear form to be more accurate, understood by the model with smaller AIC and BIC values, once more however the difference was small but more distinct than KAC results. As the difference between the models was still fairly narrow, the results were graphed. When graphed (see A1 Graphs 28-30), the results show the predominant effect to be a slightly negative linear trend although the vague inverse-U shape is more pronounced.

We also examined these relationships with a second measure of distance, that of Technological Distinctiveness. Without fail, results show they are matched to that of Technological Diversification[28].

7.1.3 ICT and Turbulence

The central purpose here is to examine if the uptake of the ICT paradigm is a second driver of increasing knowledge complexity. Hypothesis 2 is confirmed for both KAC and KSC. KAC results for Hypothesis 2 show increasing ICT has a positive effect on KAC[29], thus ICT above and beyond the technological distance between subclasses is a driver of increasing complexity. KSC results for Hypothesis 2 also confirm increasing ICT has a positive effect on KSC[30], thus ICT above and beyond the technological distance between citations is also a driver of increasing complexity. This suggests that connectivity between technological areas is systematically rising as the knowledge base is becoming more interconnected.

---

[28] A brief aside: while the results of KAC and KSC are consistent when comparing the Technological Diversification and Technological Distinctiveness weight, it is of interest to note the Technological Diversification measure demonstrates the effects of increasing complexity and increasing ICT better than using the Technological Distinctiveness measure as indicated by the larger coefficients with a reasonable amount of influence. This is not the result we expected as we anticipated the Technological Distinctiveness measure would better capture the effect of complexity from connecting technological distances. The effect size of the Technological Distinctiveness measure is essentially negligible for both KSC and KAC. Although the sign of the results are the same and indicate consistency, the effect size is marginal at best when using the Technological Distinctiveness weight. This indicates the simpler Technological Diversification weight is better able to capture the increase of complexity through connecting distant technologies than the more sophisticated Technological Distinctiveness weight.

To illustrate: using the Technological Distinctiveness weight KSC, TechDistinctivness_Cross-class (A1 Table 25) in time period 1 a 1% increase in linear technological distance, is expected to increase KSC by by 0.008 units. At the same time, using the parallel Technological Diversification weight KSC TechDiversification_Cross-class (A1 Table 21) in time period 1 a 1% increase in the linear technological distance, is expected to increase KSC by 0.929 units.

[29] To understand the meaningfulness of the correlations particularly with a large number of data points the results can be interpreted as follows. In time period 1 using KAC, TechDiversification_Co-class if we increase the share of ICT by 1%, KAC is expected to increase by 0.579% units.

[30] To understand the meaningfulness of the correlations particularly with a large number of data points the results can be interpreted as follows. In time period 1 using KSC, TechDiversification_Cross-class, if we increase the share of ICT by 1%, KSC is expected to increase by 0.406% units.

Hypothesis 2b examines if turbulence across the Tech56 fields from the uptake of the ICT paradigm is also associated with an increase in knowledge complexity. To determine this we examine the shares of each Tech56 field comparing the distribution of them in the third time period to the distribution of them in the first time period. Results indicate ICT is responsible for 42% turbulence[31] reflecting an indirect effect. Thus, ICT is a driver of complexity two-fold through both direct and indirect effects.

7.1.4 Distinguishing KAC and KSC

In this next section we seek to distinguish KAC from KSC first at a broader level and then at an individual patent level. In establishing that KAC and KSC are reflecting different characteristics of knowledge complexity, we must discuss the factors that distinguish these two measures. Although both are used to describe a single focal patent, they do so in different manners and thus becomes a starting point for investigation.

We see time is an important distinguishing characteristic of KAC and KSC because of the time period in which the shape of the parabola flips which indicates these two different types of knowledge expertise are indeed distinct. Examining the inversion present in the regression results for KAC period 3 and in KSC period 1, we first establish

[31] To reach this conclusion we first determine the regression (A1 Table 53) is statistically significant and the t-statistic of the independent variable is significantly different from zero and so the test confirms there is the possibility for turbulence to also be a driver of increasing complexity. In order to confirm this hypothesis we must take a second step and determine if the coefficient of the independent variable is also significantly different from one (see mobility effect or regression effect; c.f. Cantwell, 1989, 1991; Foreman-Peck, 1986). If it is also significantly different from one then Hypothesis 2b can be confirmed in that turbulence stemming indirectly from the effect of ICT connecting distant and as yet previously unsuccessfully connected technology fields to result in greater mobility.

Applying the same metric as a standard t-test but with the intention of determining if the result is statistically different from one (instead of zero) we calculate the significance of the turbulence effect: ((1 – Coefficient)/ Standard Error) = ((1 – 0.7197639)/ 0.1381266) = 2.02884. With 55 degree of freedom, this value is decided to be statistically different from one as well. Thus, the turbulence from time period 1 to 3 is different from one. To determine the exact value of turbulence, we calculate 1-r, taking the square root of r-squared and subtracting that result from one. This results in 42% turbulence.

that the inversion is not dramatic and second we suggest a driver behind the inversion. Beginning with KAC, we graph the data points for all three periods (A1 Graphs 28 to 30) using the TechDiversification measure because of its stronger results (see footnote 28). The clearest message from these graphs is a slight negative trend rather than an inversion[32].

It is likely that the "tipping point" from which the prior chemical driven paradigm shifted to the ICT driven paradigm is witnessed in the regression inversions first in KSC and later is KAC – which is also reflective of how firms know more than they do (c.f. Brusoni, Principe, & Pavitt, 2001). ICT complexity values are consistently the highest while interacting with the greatest diversity of fields. It is likely that the current ICT paradigm first began to appear in KSC where it reached a tipping point, thus inverting the regression; only to be later followed by KAC. This aligns with sources that suggest the current ICT paradigm began in the 1970's (c.f. Anderson, 2001).

Because firms know more than they do we expect KAC will be more narrowly focused while KSC will have a more broad purview. This is seen to be true not from the

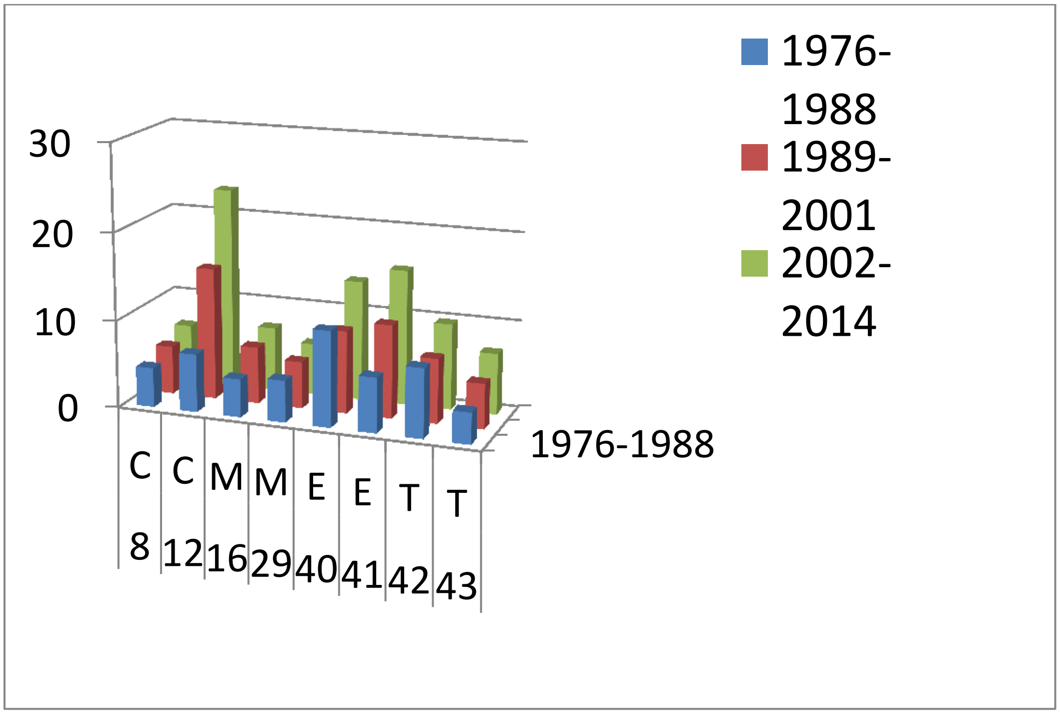
---

[32] Examining this more closely, the general shape of all three period graphs can be described as follows: it opens with the widest range of complexity results[32] before narrowing in a funnel-like fashion to a consistent band of data points (approximately 80% of the data) until the tail where changes occur. In period 1 the tail of the graph is unstable (identified by the myriad of densities present as compared to datapoints west) but shows a general downward direction, period 2 shows the tail still pointing downward but with a partial increase, and in period 3 the tail has nearly leveled off with that of the middle band of data with the except of the very tip of the tail. The mouth of the data, also being the widest part of the data, demonstrates increasingly complex values as time passes although there is very little technological distance crossed. While the mouth of the data is progressively increasing in complexity with very little distance and the tail of the data is progressively increasing in complexity while at the highest levels of distance, taken together this may explain the mechanics behind the "flip" in the results. These graphs illustrate how the flip may be a function of changes at either end of the spectrum – those at very low technological distance and those at very high technological distance – rather than any fundamental changes within the entirety of the results.

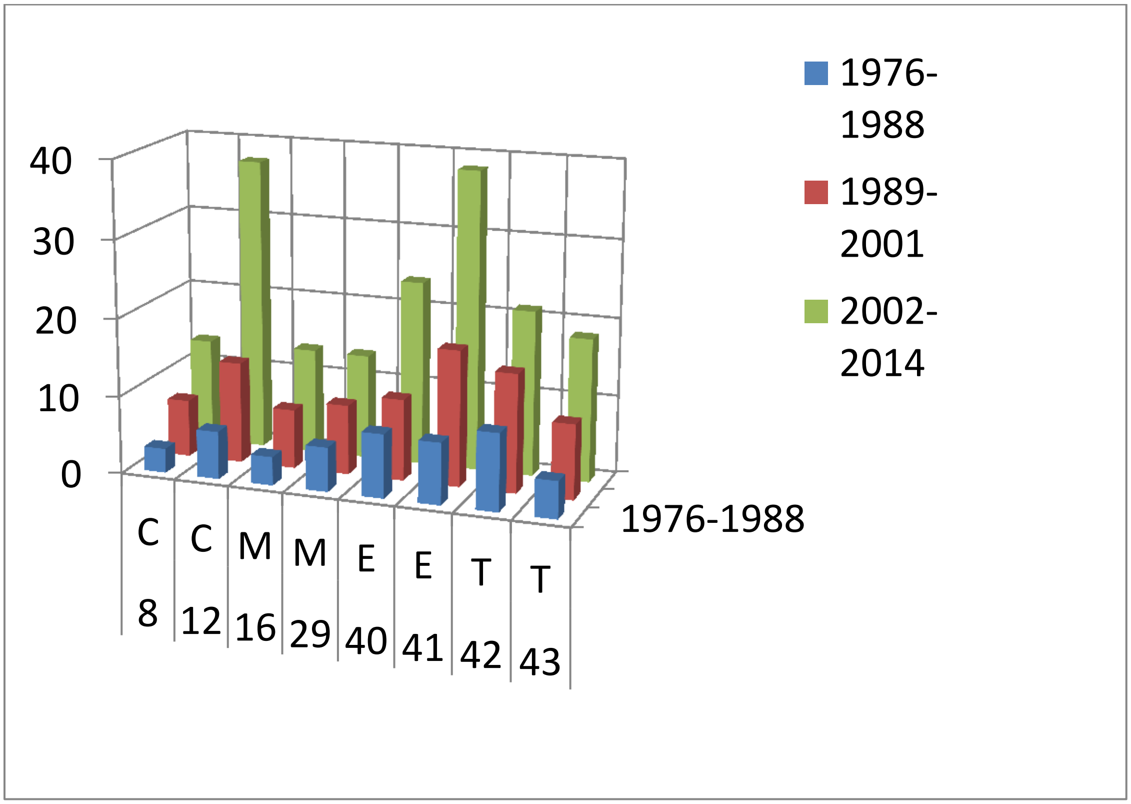
The results for KSC follow a similar pattern when graphed (A1 Graphs 31-33) for their density graphs show a funnel-like effect for all three periods although the "band" across the middle of the data (approximately 60% of the data) is clear and noticeably wider than the KAC results. This aligns with earlier suggestions that sourcing distance can have a wider influence on the complexity value of the patent than its characteristic distance.

regressions but rather from the complexity values in and of themselves. Basic graphs of KAC complexity values and KSC complexity values by period and sourcing field are located below on Graphs 1 and 2, respectively. Comparing these graphs, we see the scale of the KSC (Graph 2) values is almost double those of the KAC (Graph 1). This also may indicate how knowledge boundary regions are first sampled by cross-classification data (as identified by the higher complexity values indicating unusual or distant recombinations) rather than by co-classification data. This indicates cross-classification data show more chaos or randomness than co-classification data, because cross-classification data is inherently showing the direction of influence (Engelsman & van Raan, 1991; Olsson, 2000). Coupled with the time difference with KSC leading and KAC following, that cross-classification direction of influence is actually appearing long before that influence is reflected in the co-classification data outcome. Historically cross-classification data is thought to lag behind co-classification data (c.f. Alcacer & Gittelman, 2006; Engelsman & van Raan, 1991), however in the case of complexity the novel ideas appear to be germinating long before they are realized as a characteristic. With regards to structural changes of innovation (e.g. the uptake of the ICT era and illustrated in Hypothesis 2), the inversion appearing first in KSC then later in KAC data illustrates again how co-classification data lags behind cross-classification data.

Graph 1: Basic plot of KAC by period and sourcing field across all time periods



Graph 2: Basic plot of KSC by period and sourcing field across all time periods



We include regressions on Table 3 (for time periods 1-3, respectively) to examine the specific differences between the knowledge complexity constructs beginning with KAC as the dependent variable [33] at the level of individual patents or artifacts and illustrate again their different respective types of knowledge expertise revealed. Without changing any variables, the high and increasing explanatory power of the model, of which the high but rather stable significance of the coefficient on KSC is just one part, can be used to help decompose the explanation of KAC, and its decrease over time. This indicates these two complexity measures, while they are related, each has distinctive features, and they are becoming progressively more distinctive from one another over time. The overall correlation of 0.4771[34] (see A1 Table 34) again confirms KAC and KSC are capturing different complexity characteristics in describing the same patent and thus are distinct measures unto themselves although they are conceptually quite similar. We use six variables capturing basic patent characteristics to explain the influences on KAC after we control for any of its commonality with KSC[35]. Across all three periods, N_Subclasses are included but represent a crude measure of KAC or KSC, respectively, and thus can be thought of as a control[36]. Overall, the Share of Mechanical Engineering remains negative while growing more largely negative. As anticipated the Share of ICT is

[33] Correlation charts (A1 Tables 35-37) and descriptive statistics (A1 Tables 38-40) do not present problems. All models and the KSC variable are statistically significant and different from zero.

[34] The overall correlation of 0.4771 represents all three periods. When taken period by period, the correlation drops from 0.3646, 0.2670, to 0.1675 from periods 1-3, respectively. This is further suggestive of KAC and KSC capturing different characteristics of complexity.

[35] We examine the number of citations (N_Citations), the number of subclasses (N_Subclasses), a dummy indicating if the patent is from an Engineering field – Tech56 fields 16, 29, 42, and 43 (EngineeringDummy), the number of times the exact set of subclasses has been utilized repeatedly on other patents (Control_NumberOfTrials_Subclass), the number of unique Tech56 classes the subclasses represent (Ctrl_NumUniqueClasses_Subclass), the number of unique Tech56 classes the citations represent (Ctrl_NumUniqueClasses_Citation), the probability citation field A will cite citation field B (Ctrl_ProbFieldAcitesB_Citation), and the probability cited country A will cite cited country B (DegreeofCountryConnectivity).

[36] Across all three periods the count of subclasses are significant and show strong effects, as we would expect. N_Subclasses remains positive and grows stronger as time passes.

positive across all three periods and reflects the uptake of the ICT era is having a positive effect on KAC. Another crude form of complexity is the number of unique major patent classes which having more would suggest a particularly complex artifact as it contains characteristics from multiple technological areas. In this case, Ctrl_NumUniqueClasses_Citation is negative across all three periods although fairly stable suggesting the more patent citations from different major technology classes will have a limiting effect on the complexity of the output characteristics. In essence, the sourcing pattern is narrow for a more complex KAC pattern. The two measures for the Probability Field A will cite Field B reflects Technological Relatedness measured either through subclass or citation patterns. In the case of the former the results indicate positive coefficients, while the latter revealing negative coefficients. With the case of KAC, it was found the technological relatedness measured via citation pattern was particularly revealing.

Table 3: KAC as DV, period 1, 2, 3

	(1) KnowledgeA~y	(2) KnowledgeA~y	(3) KnowledgeA~y
KnowledgeS~y	0.432***	0.395***	0.349***
	(203.51)	(361.18)	(516.55)
N_Subclasses	0.0922***	0.103***	0.139***
	(196.10)	(388.45)	(633.99)
Ctrl_NumUn~n	-0.116***	-0.117***	-0.125***
	(-64.88)	(-153.60)	(-318.25)
Ctrl_ProbF~n	-0.767***	-0.778***	-0.570***
	(-39.20)	(-63.06)	(-65.02)
Ctrl_ProbF~s	0.0642*	1.213***	0.936***
	(2.32)	(63.97)	(65.67)
MechEng_Sh~d	-0.0214***	-0.162***	-0.284***
	(-4.40)	(-45.39)	(-88.31)
ICT_ShareP~d	0.676***	0.361***	0.449***
	(83.35)	(77.46)	(133.72)
_cons	0.748***	0.709***	0.819***
	(106.93)	(152.28)	(238.09)
N	94931	329288	657021
R-sq	0.490	0.518	0.567
f			

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

The Share of Mechanical Engineering is particularly interesting. As we progress from period 1 to period 3, this measure is remains significant and becomes progressively negative. This is likely because mechanical engineering devices are commonly known to be modularized and thus represent parts of a complex engineering system (e.g. planes, trains, and automobiles). Given this patenting pattern, complexity is more likely to emerge from interactions between the various devices or parts of the whole. When these devices are considered in isolation and not as a complete system, the components reflect more specialized parts and thus show lower KAC complexity values.

The Technological Relatedness via citations is also an interesting case and results again clearly indicate firms have different types of knowledge expertise and different types of distributed knowledge systems. Here we would presume, when more technologically related knowledge sources (KSC) have been recombined to produce a new knowledge artifact, it implies the resultant knowledge artifact will be less complex and thus we will see a negative effect on KAC – which we find to be the case in periods 1 and 3[37]. This is an especially important finding in distinguishing the aspects of complexity KAC and KSC reveal, as this finding suggests more complex knowledge artifacts (those with higher KAC) are more likely to rely on those more distant technological knowledge domains where there is little opportunity for interaction between these given areas of knowledge characteristics themselves and thus less possibility for distant knowledge sourcing (KSC).

The coefficient of KSC explaining KAC decreases over time. These results indicate as time goes by, KSC explains less and less of KAC. This may be because of the fundamentally different pursuits of each measure. KAC is becoming narrower over time and thus shows a more focused selection whereas KSC is showing as drawing upon a wider array of fields. Meanwhile the citation pattern for KSC indicates that overall these more diverse citation patterns are becoming more familiar. At the same time the explanatory power of the regression is increasing, suggesting the variables increase in explanatory power as time progresses. Overall these two measures are not substitutive but

[37] This potentially reflects exploitation behavior of the firm.

rather complementary and indicate two different types of knowledge expertise firms integrate to build a single patent.

Table 4: KSC as DV, period 1, 2, 3

	(1) KnowledgeS~y	(2) KnowledgeS~y	(3) KnowledgeS~y
KnowledgeA~y	0.494***	0.562***	0.574***
	(191.95)	(345.62)	(430.23)
N_Citations	0.146***	0.0448***	0.0170***
	(243.06)	(359.86)	(550.56)
Ctrl_NumUn~s	-0.213***	-0.219***	-0.295***
	(-92.76)	(-106.02)	(-147.72)
Ctrl_ProbF~n	1.384***	1.905***	1.734***
	(61.89)	(106.58)	(109.35)
Ctrl_ProbF~s	-2.822***	-2.981***	-2.893***
	(-70.49)	(-86.95)	(-93.42)
DegreeofCo~y	-0.0910***	0.109***	0.583***
	(-16.37)	(24.23)	(137.61)
ICT_ShareP~d	-0.508***	-0.490***	-0.741***
	(-49.73)	(-69.48)	(-117.49)
Engineerin~y	-0.368***	-0.159***	-0.446***
	(-54.14)	(-30.24)	(-86.21)
_cons	1.117***	1.259***	1.652***
	(88.15)	(123.23)	(172.76)
N	94931	329288	657021
R-sq	0.566	0.466	0.482
f			

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

With KSC as the dependent variable, the explanatory power of the model is falling as time progresses. The coefficient on KAC slowly increases suggesting it is a driver of a higher sourcing pattern – this is interesting because we previously observed how higher KAC patterns limit KSC patterns. The number of citations, a crude measure of KSC, and shows with greater clarity how it is a poor measure of complexity as it falls to almost a zero coefficient by period 3. The second crude measure, the number of unique subclasses, is also negative and consistent. The engineering dummy aligns with the results of KAC in that it is statistically significant and negative. In conjunction with this result however, it suggests that not only are the characteristics of mechanical engineering becoming simpler but also the sourced technologies used to put them together are becoming simpler. The Share of ICT is statistically significant but negative across all three periods. The sign of the coefficient is the opposite of the KAC results and represents an interesting point of comparison. In the case of KSC as the share of ICT increases the complexity of the sourcing pattern decreases sizably. This may again reflect the double edged blade of distant recombination where once a path for sourcing is established the path is no longer quite as distant as it once was. Evidence of this behavior was observed earlier in this study where no clear result could be determined when examining if increasing technological distance is a driver of complexity. Although this coefficient is stable across periods 1 and 2 it increases 50% in period 3 – this may reflect how both KAC and KSC have reached the “tipping point” for the uptake of the ICT era and this ICT is very widespread and thus very common. The degree of country connectivity is interesting in that during period 1 it is negative but becomes positive for periods 2 and 3. This may be reflecting the same pattern of how KSC reached its ICT tipping point from period 1 to 2 – as ICT connects both technology fields and geographic locations, we would naturally expect the degree of connectivity to be positive and increasing as time passes. Again, also reflecting how both KAC and KSC have surpassed their respective ICT tipping points in period 3, the coefficient of this variable grows dramatically in the third period indicating how the degree of country connectivity has increased but also how it is increasing the KSC patterns.

7.1.5 Discussion

This study puts forth the novel contributions of (1) adapting the existing complex knowledge calculation (see Fleming & Sorenson, 2001, K calculation) to be utilized across all Tech56 fields – KAC, and integrating the asymmetry of primary and secondary classification ranking to reflect the inherent importance of the primary classification, (2) identifying the uptake of ICT as a direct and indirect driver of increasing knowledge complexity, (3) building a new measure of knowledge complexity – that of KSC, and (4) examining the characteristics revealed by KAC and KSC which have previously not been compared.

We use these contributions to establish firms draw upon different types of knowledge expertise (KAC and KSC) and use different types of distributed knowledge systems during innovation. In essence, KAC reflects the present moment in that the complexity of the outcome characteristics of the focal patent is calculated. KSC is interesting in that the measure is built to reflect the antecedent (past contributing) characteristics but appears to lead the direction KAC will later progress towards. This is exemplified through various trends: KSC shows overall higher complexity levels, KSC shows greater diversity in sourcing patterns, and the regression inversion first occurs in KSC then KAC. The results suggest the outcome of knowledge building for KAC is narrower than the wider process of knowledge building (KSC) over time. Particularly with reference to the inversion, it should be no surprise that it occurs first in KSC then later in KAC. Thus while both are used to explain the same patent, there is a natural and inherent lag between the nature of the information it represents. Thus trends may be identified first in knowledge sourcing data then later observed in the characteristics of knowledge artifact data.

These results offer another way to distinguish KAC and KSC. KAC appears to reflect relatively more of the complexity from individual knowledge artifacts or component parts as well as engineering or production systems. KSC appears to provide the wider picture, reflecting the complexity of the knowledge base from which inventors drew upon to compose those individual knowledge artifacts and production systems while also appearing to reflect the fundamental underlying connections with scientific knowledge. This establishes how KAC and KSC are different measures and behave differently representing different types of knowledge expertise.

7.2 Hypotheses Testing – Study 2

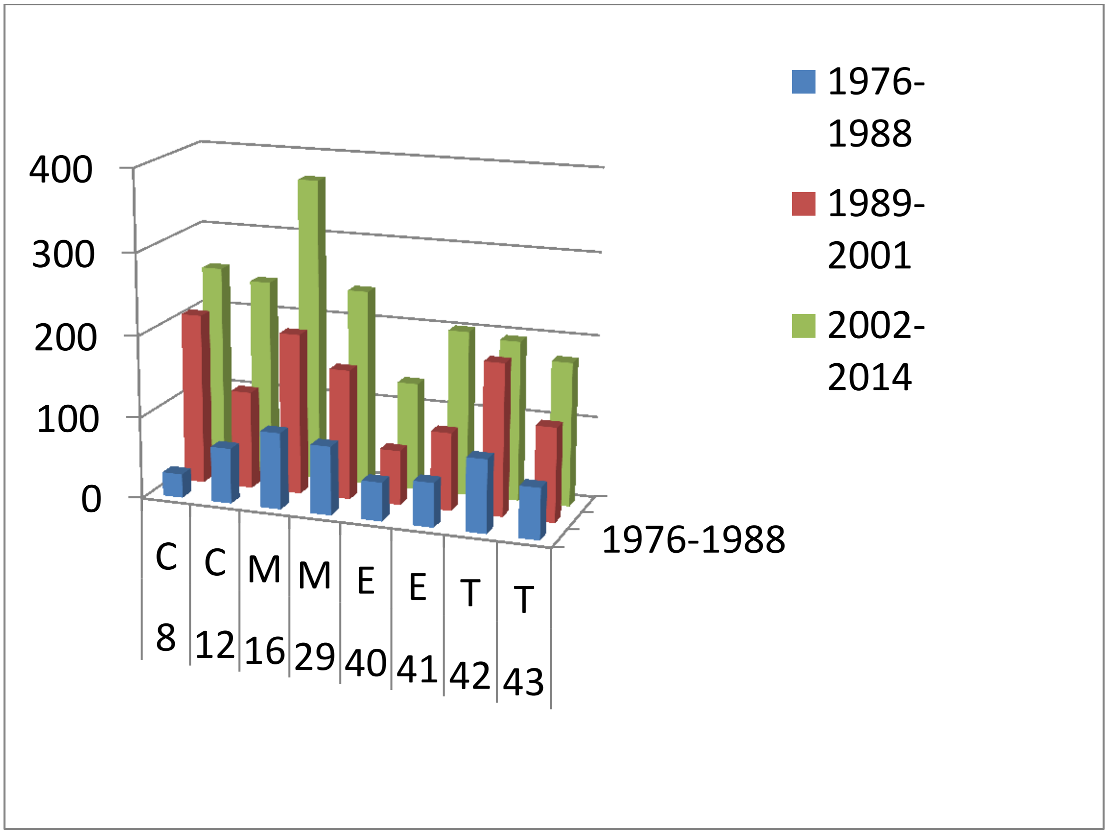
7.2.1 Location Sourcing Complexity[38]

A2 Tables 2 and 3 present the descriptive statistics and the two-tailed correlations of the relevant variables for Location Sourcing Complexity, across the data set. There are no problematic correlations observed among the variables. For comparison to KSC and KAC values by period, a basic graph of LSC complexity values by period and sourcing field are located below on Graph 1. For this study it was determined the three periods were not a necessary divisor, as such the regressions are for the data as a whole[39].

[38] Supporting Tables, Graphs, and Figures for this study are located in Appendix A2.

[39] That said, the complexity values are still calculated in 13 year periods, with latter periods building on the results of the former periods – as is the format of Study 1.

Graph 1: Basic plot of LSC by period and sourcing field across all time periods



Here the main purpose is to examine whether the joint consideration of two constructs of complexity increase the third form: beginning with testing if KAC and KSC increase LSC and if the take-up of the ICT paradigm is again a driver of increasing LSC. Secondly, we examine the distinguishing characteristics revealed by each of the three complexity constructs. The results of the regression for Location Sourcing Complexity (LSC) are presented below in Table 2. Both the regression and the models are statistically significant. Model 1 includes the dependent variable and the relevant controls of which all are significantly different from zero. Model 2 adds the independent variable Knowledge Artifact Complexity (KAC) and it is significantly different from zero; this is the first independent variable included because it was first used in the established literature and it represents the most obvious measure of complexity that of the characteristics of the outcome. Model 3 incorporates Knowledge Sourcing Complexity (KSC) and is statistically significantly different from zero; this is the second independent variable included because it is the next measure of complexity illustrating the contributing input characteristics from the antecedent patents. Model 4 includes the final independent variable Share of ICT which is shown to be significantly different from zero; this is included last because we expect the most influence on LSC will stem from the complexity of the outcome characteristics (KAC) and the input of the parent characteristics (KSC) rather than from a more generic knowledge building pattern. Model 5 incorporates the interaction effects pertaining to the Share of ICT interacted with KAC and KSC distinctly of which the results are significantly different from zero.

Table 2: LSC as DV

	(1) LocationSo~y	(2) LocationSo~y	(3) LocationSo~y	(4) LocationSo~y	(5) LocationSo~y
CountryDis~e	0.166***	0.168***	0.132***	0.162***	0.124***
	(420.20)	(432.86)	(342.47)	(414.95)	(318.03)
N_Location~n	0.0163***	0.0157***	0.00780***	0.0158***	0.00759***
	(322.75)	(316.29)	(147.21)	(317.09)	(139.71)
N_Subclasses	0.0126***	-0.0305***	0.0131***	0.0234***	0.0394***
	(29.68)	(-64.75)	(32.71)	(55.25)	(79.50)
DegreeofCo~y	-0.915***	-0.933***	-1.255***	-0.985***	-1.374***
	(-165.01)	(-171.14)	(-236.31)	(-179.83)	(-259.99)
KnowledgeA~y		0.376***			-0.127***
		(198.44)			(-43.23)
KnowledgeS~y			0.529***		0.735***
			(366.80)		(315.28)
ICT_ShareP~d				0.607***	1.479***
				(181.99)	(164.88)
Inter~ACxICT					-0.118***
					(-28.50)
Inter~SCxICT					-0.321***
					(-105.79)
_cons	2.609***	2.023***	1.990***	2.360***	1.619***
	(474.87)	(328.97)	(365.33)	(422.86)	(248.89)
N	1081240	1081240	1081240	1081240	1081240
R-sq	0.277	0.303	0.357	0.299	0.380
f					

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

The results show when both KSC and KAC are applied to the model (see Model 5), KAC is negative; thus there is an offsetting effect suggesting increasing KSC is associated with increasing LSC but increasing KAC decreases LSC. Overall this is mixed support for Hypothesis 3a. Although mixed this does follow in the path established in Study 1 in that these two measures of knowledge complexity are determinedly different. While in study 1 we examined the individual effects of KAC and KSC, here in study 2 we examine them as acting jointly. The results further drive home the contribution that these two measures truly represent different types of knowledge expertise.

Having controlled for increasing complexity via KAC and KSC, hypothesis 4a is still shown support as the effect of ICT on LSC is positive and statistically significant.

Model 5 provides evidence that ICT in and of itself is increasing complexity, but not because of KAC or KSC in LSC[40]. Thus increasing ICT has a positive and significant effect on location complexity above and beyond knowledge complexity; and ICT as a GPT is also a driver of increasing the complexity through interconnecting international knowledge sourcing origins. This is likely because ICT is not merely a connector of technology fields but also a connector of locations. These results reveal the Information Age is encouraging and facilitating the exchange of information and knowledge across technological fields (Study 1) and geographic space (Study 2). And thus, we also have further verification that while the world in which we live is growing more interdependent and interconnected it is also more spread out. Hypothesis 5a overall does not receive support in that there is not shown to be a moderating effect of ICT and knowledge complexity on LSC[41].

While the distance between countries increases has a positive effect on LSC, the degree to which those countries are connected has a negative effect. This suggests the further away and more disconnected a country is, the greater the effect will be on LSC. The number of locations cited and the number of subclasses also has a positive effect of LSC. This might not have been expected given the results of study one in which KAC and KSC revealed diverging effects.

---

[40] To understand the meaningfulness of the correlations particularly with a large number of data points the results can be interpreted as follows. In time period 1 using Model 5, increasing KAC by 1%, is expected to decrease LSC by 0.127%.

[41] The correlation between KSC and KAC is 0.4771 across all three periods. To confirm that this is not causing a multicollinearity problem and thus adversely influencing the results of Hypothesis 5, the Variable Factor Inflation test was run for all three time periods. Results are not concerning, see A2 Table 4.

7.2.2 Knowledge Artifact Complexity

A2 Tables 5 and 6 present the descriptive statistics and the two-tailed correlations of the relevant variables for KAC, across the data set. There are no problematic correlations observed among the variables.

The results of the regression for KAC are presented below in Table 3. Both the regression and the models are statistically significant. Model 1 includes the dependent variable and the relevant controls of which all are significantly different from zero. Model 2 adds the independent variable LSC and it is significantly different from zero. Model 3 adds Knowledge Sourcing Complexity (KSC) and is statistically significantly different from zero. Model 4 includes the final independent variable Share of ICT which is shown to be significantly different from zero. Model 5 incorporates the interaction effects pertaining to the Share of ICT interacted with LSC and KSC distinctly of which both are significantly different from zero.

Table 3: KAC as DV

	(1) KnowledgeA~y	(2) KnowledgeA~y	(3) KnowledgeA~y	(4) KnowledgeA~y	(5) KnowledgeA~y
Ctrl_NumUn~s	0.172***	0.191***	0.253***	0.199***	0.258***
	(138.81)	(155.26)	(234.49)	(162.59)	(241.63)
Ctrl_NumUn~n	-0.00882***	-0.0376***	-0.118***	-0.0399***	-0.118***
	(-20.12)	(-79.89)	(-272.68)	(-85.25)	(-274.77)
Ctrl_ProbF~s	2.626***	2.445***	1.562***	1.970***	1.127***
	(360.64)	(335.31)	(238.64)	(237.22)	(153.16)
LocationSo~y		0.0738***	-0.00941***	0.0683***	-0.00463***
		(156.70)	(-21.62)	(145.36)	(-9.12)
KnowledgeS~y			0.412***		0.456***
			(580.76)		(446.72)
ICT_ShareP~d				0.242***	0.578***
				(117.08)	(134.31)
Inte~LSCxICT					-0.0506***
					(-46.53)
Inte~KSCxICT					-0.0746***
					(-50.28)
_cons	1.145***	0.972***	0.662***	1.001***	0.565***
	(348.64)	(283.32)	(217.61)	(292.87)	(170.56)
N	1081240	1081240	1081240	1081240	1081240
R-sq	0.109	0.129	0.336	0.140	0.350
f					

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

These results vary in important ways from LSC as a dependent variable. Here with KAC as the dependent variable, LSC becomes negative in the final model and KSC remains positive. Thus Hypothesis 3b received mixed results, as LSC and KSC taken together reveals only KSC and not LSC positively influences KAC. Hypothesis 4b is fully supported as ICT has a positive effect on KAC. Again, Hypothesis 5b is not supported by either interaction effect.

These results reveal an interesting result is distinguishing the constructs. LSC and KSC are both measures of sourcing complexity and so naturally[42] we might expect the results to align but in fact once again the results diverge. Examined individually, each measure of sourcing complexity has a positive effect on KAC. When taken together, we

[42]As in study 1 – where we measured two forms of knowledge complexity.

consistently see (Models 3 and 5) an offsetting effect. Here LSC when used without KSC appears to proxy for KSC, we are reassured of this by the scale change in the LSC values when KSC is jointly included. We also see confirmation of the results from Study 1 in that the number of unique subclass classes and the technological relatedness of the subclasses both have a positive effect on KAC, while the number of unique citation classes has a negative effect on KAC.

7.2.3 Knowledge Sourcing Complexity

A2 Tables 8 and 9 present the descriptive statistics and the two-tailed correlations of the relevant variables for KSC, across the data set. There are no problematic correlations observed among the variables.

The results of the regression for KSC are presented below in Table 4. Both the regression and the models are statistically significant. Model 1 includes the dependent variable and the relevant controls of which all are significantly different from zero. Model 2 adds the independent variable LSC and it is significantly different from zero. Model 3 adds Knowledge Artifact Complexity (KAC) and is statistically significantly different from zero. Model 4 includes the final independent variable Share of ICT which is shown to be significantly different from zero. Model 5 includes all variables expect the interaction effects. Model 6 incorporates the interaction effects pertaining to the Share of ICT interacted with LSC and KAC distinctly of which both are significantly different from zero.

Table 4: KSC as DV

	(1) KnowledgeS~y	(2) KnowledgeS~y	(3) KnowledgeS~y	(4) KnowledgeS~y	(5) KnowledgeS~y	(6) KnowledgeS~y
DegreeofCo~y	0.537***	0.877***	0.707***	0.837***	0.709***	0.763***
	(144.79)	(251.92)	(236.25)	(240.62)	(236.32)	(259.21)
Ctrl_NumUn~n	0.254***	0.149***	0.188***	0.152***	0.188***	0.174***
	(449.53)	(263.20)	(384.82)	(269.68)	(384.69)	(360.07)
Ctrl_NumUn~s	-0.327***	-0.223***	-0.273***	-0.192***	-0.276***	-0.260***
	(-217.08)	(-159.10)	(-227.29)	(-135.70)	(-224.65)	(-216.60)
LocationSo~y		0.255***	0.191***	0.242***	0.192***	0.141***
		(451.99)	(387.49)	(425.77)	(385.34)	(253.10)
KnowledgeA~y			0.586***		0.588***	0.510***
			(629.67)		(614.28)	(407.60)
ICT_ShareP~d				0.257***	-0.0191***	-1.230***
				(119.57)	(-10.06)	(-205.88)
Inter~ACxICT						0.186***
						(89.76)
Inte~LSCxICT						0.218***
						(198.94)
_cons	1.831***	0.795***	-0.112***	0.693***	-0.108***	0.185***
	(527.14)	(202.57)	(-30.60)	(173.70)	(-29.31)	(47.81)
N	1081240	1081240	1081240	1081240	1081240	1081240
R-sq	0.195	0.323	0.505	0.332	0.505	0.528
f						

t statistics in parentheses

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

These results reveal the only time when two measures of complexity jointly considered, have a positive effect on the third complexity measure. Hypothesis 3c is supported in that KAC and LSC have a positive effect on KSC. Hypothesis 4c is not supported in that increasing ICT has a negative effect on KSC[43]. This may again be because of the double edged blade of connecting distances in that once the distance is connected it is no longer considered as novel when it is utilized in successive turns[44]. Given that the uptake of information age is being tested, it may be that the patterns of ICT connections are becoming well known and highly utilized. Hence, the connections of ICT have been effectively made with other technology fields. When the measure is taking this into account, these distances are determined to have a lower weight because they have become commonly observed[45]. Hypothesis 5c is supported in the combined effect of ICT and LSC or KAC, respectively, has a positive effect on KSC. This is also the first time this hypothesis is supported. The interaction of KAC and ICT is supported, and the interaction of LSC and ICT is supported. This is an interesting result as ICT by itself decreases KSC, but as part of an interaction effect it unanimously increases KSC. The sign change of ICT from model 4 versus models 5 and 6 is also interesting. ICT may be again be revealing the doubled-edged nature of recombination in the sense that

[43] The results of model 5 were rerun without the interaction effects to see if the negative result of ICT was generated by the interactions. The results remained unchanged.

[44] ICT does change to a negative effect in this regression not because of the correlation and not because of the interactions present in the final model as evidenced by the negative ICT coefficient in Model 5. The magnitude of the ICT coefficient in the final model is driven by the interaction effects. Overall, this result aligns with the results of study 1 where increasing technological distance has a negative effect when graphed.

[45] Two other possible explanations exist. The first being that ICT fields may be connecting primarily with itself and thus this measure might not be adequately reflecting the share of ICT with other fields. Second, it may be that ICT is proxying for KAC as an output characteristic pattern in model 4 because the share of ICT is so prominent during this period of its uptake. When they are taken together in Model 5 and 6, the joint effect shows a divergence. This aligns with the results of Hypothesis 3a and 3b where one complexity construct is positive, while the second complexity construct is negative.

connecting the distant technologies has become more familiar and thus is no longer considered complex past some point in this time horizon.

The degree of country connectivity has a positive effect on KSC. Technological relatedness in terms of unique subclass classes reveals a limited effect on the citation based KSC measure, which represents another distinguishing factor in what the constructs reveal.

7.2.4 Discussion

Further illustrating the uniqueness of KAC and KSC, let us examine the changes KAC undergoes comparing Models 2 to Models 5 in Table 2. With LSC as dependent variable, KAC is positive in Model 2 but changes to negative in Model 5. It seems likely in Model 2, KAC is in part acting as a proxy for KSC because in Model 5 the coefficient of KAC increases enormously from its starting point in Model 2. Thus, including only KAC in Model 2 is representing a more generic knowledge complexity measure in the sense that it is reflecting knowledge artifact complexity but it is simultaneously trying to proxy for KSC. The negative effect of what KAC is capturing (the narrowing focus on outcome characteristics of knowledge building) is so strong it is beginning to dominate the results of the Model 5 results. Having a negative sign on the coefficient of KAC in Model 5 indicates that it is not distant interdependencies between very different technology fields that increases locational complexity across knowledge sourcing places but rather the proximity between them when we consider the combined effects KAC and KSC taken together (as their distinctive features constitute separate but complementary effects on locational knowledge sourcing complexity, viewed as a whole). This explanation is entirely consistent with the findings of Study 1 where we directly compared KAC and KSC independently considered, finding that as pools of knowledge sourcing become more related and so more likely to mutually exchange ideas (KSC increases), artifacts tend to be less complex devices (KAC decreases); and so the places from which knowledge is sourced are also more likely to mutually exchange ideas, although the resulting products or devices reflect a very different type of distributed knowledge system, with little knowledge exchange between technologically distinct sub-components of a product, or between the primary places from which knowledge in each of those fields of technological expertise is sourced.

Here, when considering KAC and KSC jointly, we see distinct effects. As KAC increases one unit, LSC is expected to decrease; meanwhile as KSC increases one unit, LSC is expected to increase. This reflects how as artifact complexity increases, the technologies being recombined are not necessarily synthesized and they may instead be "bolted together" to make some piece of equipment or a workable device as evidenced by the decrease in LSC. Thus we see co-development in sourcing the knowledge but a consistent degree of individualized myopic focus with regards to the knowledge artifact characteristics. Within a given single patent, there are different degrees of knowledge synergy and complementarity from the distinct measures of knowledge complexity. Once again, this aligns with the theme that these are distinct measures for knowledge complexity and how firms are utilizing different types of distributed knowledge systems to build a single knowledge artifact.

In addition, when this change of KAC is paired with KSC and LSC it suggests that firms know more than they do in two different arenas. First, of course, it confirms that firms have to pay more attention to a diverse set of underlying technologies that are used in the design of any given product, while they may be becoming more focused on core products for their business. Second, and of much greater interest and novelty, it shows that *knowing more* has two aspects: a firm knows of a much wider range of technological fields and this implies that it knows of a wider range of specialized geographic origins from which knowledge can be sourced, so as to be able to bring the relevant ideas together in some form of combination. This illustrates again how firms are utilizing three types of knowledge expertise (KAC, KSC, and LSC) from different types of distributed knowledge systems (those of product or artifact characteristics, those of knowledge sourcing characteristics, and those of sourcing from across countries knowledge found in different clusters of geographically bounded innovation centers). As the products of the firm become more focused, the firm must access specialized places to obtain and bring these knowledge characteristics together.

Effectively KAC and KSC are working in an opposite direction to one another. KAC is demonstrating centrifugal characteristics whereas KSC is demonstrating centripetal forces and therefore reflects how the objective of knowledge sourcing is different from that of artifact coherence. What may be happening is firms are looking to other sources to find new ways of rearranging existing characteristics. Suggesting innovators may be searching the different types of distributed knowledge systems for new ways of reapplying or reimaging existing characteristics.

Just as the knowledge complexity constructs work is opposite directions when considered jointly, so too do the sourcing complexity constructs. When taken together in KAC as dependent variable, reveals KSC has a positive effect while LSC has a negative effect. This suggests while the knowledge sourcing pattern becomes more complex, the location sourcing pattern is becoming narrower. Said alternatively, the location sourcing patterns are becoming simpler. This also emphasizes a third pattern for a distributed knowledge system for which firms have a third type of expertise. Although KSC and LSC reveal diverging trends, the coefficient of LSC is very small and almost effectively zero. This may indicate an effect of ICT where it is connecting the locations and thus making it easier to communicate and conduct R&D across dispersed and highly complex locations patterns, even in light of the positive effect of increasing country distance. This reveals how the world is becoming more interconnected and yet spread out.

When considering KSC and LSC together we are also presented with diverging sourcing complexity patterns. In Table 3 we see how when they are taken together reveal how increasing the location sourcing pattern will hinder KAC growth while increasing the knowledge sourcing pattern will encourage KAC growth. This suggests these two measures of sourcing complexity are distinct.

7.3 Hypotheses Testing – Study 3 [46]

In study 3 the main purpose is to examine the outliers of the full regression from each complexity measure in study 2 in an attempt to learn more about what each measure is capturing and reflecting. To determine the outliers from each complexity measure as dependent variable, we examine the histogram for each set of residuals. We consider the outliers to be at the tails of the distribution, where the results level off to nearly zero. When LSC is the dependent variable we determine outliers to exist beyond +/- 7 residual units; the outliers for KSC and KAC exist beyond +/- 3 residual units. In almost every instance the plots of the residuals lie along the zero residual mark. This suggests with given a complexity construct as the dependent variable, the other two complexity constructs are unremarkable indicating there are no clear association for those high complexity values to exist – we examine this deeper to determine there are some instances where the other two complexity variables show specific trends. All three complexity measures show normal residuals to occur on all six patenting continents. A3 Table 30 presents the patenting percentage continent makes up in the data. Originally we predicted there would be two groups of outliers – one with high positive outlier complexity values and a second grouping of high negative outlier complexity; this manifests.

We would expect engineering technologies (particularly Mechanical Engineering, secondly Chemical Engineering) to compose the majority of the outlier patenting because engineering is the most traditionally mature technology which has also been diffused widely (Kuhn, 1962; Vertova, 1998; 2002). It is possible the Electrical Engineering Tech56 fields of ICT may appear as outliers as it was establishing as the next major

[46] Supporting Tables, Graphs, and Figures for this study are located in Appendix A3.

innovation paradigm in time period 1 and the “tipping point” was reached by both KAC and KSC measures by time period 3 (see Study 1).

7.3.1 Empirical Findings and Critical Analysis of LSC Outliers

In Appendix A3, Table 1 simply restates the results of the LSC full regression from Study 2 for easy reference. Graph 2 is the histogram of the results for those residuals, Graph 3 shows plotted LSC residuals by KAC, Graph 4 shows plotted LSC residuals by KSC. Table 5 illustrates the descriptive statistics of the normal range of residuals, Table 6 the descriptives for the negative outlier residuals, Table 7 for the positive outlier residuals. Table 8 presents descriptives on the patenting continents by Tech56 field for the normal range of data points, Table 9 for the negative outlier residuals, Table 10 for the positive outlier residuals.

Across the data these outliers account for less than 0.01% of the data but two distinct trends become apparent with investigation. LSC values in the normal range (A3 Table 5) of data averages 3.57, LSC negative outliers (A3 Table 6) average 6.60, and LSC positive outliers (A3 Table 7) average 10.20; indicating the values for both sets of LSC outliers are two to three times higher than the mean. The negative residual KSC patterns are approximately twice that of the normal residuals, possibly reflecting the shared sourcing pattern activities, while the positive residual KSC pattern is one-quarter that of the normal residual pattern. KAC values for positive and negative residuals are close to that of the normal residuals. The average positive residuals for the share of ICT is negligible while the normal and negative values are nearly the same at approximately 0.48. This is an unexpectedly large difference particularly in light of ICT being the current innovation paradigm. When investigated further (see A3 Table 10), we find with one exception all of the 128 positive LSC residuals stem from the three least-common patenting continents (Table 30) – Australia, South America, Africa. The results show Chemical and Mechanical Engineering fields dominate the majority of these positive LSC residuals. Concurrent to this, it may be a lack of ICT-oriented infrastructure driving this trend in the southern hemisphere and a barrier in the form of time zone changes from North America (the top patenting continent) to Australia[47]. In general, the patenting style for negative LSC residuals depends upon connecting a very great number of locations. Most of these patents cite an average between 750 – 1847 locations and stem from the most common patenting continent[48]. In the case of novel patterns (positive residuals), it may be the MNE calling upon diaspora which is increasing the sourcing complexity levels.

LSC outliers are primarily from complex location sourcing patterns and not necessarily associated with equally high KSC patterns, nor the more distantly related KAC patterns. The frequent occurrence of rare and unusual patenting continents distant locations are becoming more accessed and drawn upon as outliers suggest these locations are becoming "normalized" into the data. At the same time established locations are becoming more and more interconnected with unusual locations for sourcing knowledge. This trend suggests established patenting locations are searching for material at greater distances to integrate with their established patterns. Most of this activity is carried on the branch of those accessing and collaborating with Australian-based inventors, with a

[47] There is a 14 hour time difference between the capital of the United States (the largest patenting country) and the capital of Australia.

[48] The only exception is Chemical Engineering patents coming out of Europe where they become outliers not because of the high number of locations but rather because of the high number of subclasses associated with a moderately unusual patenting location.

distant second and third to South America and Africa. This may indicate that Australia is becoming both a more reliable location with which to innovate as well as a historically underutilized resource that is beginning to come to the fore.

What appears to be happening under LSC outliers is the common patenting locations are being paired with only a few but rare secondary locations; meanwhile the rare patenting locations are not limited in pairing with either other rare locations or with a great number of secondary locations. These results help to additionally illustrate the role of LSC as a third and distinct type of knowledge expertise as driving sourcing patterns beyond that of knowledge sourcing or knowledge characteristic patterns.

7.3.2 Empirical Findings and Critical Analysis of KSC Outliers

In Appendix A3, Table 11 simply restates the results of the LSC full regression from Study 2 for easy reference. Graph 12 is the histogram of the results for those residuals, Graph 13 shows plotted KSC residuals by KAC, Graph 14 shows plotted KSC residuals by LSC. Table 15 illustrates the descriptive statistics of the normal range of residuals, Table 16 the descriptives for the negative outlier residuals, Table 17 for the positive outlier residuals. Table 18 presents descriptives on the patenting continents by Tech56 field for the normal range of data points, Table 19 for the negative outlier residuals, Table 20 for the positive outlier residuals.

Across the data these outliers account for less than 0.08% of the data. With regards to the negative KSC residuals, the average is negative (where the normal residuals are positive) but the LSC residuals are larger than the average. With regards to the positive KSC residuals, both KSC and LSC residuals are higher than the normal residuals. Across all three groups, KAC residuals are nearly the same. The unaltered KAC residuals suggest the artifact characteristic complexity of outliers is unconnected with the sourcing pattern. Thus both novel and well-worn sourcing patterns have no clear link to KAC patterns and indicate other forces at work. Chemical Engineering is present in negative residuals but it is dominated by patents from Mechanical Engineering and Electrical Engineering. Unsurprisingly, the North American based patents show the highest average number of citing locations. Positive KSC residuals are entirely from the top 3 patenting continents (North America, Asia, Europe – in this order). The Tech56 fields 12 (Chemicals), 29 (Mechanical), and 41 (Electrical) receive a lot of patenting attention. The share of ICT patenting falls slightly from its high in normal KSC residuals to negative residuals but falls more dramatically for the positive residuals.

KSC and LSC reveal inverted outlier patterns. Positive KSC outliers are more associated with the top three patenting locations whereas positive LSC outliers are more associated with the three rare patenting locations. This may help explain the results of the regression in Study 2 when using both sourcing complexity measures as independent variables revealed how they work in opposite directions. The same is not apparent for revealing another difference between KAC and KSC. Rather they exhibit a similar tendency in that their positive residuals show similar patterns with Australia becoming more commonly introduced for KAC patterns. Instead the difference between these two constructs appears in the number of locations listed – KSC shows more while KAC shows far less. Supplemental to this, the positive KSC residuals have very few average subclasses listed whereas the positive KAC residuals have many subclasses listed. Taken together this aligns and further reinforces the precedent established in Study 1 where the sourcing pattern narrowed for artifact characteristic complexity to grow.

7.3.3 Empirical Findings and Critical Analysis of KAC Outliers

In Appendix A3, Table 21 simply restates the results of the LSC full regression from Study 2 for easy reference. Graph 22 is the histogram of the results for those residuals, Graph 23 shows plotted KAC residuals by KSC, Graph 24 shows plotted KAC residuals by LSC. Table 25 illustrates the descriptive statistics of the normal range of residuals, Table 26 the descriptives for the negative outlier residuals, Table 27 for the positive outlier residuals. Table 28 presents descriptives on the patenting continents by Tech56 field for the normal range of data points, Table 29 for the negative outlier residuals, Table 30 for the positive outlier residuals.

Across the data these outliers account for approximately 0.04% of the data. From the starting point of the normal residuals, the negative KAC residuals have an average much lower than the normal and are negative, and positive residual outliers average 1.5 times larger than their norm. KSC is similar between the normal and positive residuals while the negative residuals are approximately 50% higher. LSC is similar between normal and negative residuals but falls almost 50% for positive residuals. This is interesting because it suggests KAC will produce negative outliers when only the knowledge scouring pattern becomes more complex but will produce more positive outliers when only the LSC pattern becomes less complex. This suggests the distinction between the two sourcing patterns influences the residual outcome of the KAC pattern in addition to the characteristics themselves. Thus while output characteristics are the primary influence on KAC patterns, secondarily whether it is a high characteristic sourcing pattern or a low location sourcing pattern can also influence the outliers. ICT patterns in the outlier zones falls to nearly zero suggesting no major influence here while also being a deviation from the norm. This suggests ICT characteristics are not prevalent nor particularly influential in the outliers.

7.3.4 Discussion

There are considered to be two general models to knowledge building. First: digging deeper within a paradigm – which becomes increasingly expensive, and can reveal diminished returns to creativity (Olsson, 2000), as well as intra-context friction (Weitzman, 1998); or secondly: accessing knowledge from another field and recombining it with the core field – which is considered to be more uncertain and less reliable but has a wider scope for novelty. Patenting Tech56 field 12 is omnipresent in negative residuals across all three complexity measures which are also associated with having a great number of geographic locations listed. This may indicate this field is "digging deep" in order to discover new knowledge and exhausting the more readily discoverable knowledge. There is some verification behind this as innovation in the pharmaceuticals has become increasingly expensive (Choi, 2015) suggesting a paradigm change is eminent (c.f. Olsson, 2000). The positive residuals are more associated with fewer locations[49] suggesting MNEs are reaching across the knowledge clusters to access tacit knowledge. These patents may be offering novel updates to existing ideas or redeveloping existing ideas in or for a new context which may be associated with

---

[49] The exception being KSC.

Nonaka's (1994) expectation of circular learning where there is nothing new under the sun and the knowledge is simply being reimagined and rediscovered in new territory.

The earliest paradigm is considered to be Mechanical Engineering lasting approximately from 1770's to 1870's and as the most mature innovation paradigm it is greatly geographically spread (Vertova, 1998; 2002). Thus the information is well known, codified, and is comparatively easier to disseminate. In this case, innovating in a mature paradigm may drive firms to search in truly novel locations to advance their novel knowledge building. Mechanical Engineering fields appear as positive residuals which may add justification to this explanation. Mechanical Engineering patenting emerges as an outlier only in the sourcing pattern complexity measures and never in KAC. This is an interesting phenomenon and may be explained by the age of this innovation paradigm. Being the oldest innovation paradigm, having no KAC outliers but a consistent presence in sourcing complexity outliers suggests this industry has exhausted particularly distant characteristic patterns but can still produce highly unusual innovations from more exotic sourcing patterns. Thus innovation is not limited by the first time it is discovered, but rather novel iterations can be developed in new locations. This may also suggest how codified and thus geographically spread out this knowledge pattern is around the world.

Chemical Engineering being the second oldest innovation paradigm lasting from 1870's to 1970's may explain the presence of Tech56 field 12 in the outliers. More consistently this field appears in the negative residual range indicating a great many locations fed prior knowledge into developing the focal patent. As anticipated, the Electrical Engineering ICT fields register outliers in all but one set of outliers. This may be explained because of the two-fold effect ICT has as compared to the more traditional

Mechanical or Chemical Engineering technologies in that ICT connects technological fields and geographic locations (rather than just technological fields) simultaneously.

The findings also suggest the knowledge fields of Electrical Engineering are diffusing across geographic locations quite rapidly[50] and thus suggests an increase in LSC, which the other more mature Mechanical Engineering is not leading but is still present. Chemical Engineering is the other more common outlier which may reflect its dominance as the previous leading innovation paradigm. Drawing back our point of view, we may be observing a life cycle of complexity as has been theorized such that there is a slow oscillation between simple and complex, e.g. Gall's Law[51] (Gall, 1975). The ICT fields in Electrical Engineering are on the rise and their outlier pattern suggests ICT is in uptake where it is expected to be comparatively easier to make novel connections with other fields. Chemical Engineering being the prior innovation paradigm may be approaching a critical juncture where it is becoming more and more difficult to make complex artifacts within the traditional boundaries of the industry. Mechanical engineering, the oldest innovation paradigm, seems to have surpassed the crux of too-much-complexity and has become simpler in its knowledge building patterns. It will be worth watching in the future if Chemical Engineering also follows this pattern. Transport emerges infrequently in the outliers – once in positive KAC, in positive LSC, and negative KSC this may be the results of containing the lowest overall number of patents. Transport seems to be following a similar path to Mechanical Engineering albeit a more specialized engineering field.

[50] As the 1970's is stated to be the uptake of the ICT era and this dataset starts in 1976 lasting 39 years.

[51] "A complex system that works is invariably found to have evolved from a simple system that worked. A complex system designed from scratch never works and cannot be patched up to make it work. You have to start over with a working simple system." (Gall, 1975, p71).

A second reason for this fundamental complexity difference between Mechanical Engineering and Chemical Engineering may be differences in the patenting styles because of modularity differences. Traditionally, the Mechanical Engineering industry is considered more complex than Chemical Engineering because of the additive nature of integrating a great many number of devices or parts for a single outcome (e.g. again: planes, trains, and automobiles). There is a fundamental difference in the patenting styles of Mechanical Engineering versus Chemical Engineering. Chemical patents are required to list every single molecule that is present on the development whereas mechanical patents are not hindered by this limitation and have organically evolved into patenting small individualized parts or components of a whole development (by USPTO definition). Thus it may be that the level of aggregation is contributing to a division in the results for both complexity and for the resulting outliers.

Much of the technological knowledge in South America and Africa remains geographically bound but highly influential in terms of determining complexity. There is a degree of the staying power associated with distant location sourcing as a driver of complex knowledge building. Gaining more regular access to distant and unusual locations is in fact not weakening the power unusual locations have in developing complex technological knowledge but in fact driving and carrying on this trend allowing these rarely used locations to contribute to the global knowledge building arena as identified by their presence in both the normal range and as outliers. It may be that these unusual locations paving the way for more unusual combinations. Here we see some equalizing in the trends between common patenting and rare patenting locations such that they are not limited in pairing with either other rare locations or with a great number of secondary locations. Both have the opportunity to contribute in a highly complex way to the global knowledge environment.

CHAPTER 8: DISCUSSION AND CONCLUSION

8.0 Summary of Studies

We use USPTO patent data to investigate if the distance between technology fields and geographic locations leads to an increase in knowledge complexity and location complexity in global patenting activity. We utilized two representative technology fields from each of the four major global industries of Chemical, Electrical, Mechanical, and Transport to examine multiple measures of complexity over a long time horizon; this study is the first of its kind to apply the measure to this wide of a global base and revise the base measure for its scalability across different patenting contexts while emphasizing the inherent asymmetry between primary and secondary locations. In the first study, we examined the results of knowledge complexity as measured through subclass characteristics; Knowledge Artifact Complexity (KAC), this measure of knowledge complexity examines the difficulty in recombining the patent characteristics (co-classifications) and the commonality of those characteristics involved in producing the final patent outcome.

Next we introduce a new measure for knowledge complexity to the literature developed for this study based on the underlying principles of the initial KAC measure. This new measure examines knowledge complexity through the contributing cross-classification characteristics. Named, Knowledge Sourcing Complexity (KSC) this examines the difficulty of recombining the contributing source parent patent characteristics along with the commonality of that pattern. We determine that these two measures although related are fundamentally distinct and reveal different trends. KSC is becoming more varied as time passes while KAC is reflecting the narrowing focus of patent characteristics. KSC appears to lead and KAC appears to lag in revealing trends across time and thus as their name suggests input sourcing precedes output characteristics. This aligns with the evidence that firms know more than they do but adds to this notion in two fundamental ways. First, not only do firms know more than they do but firms know more than they do well in advance of doing it (as evidenced by KSC preceding KAC). Second, "knowing more" encompasses two aspects – that of knowing more of the available sourcing inputs and knowing the geographic locations to search for those inputs so the firm will be able to bring together and recombine relevant innovation ideas. Firms thus are utilizing two types of knowledge expertise from two different distributed knowledge systems. Because KAC and KSC are in fact different measures, they were named to reflect the patent properties they reflect. The effect of increasing distance was not found to have a clear effect on complexity for both KAC and KSC measures.

Concluding the first study, we examined if a second driver for increasing complexity stemmed from the uptake of a new paradigm – the rise of the Information Era as ICT is a GPT connecting new or previously unsuccessfully connected fields. We find this driver to be true of both KAC and KSC. Then we further examined ICT to determine that it has a two-fold effect. First that ICT itself is directly connecting fields but secondly that the turbulence from ICT is indirectly connecting fields that had not been before. This indicates ICT is contributing to an increase in knowledge complexity through both direct and indirect means.

KAC and KSC were next jointly examined as to their relationship with a third measure of complexity – that of the recombination of the antecedent cited locations from which the contributing information was sourced (LSC). This third measure of complexity is also new to the literature and developed for this second study – it is built to reflect the difficulty and commonality for the international connectivity sourcing patterns of patenting locations derived from the antecedent sourced parent characteristics. This revealed the relative international connectivity (or lack of) of the inhabitable continents through cross-country complexity. These three complexity measures were derived for each patent examined and thus we had matched data. Study 2 revealed that when taken together to explain LSC, the KSC and KAC measures worked in opposite directions such that KSC exhibits centripetal forces and KAC exhibits centrifugal forces. This aligns with the results of the first study and further distinguishes the two measures. Said alternatively, this trend suggests that while KSC becomes more diverse when LSC rises, when KAC rises LSC becomes more narrowly focused. With KAC as the DV, the sourcing patterns mirror the aforementioned results in that KSC shows centripetal forces and KAC shows centrifugal forces. With KSC as the DV, both LSC and KAC exhibit centripetal forces although this is the one context in which ICT exhibits a negative effect. This provided further evidence of how these complexity measures are distinct and represents different aspects of knowledge expertise and distributed knowledge systems.

Again here, we tested if the rise of the Information Era is not only connecting new or previously unsuccessfully connected fields because of the GPT nature of ICT but also because ICT in and of itself is simultaneously connecting new or previously unsuccessfully connected geographic locations with greater speed and reliability. This driver holds true here as well.

In the third and final study we examine the outliers of the relationship between LSC, KAC, and KSC. Here we find the outliers of the joint relationship reflecting high LSC independent of knowledge complexity. Within this we discover the types of outliers for each complexity measures – those with extremely high positive and negative residuals. Digging down further we find within the abnormally large LSC values, the outliers can be explained with one of two patterns: either there are a few rare locations being recombined or there are a great number of locations listed but from common locations. Extreme outliers are driven by either extremely high citation location counts or by extremely rare citation location patterns and not necessarily by high knowledge complexity values. LSC also shows the reader how it is a distinct third type of knowledge expertise and its own distributed knowledge system as the upper cloud of outliers is encouraging more complex knowledge building across time. In the case of KSC outliers, they reveal the opposite of LSC outliers in that the positive residuals are more associated with the top three patenting continents and the negative residuals with the three unusual patenting continents. KAC outliers align with the LSC patterns.

8.1 Contributions

This dissertation puts forth several contributions to the international business literature through the development and testing of technological knowledge complexity and location complexity in innovation. With regards to theory, we theorized the existence of two other forms of complexity (KSC and LSC) beyond the established mode (KAC). In support of this, we next provide an empirical contribution via testing the existence of these three forms of complexity and determined them to each show unique patterns and characteristics. This work also contributes to managers as it provides guidelines for the underlying structural principles or methods to increase or decrease complexity thus implying the relative difficulty or ease of recombining selected knowledge and sourcing elements in patenting activity – thus potentially influencing the speed at which the knowledge artifact can be patented.

8.2 Implications

This body of work contributes to government policy in that greater access to the rare and unusual locations provides important novel routes for patenting activity. Concurrent to this, those rare locations represent potential contributions to highly impactful complexity changes which are suggestive of a greater degree of value in the global knowledge arena. To facilitate this process, governments can reduce the complexity of the physical distance by connecting knowledge clusters. The process can also be facilitated by encouraging the use and development of ICT-oriented applications as it has the mapping effect of connecting previously difficult to connect knowledge streams and locations. Greater connections forged with other governments have been shown to decrease inter-country conflict[52] and thus facilitate economic growth.

Implications for technology policy may include, influencing the competitiveness of national (or corporate) patterns of innovation but may have to be moderated for the industry as a blanket approach may not benefit industries appropriately. To exemplify, we found how the differences in the patenting style of chemical engineering and mechanical engineering reflect different complexity patterns (of increasing complexity or increasing

---

[52] This is a fundamental reason behind the formation of the European Union after 3 major land wars stretching from the late 1800's to mid-1900's.

simplicity, respectively) but suggest this may be attributed to the differences in modularization activities of the two industries. This illustrates an implication for technology policy such that the modularity of the final product has an impact on the complexity of the contributing components (be they characteristics or sources).

Practical implications exist such that increasing the knowledge artifact complexity sourcing pattern is likely to limit the knowledge sourcing complexity pattern; in much the same way increasing the knowledge sourcing pattern is likely to limit the location sourcing pattern suggesting in both cases there is a trade-off that must be considered when seeking novel innovations. A second practical implication exists in that time is an important component of increasing complexity (and thus patent novelty) in the sense that many fields must be sourced from to both build a knowledge artifact and for those sourced characteristics to (later) appear as outcome characteristics. Thus there appears to be a gestation period in which alternative characteristic approaches are sampled before they become an output.

8.3 Limitations

Given the double-edged nature of connecting distant technologies[53], in the future scholars are encouraged to take a shorter time horizon when calculating the length of distances crossed during a given time period. Doing so may bring greater clarity to the impact of connecting distant technologies. With regards to industries, it may also be beneficial to examine the complexity of the entirety of the USPTO patenting from 1976 – 2014. This may help to further elucidate the transportation industry and add the

---

[53] See results for Study 1 in which technological distance has no clear impact on complexity and Study 2 results for ICT under the LSC as DV.

complementary electrical engineering patents not a part of the ICT fields to tease apart additional difference within the electrical fields. Although impractical, it would be fascinating to examine the complexity of the mechanical engineering patents from the take up of that innovation era with the intention of witnessing the vacillation of the industry from increasing complexity to subsequent increasing simplicity (as regularly implied by the results).

8.4 Future Research

This research also provides the opportunity to perform a critique of the original KAC measure while adding the modifications utilized here and the extensions to sourcing patterns for future methodological application. It also provides the opportunity to examine how these three established forms of complexity plays out across firms and industries in the global knowledge arena. Foundations are outlined for the development of complexity matrices for use by researchers and managers in innovation contexts. This work also provides further empirical validation of complexity theory and a boundary condition for application (that being suggested use of a shorter time horizon during calculation). Future research stemming from this body of work includes examining the role informal and indirect knowledge connections serve in increasing complexity during times of innovation testing. As the three measures do represent different aspects of complexity, it may be possible to examine those patents not examined here[54] in terms of another complexity measure for further insight into complementary analysis. Finally, the

[54] As stated earlier, this is because they contained only one subclass, citation, or location which causes the complexity calculation to fail.

role of modularization as a mediating effect between fragmentation and complexity can be examined in the future.

REFERENCES

Acs, Z., & Audretsch, D. B. (1989). Patents as a measure of innovative activity. *Kyklos*, *42*(2), 171-180.

Ahuja, G., & Lampert, C. M. (2001). Entrepreneurship in the large corporation: A longitudinal study of how established firms create breakthrough inventions. *Strategic Management Journal*, *22*(6-7), 521-543.

Alcacer, J., & Gittelman, M. (2006). Patent citations as a measure of knowledge flows: The influence of examiner citations. *The Review of Economics and Statistics*, *88*(4), 774-779.

Allen, T. J. (1977). Managing the flow of technology: technology transfer and the dissemination of technological information within the R and D organization.

Amin, A., & Cohendet, P. (2004). *Architectures of knowledge: Firms, capabilities, and communities*. Oxford University Press on Demand.

Andersen, B. (2001). Technological Change and the Evolution of Corporate Innovation: The Structure of Patenting, 1890-1990. Edward Elgar Publishing.

Andersson, U., Pedersen, T., (2010) Organizational design mechanisms for the R&D function in a world of offshoring. Scandinavian Journal Management 26,431–438.

Andersson, U., Forsgren, M., & Holm, U., (2007). Balancing subsidiary influence in the federative MNC: A business network view. Journal of International Business Studies, 38(5): 802–818.

Antonelli, C. (2009). The economics of innovation: from the classical legacies to the economics of complexity. Economics of Innovation and New Technology, 18(7), 611-646.

Antonelli, C., Krafft, J., & Quatraro, F. (2010). Recombinant knowledge and growth: The case of ICTs. Structural Change and Economic Dynamics, 21(1), 50-69.

Arndt, S. and Kierzkowski, H. (2001) 'Introduction', in S. Arndt, and H. Kierzkowski (eds), *Fragmentation: New Production Patterns in the World Economy,* Oxford: Oxford University Press, pp. 1–16.

Arthur, W. B. (2007). The structure of invention. Research Policy, 36(2), 274-287.

Austen, J. & Shapard, D. M. (2004). The Annotated Pride and Prejudice. Anchor Books.

Azmeh, S., & Nadvi, K. (2014). Asian firms and the restructuring of global value chains. *International Business Review*, *23*(4), 708-717.

Bäker, A. (2015). Non-tenured post-doctoral researchers' job mobility and research output: An analysis of the role of research discipline, department size, and coauthors. *Research Policy*, *44*(3), 634-650.

Baldwin, C. Y., & Clark, K. B. (2000). Design rules: The power of modularity (Vol. 1). MIT press.

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

Baldwin, C., & Von Hippel, E. (2011). Modeling a paradigm shift: From producer innovation to user and open collaborative innovation. Organization Science,22(6), 1399-1417.

Basberg, B. L. (1987). Patents and the measurement of technological change: a survey of the literature. *Research policy*, *16*(2), 131-141.

Baumann, O., & Siggelkow, N. (2013). Dealing with complexity: integrated vs. chunky search processes. Organization Science, 24(1), 116-132.

Bell, M. & Pavitt, K. (1993). "Technological accumulation and industrial growth: contrasts between developed and developing countries," Industrial and Corporate Change, 2, 157-201.

Bell, G. G., & Zaheer, A. (2007). Geography, networks, and knowledge flow. *Organization Science*, *18*(6), 955-972.

Berry, H. (2014). Global Integration and Innovation: Multicountry Knowledge Generation within MNCs. *Strategic Management Journal*. 35(6), 869-890.

Berry, H., Guillén, M. F., & Hendi, A. S. (2014). Is there convergence across countries? A spatial approach. *Journal of international business studies*, *45*(4), 387-404.

Berry, H., Guillén, M. F., & Zhou, N. (2010). An institutional approach to cross-national distance. *Journal of International Business Studies*, *41*(9), 1460-1480.

Beugelsdijk, S., & Hennart, J. F., Slangen, A. (2011). The impact of cultural distance on bilateral arm's length exports. *Management International Review*, *51*(6), 875-896.

Beugelsdijk, S., McCann, P., & Mudambi, R. (2010). Introduction: Place, space and organization—economic geography and the multinational enterprise.

Bhattacharya, S., & Basu, P. (1998). Mapping a research area at the micro level using co-word analysis. *Scientometrics*, *43*(3), 359-372.

Breschi, S., F. Lissoni and F. Malerba (1998) "Knowledge proximity and firms' technological Diversification" Mimeo, CESPRI-Bocconi University, Milan, Italy.

Breschi, S., Lissoni, F., & Malerba, F. (2004). The empirical assessment of firms' technological "coherence". *The economics and management of technological diversification*, *34*, 69.

Brusoni, S., Prencipe, A., & Pavitt, K. (2001). Knowledge specialization, organizational coupling, and the boundaries of the firm: why do firms know more than they make?. Administrative science quarterly, 46(4), 597-621.

Buckley, P.J., (2009). Internalisation thinking: From the multinational enterprise to the global factory. International Business Review 18, 224–235.

Buckley, P.J., (2011). International integration and coordination in the global factory. Management International Review 51(2),269–283.

Burt, R. S. (2004). Structural holes and good ideas1. American journal of sociology, 110(2), 349-399.

Cano-Kollmann, M., Cantwell, J., Hannigan, T. J., Mudambi, R., & Song, J. (2016). Knowledge connectivity: An agenda for innovation research in international business. *Journal of International Business Studies*, *47*(3), 255-262.

Cantwell, J. (1991). *Historical trends in international patterns of technological innovation* (pp. 37-72). J. Foreman-Peck (Ed.). Cambridge: Cambridge University Press.

Cantwell, J. (1995). The globalisation of technology: what remains of the product cycle model?. *Cambridge journal of economics*, *19*, 155-155.

Cantwell J.A. (2004). An historical change in the nature of corporate technological diversification, in Cantwell J.A., Gambardella A., Granstrand O. (eds.) *The Economics and Management of Technological Diversification*. Routledge: London.

Cantwell, J. A. (2006). Introduction, The Economics of Patents. Northampton: Edward Elgar.

Cantwell, J.A. (2008), "Innovation and information technology in the multinational enterprise", in Rugman, A.M. (ed., 2008), *The Oxford Handbook of International Business*, New York: Oxford University Press.

Cantwell, J., & Andersen, B. (1996). A statistical analysis of corporate technological leadership historically. *Economics of Innovation and New Technology*, *4*(3), 211-234.

Cantwell, J.A., Gambardella, A., & Granstrand, O. (2004). “Technological and corporate diversification” in J.A. Cantwell, A. Gambardella, and O. Granstrand (eds), *The Economics and Management of Technological Diversification*, Ch.1. Routledge: London.

Cantwell J. A., & Fai, F. (1999). Firms as the source of innovation and growth: the evolution of technological competence. *Journal of Evolutionary Economics*, 9(3), 331–366;

Cantwell, J., & Iammarino, S. (1998). MNCs, technological innovation and regional systems in the EU: some evidence in the Italian case. *International Journal of the Economics of Business*, *5*(3), 383-408.

Cantwell, J., & Iammarino, S. (2000). Multinational corporations and the location of technological innovation in the UK regions. *Regional Studies*, *34*(4), 317-332.

Cantwell,J.A., Iammarino, S., Noonan, C. (2001) Sticky places in slippery space – the location of innovation by MNCs in the European regions. In *Inward Investment Technological Change and Growth: The Impact of Multinational Corporations on the UK Economy*. (*Ed.) Pain, N*. Palgrave Macmillan UK.

Cantwell, J.A., & Noonan, C. (2001). *Technology Relatedness and Corporate Diversification*. Paper presented at Nelson and Winter Conference. Aalborg, Denmark. June 12-15, 2001.

Cantwell, J. A., & Noonan, C. (2004). Technology Relatedness: 1890–1995. In Academy of Management Conference (Vol. 10).

Cantwell, J.A., Noonan, C., & Zhang, F. (2008). *Technological Complexity and the Restructuring of Subsidiary Knowledge Sourcing – A 'Phantom Picture of the MNC'?* Unpublished Manuscript.

Cantwell, J., & Piscitello, L. (2000). Accumulating technological competence: its changing impact on corporate diversification and internationalization. *Industrial and Corporate Change*, *9*(1), 21-51.

Cantwell, J.A. & Piscitello, L. (2004) The Relationships Between Technological Diversification and Internalization. *The Economics and Management of Technoligical Diversification*. Cantwell, J.A., Gambardella, A. & Granstrand, O. (eds.) Routledge.

Cantwell, J. A., & Sanna-Randaccio, F. (1993). Multinationality and firm growth. Weltwirtschaftliches Archives, 129(2), 275-299.

Cantwell, J.A., Santangelo, G.D. (2002), "The new geography of corporate research in information and communications technology (ICT)", *Journal of Evolutionary Economics*, vol. 12, nos. 1-2, pp. 163-197.

Cantwell, J.A., Vertova, G. (2004). Historical evolution of technological diversification. Research Policy, 33(3), 511-529.

Cantwell J.A., & Zhang F. (2011). Knowledge accession strategies and the spatial organization of R&D, in Andersson M., Johansson B. (eds.) *Firms, Innovation and Growth*. Oxford University Press: Oxford.

Cantwell, J.A. & Zhang, F. (2011). *Technological Complexity and the Evolving Structure of MNC Subsidiary Knowledge Accumulation*. Economica e Politica Industriale – Journal of Industrial and Business Economics. 38(4), 5-33.

Celo, S., Nebus, J., & Wang, I. K. (2015). MNC structure, complexity, and performance: Insights from NK methodology. *Journal of International Management*, *21*(3), 182-199.

Chang, J., Boyd-Graber, J., Wang, C., Gerrish, S., Blei, D. (2009). Reading tea leaves: How humans interpret topic models. In *Proceedings of Neural Information Processing Systems*. Vancouver, B.C., Canada.

Choi, J. (2015) *Global Knowledge Sourcing Activities: The Choice of Research and Development (R&D) Alliance Governance Modes in the Pharmaceutical Industry (Doctoral Dissertation)*. Retrieved from RUcore Rutgers Electronic Theses & Dissertations.

Chopra, S. and Meindl, P., (2012). Supply Chain Management: strategy, planning, and operation, 5th edition, Pearson, Boston.

Contractor, F.J., Kumar, V., Kundu, S.K., Pedersen, T. (2010). Reconceptualizing the firm in a world of outsourcing and offshoring: The organizational and geographical relocation of high-value company functions. Journal Management Studies 47, 1417–1433.

Dosi, G. (1982). Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. Research policy, 11(3), 147-162.

Dosi, G. (1984). *Technical Change and Industrial Transformation: The Theory and an Application to the Semiconductor Industry*. London: Macmillan.

Dunning, J.H. (1970). *Studies in International Investment*. London: George Allen and Unwin.

Dunning, J. H. (1993). *Multinational Enterprises and the Global Economy*. Reading, MA: Addison Wesley.

Dunning, J. H., & Lundan, S. M. (2008). *Multinational enterprises and the global economy*. Edward Elgar Publishing.

Dyer, J., (1996). Specialized supplier networks as a source of competitive advantage: evidence from the automotive industry. Strategic Manage. J. 17 (4), 271–29.

Engelsman, E. C., & van Raan, A. F. (1991). *Mapping of technology: A first exploration of knowledge diffusion amongst fields of technology* (Vol. 15). Centre for Science and Technology Studies, University of Leiden.

Engelsman, E. C., & van Raan, A. F. (1994). A patent-based cartography of technology. *Research Policy*, *23*(1), 1-26.

Ethiraj, S. K., & Levinthal, D. (2004). Modularity and innovation in complex systems. *Management Science*, *50*(2), 159-173.

Ethiraj, S. K., Levinthal, D., & Roy, R. R. (2008). The Dual Role of Modularity: Innovation and Imitation. *Management Science*, 54(5), 939-955.

Feenstra, R. C. (1998). Integration of trade and disintegration of production in the global economy. *The journal of economic perspectives*, *12*(4), 31-50.

Fernandes, N. (2011). Global convergence of financing policies: Evidence for emerging-market firms. *Journal of International Business Studies*, *42*(8), 1043-1059.

Fleming, L. (2001). Recombinant uncertainty in technological search.Management science, 47(1), 117-132.

Fleming, L., & Sorenson, O. (2001). Technology as a complex adaptive system: evidence from patent data. Research policy, 30(7), 1019-1039.

Foreman-Peck, J. (1986). The motor industry. *Multinationals and World Trade, Vertical Integration and the Division of Labor in World Industries, London: Allen & Unwin*, 149-56.

Franzoni, C., Scellato, G., & Stephan, P. (2014). The mover's advantage: The superior performance of migrant scientists. Economics Letters, 122(1), 89-93.

Freeman, C. (1991). Innovation, changes of techno-economic paradigm and biological analogies in economics. Revue économique, 211-231.

Freeman, C. (1987). *Technology Policy and Economic Performance: Lessons from Japan*. London: Frances Pinter.

Freeman, C. and Louça, F. (2001). *As Time Goes By: From the Industrial Revolutions to the Information Revolution*. Oxford and New York: Oxford University Press.

Freeman, C., & Pérez, C. (1988). Structural Crises and adjustements. *G. Dosi, C. Freeman, R. Nelson, G. Silverberg. e L. Soet (Eds.), Technical Change and Economic Theory. London: Pinter*.

Frenken, K. (2006). Technological innovation and complexity theory.Economics of Innovation and New Technology, 15(2), 137-155.

Funk, R. J. (2014). Making the most of where you are: Geography, networks, and innovation in organizations. *Academy of Management Journal*, *57*(1), 193-222.

Gall, J. (1975). Systemantics: How Systems Work and Especially How They Fail. Quadrangle. The New York Times Book Co.

Ganco, M. (2013). Cutting the Gordian knot: The effect of knowledge complexity on employee mobility and entrepreneurship. *Strategic Management Journal*, *34*(6), 666-686.

Ganco, M. (2015) Complexity Theory. Wiley Encyclopedia of Management. 11:1-3.

Gavetti, G., & Levinthal, D. (2000). Looking forward and looking backward: Cognitive and experiential search. *Administrative science quarterly*, *45*(1), 113-137.

Gereffi, G. (1994). The organization of buyer-driven global commodity chains: How U.S. retailers shape overseas production networks. In *Commodity chains and global capitalism*, eds. Gereffi and M. Korzeniewicz, 95–122. Westport, CT: Praeger.

Gereffi, G. (1999). International trade and industrial upgrading in the apparel com-modity chain. Journal of International Economics, 48(1), 37–70.

Gereffi, G. (2005). The global economy: organization, governance, and development. *The handbook of economic sociology*, *2*, 160-182.

Gereffi, G., & Korzeniewicz, M. (Eds) (1994) Commodity Chains and Global Capitalism (Westport, CT: Prager Publishers).

Gereffi, G., Humphrey, J., & Sturgeon, T. (2005). The governance of global value chains. *Review of international political economy*, *12*(1), 78-104.

Goerzen, A., Asmussen, C. G., & Nielsen, B. B. (2013). Global cities and multinational enterprise location strategy. *Journal of International Business Studies*, *44*(5), 427-450.

Gooris, J., & Peeters, C. (2016). Fragmenting global business processes: A protection for proprietary information. *Journal of International Business Studies*, *47*(5), 535-562.

Ghosh, A., Martin, X., Pennings, J. M., & Wezel, F. C. (2013). Ambition is nothing without focus: Compensating for negative transfer of experience in R&D. *Organization Science*, *25*(2), 572-590.

Granstrand, O., & Oskarsson, C. (1994). Technology diversification in" mul-tech" corporations. *IEEE Transactions on Engineering Management*, *41*(4), 355-364.

Granstrand, O., Patel, P., & Pavitt, K. (1997). Multi-technology corporations: why they have "distributed" rather than "distinctive core" competencies. *California management review*, *39*(4), 8-25.

Grant, R. M. (1996). Prospering in dynamically-competitive environments: Organizational capability as knowledge integration. *Organization science*,*7*(4), 375-387.

Grigoriou, K., & Rothaermel, F. T. (2014). Structural Microfoundations of Innovation The Role of Relational Stars. Journal of Management, 40(2), 586-615.

Griliches, Z. (1988). Productivity puzzles and R & D: another nonexplanation. *The Journal of Economic Perspectives*, *2*(4), 9-21.

Griliches, Z. (1990). *Patent statistics as economic indicators: a survey* (No. w3301). National Bureau of Economic Research.

Griliches, Z. (1992). The search for R&D spillovers. The Scandinavian Journal of Economics 94 29–47. 17

Griliches, Z. (1998). P. Stoneman, ed., Handbook of the Economics of Innnovation and Technological Change. Blackwell, Oxford.

Grossman, G. M., & Rossi-Hansberg, E. (2008). Trading tasks: A simple theory of offshoring. *The American Economic Review*, *98*(5), 1978-1997.

Guilford, J. P. (1967). The nature of human intelligence. New York, NY, US: McGraw-Hill.

Hall, B. H., & Trajtenberg, M. (2004). *Uncovering GPTs with patent data* (No. w10901). National Bureau of Economic Research.

Hargadon, A.B. (1998). Firms As Knowledge Brokers: Lessons in Pursuing Continuous Innovation. California Management Review, 40(3), 209-227.

Hargadon, A., & Sutton, R. I. (1997). Technology brokering and innovation in a product development firm. *Administrative science quarterly*, 716-749.

Hagedoorn, J., & Schakenraad, J. (1992). Leading companies and networks of strategic alliances in information technologies. *Research Policy*, *21*(2), 163-190.

Helpman, E., & Trajtenberg, M. (1996). *Diffusion of general purpose technologies* (No. w5773). National bureau of economic research.

Helpman, E., & Trajtenberg, M. (1998). A Time to Sow and a Time to Reap: growth based on general purpose technologies ', in: E. Helpman (ed.) General purpose technologies and economic growth, Cambridge, Mass.: MIT press.

Henderson, R. M., & Clark, K. B. (1990). Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative science quarterly*, 9-30.

Hiskey, D. (2011, July 28). Making Tires Black, Instead of the Natural White Color of Rubber, Produces a Much Stronger and Longer Lasting Tire. Retrieved from http://www.todayifoundout.com/index.php/2011/07/making-tires-black-instead-of-the-natural-white-color-of-rubber-produces-a-much-stronger-and-longer-lasting-tire/.

History of the Passenger Tire. (n.d.). Retrieved from http://www.continentaltire.com/content/history-passenger-tire.

Hobday, M., Davies, A., & Prencipe, A. (2005). Systems integration: a core capability of the modern corporation. Industrial and corporate change, 14(6), 1109-1143.

Hughes, T. P. (1987). The evolution of large technological systems. The social construction of technological systems: New directions in the sociology and history of technology, 51-82.

Hummels, D., Rapoport, D. and Yi, K.-M. (1998) 'Vertical Specialisation and the Changing Nature of World Trade', *Federal Reserve Bank of New York Economic Policy Review,* 79–99.

Ietto-Gillies, G. (2014). The theory of the transnational corporation at 50+. *Economic Thought*, *3*(2), 38-57.

Jacobides, M. G., MacDuffie, J. P., & Tae, C. J. (2015). Agency, structure, and the dominance of OEMs: Change and stability in the automotive sector. *Strategic Management Journal*.

Jansen, J. (1996). "High Throughput Screening as a Discovery Tool in the Pharmaceutical Industry," *Labroratory Robotics and Automation*, 8, 261-265.

Johnson, B. (2008, October 26). Robert William Thomson. Retrieved from http://www.historic-uk.com/HistoryUK/HistoryofScotland/Robert-William-Thomson/.

Johnson, R. C., Noguera, G. (2012). Proximity and production fragmentation. American Economic Review, 102(3): 407-411.

Kaplan, S., & Vakili, K. (2014). The double-edged sword of recombination in breakthrough innovation. Strategic Management Journal.

Kaplinsky, R. (2005) Globalisation, Poverty and Inequality: Between a Rock and a Hard Place (Cambridge: Polity Press).

Kapoor, R., & Adner, R. (2012). What firms make vs. what they know: how firms' production and knowledge boundaries affect competitive advantage in the face of technological change. *Organization Science*, *23*(5), 1227-1248.

Katila, R., & Ahuja, G. (2002). Something old, something new: A longitudinal study of search behavior and new product introduction. Academy of management journal, 45(6), 1183-1194.

Kauffman, S.A. (1993). The origins of order: Self organization and selection in evolution. Oxford university press.

Kleinknecht, A., & Reinders, H. J. (2012). How good are patents as innovation indicators? Evidence from German CIS Data. Innovation and growth: from R&D strategies of innovating firms to economy-wide technological change. Oxford University Press, Oxford, 115-127.

Kodama, F. (1986). Technological Diversification of Japanese Industry. *Science* (233) 291-296.

Kodama, F. (1992). Technology Fusion and the New R&D. *Harvard Business Review*, July-August: 70-78.

Kodama, F. (1995). *Emerging patterns of innovation*. Harvard Business School Press.

Kodama, F. (2014). MOT in transition: From technology fusion to technology-service convergence. *Technovation*, *34*(9), 505-512.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy*, 99(3), 483-499.

Kuhn, Thomas S. (1962), The Structure of Scientific Revolutions, Chicago: University of Chicago Press. (1977), The Essential Tension, Chicago, IL: University of Chicago Press.

Kumar, K. (2001). Technology for Supporting Supply Chain Management. Communications of the ACM, 44(6), 58-61.

Langlois, R.N. (2003). The Vanishing Hand: The Changing Dynamics of Industrial Capitalism, *Industrial and Corporate Change*, 12(2): 351-385.

Langlois, R.N. (2007). The Dynamics of Industrial Capitalism: Schumpeter, Chandler, and the New Economy. London and New York: Routledge.

Langlois, R. N., & Robertson, P. L. (1992). Networks and innovation in a modular system: Lessons from the microcomputer and stereo component industries. *Research policy*, *21*(4), 297-313.

Larsen, M. M., & Pedersen, T. (2009). Vestas Wind System A/S— Exploiting global R&D synergies. Ivey Publishing, The University of Western Ontario. Case no. 9B09M079.

Laursen, K. (1996). Horizontal diversification in the Danish national system of innovation: the case of pharmaceuticals. *Research Policy*, *25*(7), 1121-1137.

Leinhard, J.H. (n.d.) Bertha Benz's Ride. Retrieved from http://www.uh.edu/engines/epi2402.htm.

Levin, R., Klevorick, A., Nelson, R., Winter, S., (1987). Appropriating the returns from industrial research and development: comments and discussion. Brookings Papers On Economic Activity 3, 783–831.

Lincoln, J. R. (1982). Intra-(and inter-) organizational networks. *Research in the Sociology of Organizations*, *1*(1), 1-38.

Lorenzen, M., & Mudambi, R. (2013). Clusters, connectivity and catch-up: Bollywood and Bangalore in the global economy. Journal of Economic Geography, 13(3): 501–534.

Madhok, A. & Phene, A. (2003) The co-evolution of advantage: Strategic management theory and the eclectic paradigm. Edited by Cantwell, J.A. & Narula, R. in International Business and the Eclectic Paradigm: Developing the OLI framework. Routledge.

Mansfield, E. (1988). The speed and cost of industrial innovation in Japan and the United States: External vs. internal technology. Management Science 34(10) 1157–1168.

Mansfield, E. (1991). Academic research and industrial innovation. Research Policy 20 1–12

March, J.G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, 2(1), Special Issue: Organizational Learning: papers in honor of (and by) James G. March, 71-87.

Marengo, L., Pasquali, C., Valente, M., & Dosi, G. (2012). Appropriability, patents, and rates of innovation in complex products industries. *Economics of Innovation & New Technology*, 21(8), 753-773.

Markusen, A. (1996) "Sticky places in slippery space: a typology of industrial districts". Economic Geography, 72(3): 293-313.

Marshall, A. (1920). *Principles of Economics,* revised 8th ed. MacMillan, London.

Mei, Q., Shen, X., & Zhai, C. (2007). Automatic labeling of multinomial topic models. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 490-499). ACM.

Miller, K. D., Zhao, M., & Calantone, R. J. (2006). Adding Interpersonal Learning and Tacit Knowledge to March's Exploration-Exploitation Model. Academy Of Management Journal, 49(4), 709-722.

Nadvi, K. (2008) Global standards, global governance and the organization of global value chain, Journal of Economic Geography, 8(3), pp. 323–343.

Nahapiet, J., & Ghoshal, S. (1998). Social capital, intellectual capital, and the organizational advantage. Academy of management review, 23(2), 242-266.

Nelson, R., Winter, S. (1982). An evolutionary theory of economic change. Harvard University Press, Cambridge, MA.

Nightingale, P. (1998). A cognitive model of innovation. Research policy, 27(7), 689-709.

Nightingale, P. (2000). Economies of scale in experimentation: knowledge and technology in pharmaceutical R&D. *Industrial and Corporate Change*,*9*(2), 315-359.

Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization science*, *5*(1), 14-37.

Noonan, C. (2002). Regional dynamics of technology sourcing: The Case of Germany. Retrieved from http://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.252237 on February 13, 2017.

Olsson, O. (2000). Knowledge as a set in idea space: An epistemological view on growth. Journal of economic growth, 5(3), 253-275.

Olsson, O., & Frey, B. S. (2002). Entrepreneurship as recombinant growth.Small Business Economics, 19(2), 69-80.

Parnas, D. L. (1972). On the criteria to be used in decomposing systems into modules. *Communications of the ACM*, *15*(12), 1053-1058.

Patel, P. & Pavitt, K. (1994). "Technological competencies in the world's largest firms: characteristics, constraints and scope for managerial choice,: ESRC paper, Science and Policy Research Unit, University of Sussex, Brighton.

Patel, P., & Pavitt, K. (1997). The technological competencies of the world's largest firms: complex and path-dependent, but not much variety. *Research policy*, *26*(2), 141-156.

Pavitt, K. (1984). “Sectoral Patterns of Technological Change: Towards a Taxonomy and a Theory,” *Research Policy*, 13, 343-374.

Pavitt, K. (1988). Uses and abuses of patent statistics. In A. F. J. van Raan (Ed.), Handbook of Quantitative Studies of Science and Technology: 509-536. Amsterdam: North Holland.

Pavitt, K. (1990). What we know about the strategic management of technology. *California management review*, *32*(3), 17-26.

Pavitt, K. (2002). Innovating routines in the business firm: what corporate tasks should they be accomplishing?. Industrial and Corporate Change, 11(1), 117-133.

Pavitt, K., Robson, M., & Townsend, J. (1987). The size distribution of innovating firms in the UK: 1945-1983. *The Journal of Industrial Economics*, 297-316.

Perez, C. (1985). Microelectronics, Long Waves and World Structural Change: New Perspectives for Developing Countries. *World Development*, 13(3): 441-463.

Piscitello, L. (2004). Corporate diversification, coherence and economic performance. *Industrial and corporate change*, *13*(5), 757-787.

Porter, M. E. (1990). The competitive advantage of notions. *Harvard business review*, *68*(2), 73-93.

Porter, M. E. (1998). Clusters and the new economics of competition. Harvard Business Review, 76(6): 77–90.

Qui, R., Cantwell, J., (2015). Revisit the Classification of General Purpose Technologies (GPTs) in Corporate Innovation Research Using Patent and Patent Citation Data. Journal of International Technology and Information Management, 24(2), 87-105.

Rezk, R., Srai, J. S., & Williamson, P. J. (2016). The impact of product attributes and emerging technologies on firms’ international configuration. *Journal of International Business Studies*, *47*(5), 610-618.

Roberts, J. (2000). From know-how to show-how? Questioning the role of information and communication technologies in knowledge transfer. Technology Analysis & Strategic Management, 12(4): 429–443.

Rosenberg, N., (1976). Perspectives on Technology. Cambridge. Cambridge University Press.

Rothaermel, F. T., & Alexandre, M. T. (2009). Ambidexterity in technology sourcing: The moderating role of absorptive capacity. Organization Science,20(4), 759-780.

Rugman, A. M., & Verbeke, A. (2001). Subsidiary-specific advantages in multinational enterprises. *Strategic Management Journal*, *22*(3), 237-250.

Rugman, A. M. (1981). *Inside the multinationals: The economics of internal markets*. Taylor & Francis.

Rugman, A.M. (2005). *The Regional Multinationals: MNEs and 'Global' Strategic Management*. Cambridge and New York: Cambridge University Press.

Santangelo, G. D. (2001). The impact of the information and communications technology revolution on the internationalisation of corporate technology. *International Business Review*, *10*(6), 701-726.

Santangelo, G.D. (2002). *Innovation in Multinational Corporations in the Information Age: The Experience of the European ICT Industry*. Cheltenham: Edward Elgar.

Saxenian, A. (1994). *Regional Advantage: Culture and Competition in Silicon Valley and Route 128.* Harvard University Press, Cambridge, MA.

Scellato, G., Franzoni, C., & Stephan, P. (2015). Migrant scientists and international networks. *Research Policy*, *44*(1), 108-120.

Scherer, F. M. (1965). Firm size, market structure, opportunity, and the output of patented inventions. *The American Economic Review*, *55*(5), 1097-1125.

Schmitt, A., & Van Biesebroeck, J. (2013). Proximity strategies in outsourcing relations: The role of geographical, cultural and relational proximity in the European automotive industry. *Journal of International Business Studies*, *44*(5), 475-503.

Schmookler, J. (1966). Patent Statistics, Inventions and Economic Growth: 18-56. Cambridge MA: Harvard University Press.

Schreiber, S.L., and Nicolaou, K.C. (1994a). Crossing the boundaries. *Chem. Biol*. 1, 1.

Schreiber, S.L., and Nicolaou, K.C. (1994b). A new journal for an expanding field. *Chem. Biol*. 1, i–ii.

Schumpeter, J. A. (1950). *Capitalism, Socialism, and Democracy. 3d Ed*. Harper and Row. New York.

Searle, J. R. (1995). *The construction of social reality*. Simon and Schuster.

Shaver, J. M. (1998). Accounting for endogeneity when assessing strategy performance: does entry mode choice affect FDI survival?. *Management Science*, *44*(4), 571-585.

Simpson, Edward H. (1951). "The Interpretation of Interaction in Contingency Tables". Journal of the Royal Statistical Society, Series B. 13: 238–241.

Simon, H. A. (1962). The Architecture of Complexity' Proceedings of the American Philosophical Society, Vol. 106, No. 6, pp. 467-482.

Soete, L. L. G. (1981). A General Test of the Technological Gap Trade Theory. Weltwirtschaftliches Archiv, 117: 638-666.

Song, J. (2014). Subsidiary absorptive capacity and knowledge transfer within multinational corporations. *Journal of International Business Studies*, *45*(1), 73-84.

Sorenson, O., Rivkin, J. W., & Fleming, L. (2006). Complexity, networks and knowledge flow. Research Policy, 35(7), 994-1017.

Stopford, John M. (1995). Competing globally to create and control resources, Mimeo. London Business School.

Storper, M., & Walker, R. (1989). The capitalist imperative: Territory, technology, and industrial growth. New York: Basil Blackwell.

Sturgeon, T.J. (2002). "Modular Production Networks: A New American Model of Industrial Organization," *Industrial and Corporate Change* 11(3), pp 451-496.

Sturgeon, T. J., Van Biesebroeck, J., & Gereffi, G. (2008). Value chains, networks and clusters: Reframing the global automotive industry. Journal of Economic Geography, 8(3): 297-321.

Takeishi, A. (2002). Knowledge Partitioning in the Interfirm Division of Labor: The Case of Automotive Product Development, Organization Science, 13(3), 321-338.

Tallman, S., Jenkins, M., Henry, N., & Pinch, S. (2004). Knowledge, clusters, and competitive advantage. *Academy of management review*,*29*(2), 258-271.

Taylor, A., & Greve, H. R. (2006). Superman or the fantastic four? Knowledge combination and experience in innovative teams. *Academy of Management Journal*, *49*(4), 723-740.

Teece, D.J., R. Rumelt, G. Dosi and S.G. Winter (1994) "Understanding Corporate Coherence: Theory and Evidence" *Journal of Economic Behaviour and Organisation* 23:1-30.

Thomas, D.C. (2016). The multicultural mind: Unleashing the hidden force for innovation in your organization. San Francisco, CA: Berrett-Koehler.

Trajtenberg, M., Henderson, R., & Jaffe, A. (1997). University versus corporate patents: A window on the basicness of invention. Economics of Innovation and new technology, 5(1), 19-50.

Vagnani, G. (2012). Exploration and long-run organizational performance: The moderating role of technological interdependence. Journal of Management, 0149206312466146.

Vernon, R. (1966). International investment and international trade in the product cycle. *The quarterly journal of economics*, 190-207.

Vertova, G. (1998), "Technological similarity in national styles of innovation in a historical perspective", *Technology Analysis and Strategic Management*, Vol. 10, No. 4, pp. 437-449.

Vertova, G. (2002), "A historical investigation of the geography of innovative activities", *Structural Change and Economic Dynamics*, Vol. 13, No. 3, pp. 259-283.

Von Tunzelmann, N., Malerba, F., Nightingale, P., & Metcalfe, S. (2008). Technological paradigms: past, present and future. Industrial and Corporate Change, 17(3), 467-484.

Weigelt, C., & Miller, D. J. (2013). Implications of internal organization structure for firm boundaries. Strategic Management Journal, 34(12), 1411-1434.

Weisberg RW. (1999). Creativity and knowledge: a challenge to theories. In *Handbook of Creativity*, Sternberg RJ (ed). Cambridge University Press: Cambridge, UK; 226–250.

Weitzman, M. L. (1996). Hybridizing growth theory. The American Economic Review, 207-212.

Weitzman, M. L. (1998). Recombinant growth. Quarterly Journal of Economics, 331-360.

Whittaker, E. & D. Jane Bower. (1994). A shift to external alliances for product development in the pharmaceutical industry. R&D Management, 24(3): 249-60.

Whittington, K. B., Owen-Smith, J., & Powell, W. W. (2009). Networks, propinquity, and innovation in knowledge-intensive industries. *Administrative science quarterly*, *54*(1), 90-122.

Winter, S. G. (1984). Schumpeterian competition in alternative technological regimes. Journal of Economic Behavior & Organization, 5(3), 287-320.

Yamin, M., & Sinkovics, R., (2010). ICT deployment and resource-based power in multinational enterprise futures. The Futures of International Business, 42(9): 952–959.

Yayavaram, S., & Chen, W. (2015). Changes in firm knowledge couplings and firm innovation performance: The moderating role of technological complexity. Strategic Management Journal, 36(3), 377-396.

Yeats, A. (2001) 'Just How Big is Global Production Sharing?' in S. Arndt and H. Kierzkowski (eds), *Fragmentation: New Production Patterns in theWorld Economy*, Oxford: Oxford University Press, pp. 108–43.

Yeung, H. W. C., & Coe, N. (2015). Toward a dynamic theory of global production networks. *Economic Geography*, *91*(1), 29-58.

Yusuf, S. (2008). Intermediating knowledge exchange between universities and businesses. *Research Policy*, 37 (8): 1167-1174.

Zaheer, S., Schomaker, M. S., & Nachum, L. (2012). Distance without direction: Restoring credibility to a much-loved construct. *Journal of International Business Studies*, *43*(1), 18-27.

Zhang, F. (2010). Emerging Innovators in the Technological Knowledge Accumulation Networks of Multinational Corporations. Retrieved from https://rucore.libraries.rutgers.edu/rutgers-lib/29865/ on February 13, 2017.

Zhang, J. (2016). Facilitating exploration alliances in multiple dimensions: the influences of firm technological knowledge breadth. *R&D Management*,*46*(S1), 159-173.

Zhang, F., Jiang, G., & Cantwell, J. A. (2015). Subsidiary exploration and the innovative performance of large multinational corporations. *International Business Review*, *24*(2), 224-234.

APPENDIX A0 – DATA OVERVIEW

Table 1: Subclass descriptives by primary Tech56 field

Subclasses 1976-1988										
Tech56	CEMT	count patents	Average count patents	# of single subclass patents	% of single subclasses	# of unique primary sub-classes	Average number of unique primary sub-classes	# of unique secondary subclasses	Average # of unique secondary subclasses	Average # of sub-classes per patent
8	C	336	0.23%	0	0.00%	39	0.116	584	1.74	5.89
12	C	26,482	18.35%	3,418	12.91%	1,507	0.057	8,267	0.31	5.72
16	M	29,979	20.78%	1,528	5.10%	3,422	0.114	16,701	0.56	4.61
29	M	44,730	31.00%	2,915	6.52%	4,407	0.099	16,103	0.36	3.79
40	E	8,187	5.67%	68	0.83%	333	0.041	2,111	0.26	4.52
41	E	18,197	12.61%	4,830	26.54%	1,015	0.056	6,332	0.35	3.38
42	T	10,909	7.56%	873	8.00%	487	0.045	2,796	0.26	3.34
43	T	5,468	3.79%	506	9.25%	408	0.075	3,771	0.69	3.46
Aggregate	CEMT	Sum	Average	Sum	Average	Sum	Average	Sum	Average	Average
		144,288	12.50%	14138	8.64%	11618	0.075	56665	0.56	4.34

Subclasses 1989-2001										
Tech56	CEMT	count patents	Average count patents	# of single subclass patents	% of single subclasses	# of unique primary sub-classes	Average number of unique primary sub-classes	# of unique secondary subclasses	Average # of unique secondary subclasses	Average # of sub-classes per patent
8	C	1,020	0.29%	0	0.00%	40	0.039	1,535	1.50	6.11
12	C	95,283	26.68%	4,717	4.95%	2,118	0.022	16,188	0.17	6.80
16	M	46,069	12.90%	1,221	2.65%	4,081	0.089	21,636	0.47	5.25
29	M	56,061	15.70%	2,622	4.68%	4,639	0.083	17,628	0.31	3.82
40	E	28,146	7.88%	333	1.18%	1,169	0.042	5,734	0.20	4.49
41	E	108,609	30.41%	14,261	13.13%	3,420	0.031	18,281	0.17	3.97
42	T	12,776	3.58%	2,425	18.98%	475	0.037	2,777	0.22	2.87
43	T	9,147	2.56%	618	6.76%	449	0.049	4,432	0.48	3.52
Aggregate	CEMT	Sum	Average	Sum	Average	Sum	Average	Sum	Average	Average
		357,111	12.50%	26,197	6.54%	16,391	0.049	88,211	0.44	4.60

Subclasses 2002-2014										
Tech56	CEMT	count patents	Average count patents	# of single subclass patents	% of single subclasses	# of unique primary sub-classes	Average number of unique primary sub-classes	# of unique secondary subclasses	Average # of unique secondary subclasses	Average # of sub-classes per patent
8	C	1,169	0.15%	0	0.00%	40	0.034	2,043	1.75	7.07
12	C	151,996	19.01%	13,752	9.05%	2,053	0.014	15,594	0.10	5.72
16	M	53,368	6.67%	2,293	4.30%	3,720	0.070	22,662	0.42	5.74
29	M	61,713	7.72%	4,830	7.83%	4,363	0.071	17,190	0.28	4.00
40	E	80,649	10.08%	2,976	3.69%	937	0.012	7,684	0.10	4.91
41	E	412,480	51.58%	66,127	16.03%	3,452	0.008	27,314	0.07	4.08
42	T	19,376	2.42%	2,556	13.19%	435	0.022	3,259	0.17	3.18
43	T	18,969	2.37%	2,708	14.28%	443	0.023	4,737	0.25	3.34
Aggregate	CEMT	Sum	Average	Sum	Average	Sum	Average	Sum	Average	Average
	CEMT	799,720	12.50%	95242	8.54%	15443	0.032	100,483	0.39	4.76

Table 2: Primary Subclass CEMT : Secondary Subclass CEMTO

Subclasses table 4											
1976-1988				1989-2001				2002-2014			
	Total CEMTO: reference CEMTO pairs	Total patents per CEMTO field	Average per CEMTO field		Total CEMTO: reference CEMTO pairs	Total patents per CEMTO field	Average per CEMTO field		Total CEMTO: reference CEMTO pairs	Total patents per CEMTO field	Average per CEMTO field
CC	120,902	126,327	95.71%	CC	527,960	557,713	94.67%	CC	620,810	658,035	94.34%
CE	173	126,327	0.14%	CE	1,595	557,713	0.29%	CE	3,158	658,035	0.48%
CM	4,926	126,327	3.90%	CM	25,212	557,713	4.52%	CM	32,549	658,035	4.95%
CO	199	126,327	0.16%	CO	1,797	557,713	0.32%	CO	838	658,035	0.13%
CT	127	126,327	0.10%	CT	1,149	557,713	0.21%	CT	680	658,035	0.10%
EC	713	71,832	0.99%	EC	1,598	419,958	0.38%	EC	6,535	1349853	0.48%
EE	60,202	71,832	83.81%	EE	386,419	419,958	92.01%	EE	1259040	1349853	93.27%
EM	9,545	71,832	13.29%	EM	23,589	419,958	5.62%	EM	54,631	1349853	4.05%
EO	313	71,832	0.44%	EO	2,936	419,958	0.70%	EO	13,452	1349853	1.00%
ET	1,059	71,832	1.47%	ET	5,416	419,958	1.29%	ET	16,195	1349853	1.20%
MC	18,317	231,543	7.91%	MC	32,403	353,997	9.15%	MC	45,527	411,583	11.06%
ME	7,398	231,543	3.20%	ME	10,980	353,997	3.10%	ME	17,132	411,583	4.16%
MM	201,518	231,543	87.03%	MM	304,044	353,997	85.89%	MM	341,379	411,583	82.94%
MO	874	231,543	0.38%	MO	1,558	353,997	0.44%	MO	1,557	411,583	0.38%
MT	3,436	231,543	1.48%	MT	5,012	353,997	1.42%	MT	5,988	411,583	1.45%
TC	163	38,880	0.42%	TC	209	46,970	0.44%	TC	191	74,143	0.26%
TE	1,994	38,880	5.13%	TE	2,583	46,970	5.50%	TE	4,626	74,143	6.24%
TM	10,777	38,880	27.72%	TM	12,397	46,970	26.39%	TM	15,333	74,143	20.68%
TO	370	38,880	0.95%	TO	678	46,970	1.44%	TO	1,128	74,143	1.52%

TT	25,576	38,880	65.78%	TT	31,103	46,970	66.22%	TT	52,865	74,143	71.30%
IntraCEMTO			83.08%				84.70%				85.46%
InterCEMTO			16.92%				15.30%				14.54%

Table 3: Citation descriptives by primary Tech56 field

Citations 1976-1988										
Tech56	CEMT	count patents	Average count patents	# of single citation patents	% of single citations	# of unique primary citations	Average number of unique primary citations	# of unique secondary citations	Average # of unique secondary citations	Average # of citations per patent
8	C	251	0.26%	51	20.32%	35	0.139	403	1.61	4.97
12	C	16,144	16.95%	5,919	36.66%	1,364	0.084	3,792	0.23	3.94
16	M	17,980	18.87%	6,180	34.37%	2,911	0.162	9,154	0.51	3.95
29	M	27,780	29.16%	9,423	33.92%	3,628	0.131	9,573	0.34	3.95
40	E	5,941	6.24%	1,409	23.72%	308	0.052	1,664	0.28	4.46
41	E	15,574	16.35%	3,045	19.55%	952	0.061	5,742	0.37	5.13
42	T	8,186	8.59%	1,760	21.50%	457	0.056	1,824	0.22	4.67
43	T	3,404	3.57%	1,230	36.13%	354	0.104	1,770	0.52	3.62
Aggregate	CEMT	Sum	Average	Sum	Average	Sum	Average	Sum	Average	Average
		95,260	12.50%	29,017	28.27%	10,009	0.099	33,922	0.51	4.33

Citations 1989-2002										
Tech56	CEMT	count patents	Average count patents	# of single citation patents	% of single citations	# of unique primary citations	Average number of unique primary citations	# of unique secondary citations	Average # of unique secondary citations	Average # of citations per patent
8	C	988	0.30%	63	6.38%	40	0.040	2,271	2.30	10.34
12	C	77,600	23.53%	16,081	20.72%	2,069	0.027	14,935	0.19	7.43
16	M	43,807	13.28%	4,238	9.67%	4,000	0.091	25,342	0.58	8.70
29	M	53,049	16.09%	5,664	10.68%	4,524	0.085	21,648	0.41	7.27
40	E	27,210	8.25%	2,464	9.06%	1,163	0.043	8,027	0.30	8.22
41	E	106,002	32.15%	6,449	6.08%	3,399	0.032	22,842	0.22	9.60
42	T	12,334	3.74%	810	6.57%	469	0.038	4,549	0.37	8.04
43	T	8,763	2.66%	799	9.12%	443	0.051	5,927	0.68	7.71
Aggregate	CEMT	Sum	Average	Sum	Average	Sum	Average	Sum	Average	Average
		329,753	12.50%	36,568	9.78%	16,107	0.051	105,541	0.63	8.41

Citations 2002-2014										
Tech56	CEMT	count patents	Average count patents	# of single citation patents	% of single citations	# of unique primary citations	Average number of unique primary citations	# of unique secondary citations	Average # of unique secondary citations	Average # of citations per patent
8	C	1,164	0.16%	89	7.65%	40	0.034	4,822	4.14	20.66
12	C	127,449	17.32%	19,584	15.37%	2,027	0.016	25,073	0.20	15.60
16	M	51,485	7.00%	2,862	5.56%	3,714	0.072	38,510	0.75	18.33
29	M	59,499	8.09%	3,145	5.29%	4,375	0.074	32,893	0.55	12.76
40	E	74,637	10.14%	8,188	10.97%	934	0.013	17,297	0.23	15.89
41	E	385,422	52.37%	29,856	7.75%	3,502	0.009	44,212	0.11	17.66
42	T	17,930	2.44%	1,488	8.30%	437	0.024	8,129	0.45	11.18
43	T	18,323	2.49%	789	4.31%	445	0.024	13,636	0.74	14.68
Aggregate	CEMT	Sum	Average	Sum	Average	Sum	Average	Sum	Average	Average
		735,909	12.50%	66,001	8.15%	15,474	0.033	184,572	0.90	15.84

Table 4: Citing and Cited Patents by Tech56

Citing Patents by Primary Tech56 Field			
	1976-1988	1989-2001	2002-2014
tech56	N(_freq)	N(_freq)	N(_freq)
8	35	40	40
12	1,364	2,069	2,027
16	2,911	4,000	3,714
29	3,628	4,524	4,375
40	308	1,163	934
41	952	3,399	3,502
42	457	469	437
43	354	443	445
Cited Patents by Primary Tech56 Field			
	1976-1988	1989-2001	2002-2014
tech56	N(_freq)	N(_freq)	N(_freq)
8	403	2,271	4,822
12	3,792	14,935	25,073
16	9,154	25,342	38,510
29	9,573	21,648	32,893
40	1,664	8,027	17,297
41	5,742	22,842	44,212
42	1,824	4,549	8,129
43	1,770	5,927	13,636

Table 5: Citing subclass CEMT : Cited subclass CEMTO

Citations table 4											
1976-1988				1989-2001				2002-2014			
	Total CEMTO: reference CEMTO pairs	Total patents per CEMTO field	Average per CEMTO field		Total CEMTO: reference CEMTO pairs	Total patents per CEMTO field	Average per CEMTO field		Total CEMTO: reference CEMTO pairs	Total patents per CEMTO field	Average per CEMTO field
CC	44,325	48,447	91.49%	CC	437,562	508,142	86.11%	CC	1575613	1883236	83.67%
CE	235	48,447	0.49%	CE	5,856	508,142	1.15%	CE	38,794	1883236	2.06%
CM	3,670	48,447	7.58%	CM	59,916	508,142	11.79%	CM	255,262	1883236	13.55%
CO	101	48,447	0.21%	CO	1,994	508,142	0.39%	CO	5,660	1883236	0.30%
CT	116	48,447	0.24%	CT	2,814	508,142	0.55%	CT	7,907	1883236	0.42%
EC	1,015	84,903	1.20%	EC	12,812	1108513	1.16%	EC	112,470	7530896	1.49%
EE	65,402	84,903	77.03%	EE	968,199	1108513	87.34%	EE	6620323	7530896	87.91%
EM	16,074	84,903	18.93%	EM	95,870	1108513	8.65%	EM	554,641	7530896	7.36%
EO	348	84,903	0.41%	EO	18,477	1108513	1.67%	EO	188,586	7530896	2.50%
ET	2,064	84,903	2.43%	ET	13,155	1108513	1.19%	ET	54,876	7530896	0.73%
MC	18,061	134,840	13.39%	MC	97,492	669,762	14.56%	MC	274,619	1591946	17.25%
ME	5,733	134,840	4.25%	ME	37,961	669,762	5.67%	ME	152,377	1591946	9.57%
MM	107,555	134,840	79.76%	MM	513,131	669,762	76.61%	MM	1109781	1591946	69.71%
MO	416	134,840	0.31%	MO	3,520	669,762	0.53%	MO	10,665	1591946	0.67%
MT	3,075	134,840	2.28%	MT	17,658	669,762	2.64%	MT	44,504	1591946	2.80%
TC	229	38,924	0.59%	TC	1,131	669,762	0.17%	TC	5,230	1591946	0.33%
TE	1,966	38,924	5.05%	TE	8,666	145,580	5.95%	TE	37,976	433,110	8.77%
TM	5,370	38,924	13.80%	TM	28,088	145,580	19.29%	TM	102,417	433,110	23.65%

TO	159	38,924	0.41%	TO	1,390	145,580	0.95%	TO	8,241	433,110	1.90%
TT	31,200	38,924	80.16%	TT	106,305	145,580	73.02%	TT	279,246	433,110	64.47%
IntraCEMTO			82.11%				80.77%				76.44%
InterCEMTO			17.89%				19.08%				23.34%

Table 6: Location descriptives by primary Tech56 field

1976-1988										
Tech56	CEMT	count patents	Average count patents	# of single location patents	% of single location patents	# of unique primary locations	Average number of unique primary locations	# of unique secondary locations	Average # of unique secondary locations	Average # of locations per patent
8	C	280	0.28%	63	22.50%	15	0.054	20	0.071	5.07
12	C	17,157	17.44%	5,784	33.71%	48	0.003	59	0.003	4.17
16	M	18,603	18.91%	6,147	33.04%	57	0.003	67	0.004	4.07
29	M	28,335	28.80%	9,231	32.58%	64	0.002	75	0.003	4.07
40	E	6,001	6.10%	1,345	22.41%	26	0.004	35	0.006	4.62
41	E	16,224	16.49%	2,582	15.91%	46	0.003	52	0.003	5.70
42	T	8,285	8.42%	1,689	20.39%	43	0.005	40	0.005	4.85
43	T	3,509	3.57%	1,186	33.80%	29	0.008	36	0.010	3.77
Aggregate	CEMT	Sum	Average	Sum	Average	Sum	Average	Sum	Average	Average
		98,394	12.50%	28,027	26.79%	328	0.010	384	0.013	4.54

1989-2001										
Tech56	CEMT	count patents	Average count patents	# of single location patents	% of single location patents	# of unique primary locations	Average number of unique primary locations	# of unique secondary locations	Average # of unique secondary locations	Average # of locations per patent
8	C	990	0.30%	48	4.85%	25	0.025	42	0.042	10.93
12	C	78,963	23.71%	15,109	19.13%	92	0.001	97	0.001	7.95
16	M	44,077	13.24%	3,845	8.72%	76	0.002	101	0.002	9.16
29	M	53,416	16.04%	5,147	9.64%	79	0.001	92	0.002	7.61
40	E	27,287	8.19%	2,289	8.39%	43	0.002	61	0.002	8.53
41	E	107,042	32.15%	4,630	4.33%	73	0.001	93	0.001	10.84
42	T	12,397	3.72%	664	5.36%	54	0.004	66	0.005	8.67
43	T	8,823	2.65%	677	7.67%	38	0.004	55	0.006	8.20
Aggregate	CEMT	Sum	Average	Sum	Average	Sum	Average	Sum	Average	Average
		332,995	12.50%	32,409	8.51%	480	0.005	607	0.008	8.99

2002-2014										
Tech56	CEMT	count patents	Average count patents	# of single location patents	% of single location patents	# of unique primary locations	Average number of unique primary locations	# of unique secondary locations	Average # of unique secondary locations	Average # of locations per patent
8	C	1,170	0.16%	82	7.01%	33	0.028	51	0.044	21.70
12	C	128,393	17.36%	18,960	14.77%	110	0.001	128	0.001	16.34
16	M	51,617	6.98%	2,697	5.23%	78	0.002	126	0.002	19.13
29	M	59,655	8.07%	2,777	4.66%	81	0.001	124	0.002	13.38
40	E	74,855	10.12%	7,986	10.67%	59	0.001	82	0.001	16.30
41	E	387,392	52.39%	26,970	6.96%	112	0.000	139	0.000	19.00
42	T	18,047	2.44%	1,279	7.09%	60	0.003	85	0.005	12.12
43	T	18,363	2.48%	699	3.81%	53	0.003	78	0.004	15.47
Aggregate	CEMT	Sum	Average	Sum	Average	Sum	Average	Sum	Average	Average
		739,492	12.50%	61,450	7.52%	586	0.005	813	0.007	16.68

Table 7: Continent : referenceContinent

Locations									
	1976-1988			1989-2001			2002-2014		
Continent: reference - Continent	Total Continent: reference-Continent pairs	Total patents per Continent field	Average per Continent	Total Continent: reference-Continent pairs	Total patents per Continent field	Average per Continent	Total Continent: reference-Continent pairs	Total patents per Continent field	Average per Continent
AfricaAfrica	38	361	10.53%	95	1,430	6.64%	302	5,370	5.62%
AfricaAsia	26	361	7.20%	143	1,430	10.00%	488	5,370	9.09%
AfricaAustralia	6	361	1.66%	18	1,430	1.26%	79	5,370	1.47%
AfricaEurope	66	361	18.28%	301	1,430	21.05%	605	5,370	11.27%
Africa - North America	221	361	61.22%	868	1,430	60.70%	3,893	5,370	72.50%
Africa - South America	4	361	1.11%	5	1,430	0.35%	3	5,370	0.06%
AsiaAfrica	34	72,521	0.05%	115	519,809	0.02%	528	1688031	0.03%
AsiaAsia	31,925	72,521	44.02%	256,381	519,809	49.32%	760,973	1688031	45.08%
AsiaAustralia	116	72,521	0.16%	885	519,809	0.17%	4,327	1688031	0.26%
AsiaEurope	12,633	72,521	17.42%	55,755	519,809	10.73%	132,367	1688031	7.84%
Asia - North America	27,785	72,521	38.31%	206,512	519,809	39.73%	789,309	1688031	46.76%
Asia - South America	28	72,521	0.04%	161	519,809	0.03%	527	1688031	0.03%
AustraliaAfrica	6	1,394	0.43%	30	10,260	0.29%	65	72,447	0.09%
AustraliaAsia	113	1,394	8.11%	1,358	10,260	13.24%	7,351	72,447	10.15%
Australia - Australia	107	1,394	7.68%	522	10,260	5.09%	4,774	72,447	6.59%
AustraliaEurope	282	1,394	20.23%	1,741	10,260	16.97%	8,283	72,447	11.43%
AustraliaNorth America	886	1,394	63.56%	6,583	10,260	64.16%	51,891	72,447	71.63%

AustraliaSouth America	0	1,394	0.00%	26	10,260	0.25%	83	72,447	0.11%
Europe Africa	82	70,427	0.12%	219	314,551	0.07%	528	1044538	0.05%
Europe Asia	9,250	70,427	13.13%	54,894	314,551	17.45%	155,620	1044538	14.90%
Europe Australia	214	70,427	0.30%	1,079	314,551	0.34%	4,723	1044538	0.45%
Europe Europe	29,513	70,427	41.91%	106,268	314,551	33.78%	252,601	1044538	24.18%
Europe North America	31,323	70,427	44.48%	151,906	314,551	48.29%	630,281	1044538	60.34%
Europe South America	45	70,427	0.06%	185	314,551	0.06%	785	1044538	0.08%
North America Africa	148	194,157	0.08%	1,179	1840698	0.06%	4,477	9434034	0.05%
North America Asia	19,423	194,157	10.00%	261,501	1840698	14.21%	1136366	9434034	12.05%
North America Australia	628	194,157	0.32%	6,224	1840698	0.34%	39,464	9434034	0.42%
North America Europe	30,393	194,157	15.65%	216,398	1840698	11.76%	830,939	9434034	8.81%
North America North America	143,443	194,157	73.88%	1354487	1840698	73.59%	7417638	9434034	78.63%
North America South America	122	194,157	0.06%	909	1840698	0.05%	5,150	9434034	0.05%
South America Africa	0	267	0.00%	7	2,056	0.34%	15	7,611	0.20%
South America Asia	25	267	9.36%	277	2,056	13.47%	810	7,611	10.64%
South America Australia	1	267	0.37%	15	2,056	0.73%	91	7,611	1.20%
South America Europe	42	267	15.73%	423	2,056	20.57%	1,160	7,611	15.24%
South America North America	181	267	67.79%	1,261	2,056	61.33%	5,314	7,611	69.82%
South America	18	267	6.74%	73	2,056	3.55%	221	7,611	2.90%

South America									
IntraContinent			30.79%			28.66%			27.17%
InterContinent			69.21%			71.34%			72.83%

Table 8: Weights of Continent to Reference Continent

Paired Continents			
continent_rcontinent	1976-1988	1989-2001	2002-2014
AfricaAfrica	0.0209%	0.0076%	0.0077%
AfricaAsia	0.0105%	0.0134%	0.0132%
AfricaAustralia	0.0039%	0.0022%	0.0027%
AfricaEurope	0.0301%	0.0220%	0.0159%
AfricaNorth America	0.0608%	0.0334%	0.0303%
AfricaSouth America	0.0026%	0.0008%	0.0002%
AsiaAfrica	0.0216%	0.0169%	0.0266%
AsiaAsia	8.9131%	11.1508%	10.5271%
AsiaAustralia	0.0719%	0.1237%	0.2076%
AsiaEurope	5.1246%	4.8696%	4.2492%
AsiaNorth America	8.4099%	9.7581%	9.9472%
AsiaSouth America	0.0183%	0.0226%	0.0308%
AustraliaAfrica	0.0039%	0.0034%	0.0036%
AustraliaAsia	0.0503%	0.0841%	0.1386%
AustraliaAustralia	0.0497%	0.0544%	0.1051%
AustraliaEurope	0.1202%	0.1214%	0.1655%
AustraliaNorth America	0.2346%	0.2130%	0.2798%
AustraliaSouth America	0.0000%	0.0039%	0.0048%
Europe Africa	0.0523%	0.0312%	0.0330%
Europe Asia	3.8884%	3.7438%	3.4391%
Europe Australia	0.1353%	0.1513%	0.2487%
Europe Europe	10.3022%	6.2156%	4.7527%
Europe North America	9.9716%	6.6824%	6.0113%
Europe South America	0.0294%	0.0266%	0.0428%
North America Africa	0.0895%	0.1580%	0.2394%
North America Asia	7.9330%	13.9997%	15.5862%
North America Australia	0.3914%	0.7937%	1.7269%
North America Europe	12.5427%	13.7431%	14.0258%
North America North America	31.3496%	27.7121%	27.7572%
North America South America	0.0751%	0.1272%	0.2635%
South AmericaAfrica	0.0000%	0.0011%	0.0005%
South America Asia	0.0118%	0.0213%	0.0238%
South America Australia	0.0007%	0.0022%	0.0026%
South America Europe	0.0196%	0.0311%	0.0299%
South America North	0.0510%	0.0489%	0.0508%

America			
South America South America	0.0098%	0.0097%	0.0099%
Aggregate			
Intracontinent	50.645%	45.150%	43.160%
Intercontinent	49.355%	54.860%	56.840%

Table 9: Top five patenting countries per time period

	1976-1988		1989-2001		2002-2014	
5	France	3.447%	Great Britain	2.607%	Taiwan	2.727%
4	Great Britain	4.046%	France	2.627%	South Korea	3.897%
3	Germany	10.121%	Germany	6.468%	Germany	5.534%
2	Japan	19.981%	Japan	22.805%	Japan	17.124%
1	United States	52.624%	United States	54.196%	United States	55.740%

Table 10: Location Complexity by Citing Country

	tech56	1976-1988 complexity(avg)	1989-2001 complexity(avg)	2002-2014 complexity(avg)
United States	8	21.18628	51.42715	111.3237
	12	20.09301	40.01593	83.02827
	16	21.13675	48.09929	113.9007
	29	18.8006	39.35887	81.22353
	40	21.71449	48.0844	109.3514
	41	25.57295	54.34984	95.39241
	42	27.96604	54.89087	88.47143
	43	18.69941	43.32477	92.51512
Japan	8	48.09307	69.21205	58.88881
	12	53.16281	56.62749	76.65489
	16	60.44552	73.7872	98.56763
	29	63.31183	70.60284	85.31982
	40	56.4829	63.80584	104.3226
	41	66.34336	76.91763	88.57243
	42	66.92973	77.43723	82.55505
	43	51.73136	69.36327	105.122
Germany	8	40.35498	129.3386	124.0534
	12	47.52963	94.39323	188.1478
	16	47.83872	107.8882	195.379
	29	54.04876	102.7194	160.1532
	40	51.01471	107.9108	220.2325
	41	66.05443	168.7566	225.9482
	42	58.35785	129.4775	156.5494
	43	38.99068	109.2649	199.2178

APPENDIX A1 – STUDY 1

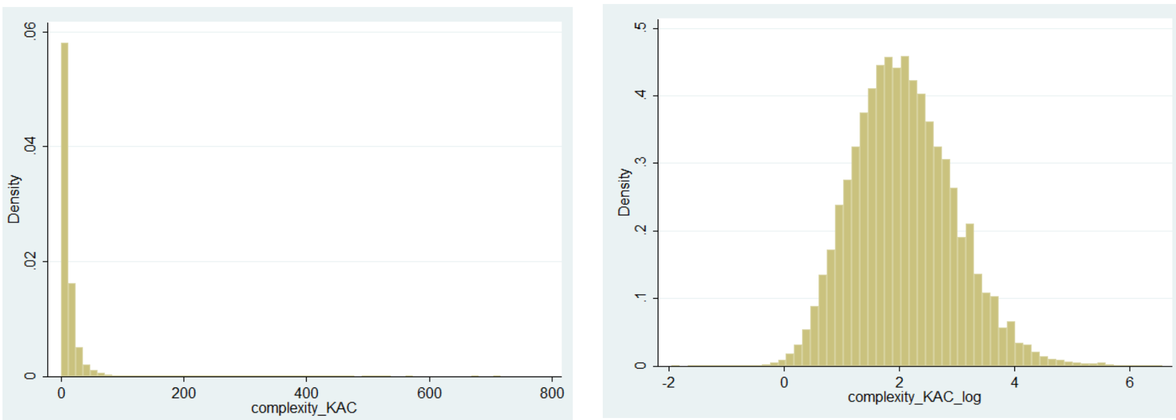
Reference for Variable Names and Definitions

Table 1: Complete List of Variable Names and Definitions

Variable Name	Definition	DV, IV, Control
Knowledge Artifact Complexity	Logged degree of technological complexity via co-classifications.	DV
Knowledge Sourcing Complexity	Logged degree of technological complexity via cross-classifications.	DV
Location Sourcing Complexity	Logged degree of technological complexity via location cross-classifications.	DV
Complexity_KAC	Degree of technological complexity via co-classifications.	DV
Complexity_KSC	Degree of technological complexity via cross-classifications.	DV
Complexity_LSC	Degree of technological complexity via location cross-classifications.	DV
TechDiversification_CoClass	A simple measure of technological diversification as a proportion of technologies per patent, calculated from co-classification (subclass) data.	IV
TechDiversificationSqd_CoClas	Squared, simple measure of technological diversification as a proportion of technologies per patent, calculated from co-classification data.	IV
TechDistinctiveness_CoClass	A sophisticated measure of technological distinctiveness via the likelihood of linkages between technology *i* and *j*, calculated from co-classification data.	IV
TechDistinctivenessSqd_CoClass	Squared, sophisticated measure of technological distinctiveness via the likelihood of linkages between technology *i* and *j*, calculated from co-classification data.	IV
TechDiversification_CrossClas	A simple measure of technological diversification as a proportion of technologies per patent, calculated from cross-classification (citation) data.	IV
TechDiversificationSqd_CrossClas	Squared, simple measure of technological	IV

	diversification as a proportion of technologies per patent, calculated from cross-classification data.	
TechDistinctiveness _CrossClas	A sophisticated measure of technological distinctiveness via the likelihood of linkages between technology *i* and *j*, calculated from cross-classification data.	IV
TechDistinctivenessSqd _CrossClas	Squared, sophisticated measure of technological distinctiveness via the likelihood of linkages between technology *i* and *j*, calculated from cross-classification data.	IV
ICT_sharePerField	The percentage of ICT classes in a given Tech56 field	IV
Interaction KACxICT	The interaction of Knowledge Artifact Complexity (logged form) and ICT_sharePerField.	IV
Interaction KSCxICT	The interaction of Knowledge Sourcing Complexity (logged form) and ICT_sharePerField.	IV
N_Subclasses	Number of subclasses on focal patent	Control
N_Citations	Number of parent patents, for subclasses	Control
N_LocationCitation	Number of parent patents, for locations	Control
Ctrl_NumUniqueClasses_Citation	A count of the number of unique Tech56 patent classes per patent from cross-classification data.	Control
Ctrl_NumUniqueClasses_Subclass	A count of the number of unique Tech56 patent classes per patent from co-classification data.	Control
Control_NumofTrials_Subclass	The number of times the same set of subclasses has appeared on other patents in the database.	Control
Ctrl_ProbFieldAcitesB_Subclass	Probability Tech56 Field A will cite Tech56 Field B.	Control
Degree of Country Connectivity	Probability citing Country A cites cited location Country B.	Control
Country Distance	Log of the sum of the miles from citing to cited countries between capital cities.	Control
T1	Dummy indicating time period 1, 1976-1988	Control
T2	Dummy indicating time period 2, 1989-2001	Control
T3	Dummy indicating time period 3, 2002-2014	Control

Graph 2: Raw KAC Data vs Normalized KAC data



Graph 3: Raw KSC Data vs Normalized KSC data

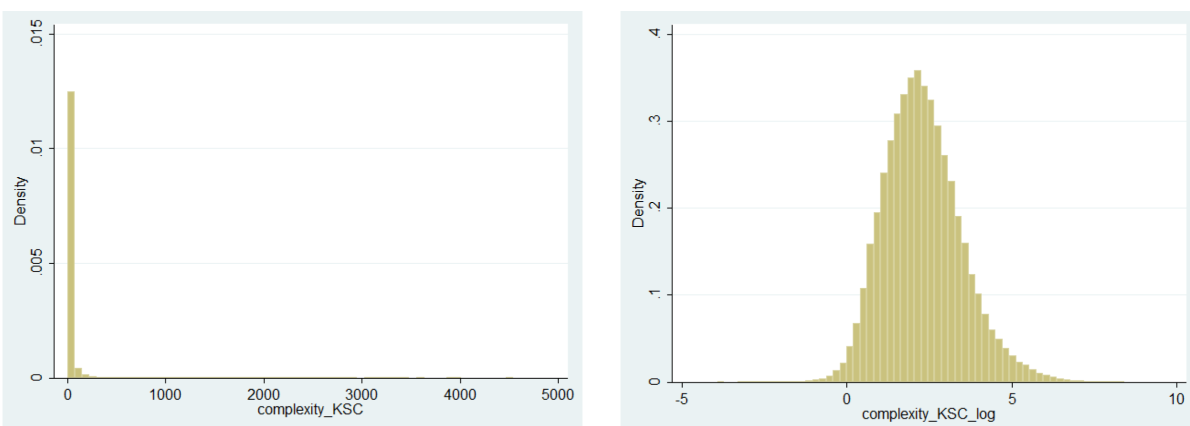


Table 4: KAC Descriptive Statistics, Period 1

```
. summarize KnowledgeArtifactComplexity TechDiversification_CoClass TechDiversifica
> tionSqd_CoClass N_Subclasses TechDistinctiveness_CoClass TechDistinctivenessSqd_C
> oClass ICT_SharePerField N_Citations Control_NumberOfTrials_Subclass Ctrl_NumUniq
> ueClasses_Citation Ctrl_ProbFieldAcitesB_Subclass
```

Variable	Obs	Mean	Std. Dev.	Min	Max
KnowledgeA~y	94931	1.560074	.6727414	-1.358123	4.727853
Te~n_CoClass	94931	.2426053	.2229162	.0429441	.9989946
TechDivers..	94931	.1085484	.1990324	.0018442	.9979903
N_Subclasses	94931	4.371533	3.430617	2	215
Te~s_CoClass	94931	2.587912	3.162761	0	14.9
TechDistin..	94931	16.70024	29.35331	0	222.01
ICT_ShareP~d	94931	.1848164	.3355027	0	.8528234
N_Citations	94931	4.22845	2.813293	2	98
Control_Nu~s	94931	3.116843	13.50072	1	749
Ctrl_NumUn~n	94931	1.781673	.902079	1	12
Ctrl_ProbF~s	94931	.1991886	.1006449	.0037681	.3983796

Table 5: KAC Descriptive Statistics, Period 2

```
. summarize KnowledgeArtifactComplexity TechDiversification_CoClass TechDiversifica
> tionSqd_CoClass N_Subclasses TechDistinctiveness_CoClass TechDistinctivenessSqd_C
> oClass ICT_SharePerField N_Citations Control_NumberOfTrials_Subclass Ctrl_NumUniq
> ueClasses_Citation Ctrl_ProbFieldAcitesB_Subclass
```

Variable	Obs	Mean	Std. Dev.	Min	Max
KnowledgeA~y	329288	1.908707	.8136922	-1.403389	5.750908
Te~n_CoClass	329288	.1818846	.2156498	.0533482	.9979398
TechDivers..	329288	.0795867	.1952841	.002846	.9958839
N_Subclasses	329288	4.803843	3.898909	2	208
Te~s_CoClass	329288	1.840482	2.856037	0	15.5
TechDistin..	329288	11.5443	25.72533	0	240.25
ICT_ShareP~d	329288	.3682409	.4418591	.000408	.9114777
N_Citations	329288	8.377114	10.10707	2	634
Control_Nu~s	329288	6.67366	34.99919	1	634
Ctrl_NumUn~n	329288	2.188039	1.390765	1	31
Ctrl_ProbF~s	329288	.2576448	.1128716	.0033211	.3983796

Table 6: KAC Descriptive Statistics, Period 3

```
. summarize KnowledgeArtifactComplexity TechDiversification_CoClass TechDiversifica
> tionSqd_CoClass N_Subclasses TechDistinctiveness_CoClass TechDistinctivenessSqd_C
> oClass ICT_SharePerField N_Citations Control_NumberOfTrials_Subclass Ctrl_NumUniq
> ueClasses_Citation Ctrl_ProbFieldAcitesB_Subclass
```

Variable	Obs	Mean	Std. Dev.	Min	Max
KnowledgeA~y	657021	2.264086	.8855995	-1.94591	6.574169
Te~n_CoClass	657021	.1556963	.2096168	.0565699	.9989666
TechDivers..	657021	.0681805	.193726	.0032002	.9979343
N_Subclasses	657021	4.552115	3.338053	2	177
Te~s_CoClass	657021	1.500094	2.664883	0	15.5
TechDistin..	657021	9.351874	23.79525	0	240.25
ICT_ShareP~d	657021	.5592496	.4411181	.0017501	.925244
N_Citations	657021	16.54789	36.1849	2	5322
Control_Nu~s	657021	16.14812	97.40364	1	1452
Ctrl_NumUn~n	657021	2.626972	2.120079	1	43
Ctrl_ProbF~s	657021	.2904507	.1137086	.0033197	.3983796

Table 7: KAC Correlation Table, Period 1, Technological Diversification – Co-classification

```
. corr KnowledgeArtifactComplexity TechDiversification_CoClass TechDiversificationSqd_CoClass ICT_S
> harePerField N_Subclasses N_Citations Control_NumberOfTrials_Subclass Ctrl_NumUniqueClasses_Citat
> ion Ctrl_ProbFieldAcitesB_Subclass
(obs=94931)
```

	K~Arti~y	TechDi..	TechDi..	ICT_Sh~d	N_Subc~s	N_Cita~s	Contro~s	Ctrl_N~n	Ctrl_P~s
KnowledgeA~y	1.0000								
Te~n_CoClass	0.0092	1.0000							
TechDivers..	-0.0307	0.9615	1.0000						
ICT_ShareP~d	0.2093	0.0786	0.0445	1.0000					
N_Subclasses	0.4101	-0.0577	-0.0484	-0.0978	1.0000				
N_Citations	0.1230	0.0829	0.0747	0.1369	0.0586	1.0000			
Control_Nu~s	0.0213	-0.0195	-0.0146	-0.0271	-0.0996	-0.0139	1.0000		
Ctrl_NumUn~n	0.0027	0.1951	0.1882	0.1639	0.1039	0.4968	-0.0530	1.0000	
Ctrl_ProbF~s	0.0407	-0.3517	-0.2519	0.4389	0.0738	0.0464	-0.0121	0.0678	1.0000

Table 8: KAC Correlation Table, Period 2, Technological Diversification – Co-classification

```
. corr KnowledgeArtifactComplexity TechDiversification_CoClass TechDiversificationSqd_CoClass ICT_S
> harePerField N_Subclasses N_Citations Control_NumberOfTrials_Subclass Ctrl_NumUniqueClasses_Citat
> ion Ctrl_ProbFieldAcitesB_Subclass
(obs=329288)
```

	K~Arti~y	TechDi..	TechDi..	ICT_Sh~d	N_Subc~s	N_Cita~s	Contro~s	Ctrl_N~n	Ctrl_P~s
KnowledgeA~y	1.0000								
Te~n_CoClass	-0.1036	1.0000							
TechDivers..	-0.0835	0.9669	1.0000						
ICT_ShareP~d	0.1376	-0.1345	-0.0647	1.0000					
N_Subclasses	0.4913	-0.0318	-0.0329	-0.1569	1.0000				
N_Citations	0.0307	0.0350	0.0474	0.0777	0.0370	1.0000			
Control_Nu~s	0.0702	-0.0024	-0.0138	0.0379	-0.0980	-0.0032	1.0000		
Ctrl_NumUn~n	-0.0473	0.2122	0.2007	-0.0054	0.0726	0.5633	-0.0575	1.0000	
Ctrl_ProbF~s	0.2640	-0.4156	-0.2866	0.4889	0.0655	0.0377	0.0060	-0.1424	1.0000

Table 9: KAC Correlation Table, Period 3, Technological Diversification – Co-classification

```
. corr KnowledgeArtifactComplexity TechDiversification_CoClass TechDiversificationSqd_CoClass ICT_S
> harePerField N_Subclasses N_Citations Control_NumberOfTrials_Subclass Ctrl_NumUniqueClasses_Citat
> ion Ctrl_ProbFieldAcitesB_Subclass
(obs=657021)
```

	K~Arti~y	TechDi..	TechDi..	ICT_Sh~d	N_Subc~s	N_Cita~s	Contro~s	Ctrl_N~n	Ctrl_P~s
KnowledgeA~y	1.0000								
Te~n_CoClass	-0.1601	1.0000							
TechDivers..	-0.1247	0.9698	1.0000						
ICT_ShareP~d	0.2225	-0.2070	-0.1098	1.0000					
N_Subclasses	0.4892	0.0397	0.0099	-0.1333	1.0000				
N_Citations	0.0210	0.0065	0.0221	0.0327	0.0241	1.0000			
Control_Nu~s	0.1307	-0.0458	-0.0399	0.0608	-0.0872	0.0041	1.0000		
Ctrl_NumUn~n	-0.0897	0.1874	0.1762	-0.0677	0.0828	0.6412	-0.0427	1.0000	
Ctrl_ProbF~s	0.2816	-0.4086	-0.2874	0.5071	-0.0462	0.0280	0.0782	-0.1894	1.0000

Table 10:
Knowledge Artifact Complexity regressions,
Technological Diversification in time period 1, 2, and 3

	(1) KnowledgeA~y	(2) KnowledgeA~y	(3) KnowledgeA~y
Te~n_CoClass	1.010***	0.718***	-1.194***
	(27.74)	(27.92)	(-58.27)
TechDivers..	-1.165***	-0.728***	1.040***
	(-30.34)	(-27.00)	(49.15)
ICT_ShareP~d	0.579***	0.254***	0.354***
	(83.29)	(81.67)	(154.19)
N_Subclasses	0.0906***	0.108***	0.146***
	(162.94)	(348.60)	(560.01)
N_Citations	0.0320***	0.00338***	0.00189***
	(41.98)	(23.96)	(60.24)
Control_Nu~s	0.00346***	0.00251***	0.00135***
	(25.03)	(74.92)	(152.98)
Ctrl_NumUn~n	-0.112***	-0.0477***	-0.0539***
	(-45.76)	(-45.24)	(-97.80)
Ctrl_ProbF~s	-0.564***	1.282***	1.007***
	(-21.84)	(87.42)	(98.24)
_cons	1.104***	0.953***	1.314***
	(121.49)	(161.02)	(290.89)
N	94931	329288	657021
R-sq	0.278	0.327	0.394
f			

t statistics in parentheses

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

Table 11: KAC Correlation Table, Period 1 Technological Distinctiveness – Co-Classification

```
. corr KnowledgeArtifactComplexity TechDistinctiveness_CoClas TechDistinctivenessSqd_CoClas ICT_Sha
> rePerField N_Subclasses N_Citations Control_NumberOfTrials_Subclass Ctrl_NumUniqueClasses_Citatio
> n Ctrl_ProbFieldAcitesB_Subclass
(obs=94931)
```

	K~Arti~y	TechDi..	TechDi..	ICT_Sh~d	N_Subc~s	N_Cita~s	Contro~s	Ctrl_N~n	Ctrl_P~s
KnowledgeA~y	1.0000								
Te~s_CoClass	-0.0549	1.0000							
TechDistin..	-0.0886	0.9349	1.0000						
ICT_ShareP~d	0.2093	0.0103	0.0083	1.0000					
N_Subclasses	0.4101	0.0571	-0.0199	-0.0978	1.0000				
N_Citations	0.1230	0.0379	0.0259	0.1369	0.0586	1.0000			
Control_Nu~s	0.0213	-0.0485	-0.0362	-0.0271	-0.0996	-0.0139	1.0000		
Ctrl_NumUn~n	0.0027	0.3017	0.2364	0.1639	0.1039	0.4968	-0.0530	1.0000	
Ctrl_ProbF~s	0.0407	-0.2421	-0.2458	0.4389	0.0738	0.0464	-0.0121	0.0678	1.0000

Table 12: KAC Correlation Table, Period 2, Technological Distinctiveness – Co-Classification

```
. corr KnowledgeArtifactComplexity TechDistinctiveness_CoClas TechDistinctivenessSqd_CoClas ICT_Sha
> rePerField N_Subclasses N_Citations Control_NumberOfTrials_Subclass Ctrl_NumUniqueClasses_Citatio
> n Ctrl_ProbFieldAcitesB_Subclass
(obs=329288)
```

	K~Arti~y	TechDi..	TechDi..	ICT_Sh~d	N_Subc~s	N_Cita~s	Contro~s	Ctrl_N~n	Ctrl_P~s
KnowledgeA~y	1.0000								
Te~s_CoClass	-0.0765	1.0000							
TechDistin..	-0.1026	0.9341	1.0000						
ICT_ShareP~d	0.1376	-0.1312	-0.0886	1.0000					
N_Subclasses	0.4913	0.0655	-0.0101	-0.1569	1.0000				
N_Citations	0.0307	0.0291	0.0261	0.0777	0.0370	1.0000			
Control_Nu~s	0.0702	-0.0720	-0.0457	0.0379	-0.0980	-0.0032	1.0000		
Ctrl_NumUn~n	-0.0473	0.2818	0.2343	-0.0054	0.0726	0.5633	-0.0575	1.0000	
Ctrl_ProbF~s	0.2640	-0.3779	-0.3301	0.4889	0.0655	0.0377	0.0060	-0.1424	1.0000

Table 13: KAC Correlation Table, Period 3, Technological Distinctiveness – Co-Classification

```
. corr KnowledgeArtifactComplexity TechDistinctiveness_CoClas TechDistinctivenessSqd_CoClas ICT_Sha
> rePerField N_Subclasses N_Citations Control_NumberOfTrials_Subclass Ctrl_NumUniqueClasses_Citatio
> n Ctrl_ProbFieldAcitesB_Subclass
(obs=657021)
```

	K~Arti~y	TechDi..	TechDi..	ICT_Sh~d	N_Subc~s	N_Cita~s	Contro~s	Ctrl_N~n	Ctrl_P~s
KnowledgeA~y	1.0000								
Te~s_CoClass	-0.1318	1.0000							
TechDistin..	-0.1477	0.9325	1.0000						
ICT_ShareP~d	0.2225	-0.1547	-0.1113	1.0000					
N_Subclasses	0.4892	0.1087	0.0234	-0.1333	1.0000				
N_Citations	0.0210	0.0099	0.0092	0.0327	0.0241	1.0000			
Control_Nu~s	0.1307	-0.0751	-0.0532	0.0608	-0.0872	0.0041	1.0000		
Ctrl_NumUn~n	-0.0897	0.2227	0.1895	-0.0677	0.0828	0.6412	-0.0427	1.0000	
Ctrl_ProbF~s	0.2816	-0.4217	-0.3488	0.5071	-0.0462	0.0280	0.0782	-0.1894	1.0000

Table 14: Knowledge Artifact Complexity regressions,
using Technological Distinctiveness measure in time period 1, 2, and 3.

	(1) KnowledgeA~y	(2) KnowledgeA~y	(3) KnowledgeA~y
Te~s_CoClass	-0.00103	0.0129***	-0.0222***
	(-0.59)	(10.54)	(-23.18)
TechDistin..	-0.00201***	-0.00178***	0.000164
	(-10.95)	(-13.62)	(1.60)
ICT_ShareP~d	0.662***	0.257***	0.386***
	(103.97)	(82.46)	(170.09)
N_Subclasses	0.0912***	0.108***	0.146***
	(160.81)	(340.06)	(547.29)
N_Citations	0.0294***	0.00323***	0.00187***
	(38.17)	(22.70)	(59.35)
Control_Nu~s	0.00331***	0.00258***	0.00132***
	(23.89)	(76.74)	(149.76)
Ctrl_NumUn~n	-0.0928***	-0.0461***	-0.0515***
	(-36.17)	(-42.62)	(-92.66)
Ctrl_ProbF~s	-1.054***	1.060***	1.138***
	(-48.58)	(81.39)	(117.19)
_cons	1.316***	1.078***	1.169***
	(215.71)	(261.78)	(349.21)
N	94931	329288	657021
R-sq	0.278	0.326	0.393
f			

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

.

Table 15: KSC Descriptive Statistics, Period 1

```
. summarize KnowledgeSourcingComplexity TechDiversification_CrossClass Tec
> hDiversificationSqd_CrossClas TechDistinctiveness_CrossClas TechDistinct
> ivenessSqd_CrossClas ICT_SharePerField N_Subclasses N_Citations Control_
> NumberOfTrials_Subclass Ctrl_NumUniqueClasses_Subclass Ctrl_ProbFieldAci
> tesB_Subclass
```

Variable	Obs	Mean	Std. Dev.	Min	Max
KnowledgeS~y	94931	1.504951	.7715607	-1.446919	5.225226
TechD~sClass	94931	.3018705	.2251659	.0850827	.9979153
TechDiver~as	94931	.141825	.2224223	.0072391	.9958349
~s_CrossClas	94931	2.932281	3.549385	0	15.5
TechDistin..	94931	21.19627	35.24185	0	240.25
ICT_ShareP~d	94931	.1848164	.3355027	0	.8528234
N_Subclasses	94931	4.371533	3.430617	2	215
N_Citations	94931	4.22845	2.813293	2	98
Control_Nu~s	94931	3.116843	13.50072	1	749
Ctrl_NumUn~s	94931	1.693841	.782093	1	10
Ctrl_ProbF~s	94931	.1991886	.1006449	.0037681	.3983796

Table 16: KSC Descriptive Statistics, Period 2

```
. summarize KnowledgeSourcingComplexity TechDiversification_CrossClass Tec
> hDiversificationSqd_CrossClas TechDistinctiveness_CrossClas TechDistinct
> ivenessSqd_CrossClas ICT_SharePerField N_Subclasses N_Citations Control_
> NumberOfTrials_Subclass Ctrl_NumUniqueClasses_Subclass Ctrl_ProbFieldAci
> tesB_Subclass
```

Variable	Obs	Mean	Std. Dev.	Min	Max
KnowledgeS~y	329288	2.045397	.9705723	-1.89712	7.333428
TechD~sClass	329288	.2702486	.1953408	.1265786	.9983113
TechDiver~as	329288	.1111922	.1757841	.0160221	.9966255
~s_CrossClas	329288	2.579088	3.022692	0	15.7
TechDistin..	329288	15.78833	27.89377	0	246.49
ICT_ShareP~d	329288	.3682409	.4418591	.000408	.9114777
N_Subclasses	329288	4.803843	3.898909	2	208
N_Citations	329288	8.377114	10.10707	2	634
Control_Nu~s	329288	6.67366	34.99919	1	634
Ctrl_NumUn~s	329288	1.528188	.7067478	1	9
Ctrl_ProbF~s	329288	.2576448	.1128716	.0033211	.3983796

Table 17: KSC Descriptive Statistics, Period 3

```
. summarize KnowledgeSourcingComplexity TechDiversification_CrossClass Tec
> hDiversificationSqd_CrossClas TechDistinctiveness_CrossClas TechDistinct
> ivenessSqd_CrossClas ICT_SharePerField N_Subclasses N_Citations Control_
> NumberOfTrials_Subclass Ctrl_NumUniqueClasses_Subclass Ctrl_ProbFieldAci
> tesB_Subclass
```

Variable	Obs	Mean	Std. Dev.	Min	Max
KnowledgeS~y	657021	2.560702	1.240508	-3.912023	8.421334
TechD~sClass	657021	.2609819	.1991833	.1209117	.9969946
TechDiver~as	657021	.1077855	.1750843	.0146196	.9939982
~s_CrossClas	657021	2.550265	2.939142	0	15.7
TechDistin..	657021	15.14239	26.86581	0	246.49
ICT_ShareP~d	657021	.5592496	.4411181	.0017501	.925244
N_Subclasses	657021	4.552115	3.338053	2	177
N_Citations	657021	16.54789	36.1849	2	5322
Control_Nu~s	657021	16.14812	97.40364	1	1452
Ctrl_NumUn~s	657021	1.428531	.6598886	1	13
Ctrl_ProbF~s	657021	.2904507	.1137086	.0033197	.3983796

Table 18: KSC Correlation Table, Period 1, Technological Diversification – Cross-classification

```
. corr KnowledgeSourcingComplexity TechDiversification_CrossClass TechDiversificationSqd_CrossClas
> ICT_SharePerField N_Subclasses N_Citations Control_NumberOfTrials_Subclass Ctrl_NumUniqueClasses_S
> ubclass Ctrl_ProbFieldAcitesB_Subclass
(obs=94931)
```

	K~Sour~y	T~sClass	TechDi..	ICT_Sh~d	N_Subc~s	N_Cita~s	Contro~s	Ctrl_N~s	Ctrl_P~s
KnowledgeS~y	1.0000								
TechD~sClass	-0.0663	1.0000							
TechDiver~as	-0.1094	0.9743	1.0000						
ICT_ShareP~d	0.1904	0.1254	0.0542	1.0000					
N_Subclasses	-0.0506	-0.0270	0.0094	-0.0978	1.0000				
N_Citations	0.5831	0.0945	0.0332	0.1369	0.0586	1.0000			
Control_Nu~s	0.1411	-0.0578	-0.0477	-0.0271	-0.0996	-0.0139	1.0000		
Ctrl_NumUn~s	-0.0837	0.2057	0.2127	-0.0101	0.2933	0.0582	-0.0626	1.0000	
Ctrl_ProbF~s	0.0447	-0.1567	-0.1096	0.4389	0.0738	0.0464	-0.0121	-0.1247	1.0000

Table 19: KSC Correlation Table, Period 2, Technological Diversification – Cross-classification

```
. corr KnowledgeSourcingComplexity TechDiversification_CrossClass TechDiversificationSqd_CrossClas
> ICT_SharePerField N_Subclasses N_Citations Control_NumberOfTrials_Subclass Ctrl_NumUniqueClasses_S
> ubclass Ctrl_ProbFieldAcitesB_Subclass
(obs=329288)
```

	K~Sour~y	T~sClass	TechDi..	ICT_Sh~d	N_Subc~s	N_Cita~s	Contro~s	Ctrl_N~s	Ctrl_P~s
KnowledgeS~y	1.0000								
TechD~sClass	-0.0774	1.0000							
TechDiver~as	-0.0946	0.9676	1.0000						
ICT_ShareP~d	0.1599	-0.2231	-0.1439	1.0000					
N_Subclasses	0.0216	-0.0103	0.0059	-0.1569	1.0000				
N_Citations	0.4853	0.0655	0.0328	0.0777	0.0370	1.0000			
Control_Nu~s	0.1427	-0.0370	-0.0369	0.0379	-0.0980	-0.0032	1.0000		
Ctrl_NumUn~s	-0.0497	0.2777	0.2615	-0.1893	0.3006	0.0296	-0.0955	1.0000	
Ctrl_ProbF~s	0.1722	-0.3915	-0.2843	0.4889	0.0655	0.0377	0.0060	-0.2713	1.0000

Table 20: KSC Correlation Table, Period 3, Technological Diversification – Cross-classification

```
. corr KnowledgeSourcingComplexity TechDiversification_CrossClass TechDiversificationSqd_CrossClas
> ICT_SharePerField N_Subclasses N_Citations Control_NumberOfTrials_Subclass Ctrl_NumUniqueClasses_S
> ubclass Ctrl_ProbFieldAcitesB_Subclass
(obs=657021)
```

	K~Sour~y	T~sClass	TechDi..	ICT_Sh~d	N_Subc~s	N_Cita~s	Contro~s	Ctrl_N~s	Ctrl_P~s
KnowledgeS~y	1.0000								
TechD~sClass	-0.1433	1.0000							
TechDiver~as	-0.1554	0.9649	1.0000						
ICT_ShareP~d	0.1194	-0.3526	-0.2310	1.0000					
N_Subclasses	0.0167	0.0493	0.0451	-0.1333	1.0000				
N_Citations	0.5179	0.0407	0.0173	0.0327	0.0241	1.0000			
Control_Nu~s	0.1264	-0.0580	-0.0477	0.0608	-0.0872	0.0041	1.0000		
Ctrl_NumUn~s	-0.0957	0.3009	0.2811	-0.2078	0.3596	0.0152	-0.0872	1.0000	
Ctrl_ProbF~s	0.2094	-0.4595	-0.3552	0.5071	-0.0462	0.0280	0.0782	-0.3429	1.0000

Table 21:
Knowledge Sourcing Complexity regressions,
Technological Diversification in time period 1, 2, and 3.

	(1) KnowledgeS~y	(2) KnowledgeS~y	(3) KnowledgeS~y
TechD~sClass	-0.929***	1.677***	1.753***
	(-19.89)	(49.37)	(59.32)
TechDiver~as	0.485***	-2.103***	-2.541***
	(10.52)	(-58.20)	(-80.38)
ICT_ShareP~d	0.406***	0.165***	0.0764***
	(54.73)	(42.86)	(21.31)
N_Subclasses	-0.00709***	0.00665***	0.0141***
	(-11.65)	(16.66)	(34.61)
N_Citations	0.163***	0.0446***	0.0174***
	(224.57)	(308.60)	(496.39)
Control_Nu~s	0.00774***	0.00392***	0.00140***
	(53.66)	(95.75)	(108.01)
Ctrl_NumUn~s	-0.0833***	-0.000133	-0.0537***
	(-31.06)	(-0.06)	(-24.32)
Ctrl_ProbF~s	-0.721***	1.197***	1.819***
	(-29.46)	(71.56)	(125.49)
_cons	1.246***	1.026***	1.509***
	(110.67)	(113.81)	(177.09)
N	94931	329288	657021
R-sq	0.406	0.292	0.333
f			

t statistics in parentheses

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

Table 22: KSC Correlation Table, Period 1, Technological Distinctiveness – Cross-Classification

```
. corr KnowledgeSourcingComplexity TechDistinctiveness_CrossClas TechDistinctivenessSqd_CrossClas  I
> CT_SharePerField N_Subclasses N_Citations Control_NumberOfTrials_Subclass Ctrl_NumUniqueClasses_Su
> bclass Ctrl_ProbFieldAcitesB_Subclass
(obs=94931)
```

	K~Sour~y	TechDi..	TechDi..	ICT_Sh~d	N_Subc~s	N_Cita~s	Contro~s	Ctrl_N~s	Ctrl_P~s
KnowledgeS~y	1.0000								
~s_CrossClas	-0.1512	1.0000							
TechDistin..	-0.2104	0.9476	1.0000						
ICT_ShareP~d	0.1904	0.0773	0.0262	1.0000					
N_Subclasses	-0.0506	0.0627	0.0476	-0.0978	1.0000				
N_Citations	0.5831	0.0799	-0.0078	0.1369	0.0586	1.0000			
Control_Nu~s	0.1411	-0.0662	-0.0591	-0.0271	-0.0996	-0.0139	1.0000		
Ctrl_NumUn~s	-0.0837	0.3575	0.3089	-0.0101	0.2933	0.0582	-0.0626	1.0000	
Ctrl_ProbF~s	0.0447	-0.0411	-0.0620	0.4389	0.0738	0.0464	-0.0121	-0.1247	1.0000

Table 23: KSC Correlation Table, Period 2, Technological Distinctiveness – Cross-Classification

```
. corr KnowledgeSourcingComplexity TechDistinctiveness_CrossClas TechDistinctivenessSqd_CrossClas  I
> CT_SharePerField N_Subclasses N_Citations Control_NumberOfTrials_Subclass Ctrl_NumUniqueClasses_Su
> bclass Ctrl_ProbFieldAcitesB_Subclass
(obs=329288)
```

	K~Sour~y	TechDi..	TechDi..	ICT_Sh~d	N_Subc~s	N_Cita~s	Contro~s	Ctrl_N~s	Ctrl_P~s
KnowledgeS~y	1.0000								
~s_CrossClas	-0.0852	1.0000							
TechDistin..	-0.1335	0.9355	1.0000						
ICT_ShareP~d	0.1599	-0.0843	-0.1047	1.0000					
N_Subclasses	0.0216	0.0365	0.0340	-0.1569	1.0000				
N_Citations	0.4853	0.1051	0.0445	0.0777	0.0370	1.0000			
Control_Nu~s	0.1427	-0.0672	-0.0545	0.0379	-0.0980	-0.0032	1.0000		
Ctrl_NumUn~s	-0.0497	0.3770	0.3336	-0.1893	0.3006	0.0296	-0.0955	1.0000	
Ctrl_ProbF~s	0.1722	-0.2904	-0.2493	0.4889	0.0655	0.0377	0.0060	-0.2713	1.0000

Table 24: KSC Correlation Table, Period 3, Technological Distinctiveness – Cross-Classification

```
. corr KnowledgeSourcingComplexity TechDistinctiveness_CrossClas TechDistinctivenessSqd_CrossClas  I
> CT_SharePerField N_Subclasses N_Citations Control_NumberOfTrials_Subclass Ctrl_NumUniqueClasses_Su
> bclass Ctrl_ProbFieldAcitesB_Subclass
(obs=657021)
```

	K~Sour~y	TechDi..	TechDi..	ICT_Sh~d	N_Subc~s	N_Cita~s	Contro~s	Ctrl_N~s	Ctrl_P~s
KnowledgeS~y	1.0000								
~s_CrossClas	-0.1292	1.0000							
TechDistin..	-0.1713	0.9358	1.0000						
ICT_ShareP~d	0.1194	-0.1094	-0.1171	1.0000					
N_Subclasses	0.0167	0.0789	0.0701	-0.1333	1.0000				
N_Citations	0.5179	0.0803	0.0391	0.0327	0.0241	1.0000			
Control_Nu~s	0.1264	-0.0727	-0.0564	0.0608	-0.0872	0.0041	1.0000		
Ctrl_NumUn~s	-0.0957	0.3601	0.3214	-0.2078	0.3596	0.0152	-0.0872	1.0000	
Ctrl_ProbF~s	0.2094	-0.4074	-0.3382	0.5071	-0.0462	0.0280	0.0782	-0.3429	1.0000

Table 25: Knowledge Sourcing Complexity regressions,
using Technological Distinctiveness measure in time period 1, 2, and 3.

	(1) KnowledgeS~y	(2) KnowledgeS~y	(3) KnowledgeS~y
~s_CrossClas	-0.00783***	0.0411***	0.0886***
	(-4.35)	(29.23)	(68.85)
TechDistin..	-0.00341***	-0.00835***	-0.0149***
	(-19.25)	(-56.73)	(-111.10)
ICT_ShareP~d	0.333***	0.127***	0.00713*
	(50.92)	(33.24)	(2.13)
N_Subclasses	-0.00664***	0.00469***	0.0136***
	(-11.13)	(11.83)	(33.72)
N_Citations	0.157***	0.0455***	0.0174***
	(222.27)	(316.94)	(501.51)
Control_Nu~s	0.00771***	0.00389***	0.00140***
	(54.27)	(95.17)	(109.02)
Ctrl_NumUn~s	-0.0442***	0.0212***	-0.0474***
	(-15.94)	(9.04)	(-21.28)
Ctrl_ProbF~s	-0.447***	0.907***	1.685***
	(-20.68)	(58.01)	(116.57)
_cons	1.042***	1.329***	1.762***
	(151.32)	(233.63)	(308.61)
N	94931	329288	657021
R-sq	0.422	0.298	0.341
f			

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

.

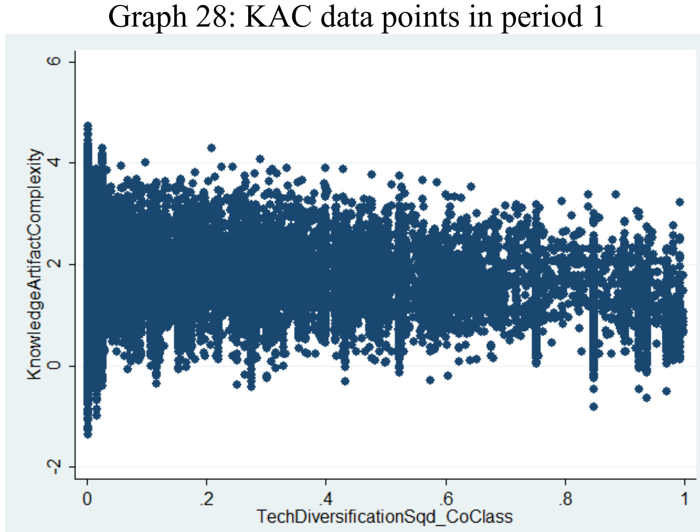
Table 26: Correlation Table, KAC, TechDiversification and TechDistinctiveness in Co-classification, All time periods

```
. corr KnowledgeArtifactComplexity TechDistinctiveness_CoClass TechDiversification_CoClass
(obs=1081240)

             | Knowle~y T~stin~s T~vers~s
-------------+---------------------------
KnowledgeA~y |   1.0000
TechDistin~s |  -0.1322   1.0000
TechDivers~s |  -0.1544   0.7136   1.0000
```

Table 27: Correlation Table KSC, TechDiversification and TechDistinctiveness in Cross-classification, All time periods

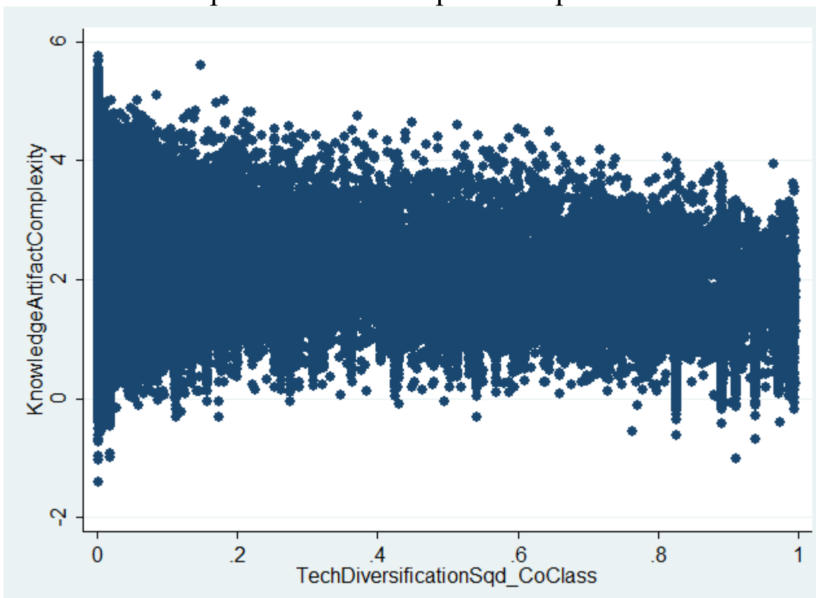
```
. corr KnowledgeSourcingComplexity TechDistinctiveness_CrossClas TechDiversification_CrossClas
(obs=1081240)

             | K~Sour~y TechDi.. TechDi..
-------------+---------------------------
KnowledgeS~y |   1.0000
TechDisti~as |  -0.1200   1.0000
TechDiver~as |  -0.1294   0.7330   1.0000
```

Graph 28: KAC data points in period 1



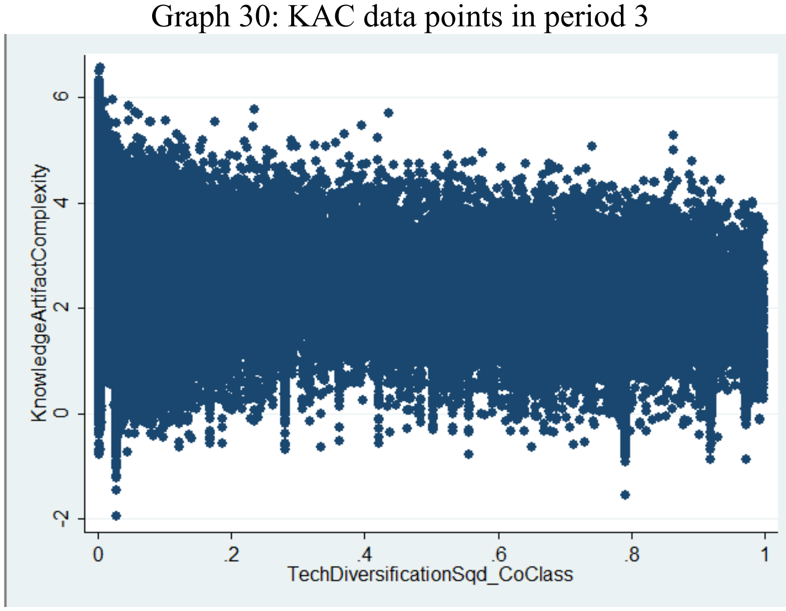
	KAC	TechDiversificationSquared_Co-class
Mean	1.560074	0.1085484
Median	1.539584	0.0168152

Graph 29: KAC data points in period 2



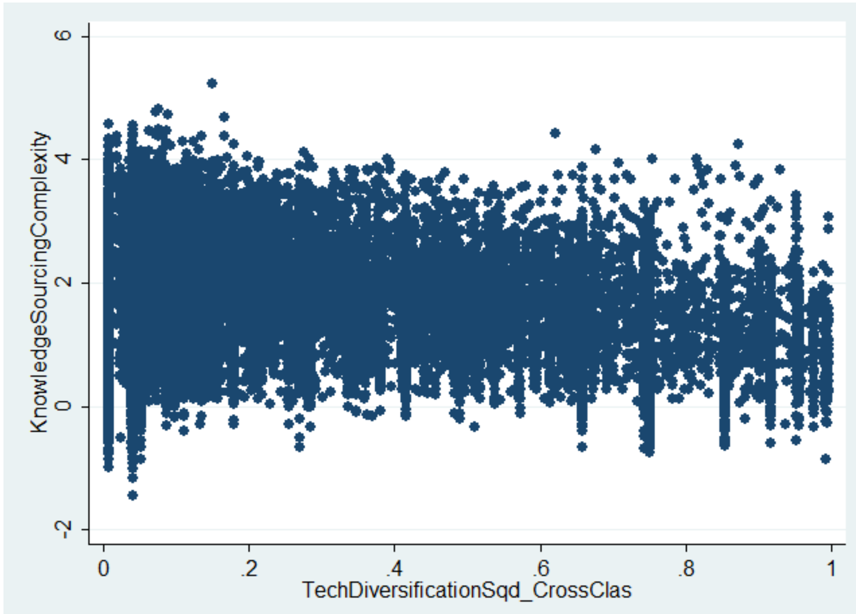
	KAC	TechDiversificationSquared_Co-class
Mean	1.908707	0 .0795867
Median	1.837505	0 .0063781

Graph 30: KAC data points in period 3



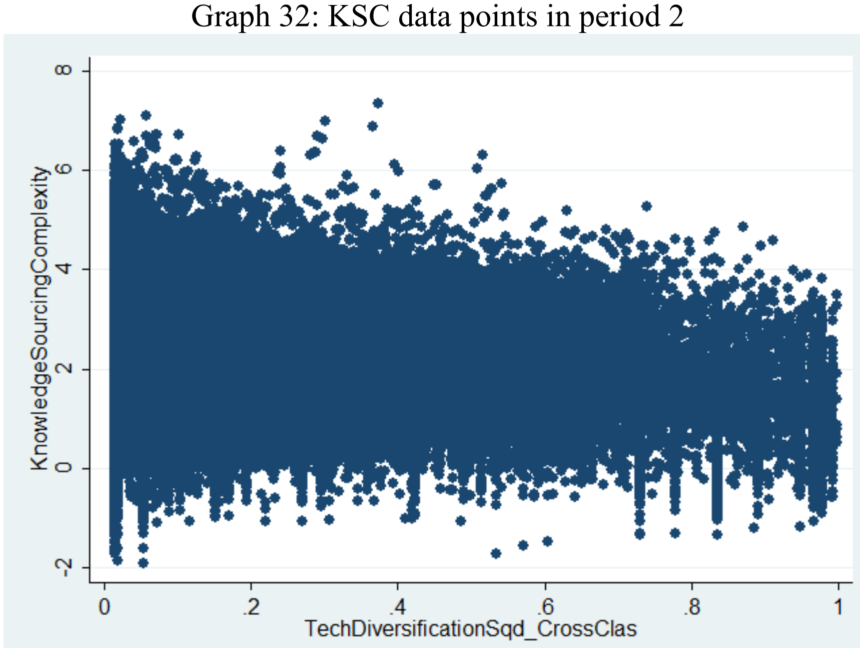
	KAC	TechDiversificationSquared_Co-class
Mean	2.264086	0.0681805
Median	2.223637	0.0045261

Graph 31: KSC data points in period 1



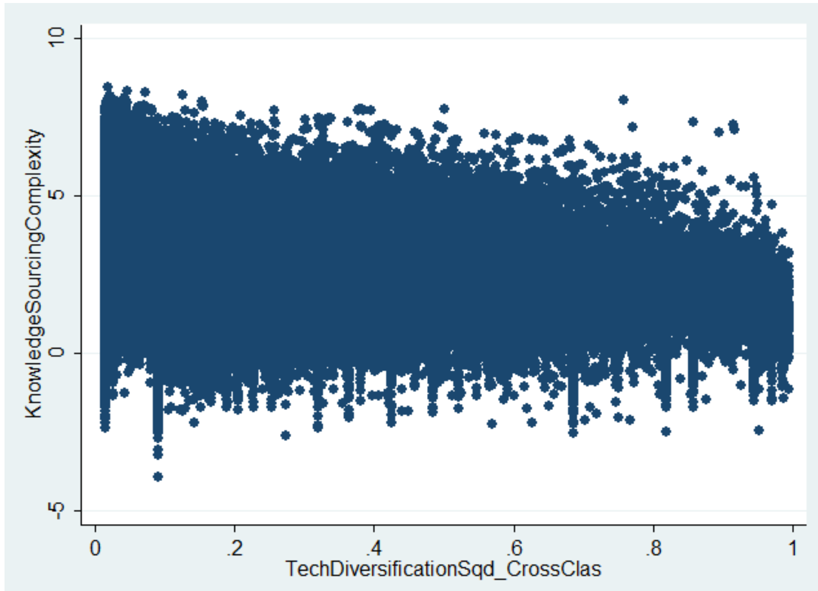
	KSC	TechDiversificationSquared_Cross-class
Mean	1.504951	0.141825
Median	1.44322	0.0409459

Graph 32: KSC data points in period 2



	KSC	TechDiversificationSquared_Cross-class
Mean	2.045397	0.1111922
Median	2.010449	0 .0537298

Graph 33: KSC data points in period 3



	KSC	TechDiversificationSquared_Cross-class
Mean	2.560702	0.1077855
Median	2.480552	0.0266826

Table 34: Correlation Table KAC, KSC, and LSC, for All Periods

```
. corr KnowledgeArtifactComplexity KnowledgeSourcingComplexi
> ty LocationSourcingComplexity
(obs=1081240)
```

	K~Arti~y	K~Sour~y	Locati~y
KnowledgeA~y	1.0000		
KnowledgeS~y	0.4771	1.0000	
LocationSo~y	0.1602	0.4593	1.0000

Table 35: Correlation Table KAC as DV, Period 1

```
. corr KnowledgeArtifactComplexity KnowledgeSourcingComplexity N_Subclasses Ctrl_NumUniqueClasses_Citation
>  Ctrl_ProbFieldAcitesB_Citation Ctrl_ProbFieldAcitesB_Subclass MechEng_SharePerField ICT_SharePerField
(obs=94931)
```

	K~Arti~y	K~Sour~y	N_Subc~s	Ctrl_N~n	Ctrl_P~n	Ctrl_P~s	MechEn~d	ICT_Sh~d
KnowledgeA~y	1.0000							
KnowledgeS~y	0.4849	1.0000						
N_Subclasses	0.4101	-0.0506	1.0000					
Ctrl_NumUn~n	0.0027	0.1542	0.1039	1.0000				
Ctrl_ProbF~n	0.0780	0.1450	-0.0602	0.1135	1.0000			
Ctrl_ProbF~s	0.0407	0.0447	0.0738	0.0678	0.7823	1.0000		
MechEng_Sh~d	-0.2213	-0.2458	-0.0417	-0.0811	-0.5018	-0.4263	1.0000	
ICT_ShareP~d	0.2093	0.1904	-0.0978	0.1639	0.7521	0.4389	-0.5445	1.0000

Table 36: Correlation Table KAC as DV, Period 2

	K~Arti~y	K~Sour~y	N_Subc~s	Ctrl_N~n	Ctrl_P~n	Ctrl_P~s	MechEn~d	ICT_Sh~d
KnowledgeA~y	1.0000							
KnowledgeS~y	0.4579	1.0000						
N_Subclasses	0.4913	0.0216	1.0000					
Ctrl_NumUn~n	-0.0473	0.2600	0.0726	1.0000				
Ctrl_ProbF~n	0.1865	0.2049	-0.0841	-0.1109	1.0000			
Ctrl_ProbF~s	0.2640	0.1722	0.0655	-0.1424	0.8158	1.0000		
MechEng_Sh~d	-0.2925	-0.1886	-0.0681	0.0812	-0.5526	-0.5728	1.0000	
ICT_ShareP~d	0.1376	0.1599	-0.1569	-0.0054	0.8005	0.4889	-0.5491	1.0000

Table 37: Correlation Table KAC as DV, Period 3

	K~Arti~y	K~Sour~y	N_Subc~s	Ctrl_N~n	Ctrl_P~n	Ctrl_P~s	MechEn~d	ICT_Sh~d
KnowledgeA~y	1.0000							
KnowledgeS~y	0.4243	1.0000						
N_Subclasses	0.4892	0.0167	1.0000					
Ctrl_NumUn~n	-0.0897	0.3864	0.0828	1.0000				
Ctrl_ProbF~n	0.2451	0.1846	-0.1229	-0.1760	1.0000			
Ctrl_ProbF~s	0.2816	0.2094	-0.0462	-0.1894	0.8437	1.0000		
MechEng_Sh~d	-0.3314	-0.2137	0.0250	0.1271	-0.5205	-0.5054	1.0000	
ICT_ShareP~d	0.2225	0.1194	-0.1333	-0.0677	0.7852	0.5071	-0.5742	1.0000

Table 38: Descriptive Statistics of KAC as DV, Period 1

```
. summarize KnowledgeArtifactComplexity KnowledgeSourcingComplexity N_Sub
> classes Ctrl_NumUniqueClasses_Citation Ctrl_ProbFieldAcitesB_Citation C
> trl_ProbFieldAcitesB_Subclass MechEng_SharePerField ICT_SharePerField
```

Variable	Obs	Mean	Std. Dev.	Min	Max
KnowledgeA~y	94931	1.560074	.6727414	-1.358123	4.727853
KnowledgeS~y	94931	1.504951	.7715607	-1.446919	5.225226
N_Subclasses	94931	4.371533	3.430617	2	215
Ctrl_NumUn~n	94931	1.781673	.902079	1	12
Ctrl_ProbF~n	94931	.2141308	.1891629	.0032201	.6940107
Ctrl_ProbF~s	94931	.1991886	.1006449	.0037681	.3983796
MechEng_Sh~d	94931	.4346675	.403792	.013065	.8845263
ICT_ShareP~d	94931	.1848164	.3355027	0	.8528234

Table 39: Descriptive Statistics of KAC as DV, Period 2

```
. summarize KnowledgeArtifactComplexity KnowledgeSourcingComplexity N_Sub
> classes Ctrl_NumUniqueClasses_Citation Ctrl_ProbFieldAcitesB_Citation C
> trl_ProbFieldAcitesB_Subclass MechEng_SharePerField ICT_SharePerField
```

Variable	Obs	Mean	Std. Dev.	Min	Max
KnowledgeA~y	329288	1.908707	.8136922	-1.403389	5.750908
KnowledgeS~y	329288	2.045397	.9705723	-1.89712	7.333428
N_Subclasses	329288	4.803843	3.898909	2	208
Ctrl_NumUn~n	329288	2.188039	1.390765	1	31
Ctrl_ProbF~n	329288	.318821	.237012	.0031081	.6940107
Ctrl_ProbF~s	329288	.2576448	.1128716	.0033211	.3983796
MechEng_Sh~d	329288	.2632403	.3766023	.0034161	.8991429
ICT_ShareP~d	329288	.3682409	.4418591	.000408	.9114777

Table 40: Descriptive Statistics of KAC as DV, Period 3

```
. summarize KnowledgeArtifactComplexity KnowledgeSourcingComplexity N_Sub
> classes Ctrl_NumUniqueClasses_Citation Ctrl_ProbFieldAcitesB_Citation C
> trl_ProbFieldAcitesB_Subclass MechEng_SharePerField ICT_SharePerField
```

Variable	Obs	Mean	Std. Dev.	Min	Max
KnowledgeA~y	657021	2.264086	.8855995	-1.94591	6.574169
KnowledgeS~y	657021	2.560702	1.240508	-3.912023	8.421334
N_Subclasses	657021	4.552115	3.338053	2	177
Ctrl_NumUn~n	657021	2.626972	2.120079	1	43
Ctrl_ProbF~n	657021	.4091731	.2499718	.0023668	.6940107
Ctrl_ProbF~s	657021	.2904507	.1137086	.0033197	.3983796
MechEng_Sh~d	657021	.1475586	.3009121	.0028396	.8916817
ICT_ShareP~d	657021	.5592496	.4411181	.0017501	.925244

Table 41: VIF Testing for KAC as DV in T1

Variable	VIF	1/VIF
Ctrl_ProbF~n	5.63	0.177492
Ctrl_ProbF~s	3.20	0.312794
ICT_ShareP~d	3.05	0.328148
MechEng_Sh~d	1.59	0.628747
KnowledgeS~y	1.11	0.904530
N_Subclasses	1.07	0.935050
Ctrl_NumUn~n	1.06	0.939493
Mean VIF	2.39	

Table 42: VIF Testing for KAC as DV in T2

Variable	VIF	1/VIF
Ctrl_ProbF~n	8.81	0.113490
Ctrl_ProbF~s	4.72	0.211740
ICT_ShareP~d	4.37	0.228703
MechEng_Sh~d	1.87	0.533794
KnowledgeS~y	1.16	0.861542
Ctrl_NumUn~n	1.15	0.869363
N_Subclasses	1.09	0.916332
Mean VIF	3.31	

Table 43: VIF Testing for KAC as DV in T3

Variable	VIF	1/VIF
Ctrl_ProbF~n	9.27	0.107891
Ctrl_ProbF~s	5.08	0.196870
ICT_ShareP~d	4.24	0.235982
MechEng_Sh~d	1.81	0.552163
KnowledgeS~y	1.36	0.736894
Ctrl_NumUn~n	1.35	0.742344
N_Subclasses	1.04	0.966016
Mean VIF	3.45	

Table 44: Correlations for KSC as DV, Period 1

```
. corr KnowledgeSourcingComplexity KnowledgeArtifactComplexity N_Citations Ctrl_NumUniqueClasses_Subclass
> Ctrl_ProbFieldAcitesB_Citation Ctrl_ProbFieldAcitesB_Subclass  DegreeofCountryConnectivity ICT_SharePerF
> ield EngineeringDummy
(obs=94931)
```

	K~Sour~y	K~Arti~y	N_Cita~s	Ctrl_N~s	Ctrl_P~n	Ctrl_P~s	Degree~y	ICT_Sh~d	Engine~y
KnowledgeS~y	1.0000								
KnowledgeA~y	0.4849	1.0000							
N_Citations	0.5831	0.1230	1.0000						
Ctrl_NumUn~s	-0.0837	0.0940	0.0582	1.0000					
Ctrl_ProbF~n	0.1450	0.0780	0.1235	0.0083	1.0000				
Ctrl_ProbF~s	0.0447	0.0407	0.0464	-0.1247	0.7823	1.0000			
DegreeofCo~y	-0.0468	-0.0452	0.0679	0.0431	0.0145	0.0259	1.0000		
ICT_ShareP~d	0.1904	0.2093	0.1369	-0.0101	0.7521	0.4389	0.0409	1.0000	
Engineerin~y	-0.1595	-0.1801	-0.0853	-0.0483	-0.6980	-0.7553	0.0204	-0.6600	1.0000

Table 45: Correlations for KSC as DV, Period 2

	K~Sour~y	K~Arti~y	N_Cita~s	Ctrl_N~s	Ctrl_P~n	Ctrl_P~s	Degree~y	ICT_Sh~d	Engine~y
KnowledgeS~y	1.0000								
KnowledgeA~y	0.4579	1.0000							
N_Citations	0.4853	0.0307	1.0000						
Ctrl_NumUn~s	-0.0497	0.1019	0.0296	1.0000					
Ctrl_ProbF~n	0.2049	0.1865	0.0668	-0.1775	1.0000				
Ctrl_ProbF~s	0.1722	0.2640	0.0377	-0.2713	0.8158	1.0000			
DegreeofCo~y	0.1046	0.0098	0.1676	0.0215	0.0672	0.1114	1.0000		
ICT_ShareP~d	0.1599	0.1376	0.0777	-0.1893	0.8005	0.4889	0.0239	1.0000	
Engineerin~y	-0.1587	-0.3025	-0.0343	0.0799	-0.6558	-0.7578	-0.0789	-0.6097	1.0000

Table 46: Correlations for KSC as DV, Period 3

	K~Sour~y	K~Arti~y	N_Cita~s	Ctrl_N~s	Ctrl_P~n	Ctrl_P~s	Degree~y	ICT_Sh~d	Engine~y
KnowledgeS~y	1.0000								
KnowledgeA~y	0.4243	1.0000							
N_Citations	0.5179	0.0210	1.0000						
Ctrl_NumUn~s	-0.0957	0.0256	0.0152	1.0000					
Ctrl_ProbF~n	0.1846	0.2451	0.0268	-0.2297	1.0000				
Ctrl_ProbF~s	0.2094	0.2816	0.0280	-0.3429	0.8437	1.0000			
DegreeofCo~y	0.2361	0.0479	0.1525	-0.0407	0.1988	0.2290	1.0000		
ICT_ShareP~d	0.1194	0.2225	0.0327	-0.2078	0.7852	0.5071	0.0775	1.0000	
Engineerin~y	-0.1974	-0.3393	-0.0240	0.1168	-0.6209	-0.6912	-0.1277	-0.6374	1.0000

Table 47: Descriptive Stats for KSC as DV, Period 1

```
. summarize KnowledgeSourcingComplexity KnowledgeArtifactComplexity N_Cit
> ations Ctrl_NumUniqueClasses_Subclass Ctrl_ProbFieldAcitesB_Citation Ct
> rl_ProbFieldAcitesB_Subclass DegreeofCountryConnectivity ICT_SharePerFi
> eld EngineeringDummy
```

Variable	Obs	Mean	Std. Dev.	Min	Max
KnowledgeS~y	94931	1.504951	.7715607	-1.446919	5.225226
KnowledgeA~y	94931	1.560074	.6727414	-1.358123	4.727853
N_Citations	94931	4.22845	2.813293	2	98
Ctrl_NumUn~s	94931	1.693841	.782093	1	10
Ctrl_ProbF~n	94931	.2141308	.1891629	.0032201	.6940107
Ctrl_ProbF~s	94931	.1991886	.1006449	.0037681	.3983796
DegreeofCo~y	94931	.6311336	.3019074	.0000309	.916511
ICT_ShareP~d	94931	.1848164	.3355027	0	.8528234
Engineerin~y	94931	.6021215	.4894627	0	1

Table 48: Descriptive Stats for KSC as DV, Period 2

```
. summarize KnowledgeSourcingComplexity KnowledgeArtifactComplexity N_Cit
> ations Ctrl_NumUniqueClasses_Subclass Ctrl_ProbFieldAcitesB_Citation Ct
> rl_ProbFieldAcitesB_Subclass DegreeofCountryConnectivity ICT_SharePerFi
> eld EngineeringDummy
```

Variable	Obs	Mean	Std. Dev.	Min	Max
KnowledgeS~y	329288	2.045397	.9705723	-1.89712	7.333428
KnowledgeA~y	329288	1.908707	.8136922	-1.403389	5.750908
N_Citations	329288	8.377114	10.10707	2	634
Ctrl_NumUn~s	329288	1.528188	.7067478	1	9
Ctrl_ProbF~n	329288	.318821	.237012	.0031081	.6940107
Ctrl_ProbF~s	329288	.2576448	.1128716	.0033211	.3983796
DegreeofCo~y	329288	.6449825	.2803273	.0000288	.916511
ICT_ShareP~d	329288	.3682409	.4418591	.000408	.9114777
Engineerin~y	329288	.3577264	.4793317	0	1

Table 49: Descriptive Stats for KSC as DV, Period 3

```
. summarize KnowledgeSourcingComplexity KnowledgeArtifactComplexity N_Cit
> ations Ctrl_NumUniqueClasses_Subclass Ctrl_ProbFieldAcitesB_Citation Ct
> rl_ProbFieldAcitesB_Subclass DegreeofCountryConnectivity ICT_SharePerFi
> eld EngineeringDummy
```

Variable	Obs	Mean	Std. Dev.	Min	Max
KnowledgeS~y	657021	2.560702	1.240508	-3.912023	8.421334
KnowledgeA~y	657021	2.264086	.8855995	-1.94591	6.574169
N_Citations	657021	16.54789	36.1849	2	5322
Ctrl_NumUn~s	657021	1.428531	.6598886	1	13
Ctrl_ProbF~n	657021	.4091731	.2499718	.0023668	.6940107
Ctrl_ProbF~s	657021	.2904507	.1137086	.0033197	.3983796
DegreeofCo~y	657021	.6646314	.2714147	1.19e-06	.916511
ICT_ShareP~d	657021	.5592496	.4411181	.0017501	.925244
Engineerin~y	657021	.2092962	.4068065	0	1

Table 50: VIF Testing for KAC as DV in T1

Variable	VIF	1/VIF
Ctrl_ProbF~n	6.58	0.152077
Ctrl_ProbF~s	5.97	0.167636
ICT_ShareP~d	4.31	0.232111
Engineerin~y	4.06	0.246495
Ctrl_NumUn~s	1.18	0.844649
KnowledgeA~y	1.10	0.906737
N_Citations	1.04	0.959452
DegreeofCo~y	1.03	0.966224
Mean VIF	3.16	

Table 51: VIF Testing for KAC as DV in T2

Variable	VIF	1/VIF
Ctrl_ProbF~n	11.75	0.085131
Ctrl_ProbF~s	9.80	0.102027
ICT_ShareP~d	6.35	0.157590
Engineerin~y	4.16	0.240665
Ctrl_NumUn~s	1.40	0.715234
KnowledgeA~y	1.15	0.872457
DegreeofCo~y	1.05	0.953848
N_Citations	1.04	0.963007
Mean VIF	4.58	

Table 52: VIF Testing for KAC as DV in T3

Variable	VIF	1/VIF
Ctrl_ProbF~n	12.95	0.077191
Ctrl_ProbF~s	10.22	0.097824
ICT_ShareP~d	6.39	0.156562
Engineerin~y	3.65	0.274003
Ctrl_NumUn~s	1.43	0.698532
KnowledgeA~y	1.15	0.868279
DegreeofCo~y	1.09	0.918066
N_Citations	1.03	0.973428
Mean VIF	4.74	

Table 53: Turbulence with ICT

```
. regress block3 block1

      Source |       SS       df       MS              Number of obs =      56
-------------+------------------------------           F(  1,    54) =   27.15
       Model |  683.224191     1  683.224191           Prob > F      =  0.0000
    Residual |    1358.724    54  25.1615556           R-squared     =  0.3346
-------------+------------------------------           Adj R-squared =  0.3223
       Total |  2041.94819    55  37.1263308           Root MSE      =  5.0161

------------------------------------------------------------------------------
      block3 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      block1 |   .7197639   .1381266     5.21   0.000     .4428368     .996691
       _cons |   .5004216   .7142492     0.70   0.487    -.9315619    1.932405
------------------------------------------------------------------------------
```

APPENDIX A2 – STUDY 2

Graph 1: Raw LSC Data vs Normalized LSC data

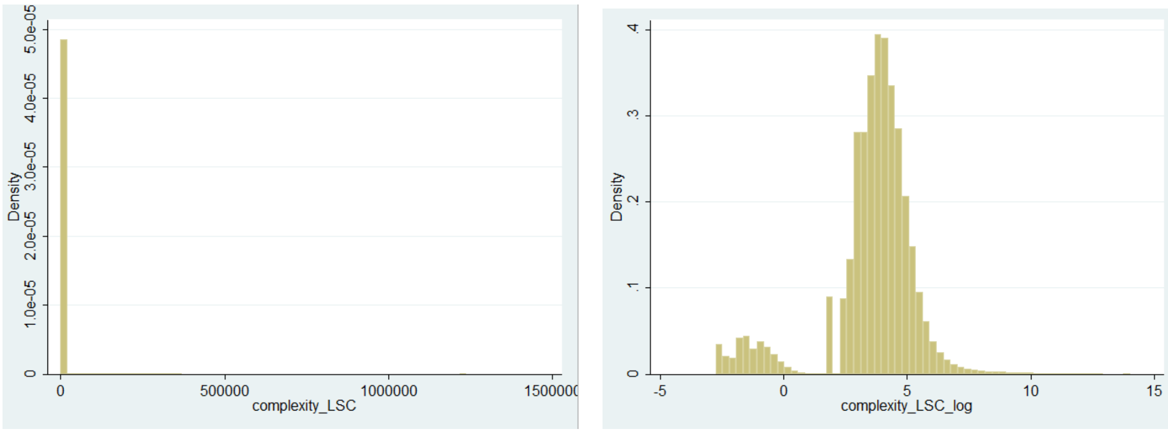


Table 2: LSC as DV: Descriptive Statistics, all periods

```
. summarize LocationSourcingComplexity KnowledgeArtifactComplexity Knowl
> edgeSourcingComplexity ICT_SharePerField Interaction_KACxICT Interacti
> on_KSCxICT CountryDistance N_LocationCitation N_Subclasses DegreeofCou
> ntryConnectivity
```

Variable	Obs	Mean	Std. Dev.	Min	Max
LocationSo~y	1081240	3.574587	1.835641	-2.752054	14.02869
KnowledgeA~y	1081240	2.094046	.8780738	-1.94591	6.574169
KnowledgeS~y	1081240	2.311074	1.179264	-3.912023	8.421334
ICT_ShareP~d	1081240	.468204	.4502376	0	.925244
Inter~ACxICT	1081240	1.080811	1.162463	-.6413302	6.08271
Inte~KSCxICT	1081240	1.188928	1.396714	-2.413817	7.791789
CountryDis~e	1081240	7.712113	3.974151	0	15.81663
N_Location~n	1081240	13.88387	31.18193	2	5784
N_Subclasses	1081240	4.612923	3.528763	2	215
DegreeofCo~y	1081240	.6557063	.2771908	1.19e-06	.916511

Table 3: LSC as DV: Correlation Table, all periods

	Locati~y	K~Arti~y	K~Sour~y	ICT_Sh~d	I~ACxICT	I~SCxICT	Countr~e	N_Loca~n	N_Subc~s	Degree~y
LocationSo~y	1.0000									
KnowledgeA~y	0.1602	1.0000								
KnowledgeS~y	0.4593	0.4771	1.0000							
ICT_ShareP~d	0.1775	0.2539	0.2013	1.0000						
Inter~ACxICT	0.1856	0.5073	0.3271	0.8997	1.0000					
Inter~SCxICT	0.3003	0.3722	0.5816	0.8256	0.8548	1.0000				
CountryDis~e	0.4463	-0.0084	0.3070	0.0660	0.0622	0.1822	1.0000			
N_Location~n	0.3479	0.0592	0.5010	0.0799	0.0860	0.3102	0.2499	1.0000		
N_Subclasses	0.0283	0.4618	0.0114	-0.1351	0.0002	-0.1053	0.0088	0.0182	1.0000	
DegreeofCo~y	-0.1431	0.0392	0.1834	0.0668	0.0428	0.1158	-0.1237	0.1403	0.0304	1.0000

Table 4: LSC as DV: VIF, all periods

Variable	VIF	1/VIF
Inter~ACxICT	12.00	0.083346
Inte~KSCxICT	9.28	0.107704
ICT_ShareP~d	8.45	0.118397
KnowledgeS~y	3.91	0.255550
KnowledgeA~y	3.44	0.290858
N_Subclasses	1.59	0.629914
N_Location~n	1.49	0.672019
CountryDis~e	1.23	0.810127
DegreeofCo~y	1.11	0.900193
Mean VIF	4.72	

Table 5: KAC as DV: Descriptive Statistics, all periods

```
. summarize KnowledgeArtifactComplexity KnowledgeSourcingComplexity Loca
> tionSourcingComplexity ICT_SharePerField Interaction_LSCxICT Interacti
> on_KSCxICT Ctrl_NumUniqueClasses_Subclass Ctrl_NumUniqueClasses_Citati
> on Ctrl_ProbFieldAcitesB_Subclass
```

Variable	Obs	Mean	Std. Dev.	Min	Max
KnowledgeA~y	1081240	2.094046	.8780738	-1.94591	6.574169
KnowledgeS~y	1081240	2.311074	1.179264	-3.912023	8.421334
LocationSo~y	1081240	3.574587	1.835641	-2.752054	14.02869
ICT_ShareP~d	1081240	.468204	.4502376	0	.925244
Inte~LSCxICT	1081240	1.820369	1.973125	-2.347016	12.97996
Inte~KSCxICT	1081240	1.188928	1.396714	-2.413817	7.791789
Ctrl_NumUn~s	1081240	1.482175	.6904734	1	13
Ctrl_NumUn~n	1081240	2.419081	1.862778	1	43
Ctrl_ProbF~s	1081240	.2724471	.1155893	.0033197	.3983796

Table 6: KAC as DV: Correlation Table, all periods

	K~Arti~y	K~Sour~y	Locati~y	ICT_Sh~d	I~LSCx~T	~KSCxICT	Ctrl_N~s	Ctrl_N~n	Ctrl_P~s
KnowledgeA~y	1.0000								
KnowledgeS~y	0.4771	1.0000							
LocationSo~y	0.1602	0.4593	1.0000						
ICT_ShareP~d	0.2539	0.2013	0.1775	1.0000					
Inte~LSCxICT	0.2583	0.3494	0.4034	0.8970	1.0000				
Inte~KSCxICT	0.3722	0.5816	0.3003	0.8256	0.8686	1.0000			
Ctrl_NumUn~s	0.0218	-0.1096	-0.0633	-0.2079	-0.1910	-0.2080	1.0000		
Ctrl_NumUn~n	-0.0330	0.3774	0.3565	-0.0009	0.1324	0.1535	0.2101	1.0000	
Ctrl_ProbF~s	0.3053	0.2429	0.1364	0.5275	0.4830	0.4925	-0.3163	-0.1235	1.0000

Table 7: KAC as DV: VIF, all periods

Variable	VIF	1/VIF
Inte~LSCxICT	9.94	0.100575
Inte~KSCxICT	9.27	0.107894
ICT_ShareP~d	8.09	0.123589
KnowledgeS~y	3.12	0.320477
LocationSo~y	1.88	0.532492
Ctrl_ProbF~s	1.56	0.640379
Ctrl_NumUn~n	1.39	0.718542
Ctrl_NumUn~s	1.18	0.850357
Mean VIF	4.55	

Table 8: KSC as DV: Descriptive Statistics, all periods

```
. summarize KnowledgeSourcingComplexity LocationSourcingComplexity Know
> ledgeArtifactComplexity  ICT_SharePerField Interaction_KACxICT Intera
> ction_LSCxICT  DegreeofCountryConnectivity Ctrl_NumUniqueClasses_Cita
> tion Ctrl_NumUniqueClasses_Subclass
```

Variable	Obs	Mean	Std. Dev.	Min	Max
KnowledgeS~y	1081240	2.311074	1.179264	-3.912023	8.421334
LocationSo~y	1081240	3.574587	1.835641	-2.752054	14.02869
KnowledgeA~y	1081240	2.094046	.8780738	-1.94591	6.574169
ICT_ShareP~d	1081240	.468204	.4502376	0	.925244
Inter~ACxICT	1081240	1.080811	1.162463	-.6413302	6.08271
Inte~LSCxICT	1081240	1.820369	1.973125	-2.347016	12.97996
DegreeofCo~y	1081240	.6557063	.2771908	1.19e-06	.916511
Ctrl_NumUn~n	1081240	2.419081	1.862778	1	43
Ctrl_NumUn~s	1081240	1.482175	.6904734	1	13

Table 9: KSC as DV: Correlation Table, all periods

```
. corr KnowledgeSourcingComplexity LocationSourcingComplexity KnowledgeArtifactComplexity  ICT_Sh
> arePerField Interaction_KACxICT Interaction_LSCxICT  DegreeofCountryConnectivity Ctrl_NumUnique
> Classes_Citation Ctrl_NumUniqueClasses_Subclass
(obs=1081240)
```

	K~Sour~y	Locati~y	K~Arti~y	ICT_Sh~d	I~ACxICT	I~LSCx~T	Degree~y	Ctrl_N~n	Ctrl_N~s
KnowledgeS~y	1.0000								
LocationSo~y	0.4593	1.0000							
KnowledgeA~y	0.4771	0.1602	1.0000						
ICT_ShareP~d	0.2013	0.1775	0.2539	1.0000					
Inter~ACxICT	0.3271	0.1856	0.5073	0.8997	1.0000				
Inte~LSCxICT	0.3494	0.4034	0.2583	0.8970	0.8274	1.0000			
DegreeofCo~y	0.1834	-0.1431	0.0392	0.0668	0.0428	0.0124	1.0000		
Ctrl_NumUn~n	0.3774	0.3565	-0.0330	-0.0009	-0.0176	0.1324	0.1350	1.0000	
Ctrl_NumUn~s	-0.1096	-0.0633	0.0218	-0.2079	-0.1862	-0.1910	-0.0164	0.2101	1.0000

Table 10: KSC as DV: VIF, all periods

Variable	VIF	1/VIF
ICT_ShareP~d	11.92	0.083903
Inter~ACxICT	9.54	0.104839
Inte~LSCxICT	7.71	0.129636
KnowledgeA~y	1.98	0.504187
LocationSo~y	1.71	0.583905
Ctrl_NumUn~n	1.33	0.753326
Ctrl_NumUn~s	1.13	0.881091
DegreeofCo~y	1.10	0.912207
Mean VIF	4.55	

APPENDIX A3 – STUDY 3

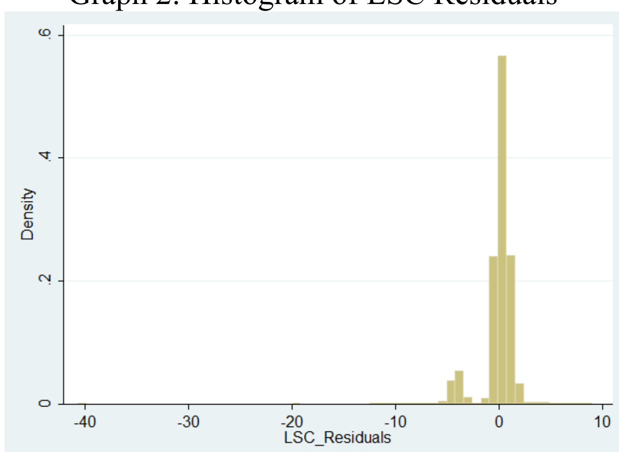
Table 1: LSC as DV Full Regression (Model 5)

Source	SS	df	MS
Model	1386173.82	9	154019.314
Residual	2257146.26	1081230	2.08757273
Total	3643320.09	1081239	3.36957887

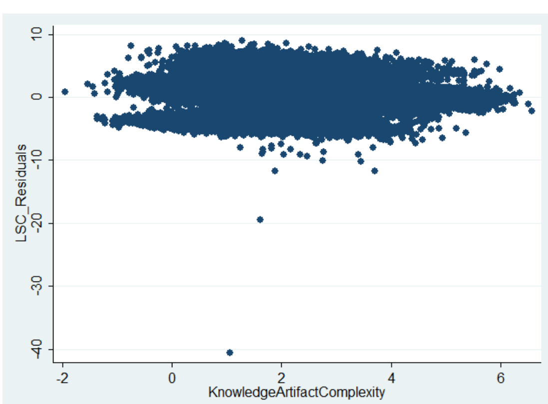
Number of obs	= 1081240
F(9,1081230)	=73779.14
Prob > F	= 0.0000
R-squared	= 0.3805
Adj R-squared	= 0.3805
Root MSE	= 1.4448

LocationSourcingComplexity	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
KnowledgeArtifactComplexity	-.1268554	.0029342	-43.23	0.000	-.1326063	-.1211044
KnowledgeSourcingComplexity	.7348574	.0023308	315.28	0.000	.7302891	.7394258
ICT_SharePerField	1.478834	.0089691	164.88	0.000	1.461254	1.496413
Interaction_KACxICT	-.1179875	.0041403	-28.50	0.000	-.1261024	-.1098725
Interaction_KSCxICT	-.3206777	.0030313	-105.79	0.000	-.3266191	-.3147364
CountryDistance	.1235417	.0003885	318.03	0.000	.1227804	.1243031
N_LocationCitation	.0075942	.0000544	139.71	0.000	.0074877	.0077008
N_Subclasses	.0394424	.0004961	79.50	0.000	.03847	.0404148
DegreeofCountryConnectivity	-1.37364	.0052834	-259.99	0.000	-1.383995	-1.363285
_cons	1.618861	.0065043	248.89	0.000	1.606112	1.631609

Graph 2: Histogram of LSC Residuals



Graph 3: Graph of LSC Residuals by KAC



Graph 4: Graph of LSC Residuals by KSC

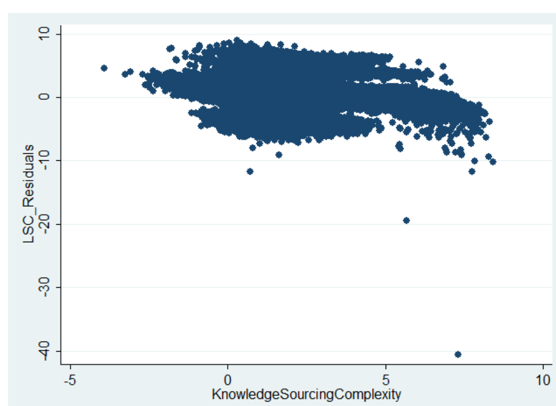


Table 5: Descriptives of LSC Normal Residuals

Variable	Obs	Mean	Std. Dev.	Min	Max
LocationSo~y	1081084	3.573725	1.834153	-2.752054	14.02869
KnowledgeS~y	1081084	2.311182	1.178981	-3.912023	8.292603
KnowledgeA~y	1081084	2.09413	.8780352	-1.94591	6.574169
ICT_ShareP~d	1081084	.4682572	.4502359	0	.925244
CountryDis~e	1081084	7.711886	3.974051	0	15.15303
N_Location~n	1081084	13.85079	29.95792	2	1122
N_Subclasses	1081084	4.612272	3.516222	2	208
DegreeofCo~y	1081084	.6557389	.277176	.0000255	.916511
Inter~ACxICT	1081084	1.08094	1.162478	-.6413302	6.08271
Inte~KSCxICT	1081084	1.18901	1.396617	-2.413817	7.67268

Table 6: Descriptives of LSC Negative Residual Outliers

Variable	Obs	Mean	Std. Dev.	Min	Max
LocationSo~y	28	6.585361	4.282881	-2.752054	10.04992
KnowledgeS~y	28	5.785041	2.505992	.7205462	8.421334
KnowledgeA~y	28	2.442424	.9166489	1.067262	4.443154
ICT_ShareP~d	28	.4940992	.4660652	.000172	.925244
CountryDis~e	28	12.83266	4.168714	0	15.81663
N_Location~n	28	1315.071	1116.066	2	5784
N_Subclasses	28	29.28571	54.24421	2	215
DegreeofCo~y	28	.7937512	.2150259	.1303003	.916511
Inter~ACxICT	28	.9808184	1.009126	.0005842	3.195497
Inte~KSCxICT	28	3.360441	3.237311	.0001239	7.791789

Table 7: Descriptives of LSC Positive Residual Outliers

Variable	Obs	Mean	Std. Dev.	Min	Max
LocationSo~y	128	10.19195	.7784857	8.278176	12.72807
KnowledgeS~y	128	.6443403	.7740527	-1.830226	2.430637
KnowledgeA~y	128	1.303806	.8302609	-.7396672	4.098377
ICT_ShareP~d	128	.0128149	.0813948	.000408	.925244
CountryDis~e	128	8.504746	4.012353	0	13.18701
N_Location~n	128	8.617188	10.04958	2	69
N_Subclasses	128	4.710938	2.755531	2	16
DegreeofCo~y	128	.3503254	.237231	1.19e-06	.6910384
Inter~ACxICT	128	.0168963	.1182705	-.0045939	1.34293
Inte~KSCxICT	128	.0206672	.1985698	-.0113671	2.248932

Table 8: Normal LSC, Location Descriptives

tech56	Primary Continent	Average # of Locations	Average # of Subclasses	_freq
8	Africa	6.2	6.6	5
8	Asia	7.081633	7.897959	198
8	Australia	8.25	6	20
8	Europe	11.63314	9.189349	441
8	North America	35.72274	8.735675	1691
8	South America	5.6	6.8	5
12	Africa	9.125	6.6875	156
12	Asia	37.14654	15.05263	25068
12	Australia	23.56477	8.238342	2014
12	Europe	42.81807	16.23537	48629
12	North America	93.82921	15.27495	134496
12	South America	11.56589	7.356589	332
16	Africa	10.82955	5.443182	194
16	Asia	33.77121	11.26786	24013
16	Australia	39.34842	7.936652	789
16	Europe	35.41259	10.89277	24078
16	North America	93.58113	10.77523	61939
16	South America	10.97	6.74	207
29	Africa	7.457627	4.932203	133
29	Asia	25	7.795918	28866
29	Australia	19.86735	5.627551	927
29	Europe	29.06195	7.842478	33490
29	North America	84.26217	7.637897	72666
29	South America	12.78261	4.880435	295
40	Africa	6.4375	5.5625	17
40	Asia	72.17748	9.286896	53988
40	Australia	15.4375	5.640625	79
40	Europe	37.26154	8.702564	9162
40	North America	112.6153	7.973917	42038
40	South America	4	3.714286	7
41	Africa	23.32381	4.809524	189
41	Asia	69.01104	9.629399	130122
41	Australia	46.94978	6.305677	1845
41	Europe	85.04357	8.084329	43867
41	North America	187.836	8.970702	274135
41	South America	16.99123	5.175438	214
42	Africa	9.555555	2.888889	11
42	Asia	24.97378	5.573034	13955

42	Australia	10.90541	3.878378	243
42	Europe	19.53419	5.277778	8893
42	North America	64.33488	5.178295	13548
42	South America	11.30769	4	59
43	Africa	6.571429	3.714286	17
43	Asia	25.27723	5.452145	8355
43	Australia	10.48387	3.903226	104
43	Europe	19.93061	5.342857	5454
43	North America	61.29508	5.400273	14105
43	South America	9.625	3.416667	25

Table 9: Negative LSC, Location Descriptives

tech56	Primary Continent	Average # of Locations	Average # of Subclasses	_freq
12	Europe	5.25	103.5	4
12	North America	749.75	81.25	4
16	North America	1282.25	3.5	5
40	North America	1474	3	1
41	North America	1847.07	4.35714	14

Table 10: Positive LSC, Location Descriptives

Tech56	Primary Continent	Average # of Locations	Average # of Subclasses	_freq
12	Africa	6	4.357143	18
12	South America	6	5.777778	13
16	Africa	9.555555	5.259259	31
16	Europe	2	7	1
16	South America	11.10345	5.241379	31
29	Africa	10.10526	3.263158	19
29	South America	5.888889	4.222222	11
41	Africa	69	3	1
42	Africa	4	2	1
43	Africa	7	10	1
43	Australia	2	3	1

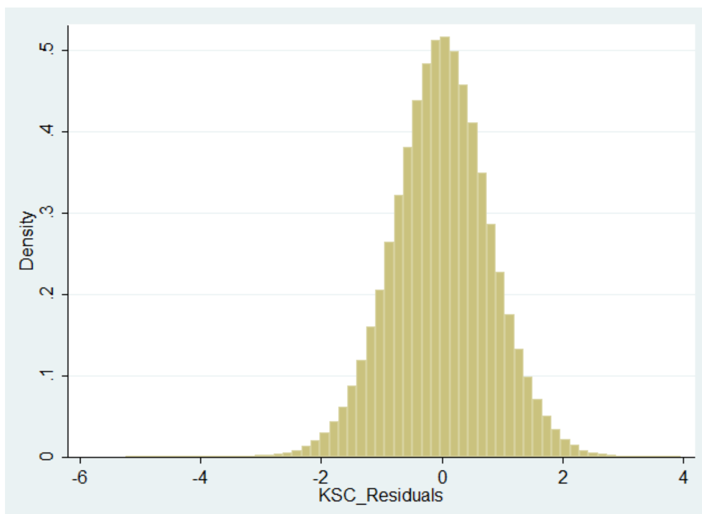
Table 11: KSC as DV Full Regression (Model 5)

Source	SS	df	MS
Model	793402.877	8	99175.3596
Residual	710237.169	1081231	.656878289
Total	1503640.05	1081239	1.3906639

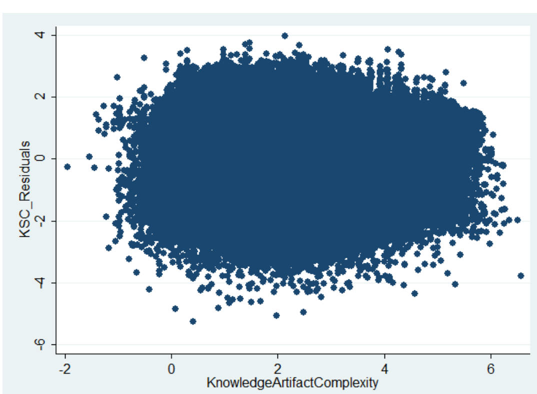
Number of obs = 1081240
F(8,1081231) = .
Prob > F = 0.0000
R-squared = 0.5277
Adj R-squared = 0.5277
Root MSE = .81048

KnowledgeSourcingComplexity	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
LocationSourcingComplexity	.1406441	.0005557	253.10	0.000	.139555	.1417332
KnowledgeArtifactComplexity	.5095466	.0012501	407.60	0.000	.5070964	.5119968
ICT_SharePerField	-1.230466	.0059766	-205.88	0.000	-1.24218	-1.218752
Interaction_KACxICT	.1858726	.0020708	89.76	0.000	.1818139	.1899313
Interaction_LSCxICT	.2182697	.0010971	198.94	0.000	.2161193	.2204201
DegreeofCountryConnectivity	.7631554	.0029441	259.21	0.000	.757385	.7689258
Ctrl_NumUniqueClasses_Citation	.1735882	.0004821	360.07	0.000	.1726433	.1745331
Ctrl_NumUniqueClasses_Subclass	-.2604873	.0012026	-216.60	0.000	-.2628444	-.2581302
_cons	.1849587	.0038687	47.81	0.000	.1773761	.1925413

Graph 12: Histogram of KSC Residuals



Graph 13: Graph of KSC Residuals by KAC



Graph 14: Graph of KSC Residuals by LSC

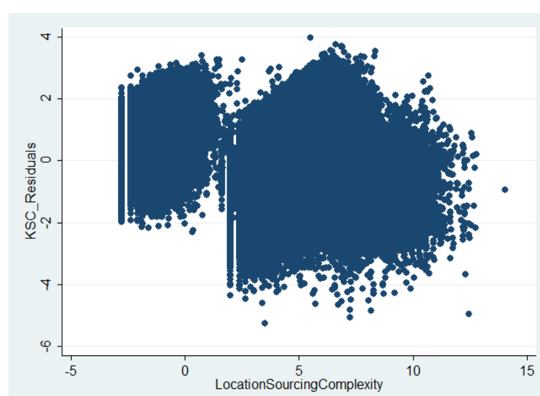


Table 15: Descriptives of KSC Normal Residuals

Variable	Obs	Mean	Std. Dev.	Min	Max
KnowledgeS~y	1080412	2.312393	1.176454	-2.515678	8.421334
LocationSo~y	1080412	3.573585	1.834894	-2.752054	14.02869
KnowledgeA~y	1080412	2.094048	.8779352	-1.94591	6.505368
ICT_ShareP~d	1080412	.4683085	.4502319	0	.925244
DegreeofCo~y	1080412	.6556897	.2772194	1.19e-06	.916511
Ctrl_NumUn~n	1080412	2.417689	1.856235	1	36
Ctrl_NumUn~s	1080412	1.482043	.6903185	1	13
Inter~ACxICT	1080412	1.081037	1.162446	-.6413302	5.462739
Inte~LSCxICT	1080412	1.820397	1.972365	-2.347016	12.97996

Table 16: Descriptives of KSC Negative Residual Outliers

Variable	Obs	Mean	Std. Dev.	Min	Max
KnowledgeS~y	694	-.4391304	1.294595	-3.912023	7.31718
LocationSo~y	694	4.697198	2.285596	1.974822	12.45072
KnowledgeA~y	694	2.096435	1.064771	-.7985078	6.574169
ICT_ShareP~d	694	.3643693	.4465498	.000408	.925244
DegreeofCo~y	694	.6695491	.2335763	.0034276	.916511
Ctrl_NumUn~n	694	4.409222	6.123522	1	43
Ctrl_NumUn~s	694	1.635447	.904349	1	7
Inter~ACxICT	694	.8731838	1.193237	-.0049593	6.08271
Inte~LSCxICT	694	1.995043	2.887499	.0008064	11.51995

Table 17: Descriptives of KSC Positive Residual Outliers

Variable	Obs	Mean	Std. Dev.	Min	Max
KnowledgeS~y	134	5.920385	1.042183	3.061534	7.728779
LocationSo~y	134	5.833288	2.077846	-1.142616	8.353377
KnowledgeA~y	134	2.065447	.9326264	-.5031036	4.326877
ICT_ShareP~d	134	.1627733	.3386604	.000172	.925244
DegreeofCo~y	134	.7178768	.245207	.0736254	.916511
Ctrl_NumUn~n	134	3.335821	1.536249	1	9
Ctrl_NumUn~s	134	1.753731	.4971583	1	4
Inter~ACxICT	134	.3356967	.7308676	-.0000865	2.355221
Inte~LSCxICT	134	.6878997	1.92155	-.4974305	7.074971

Table 18: Normal KSC

tech56	Primary Continent	Average # of Locations	Average # of Subclasses	_freq
8	Africa	6.2	6.6	5
8	Asia	7.081633	7.897959	198
8	Australia	8.25	6	20
8	Europe	11.63314	9.189349	441
8	North America	35.72274	8.735675	1691
8	South America	5.6	6.8	5
12	Africa	9.259259	6.444445	173
12	Asia	37.11054	15.06612	25061
12	Australia	23.56477	8.238342	2010
12	Europe	42.2707	16.49745	48604
12	North America	93.72511	15.4215	134387
12	South America	11.53383	7.360902	345
16	Africa	11.30303	5.515152	223
16	Asia	33.90583	11.18386	23935
16	Australia	38.83256	7.190698	775
16	Europe	35.10668	10.85698	24005
16	North America	94.24951	10.68027	61788
16	South America	11.25455	6.654545	233
29	Africa	7.939394	4.818182	151
29	Asia	23.77823	7.829569	28855
29	Australia	19.86735	5.627551	926
29	Europe	29.06195	7.842478	33484
29	North America	83.62915	7.625895	72643
29	South America	12.57447	4.925532	306
40	Africa	6.615385	5.230769	14
40	Asia	72.17748	9.286896	53981
40	Australia	15.63492	5.555555	78
40	Europe	37.26712	8.702055	9161
40	North America	113.2491	7.971601	42037
40	South America	4	3.714286	7

41	Africa	24.79167	4.479167	165
41	Asia	69.07538	9.612724	130070
41	Australia	47.73942	5.792873	1805
41	Europe	85.04357	8.084329	43819
41	North America	191.5367	8.952576	274050
41	South America	16.83809	4.971428	195
42	Africa	9.555555	2.888889	12
42	Asia	24.97378	5.573034	13954
42	Australia	10.90541	3.878378	243
42	Europe	19.53419	5.277778	8890
42	North America	63.95186	5.167702	13546
42	South America	11.30769	4	59
43	Africa	6.6	4.133333	18
43	Asia	25.27723	5.452145	8355
43	Australia	10.48387	3.903226	105
43	Europe	19.93061	5.342857	5454
43	North America	61.29508	5.400273	14105
43	South America	9.625	3.416667	25

Table 19: Negative KSC

tech56	Primary Continent	Average # of Locations	Average # of Subclasses	_freq
12	Africa	6	28	1
12	Asia	5	11.66667	3
12	Australia	4.5	10.75	4
12	Europe	3.333333	8.111111	10
12	North America	41.7037	10.7037	30
16	Africa	2	8.5	2
16	Asia	3.659091	9.818182	78
16	Australia	31.16667	20.41667	14
16	Europe	16.26415	9.716981	74
16	North America	78.51305	10.42609	156
16	South America	5.25	9	5
29	Africa	47	4	1

29	Asia	60.9	4.7	10
29	Australia	2	5	1
29	Europe	7.6	4.2	6
29	North America	82.55	7.25	21
40	Africa	5.666667	7	3
40	Asia	6.142857	6.571429	7
40	Australia	3	11	1
40	Europe	34	9	1
40	North America	109.5	11	2
41	Africa	6.217391	5.652174	25
41	Asia	8.138889	8.138889	47
41	Australia	6.6	13.33333	40
41	Europe	5.558824	7.5	46
41	North America	156.7937	8.507936	81
41	South America	12.21053	5.736842	19
42	Asia	5	8	1
42	Europe	4	3.333333	3
42	North America	158	7	2

Table 20: Positive KSC

tech56	Primary Continent	Average # of Locations	Average # of Subclasses	_freq
12	Asia	64	2.333333	4
12	Europe	101.6842	3.631579	19
12	North America	157.2639	3.041667	83
29	Asia	98	2	1
29	North America	31.5	2	2
41	Asia	16.8	2.8	5
41	Europe	10	2.5	2
41	North America	241.7778	2.333333	18

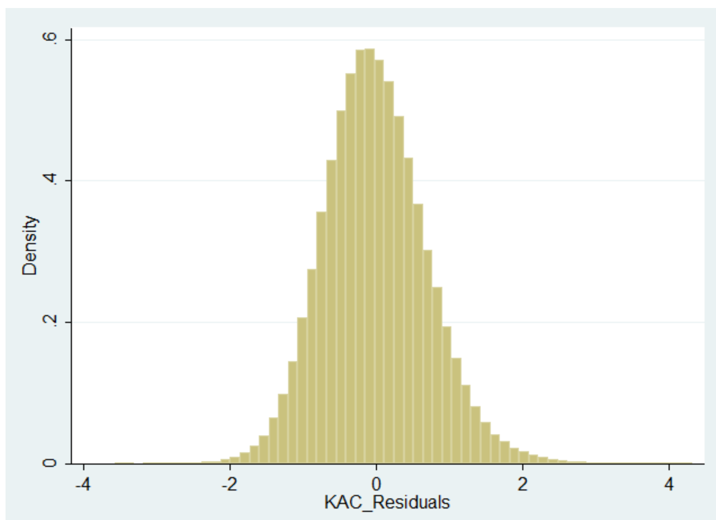
Table 21: KAC as DV Full Regression (Model 5)

Source	SS	df	MS
Model	296447.439	9	32938.6043
Residual	537202.585	1081230	.496843951
Total	833650.024	1081239	.771013646

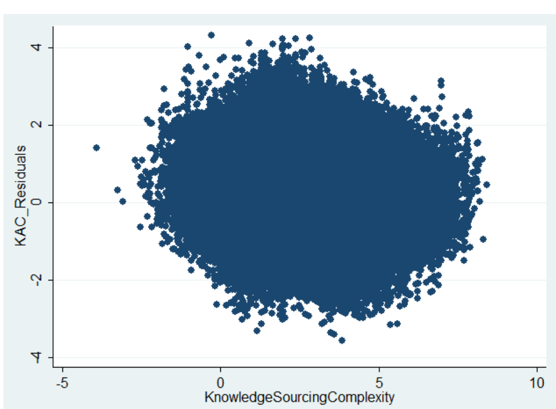
Number of obs	= 1081240
F(9,1081230)	=66295.67
Prob > F	= 0.0000
R-squared	= 0.3556
Adj R-squared	= 0.3556
Root MSE	= .70487

KnowledgeArtifactComplexity	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
KnowledgeSourcingComplexity	.471204	.0010283	458.25	0.000	.4691887	.4732194
LocationSourcingComplexity	-.0161313	.0005199	-31.03	0.000	-.0171503	-.0151124
ICT_SharePerField	.6080385	.0042942	141.59	0.000	.599622	.6164551
Interaction_LSCxICT	-.0582302	.0010862	-53.61	0.000	-.0603591	-.0561014
Interaction_KSCxICT	-.077458	.0014778	-52.41	0.000	-.0803545	-.0745614
DegreeofCountryConnectivity	-.2574291	.0026666	-96.54	0.000	-.2626556	-.2522026
Ctrl_NumUniqueClasses_Subclass	.2589002	.0010647	243.18	0.000	.2568135	.2609869
Ctrl_NumUniqueClasses_Citation	-.110635	.0004369	-253.23	0.000	-.1114913	-.1097787
Ctrl_ProbFieldAcitesB_Subclass	1.257897	.0074523	168.79	0.000	1.24329	1.272503
_cons	.6861145	.0035285	194.45	0.000	.6791988	.6930302

Graph 22: Histogram of KAC Residuals



Graph 23: Graph of KAC Residuals by KSC



Graph 24: Graph of KAC Residuals by LSC

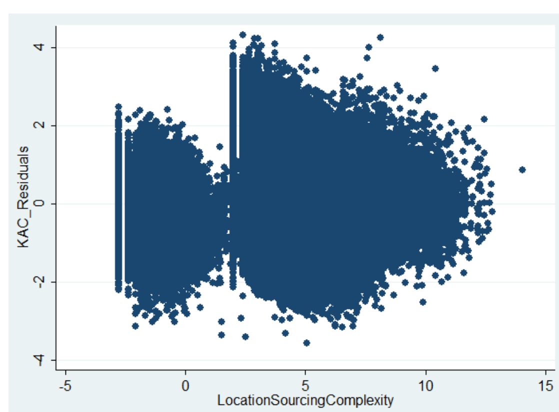


Table 25: Descriptives of KAC Normal Residuals

Variable	Obs	Mean	Std. Dev.	Min	Max
KnowledgeA~y	1080777	2.09273	.8756745	-1.94591	6.041676
KnowledgeS~y	1080777	2.311176	1.179323	-3.912023	8.421334
LocationSo~y	1080777	3.574859	1.835816	-2.752054	14.02869
CountryDis~e	1080777	7.713711	3.973176	0	15.81663
Ctrl_NumUn~s	1080777	1.482143	.6905269	1	13
Ctrl_NumUn~n	1080777	2.419412	1.862556	1	43
Ctrl_ProbF~s	1080777	.272421	.1156049	.0033197	.3983796
ICT_ShareP~d	1080777	.4683989	.4502308	0	.925244
Inte~LSCxICT	1080777	1.821124	1.973183	-2.347016	12.97996

Table 26: Descriptives of KAC Negative Residual Outliers

Variable	Obs	Mean	Std. Dev.	Min	Max
KnowledgeA~y	13	-.5852912	.5943401	-1.403389	.3022808
KnowledgeS~y	13	3.255229	1.617182	.6931472	5.589104
LocationSo~y	13	3.333904	3.246436	-2.058907	6.958786
CountryDis~e	13	9.629912	4.613576	0	13.80248
Ctrl_NumUn~s	13	1.923077	.2773501	1	2
Ctrl_NumUn~n	13	2.461538	1.198289	1	4
Ctrl_ProbF~s	13	.2811902	.0271264	.2273428	.3497039
ICT_ShareP~d	13	.0008153	.0007737	.000172	.0017501
Inte~LSCxICT	13	.0046803	.0055462	-.0003541	.0121786

Table 27: Descriptives of KAC Positive Residual Outliers

Variable	Obs	Mean	Std. Dev.	Min	Max
KnowledgeA~y	450	5.331803	.4423981	3.306704	6.574169
KnowledgeS~y	450	2.041047	.9658013	-1.143564	6.984927
LocationSo~y	450	2.926808	1.106238	1.974822	10.42102
CountryDis~e	450	3.817502	4.366336	0	13.5496
Ctrl_NumUn~s	450	1.546667	.5493872	1	4
Ctrl_NumUn~n	450	1.622222	2.212426	1	30
Ctrl_ProbF~s	450	.3349943	.0338984	.03585	.3983796
ICT_ShareP~d	450	.0135826	.1015873	.000408	.925244
Inte~LSCxICT	450	.0588682	.5035487	.0008064	7.507423

Table 28: Normal KAC

tech56	Primary Continent	Average # of Locations	Average # of Subclasses	_freq
8	Africa	6.2	6.6	5
8	Asia	7.081633	7.897959	198
8	Australia	8.25	6	20
8	Europe	11.63314	9.189349	441
8	North America	35.72274	8.735675	1691
8	South America	5.6	6.8	5
12	Africa	9.219512	6.707317	174
12	Asia	37.18285	14.90186	25048
12	Australia	23.61559	8.2	2007
12	Europe	42.79911	16.26319	48568
12	North America	94.56042	15.26647	134150
12	South America	11.53383	7.360902	345
16	Africa	11.21	5.56	225
16	Asia	33.8757	11.20605	24010
16	Australia	39.15454	7.740909	788
16	Europe	35.45158	10.87748	24077
16	North America	95.644	10.70424	61936
16	South America	11.19643	6.741071	238
29	Africa	8.522388	4.80597	152
29	Asia	25	7.795918	28866
29	Australia	19.86735	5.627551	927
29	Europe	29.06195	7.842478	33490
29	North America	84.26217	7.637897	72666
29	South America	12.57447	4.925532	306
40	Africa	6.4375	5.5625	17
40	Asia	72.21515	9.251432	53987
40	Australia	15.4375	5.640625	79
40	Europe	37.31849	8.660959	9161
40	North America	113.2491	7.971601	42038
40	South America	4	3.714286	7
41	Africa	23.32381	4.809524	190
41	Asia	69.01104	9.629399	130122
41	Australia	47.03501	5.932166	1844
41	Europe	85.04357	8.084329	43867
41	North America	193.1767	8.942696	274147
41	South America	16.99123	5.175438	214
42	Africa	9.555555	2.888889	12
42	Asia	25.05639	5.454887	13954
42	Australia	10.90541	3.878378	243

42	Europe	19.53419	5.277778	8893
42	North America	64.33488	5.178295	13548
42	South America	11.30769	4	59
43	Africa	6.6	4.133333	18
43	Asia	25.27723	5.452145	8355
43	Australia	10.48387	3.903226	105
43	Europe	19.93061	5.342857	5454
43	North America	61.29508	5.400273	14105
43	South America	9.625	3.416667	25

Table 29: Negative KAC

tech56	Primary Continent	Average # of Locations	Average # of Subclasses	_freq
12	Asia	24.5	2	46
12	Europe	34.6	2.2	40
12	North America	74.4	2	490

Table 30: Positive KAC

tech56	Primary Continent	Average # of Locations	Average # of Subclasses	_freq
12	Asia	2.8	20.6	30
12	Australia	5.8	13.2	26
12	Europe	3.27907	24.30233	115
12	North America	7.801802	17.42342	1486
16	Asia	2.666667	29.66667	5
16	Australia	82	51	81
16	Europe	2	18.5	2
16	North America	14.375	31.125	107
40	Asia	13	65	12
40	Europe	4	33	3
40	North America	6	23	5
41	Australia	8	177	7
41	North America	13	38.5	24
42	Asia	3	37	2

Table 30: Patenting Percentage by Continent

continent	_freq	Average
Africa	6997	0.046%
Asia	2242366	14.856%
Australia	83409	0.553%
Europe	1396400	9.251%
North America	11354899	75.228%
South America	9809	0.065%
Total	15093880	100%