Essays on Firm Innovation in Dynamic Product Markets: Examining Competitive Interactions During Technological Commercialization

by

Kevin Kai Du

Business Administration
Duke University

Date: _____
Approved:

_____
Ashish Arora, Supervisor

_____
William Mitchell, Co-Supervisor

_____
Sharon Belenzon

_____
Jose Ramon Lecuona Torras

_____
Giuseppe Lopomo

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Business Administration
in the Graduate School of Duke University
2018

Abstract

Essays on Firm Innovation in Dynamic Product Markets: Examining Competitive Interactions During Technological Commercialization

by

Kevin Kai Du

Business Administration
Duke University

Date: _____
Approved:

_____
Ashish Arora, Supervisor

_____
William Mitchell, Co-Supervisor

_____
Sharon Belenzon

_____
Jose Ramon Lecuona Torras

_____
Giuseppe Lopomo

An abstract of a dissertation submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in Business Administration
in the Graduate School of Duke University
2018

Abstract

How can firms gain competitive advantage from available technologies is a key question in strategy. In my dissertation, I develop new theory and provide evidence to show that a firm's focus in selective technological areas may play a central role of creating competitive advantage in industries with rapid product turnover. Firms commit limited resources when selecting which technologies to develop, affecting the composition of their product portfolios and allowing some firms to subsequently capture greater value relative to others.

I examine how firm attention to technologies within an industry affect their ability to swiftly incorporate them into products (essay 1); establish a theoretical foundation for firm-to-firm matching in the market for alliances (essay 2); develop an econometric methodology based on the insights from a firm-to-firm matching model (essay 3); and investigate how firm technological interests attract partners in the market for inter-firm collaboration (essay 4). Across four essays, I find that competitive advantage varies with the firm's technological composition, its current focal area of technological development, and the collection of potential alliance partners. These findings contribute to understanding conditions under which a firm captures value from the component technologies scattered across its industry, and the key trade-offs associated with allocating its technological focus.

For You An Du (dec.), Hong Lian Li (dec.), Jiang Zao Zhu, & Wei Xiu Huang

Contents

List of Tables

List of Figures

Acknowledgements

I am very grateful to my advisor Ashish Arora and co-advisor Will Mitchell for our extensive discussions, their influential guidance and great patience. I want to thank my committee members Sharon Belenzon, Ramon Lecuona Torras, and Pino Lopomo for their tremendous help, many insightful comments, and invaluable support throughout the process. During my time at Duke, I have had the great fortune of meeting many amazing people. I'm highly appreciative of the Strategy Department at the Fuqua School of Business and want to take this opportunity to express my sincere thanks to the faculty group for the nurturing environment it provides to early-stage researchers like myself.

On a personal note, I thank my family for their life-long love and support. To my parents, Matthew Du and Jennifer Zhu: I would not be the person I am today if not for your sacrifices.

1

Introduction

A central area of focus for strategy scholars is understanding how firms consistently outcompete rivals. While maintaining low costs works well for undifferentiated goods, industries characterized by high levels of innovation and product turnover require companies to effectively extract value from the wide range of component technologies within their environment. This is no easy task and understanding how this occurs is further complicated by the endogenous nature of firm interactions. That is, the decisions or outcomes of some firms will affect the options and consequences of others. This brings traditional empirical work on companies that have relied on single agent-based approaches to question. The thesis presented here consists of four essays that are motivated to answer the question: how can firms extract value from the available set of technologies in their environment given the presence of such externalities? The set of essays contributes to the areas of innovation, technological market entry, and inter-firm alliances within the field of strategic management by investigating the nature of how firm decisions affect one another, re-examining prior results under a novel methodology and deriving new insights.

In my first study, I examine how the speed of introducing a product with a new

focal technology for follower firms after a pioneer's entry is affected by managerial attention, considering the role of both the follower firm's attentional dispersion to latent industry-level technological topics as well as its attentional similarity with respect to the pioneer firm. Using patent application abstracts filed by top mobile device firms from 2004 to 2008, I implement probabilistic topic modeling to reveal latent themes across technologies, develop novel measures of firm attention and test their impact on the speed of launching new mobile phone features in the German market. I find that both attentional dispersion and attentional similarity are related to faster entry following a pioneer. Moreover, attentional dispersion not only help follower firms commercialize market-tested and successful technologies quickly, but may also provide alternatives for solving earlier flops. This work contributes to the literatures on entry timing and managerial cognition.

For the second essay, I setup a nontransferable utility (NTU) two-sided matching model that can be used to characterize a firm-to-firm matching market for collaborations. Within this framework, all potential matches are assigned values corresponding to a net realized profit from an alliance between two firms. A Greedy algorithm generates the equilibrium, whereby the highest valued alliances are formed in sequence so long as no firm capacity constraints are violated at each stage. After ensuring existence and uniqueness of the matching equilibrium, I show a set of Pareto efficient properties that characterize the equilibrium outcome. This work contributes to economic matching theory by extending a set of results to a many-to-many (NTU) matching market with a fixed sharing rule between partners.

Next, I propose a score estimator based on the insights generated from the NTU matching model of essay 2. The score estimator uses a simple set of inequality conditions implied by the model to estimate the effects of observable match characteristics on the surplus generated within a match. I start off by making a set of assumptions on the functional form of the match surplus function and the distri-

bution of the observable characteristics underlying the data generating process to ensure identification. The score estimator is then proposed and the intermediate result on identification is used to prove its consistency under random sampling. This work contributes to the matching econometrics literature by furnishing a convenient semi-parametric estimator for a class of NTU matching models.

In my last chapter, I use the insights and tools developed in the preceding 2 chapters to investigate the impact of firms' technological interests on alliance value creation among Software and Information & Communication Technology (ICT) firms. With the newly developed matching framework, I'm able to account for the nature of incomplete contracts that are symptomatic of inter-firm relations. Based on my findings, firms seek partners that have common technological interests while possessing breadth in technological specializations/expertise that can be used to realize commercial value from the joint knowledge produced within the alliance. Secondly, most of the net value created within an alliance comes from having common interests in technologies that neither partner owns. This is because when a firm owns a technology, it is better able to commercialize it in the absence of collaboration, thus reducing the potential net benefits from an alliance. In addition, firms with relatively better outside options compared to their partners can extract a greater proportion of the total joint value produced. This pattern of findings is not detected using other methodologies that have been employed and contributes to the alliance literature by re-examining the determinants of inter-firm collaboration under an innovative matching-based approach.

2

Following Faster And Smarter: Attention Focus And Entry Timing In Mobile Devices

Organizational and strategy scholars who seek to understand innovation and market entry are paying increasing heed to the role of managerial attention. Attention is a scarce managerial resource (Simon, 1947; Ocasio, 1997) and the technology-related problems, opportunities, and solutions (more generally, topics) to which managers attend can have far-reaching consequences for whether and when firms introduce new technologies (Cho and Hambrick, 2006; Dahlander, O'Mahony, and Gann, 2016; Wilson and Joseph, 2015). Managers' attention can direct firm investments in new technologies (Barnett, 2008; Kaplan, 2008; Maula, Keil, and Zahra, 2013), increase the rate of new product introductions (Li, Magitti, Smith, Tesluk, and Katila, 2013; Nadkarni and Chen, 2014; Yadav, Prabhu, and Chandy, 2007), and facilitate product market entry (Eggers and Kaplan, 2009). Given that speed of introducing successful technological innovations is associated with revenue premiums (Banbury and Mitchell, 1995; Klingebiel and Joseph, 2016), the distribution of attention within organizations and its impact on timing of technology introductions may have important

implications for overall firm performance (Williams and Mitchell, 2004).

Key issues remain open in how attention affects market entry. Prior studies of cognition and innovation have emphasized attention patterns within firms and firms' selective focus on a given technology in predicting the corresponding entry into that technology (e.g., Eggers and Kaplan, 2009). Yet firms face a wide array of decision alternatives (Klingebiel, 2017), with competing firms attending to varying subsets of technology-related topics. Notably absent in the extant research is consideration of how an overlap in competing firms' attention patterns (such as between technology pioneer and follower) might affect technology introductions. Indeed, variation in firms' attention patterns to topics may be consequential for their entry timing.

Much of the work on early mover advantages and disadvantages (Lieberman and Montgomery, 1988; Mitchell, 1989) has ignored cognitive explanations, favoring an emphasis on industry dynamics or firm capabilities (cf., Fosfuri, Lanzolla, and Suarez, 2013). Yet early entry decisions reflect firms' efforts to choose among a wide array of feature technology alternatives and to balance commercial uncertainty with potential competitive preemption (Agarwal and Bayus, 2004; Klingebiel and Joseph, 2016; McCardle, 1985). Given that the problem in these choice situations is to make the most of limited available attention and to use resources for features that could provide the greatest returns, attention-based theory offers a window on such decisions (Ocasio, 1997).

This study aims to address these gaps by examining the relationship between entry timing and topic attention. I link theories of attention (Ocasio, 1997, 2011) with perspectives on organizational learning (Levinthal and Rerup, 2006; Weick and Sutcliffe, 2006) to explore whether entry timing following a technology pioneer's entry is affected by the relationship of attentional dispersion across industry topics and attentional similarity to industry topics relative to the pioneer. My thesis is that in order to understand entry timing, it is necessary to understand how attentional

dispersion and similarity affect the firm's recognition of environmental cues and trajectory of organizational learning - the identification, development, and utilization of knowledge in support of integrating new technology features into new products.

To test these arguments, I examine firm attention patterns and entry timing of technology (i.e., feature) innovations in the mobile device industry from 2004 to 2008, during what is known as the feature phone era (Klingebiel and Joseph, 2016). As the focal outcome, I examine the likelihood of commercializing a new technology in a product following the introduction of the technology by a pioneer. My conjectures shift from the concerns of the pioneer and, instead, focus on the speed of entry displayed by other early movers. Technological advancement of products is a key concern of managers and a source of performance heterogeneity (Conti, Gambardella, and Novelli, 2013; Eggers, 2012a; Paulson Gjerde, Slotnick, and Sobel, 2002). During this era, performance in the mobile device industry required firms to offer new products with novel but rapidly outdated technologies (Giachetti and Lampel, 2010; Koski and Kretschmer, 2010), such as newer displays, cameras, and memory.

I explore attention to technology topics articulated within company patents by way of a text analysis technique, topic modeling (Blei, 2012; Kaplan and Vakili, 2015). Topic modeling identifies topics (themes) that are latent in a collection of documents and captures the frequency distribution of generated topics that best accounts for each document (Blei, Ng, and Jordan, 2003). Recently, social scientists have begun to recognize the scalability and applicability of this text mining technique and have adopted it in various empirical settings to capture cognitive breakthroughs in the corpus of patent abstracts (Kaplan and Vakili, 2015), and specifically to examine the attentional patterns of firms (Wilson and Joseph, 2015).

This paper makes two contributions to studies of managerial cognition and technology strategy. First, it augments cognitive perspectives of technology strategy by

explicitly addressing the entry timing consequences of attention patterns to both industry-wide and pioneer topics. Prior work has examined the effects of attention to external information sources on the rate of new product or patent introductions (Dahlander et al., 2016; Li, Maggitti, Smith, Tesluk, and Katila, 2013; Nadkarni and Barr, 2008; Yadav et al., 2007). Here, I broaden the attentional possibilities of the firm and consider its attention to the array of potential topics within the industry and to the first movers of new technologies. Second, I contribute to the literature on entry timing and early mover advantages (Suarez and Lanzolla, 2007). In addition to offering an attention-based lens for such decisions, I make heretofore unrecognized distinctions in feature technologies. In particular, I acknowledge that not all technologies are worth pursuing and highlight the variations caused by topic attention patterns in following hit vs. flop technologies. I also examine the capacity of firms to recover from flop introductions, an important aspect of technology strategy with managerial implications given that performance in many industries depends on fast entry of a stream of new features.

2.1   Organizational Attention and Entry

The notion of topics to which organizations attend is central to the attention-based view (ABV) of strategy (Ocasio and Joseph, 2005). The ABV defines attention as the noticing, encoding, interpreting, and focusing of time and effort by organizational decision makers on issues and initiatives (Ocasio, 1997: 189). I use the term topic to include combinations of issues (problems, opportunities) and initiatives (solutions). A firm's choice of topics not only reflects an assessment of expected value to the firm (Nickerson and Zenger, 2004) but also is shaped by managerial experience (Barreto and Patient, 2013; Levy, 2005; Tuggle, Sirmon, Reutzel, and Bierman, 2010) and organizational structures (Joseph and Ocasio, 2012; Maula et al., 2013; Palmi, Lingens, and Gassmann, 2016; Stevens, Moray, Bruneel, and Clarysse, 2015). Although

a large number of technology topics will be proposed by organizational members, only a subset will establish the often limited foci of attention within the organization (Dutton, Ashford, O'Neill, Hayes and Wierba, 1997).

The topics to which organizations attend serve as an impetus for strategic decision making (Mintzberg, Raisinghini, and Theoret, 1976; Nutt, 1984) and are a prerequisite to managerial activity aimed at improving firm performance (Keisler and Sproull, 1982). Attention patterns shape the environmental signals to which managers are most attuned, bringing salience to certain segments of the information environment and filtering out peripheral environmental stimuli. Accordingly, a firms' patterns of attention to particular technology-related topics guide how resources are allocated in the acquisition and development of knowledge and coordination of activities in support of utilizing and integrating that knowledge - such as developing new technologies and integrating them into platforms and products.

This study focuses on how attention affects entry timing in response to a pioneer entrant. I define a follower's entry timing as the time between a pioneer's launch of a new technology feature (e.g., Bluetooth, USB) and the focal firm's launch of the same technology for its own product. Prior research has identified several inter-firm differences in entry timing, including efforts to capture uncertain opportunities (Eggers, 2012b; Wernerfelt and Karnani, 1987), firm capabilities (De Figueiredo and Kyle, 2006; Lee, 2008; Mitchell, 1991; Hawk, Pacheco-de-Almeida and Yeung, 2013; Robinson and Chiang, 2002), industry characteristics (Giachetti and Lampel, 2010; Koski and Kretschmer, 2010; Putsis and Bayus, 2001), and competitors' attempts to capture new product categories (Giachetti and Dagnino 2013; Sorenson, McEvily, Ren, and Roy, 2006). Here, I complement with cognitive explanations that consider organizations' attention.

Entry timing is an important component of technology strategy because moving early or late accounts for significant variations in firm performance outcomes (e.g.,

Gomez and Macas, 2011; Lieberman and Montgomery, 1988; Mitchell, 1991; Suarez and Lanzolla, 1998). Although findings are mixed as to whether early or late entry is better for performance (Golder and Tellis, 1993; Klingebiel and Joseph, 2016; VanderWerf and Mahon, 1997), the literature agrees that both strategies are fraught with risk (Klingebiel, 2017). Late movers often cede revenue premiums or market share to early entrants (Banbury and Mitchell, 1995). Early movers face technical and demand uncertainty, and firms might easily allocate attention to the wrong technologies and products (Klingebiel and Rammer, 2014; Vidal and Mitchell, 2013).

Implicit in much of the extant research is that attention to a fixed set of issues provides criteria and shared parameters for directing resources in support of new technologies (e.g., Eggers and Kaplan, 2009). Much of it focuses on how top management attention directs activities within the organization and emphasizes attention to a given set of external topics, technologies, or information sources (e.g., Li et al., 2013). Dynamic industries offer a vast and changing array of technology topics that compete for managerial attention. Hoffman and Ocasio (2001) note that firms demonstrate industry-level attention, which reflects how industry participants selectively focus their attention on a limited set of topics that represent potential opportunities for the industry. Consistent with this notion, I consider all technology topics available in the industry and firms' concentration vs. dispersion of attention to those topics.

Also, prior research commonly prioritizes the innovation implications of attention within the boundary of a single firm (e.g., Dahlander et al., 2016; Wilson and Joseph, 2015). The idea of industry-level attention opens the potential for firms within an industry to have similar attention patterns. Research shows that industry players often develop similar cognitive maps (Barr , Stimpert, and Huff, 1992; Nadkarni and Barr, 2008) or perceptions of competitors (Porac, Thomas, and Baden-Fuller,, 1989), which guide subsequent and often similar competitive actions (Miller and

9

Chen, 1994). The same may be true for attention focus within an industry. In particular, no studies have examined the implications of a common attention focus on industry topics between a focal firm and the pioneer of an industry technology, which may be especially important for entry timing.

In what follows, I examine the links between the market introduction of new technologies and the configuration of a firm's attention within the research domain. I emphasize attention to technology topics and discuss mechanisms that directly relate topic attention to entry timing. In particular, I introduce two types of attentional focus: attentional dispersion and attentional similarity. Attentional dispersion (conversely, concentration) reflects the degree to which the firm focuses broadly (conversely, narrowly) on a set of technology topics or themes. Attentional similarity is defined as the degree to which two firms share a similar focus of attention on technology topics.

2.2 Hypotheses

2.2.1 Attentional Dispersion and Similarity

Attentional dispersion is likely to be consequential for entry timing for two reasons. First, more dispersed attention may improve the detection of environmental cues, trends, and events. Greater dispersion of attention may help liberate the firm from parochial vision and increase the probability that subtle cues that exist throughout the technological landscape will be encoded appropriately rather than overlooked or misinterpreted (Levinthal and Rerup, 2006; Weick and Sutcliffe, 2006). In technological spheres, these cues are often ambiguous and unanticipated (Rerup, 2009) and require attention on multiple topics to increase their salience. Picking up such cues helps the firm respond to emergent technologies and enables faster allocation of resources to developing those opportunities (Van de Ven, 1986; Yadav et al., 2007).

Accordingly, attentional dispersion will increase the amount of information to

which managers are exposed and so improves the odds that those managers will select the most valuable knowledge for developing new technologies (Laursen and Salter, 2006; Leiponen and Helfat, 2010; Li et al., 2013). It may also yield a better understanding of how new technologies fit with existing platforms or product lines. For example, researchers found that external knowledge was only useful to the firm when managers directed more attention to those sources, because it required more effort to understand how it fit with internal knowledge (Dahlander et al., 2016).

Second, greater dispersion of attention is likely to provide a better understanding of opportunities for recombining knowledge and routines. The speed of technology innovation relies heavily on the ability of firms to not only identify new knowledge (Ahuja, 2000), but also connect previously unrelated ideas and recombine previous combinations of knowledge in novel ways (Levinthal and Rerup, 2006). For example, in the mobile device industry, a focus on radio frequency transmission and on photo image quality (two topics) may lead a firm to make or buy an upgraded camera, and create a better understanding of how that camera should be combined with existing platforms and technologies that can be incorporated onto a new device. Mindful attention to a more diverse set of topics may allow for more rapid recombinations of knowledge and integration of the most promising technologies into new products. Conversely, although high focus on a given topic may increase the speed of introduction for the relevant technologies, dispersion may result in greater speed on average when taken the whole spectrum of technologies into account. This suggests the first hypothesis.

Hypothesis 2.1 [H2.1]: Greater attentional dispersion to industry topics is related to faster introduction of new technology features following a pioneer's entry.

Greater similarity in attention to an industry pioneer increases the chances that the time and effort given to developing particular technologies will parallel that of

11

a pioneer. Similarity in attention improves the likelihood that the focal firm and pioneer notice the same changes in the technological environment and incorporate them into the focal firms' technology roadmaps and plans. Attention to similar topic areas also suggests that time and effort of the focal firm is devoted to finding and exploiting similar knowledge as the pioneer, which leads to similar discoveries. It also aids in detecting components and combinations to avoid and gives intuition about knowledge and routine recombinations that are likely to yield the best results.

Although typically overlooked in studies of attention, attentional similarity can arise from either chance or interactions within the industry. Common focus may be purely stochastic, with overlaps reflecting a confluence of various factors. However, attentional similarity may also be derived from the active efforts of managers to follow technology leaders and the interaction of managers from different firms. Research has shown that representatives from different firms in the same industry often attend the same industry meetings and that this is consequential for sharing information and opportunities between firms (Rosenkopf, Metiu, and George, 2001). Also, industries are frequently characterized by movement of engineers or scientists between firms (e.g., Almeida and Kogut, 1999; Argote and Ingram, 2000; Song, Almeida, and Wu, 2003). These people are likely to bring their attentional patterns with them, and consequently their corresponding knowledge focus, even if there are NDAs in place (Mawdsley and Somaya, 2016). Finally, firms within the same industry often share suppliers (Martin, Mitchell, and Swaminathan, 1995); this is particularly true of the mobile device industry (Klingebiel and Joseph, 2016) where often only a few suppliers are capable of meeting the necessary quality and quantity demands. Suppliers such as providers of chips, hardware, and software may serve as a conduit of attention, directing manufacturers toward particularly important topics.

Even if the firms do not share technology roadmaps (due to other factors such as different experiences, resources, capabilities, or priorities), similar attention patterns

suggest the focal firm is well positioned to respond quickly to the pioneer. If the firm is caught off guard by a pioneer's launch, similar attention patterns as the pioneer will facilitate faster recovery and quicker launch of the same technology for their own product. This suggests a second hypothesis.

Hypothesis 2.2 [H2.2]: Greater attentional similarity to industry topics between a focal firm and the pioneering firm is related to faster introduction of new technology features following a pioneer's entry.

By extension, I expect the combination of high attentional dispersion and similarity to pioneers to be especially helpful in moving early. Attentional similarity to a pioneer's technology positions the focal firm to discern the integrative solutions developed by pioneers, while attentional dispersion allows the firm to better integrate features with other existing platforms, technologies and products. Hence, firms with both patterns can more quickly develop a particular technology and follow the pioneer. This suggests Hypothesis 3.

Hypothesis 2.3 [H2.3]: The greater the combination of attentional dispersion and similarity, the faster its introduction of new product features following a pioneer's entry.

2.2.2 Hits and Flops

An overlooked aspect of entry timing studies is the idea that some features are not worth following (Klingebiel and Joseph, 2016). Only a subset of technologies eventually become hits, i.e., technologies that generate feature revenue premiums because they either command a high price or generate greater sales. These premiums are generally short-lived and available only during the first few months following the launch of a feature. By contrast, flop features achieve only average returns (reaching average returns is viewed as a flop because products require superior returns to justify ongoing investments in dynamic technological sectors) or below-average returns over

the life of the product.

The distinction between hits and flops is important in high-tech industries, which are characterized by continuous technology innovations, frequent product introductions, and fast product obsolescence. These patterns reflect managerial expectations that only a subset of technologies and products will succeed, and that it is often necessary to launch many products early in hopes of creating a few hits amid many flops. This is especially true among early movers who compensate for the higher risks of launching early by launching a variety of features (Klingebiel and Joseph, 2016). Attention patterns among these early movers may provide insights about differences between entry of both hits and flops.

First consider hits. I expect that firms with greater attentional similarity to the pioneer of a given feature (i.e., such that similarity relates to a particular combination of feature and pioneer) will introduce the same feature more quickly. Since similarity in attention patterns attracts similar environmental stimuli and directs similar learning activities and resource investments, pioneers and other early movers are likely to launch in a similar fashion. Nonetheless, the effects for hit technologies should be particularly strong. Hit products may provide stronger incentive for firms to enter quickly, because they only offer excess returns during the first few months following launch, before dropping to more modest levels (Klingebiel and Joseph, 2016). Similar attention patterns will allow the firm to make the most of market information to allocate resources to features with the greatest expected returns.

Hypothesis 2.4 [H2.4]: Among hit features, firms with greater attentional similarity relative to the feature-pioneering firm will introduce the features faster following a pioneer's entry.

Now consider flops. Although early evidence of a flop may disincentivize attentionally similar firms from launching the feature, their technology trajectory and

resource commitments to suppliers and operators will make it difficult to execute a complete withdrawal (Klingebiel and Joseph, 2016). However, one mechanism that may provide the firm with the necessary flexibility to shift away from a potential flop is attentional dispersion. Among firms well positioned to follow quickly (i.e., they are attentionally similar to a feature-flop pioneer), those that are attentionally dispersed are also better positioned to pivot away from the flop.

Attention dispersion's capacity for providing the firm with access to greater and more appropriate knowledge means they may have a greater number of alternative technologies near launch. Dispersion provides the capacity for spontaneous recombination of existing knowledge, which means they may have more integration opportunities than a highly focused firm. Plus, rather than merely random combinations, they are likely to have a better sense of more apt combinations (Levinthal and Rerup, 2006). Hence, I propose the following hypothesis.

Hypothesis 2.5 [H2.5]: The greater the attentional similarity to a pioneer, the more that attentional dispersion will reduce the likelihood of introducing feature flops following a pioneer's entry.

2.2.3   Recovery from Flops

Flops are inevitable, especially in industries facing irreversible investments in R&D (DiMasi, Hansen, and Grabowski, 2003) and characterized by complex, dynamic environments (Deeds, Decarolis, and Coombs, 2000; Keil and Robey, 1999; McGrath, Keil, and Tukiainen, 2006). Commercial challenges may be a function of low demand for the particular technology, better alternatives, or limited supply. Flops are problematic for firms in that they divert resources from more productive uses, delaying the introduction of future hit features (Eggers, 2012b).

The implications of introducing a flop for what I call recovery - i.e., the ability to undertake subsequent entry - are threefold. First, focusing attention on a losing

technology suggests that other alternatives will receive less attention. Managers are limited in what they attend to, and a focus on a particular technology may preclude time and attention being dedicated to other potentially more promising options. Their efforts to introduce a flop may reflect attention to the wrong topics and crowd out focus on areas that might have allowed them to pursue more commercially successful technologies. This attentional foreclosure on alternatives means that effort and resources that were targeted at the losing feature now have to be repurposed for alternatives that will inevitably delay their subsequent launch.

Second, failure may impede learning (Cannon and Edmonson, 1999; Eggers, 2012b), a necessary condition for subsequent hit introduction. Learning from failure is difficult because managers often make incorrect causal references and make false conclusions when trying to assess the source of failure or, in many cases, leave the firm after the failure. The impediments stem from the complex nature of interactions between internal and external factors, and the fact that it is often difficult to pinpoint the actual problem (Lant, Milliken, and Batra, 1992). The probability of success assigned to lesser-known alternatives may be discounted (Denrell and March, 2001), making those alternatives less attractive to decision makers.

Third, subsequent hit introductions may be delayed by internal politics or demands for greater due diligence prior to launch. Research shows that investments in failing technologies can often lead to further path-dependent investments in those same technologies and may create resistance to alternatives (Guler, 2007). Relatedly, senior managers may demand greater due diligence on subsequent feature launches. For example, they may demand greater scrutiny in the form of more extensive development and testing, more extensive marketing research, or larger trials. They may also demand escalation of those approvals, which could further delay action on a potential subsequent hit. In all, this suggests the next hypothesis.

Hypothesis 2.6 [H2.6]: The longer a firm takes to introduce a flop feature following a pioneer's entry, the longer on average the firm will be to introduce a future hit feature.

Attentional dispersion may help offset the negative effect of launching a flop feature, for three reasons. First, dispersion of attention may provide a mechanism to counter some of the heretofore mentioned impediments to following a flop with a hit. If firms are attentionally diverse, they are less likely to have to completely repurpose existing resources and capabilities (Keil, McGrath, and Tukiainen, 2009). This is primarily because their breadth of focus allows them to break out of the current losing technologies and pursue more promising alternative features. Diverse attention may serve as a way to break the inertia of existing routines (Leonard-Barton, 1992) and alert managers of alternatives outside core or established features. More diverse cues may alert firms to alternatives, even if not initially prioritized, and make it easier for the firm to recover from a flop more quickly.

Second, attentional diversity may provide greater insights as to the source of the earlier failure. Such diversity offers a greater spectrum of source of potential problems and may provide greater transparency on causal relationships (Nadkarni and Narayanan, 2007). As a result, attentional diversity means firms are less subject to bias in understanding such relationships and interpreting the cause of the flop incorrectly.

Third, less due diligence may be required following the flop since the firm will likely already have invested to probe the general space and test the technology. Dispersion may reflect a top or middle management that has already approved, legitimatized, and supported the new technology topics (Kaplan, 2008), which makes it easier to recover with a new more successful alternative. Hence, I suggest the following hypothesis.

17

Hypothesis 2.7 [H2.7]: The greater the attentional dispersion of a firm that launches a flop feature following a pioneer's entry, the faster it will introduce future hit features.

2.3   Methods

2.3.1   Data

The setting of the analysis is the German mobile device market between 2004 and 2008, which I selected for three reasons. First, the basis for competition during the period of study was differentiation grounded in the introduction of new technology features. The setting reflects a time when the bulk of innovation activity within the industry was focused on adding functionality and equipping products with new features (Giachetti, 2013). Second, limiting the analysis to Germany eliminates concerns about heterogeneity across countries since phone manufacturers tended to develop marketing strategies for countries in each major region, due to variations in network standards and consumer preferences (Walkley and Ramsay, 2011). Interviews with industry contacts confirmed that, among European countries, feature-phone products typically launched first in Germany, limiting problems if entry behavior depended on market experience beyond our observations. Third, a focus on a single industry allows for more accurate identification of corresponding technology topics and measurement of firm attention to those topics.

Monthly device-level data was provided by GfK, one of two market research organizations (the other is Informa) that cover the mobile telecommunications sector. GfK's product database provides handset sales units, price, retailer penetration, and multiple feature dimensions. I compared data from Informa, which uses different methodologies for collection, to check reliability. To add variables relating to handset features, I supplemented the GfK from advice Ibsites such as GSMArena.com, Inside-Handy.de, Handy-MC.de, PDAdb.net, and PhoneArena.com, which provided

consumer-relevant technological aspects of mobile phones. The final sample consisted of 54 mobile phone features with accompanying month-level data for nine firms that have a combined market share of 90 percent at the start of the sample period based on handset revenues in the prior 18 months.

Patent information for firms was extracted from the European Patent Office's Master Documentation Database (DOCDB), which contains worldwide coverage of bibliographic data, abstracts, and classifications from over 90 countries. I collected patents at the application level rather than restricting on the set of granted patents to be more comprehensive in the scope of measurement for firm attention. R&D and sales data was derived from COMPUSTAT and publicly available data.

2.3.2   Variables and Empirical Approach

Dependent variables: Feature technological entry timing. The estimation of entry timing uses a Cox hazard rate model with entry rate as the dependent variable. I define an entry as the first appearance of a particular technology feature (e.g., camera, USB) on a new product of the focal firm. Firms enter the risk set when the pioneer introduces a particular technology feature. The duration variable is the number of months elapsed since the pioneer introduced the feature or when right censoring occurs. The database shows whether and when each firm introduced a feature into its mobile phone portfolio during the 36 months following the pioneer's launch.

Topic modeling of attention to technology topics. A major empirical challenge was to construct a reasonable measure of a firm's attention focus on industry and pioneer technology topics. Prior attention research has relied heavily on text analysis, since the words used in company documents are reflective of the firm's attention to particular themes or categories (e.g., Bettman and Iitz, 1983; Abrahamson and Park, 1994; Cho and Hambrick, 2006; Kaplan, 2008). This approach required

19

identifying all possible technology topics available in the mobile device industry and creating a measure that would capture the attention patterns of firms' technology organization rather than that of the CEO or top management. Also, it was necessary to capture firm attention patterns on a monthly basis over the period of study because the dependent variable (technology entry timing) was measured in months. I chose to analyze patent abstracts, which are appropriate because they reflect a cognitive focus distinct from that of quantitative measures of patent classes or claims (cf. Kaplan and Vakili, 2015), and summarize the novel aspects of an invention.

Given the need for a quantitative measure of topic attention over a large amount of text, I use a technique known as topic modeling to properly identify the technology topics attended to in patent abstracts. Probabilistic topic modeling is a text analysis tool that is based on the Bayesian statistical technique of latent Dirichlet allocation (LDA; Blei et al., 2003), which uses the co-occurrence of words in documents to identify the latent themes underlying those documents. The method allows for the identification of the main topics that pervade a large and otherwise unstructured collection of documents (Blei, 2012). This technique has benefits over direct word counts in that it allows for coding of a large collection of documents (i.e., unsupervised algorithm), accounts for co-occurrences of words, and acknowledges that some words may mean different things depending on their co-occurrence. The technique, which originated in computer science, has been effectively used by strategy scholars to measure the emergence of new technology topics (e.g., Kaplan and Vakili, 2015) and characterize attention patterns in patent text (Wilson and Joseph, 2015).

To ensure precision of measurement, I first isolated a core set of patent applications from which I constructed the list of abstracts to be inputted into the topic model. I started with 585,785 patent applications filed by mobile devices firms from ten years prior to the beginning of the sample period up until the final observational year. Acknowledging that multiple jurisdictional patents may be filed on the same

invention (e.g., in U.S., Europe, and Korea), I selected the application with the earliest filing date. Given that the sample is populated by multi-business firms that are engaged in activities outside the focal product category, I used the Derwent World Patent Index (DWPI) along with international patent classifications (IPC) to further narrow the set of applications that were used for mobile devices (Stembridge, 1999). The corresponding DWPI category is telephone and communications (W01). I then use the DWPI guide's crosswalk to isolate the patents with the corresponding relevant IPCs in the data (H04L,M,Q). After this process, the input to the topic model (our corpus) is 74,021 patent abstracts.

For each firm in each month, I constructed a technological portfolio based on its set of previously filed patent applications. I used a two-step process to improve accuracy of measurement. First, given that patents may be sequentially assigned multiple IPCs over time as its technological applications expand, I select the set of core mobile device patents based on relevance of the first assigned IPC. This further limits the sample to 46,995 patent applications and provides assurance that the technological ideas contained in these applications originated from the mobile divisions of the firms within the sample. Secondly, to ensure that the latent topics present in the firm's portfolio of technologies are actively attended to at a given point in time, I utilized the legal event tables from PATSTAT and removed all applications that have been withdrawn, as well as granted patents that expired or were invalidated in court.

To implement LDA, I utilized the available topic models package for R (Grun and Hornik, 2011). I applied the topic model to the set of patent abstracts as follows. Following the conventional process of preparing the documents (patent abstracts), I first removed all stop words (such as and, or) and removed words with high and low frequency to aid in computation. I also analyze based on word stems (e.g., frequenc from frequency or frequencies). For the key parameters of the model - number of

topics K, and the distribution of topics to documents α - I follow prior convention (Kaplan and Vakili, 2015) and selected $K = 100$ and $\alpha = 0.01$. I then applied the topic model to the patent abstracts to find the hidden topic structure and, in particular, the joint distribution of variables, as follows:

$$p(\beta_{1:K}, \theta_{1:D}, z_{1:D} | w_{1:D})$$

In this representation, each β_k is a topic (distribution over words), for K number of topics; $\theta_{d,k}$ is the topic proportion for topic k in document d (topic mixture in each doc); $z_{d,n}$ is the topic assignment for the nth word in document d (word mixture in each topic); and $w_{d,n}$, is the nth word in document d. Using the topic modeling algorithm, I computed the inferred topic distribution across all patents and the most probable terms from each topic associated with each document. The model uses a variational expectation maximization procedure to compute the per-word topic assignment distribution of topics per document as well as the distribution of words per topic. Blei et al. (2003) provide additional details on the procedure.

From the topic distribution per document output, I assigned the topic with the highest probability of occurrence to each patent abstract. I then aggregated the assigned topics for patent abstracts over the firm's entire technological portfolio (all patent applications), observing the distribution of the firm's attention across topics on a moving time window through the sample period of analysis as described below.

Explanatory variables: Dispersion and similarity to technological pioneer. From the distribution of topics for a given firm's technological portfolio, the main explanatory variables are measured at the time the technological pioneer launches a focal feature, operationalized as follows. I characterize attentional dispersion as one minus the Herfindahl concentration ratio. That is, suppose I let i denote each topic in the set of n topics ($n = 100$ in our case). Then the attentional

dispersion of a given firm is defined as:

$$\textbf{Dispersion}_{firm} = 1 - \sum_{i=1}^{n} (\textbf{Proportion of abstracts in topic i})^2 = 1 - \sum_{i=1}^{n} \textbf{firm}_i^2$$

This allows my measure to range between 0 and 1. Firms that are highly concentrated in any topic (i.e., many patents reflect only a few topics) will have dispersion close to 0 while those that are evenly spread out across topics will have dispersion values close to 1.

I define attentional similarity between a focal firm and the feature pioneer's technological portfolios as one minus the sum of absolute differences in their topic proportions.

$$\textbf{Similarity}_{firm} = 1 - \sum_{i=1}^{n} |\textbf{firm}_i - \textbf{pioneer}_i|$$

The similarity measure ranges between -1 and 1. In this range, high (low) values correspond to high (low) overlap in the two firm's topic allocations. Zero is a neutral value denoting expected overlap of two random topic allocations (i.e., firms neither similar to nor dissimilar from each other). By using absolute differences in the topic proportions, the measure is invariant to the distribution of a focal firm's dissimilar topics - a desirable property for a similarity construct.

As a clarifying example, suppose Alcatel is focused on two topics (Topics A and B) that are outside the pioneer's domain (which commits all its attention to Topic C). Then the similarity measure remains unchanged whether Alcatel allocates its attention evenly across the two topics (0.5 on Topic A and 0.5 on Topic B) or chooses to put more relative weight on one topic (e.g., 0.9 on Topic A and 0.1 on Topic B). Note that although Alcatel may be more dispersed in the former case than the latter (Dispersion = 0.50 vs. 0.18), its similarity to the pioneer remains the same (Similarity = 0), which is conceptually valid.

Controls. Several feature- and firm-level control variables address potential confounding effects. Competitor adoption measures the number of outside firms that have commercialized a focal feature at the monthly level. This controls for the popularity of a given feature and the influence of competitor actions should externalities in commercialization decisions exist. R&D intensity accounts for the rate of knowledge input. This is measured as the firm's R&D expenditure divided by its sales in the preceding 12 months, which was taken from annual reports. To alleviate concerns that firms may serve different customer segments, I add a variable for the firm's price position, measured as the firm's scaled deviation from the average handset market price (centered on 0). Next, I include a measure of the firm's market presence (its penetration in the German market) to take into consideration the size and reach of its handset portfolio among active users. The market penetration variable is measured as a weighted percentage of the post-paid market tapped by the handset portfolio of the focal firm.

Given that the similarity measure relates to the degree of overlap between two firms' attention allocation over a set of common technological issues, I want to distinguish this construct from inter-firm knowledge transfer, which is traditionally measured by patent citations (Mowery, Oxley, and Silverman, 1996). This is important since I'm concerned with whether firms are solving similar technological problems, rather than how such problems are solved. (Although how firms solve problems is important in understanding the rate of technological development, it is beyond the scope of this paper.) To establish a size-neutral measure, I normalize the citations by the firm's number of patents.

Table 1 reports descriptive statistics. Among the controls, market penetration and price position have a moderate correlation (r=0.52). Among the conceptual variables, the main relationship of note is that dispersion and similarity have moderate correlation (r=0.58), leaving sufficient variation for analysis of both effects.

24

Table 2.1: Descriptive Statistics and Variable Correlations

		Descriptive Statistics					Correlations			
	A. Variables: Firm-year	Cases	Mean	s.d.	Min	Max	A1	A2	A3	A4
A1	Dispersion	68	0.917	0.085	0.728	0.981	1			
A2	R&D intensity	68	7.938	3.254	2.687	15.046	0.20	1		
A3	Price position	68	-0.176	1.722	-2.764	3.494	-0.02	-0.27	1	
A4	Market penetration	68	40.096	21.458	3.184	100.0	-0.11	0.09	0.52	1
	B. Variables: Firm-feature						B1	B2	B3	B4
B1	Similarity	383	-0.154	0.501	-0.996	1	1			
B2	Importance	400	0.112	0.109	0.002	0.515	-0.36	1		
B3	Pioneer citations	383	0.005	0.011	0	0.073	0.47	-0.20	1	
B4	Entry time	140	15.957	11.167	0	36	-0.66	0.05	-0.40	1
	B. Variables: Feature						C1	C2	C3	C4
C1	Hit feature	54	0.278	0.452	0	1	1			
C2	Competitor adoption	54	6.315	5.508	1	20	0.27	1		
C3	Entry time: mean	54	14.312	11.576	0	36	0.30	0.44	1	
C4	Entry time: standard deviation	54	4.212	4.792	0	14.4	-0.02	0.41	0.23	1

D. Firm-feature-month: Correlations		D1	D2	D3	D4	D5	D6	D7	D8
D1	Dispersion	1							
D2	Similarity	0.58	1						
D3	Hit feature	-0.01	-0.03	1					
D4	R&D intensity	0.16	0.14	0.03	1				
D5	Price position	0.05	0.00	-0.01	-0.34	1			
D6	Market penetration	-0.14	-0.18	0.06	-0.01	0.48	1		
D7	Importance	-0.31	-0.36	0.00	-0.25	0.01	0.26	1	
D8	Pioneer citations	0.25	0.47	0.02	-0.02	0.09	-0.13	-0.20	1
D9	Adoption firms	-0.10	-0.19	0.13	0.03	0.00	0.18	0.04	-0.07

2.3.3   Empirical Specification

In order to examine the baseline relationship between attention and entry (H1-H4), I applied a Cox proportional hazards regression model (Cox, 1972), a common model for survival data and entry studies in particular (e.g., Fuentelsaz, Gomez, and Polo, 2002). The Cox model is ideal for our analysis because the model makes no assump-

tions about the shape of hazard over time, and takes the following form:

$$h(t|x) = h_0(t) * exp(x\beta)$$

where $h_0(t)$ is the baseline hazard, x is the vector of explanatory variables, and β is the vector of regression coefficients.

Analyzing timing of technological entry requires addressing the unobservability of non-entrants, which are not incorporated into standard duration models. Therefore, I test the first set of hypotheses (H1-H3) by estimating the coefficients under a split-population hazard model (Schmidt and Witte, 1989). The split-population framework estimates an ex-ante probability of non-entry along with the marginal effects on entry timing for the subset of entrants.

To capture the probability of launch a flop following a pioneer's entry (H5), I use a linear probability model. To test for recovery from flops and the speed of introducing future hits (H6, H7), I utilize an ordinary least squares (OLS) regression.

2.4   Results

Table 2 reports results with the dependent variable measured as the number months from the pioneering date until the focal firm launches the technology. I present values in the form of hazard ratios to better allow comparison of magnitude. Coefficients take the form of positive numbers: those less than 1 imply a negative marginal effect on risk of failure (slower entry), while values greater than 1 imply less average time to launch (faster entry).

I find statistical support for the first three hypotheses, which concern the relationship between attentional dispersion and attentional similarity on entry timing relative to the pioneer. Firms that are more attentionally dispersed across technological issues tend to be faster to respond to pioneers (H1; Model 2: $\beta = 463.46$; $p < 0.001$). This is also true for feature entries where the launch firm is similar to the

26

Table 2.2: Split Population Hazard Model of Effect of Dispersion and Similarity on Follower Entry Timing

	1	2	3	4	5
Dispersion (H1 +)		463.46***		128.77***	388250.2***
		(778.10)		(241.15)	(1878998.0)
Similarity (H2 +)			3.004***	1.837*	0.00013*
			(0.976)	(0.659)	(0.000083)
Disperion*Similarity (H3 +)					222551.1*
					(1444340.0)
Competitor adoption	1.235***	1.269***	1.269***	1.282***	1.274***
	(0.030)	(0.034)	(0.033)	(0.035)	(0.038)
R&D intensity	0.935*	0.912**	0.900***	0.898***	0.908**
	(0.037)	(0.035)	(0.037)	(0.036)	(0.036)
Price position	0.913	0.883*	0.894	0.882*	0.888
	(0.068)	(0.065)	(0.067)	(0.066)	(0.066)
Market penetration	1.007	1.011*	1.010	1.011*	1.017**
	(0.006)	(0.007)	(0.006)	(0.007)	(0.007)
Pioneer citations	7259749.0	143.5	0.003	0.005	0.012
	(96400000.0)	(2001.0)	(0.039)	(0.072)	(0.180)
Constant	0.008***	0.00028***	0.0113***	0.0011***	.000004***
	(0.003)	(0.00005)	(0.005)	(0.00021)	(0.0000002)
Cure probability: constant	0.133**	0.185**	0.142**	0.183**	0.165**
	(0.112)	(0.124)	(0.110)	(0.122)	(0.130)
Log likelihood	-463.4	-455.3	-457.4	-453.8	-452.0
LL ratio test vs. Model 1 (df)		16.0 (1) ***	12.0 (1) ***	19.0 (2) ***	22.9 (3) ***
Observations (feature-firm-months)	7159	7159	7159	7159	7159

*** p<0.01 ** p<0.05 * p<0.10 (two-tailed tests; standard errors in parenthesis)

feature pioneer (H2; Model 3: $\beta = 3.004$; $p < 0.001$), although attentional dispersion has a larger average hazard ratio. When both variables are included in the specification (Model 4), I see that dispersion continues to dominate similarity in magnitude and significance. Nonetheless, even in the combined model, the relation between similarity and entry timing is at least marginally significant (Model 4: $\beta = 1.837$; $p = 0.09$). H3 receives moderate support (Model 5: $\beta = 222551.1$; $p = 0.058$); here, the results suggest dispersion and similarity may work together to speed up entry.

To test our next set of hypotheses concerning hits and flops, I utilized information on the success of features based on the revenue premiums generated in the early

months following a pioneer launch (Klingebiel and Joseph, 2016). In the sample, 15 of 54 features are classified as hits, generating monthly peak premiums over 2m euros (the benchmark for features to be classed as commercial successes); the remaining features are designated as flops. To test H4 and resolve the issue of non-entrants, I confine the analysis to hit features only. Since non-pioneers can observe the market outcome of feature introductions, one reason to avoid entry may stem from poor demand that is only revealed ex-post feature launch. Because this issue is absent among the subset of hit features, I analyze the impact of attentional similarity on the firm's launch speed in a traditional Cox-proportional hazard model. Assuming all firms have incentives to launch hit features, the results are more indicative of the firm's ability to launch faster.

Table 3 presents the results. From our hit sample analysis, I find support for H4. Firms that are more attentionally similar to the focal pioneer tend to launch hit features faster, even when controlling for dispersion (Model 4: $\beta = 4.662$; $p < 0.001$). Considering that the results are statistically significant in a greatly reduced sample, this gives us some assurance that similarity is potentially related to the firm's ability for quickly commercializing new technologies. I also compare the full specification to the sample counterpart for flop technologies (Model 5). Although the relationship still exists, it is noticeably less in magnitude, reinforcing support for H4 (Model 5: $\beta = 3.515$; $p = 0.016$). Additionally, it is statistically less significant among the larger flop sample, which can be symptomatic from unobserved non-entry.

Given our finding on the relationship between similarity and entry timing for hit features, I can segment the firms into two categories based on whether they are above or below the mean similarity to a pioneer for each feature. To test H5, I use this segmentation and confine the analysis to flop features. Specifically, I test whether, among similarly positioned firms in the technological space, those that are more attentionally dispersed are less likely to adopt an observed flop feature, perhaps due

28

Table 2.3: Cox Proportional Hazard Models of Impact of Dispersion and Similarity on Follower Entry Timing for Hit Features

	1. Hits	2. Hits	3. Hits	4. Hits	5. Comparison: Flops
Similarity (H4 +)		5.330***		4.662***	3.515**
		(2.373)		(2.115)	(1.831)
Dispersion			45.415*	11.694	1.108
			(102.429)	(29.577)	(2.764)
Competitor adoption	1.160***	1.247***	1.165***	1.250***	1.192***
	(0.049)	(0.058)	(0.049)	(0.059)	(0.040)
R&D intensity	0.988	0.931	0.971	0.926	0.837***
	(0.047)	(0.046)	(0.046)	(0.046)	(0.046)
Price position	1.123	1.048	1.117	1.052	0.790**
	(0.129)	(0.123)	(0.128)	(0.123)	(0.079)
Market penetration	0.990	1.000	0.995	1.003	1.021**
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Pioneer citations	1962.0	0.000	51.370	0.000	0.000
	(28524.1)	(0.000)	(745.724)	0.000	(0.000)
Log likelihood	-188.6	-181.3	-186.9	-180.8	-231.4
LL ratio test vs. Model 1 (df)		14.5 (1) ***	3.3 (1) *	15.5 (2) ***	
Firm-feature combinations	122	122	122	122	239
Number of launches	44	44	44	44	50
Observations (firm-feature-months)	1950	1950	1950	1950	5003
Time at risk (months)	3209	3209	3209	3209	5939

*** p<0.01 ** p<0.05 * p<0.10 (two-tailed tests; standard errors in parenthesis)

to more outside options.

Table 4 presents the results of a linear probability model on the likelihood of adopting a flop feature among attentionally similar firms. Contrary to H5, I find that greater dispersion is associated with a higher likelihood of adopting a flop feature (Model 2: $\beta = 10.360$; $p < 0.001$). One possible reason for this unexpected finding is that firms that are on the same trajectory as a pioneer (i.e., similar topics) yet have a particularly wide range of technological search (i.e., greater dispersion of technological topics) may believe that they can overcome the problems that caused the pioneer's launch to flop.

I compare the findings to a reference group, composed of firms that have similarity measures less than the mean (Model 3: $\beta = -0.551$; $p = 0.118$). The absence of a

Table 2.4: Linear Probability Model of Effect of Dispersion on Decision by Potential Followers to Launch Flops

	1. Similar Firms	2. Similar Firms	3. Comparison: Dissimilar Firms
Dispersion (H5 -)		10.360***	-0.551
		(2.241)	(0.350)
Competitor adoption	0.020*	0.0256***	0.036***
	(0.010)	(0.010)	(0.007)
R&D intensity	-0.015	0.000	-0.009
	(0.010)	(0.010)	(0.010)
Price position	-0.039	-0.004	0.019
	(0.034)	(0.032)	(0.019)
Market penetration	0.012***	0.019***	0.002
	(0.002)	(0.003)	(0.002)
Pioneer citations	1.444	3.694	-3.862
	(4.392)	(4.060)	(4.349)
Constant	0.008	-10.324***	0.474
	(0.135)	(2.238)	(0.332)
Adjusted R-square	0.18	0.31	0.23
Mean launch time (months)	18.1	18.1	23.7
Launches (sample = 50)	32	32	18
Observations (feature-firm-months; sample = 240)	115	115	125

*** $p<0.01$ ** $p<0.05$ * $p<0.10$ (two-tailed tests; standard errors in parenthesis)

dispersion effect for dissimilar firms conforms to our intuition that those unrelated to the pioneer can more easily reject underperforming features regardless of their level of attentional dispersion, potentially stemming from the absence of commitments to outside parties as aforementioned.

With knowledge of the subset of hit and flop features, I test the final two hypotheses, on recovery, by looking at the relation between adopting a flop feature and the time it takes to adopt a subsequent hit feature. The objective is to see whether launching features with lackluster returns had measurable market consequences in the form of delaying the future commercialization of high demand features. If it did, I wanted to know whether being dispersed in attention to various technological topics would offset the incurred disadvantage.

30

Table 2.5: OLS Estimates of Impact of Follower Flop Adoption and Dispersion on Launch Timing of Subsequent Hit Features

	1	2	3	4
Adopt flop	1.825	-9.341**		-7.031
	(2.187)	(4.595)		(4.567)
Time to adopt flop (months) (H6 +)		0.561***		0.489**
		(0.204)		(0.201)
Dispersion (H7 -)			-36.603***	-35.410***
			(11.158)	(11.192)
R&D intensity	0.276	0.257	0.452*	0.439*
	(0.244)	(0.241)	(0.245)	(0.243)
Price position	-1.106 *	-1.095*	-0.863	-0.765
	(0.567)	(0.560)	(0.558)	(0.559)
Market penetration	0.052	0.040	0.010	-0.018
	(0.057)	(0.056)	(0.056)	(0.058)
Pioneer citations	-150.68	-139.81	-68.169	-68.780
	(107.65)	(106.24)	(107.853)	(106.632)
Constant	4.851	5.282*	38.372***	37.708***
	(2.974)	(2.938)	(10.636)	(10.646)
Adjusted R-square	0.02	0.05	0.06	0.08
Observations (feature-firm-months; full population)	240	240	240	240

*** p<0.01 ** p<0.05 * p<0.10 (two-tailed tests; standard errors in parenthesis)

Table 5 presents the linear regression results using 240 firm-feature flops as the unit of observation. The dependent variable is the number of months since the flop's pioneering debut until the focal firm launches its next hit feature. Model 2 reports evidence for H6 that the more time a firm spends on commercializing a flop feature, the slower it is to launch the next hit feature (Model 2: $\beta = 0.561$; $p = 0.006$). In turn, Model 3 provides support for H7 that being attentionally dispersed helps to offset the negative effect of launching a flop feature (Model 3: $\beta = -36.603$; $p = 0.001$). Both H6 and H7 hold in the combined model (Model 4).

Combining the results with information from our statistics on mean launch time (14.3 months), I calculate that launching a flop delays introduction of future hit features by seven months on average. Meanwhile, an increase in dispersion of one standard deviation can speed up the launch of subsequent hits by three months. Such delays and accelerants are highly material given the steep decline in a hit feature's revenue premiums over the twelve months after debut.

31

2.5 Discussion

This paper examines how attention patterns across firms and industry topics may be consequential for a firm's early entry of feature technologies. I used data about features in the mobile device industry from 2004 to 2008; during this era, feature innovation was a key aspect of competition (Giachetti, 2013). The results support almost all predictions. I find that attentional dispersion and attentional similarity provide the means to be an early mover and, in particular, the capacity to quickly launch new technology features following the pioneer. In combination, attentional dispersion and similarity create a potent mechanism to increase the speed with which technologies can be launched; hence the results demonstrate an important role for multiple types of attention patterns in predicting innovation outcomes. I also find that similarity is related to faster entry timing of both hits and flops. Though the effect for hits is stronger, the effect for flops is significant and material. Particularly novel is the finding that attentional dispersion both speeds the entry timing of subsequent hits for all firms but, unexpectedly, increases the likelihood of launching an observed flop for similarly positioned firms; I discuss this result in greater detail below. The study offers contributions to the literature on managerial cognition and to technology strategy studies of market entry timing.

2.5.1 Contributions to Managerial Cognition

The emphasis on attentional dispersion and similarity expands conceptual understanding of how managerial attention affects the development of new technologies. Prior research has implicitly assumed that firms narrowly focus attention on either a particular technology or alternative, or on external vs. internal sources of information (Dahlander et al., 2016). Missing from the conversation is the idea that there are always many topics, themes, and ideas in the air in a given industry at any point

in time (Nelson and Winter, 1982). The challenge for any firm is the breadth of topics to attend to. The unique empirical approach utilized allows me to address this conceptual issue by assessing the emergence and shifts in the totality of potential topics within the mobile device industry during the period. Correspondingly, I can develop insights about firms' similarity in the subset of industry topics they attend to and how this affects entry timing.

The contribution aids in resolving a division in the literature concerning the merits of attention focus. Collectively, this research is split on whether a narrow or broad attentional focus is beneficial for firm performance. Some studies suggest that a narrower CEO focus is especially important for entry timing in new technologies (e.g., Kaplan, 2008), because it provides legitimation and support for what would otherwise be a risky move (Eggers and Kaplan, 2009). Other studies suggest that a broader focus limits biases in decision making (Nadkarni and Narayanan, 2007) and provides the necessary exposure to cues and knowledge that would be otherwise ignored, increasing the rate of new technologies (Li et al., 2013). Approaching the question of attentional diversity by contextualizing it in the realm of industry attention, I help resolve these differences.

My analysis of hits and flops highlights a conceptual contribution of attention focus - that of aiding recovery from a flop. I find that firms attentionally similar to the feature pioneer are quicker to introduce both hits and flops. Among similar firms, dispersion increases the likelihood of technological entry with respect to flops.

Although the relationship between dispersion and flop entry likelihood was un- expected, it may reflect the wider context of all available technologies. Taken in isolation, there may be little benefit or even detriment to copying a focal technology, especially those that exhibit poor market performance. However, attentional disper- sion may help a firm refine a failed technology introduced by a competitor, either by improving its quality or by pairing it with another new technology. Given the

nature of complementarities between component technologies in our setting, such as Bluetooth and MP3 playback or upgraded cameras and high-resolution screens, attention to a wider array of technological topics can increase the marginal value of incorporating other component technologies. Firms that are attentionally dispersed may be better able to harness the value of other phone features and would choose to integrate them into their product portfolios rather than abandon them. In this way, our conceptualization of recombinant possibilities applies not only to technological alternatives outside of flops, but also to flops themselves. Given that most phone features are not hits, the recombination possibilities may be greater among the set of flops.

While full analysis of interactions between component technologies is beyond the scope of this paper (and our data), I do find that firms tend to bundle features. In supplementary analysis, I examined the case of Bluetooth Audio (A2DP) as a long-lasting hit feature that was pioneered early enough that I can track the subsequent actions of follower firms over an extended period. Among the seven firms in our sample that eventually adopted A2DP, the four firms that were earliest to do so (HTC, LG, Motorola, Nokia) are also pioneers of subsequent features, whereas the non-adopters - Alcatel, Blackberry, and Sagem - are not subsequent pioneers. Moreover, three out of the four future pioneers commercialized their next technology quickly (all but HTC), within six months of the adoption of A2DP.

More generally, these findings inform the literature on learning from failure. Prior work has generally argued that failure tends to dampen efforts to generate new knowledge regarding other risky or uncertain alternatives (Denrell and March, 2001). For example, Eggers (2012b) finds that firms which initially pursue a failing technology will subsequently invest less in developing knowledge in a winning alternative. Likewise, Guler (2007) finds that managers often increase investment in losing technologies in which they have already invested - rather than more successful alternatives

- an argument in line with theories of escalation of commitment. I add to this discussion in that high levels of attentional dispersion may provide the mechanism to advance recovery from such choices. It aids in keeping the firm flexible and less likely to rely on existing routines and knowledge. Attention dispersion may facilitate the acquisition and development of new knowledge and recombination of existing routines; that is, it allows for increasing learning at a time when learning normally may constrict information flow, making it easier to pivot to an alternative technology or novel combinations.

Empirically, the study has implications for researchers interested in exploring attention. Prior studies have used multiple methods including case studies (e.g., Rerup, 2009), surveys (Li et al., 2013), and archival data (Kaplan, 2008). Here I chose text analysis based on topic modeling because it allows large-scale quantitative analysis of attention based on patents. Wilson and Joseph (2015) used this approach in their analysis of attention focus at a single firm, Motorola. Here, I show it can be used in a multi-firm sample.

2.5.2 Contributions to Entry Timing

The study also provides insights into the literature on market entry timing. Prior work typically focuses on advantages and disadvantages of entry timing as well as the relationship between innovation timing and firm performance (Lieberman and Montgomery, 1988; Mitchell, 1991). Other research in this domain has examined firm-level resources and environmental factors that condition early-mover advantages (cf. Fosfuri et al., 2013). The role of managers is largely implicit in this research. Given that early entry depends heavily on a firm's assets (Vidal and Mitchell, 2013; Franco, Sarkar, Agarwal, and Echambadi, 2009, Klepper and Simons, 2000) and the managerial cognitive capabilities with which the firm wields those assets (Helfat and Peteraf, 2015), cognitive explanations merit more light. While I begin to make inroads with

35

this paper, future research may want to make the links between managerial attention and specific resources more concrete. For example, to what degree does attentional dispersion and similarity affect the firm's ability to draw on complementary assets to speed entry (Mitchell, 1989)?

An additional contribution to this literature is our attention to hits and flops. Much of the literature examines entry timing either by selecting successful products or with little regard for whether the products, markets, or technologies are worth following. The emphasis is on isolating mechanisms that give early movers advantages (Lieberman and Montgomery, 1988; Gomez and Macas, 2011). However, in the mobile devices industry and others like it, a continual stream of new products helps firms to differentiate. The hit rate for such industries varies widely, as it does in other sectors such as pharmaceuticals and movies, and inevitably some features will be unsuccessful. The finding that introducing flops delays introduction of future hits is important, suggesting it will make firms laggards in subsequent rounds of entry that are more relevant for success.

Relatedly, I elaborate on the benefits of recovery. Recovery through attentional dispersion appears to be an important managerial cognitive capability that could aid firms in such industries. Of course, there are also costs to dispersion since the firm may not be able to capture the same economies of attention and future work may want to consider such factors.

2.5.3   Managerial Implications

The findings have implications for managers. Results for recovery suggest that executives should consider carefully the upstream attentional patterns of its engineers and scientists. More generally, I show that topic focus is a worthwhile strategic consideration. Topic focus goes beyond specific patents, claims, or classes to identify the underling themes that unite them in support of developing new products.

Similarity suggests executives should be aware that attention focus, like ideas, may be transferable and spill over. People attend the same conferences, patronize the same coffee shops, or move between firms and take their attentional perspective with them (Ocasio, 2011). This leads to similar attention patterns at rival firms, possibly shortening the window of revenue premiums that pioneers can enjoy. Dispersion reflects a broader cognitive pattern, providing the foundation for fast followership and recovery. The pattern indicates that firms may be selective in their efforts following flops and dispersion provides them more discretion concerning subsequent feature launches.

2.5.4 Limitations

This study has four limitations that offer avenues for future research. First, I focus on a specific period in a single industry. This provided us with a setting in which the basis for competition was based on feature innovation and a time frame within which this environment remained largely unchanged. Nonetheless, there may be different dynamics in industries that require significant breakthroughs or have extensive network externalities (which came later in the mobile device industry, with the arrival of Apple and Android).

Second, this paper is a correlational analysis, and I'm careful not to draw strong causal relationships in the results. Although the temporal sequencing of attention and entry accommodates some of these concerns, I cannot entirely rule out endogeneity - firms may plan to launch technologies and therefore attend to certain areas. However, given that the timing of development in this industry was 12 to 18 months, and starting a major research program within the R&D organizations of these firms would have been a considerable undertaking, it is unlikely this was the case. Moreover, even if the firm's intention to launch technologies influenced its attention to topics, such attention may affect the speed of subsequent technological

introductions, which our empirical framework will detect. A more problematic case of endogeneity can arise if an unobserved factor influences a firm's mimetic behavior with respect to both intention to launch similar technologies and attention to similar topics as a pioneer; this could create spurious correlations that might be mistaken for causal relationships.

Third, the analysis rests on the construction of the firm's technological portfolio at particular points in time. From sensitivity analysis, I see that both lagging the firm's technology portfolio by one year and confining our measures to only new technologies in any given year would greatly diminish the explanatory power of our main variables. This indicates the importance of precise measurements in time so that our hypothesized constructs may take effect as well as the critical nature of the firm's past and present technologies. At the same time, because technology portfolios do not change much over the short duration of a five-year sample horizon, I lose intertemporal variation. This prevents me from picking up significance with the inclusion of firm fixed effects. Given the nature of shifting technological paradigms (Dosi, 1982), I accepted this trade-off to focus on a carefully chosen period when the rules of the game for innovation in the mobile device industry remained largely unchanged.

Fourth, an implementation of the split population hazard model to tease out a mechanism does not completely solve the problem of non-entry. I get close by restricting our analysis to hit features, but cannot say with absolute certainty about the firm's intention to follow the pioneer. Aside from the follower firm's intention, I cannot be certain whether firms were permitted to follow based on a bevy of outside factors such as government regulation. However, the fact that during the time of this study, intellectual property rights to the technologies resided in upstream firms rather than the downstream competitors from which the sample was drawn provides a degree of confidence in the results.

3

A Theory on Firm to Firm Matching

Understanding the determinants of value creation among firms and the process of managerial decision making that lead firms to realize such value is a central issue for strategy scholars. Given that firms interact within dynamic environments, the externalities from the decisions of some firms on the choices of others have received scant attention. In the context of the market for inter-firm collaboration, the formation of alliances between companies may limit the choices or even foreclose opportunities for remaining firms seeking potential partners (McDonald and Ryall, 2004). Accounting for congestion of potential partners, recent works have begun utilizing matching models in various contexts (Yang, Shi and Goldfarb, 2009; Fox, Yang, and Hsu 2017). Matching models work by limiting the number of potential "matches" for each participant within a market and characterize the outcomes that would emerge given a set of payoffs for each match.

However, empirical work has mostly relied on the matching framework of Shapley and Shubik (1972), which permits the division of total joint value between partners within a potential match to be committed prior to the formation of the match. While the endogenously determined "transfers" of utility (TU) between partners can be

appropriate in certain settings such as mergers and acquisitions, which more closely resemble the underlying processes of an auction (Sotomayor, 2009), the market for alliances is fundamentally different. In an alliance, the value is not realized at the moment two firms are matched, but only after the collective actions of partners have been executed. Firms typically cannot commit to a division of value because contracts are incomplete and prior agreements are renegotiated (Arora, Belenzon, and Patacconi, 2016).

As collaboration projects between firms typically require substantial time and project-specific investments for execution (Williamson, 1979), alternative potential partners cease to function as outside options for those currently engaged within an alliance. This means the role of renegotiation becomes important since the division of surplus (i.e. net value from an alliance) between partners will depend on the outside options of each partner at time the surplus is realized. Given that outside options will be different from those available before the alliance is actually formed, firms select partners with the expectation that division of value within any match cannot be affected by the availability of alternative alliances. This leads naturally to a non-transferable utility matching framework where the division of value within a match is exogenous to the matching process (e.g. Hatfield and Milgrom, 2005).

I develop a non-transferable utility matching model between firms from a many-to-many two-sided market, which I will categorize into upstream and downstream firms. A key assumption is that the division of the surplus will follow a fixed-sharing rule, which subsumes the case of Nash-Bargaining between partners that would result in a one-half split of the generated surplus within an alliance. I utilize an intuitive equilibrium concept (Roth, 1984), which states that given the alliance market outcome, all matches have positively-valued surplus and no subset of firms can make themselves better off by forming matches among themselves (possibly at the expense of dissolving matches with firms outside the subset). I adapt a

Greedy algorithm to the many-to-many market setting and show that it terminates in the equilibrium outcome, demonstrating that such an equilibrium exists. Moreover, since the Greedy algorithm always terminates in a unique outcome given the model assumptions and any violation will create a violation of the equilibrium criteria, the existing equilibrium is unique.

This work contributes to the alliance literature by shedding light on the existence of matching frameworks with non-transferable utility (NTU) that are more relevant for an inter-firm collaboration market. More broadly, it contributes to matching markets where agents' payoffs within a match are not immediately realized and cannot be credibly committed at the time of formation. In this paper, the presence of portfolio effects is not addressed. I focus on the base case when the total value of an alliance set is a sum of individual alliance values. Understanding the basic matching framework without portfolio effects helps to lay the foundation for incorporating additional features such as complementarities among matches, which has recently been studied for the case of a many-to-one matching market (Pycia, 2012).

This work also contributes to the matching theory literature. Although the case of a fixed-sharing rule between partners can be classified more generally under aligned preferences (Sorensen, 2005) with the same corresponding equilibrium existence/uniqueness theorems, I formally extend these results to a many-to-many matching market. This is a nontrivial extension and the most relevant works on the stability properties of such markets have required a refinement of individual preference characterizations (Echenique and Oviedo, 2006) or the notion of stability (Hatfield and Kominers, 2017) to ensure the existence of a stable outcome. I demonstrate that in the special case where a fixed sharing rule determines the division of value between partners, the equilibrium (i.e. the stable outcome) is unique. Secondly, having established the existence and uniqueness results, this paper shows a set of Pareto optimal characteristics that are not immediately obvious while demonstrating

plausibility of the equilibrium outcome.

The next section provides a more extensive background on the matching theory before proceeding to the formal analysis of section 3. Within section 3, I start off with a formal result by showing the equilibrium can deviate from the transferable utility outcome. Next, I establish a result that sufficiently characterizes the equilibrium by only considering individual firms and pairs of potential matches, setting up a criterion that will be used for establishing an estimator in the next chapter of the thesis. After showing the existence and uniqueness of the equilibrium, I demonstrate Pareto efficiency for single-partner firms in the special case of a many-to-one matching market before showing the Pareto efficiency of all firms within a many-to-many market.

3.1   Two-Sided Matching Markets

The extensive work on matching markets, dating back to the foundational works of David Gale and Lloyd Shapley (1962) and culminating in the classic monograph by Alvin Roth and Marilda Sotomayor (1992), has found numerous applications over the years. Analysis of match characteristics like stability along with computational algorithms that resulted in such outcomes have helped to augment policies and improve lives in various contexts ranging from School Choice (Abdulkadiroglu, Pathak, Roth, and Sonmez 2005) to kidney exchange (Roth, Sonmez, and Unver, 2004). While the application of matching models are broad, the implementation of such models to characterize inter-firm behavior is nascent.

In addition, although prior work has examined markets in which all participants or a subset are confined to one match (respective examples being men-women and worker-firms), I observe many cases in which this restriction does not apply. Examples such as upstream manufacturers and downstream distributors or the formation of business-to-business alliances for joint innovation permeate the industrial orga-

nization and strategy literatures. A many to many matching model is appropriate to analyze the features of these markets, which has often been overlooked in the literature to direct attention on contexts that have greater historical focus such as marriages (Becker, 1981), college admissions (Roth, 1985), and employment (Kelso and Crawford, 1982).

The matching literature has developed along two strands, with origins rooted in the marriage model (Gale and Shapley, 1962) and the assignment game (Shapley and Shubik, 1972). The principal distinction concerns whether agents can coordinate among themselves in the form of side-payments, with the implication that utility from a match is both divisible and transferable. Naturally, when side payments are permitted under a transferable utility (TU) matching game, I often see applications in markets that involve money such as auctions and the assignment of sellers to buyers. On the other hand, in the non-transferrable utility (NTU) case, more work has been focused on the preference characteristics of individuals over potential matches that have an exogenously fixed utility or payoff for each partner. Since I develop an NTU model with firms that make decisions based on the amount of value or money generated in a match, this framework allows me to illuminate the difference between a stable NTU outcome and a TU assignment.

For the TU assignment game, the allowance of side-payments means that the share of value each individual can capture will be endogenous to the matching process. The resulting outcome of matches maximizes the total value generated within the market. This has an intuitive interpretation since the optimal set of matches allows individual payments to be subdivided from the total in a way that makes all participants strictly better off than any sub-optimal assignment. Thus, an equilibrium will constitute a set of matches along with a vector of individual payoffs that is endogenously determined by the transfers among firms. Shapley and Shubik (1972) provides a precise definition of stability that is satisfied if the sum of individual pay-

offs for any potential pair of firms is no less than the total payoff generated from their match. If this were not the case, both firms can deviate from the market by forming a match and dividing the resulting total payoffs to be strictly better off.

In the NTU case where the division of payoffs within matches is exogenously fixed, the notion of a "stable" matching between two sets of economic agents has generally centered around a set of matches that satisfy individual rationality and pair-wise blocking. Within the simple example of a marriage market as envisioned by Gale and Shapley, marriages are stable when no married individual prefers to be single (i.e. all agents are individually rational) and there are no two individuals who would prefer the fixed utility each gains from being together over their current partners (i.e. current matches cannot be disrupted or blocked). I use this feature of stability to characterize an equilibrium in the model, which is defined as an outcome that is not blocked by coalitions of any size.

The fundamental difference between the NTU and TU outcomes concern whether the division of value within a match can be credibly adjusted given the presence of other firms (Sotomayor, 2009; Greve, Mitsuhashi and Baum, 2013). When the joint value generated within a match is not realized at the point of formation and the division of shares between partners committed to a given match are subjected to subsequent renegotiation, an NTU matching outcome becomes more relevant (Hart and Moore, 1988). Next, I formally set up the NTU matching model where the firm's payoffs correspond to a fixed share of the joint value that is generated from the match.

3.2   Formal Model

Let there be a set $\mathcal{U} = \{1, ..., u, ..., U\}$ of upstream firms, and set $\mathcal{D} = \{1, ..., d, ..., D\}$ of downstream firms.

A matching function $\alpha : \mathcal{U} \times \mathcal{D} \rightarrow \{0, 1\}$ characterizes the relationship between

both sets of firms. For each pair $(u, d) \in \mathcal{U} \times \mathcal{D}$, if both firms agree, an alliance is formed: $\alpha(u, d) = 1$

Each firm $u \in \mathcal{U}$ can form at most k^u alliances and each firm $d \in \mathcal{D}$ can form at most k^d alliances.

An **outcome** A is a subset of the Cartesian product $\mathcal{U} \times \mathcal{D}$ where $\alpha(u, d) = 1$:

$$A = \{(u, d) \in \mathcal{U} \times \mathcal{D} : \alpha(u, d) = 1\}$$

Let $\mathcal{A}^f$ denote the set of feasible outcomes. An outcome (the subset A or its function α) is **feasible** (i.e. $A \in \mathcal{A}^f$) if all capacity constraints are satisfied

$$\sum_{d \in \mathcal{D}} \alpha(u, d) \leqslant k^u, \quad \forall u \in \mathcal{U} \tag{3.1}$$

and

$$\sum_{u \in \mathcal{U}} \alpha(u, d) \leqslant k^d, \quad \forall d \in \mathcal{D} \tag{3.2}$$

If u and d unite, a surplus value $\pi_{u,d}$ is generated, from which u earns $\sigma_d^u \cdot \pi_{u,d}$ and d earns $\sigma_u^d \cdot \pi_{u,d} = (1 - \sigma_d^u) \cdot \pi_{u,d}$. The shares σ_d^u, σ_u^d are exogenously given and may reflect the bargaining power of each agent in an alliance.

The total payoff of any agent is a function of the matching outcome α defined as follows:

$$\Pi_u(\alpha) \equiv \sum_{d \in \mathcal{D}} \alpha(u, d) \cdot \sigma_d^u \cdot \pi_{u,d} \quad \forall u \in \mathcal{U}$$

and

$$\Pi_d(\alpha) \equiv \sum_{u \in \mathcal{U}} \alpha(u, d) \cdot \sigma_u^d \cdot \pi_{u,d} \quad \forall d \in \mathcal{D}$$

A standard assumption in the literature characterizes preferences as strict. For firms, this translates to distinct total payoffs from being allied with different subsets

of partners. This is a crucial assumption because indifference across sets of part-
ners would lead to ambiguous outcomes, giving rise to multiple potential matching
equilibria.

Assumption 1 (Strict Differences in Payoffs Over Partners)**.** *For any two outcomes*
α *and* α', *we have:*

- *for each* $u \in \mathcal{U}$ *if* $\{(u,d,\alpha(u,d)) : d \in \mathcal{D}\} \neq \{(u,d,\alpha'(u,d)) : d \in \mathcal{D}\}$ *then*
  $\Pi_u(\alpha) \neq \Pi_u(\alpha')$

- *for each* $d \in \mathcal{D}$ *if* $\{(u,d,\alpha(u,d)) : u \in \mathcal{U}\} \neq \{(u,d,\alpha'(u,d)) : u \in \mathcal{U}\}$ *then*
  $\Pi_d(\alpha) \neq \Pi_d(\alpha')$

3.2.1   Equilibrium

Definition (Equilibrium)**.** *We say that a feasible outcome* α^E *is an equilibrium if*
and only if:

1. *$\forall(u,d) \in \mathcal{U} \times \mathcal{D}$,   if $\alpha^E(u,d) = 1$ then $\pi_{u,d} > 0$;*

2. *there is no other feasible outcome α' such that $\exists(u,d) \in \mathcal{U} \times \mathcal{D}$ with $\alpha^E(u,d) = 0$*
   and $\alpha'(u,d) = 1$, and for all such (u,d):

   - $\Pi_u(\alpha') > \Pi_u(\alpha^E)$

   - $\Pi_d(\alpha') > \Pi_d(\alpha^E)$

The equilibrium outcome is characterized by two conditions. First, all matches
are positive-valued. Otherwise, each firm can do better on their own by unilaterally
dissolving the negative-valued match. Secondly, there does not exist a subset of firms
that can coordinate to form new matches among themselves, making each of them
better off in the process. If this subset existed, then there is nothing to prevent

its members from forming these alliances since such firms can unilaterally break off matches with non-members to pursue a mutually beneficial outcome amongst themselves. This is different from the socially efficient outcome, which seeks to maximize the total value generated in the market, as defined below.

Definition (Social Efficiency). *An outcome α^* is socially efficient if and only if it is a solution to the following linear program*

$$\max_{(u,d)\,\in\,\mathcal{U}\times\mathcal{D}} \sum \alpha(u,d)\,\pi_{u,d}$$

subject to: (3.1) and (3.2)

I start off by demonstrating a standard result in the literature (Gale and Shapley, 1962; Shapley and Schubik, 1972), acknowledging the difference between such outcomes with a simplified example.

Let $\mathcal{A}^E$ and $\mathcal{A}^*$ respectively denote the sets of all equilibrium and socially efficient outcomes.

Theorem 1 (Equilibrium Inefficiency). *The set of equilibrium outcomes $\mathcal{A}^E$ is not equivalent to the set of socially efficient outcomes $\mathcal{A}^*$.*

Proof. It suffices to provide a simplified example in a 1-to-1 matching market. Let $\sigma_d^u = \sigma_u^d = 1/2$, $\mathcal{U} = \{u_1, u_2\}$, $\mathcal{D} = \{d_1, d_2\}$ and the match values π be given below:

π	d_1	d_2
u_1	1	3
u_2	3	4

We observe that the social optimal outcome is $\{(u,d) \in \mathcal{U} \times \mathcal{D} : \alpha^*(u,d) = 1\} = \{(u_1, d_2), (u_2, d_1)\}$. However, the equilibrium outcome is given by $\{(u,d) \in \mathcal{U} \times \mathcal{D} : \alpha^E(u,d) = 1\} = \{(u_1, d_1), (u_2, d_2)\}$. $\square$

47

Note that the social optimal outcome is not an equilibrium since firms u_2 and d_2 can form a match between each other at the expense of their current partners to become mutually better off. Once this occurs, the remaining unmatched firms u_1 and d_1 can form a mutually beneficial match, resulting in the equilibrium matching outcome. I next propose a simplified way of characterizing the equilibrium given the model setup, which also serves as an intermediate result to an estimation criteria to be developed later. This result is also apparent in Echenique and Oviedo (2006), who show the equivalence of "setwise" and "pairwise" stability under a more general set of conditions on preferences. The result presented below is useful since it takes the form of surplus value functions, which appropriately characterizes a theory firm-to-firm matching.

Theorem 2 (Equilibrium Characterization). *Suppose that $\sigma_d^u = k \in (0, 1) \quad \forall (u, d) \in \mathcal{U} \times \mathcal{D}$.*

Then, an outcome α is an equilibrium, i.e. $\alpha \in \mathcal{A}^E$ if and only if

1. $\forall (u, d) \in \mathcal{U} \times \mathcal{D}$ *if* $\alpha(u, d) = 1$, *then* $\pi_{u,d} > 0$ *and*

2. $\forall (u, d) \in \mathcal{U} \times \mathcal{D}$ *such that* $\alpha(u, d) = 0$, *at least one of the following holds:*

    (a) $\pi(u, d) < 0$

    (b) $\pi(u, d) < \min_{d' \in \mathcal{D}} \{\pi(u, d') : \alpha(u, d') = 1\}$ *and* $\sum_{d' \in \mathcal{D}} \alpha(u, d') = k^u$

    (c) $\pi(u, d) < \min_{u' \in \mathcal{U}} \{\pi(u', d) : \alpha(u', d) = 1\}$ *and* $\sum_{u' \in \mathcal{U}} \alpha(u', d) = k^d$

Proof. We first show that if α is an equilibrium, condition 2 must hold. Suppose not. Then there exists an unmatched pair (u, d) such that:

1. $\pi(u, d) > 0$

2. $\alpha(u, d') = 1$ and $\pi(u, d') < \pi(u, d)$ for some $d' \in \mathcal{D}$ if $\sum_{d_j \in \mathcal{D}} \alpha(u, d_j) = k^u$

48

3. $\alpha(u', d) = 1$ and $\pi(u', d) < \pi(u, d)$ for some $u' \in \mathcal{U}$ if $\sum_{u_i \in \mathcal{U}} \alpha(u_i, d) = k^d$

We can construct another feasible matching α' such that:

1. $\alpha'(u, d') = 0$, $\alpha'(u', d) = 0$ and $\alpha(u, d) = 1$

2. $\alpha'(u_i, d_j) = \alpha(u_i, d_j)$ for all other pairs $(u_i, d_j) \in \mathcal{U} \times \mathcal{D}$

This violates the second equilibrium criteria. By contrapositive logic, an equilibrium outcome implies that both characterization conditions are true.

Now suppose α satisfies both characterization conditions and there exists an alternative matching α' that violates equilibrium criteria 2. Let (u^*, d^*) denote the new match under α' with the highest value:

$$\pi(u^*, d^*) = \max_{(u,d) \in \mathcal{U} \times \mathcal{D}} \{\pi(u, d) : \alpha(u, d) = 0, \, \alpha'(u, d) = 1\}$$

Case 1: Suppose $\pi(u^*, d^*) < 0$. Then $\pi(u_i, d_j) < 0 \; \forall (u_i, d_j) \in \mathcal{U} \times \mathcal{D}$ s.t. $\alpha(u_i, d_j) = 0$ and $\alpha'(u_i, d_j) = 1$. We have $\Pi_{u^*}(\alpha') > \Pi_{u^*}(\alpha)$ implying $\exists \, d' \in \mathcal{D}$ s.t. $\alpha(u^*, d') = 1$, $\alpha'(u^*, d') = 0$, and $\pi(u^*, d') < \pi(u^*, d)$. But this violates the first characterization condition.

Case 2: WLOG Suppose $\pi(u^*, d^*) < \min_{d_j \in \mathcal{D}}\{\pi(u^*, d_j) : \alpha(u^*, d_j) = 1\}$ and $\sum_{d_j \in \mathcal{D}} \alpha(u^*, d_j) = k^{u^*}$. Then $\pi(u^*, d') < \min_{d_j \in \mathcal{D}}\{\pi(u^*, d_j) : \alpha(u^*, d_j) = 1\} \; \forall \, d' \in \mathcal{D}$ s.t. $\alpha(u^*, d') = 0$ and $\alpha'(u^*, d') = 1$. Thus $\Pi_{u^*}(\alpha') > \Pi_{u^*}(\alpha)$ implies $\min_{d_j \in \mathcal{D}}\{\pi(u^*, d_j) : \alpha(u^*, d_j) = 1\} < 0$. This violates the first characterization condition. $\square$

The next result follows almost directly from Theorem 2. It can be used as an estimation criteria when observing a real world matching outcome. Note that Theorem 3 below is implied by stability of a matching model with a fixed sharing rule,

but that the reverse is not necessarily true. That's because the following condition would still hold under a more general classification of aligned preferences, which is shown for a many-to-one college admissions matching model (Sorensen, 2005). Here, the result also holds for a many-to-many matching market for firms.

Theorem 3 (Estimation Criterion). *In the Equilibrium Outcome α^E, given $u_1, u_2 \in \mathcal{U}$ and $d_1, d_2 \in \mathcal{D}$ such that*

1. *$\alpha^E(u_1, d_1) = 1$ and $\alpha^E(u_2, d_2) = 1$*

2. *$\alpha^E(u_1, d_2) = 0$ and $\alpha^E(u_2, d_1) = 0$*

We have the following result:

$$\max\{\pi(u_1, d_1),\ \pi(u_2, d_2)\} > \max\{\pi(u_1, d_2),\ \pi(u_2, d_1)\}$$

Proof. Suppose $\max\{\pi(u_1, d_1), \pi(u_2, d_2)\} < \max\{\pi(u_1, d_2), \pi(u_2, d_1)\}$. WLOG, let $\pi(u_1, d_2) > \pi(u_1, d_1), \pi(u_2, d_2)$. Equilibrium condition 1 implies that $\pi(u_1, d_2) > 0$. We also know that $\pi(u_1, d_2) > \min_{d_j \in \mathcal{D}}\{\pi(u_1, d_j) : \alpha^E(u_1, d_j) = 1\}$ and $\pi(u_1, d_2) > \min_{u_i \in \mathcal{U}}\{\pi(u_i, d_2) : \alpha^E(u_i, d_2) = 1\}$. The pair (u_1, d_2) violates the equilibrium characterization condition 2, leading to a contradiction of an equilibrium outcome α^E. □

I proceed to show that the above equilibrium exists in every matching market by constructing an algorithm that always terminates in this outcome. Furthermore, the equilibrium is unique under the strict differences in payoffs assumption.

3.2.2 Existence and Uniqueness

I now present a Greedy algorithm (GA) for finding the equilibrium outcome. Consider the following steps:

1. In the first step, locate an ordered pair $(u,d)_1 \in \mathcal{U} \times \mathcal{D}$ with the highest positive value i.e. $\pi((u,d)_1) = \max\{\pi(u,d) : \pi(u,d) > 0\}$ and construct the matching outcome A^1, represented by function α^1 with:

$$\alpha^1((u,d)_1) = 1 \text{ and } \alpha^1((u,d)) = 0 \ \forall(u,d) \neq (u,d)_1$$

2. At step k, with matching outcome $A^{k-1} = \{(u,d)_1, ..., (u,d)_{k-1}\}$, locate the next ordered pair $(u,d)_k \in \mathcal{U} \times \mathcal{D}$ with the highest positive value among remaining feasible matches:

$$\pi((u,d)_k) = \max\{\pi(u,d) : \pi(u,d) > 0, \ \alpha^{k-1}(u,d) = 0 \text{ such that}$$

$$\sum_{d_j \in \mathcal{D}} \alpha^{k-1}(u,d_j) < k^u, \ \sum_{u_i \in \mathcal{U}} \alpha^{k-1}(u_i,d) < k^d\}$$

Construct a new matching outcome A^k, represented by function α^k with:

- $\alpha^k(u,d) = 1$ for all $\alpha^{k-1}(u,d) = 1$

- $\alpha^k((u,d)_k) = 1$ and $\alpha^k((u,d)) = 0 \ \forall(u,d) \notin A^{k-1} \bigcup \{(u,d)_k\}$

Terminate when all firms on any given side are at capacity or all remaining feasible matches have negative value. Although the fixed-sharing rule of the NTU matching model is a special case of the "Matching with Contracts" model of Hatfield and Kominers (2017), who show the existence of an equilibrium, the following proof utilizes the Greedy algorithm to conveniently show this result.

Theorem 4 (Non-emptiness of Equilibrium). *The Greedy algorithm (GA) results in an equilibrium matching.*

Proof. Let K denote the last step of GA, resulting the matching outcome α^K.

From Theorem 2, we only need to check for cases that violate the two characterization criteria. Immediately, criterion 1 is satisfied; otherwise this would imply the existence of matches with negative value, which the algorithm does not assign.

Next, suppose the unmatched ordered pair (u, d) violates criterion 2. $\pi(u, d) > 0$ implies either $\sum_{d_j \in \mathcal{D}} \alpha^K (u, d_j) = k^u$ or $\sum_{u_i \in \mathcal{U}} \alpha^K (u_i, d) = k^d$, otherwise the algorithm would not have terminated.

1. If one of the firms is at capacity (WLOG let $\sum_{d_j \in \mathcal{D}} \alpha^K (u, d_j) = k^u$, $\sum_{u_i \in \mathcal{U}} \alpha^K (u_i, d) < k^d$), then there exists a firm $d' \in \{d_j \in \mathcal{D} : \alpha^K(u, d_j) = 1\}$ such that $\pi(u, d) > \pi(u, d')$. This contradicts the step in which (u, d') was matched instead of (u, d).

2. If both firms are at capacity ($\sum_{d_j \in \mathcal{D}} \alpha^K (u, d_j) = k^u$, $\sum_{u_i \in \mathcal{U}} \alpha^K (u_i, d) = k^d$), then there exists firms $d' \in \{d_j \in \mathcal{D} : \alpha^K(u, d_j) = 1\}$ and $u' \in \{u_i \in \mathcal{U} : \alpha^K(u_i, d) = 1\}$ such that $\pi(u, d) > \pi(u, d')$ and $\pi(u, d) > \pi(u', d)$. Then the earlier-matched/higher-valued of the two pairs (u, d') and (u', d) generates a contradiction when compared with (u, d).

Thus $\alpha^K = \alpha^E$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Having established the existence of an equilibrium, I can now use the Greedy algorithm to prove a novel uniqueness result for the many-to-many matching market. This will be followed by some new results on the Pareto efficient properties of the equilibrium outcome.

Theorem 5 (Uniqueness of Equilibrium). *The matching equilibrium α^E is unique.*

Proof. We establish this result by showing that all the matches resulting from GA are required to prevent the existence of an unmatched pair that contradicts the second criterion of theorem 2. The argument will follow by induction. Assume for simplicity that all ordered pairs considered have positive value. In step 1, the first ordered pair $(u, d)_1$ is matched in the equilibrium A^E, otherwise $(u, d)_1$ clearly contradicts criterion 2. At step k, the same argument applies to the ordered pair $(u, d)_k$ since, given A^{k-1}:

- $\sum\limits_{d_j \in \mathcal{D}} \alpha^{k-1}(u, d_j) < k^u$

- $\sum\limits_{u_i \in \mathcal{U}} \alpha^{k-1}(u_i, d) < k^d$

- $\pi((u, d)_k) = \max\{\pi(u, d) : \pi(u, d) > 0, \alpha^{k-1}(u, d) = 0, \sum\limits_{d_j \in \mathcal{D}} \alpha^{k-1}(u, d_j) < k^u, \sum\limits_{u_i \in \mathcal{U}} \alpha^{k-1}(u_i, d) < k^d\}$

Given all matches that have been formed under GA (which by construction of the algorithm, contains only positive matches), no other matches can be included in A^E since remaining feasible matches are negative (otherwise the GA would not have terminated), which would violate the first criterion. □

3.2.3 Pareto Efficiency

Having shown the existence and uniqueness of the equilibrium, I prove some of its desirable properties. Although the decentralized formation process results in an outcome that is not necessarily socially efficient (i.e. Theorem 1), some Pareto efficiency properties will hold. These results are novel in two ways. First, it analyzes the properties of a unique outcome under the fixed sharing rule within a many-to-many matching market. Secondly, the Pareto efficiency results are in the form of "strong" Pareto efficiency, which says that there is no alternative "feasible" outcome

in which firms are no worse off and one is strictly better off. This contrasts with "weak" Pareto efficiency that only requires no feasible outcome such that all firms are strictly better off and has been shown in prior work on many-to-one matching models (Roth and Sotomayor, 1992).

Definition (Pareto Efficiency)**.** *We say that a matching outcome α is Pareto efficient for a set of firms $\mathcal{F} \subseteq \mathcal{U} \cup \mathcal{D}$ if for any alternative feasible outcome α':*

$$\Pi_f(\alpha) > \Pi_f(\alpha') \text{ for some } f \in \mathcal{F}$$

Suppose all downstream firms sign exclusive distribution rights with upstream producers, limiting each upstream firm to at most one match. i.e. $k^u = 1 \ \forall u \in \mathcal{U}$ In this special case of a many to one matching market, it can also be shown that the equilibrium is Pareto Efficient for the side of single-match constrained firms ($\mathcal{U}$).

Theorem 6 (Pareto Efficiency for Side of Single Partner Firms)**.** *When $k^u = 1 \ \forall u \in \mathcal{U}$, the matching equilibrium α^E is Pareto Efficient for all upstream firms $\mathcal{U}$. That is, there is no alternative feasible matching outcome α' such that $\Pi_u(\alpha') \geqslant \Pi_u(\alpha^E) \ \forall u \in \mathcal{U}$.*

Proof. We show that any alternative feasible matching outcome α' results in at least one firm $u \in \mathcal{U}$ being worse off.

1. If $A^{\alpha'} \supset A^{\alpha^E}$, then $A^{\alpha'}$ consists of at least one ordered pair with negative value, thus making at least one previously unmatched firm in $\mathcal{U}$ worse off.

2. If $A^{\alpha'} \not\supseteq A^{\alpha^E}$, let (u^*, d^*) denote the highest valued match in A^{α^E} that is not in $A^{\alpha'}$:

$$\pi(u^*, d^*) = \max\{\pi(u, d) : (u, d) \in \mathcal{U} \times \mathcal{D}, \ \alpha^E(u, d) = 1, \ \alpha'(u, d) = 0\}$$

Then u^* is necessarily made worse off since at the time of selection under GA, d^* was its best match given all prior GA matches having been formed.

□

The above result cannot be extended to the side of downstream firms. To demonstrate this, consider a simplified example consisting of 4 upstream firms $\mathcal{U} = \{A, B, C, D\}$ and 2 downstream firms $\mathcal{D} = \{X, Y\}$ with $k^u = 1 \ \forall u \in \mathcal{U}$ and $k^d = 2 \ \forall d \in \mathcal{D}$. Let the surplus $\pi_{u,d}$ of each match be given by the table below:

$\pi_{u,d}$	A	B	C	D
X	13	1	9	6
Y	12	7	10	8

Comparing the equilibrium outcome α^E and an alternative feasible outcome α' characterized by:

$$A^{\alpha^E} = \{(u, d) \in \mathcal{U} \times \mathcal{D} : \alpha^E(u, d) = 1\} = \{(A, X), (B, X), (C, Y), (D, Y)\}$$

$$A^{\alpha'} = \{(u, d) \in \mathcal{U} \times \mathcal{D} : \alpha'(u, d) = 1\} = \{(C, X), (D, X), (A, Y), (B, Y)\}$$

We can see that the equilibrium outcome does not result in a Pareto Efficient match for the set of downstream firms since both X and Y are strictly better off under the match outcome α'.

$$\Pi_X(\alpha') = \sigma_d(9 + 6) = 15\sigma_d > 14\sigma_d = \sigma_d(13 + 1) = \Pi_X(\alpha^E)$$

$$\Pi_Y(\alpha') = \sigma_d(12 + 7) = 19\sigma_d > 18\sigma_d = \sigma_d(10 + 8) = \Pi_Y(\alpha^E)$$

The natural question then arises regarding whether the equilibrium outcome is Pareto Efficient for all firms in the market. By implication of Theorem 6, this is true in the special case of a many to one matching market. To show that an equilibrium outcome is Pareto Efficient for the general case of a many to many matching market, I first prove that the equilibrium outcome implies a desirable property.

Given a matching outcome α, a subset of firms $S = S_u \cup S_d$ where $S_u \subseteq \mathcal{U}$ and $S_d \subseteq \mathcal{D}$ forms a **dominating coalition** if there exists a feasible outcome α' such that:

1. $\Pi_i(\alpha') \geqslant \Pi_i(\alpha) \ \forall i \in S$

2. $\exists j \in S$ such that $\Pi_j(\alpha') > \Pi_j(\alpha)$

3. $\{(u,d) \in \mathcal{U} \times \mathcal{D} : \alpha'(u,d) = 1, u \in S_u, d \notin S_d\} \subseteq \{(u,d) \in \mathcal{U} \times \mathcal{D} : \alpha(u,d) = 1, u \in S_u, d \notin S_d\}$

4. $\{(u,d) \in \mathcal{U} \times \mathcal{D} : \alpha'(u,d) = 1, u \notin S_u, d \in S_d\} \subseteq \{(u,d) \in \mathcal{U} \times \mathcal{D} : \alpha(u,d) = 1, u \notin S_u, d \in S_d\}$

According to the above definition, a dominating coalition exists for a given matching outcome α if there is a subset of firms S along with an alternative feasible matching α' such that everyone in the subset is weakly better off with at least one being strictly better off without forming new matches to firms outside the subset. This implies that coalition members can coordinate alternative matches among themselves to create a better aggregate group outcome while holding the decisions of non-members fixed.

Theorem 7 (Equilibrium is in the CORE). *The matching equilibrium α^E does not have dominating coalitions of any size.*

Proof. We will prove this result by contradiction. Suppose the subset of firms S forms a dominating coalition and let α' denote the corresponding match outcome. Then, by assumption of strict differences in payoffs:

1. If $u \in S_u$ and $\Pi_u(\alpha') = \Pi_u(\alpha^E)$, then $\alpha'(u,d) = \alpha^E(u,d) \ \forall d \in \mathcal{D}$.

2. If $d \in S_d$ and $\Pi_d(\alpha') = \Pi_d(\alpha^E)$, then $\alpha'(u,d) = \alpha^E(u,d) \ \forall u \in \mathcal{U}$.

Since all coalition members that are not strictly better off must have the same matches, there exists another dominating coalition $S' \subseteq S$ under α' such that $\forall i \in S'$, $\Pi_i(\alpha') > \Pi_i(\alpha^E)$. Moreover, by criterion 1 of the equilibrium definition, if $i \in S'$, then it must have formed a new match since a proper subset of matches from the equilibrium α^E would imply $\Pi_i(\alpha') < \Pi_i(\alpha^E)$. Now construct the feasible matching α^c where:

$$\alpha^c(u, d) = \begin{cases} \alpha'(u, d), & \text{if } u \in S' \text{ or } d \in S' \\ \alpha^E(u, d), & \text{if } u \notin S' \text{ and } d \notin S' \end{cases}$$

Since α^c does not affect the composition of matches for members of S' under α' while resulting in only coalition members forming new matches different from α^E, it violates the second equilibrium criterion, creating a contradiction that α^E is the equilibrium. $\square$

From theorem 7, the Pareto Efficiency result for all firms in the matching market directly follows.

Corollary 1 (Pareto Efficiency in Many-to-Many Matching Market)**.** *The matching equilibrium α^E is Pareto Efficient for all firms. That is, there is no alternative feasible matching outcome α' such that $\Pi_i(\alpha') \geqslant \Pi_i(\alpha^E) \ \forall i \in \mathcal{U} \cup \mathcal{D}$.*

Proof. By theorem 7, there is no dominating coalition under the equilibrium α^E for the subset of firms $S = \mathcal{U} \cup \mathcal{D}$. $\square$

Next, I use this theoretical framework to show how features of the firms' choice characteristics can be extracted from the observation of match outcomes. In particular, I will focus on the case where an observation consists of two matched pairs and two unmatched counter-factual pairs among 4 firms. Theorem 3 gives the implications of such observations within a matching market.

3.3 Discussion

Although the assumption of a fixed sharing rule is quite strong, it is necessary to ensure the existence of a unique equilibrium that is obtained by the Greedy algorithm. If this assumption were relaxed to permit an unknown firm specific bargaining power parameter λ_x for firm x with the share of the surplus to be divided between firms x and y in a match given by $\frac{\lambda_x}{\lambda_x + \lambda_y}$ for firm x and $\frac{\lambda_y}{\lambda_x + \lambda_y}$ for firm y, then the Greedy algorithm no longer guarantees an equilibrium outcome. Going back to the example from theorem 1 (reproduced below), we can observe that if all firms from the same side of the market have the same parameter value $(\lambda_{u_1} = \lambda_{u_2})$ and $(\lambda_{d_1} = \lambda_{d_2})$, the fixed sharing rule holds and the Greedy algorithm generates the equilibrium outcome $A^1 = \{(u_1, d_1), (u_2, d_2)\}$.

However, suppose the unobserved firm-specific parameters were given by $\lambda_{u_1} = \lambda_{d_1} = 1$ and $\lambda_{u_2} = \lambda_{d_2} = 3$, then a match between (u_1, d_2) would result in individual payoffs of $\Pi_{u_1} = \frac{1}{4} * 3 = 0.75$ and $\Pi_{d_2} = \frac{3}{4} * 3 = 2.25$ (symmetrically for the match between u_2 and d_1) while the matches from a Greedy algorithm would still result in an even split of the surplus. Thus, the new equilibrium $A^2 = \{(u_1, d_2), (u_2, d_1)\}$ would be different from the algorithm since firms u_1 and d_2 would form a blocking pair under the original set of matches A^1. Consequently, uniqueness of the equilibrium is not assured by the weaker assumption.

The algorithms that generate unique equilibrium outcomes are interesting since they may suggest the generation of firm heterogeneity through a matching process.

π	d_1	d_2
u_1	1	3
u_2	3	4

FIGURE 3.1: Example of How Firm-Specific Bargaining Parameters Can Affect the Equilibrium Outcome.

The Greedy algorithm guarantees that the pair of firms that generate the highest value will always be formed, but does not ensure that the lowest value pair is matched in the end. In the market for mergers and acquisitions, algorithms that generate buyer optimal and seller optimal prices are also available (Sotomayor, 2009). The prices reflect the values that firms can capture within such markets. Examining how the conditions and contexts of firm interaction may influence the variation in firm performance outcomes seems to be a promising area for future analysis.

Lastly, the given example (reproduced below) in the analysis of Pareto optimality presents an interesting case for a many-to-one market. In this 2-to-1 matching market, upstream producers $\{A, B, C, D\}$ are each seeking one downstream partner. Meanwhile, the distributors $\{X, Y\}$ each have capacity to contract with 2 upstream firms.

According to the example, the unique equilibrium outcome is

$$A^E = \{(A, X), (B, X), (C, Y), (D, Y)\}$$

While the equilibrium is Pareto efficient for the side of single partner firms $\{A, B, C, D\}$, is not Pareto optimal for the side of multi-partner firms $\{X, Y\}$ since the outcome

$$A' = \{(C, X), (D, X), (A, Y), (B, Y)\}$$

constitutes a Pareto improvement for multi-partner firms. This suggests that if multi-partner firms can coordinate, then they can make each other strictly better off by trading partners. Note however, that in the absence of such coordination, A' is not

$\pi_{u,d}$	A	B	C	D
X	13	1	9	6
Y	12	7	10	8

FIGURE 3.2: The Richness of Potential Match Outcomes Under Various Conditions

stable since the pair (A, X) would partner up while breaking up their existing ties (i.e. they would block the formation of (A, Y) and (D, X)).

A better motivated example may concern the one-to-many matching of managers to internal divisions within a firm. Suppose managers are compensated based on a fixed share of the division's performance. Then the example would suggest that the decentralized matching of managers to divisions can be different than the case when division heads can coordinate among themselves. At the same time, the example also shows that the Pareto improved outcome A' is not the socially optimal (TU) outcome given by $A^* = \{(A, X), (C, X), (B, Y), (D, Y)\}$, implying that the central planner or CEO could allocate managers differently than the resulting matches from a coordination among the heads of division. Note that among the three matchings, A^E has the lowest total value, but is not the worst outcome possible, suggesting that a Greedy algorithm matching managers to divisions has some value maximizing properties but is far from collectively optimal. This is because each manager is strictly looking to increase its own value without regards for other matches. This contrasts starkly with the TU outcome A^* where the impact of each match on the possible total values of all remaining matches is taken into account. In the intermediate case A', the total value is greater than the equilibrium A^E and less than the TU outcome A^* because divisions are concerned with the combined efforts of its own managers, but disregard the matching of managers to other divisions (a case of semi-coordination). Altogether, these possibilities suggest a high potential for the utilization of matching models in future research on organizational and inter-firm behavior.

4

Estimating Matching Games Without Transfers

An important issue for scholars seeking to understand firm interactions is how to test factors that drive partner selection decisions from observations on relationship data. There are many situations when researchers can observe the alliances within a market but have no information on a given firm's choice set of potential partners outside the observed alliances. This is problematic for past empirical work that have relied on traditional binary choice methods (Mowery, Oxley and Silverman, 1998; Rothaermel and Boeker, 2008) since an unmatched pair of firms cannot be inferred to have negative match surplus value if one of the firms is simply precluded from forming additional alliances due to its commitment to other partners. Since the alliance capacities of firms are unobserved, I cannot discount this possibility.

Acknowledging the limitations of binary choice, emergent work has begun to adopt a matching model approach to various contexts of firm interaction such as strategic alliances and venture capital markets (Sorensen, 2007; Mindruta et al, 2016). Although the matching theory literature is rich and extensive, the body of work on the econometrics of matching is recent and underdeveloped by comparison. Given the relatively high level of attention directed toward transferable utility (TU)

matching estimation, the scarcity of formal matching methodology is a greater issue for nontransferable utility (NTU) matching models. Since the market for inter-firm collaboration conforms closely to an NTU matching framework, the formal development of an estimation method is needed.

I propose a semiparametric NTU matching score estimator. The issues with multiplicity of stable matchings is resolved by confining the econometric application to a special set of NTU models with a unique equilibrium solution. Specifically, I assume that the division of match surplus is given by a fixed sharing rule, which is applicable to cases where the payoffs cannot be realized at the time of a match and is subjected to renegotiation in the subsequent period without the possibility of re-matching for current partners. The fixed sharing rule restriction also ensures that preferences for firms on both sides of the market are aligned with the joint surplus generated from a match. This allows the estimator to focus solely on the surplus value of potential matches with the payoffs of agents over potential matches accordingly defined by their fixed share.

The matching model-based estimation approach contributes to the strategy literature on alliances by providing a tool to test factors underlying alliance value creation. This bridges the gap between the existing empirical literature that require an appropriate methodology to understand the formation of inter-firm collaboration ties and the relevant theoretical foundations of an NTU matching model. The class of NTU matching models along with the associated estimator may also be broadly applicable other conceivable contexts characterized by incomplete contracts between matches and future renegotiation of value division given the changing nature of outside options.

This paper also contributes to the methodological literature on matching by providing a score estimator for a class of NTU models. This has two benefits over existing approaches to matching estimation. First, the semi-parameterization allows

a relaxation of strong distributional assumptions on the error terms of the match surplus function (Choo and Siow, 2006). Next, the computationally simple calculation of scores based on a set of inequalities (as noted by Fox, 2017) is straightforward and a relatively easy way of obtaining consistent estimation for applied researchers. This neither requires numerically integrating out un-observables to calculate choice probabilities (Galichon and Salanie, 2015) nor the use of more complicated techniques such as a nested fixed-point procedure to compute the equilibrium to a game (Ciliberto and Tamer, 2009).

The next section will provide a background on matching estimation before proceeding to a formal characterization of the econometric model. The model section starts off by providing some primitive assumptions on the functional form of the joint surplus function along with the nature of the parameters and observable variables. An intermediate result on the identification of the conditional probability for the observed outcome variable is used to establish the novel theorems on consistency of the score estimator. A brief discussion section concludes the chapter.

4.1 Econometrics of Matching

Utilizing the theoretical matching model, applied researchers can study the interactions of individuals, firms, and institutions. The theoretical foundation analyzes how outcomes emerge based on the set of agents and their underlying payoffs from potential matches in a given market. Since the econometric literature on matching aims to infer the determinants of individual payoffs from observed matching patterns, the theoretical insights provide the necessary restrictions to enable identification of underlying parameters. The nature of such restrictions will depend on the type of matching model under consideration, which should reflect the context of the empirical study. For example, a transferable utility (TU) matching model may be appropriate for matching buyers and sellers whereas a non-transferable utility

(NTU) matching framework may better capture the matching process in the market for inter-firm alliances.

Unlike the TU model which generally produces a unique equilibrium under innocuous conditions, estimation for the NTU model poses several challenges. First, the presence of multiple equilibria even in the simple case of one-to-one matching market (c.f. Irving and Leather, 1986) complicates econometric applications as statistical likelihood functions are not well defined for such games (Bresnahan and Reiss, 1990). Secondly, the presence of an unknown exogenous split in the division of surplus between firms in any potential match creates two distinct payoff functions for each side of the market. Menzel (2015) takes the large market limit for the one-to-one case of marriages and shows that data on observed matches alone is not enough to separate the preferences of both sides of the market.

Prior empirical work based on NTU models have attempted to circumvent these obstacles by either utilizing data beyond the observation of matches or imposing more model restrictions that confine the set of equilibria. In the first case, Hitsch, Hortacsu, and Ariely (2010) utilize data on online dating sites that allow them to observe the choice sets of individuals based on potential partners they've contacted. This reduces the estimation of individual preferences to a fixed effects logit. Del Boca and Flinn (2014) utilizes observations on household behavior to recover the preferences of spouses given some fixed exogenous rule on household allocations between partners. Using a likelihood-based metric to compare the preference-generated stable matches and observed matches, they recover the rule that optimally fit the data. These methods are different from the semi-parametric approach proposed in this paper since the use of a fixed-effects logit or a likelihood-based metric both impose strong distributional assumptions on the unobservables.

For the second case, Boyd et al. (2013) assume the equilibrium results from the Gale-Shapley deferred acceptance algorithm with firms proposing to workers

while Gordon and Knight (2009) impose symmetry assumptions between both sides of the market that guarantee the existence of a unique equilibrium. Both proceed to use moment-based simulation approaches, which can become computationally demanding as the set of agents becomes large. Other added assumptions such as an agreed-upon vertical differentiation of agents on one side (e.g. hospitals can agree on the ranking of residents) or restrictions concerning the structure/distribution of error terms have also been imposed to make estimation more tractable (Sorensen, 2007; Hsieh, 2012; Agarwal, 2015). The extensive surveys by Graham (2011) and Chiappori and Salanie (2016) provide additional examples of the empirical NTU matching literature. The methodology proposed in the next section relies on a feature of the underlying matching model (i.e. a fixed sharing rule for value division between partners) to ensure the existence of a unique equilibrium and proceeds to estimate the parameters of interest from a set of computationally simple inequalities implied by the equilibrium property.

4.2 Estimation

I now build on the NTU model with fixed surplus sharing from the previous section by developing a way to estimate the determinants of payoffs. The assumed fixed share division of the surplus permits the characterization of the net payoffs of each firm from a match by only considering the joint surplus. To adapt the matching model for econometric analysis, the joint surplus of a given pair (u, d) is reformulated as a random variable with an associated measure $f(\cdot)$:

$$\pi_{u,d} \sim f(\pi | \boldsymbol{x})$$

Where x is the set of observable characteristics that may include pairwise and individual variables:

$$\boldsymbol{x} = (\boldsymbol{x}_{u,d}, \boldsymbol{x}_u, \boldsymbol{x}_d)$$

In a special case, the joint surplus can be decomposed into two components:

$$\pi_{u,d} = s(\boldsymbol{x}, \beta) + \epsilon_{u,d}$$

The first term, $s(\boldsymbol{x}, \beta)$, is a deterministic component with $s(\cdot)$ being a function of observable characteristics $\boldsymbol{x}$, and a set of parameters β. The stochastic component $\epsilon_{u,d}$ is observed by the firm, but unobserved by the researcher. Although prior works have specified the distribution for the unobserved term before estimating the parameters, this is a strong assumption. The semi-parametric approach provides an alternative by restricting the relationship between observable characteristics and the conditional distribution of the joint surplus $f(\cdot|\boldsymbol{x})$. This will allow a relaxation on the distributional specification of unobservable factors, in particular $\epsilon_{u,d}$ for the additively separable case above. I will follow this approach and outline a set of parsimonious assumptions to enable identification of β from the observed pattern of matches.

Additionally, I assume that the structure of the data encompasses multiple markets. Specifically, given N markets, let A_i denote the matching outcome of market $i \in \{1, 2, , N\}$. I impose the surplus function and parameters to remain the same in all markets along with the exclusion restriction that firms do not observe one another across markets but observe everyone within their own market. In practice, the multiple market perspective accommodates the common data format that consists of matching markets across geographies or time. The convergence results from the multiple market perspective will also apply in a straightforward way to one large cross-sectional market.

On a technical note, the multiple market framework reconciles an issue that arises

when an observable characteristic $x_1 \in \boldsymbol{x}$ can take on an unbounded set of potential values. Since the origins of matching econometrics are rooted in single-agent discrete choice models, choosing among a finite set of partners implies that characteristics influencing the match outcome should take on a finite set of values. Thus, if the set of meaningful potential values are unbounded, then the single market becomes problematic as an agent's utility would approach infinity with greater choices of partners. To overcome this limitation, I allow a distribution of matches over the continuum of observable values $\boldsymbol{x}$ to be characterized by the measure $\eta(\boldsymbol{x})$. Each market is then a set of matched firm-pair draws from the distribution, following the exclusion restriction. Thus, the measure characterizes the population of matches given observables $\boldsymbol{x}$ from which the researcher draws observations when sampling (Graham, 2011 - discrete case; Dupuy and Galichon, 2014 - a continuous case).

4.2.1 Identification

I now develop a formal econometric framework from a set of primitive features underlying the data generation process. The following assumption specifies a functional form of the observable variables and relates it to the conditional distribution of the joint surplus.

Assumption 2 (Functional Specification). *There exists a set of K-element parameter vectors $\boldsymbol{\beta} \subset \mathbb{R}^K$ such that for a vector $\beta \in \boldsymbol{\beta}$, if $\boldsymbol{x}_1\beta > \boldsymbol{x}_2\beta$, then*

$$Pr(\pi > r|\boldsymbol{x}_1) > Pr(\pi > r|\boldsymbol{x}_2) \text{ for all } r \in \mathbb{R}$$

Assumption 2 states that there exists a linear combination of the observable variables such that the distribution of the joint surplus conditional on a value of the linear combination will first-order stochastically dominate all other conditional distributions of less value. Common frameworks that satisfy this assumption include the linear regression model as well as discrete choice-based latent variable models.

This assumption leads to some intermediate results that help lay the foundation for key theorems on identification and estimation.

Lemma 1. *Under Assumption 2,*

1. *If $\beta \in \boldsymbol{\beta}$, then $a\beta \in \boldsymbol{\beta}$ for all $a > 0$*

2. *Given $\beta \in \boldsymbol{\beta}$ and $\boldsymbol{x}_1, \boldsymbol{x}_2$ that respectively correspond to the random variables π_1, π_2, $Pr(max\{\pi_1, \pi_2\} > r)$ is increasing in $max\{\boldsymbol{x}_1\beta, \boldsymbol{x}_2\beta\}$ for all $r \in \mathbb{R}$*

Proof.

1. $\boldsymbol{x}_1\beta > \boldsymbol{x}_2\beta$ if and only if $\boldsymbol{x}_1(a\beta) = a(\boldsymbol{x}_1\beta) > a(\boldsymbol{x}_2\beta) = \boldsymbol{x}_2(a\beta)$

2. WLOG, let $\boldsymbol{x}_1\beta = max\{\boldsymbol{x}_1\beta, \boldsymbol{x}_2\beta\}$. Then for $r \in \mathbb{R}$,

$$Pr(max\{\pi_1, \pi_2\} > r|\boldsymbol{x}_1\beta)$$
$$= E[Pr(\pi_1 > r|\boldsymbol{x}_1\beta) + (1 - Pr(\pi_1 > r|\boldsymbol{x}_1\beta))(Pr(\pi_2 > r|\boldsymbol{x}_2\beta < \boldsymbol{x}_1\beta))]$$
$$= Pr(\pi_1 > r|\boldsymbol{x}_1\beta) + (1 - Pr(\pi_1 > r|\boldsymbol{x}_1\beta))E[(Pr(\pi_2 > r|\boldsymbol{x}_2\beta < \boldsymbol{x}_1\beta))]$$

As $\boldsymbol{x}_1\beta = max\{\boldsymbol{x}_1\beta, \boldsymbol{x}_2\beta\}$ increases, so does $Pr(\pi_1 > r|\boldsymbol{x}_1\beta)$ and $E[(Pr(\pi_2 > r|\boldsymbol{x}_2\beta < \boldsymbol{x}_1\beta))]$. The result follows naturally. Note that the expectation is taken only over $(Pr(\pi_2 > r$ since $max\{\boldsymbol{x}_1\beta, \boldsymbol{x}_2\beta\}$ is used, but not $min\{\boldsymbol{x}_1\beta, \boldsymbol{x}_2\beta\}$.

□

The first part of Lemma 1 suggests that a scale-invariant normalization is needed for point identification. I use the normalization proposed in Horowitz(1992). For any vector in the set $\{b \in \mathbb{R}^K : b_1 \neq 0\}$, define the function $g : \mathbb{R}^K \to \mathbb{R}^K$ as follows:

$$g(b) = \frac{b}{|b_1|}$$

It's clear that $g(b)$ maps to a unique vector for each corresponding positively-scaled set of vectors. For convenience, let $\beta^* = g(\beta)$.

Using the result of part 2, I can simplify the notation. Let $Z_{12} = max\{\pi_1, \pi_2\}$ and $\boldsymbol{x}_{12|b} = \underset{\boldsymbol{x} \in \{\boldsymbol{x}_1, \boldsymbol{x}_2\}}{\arg\max}\{\boldsymbol{x}b\}$ with $\boldsymbol{x}_{12}^* = \boldsymbol{x}_{12|\beta}$ and the associated measure:

$$Z_{12} \sim h(z|\boldsymbol{x}_{12}^*\beta^*)$$

The researcher observes either the event $Z_{12} > Z_{34}$ or $Z_{12} < Z_{34}$.[1] Denote an indicator function Z^* defined by:

$$Z^* = \begin{cases} 1, & \text{if } Z_{12} > Z_{34} \\ -1, & \text{if } Z_{12} < Z_{34} \end{cases}$$

The assumption on strict differences in payoffs from the theoretical matching model in the preceding section does not include the possibility that $Z_{12} = Z_{34}$. In an econometric model, I need to make additional assumptions on the data generating process to justify the zero probability of this event. Specifically, assumptions are needed on the distributions of the observable variables from which observations can be drawn. More importantly, these assumptions are used for identification of the parameters β^*.

Definition (Identification). *Given the joint distribution $\boldsymbol{H}(X)$ over the set of observables $X = (\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4)$, the set of parameters β is identified up to a positive scale if for any $b \in \{\mathbb{R}^K : b_1 \neq 0\}$ where $g(b) \neq g(\beta)$, there exists a set of values on the observables $\boldsymbol{X}^b$ such that:*

1. $\int_{\boldsymbol{X}^b} d\boldsymbol{H}(X) > 0$ and

2. $Pr(Z^* = 1|\boldsymbol{x}_{12|b}, \boldsymbol{x}_{34|b}; g(b)) \neq Pr(Z^* = 1|\boldsymbol{x}_{12}^*, \boldsymbol{x}_{34}^*; \beta^*)$ for all $X \in \boldsymbol{X}^b$

The definition above states that the unknown parameters β^* can be identified from the data if other candidate values $g(b)$ would result in different distributions

---

[1] The relationship between Z and the observation set of matches will be made clear in the next section. Here, I continue to focus on the observation unit consisting of 4 potential pairs, two of which correspond to observed matches.

of the observed event Z^* conditional on some observations X. For the definition to be meaningful, quantitative characterizations of the conditional probability for the indicator variable Z^* under β^* are necessary The following result establishes this characterization.

Lemma 2. *Under Assumption 2, when the distributions of Z_{12} and Z_{34} are independent:*

1. *$Pr(Z^* = 1 | \boldsymbol{x}_{12}^*, \boldsymbol{x}_{34}^*; \beta^*) = 0.5$ and $E[Z^* | \boldsymbol{x}_{12}^*, \boldsymbol{x}_{34}^*] = 0$ for $\boldsymbol{x}_{12}^* \beta^* = \boldsymbol{x}_{34}^* \beta^*$*

2. *$Pr(Z^* = 1 | \boldsymbol{x}_{12}^*, \boldsymbol{x}_{34}^*; \beta^*) \gtreqless 0.5$ if and only if $\boldsymbol{x}_{12}^* \beta^* \gtreqless \boldsymbol{x}_{34}^* \beta^*$*

3. *$E[Z^* | \boldsymbol{x}_{12}^*, \boldsymbol{x}_{34}^*] \gtreqless 0$ if and only if $(\boldsymbol{x}_{12}^* - \boldsymbol{x}_{34}^*)\beta^* = \boldsymbol{x}^* \beta^* \gtreqless 0$*

Proof. Under independence, the joint density of (Z_{12}, Z_{34}) is equivalent to the product of each conditional densities:

$$\mathbf{h}(z_{12}, z_{34} | \boldsymbol{x}_{12}^* \beta^*, \boldsymbol{x}_{34}^* \beta^*) = h(z_{12} | \boldsymbol{x}_{12}^* \beta^*) h(z_{34} | \boldsymbol{x}_{34}^* \beta^*)$$

1. For $\boldsymbol{x}_{12}^* \beta^* = \boldsymbol{x}_{34}^* \beta^*$, we have $h(z_{12} | \boldsymbol{x}_{12}^* \beta^*) = h(z_{34} | \boldsymbol{x}_{34}^* \beta^*)$. Let $h^*(z)$ denote this special case. The resulting probabilities are:

$$Pr(Z^* = 1 | \boldsymbol{x}_{12}^* \beta^*, \boldsymbol{x}_{34}^* \beta^*) = \int_{-\infty}^{\infty} \int_{Z_{34}}^{\infty} h^*(z_{12}) h^*(z_{34}) dz_{12} dz_{34}$$

$$= \int_{-\infty}^{\infty} h^*(z_{34}) \int_{Z_{34}}^{\infty} h^*(z_{12}) dz_{12} dz_{34} = \int_{-\infty}^{\infty} h^*(z_{12}) \int_{Z_{12}}^{\infty} h^*(z_{34}) dz_{34} dz_{12}$$

$$= \int_{-\infty}^{\infty} \int_{Z_{12}}^{\infty} h^*(z_{12}) h^*(z_{34}) dz_{34} dz_{12} = Pr(Z^* = -1 | \boldsymbol{x}_{12}^* \beta^*, \boldsymbol{x}_{34}^* \beta^*)$$

This leads to $Pr(Z^* = 1 | \mathbf{x}_{12}^*, \mathbf{x}_{34}^*; \beta^*) = 0.5$. $E[Z^* | \mathbf{x}_{12}^*, \mathbf{x}_{34}^*] = 0$ follows directly.

2. For the second result, observe that:

$$Pr(Z^* = 1 | \boldsymbol{x}_{12}^* \beta^* > \boldsymbol{x}_{34}^* \beta^*) = \int_{-\infty}^{\infty} \int_{Z_{34}}^{\infty} h(z_{12} | \boldsymbol{x}_{12}^* \beta^*) h(z_{34} | \boldsymbol{x}_{34}^* \beta^*) dz_{12} dz_{34}$$

$$= \int_{-\infty}^{\infty} h(z_{34} | \boldsymbol{x}_{34}^* \beta^*) \int_{Z_{34}}^{\infty} h(z_{12} | \boldsymbol{x}_{12}^* \beta^*) dz_{12} dz_{34}$$

70

By Assumption 2, $\int_{Z_{34}}^{\infty} h(z_{12}|\boldsymbol{x}_{12}^*\beta^*)dz_{12} > \int_{Z_{34}}^{\infty} h^*(z_{12})dz_{12}$ at each fixed value Z_{34} since $\boldsymbol{x}_{12}^*\beta^* > \boldsymbol{x}_{34}^*\beta^*$. Thus, $Pr(Z^* = 1|\boldsymbol{x}_{12}^*\beta^* > \boldsymbol{x}_{34}^*\beta^*) > Pr(Z^* = 1|\boldsymbol{x}_{12}^*\beta^* = \boldsymbol{x}_{34}^*\beta^*) = 0.5$. By symmetry, we also know that $Pr(Z^* = -1|\boldsymbol{x}_{12}^*\beta^* < \boldsymbol{x}_{34}^*\beta^*) > 0.5$, implying $Pr(Z^* = 1|\boldsymbol{x}_{12}^*\beta^* < \boldsymbol{x}_{34}^*\beta^*) < 0.5$.

3. This result directly follows from definition of expected value and a rearrangement of inequalities $\boldsymbol{x}_{12}^*\beta^* \gtrless \boldsymbol{x}_{34}^*\beta^*$.

$\square$

The first two parts of Lemma 2 allow me to characterize the distribution of the observed dependent variable conditional on the observable characteristics. This plays a key role in the identification of β^* since for any b where $g(b) \neq g(\beta)$, a deviation in expected conditional outcomes occurs when $\boldsymbol{x}_{12}^*\beta^* > \boldsymbol{x}_{34}^*\beta^*$ and $\boldsymbol{x}_{12|b}b < \boldsymbol{x}_{34|b}b$ or vice versa. Part 3 of Lemma 2 is a convenient result that will be used later in the estimation section. I now make a set of primitive assumptions on the space and distributions of the observable characteristics that will allow identification of the parameter β up to a positively-scaled proportion.

Assumption 3 (Distribution of $\boldsymbol{x}$).

1. *The support of $\eta(\boldsymbol{x})$ is not contained in any proper linear subspace of $\mathbb{R}^K$*

2. *There exists at least one $k \in \{1, 2, , K\}$ such that x_k is a pairwise characteristic with $\beta_k \neq 0$. Moreover, for all values of the remaining observables $\tilde{x} = (x_1, , x_{k-1}, x_{k+1}, , x_K)$, the conditional distribution $\eta_k(x_k|\tilde{x})$ has support on $\mathbb{R}_{>0}$.*

Part 1 of Assumption 3 broadly states that the dimension of the sample space for observable characteristics $\boldsymbol{x}$ must be equal to the number of characteristics K. In other words, there is no perfect multicollinearity among observable variables. This

also rules out the identification of constants within the joint surplus function. This is intuitive since comparing differences in joint surplus across different potential pairs results in mutual cancellation of the constant term. Part 2 states that at least one characteristic say x_1 must vary across pairs with an associated nonzero coefficient. This means that there is at least one characteristic among the set of observable variables that affects the joint surplus in every match. The normalization function $g(b)$ is justified as I can set the normalizing coefficient to b_1 without loss of generality. In addition, it also states that the sample space of x_1 must have positive density on $\mathbb{R}_{>0}$ conditional on all values of remaining observables $\tilde{x} = (x_2, , x_K)$. This allows the density of $\boldsymbol{x}b$ conditional on $\tilde{x}$ to have support on $\mathbb{R}_{>c}$ or $\mathbb{R}_{<c}$ for some fixed constant c depending respectively on whether $b_1 > 0$ or $b_1 < 0$ and plays a role in identification of β. Note that this also implies the event $\tilde{x}_1 b = \tilde{x}_2 b$ occurs with zero probability.

Theorem 8. *Under Assumptions 2, 3, and independence of Z_{ij},[2] β is identified up to a positive scale.*

Proof. Let $b \in \{\mathbb{R}^K : b_1 \neq 0\}$ such that $b^* = g(b) \neq g(\beta)$. Following the definition, we need to show that there exists $X \in \mathbf{X}^b$ such that:

$$Pr(Z^* = 1 | \mathbf{x}_{12|b}, \mathbf{x}_{34|b}; b^*) \neq Pr(Z^* = 1 | \mathbf{x}_{12}^*, \mathbf{x}_{34}^*; \beta^*)$$

From Lemma 2, this occurs when $(\boldsymbol{x}_{12}^* - \boldsymbol{x}_{34}^*)\beta^* < 0 < (\boldsymbol{x}_{12|b} - \boldsymbol{x}_{34|b})b^*$ or vice versa. Thus, we can show that:

$$\int_{\mathbf{X}^b} d\mathbf{H}(X) = Pr((\boldsymbol{x}_{12}^* - \boldsymbol{x}_{34}^*)\beta^* < 0 < (\boldsymbol{x}_{12|b} - \boldsymbol{x}_{34|b})b^*)$$

$$+ Pr((\boldsymbol{x}_{12|b} - \boldsymbol{x}_{34|b})b^* < 0 < (\boldsymbol{x}_{12}^* - \boldsymbol{x}_{34}^*)\beta^*) > 0$$

---

[2] This assumption requires i.i.d. stochastic components across surplus functions. Additionally, the deterministic components consist of only pair-wise characteristics. This is quite restrictive. Theorem 10 will provide the general conditions under which individual characteristics may be incorporated.

It suffices to show that either of the component probabilities is greater than 0.

For simplification, let $\boldsymbol{x}^* = \boldsymbol{x}_{12}^* - \boldsymbol{x}_{34}^* = (x_1^*, \tilde{x}^*)$ and $\boldsymbol{x}^b = \boldsymbol{x}_{12|b} - \boldsymbol{x}_{34|b} = (x_1^b, \tilde{x}^b)$ with corresponding vectors $\beta^* = (\beta_1^*, \tilde{\beta}^*)$ and $b^* = (b_1^*, \tilde{b}^*)$. By part 2 of Assumption 3, the distributions for the elements x_1^* and x_1^b have support on $\mathbb{R}$ since they are the differences of two random variables each with positive densities on $\mathbb{R}_{>0}$.

Suppose $\beta_1^* = 1$ (the case for $\beta_1^* = -1$ is symmetric)

For $b_1^* = -1$, conditioning on $\tilde{x}^*$ and $\tilde{x}^b$:

$$Pr(\boldsymbol{x}^b b^* < 0 < \boldsymbol{x}^* \beta^*) = Pr(x_1^b > \tilde{x}^b \tilde{b}^*, x_1^* > -\tilde{x}^* \tilde{\beta}^*) > 0$$

For $b_1^* = 1$, conditioning on $\tilde{x}^*$ and $\tilde{x}^b$:

1. $Pr(\boldsymbol{x}^b b^* < 0 < \boldsymbol{x}^* \beta^*) = Pr(-\tilde{x}^* \tilde{\beta}^* < x_1^b, x_1^* < -\tilde{x}^b \tilde{b}^*)$

2. $Pr(\boldsymbol{x}^* \beta^* < 0 < \boldsymbol{x}^b b^*) = Pr(-\tilde{x}^b \tilde{b}^* < x_1^b, x_1^* < -\tilde{x}^* \tilde{\beta}^*)$

When either $-\tilde{x}^* \tilde{\beta}^* > -\tilde{x}^b \tilde{b}^*$ or $-\tilde{x}^* \tilde{\beta}^* < -\tilde{x}^b \tilde{b}^*$, an open interval exists and either

$$Pr(\boldsymbol{x}^b b^* < 0 < \boldsymbol{x}^* \beta^*) > 0 \text{ or } Pr(\boldsymbol{x}^* \beta^* < 0 < \boldsymbol{x}^b b^*) > 0$$

To ensure this occurs with positive probability, $Pr(\tilde{x}^* \tilde{\beta}^* = \tilde{x}^b \tilde{b}^*) < 1$ needs to hold and is a direct result by part 1 of Assumption 3. $\square$

4.2.2   Consistency of the Score Estimator

The identification result suggests a consistent estimation approach provided the possibility of sampling all observable characteristics $X = (\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4)$ where $\boldsymbol{x}_1, \boldsymbol{x}_2$ respectively correspond to the pairs (u_1, d_1) and (u_2, d_2) while $\boldsymbol{x}_3, \boldsymbol{x}_4$ correspond to the pairs (u_1, d_2) and (u_2, d_1). Based on theorem 3 of the theoretical model in the previous chapter, I also observe the associated indicator variable Z^*. Combining these two observations, I construct a score estimator of the following form:

$$\hat{\beta} = \arg\max_{b \in \mathbf{B}} \frac{1}{N} \sum_{n=1}^{N} Z_n^* sgn(\boldsymbol{x}^b b^*)$$

73

Where b is selected from the set $\mathbf{B} = \{b \in \mathbb{R}^K : |b_1| = 1\}$ and

$$sgn(\boldsymbol{x}^b b^*) = \begin{cases} 1, & \text{if } \boldsymbol{x}^b b^* > 1 \\ -1, & \text{if } \boldsymbol{x}^b b^* < 1 \end{cases}$$

Building from theorem 8 and the Law of Large Numbers, I make a necessary assumption on the sampling of observations that will ensure convergence of the estimator to β^*.[3]

Assumption 4 (Sampling). *The set of observations* (Z_n^*, X_n), $n = 1, 2, ..., N$ *are randomly sampled from a joint distribution* $\boldsymbol{H}(X)$ *characterized by A3 with a relationship between* Z_n^* *and* X_n *that holds under A2.*

The above assumption ties together the properties on the sampling distribution of the observables X_n as stated from Assumption 3 and the implied conditional distribution properties of Z_n^* under Assumption 2. Below, I will define the relationship between the score estimate $\hat{\beta}$ and the underlying parameters β^*.

Definition (Consistency). *An estimator exhibits strong consistency for* β^* *that is contained within a set of candidate values* $\boldsymbol{B}$ *if:*

$$Pr\left[\lim_{N \to \infty} \sup_{b \in \tilde{\boldsymbol{B}}_N} ||b - \beta^*|| = 0 \right] = 1$$

where $\tilde{\boldsymbol{B}}_N$ *is the set of all maximizers of the score estimator contained in* $\boldsymbol{B}$ *with sample size* N.

This definition of consistency is a desirable property for maximum score estimation. Since the optimum of any step-based function creates a set of solution values, the consistency criteria above guarantees that the solution set collapses to the point

---

[3] For $sgn(\boldsymbol{x}^b b^*)$, the case with values 1 and 0 also works. Indicator values 1 and -1 is used for algebraic convenience.

estimate β^* with increasing sample size. I show that this is indeed the case under the given set of assumptions for the defined score estimator.

Theorem 9. *Under Assumptions 2-4 and independence of Z_{ij}, the maximum score estimate is strongly consistent for β^* within the set $\boldsymbol{B} = \{b \in \mathbb{R}^K : |b_1| = 1, ||b|| \leqslant ||\beta^*||\}$. (Alternatively, within the set $\boldsymbol{B} = \{b \in \mathbb{R}^K : |b_1| = 1, ||b|| \leqslant \gamma\}$ for any positive constant $\gamma \geqslant ||\beta^*||$)*

Proof. Let the objective function be denoted by $S_N(b) = \frac{1}{N} \sum_{n=1}^{N} Z_n^* sgn(\boldsymbol{x}^b b^*)$ and $S(b) = \lim_{N \to \infty} S_N(b)$. Additionally, let $Z = Z_{12} - Z_{34}$ denote the latent dependent variable, from which I only observe Z^* ($Z > 0$ or $Z < 0$). This result is shown in three steps and will be broken down into three component results.

1. β^* is the unique maximizer of $S(b)$:

   By Law of Large Numbers, $S(b) = E[Z^* sgn(\boldsymbol{x}^b b^*)]$. For $b \neq \beta^*$,

   $$S(\beta^*) - S(b) = E[Z^*(sgn(\boldsymbol{x}^* \beta^*) - sgn(\boldsymbol{x}^b b^*))] = 2 \int_{\boldsymbol{X}^b} E[Z^* | \boldsymbol{x}^*] sgn(\boldsymbol{x}^* \beta^*) d\mathbf{H}(X)$$

   From part 3 of Lemma 2, $E[Z^* | \boldsymbol{x}^*] sgn(\boldsymbol{x}^* \beta^*) = |E[Z^* | \boldsymbol{x}^*]| > 0$. Hence,

   $$S(\beta^*) - S(b) = 2 \int_{\boldsymbol{X}^b} |E[Z^* | \boldsymbol{x}^*]| d\mathbf{H}(X) > 0$$

   since $\int_{\boldsymbol{X}^b} d\mathbf{H}(X) > 0$ under theorem 8.

2. $S_N(b)$ converges to $S(b)$ uniformly on $\mathbf{B}$, almost surely:

   $$S_N(b) = Pr_N[Z > 0, \boldsymbol{x}^b b^* > 0] + Pr_N[Z < 0, \boldsymbol{x}^b b^* < 0]$$
   $$- Pr_N[Z < 0, \boldsymbol{x}^b b^* > 0] - Pr_N[Z > 0, \boldsymbol{x}^b b^* < 0]$$

75

By random sampling from Assumption 4 and Law of Large Numbers, each component probability converges pointwise to the respective component probabilities of S(b):

$$S(b) = Pr[Z > 0, \boldsymbol{x}^b b^* > 0] + Pr[Z < 0, \boldsymbol{x}^b b^* < 0]$$

$$- Pr[Z < 0, \boldsymbol{x}^b b^* > 0] - Pr[Z > 0, \boldsymbol{x}^b b^* < 0]$$

Uniform convergence of $S_N(b)$ to $S(b)$ follows from the result by Rao (1962), which shows that for any finite number m:

$$\lim_{N \to \infty} \left[\sup_{Z \in H_m} |Pr_N(Z) - Pr(Z)| \right] = 0$$

Where H_m is the class of all intersections of m half-spaces of a random vector space $\mathbb{R}^M$. Uniform convergence follows.

3. $S(b)$ is continuous for all b such that $b_1 \neq 0$:

To show this property, decompose $\boldsymbol{x}^b b^*$ into components $x_1^b b_1^* + \tilde{x}^b \tilde{b}^*$ and rewrite $S(b)$:

$$S(b) = Pr[Z > 0, x_1^b b_1^* + \tilde{x}^b \tilde{b}^* > 0] + Pr[Z < 0, x_1^b b_1^* + \tilde{x}^b \tilde{b}^* < 0]$$

$$- Pr[Z < 0, x_1^b b_1^* + \tilde{x}^b \tilde{b}^* > 0] - Pr[Z > 0, x_1^b b_1^* + \tilde{x}^b \tilde{b}^* < 0]$$

$$= S^1(b) + S^2(b) + S^3(b) + S^4(b)$$

Let $\Phi(\tilde{x}^b)$ be the distribution of $\tilde{x}^b$ and $\phi_1(x_1^b | \tilde{x}^b)$ be the density of x_1^b conditional on $\tilde{x}^b$. Take the first component of $S(b)$ and suppose $b_1^* > 0$ (all other components and the case of $b_1^* < 0$ is demonstrated similarly):

$$Pr[Z > 0, x_1^b b_1^* + \tilde{x}^b \tilde{b}^* > 0] = \int_{\tilde{x}^b} \int_{\frac{-\tilde{x}^b \tilde{b}^*}{b_1^*}}^{\infty} Pr(Z > 0 | \boldsymbol{x}^b) \phi_1(x_1^b | \tilde{x}^b) dx_1^b d\Phi(\tilde{x}^b)$$

Note that the inner integral:

$$s^1(b, \tilde{x}^b) = \int_{\frac{-\tilde{x}^b \tilde{b}^*}{b_1^*}}^{\infty} Pr(Z > 0 | \boldsymbol{x}^b) \phi_1(x_1^b | \tilde{x}^b) dx_1^b$$

76

Is continuous in b, measurable in $\tilde{x}^b$, and uniformly bounded. The Lebesgue dominated convergence theorem implies continuity for $S^1(b)$:

$$S^1(b) = \int_{\tilde{x}^b} s^1(b, \tilde{x}^b) d\Phi(\tilde{x}^b)$$

The continuity of $S(b)$ follows.

Having shown that β^* is the unique maximizer of the continuous limiting function $S(b)$ and that $S_N(b)$ converges to $S(b)$ uniformly over the compact set $\mathbf{B} = \{b \in \mathbb{R}^K : |b_1| = 1, ||b|| \leq ||\beta^*||\}$ containing β^*, strong consistency follows from Manski (1983). $\qquad\square$

The independence of Z_{ij} is a strong condition. Assumption 2 gives a primitive characterization of the conditional distribution for the joint surplus. Together with the independence of Z_{ij}, they establish a more generalizable condition for identification and consistency of score estimation. I will state the more generalized theorem required for strong consistency, the proof of which is clear from the steps shown above.

Theorem 10. *Given Assumptions 3-4 and the following conditions:*

1. *$MED[max\{\pi_{u,d}, \pi_{u',d'}\} - max\{\pi_{u,d'}, \pi_{u',d}\}|\boldsymbol{x}^*\beta^* = 0] = 0$*

2. *$MED[max\{\pi_{u,d}, \pi_{u',d'}\} - max\{\pi_{u,d'}, \pi_{u',d}\}]$ is monotonically increasing in $\boldsymbol{x}^*\beta^*$*

The maximum score estimate is strongly consistent for β^* within the set $\boldsymbol{B} = \{b \in \mathbb{R}^K : |b_1| = 1, ||b|| \leq ||\beta^*||\}$.

Note that the combined conditions in theorem 10 is a weaker version of the assumption by Manski (1985), which require the median to be equal to $\boldsymbol{x}^*\beta^*$. The example below illustrates the above theorem:

Example. *Let $\pi_{u,d} = \beta_0 + \beta_1 * x_{u,d} + \beta_2 * x_u + \beta_3 * x_d + \beta_4 * x_u * x_d + \epsilon_{u,d} = \boldsymbol{x}_1\beta$ and*

$\pi_{u',d'} = \beta_0 + \beta_1 * x_{u',d'} + \beta_2 * x'_u + \beta_3 * x'_d + \beta_4 * x'_u * x'_d + \epsilon_{u',d'} = \boldsymbol{x}_2\beta$

Similarly, define $\pi_{u,d'} = \boldsymbol{x}_3\beta$ and $\pi_{u',d} = \boldsymbol{x}_4\beta$

Also suppose $Max\{\boldsymbol{x}_1\beta, \boldsymbol{x}_2\beta\} = \boldsymbol{x}_1\beta$ and $Max\{\boldsymbol{x}_3\beta, \boldsymbol{x}_4\beta\} = \boldsymbol{x}_3\beta$

Given the following conditions:

1. *$Pr[max\{\pi_{u,d}, \pi_{u',d'}\} > max\{\pi_{u,d'}, \pi_{u',d}\}|\boldsymbol{x}_1\beta = \boldsymbol{x}_3\beta] = 0.5$*

2. *$Pr[max\{\pi_{u,d}, \pi_{u',d'}\} > max\{\pi_{u,d'}, \pi_{u',d}\}|(\boldsymbol{x}_1 - \boldsymbol{x}_3)\beta]$ is increasing in $(\boldsymbol{x}_1 - \boldsymbol{x}_3)\beta$*

The score estimator is consistent for β up to a positive scale.

4.3   Discussion

Matching estimation addresses a form of externality that occurs when the partner selection decisions of some firms limit the potential partner choice set of other firms. The preceding section provides the formal results that show how a consistent estimator can be constructed under a nontransferable utility (NTU) matching framework with payoffs given by a fixed sharing rule. According to theorem 2 from the previous chapter, an equilibrium can be sufficiently characterized by conditions involving only individual and dyadic pairs of firms. In the case of individual firms, the equilibrium implies that the payoffs of firms within all observed matches is positive. Since this is true for both firms within an observed match, this condition is equivalent to all observed matches having positive surplus. Unfortunately, we're unable to use the unobserved matches to infer that the joint surplus between counterfactual pairs is negative. This is because the nature of a matching market allows pairs to be positive valued and unmatched due to congestion from other matches. This prevents me from estimating the coefficients using a binary choice model on dyadic pairs. Likewise, I'm also unable to use multinomial models from the perspective of any firm since I don't observe their choice sets.

The remaining alternative is to utilize the equilibrium characterization on dyadic pairs, which says that no unmatched pair of firms would prefer to form a match while forgoing one existing match on each side. Under a fixed sharing rule, this criterion translates to the condition that no unmatched pair of firms have a joint surplus value greater than the surplus of two specific observed matches. The maximum score estimator is thus constructed by comparing the scores of two observed matches with those of their counterfactual matches. One implication of this procedure is that since I'm comparing the values between matches, I cannot recover the absolute level of surplus. This contrasts with binary choice models that estimate a constant in the latent value function, which may represent the reservation utility of individuals relative to a default option. Another issue arises concerning which pairs of observed matches should the researcher select as inputs for the estimator. In small samples where variation in subsets of observations can meaningfully impact estimated values, the use of all pairs can provide reliably replicable results. However, as Fox (2017) notes, the combinatorial nature of all pairs of observed matches makes it difficult to implement as the market grows large and a random selection of pairs is advised when working with large datasets. It is important to note that random sampling is implemented on the pairs of observed matches since these are the units of observation whose realized joint values can be characterized under the matching framework.

Given a set of optimal solutions from maximum score estimation, inference can be done using either a smoothing composition function as proposed in Horowitz (1992) or constructing confidence regions on the set of optimal solutions to the step-based score function. A smoothing function provides point estimates for a given set of observations and allows the use of bootstrapping to output standard errors. A method for constructing confidence regions on the set of optimal coefficient estimates was proposed by Romano and Shaikh (2010) using a subsampling procedure on the score function. Since the smoothing methodology does not have matching theoret-

ical justification for selecting an optimal coefficient value, set inference is generally preferred.

Lastly, the methodology is more generally applicable than the two-sided many-to-many matching market. A within-group matching of individuals or firms such as co-authorship networks or horizontal alliances among firms can also utilize the maximum score estimation procedure. One important difference is that the fixed sharing rule is now strictly one half between partners whereas in a two-sided model, it is only restricted to be constant for all firms on the same side. Given a fixed sharing rule of one half and strict preferences across potential matches, the Greedy algorithm continues to generate the unique stable set of matches. Methodologically, a greater set of inequalities can be utilized since any two observed matches can potentially generate two sets of counterfactual pairs (illustrated by the example below of within-group matching among software firms versus 2-sided matching between software and communication firms).
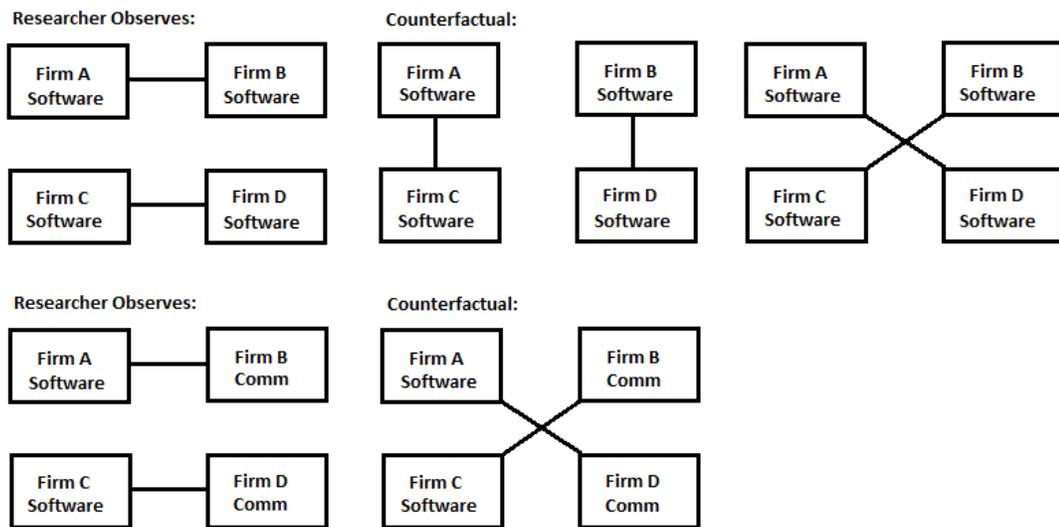
FIGURE 4.1: Example of a 2-Sided and Within-Group Matching Market

The Interplay Among Alliance Value Creation, Partner Value Capture, and the Technological Characteristics of Firms

The longstanding alliance literature acknowledges that collaboration networks among firms confer significant strategic advantages. Alliances have been shown to be instrumental in firm survival (Mitchell and Singh, 1996), growth (Powell, Koput, and Smith-Doerr, 1996), and long-term prosperity (Hagedoorn, 1993; Hamel 1991; Stuart, Hoang, and Hybels, 1999). Particularly in high technology sectors, alliances have become routine strategic initiative, drawing scholars to study the antecedent conditions that lead to formation. Recognizing the important role of firm technological compositions (Mowery, Oxley and Silverman, 1998) and the distinctive dynamics of the alliance formation process (Arora, Belenzon, and Patacconi, 2016), two critical issues exist that the empirical literature has yet to address.

First, while the role of technological overlap on inter-firm collaboration has gained wide acceptance in the empirical literature on alliance formation (Rothaermel and Boeker, 2008; Kavusan, Noorderhaven, and Duysters, 2016; etc.), the technologies in the conceptualizations of overlap may be manifested as either technological interests

or technological domains with each having different implications for how alliance value is generated. Technological interests refer to the knowledge base that a firm possess, which plays a critical role in directing its focal development activities within a broader technological space where organizations invest their limited time and resources to explore and discover new knowledge that may hold potential value (Ocasio, 1997). On the other hand, technological domains correspond to different product areas where the firm is present and is a form of resource or capability that enables it to derive commercial or realized value from its technological development activities (Penrose, 1959; Barney 1991; Teece, Pisano and Shuen, 1997). These conceptual distinctions can be clarified with an illustrative example like Apple Computers, which spent three years developing mobile technologies for its first handset debut in 2007. During this time, Apple's technological interests or knowledge of mobile computing and portable communication directed its focus toward smartphones, resulting in a reallocation of R&D spending and key personnel away from its own personal computer division. The inventive outputs from its efforts on the iPhone such as capacitive touch screens were also used to enhance the commercial value within other technological domains such as portable music players and laptop computers.

Alliance partner selection needs to consider both the technological interests of other firms as well as the collective breadth of technological domains across all partners. Aligned or common technological interests of firms create mutual understanding and facilitate knowledge exchange from similar internal activities of both partners (Cohen and Levinthal, 1990) while technological domains reflect the diversity of market channels that become accessible within an alliance (Teece, 1989; Bucklin and Sengupta, 1993). I study the impact of both technological characteristics on alliance formation in the software and communications industry. Recognizing that alliances impose coordination costs and resource commitments that affect the number of partnerships each firm can sustain at any point in time (Gulati and Singh, 1998; Arts

and Brush, 2000; Cummings and Kiesler, 2007), I employ a matching methodology that accounts for the existence of capacity constrained firms.

In addition to distinguishing between the firm's technological interests and technological domain, I also examine how value creation within alliances is affected by the location of partners' common interests within the broader technological space. Specifically, the common interests of two firms may reside in technologies owned by alliance members or be co-located in technologies that are external to all partners. A rich literature have highlighted the benefits of external knowledge and the decisions of firms to search for new technology outside their organization (Nelson and Winter, 2009; March 1991). Indeed, within the high tech setting, firms may potentially realize high commercial value from the technologies owned and developed in other companies as evidenced by the case of Xerox and the Graphical User Interface (Chesbrough and Rosenbloom, 2002). Although prior work has suggested firms can form alliances to learn from outside partner's technologies (Rosenkopf and Almeida, 2003), recent work have only begun to explore alliances around technologies that are unfamiliar to both partners (Anand, Oriani, and Vassolo, 2010; Schilling 2015).

A second central issue is that while firms compete with one another for scarce alliance partners (Macdonald and Ryall, 2004), potential partners are compared based on the private value the focal firm can capture within an alliance (Khanna, Gulati, and Nohria, 1998). Although the firm's opportunity cost has been directly related to the bargaining positions of partners in determining private value capture, the outside options of a firm within an alliance have been focused on alternative potential partners (Yan and Gray, 1994; Lavie, 2007; Greve, Mitsuhashi, and Baum, 2013). This may be reasonable if switching costs were low, but given the nature of incomplete contracts (Hart and Moore, 1988), the evolution of learning (Hamel, 1991) and commitment to relation-specific investments (Williamson, 1979), opportunity costs must reflect the conditions after an alliance market but prior to the completion of the

alliance agreement. For two firms with common technological interests that engage in a collaboration, each partner's subsequent opportunity cost will correspond to the value it can realize by itself given its own composition of technological interests and breadth of technological domains.

I tie together the conceptualizations of technological interest and technological domains in a model that draws its assumptions from the theory of recombinant search (Fleming and Sorenson, 2004; Kauffman et al., 2000). Within the model, firms produce joint knowledge as an increasing function of their common technological interests. Once the knowledge is produced (i.e. a higher technological quality has been reached), its value is realized by applying it to the partners' breadth of technological domains. Two principal assumptions concern the relative rates of knowledge production. Based on the interplay between common technological interests of partners and inter-firm alliance behavior, some features of the technological landscape can be inferred along with managerial implications. This may be particularly important for young firms that are unfamiliar with the technological landscape that they inhabit and can thus learn from observing interactions among more established players.

To test my predictions while building on the perspective of competing for scarce partners in an alliance market (Mindruta, Moeen, and Agarwal, 2016), I employ a new matching model that is more suited to a decentralized alliance formation process. In addition to accounting for alliance congestion among potential partners, the form of matching model used also considers the existence of firm outside options over the duration of an alliance. This is important as alliance contracts tend to be incomplete, allowing each partner to exercise their outside option and terminate an agreement when its own value capture from an alliance is unfavorable. This means firms with higher valued opportunity costs are in better bargaining positions to extract a greater share of joint value and will thus choose partners that maximize the alliance *net value*

or *surplus*, which is defined as the total value generated within an alliance that is net of the total value the partners can create without each other. This allows me to also examine firm individual characteristics such as size, which may relate to the value of each partner's opportunity costs and affect its share of value captured within an alliance, a topic of recent interest (Diestre and Rajagopalan, 2012).

Utilizing the matching model, I examine alliance formation among software and communication firms over the period 1990-2005, which experienced the growth of the home computing industry, increasing mobile phone usage, and most importantly the shock of the internet. This generated substantial alliance activity within a fixed industry setting where firms were partnering to work on each other's technologies as well as to jointly explore an uncharted general purpose technology that was the world wide web. I find that alliance surplus is positively affected by firms that have common technological interest and high technological breadth, conforming to the theoretical predictions. Moreover, external common interests generate more value than internal common interests. Lastly, relatively large firms extract a greater share of the total value generated within an alliance.

This work contributes to three streams of literature. First, it contributes to the alliance literature by examining the antecedents of formation, taking care to distinguish between the key technological traits of firms that were hitherto characterized under the broad umbrella of overlap. It also accounts for firm opportunity costs, providing an intuitive rationale for why big firms can reap a greater share of alliance value. Secondly, it contributes to the literature on technology and innovation management, acknowledging that many commercially valuable ideas may have origins outside the firm's boundaries and thus considering the relative alliance value contributions of common technological interests that are external to both partners versus those that are owned within the alliance. Lastly, it contributes to the literature on recombinant search by linking the potential of collaborative knowledge production

with implications for partner selection.

The paper proceeds as follows. The next section describes related literature and key distinctions that characterize the alliance market while drawing on matching theory and explaining the differences compared to alternative models (Shapley and Shubik, 1972). I then proceed to develop the hypotheses related to the role of overlapping technological interests and technological breadth on alliance formation. This is followed by data descriptions that clarify the empirical context. Next, I outline the methodological approach that is used to test my hypotheses, which directly precedes an analysis of results. A brief discussion section is given before the concluding section.

5.1 Alliance Formation

5.1.1 Related Literature

Alliance scholars have studied the determinants of formation from various perspectives. One set of studies focuses on the perspective of the focal firm, seeking to answer the question which partners would the focal firm select? This body of work examines individual traits such as product development capability (Rothaermel, 2002), status (Stern et al., 2014), or firm-specific uncertainty (Beckman et al., 2004) as well as pairwise specific characteristics such as technological/market overlap (Mowery et al., 1998; Dushnitsky and Shaver, 2009; Diestre and Rajagopalan, 2012), geographic proximity (Alcacer et al., 2009) and reputational similarity (Li and Rowley, 2002; Hallen, 2008). Another set of work uses the perspective of the dyadic pair, focusing on the question of which paired firms are likely to form alliances? and investigating factors such as resource similarity (Rothaermel and Boeker, 2008) and network positions (Ahuja et al., 2009). Both sets of work assume that decisions to form alliances with all the potential partners in a choice set are independent of one another. However, if some firms are at capacity given their commitment to current

alliance partners, then the independence assumption is violated.

To address the existence of capacity constrained firms, an emergent set of work have begun to examine alliances through the perspective of a two-sided matching market with transferable utility (Fox, 2010). This framework has been applied to match alliances of venture capital investors with biotech companies (Fox et al., 2012), biotech with pharma (Mindruta et al., 2016), and professional athletes with sports teams (Yang et al., 2009). Rather than consider individual firm-pairs as observations, the matching model considers two observed alliances among four distinct firms with two on each side. The counterfactual is constructed by swapping partners among the four firms so that capacities are not affected. This allows the researcher to recover the effects of pair-wise characteristics on a latent alliance value without being subjected to the biases of traditional binary choice models imposed by capacity constrained firms. Moreover, since the counterfactuals do not increase the original observed capacities of firms in the data, no information on the capacity limitations of firms in the market is required for the estimation to be implemented.

A central question arises concerning how the counterfactual pair of alliances would compare to the observed pair of alliances. Under transferable utility, the equilibrium maximizes the aggregate value of all alliances formed in the market. If this were true, then the sum of values of the counterfactual alliance pair would be less than the sum of values of the observed alliance pair. Such a framework would fit well within centralized settings such as the planning of hospital matches between medical interns and residency programs or kidney donors and patients (Roth and Sotomayor, 1992; Abdulkadiroglu and Snmez, 2013). On the other hand, if we assume that matches within an alliance market materializes through a decentralized process (i.e. greedy algorithm), whereby highest valued alliances form in sequential order given the feasible space of matches at each stage, then the comparison of sums no longer characterizes the resulting equilibrium.

Transferable utility also implies that firm individual characteristics do not change the equilibrium sorting of alliances since its contribution value in one alliance or another does not affect the aggregate sum of all realized alliance values. However, identification of individual characteristics matter when considering private value capture within alliances since they may correspond to the availability of outside options after the formation of an alliance but prior to its completion. This is important because the value of each partner's opportunity costs determine its relative bargaining position concerning how the joint alliance value is divided. Firms in the alliance formation stage would anticipate this outcome and factor it into its decision process when selecting partners.

Lastly, a two-sided market approach would severely limit the set of alliances that are analyzed. Although many contextual cases conform well to a two-sided approach, the literature has yet to examine a within group matching of alliances. Given the trending phenomena of increasing intra-industry alliances relative to inter-industry alliances in high tech markets such as software and biotech, a matching design needs to be adapted to a within-group setting.

5.1.2 A Proposed Matching Market Perspective of Alliances

Matching models solve the problem of finding suitable assignments among participants in markets where agents compete for a limited supply of partners (Roth and Sotomayor, 1992). Oftentimes, the process of matching is invisible to the outsider (Stovel and Fountain, 2009), but the final configurations of pairings is observed. Therefore, to properly analyze an alliance market requires a theoretical formulation of sensible firm preferences and rules that govern the matching process under which an equilibrium outcome may emerge. The equilibrium outcome would then exhibit properties that reflect the underlying matching process.

I formulate the matching model as follows. For any two firms A and B, let V_{AB}

designate the joint value from an alliance and v_A, v_B respectively denote the opportunity costs of each partner after the formation stage but prior to its completion. Each party may terminate the alliance and capture its respective outside option value at any time prior to completing the agreement. Over the course of an alliance, the values v_A, v_B may be referred to as rival benefits since they can only be obtained for each firm by prematurely breaking off the alliance (Arora, Belenzon and Patacconi, 2016).

Given that v_A, v_B can be captured by each respective firm without completing the alliance, the value that remains to be divided up between the partners upon alliance completion is the surplus π_{AB}, which is:

$$\pi_{AB} = V_{AB} - v_A - v_B$$

From Nash Bargaining, $\frac{1}{2}$ is used to divide up the total surplus between partners. Therefore, from the total alliance joint value V_{AB}, each firm captures a private value that is equal to:

$$\pi_{AB}^{A} = \frac{1}{2}(V_{AB} - v_A - v_B) + v_A$$

$$\pi_{AB}^{B} = \frac{1}{2}(V_{AB} - v_A - v_B) + v_B$$

The first term represents the added value that the firm gains from completing the alliance. This is obtained by differencing out each firm's rival benefit from the total joint value and applying Nash Bargaining, leaving us with each firm's share of the surplus π_{AB} that it captures. The second term is each firm's rival benefit, which can be captured regardless of the outcome and therefore is not subjected to negotiation. When rival benefits are positive, partners face opportunity costs of staying in the alliance; this may reflect a variety of circumstances where firms have more than one way of reaching their business objectives (Capron and Mitchell, 2012). To be incentivized to fulfill the alliance agreement, each firm would need to capture

at a minimum its opportunity cost and under Nash bargaining, would divide up the remaining surplus by $\frac{1}{2}$.

Now suppose we have three firms A, B, and C. From firm A's perspective, the decision to form an alliance with either B or C depends on its private value from the potential alliances: π^A_{AB}, π^A_{AC}. When firms are unconstrained, these decisions may be treated independently (i.e. $\pi^A_{AB} \gtrless 0, \pi^A_{AC} \gtrless 0$). If firm A is constrained to select one partner, then its decision depends on the relative differences between the two private values, leading to a comparison of the match surpluses: $\pi_{AB} \gtrless \pi_{AC}$.

Under this theoretical setup, if we observe two disjoint alliances with firms (A,B) in alliance 1 and firms (C,D) in alliance 2, a counterfactual alliance such as (A,C) will not form if either such an alliance is unprofitable ($\pi_{AC} < 0$) or a constrained firm (either A or C) obtained higher value from their observed partner ($\pi^A_{AB} > \pi^A_{AC}$ or $\pi^C_{CD} > \pi^C_{AC}$). Since it is necessary for the observed alliances to have positive value ($\pi^A_{AB} > 0$ and $\pi^C_{CD} > 0$), this leads to the following proposition.[1]

Proposition 1. *Given firms A, B, C, D with observed alliances between (A,B) and (C,D) but no observed alliance between (A,C), then either:*

$$\pi^A_{AB} > \pi^A_{AC} \ or \ \pi^C_{CD} > \pi^C_{AC}$$

An alternative way of writing this condition is that the maximum of the surpluses between the two observed alliances must be greater than the counterfactual alliance.

$$max\{\pi_{AB}, \pi_{CD}\} > \pi_{AC}$$

Since this is true for any counterfactual alliance that can be constructed from two observed alliances, we can formulate a proposition that compares two observed alliances with their corresponding pair of counterfactuals obtained by switching partners.

---

[1] The formal proofs of the propositions are provided in the Appendix.

Proposition 2. *Given firms A, B, C, D with observed alliances between (A,B) and (C,D) but no other observed alliances among them, two conditions must be true:*

1. $max\{\pi_{AB}, \pi_{CD}\} > max\{\pi_{AC}, \pi_{BD}\}$

2. $max\{\pi_{AB}, \pi_{CD}\} > max\{\pi_{AD}, \pi_{BC}\}$

Note that this is different from alternative matching models that embed availability of outside partners into the firm's opportunity cost (Andersson, Gudmundsson, Talman, and Yang, 2013) or assume transferrable utility (Fox, 2010; Mindruta et al. 2016). One important empirical implication is that since firms maximize their own value capture across potential partners rather than the sum value of all alliances in the market, the partner's individual characteristics play an important role in determining the equilibrium outcome. Firms seek to select partners that generate a high joint value while avoiding those that would capture a large proportion of the value for themselves. I use these conditions as estimation criteria and formulate an extension of the methodology developed in the preceding chapter to estimate the effects of overlapping technological interests and technological breadth on alliance formation.

5.2   Value Creation and Appropriation Within Alliances

Building off the previous section, alliances are worthwhile arrangements between two firms if their collaborative efforts can generate greater value than the sum of values that each partner can achieve on their own. A meaningful analysis of the added value from collaboration (i.e. alliance surplus) must account for both the relative benefits and deficiencies of alliance formation. On one hand, alliances enable partners to:

1. Share and exploit each other's knowledge for technological development (Zahra and George, 2002; Arora et al, 2016)

2. Expand the commercialization of technologies to new markets (Bucklin and Sengupta, 1993)

On the other hand, alliances require the establishment of communication channels to divide tasks across organizations and coordinate ongoing activities (Gulati and Singh, 1998). This implies that the tasks of technological development and commercialization can not only be carried out more efficiently within the firm where communication channels and coordination mechanisms already exist but that organizational resources from coordination cost savings can be redeployed to new internal projects.

I formalize these ideas in a model that furnishes a set of hypotheses relating the concepts of technological interests, technological domains, and firm size with alliance value surplus. Common technological interests between partners indicate the existence of overlapping knowledge bases that enable firms to learn from each other's technological development under an alliance (Cohen and Levinthal, 1990). Technological domains represent the number of markets that a firm's technology can be commercialized. Thus, non-overlapping technological domains between partners reflect the potential of spreading technologies co-developed within an alliance to new product areas. Figure 1 provides diagrams that illustrate the concepts of Firm Technological Interests, Common Technological Interests of Partners, and Technological Domains. Lastly, firm size is conceptualized as the size of each partner's customer base, which corresponds positively to the value of new internal projects that each partner can capture by forgoing alliance formation.

To characterize the process of technological development, I follow the literature on recombinant search, in which the term recombinant refers to the idea that invention or the creation of new technologies is most often the result of novel re-combinations or reconfigurations of basic sub-components (Penrose, 1959; Henderson and Clark,
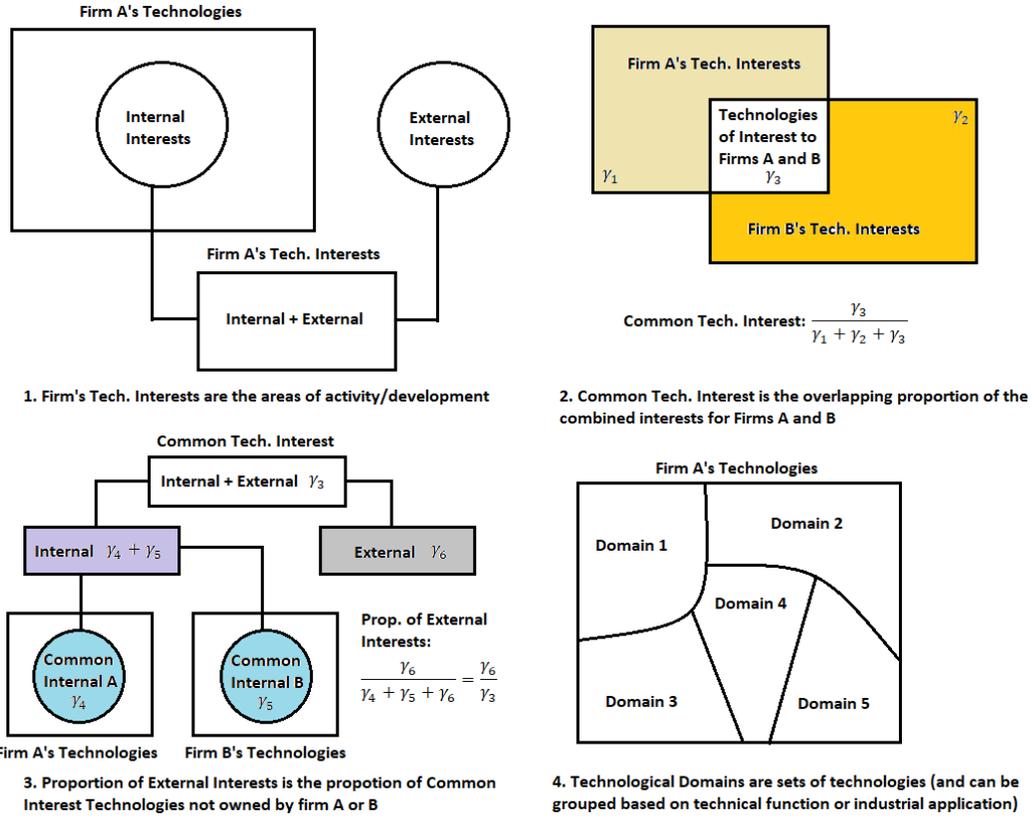
FIGURE 5.1: An Illustration of Firm Technological Interests, Common (Internal and External) Interests of Partners, and Technological Domains

1990; Fleming and Sorenson, 2004; Nelson and Winter, 2009). Through the process of search over a technological landscape, firms can discover the quality or value of different combinations that were initially unknown (Kauffman et al., 2000). Here, the developed hypotheses from the model will provide insight on how features of the technological landscape will correspond to certain relationships between the composition of common technological interests and alliance surplus value.

5.2.1 Model Setup

Let the market presence of firm A be represented by a vector $\boldsymbol{\beta_A}$:

$$\boldsymbol{\beta_A} = (\beta_A^1, \beta_A^2, ..., \beta_A^M) \text{ where}$$

$$\beta_A^i = \begin{cases} 1, & \text{if firm A is present in Market i} \\ 0, & \text{if firm A is not present in Market i} \end{cases}$$

Also let the quality of firm A's technology be given by the function $\theta(\boldsymbol{\omega_A})$, where the input $\boldsymbol{\omega_A}$ is a configuration that summarizes firm A's position on a technological landscape:

$$\boldsymbol{\omega_A} = (\omega_A^1, \omega_A^2, ..., \omega_A^N)$$

with each element ω_A^i having S possible states. The idea is that the firm is unaware of the values of other possible and unexplored positions on the landscape. Firm A will thus conduct an experiment to see if it can improve its technological position by changing the state of an element within its current configuration. Let $\boldsymbol{\omega_A'}$ denote the experimental configuration from a walk on the technological landscape:

$$\boldsymbol{\omega_A'} = (\omega_A'^1, \omega_A'^2, ..., \omega_A'^N) \text{ such that}$$

$$\omega_A'^j \neq \omega_A^j \text{ for some } j \in \{1, 2, ..., N\} \text{ and } \omega_A'^i = \omega_A^i \text{ for all } i \neq j$$

If firm A improves its technological position (i.e. $\theta(\boldsymbol{\omega_A'}) > \theta(\boldsymbol{\omega_A})$), it adopts the new technological configuration. Otherwise, it remains in its current state.

Let the same corresponding terms be defined for firm B. The market overlap between both firms is given by a dot product:

$$O(\boldsymbol{\beta_A}, \boldsymbol{\beta_B}) = \boldsymbol{\beta_A} \cdot \boldsymbol{\beta_B}$$

While the number of Non-Overlapping Markets for each firm is given by the difference between the L1-norm and dot product:

Non-overlapping Markets for Firm A: $NO(\boldsymbol{\beta_A}, \boldsymbol{\beta_B}) = ||\boldsymbol{\beta_A}||_1 - \boldsymbol{\beta_A} \cdot \boldsymbol{\beta_B}$

Non-overlapping Markets for Firm B: $NO(\boldsymbol{\beta_B}, \boldsymbol{\beta_A}) = ||\boldsymbol{\beta_B}||_1 - \boldsymbol{\beta_B} \cdot \boldsymbol{\beta_A}$

If firm A and firm B independently conduct experiments on the technological landscape, their realized values are given by:

$$\lambda_A = (max\{\theta(\boldsymbol{\omega'_A}), \theta(\boldsymbol{\omega_A})\})(NO(\boldsymbol{\beta_A}, \boldsymbol{\beta_B})$$

$$+ \mathbb{1}[max\{\theta(\boldsymbol{\omega'_A}), \theta(\boldsymbol{\omega_A})\} > max\{\theta(\boldsymbol{\omega'_B}), \theta(\boldsymbol{\omega_B})\}]O(\boldsymbol{\beta_A}, \boldsymbol{\beta_B}))$$

$$\lambda_B = (max\{\theta(\boldsymbol{\omega'_B}), \theta(\boldsymbol{\omega_B})\})(NO(\boldsymbol{\beta_B}, \boldsymbol{\beta_A})$$

$$+ \mathbb{1}[max\{\theta(\boldsymbol{\omega'_A}), \theta(\boldsymbol{\omega_A})\} < max\{\theta(\boldsymbol{\omega'_B}), \theta(\boldsymbol{\omega_B})\}]O(\boldsymbol{\beta_A}, \boldsymbol{\beta_B}))$$

The above equation states that each firm's value is a product of its technological quality (post-experiment) and the number of markets where it has dominant presence. The number of markets where the firm has dominant presence is equal to the sum of its non-overlapping markets and the overlapping markets where the focal firm has superior technology, allowing it to outcompete the opposing firm.

Both firms can observe each other's current technological positions, but do not observe the corresponding value (quality) of those positions nor each other's experimental configurations, which is information that external firms are generally not privy to access. On the other hand, when an alliance is formed, everything is observed. The joint value is thus given by the product of what both firms learn from observing each other's qualities $L(\cdot)$, and the combined market presence of both firms:

$$V_{AB} = L(\boldsymbol{\omega_A}, \boldsymbol{\omega'_A}, \boldsymbol{\omega_B}, \boldsymbol{\omega'_B})[NO(\boldsymbol{\beta_B}, \boldsymbol{\beta_A}) + NO(\boldsymbol{\beta_A}, \boldsymbol{\beta_B}) + \rho O(\boldsymbol{\beta_A}, \boldsymbol{\beta_B})]$$

In overlapping markets, the value will be less due to competition between firms with same quality technologies ($0 < \rho < 1$).

5.2.2 Coupled Versus Modular Technological Components

Technologies are modular when each component's contribution to quality is independent of all other components. In this case, the technological value function θ can

be written as a sum of N independent functions:

$$\theta(\boldsymbol{\omega}) = \theta_1(\omega^1) + \theta_2(\omega^2) + ... + \theta_N(\omega^N)$$

This means that firms can systematically reach the global optimum on the landscape by searching for the optimum of each component separately. The modular property of technologies would consequently favor a search strategy of experimenting with one component at a time within a given technological configuration to learn its individual effect on overall quality.

On the other hand, technologies are coupled when each component's contribution to quality is dependent on the configuration states of all other components. This implies that the only way for firms to be sure that a global optimum is reached is by having experimented with all other possible configurations.

Example. *Consider a technological landscape defined by two elements* $\boldsymbol{\omega} = (\omega^1, \omega^2)$ *with each element taking on two potential values* $\omega^i \in \{1, 2\}$ *An example of a value function* $\theta(\boldsymbol{\omega})$ *when the landscape is modular versus coupled is given below:*

$\theta(\boldsymbol{\omega})$	*Modular*	*Coupled*
$\boldsymbol{\omega} = (1, 1)$	*0*	*5*
$\boldsymbol{\omega} = (1, 2)$	*3*	*2*
$\boldsymbol{\omega} = (2, 1)$	*4*	*15*
$\boldsymbol{\omega} = (2, 2)$	*7*	*1*

Assuming the firm starts off in position $\boldsymbol{\omega} = (1, 1)$*, then discovering the values of* $\boldsymbol{\omega} = (1, 2)$ *and* $\boldsymbol{\omega} = (2, 1)$ *allows the firm to infer the value of* $\boldsymbol{\omega} = (2, 2)$ *on a modular landscape. This is not the case when components are coupled; prior experience does not provide any information on unexplored configurations.*

In the case of coupled technological components, the learning of partners from observing each other's qualities would simply be the highest value:

$$L(\boldsymbol{\omega_A}, \boldsymbol{\omega'_A}, \boldsymbol{\omega_B}, \boldsymbol{\omega'_B}) = Max\{\theta(\boldsymbol{\omega_A}), \theta(\boldsymbol{\omega'_A}), \theta(\boldsymbol{\omega_B}), \theta(\boldsymbol{\omega'_B})\}$$

Thus, the only motive for firms to ally is to access wider markets by diffusing a superior technology that otherwise the technologically inferior firm would not have obtained for its non-overlapping markets. I will refer to this benefit from collaboration as **Technological Diffusion**.

When technologies are modular, we observe an additional positive relationship between the knowledge base of partners and alliance formation. As the knowledge bases of partners become more common, a greater number of components in their current technological position are likely to be identically configured since their placements $\boldsymbol{\omega_A}, \boldsymbol{\omega_B}$ within the broader landscape represent the states of their accumulated knowledge. Therefore, by observing the results of each other's experimentation on the landscape, there is greater potential of observing the effect from adjusting two different components that were both in the same initial setting. When both component adjustments have positive effect on quality, we get a case where:

$$L(\boldsymbol{\omega_A}, \boldsymbol{\omega'_A}, \boldsymbol{\omega_B}, \boldsymbol{\omega'_B}) > Max\{\theta(\boldsymbol{\omega_A}), \theta(\boldsymbol{\omega'_A}), \theta(\boldsymbol{\omega_B}), \theta(\boldsymbol{\omega'_B})\}$$

This is because when both components are revealed to improve the initial state, each firm can apply the results from their partner in addition to benefiting from their own experimentation on the landscape. I will refer to the expected added benefit from the possibility of mutual learning as **Potential Complementarities**. This leads to the first set of hypotheses.[2]

Hypothesis 5.1 [H5.1]:

A. If Technological components are coupled, there is no relationship between common technological interests and alliance surplus value.

B. If Technological components are modular, greater common technological interests is positively related to higher alliance surplus value.

---

[2] Hypotheses 5.1-5.4 have formal propositions based on the model. The appendix provides corresponding propositions (respectively Prop. 3-6) along with their proofs.

5.2.3 Technological Exhaustion and Prospect of Outside Opportunities

It has been argued that firms are inclined to engage in local search due to familiarity and likelihood of finding incremental improvements (Cyert and March, 1963; Dosi, 1988). However, local search can lead to traps where the prospect of greater advances from distant regions elude the firm. From the perspective of technological landscapes, we can designate disjoint sets of components as local S_L and distant S_D. The expected improvement from experimenting with local versus distant components is given by the following:

$$E[max\{\theta(\boldsymbol{\omega}'), \theta(\boldsymbol{\omega})\} - \theta(\boldsymbol{\omega}) \mid j \in S_L]$$
$$= P_L(\theta(\boldsymbol{\omega}') > \theta(\boldsymbol{\omega})) \times E_L[\theta(\boldsymbol{\omega}') - \theta(\boldsymbol{\omega}) \mid \theta(\boldsymbol{\omega}') > \theta(\boldsymbol{\omega})]$$

$$E[max\{\theta(\boldsymbol{\omega}'), \theta(\boldsymbol{\omega})\} - \theta(\boldsymbol{\omega}) \mid j \in S_D]$$
$$= P_D(\theta(\boldsymbol{\omega}') > \theta(\boldsymbol{\omega})) \times E_D[\theta(\boldsymbol{\omega}') - \theta(\boldsymbol{\omega}) \mid \theta(\boldsymbol{\omega}') > \theta(\boldsymbol{\omega})]$$

The literature assumes that:

$$P_L(\theta(\boldsymbol{\omega}') > \theta(\boldsymbol{\omega})) > P_D(\theta(\boldsymbol{\omega}') > \theta(\boldsymbol{\omega}))$$

However, as the potential for improvement of familiar local (modular) components approach exhaustion, firms may begin to search distant external components. This occurs when:

$$\frac{E_D[\theta(\boldsymbol{\omega}') - \theta(\boldsymbol{\omega}) \mid \theta(\boldsymbol{\omega}') > \theta(\boldsymbol{\omega})]}{E_L[\theta(\boldsymbol{\omega}') - \theta(\boldsymbol{\omega}) \mid \theta(\boldsymbol{\omega}') > \theta(\boldsymbol{\omega})]} > \frac{P_L(\theta(\boldsymbol{\omega}') > \theta(\boldsymbol{\omega}))}{P_D(\theta(\boldsymbol{\omega}') > \theta(\boldsymbol{\omega}))}$$

The analogue for this result when applied to common technological interests between firms would suggest that when potential complementarities for common external technological components relative to common internal technological components are great enough, the effect of common external technological interests on alliance formation is higher than common internal technological interests. This can occur

when either the firm has exhausted technological improvements from repeated local search of familiar components or from the emergence of new components , bringing new opportunities to light for both potential partners.

Hypothesis 5.2 [H5.2]:

A. Given modular components, if potential complementarities from common external interests relative to common internal interests are high, common external interests affects alliance surplus value more positively than common internal interests.

B. Given modular components, if potential complementarities from common external interests relative to common internal interests are low, common external interests affects alliance surplus value less positively than common internal interests.

5.2.4   Trade-offs from Overlapping Technologies

It's reasonable to consider that as two firms become identical to each other in all technological components or knowledge base, their markets will tend to converge as well. Defining the technological interest overlap between two firms on the landscape as the following:

$$Overlap(\boldsymbol{\omega_A}, \boldsymbol{\omega_B}) = 1 - TechDist(\boldsymbol{\omega_A}, \boldsymbol{\omega_B}) = 1 - \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}[\omega_A^i \neq \omega_B^i]$$

If we assume that:

$$Overlap(\boldsymbol{\omega_A}, \boldsymbol{\omega_B}) \to 1 \text{ as } \frac{O(\boldsymbol{\beta_A}, \boldsymbol{\beta_B})}{NO(\boldsymbol{\beta_A}, \boldsymbol{\beta_B}) + NO(\boldsymbol{\beta_B}, \boldsymbol{\beta_A}) + O(\boldsymbol{\beta_A}, \boldsymbol{\beta_B})} \to 1$$

Then we obtain a relationship that says alliance surplus value is quadratic and may even have an inverted U-relationship with technological overlap. While the value

from potential complementarities increases with greater technological overlap, higher market overlap between firms means that the positive firm benefits from technological diffusion approaches zero and is replaced by the negative effects of competition between partners in more overlapping markets. Depending on the intensity of competition (as captured by ρ), the existence of an inverted U effect and its curvature may vary. In summary, this suggests a third empirical hypothesis:

Hypothesis 5.3 [H5.3]: All else equal, more non-overlapping technological domains of partners is positively related to alliance surplus value.

5.2.5 Firm Size and Joint Value Appropriation

Firm size has started to generate considerable attention as recent studies have noted the tendency of large companies to extract a greater share of value from partnering with smaller firms due to resource misappropriation (Katila et al, 2008; Li et al, 2008; Diestre et al, 2012) or abundance of outside partners (Mason and Drakeman, 2014). Here, I provide an alternative rationale for why larger firms can capture a greater share of the total joint value generated within an alliance by linking size of the firm's customer base to the concept of coordination costs.

Alliance coordination costs can be substantial since collaborations involve the decomposition of tasks across firm boundaries, the complexities of which require active inter-firm communication and decision making. This is compounded by the firm's need to put systems in place that restrict the spillover of information stemming from its non-alliance activities. Such costs can be saved and invested in new internal projects when the firm decides to develop and commercialize technologies by itself.

To account for this feature in the model, the private rival benefits of each partner will include an additional component, the *Value of New Interests (VONI)*, which summarize the firm's value of new internal projects that arise from alliance coordi-

nation cost savings. Moreover, the value of the firm's new interests will depend on its size **s**:

$$v_A = \lambda_A + VONI(s_A)$$

$$v_B = \lambda_B + VONI(s_B)$$

The economics of innovation literature has noted that firm size may be associated with the value that companies can capture from R&D projects. Large firms not only have more project options due to enhanced ability to absorb risk as classic Schumpeterian perspective has proclaimed, but they may also realize greater value from the same projects compared to relatively smaller firms. Similar to the idea of cost spreading (Cohen and Klepper, 1996), large firms can generate higher revenues from access to broader customer bases. This leads to the assumption that the value of new interests for each partner increases with respect to its own size.

$$\frac{dVONI}{ds} > 0 \text{ for each firm}$$

Given incomplete contracts of alliances (Hart and Moore, 1988), partners with higher opportunity costs are better positioned to negotiate a greater share of the joint alliance value. Since relatively large firms have more valuable outside options arising from new internal projects, such partners would command a greater share of the value generated within an alliance. From the model, a straightforward implication is that firm size would correspond to higher private rival benefit of partners, thus reducing joint alliance surplus.

Hypothesis 5.4 [H5.4]: All else equal, firm size of partners is negatively related to alliance surplus value.

5.3   Empirical Setting

5.3.1   Sample

The study is conducted in the specialized setting of an alliance market among software and communication firms from 1990 to 2005. Alliances were pulled from the SDC Platinum database with primary SIC classifications that correspond to Software (7372) and various forms of Communication Technology (3661, 3663, 3669, 4812, 4813, 4822, 4899). After eliminating pure licensing agreements and confining the set to dyads that consisted of at least 1 US partner to partially control for differences in US patenting activity between US and non-US firms, the final sample consisted of 180 alliances. The nationalities of non-US partner firms is shown in Figure 2, which make up only 21% of the sample.
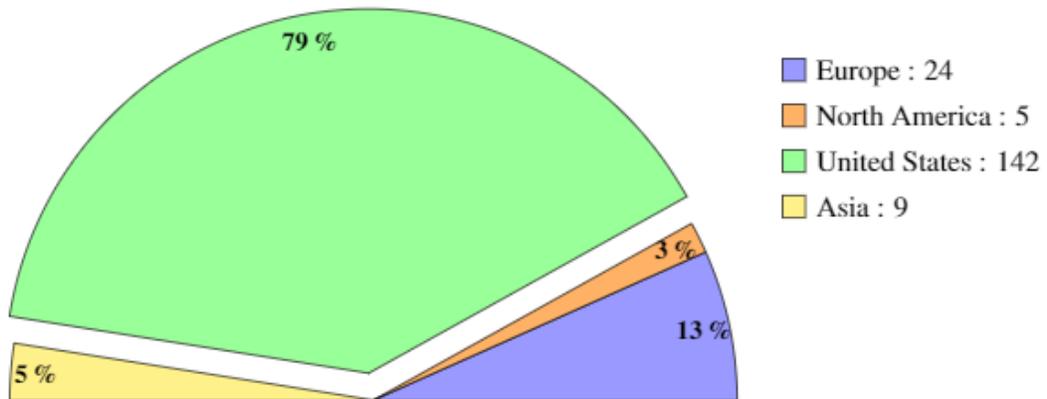
FIGURE 5.2: Nationality of the U.S. Firm's Alliance Partners (180 Total Alliances)

Alliances consisted of 103 software alliances, 26 communication alliances and 53 cross-industry alliances. The activities among alliances vary widely and include R&D, manufacturing, marketing, joint ventures, and supplier agreements. R&D and Marketing comprise the largest categories with each making up around 25%.

Two events during the sample period were of critical importance. First, the advent of the internet reached 3% market penetration in 1993 and was increasing exponentially onward (Rogers 1995), with numerous studies having identified this growth as a major technology shock (e.g. Alexopoulos, 2011). This is reflected in a spike of overall technology alliance activity as can be confirmed across multiple data sources (Schilling, 2015). The most pronounced effect can be seen in alliances among software firms, an industry that was also simultaneously propelled by the growth of the home and personal computing markets (Figure 3).
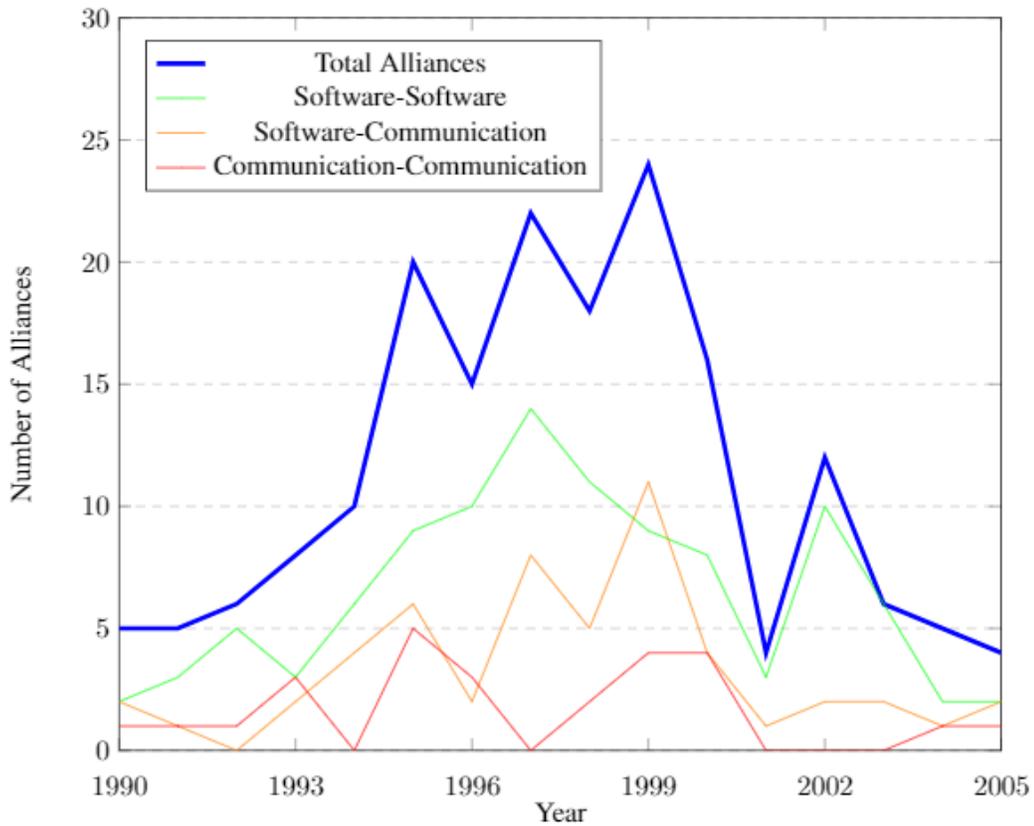
FIGURE 5.3: Alliances Among Software and ICT Firms, 1990-2005

The creation of business opportunities brought on by the internet was not confined to only software firms as more moderate jumps in inter-industry alliance ac-

tivity between software and communication firms is also observed. A second event occurred in a series of court cases between 1994 and 1995 that greatly strengthened intellectual property rights of software patents beginning in 1996 (Cockburn, 2011). This provided greater validity for the measures that were used to characterize the technological specializations of firms.

Firm patent information was initially matched using the NBER's Patent Data Project and checked manually to account for subsidiaries and ensure completeness of coverage. Characteristics of patents was then obtained from PATSTAT, while firm characteristics such as sales, employees and R&D were provided by Compustat and supplemented by Corptech for missing years.

5.3.2 Variables

Overlap in technological interests was measured using citation data on company patents. For a given alliance in year t, citations were only considered for patents up to year t-1 so that the overlap preceded alliance formation. Following the traditional conceptualizations of technological overlap in the literature (Mowery et al, 1996; Stuart and Podolny, 1996), I employ cross citation rates and common citation rates. The cross citation rate between Firm i and Firm j is given by:

$$Cross-Citation\ Rate(Firm_i, Firm_j)$$

$$= Citation\ Rate(Firm_i, Firm_j) + Citation\ Rate(Firm_j, Firm_i)$$

where $Citation\ Rate(Firm_i, Firm_j)$

$$= \frac{\text{Citations to Firm j's patents in Firm i's patents}}{\text{Total Citations in Firm i's patents}}$$

The cross-citation rate measures the relative importance of Firm j's patents relative to Firm i's technology portfolio and vice versa. An alternative form of technological overlap is given by the common citation rate and defined as follows:

$$Common\ Citation\ Rate(Firm_i, Firm_j)$$

$$= \frac{\text{Citations in Firm i's patents to patents cited by Firm j's patents}}{\text{Total Citations in Firm i's patents}}$$

$$+ \frac{\text{Citations in Firm j's patents to patents cited by Firm i's patents}}{\text{Total Citations in Firm j's patents}}$$

The common citation rate measures the extent to which two firms draw on the same knowledge pools. Since these knowledge pools are comprised of prior technologies which may or may not reside in the patent portfolios of partner firms, I define two additional forms of technological overlap in the following way.

$$\textit{External Citation Rate}(Firm_i, Firm_j)$$

$$= \frac{\text{Citations in Firm i's patents to external patents cited by Firm j's patents}}{\text{Total Citations in Firm i's patents}}$$

$$+ \frac{\text{Citations in Firm j's patents to external patents cited by Firm i's patents}}{\text{Total Citations in Firm j's patents}}$$

$$\textit{Internal Citation Rate}(Firm_i, Firm_j)$$

$$= \frac{\text{Citations in Firm i's patents to internal patents cited by Firm j's patents}}{\text{Total Citations in Firm i's patents}}$$

$$+ \frac{\text{Citations in Firm j's patents to internal patents cited by Firm i's patents}}{\text{Total Citations in Firm j's patents}}$$

External patents refer to those that are neither owned by Firm i nor Firm j and therefore represent common interests on technologies that neither firm is directly extracting value from. This is contrasted with internal patents, which are technologies that are owned by at least one partner and may potentially represent a source of market rents to the owner. A key difference is that while both Common Internal and Common External citations indicate overlapping interests in technological areas, if the overlap is centered on technologies that are internal to one of the partners, the owner may have greater knowledge of how to utilize the technology in the absence of an alliance, reducing its reliance on outside firms to add value.

I measure technological breadth as the number of distinct International Patent Classification (IPC) categories under which the firm's patents are organized. IPC categories represent distinct technological domains where the firm is present and often maps well into areas of product application[3] (Stembridge, 1998). Since this study has a focused industry context, I use the most detailed IPC level (Sub Group). As shown by descriptive statistics (Table 1), this has a range from 0 to 3770, providing adequate granularity for my construct.

Table 5.1: Descriptive Statistics

Variable	Measure	Mean	SD	Min	Max
Alliance (Dependent Variable)	Alliance Announcement	0.067	0.25	0	1
Common Technological Domains	# Common IPC Subgroups	13.965	61.6	0	837
Common Technological Interests	Cross Citation Rate	0.008	0.035	0	1
Common Technological Interests	Common Citation Rate	0.028	0.077	0	0.9
Common External Interests	Common External Citation Rate	0.024	0.061	0	0.8
Common Internal Interests	Common Internal Citation Rate	0.005	0.025	0	0.9
Proportion of External Interests	External Cite / Common Cite	0.307	0.439	0	1
Same Industry Control	Same 2-Digit SIC	0.499	0.5	0	1
Technological Domains	# of IPC Subgroups	397.415	679.449	0	3770
Firm Size	Annual Net Sales (Millions USD)	24,251	31,434.6	2.2	185,504
Alliance Experience	# of Previous Alliances	7.205	9.301	0	71
Number of Dyadic Obs: 2781					

Note:
1. IPC: International Patent Classification
2. SIC: Standard Industrial Classification

As control variables, I also construct the number of previous alliances for each firm-year observation as well as the number of overlapping IPC classes between potential partners. Table 2 presents the correlation matrix for all variables. Given the high correlation between common external and common internal citation rate along with the need to conceptually distinguish between them, I construct an alternative measure as the proportion of common citations that are external to both partners.

[3] An alternative system is the United States Patent Classification (USPC). In 1971, the United States Patent Office (USPTO) remarked the differences between USPC and IPC by stating: It is well recognized that the two systems are conceptually different; the U.S. system being based primarily on structure and function while the International Classification is primarily industry and profession oriented.

Table 5.2: Variable Correlation Matrix

Measure	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) Alliance (DV)	1										
(2) Common IPC Subgroups	0.034	1									
(3) Cross Citation Rate	0.076	0.187	1								
(4) Common Citation Rate	0.098	0.254	0.808	1							
(5) Common External Citation Rate	0.105	0.245	0.635	0.961	1						
(6) Common Internal Citation Rate	0.047	0.187	0.938	0.738	0.522	1					
(7) External Cite / Common Cite	0.105	0.255	0.216	0.442	0.492	0.162	1				
(8) Same 2-Digit SIC	0.077	-0.074	0.149	0.150	0.138	0.125	-0.068	1			
(9) # of IPC Subgroups	0.093	0.513	0.513	0.603	0.583	0.434	0.426	-0.068	1		
(10) Annual Net Sales	0.070	0.326	0.381	0.424	0.406	0.318	0.255	-0.216	0.668	1	
(11) # of Previous Alliances	0.089	0.425	0.435	0.556	0.551	0.370	0.462	0.097	0.690	0.551	1

Note:
1. IPC: International Patent Classification
2. SIC: Standard Industrial Classification

5.4   Methodology

To test hypothesis 1 and demonstrate methodological differences between binary choice and chosen matching estimator, my first set of specifications estimate the parameters to the following latent value function:

$$\pi_{i,j} = \beta_0 CommIPC_{i,j} + \beta_1 CommInterest_{i,j} + \beta_2 CommInterest_{i,j}^2 + \epsilon_{i,j}$$

Where $\pi_{i,j}$ is the match surplus generated from an alliance between Firm i and Firm j as defined previously and $CommInterest_{i,j}$ is measured through citation-rate construction. Like prior work (Yang, Shi, and Goldfarb, 2009), I estimate these parameters using maximum score (Manski, 1985). This has two principle benefits. First, the estimation relaxes the assumption on the specific distribution of the stochastic term $\epsilon_{i,j}$ which Bayesian techniques require (e.g. Sorensen, 2007) and may bias all parameters if the restrictive assumption does not hold. Secondly, the score estimation greatly reduces the curse of dimensionality, relieving the computational burden of evaluating a series of numerical integrals at each iteration.

However, unlike prior matching frameworks designed for contexts with central planning (Fox, 2010), I modify a Non-Transferrable Utility (NTU) matching estimator, which is more suitable for markets where decentralized agents form matches among themselves. The NTU matching estimation is aptly applied to alliance mar-

kets under Nash Bargaining, resulting in a fixed sharing rule governing the division of the surplus value generated between partners. This means that proposition 2 provides an estimation property to recover marginal effects on alliance surplus value. Applying this property to the set of all possible observations involving two disjoint alliances among 4 distinct firms, the estimator finds the parameter values that maximizes the number of criteria satisfied. Additionally, since the alliance market does not necessarily adhere to a two-sided separation of firms, I accommodate this by constructing two counterfactual sets of alliances for each alliance-pair observation, increasing the number of criteria and adding efficiency to the estimation.

To find the optimal values, I implemented a genetic evolution algorithm that is more suitable for optimization problems involving non-differentiable functions. The algorithm is particularly useful since it provides a random scatter of starting points that co-evolve over time with the highest (or lowest) point becoming a candidate estimate. The genetic evolution algorithm is reinforced using a pattern search algorithm to make sure that improvements cannot be obtained from neighboring points. Since we can never be certain that the starting point leading to the true global optimal value is covered, I iterated this procedure 20-100 times depending on the computational time of each iteration, which largely depended on the number of variables in the estimation. After obtaining estimates, hypothesis testing is then performed using the sub-sampling procedure outlined in Romano and Shaikh (2010).

For comparison to prior methods, I also construct a random sample from the set of all dyadic pairs among firms with at least 1 observed alliance to estimate the parameters using a binary logit. The random sample is constructed by taking the set of all alliances and matching a random counterfactual dyad to each. The counterfactual dyad must have the same year and be the same alliance type (i.e. software-software, software-communication, or communication-communication). Drawing pairs from the set of dyadic observations has the advantage of pairing firms known to be active

in the alliance market while matching on alliance type resolves an earlier issue of pairing firms in unrelated industries with low citation rates for reasons irrelevant to collaboration or non-collaboration (Mowery et al., 1998).

After comparing the baseline matching model with the binary logit, I test hypotheses 2-4 by estimating the full model based on the following specification:

$$\pi_{i,j} = \beta_0 CommIPC_{i,j} + \beta_1 CommInterest_{i,j} + \beta_2 Proportion\_ExternalInterest_{i,j}$$
$$+ \beta_3(numIPC_i + numIPC_j) + \beta_4(Size_i + Size_j)$$
$$+ \beta_5(PrevAlliances_i + PrevAlliances_j) + \epsilon_{i,j}$$

Where $Proportion\_ExternalInterest_{i,j}$ is the proportion of the common pool of technological interests that neither partner owns. $numIPC_i$ is the number of IPC Subgroups that Firm i has presence and is measured by a 4-year moving window ending in the year prior to the alliance, which is long enough to attenuate fluctuations and short enough to capture recent and relevant knowledge (Diestre at al. 2012; Rothaermel et al., 2008). Firm size is measured by net sales and $PrevAlliances$ controls for alliance experience, measured by the number of prior alliances for each firm since the beginning of the sample.

In addition, I also control for alliance type by further refining the matching estimation to use a more focused selection of criteria in the following way. For two observed intra-industry firm alliances (e.g. software-software) that involve disjoint pairs of firms, I include all the criteria in the score estimator if both alliances are within the same industry. Alternatively, for two observed inter-industry firm alliances (i.e. software-communications) that involve disjoint pairs of firms, I only consider the criteria for the counterfactual set that maintains the alliance types. All other criteria are excluded. This has the benefit of constructing more realistic counterfactual alliance pairs for comparison while demonstrating the flexibility of the matching methodology. Since the distinction between software and communication

firms is based on the 2-digit SIC classification, I refer to this modified estimation as SIC-Adjusted Matching.

5.5 Results

Table 3 presents results for the baseline model of alliance formation using different measurements for overlapping technological interests. We see that the matching model estimates support hypothesis 1 across all 4 measures of overlapping technological interests. When compared with the binary choice model, the matching estimates only exhibit a statistically significant positive linear effect across all forms of common technological interest whereas the binary logit suggests a curvilinear effect in the form of an inverted U. To reconcile the findings, I examine the reasoning behind the significant negative coefficient of the quadratic term under a binary logit. I also develop insight on why a logit model may produce a biased negative estimate in the absence of a true negative effect and how the matching model corrects for it.

A fundamental difference between the two methodological approaches concerns the type of counterfactuals that are being compared to the observed outcomes. The matching model treats counterfactuals as a *pair* of alliance outcomes across a space of 4 dyads so that when counterfactual pairings are constructed, none of the firms exceed their original observed capacity. This is convenient since it allows the method to accommodate conditions when capacity constrained firms are present, which oftentimes is unknown to the researcher. On the other hand, binary logit models assume that all firms are unconstrained. If this condition is true, then an alliance between each firm pair can be treated as an independent decision provided both partners agree on the choice to collaborate.

Table 4 shows the proportion of alliances within each citation quantile. We see that this proportion goes up with the level of citations, before declining at the highest quantiles. From the binary logit models, this means a curvilinear effect of common

Table 5.3: Overlap In Technological Interests And Alliance Formation

Specification	(1)	(2)	(3)	(4)
Tech Interest Measure:	Cross Citations	Common Citations	External Com. Citations	Internal Com. Citations
Binary Logit:				
Common Tech. Domains	-2.632	-3.003	-1.891	-2.708
	(4.230)	(4.140)	(4.082)	(4.335)
Common Tech. Interests (H1 +)	27.344**	10.825**	10.307**	43.714**
	(9.531)	(3.429)	(3.813)	(16.265)
Common Tech. Interests^2	-142.884**	-22.981**	-22.05*	-395.197**
	(54.654)	(9.027)	(12.038)	(155.595)
Constant	-0.112	-0.197	-0.194	-0.074
	(0.115)	(0.121)	(0.122)	(0.114)
Matching Estimation:				
Common Tech. Domains	1	1	1	1**
Common Tech. Interests (H1 +)	92.43**	437.487**	1080.262**	4999.788**
Common Tech. Interests^2	-817.775	-1220.053	-6064.253	-54765.914
Obs (# of Alliances)	180	180	180	180

** $p < 0.05$, * $p < 0.1$

For Binary Logit, Statistical Inference Based on Two-Tailed Test with Standard Errors Given in Parenthesis.
For Matching Model Estimation, Statistical Inference Based on Sub-Sampling.
Notes:
1. Matching estimator uses maximum score, requiring a normalization coefficient (set to coefficient on common tech. domains)
2. Hypothesis testing on matching model normalization coefficient is based on null H_0: $\beta < 0$; all other coefficients are based on null H_0: $\beta = 0$

technological interests on alliance formation. If all firms were unconstrained, then observing lower proportions of alliance formation at the high end implies that a mid-level of overlapping technological interests is most favorable.

On the other hand, if firms were constrained and the alliances we observe reflect the maximum number of alliances that each firm can sustain, then it's possible we may see lower proportions of alliance formation at high levels of overlap even if only a positive effect is present. Under the present matching framework used, the equilibrium of alliances is generated by pairing up the alliances in sequence of highest value until capacities of each firm are reached. If this is the case, then the high proportions of non-allied firms at the highest levels of overlap are symptomatic of alliances that could not have been formed due to capacity limitations and thus should not have been included as an observation under a binary choice model.

To test for this possibility, I re-examined the dataset of all dyadic pairs and the

Table 5.4: Relative Proportions of Alliances Across Common Technological Interest Quantiles

Quantiles	Cross Citations	Common Citations	External Common Citations	Internal Common Citations
1-6/7/8/9	0.4375	0.4095	0.4123	0.4464
	(256)	(210)	(211)	(280)
7		0.5714	0.5	
		(7)	(6)	
8		0.4516	0.5161	
		(31)	(31)	
9	0.5652	0.5806	0.5161	
	(23)	(31)	(31)	
10	0.7097	0.8065	0.8064	0.7333
	(31)	(31)	(31)	(30)
11	0.7419	0.7097	0.6129	0.7419
	(31)	(31)	(31)	(31)
12	0.5483	0.5806	0.6774	0.5483
	(31)	(31)	(31)	(31)

Number of Obs. in Parenthesis
Notes:
1. Binary choice estimates from table 3 imply inverted U-shaped effect of common tech. interests on alliance formation
2. Estimates from binary choice are partly due to non-monotonic increase in proportion of alliance formations across quantiles

relationship between the alliance pair citation rate and the mean counter-factual citation rate for each partner firm within the same year and alliance type. If citation rates were independent and randomly drawn, then the relationship between overlap of allied firms and counterfactuals should be minimal. However, if high overlap firms eliminate high overlapping potential alliances due to capacity limitations, then we should see a positive relationship between the citation rates of allied firm-pairs and the counterfactuals of the partners. After de-meaning the values of the citation rates, I ran an OLS (without a constant) to test the strength of the relationship and the results are shown in Table 5. There is mostly a positive relationship between an alliance pair's overlapping interests and the overlap interests of its counterfactuals across all citation categories. The only insignificant relationship comes from alliances within the communications industry, which comprises less than 15% of all alliances.

Next, I test hypotheses 2-4 by estimating the full model using the refined SIC-adjusted matching approach to account for intra-industry and inter-industry alliance types. Under hypothesis 2, external common interests would increase alliance value more than internal common interests. Therefore, using common citation rate as a

112

Table 5.5: Testing The Relationship Between Common Tech. Interest of Allied Pairs and Counterfactuals (OLS)

	Cross Citations	Common Citations	External Common Citations	Internal Common Citations
All Alliances	0.4299***	0.5349***	0.5363***	0.4387***
	(0.0891)	(0.0926)	(0.1015)	(0.0759)
Obs.	369	369	369	369
Software-Software	0.4289***	0.5257***	0.5403***	0.4388***
	(0.1026)	(0.1112)	(0.1192)	(0.0906)
Obs.	212	212	212	212
Software-Comm	0.9429***	0.8402**	0.7645**	0.7952**
	(0.2912)	(0.3237)	(0.3478)	(0.3423)
Obs.	105	105	105	105
Comm-Comm	-0.2685	0.315	0.1554	0.0641
	(0.7960)	(0.2945)	(0.3035)	(0.4459)
Obs.	52	52	52	52

Standard Errors Given in Parenthesis. *** p <0.01, ** p <0.05, * p <0.1 (Based on Two-Tailed Test)
Notes:
1. Binary choice estimates can be biased due to inclusion of invalid counterfactual alliance pairs between capacity constrained firms
2. Positive relationship between common tech. interest of allied firms and their counterfactuals suggest that the decrease in proportion of alliances across the highest common tech. interest quantiles (Table 4) is due to inclusion of such invalid counterfactuals, hence resulting in an implied inverted-U effect from binary choice estimation

measure of general common interest in the baseline model, I add in the proportion of external common citations to test the second hypothesis. To test hypotheses 3 and 4, I include technological breadth and firm size measures as individual characteristics. A significant positive coefficient on the individual characteristics reflects a greater potential to add value (i.e. surplus or π) within an alliance as a function of the partners' own attributes while a negative coefficient may potentially indicate a partner's tendency to lower surplus based on the availability of its outside options.

Table 6 presents the results of the full model, which also includes the binary logit estimates for comparison. We first note that the SIC-adjusted matching has similar baseline results, continuing to support hypothesis 1. In the second specification, the decomposition of overlap in technological interests reveals that most of the value resides in external common interests, in line with hypothesis 2. Adding in the individual characteristics, we see a positive effect of technological breadth and a negative effect of firm size, conforming to hypotheses 3 and 4. Since all specifications include

a control for common IPC classes between partners, the individual firm number of IPC classes is more reflective of each partner's contribution of its own breadth to the alliance, which is more reflective of the theory and adds confidence to hypothesis 3. The negative effect of individual firm size may indicate a lower alliance surplus due to either lower joint value creation or higher partner value appropriation. I do not detect size complementarity in alternative specifications, suggesting the latter case and supporting the theoretical foundations of hypothesis 4.

Table 5.6: Technological Specialization and Alliance Formation (Full Model Results)

Specification:	(1)	(2)	(3)
SIC Adj. Matching (Full Model)			
Common Tech. Domains	-1	1	-1
Common Tech. Interests (H1 +)	0.45**	-22.698	-0.192
Prop. Common External Interests (H2 +)		10.551**	0.063**
Technological Domains (H3 +)			0.748**
Firm Size (H4 -)			-0.863**
Alliance Experience			0.031
Binary Logit (Full Model)			
Common Tech. Domains	1	-1	-1
	(9.029)	(2.191)	(0.893)
Common Tech. Interests (H1 +)	5.217*	0.228	-0.239
	(2.798)	(0.724)	(0.319)
Prop. Common External Interests (H2 +)		0.52**	0.179**
		(0.155)	(0.056)
Technological Domains (H3 +)			0.046
			(0.134)
Firm Size (H4 -)			0.129
			(0.104)
Alliance Experience			0.039
			(0.103)
Obs. (# of Alliances)	180	180	180

** $p < 0.05$, * $p < 0.1$

I revisit the binary logit for comparison. Note that while hypotheses 1 and 2

114

are also supported, the magnitudes are very different from the matching estimation. This could be symptomatic of the high variability in the coefficient estimate for the reference variable of common IPC classes between partners from which all other coefficients are normalized. A marked difference between the two methods is shown in the coefficients on individual characteristics. The matching estimation is better able to detect the effects of firm breadth and size whereas the binary logit does not find evidence of hypotheses 3 and 4 due to potential biases in the methodology.

To draw a final comparison of the two methods, Table 7 shows the predictive power of both models using all specifications of Table 6. We see that the binary logit predicts more alliances that were observed in the data, but this is misleading since it also predicts many incorrect alliances as well. As a base comparison, if every possible dyad were predicted to be an alliance, the accuracy would be around 6.5%. The logit predicts the outcome with some improvements in accuracy. The matching model is roughly twice as accurate as the logit and the SIC-adjusted matching model makes a substantive improvement in prediction over the original matching model.

Table 5.7: Prediction Results Comparison Across Model Specifications

	Total Predicted	Correct Alliances	% Correct
Specification 1: Common Tech. Interests			
Binary Logit	435	63	14.48
Matching	177	37	20.90
SIC Adj. Matching	175	59	33.71
Specification 2: Including Common External Tech. Interests			
Binary Logit	877	95	10.83
Matching	174	39	22.41
SIC Adj. Matching	175	50	28.57
Specification 3: Full Model with Individual Characteristics			
Binary Logit	946	100	10.57
Matching	167	40	23.95
SIC Adj. Matching	173	58	33.53

Null Model Prediction: 6.5% of all dyadic pairings between firms in the data resulted in an alliance.
Notes:
1. Binary choice over-predicts alliance formation among non-allied dyadic pairs and is about twice as accurate as a null model
2. Matching models predict less actual alliances, but are more accurate than binary choice

5.6 Discussion

This paper set out to decouple technological interests and technological domains in examining the conditions that lead to alliance formation. From the theory and empirical analysis, I show that the benefits of technological overlap are mainly exhibited through common technological interests of partners. Meanwhile, the firm's tendency to seek partners that occupy new technological domains also addresses the resource redundancy drawbacks of too much overlap. From a methodological standpoint, treating alliances within the perspective of a matching framework that accounts for scarcity of partners and incomplete collaboration contracts is consequential as the pattern of results differ from traditional binary choice models, which may lead to misleading findings such as an inverted-U effect of overlapping technological interests on inferred alliance value.

Although all forms of overlapping technological interests did not exhibit eventual negative returns under the matching estimation, diminishing returns may still exist. From the theory, diminishing returns is possible when the technological interests of partners converge. This is because as the compositions of firm technologies become identical, so would their product markets. Thus the benefits arising from diffusing technologies to new markets would be replaced with competition for customers within common markets. An added consideration is that the rate of alliance knowledge production can potentially diminish at higher levels of common technological interests in the presence of increasing inter-firm coordination costs across wider pools of relevant technologies.

Decomposing common technological interests, the results also show that external overlapping interests between firms contribute more to alliance surplus than internal overlapping interests. Thus, aggregate measures of general overlapping interest such as common citations can erroneously pick up diminishing returns to alliance surplus

if the decline in proportions of external overlapping interests is sufficiently stark. In the data, greater levels of general overlapping interests are only moderately related to lower proportions of external common interests, decreasing from 1 to 0.82 at the highest quantile.

By considering the various forms of overlapping technological interests, this study showed how the form of overlap may help distinguish between external and internal common interests. This is important for empirical work that focuses on the locus of firm technological interests as it may lead to erroneous interpretations of findings when different forms of overlap (e.g. cross citations versus common citations) are used interchangeably. The forms of overlap considered here are centered around citation measures. Other forms of overlap may exist as measures become more fine-grained. Developments in text analysis enable rich maps of a firm's technological position (Gerken and Moehrle, 2012). As these measures evolve, so must the constructs to keep future scholars grounded on what characteristics are being detected in the data.

From the full model results, we also see that individual effects are not detected under a binary logit. As noted earlier, this is fundamentally a symptom of invalid observations in the dataset, namely the non-alliances of constructed firm-pairs. Qualitatively, the results from the binary logit do not change whether it is performed on the set of matched pairs or the full set of dyadic pairs. A potential interesting null result is the finding that previous alliance experience is unrelated to alliance value. This may be due to issues of sample selection both at the firm level and the alliance level. At the firm level, all companies were public at some point in time, which meant they may have all accumulated experience from alliances, thus reducing the variation. This is less of a concern since I observe these firms at different points in time throughout the sample period. At the alliance level, experience was measured based on total ties within the software-ICT firm context, but can also potentially be reflected in total overall ties beyond the restricted empirical setting considered here.

The analytical comparisons made in this paper question the validity of the methodological tools that are used to gain insight from observations. Binary choice models, though widely applicable in various facets of empirical work, carry a specific set of conditions that does not necessarily hold when analyzing interactive dynamics of firms. Biases in estimation, as shown in the results, not only detect false effects, but may also hide true effects. Future development and use of proper methodologies should not be done in isolation from other areas of research. Instead, methods should speak to theoretical insight by considering how the data is generated and what features are present but unobservable that may render certain forms of analyses invalid.

The class of NTU matching models employed in this paper addressed the presence of capacity constrained firms but was also purposefully chosen to reflect the dynamics of alliance formation. Unlike transferable utility (TU) matching models which seek to maximize the aggregate value of the market, the NTU model with fixed sharing rule on the alliance surplus results in an equilibrium where the highest valued alliances are formed in order until either all positive-valued alliances are formed or capacity constraints are reached. This is more realistic when firms are atomistic and coordination on partner selection does not exist. Most importantly, selecting the highest valued alliances in sequence requires taking individual firm characteristics into consideration, enabling estimation of private value capture. By contrast, TU models neglect individual effects since they play no role in adding or reducing aggregate value within the market.

6

Conclusion

The preceding chapters present a set of studies that address how firms gain competitive advantage in dynamic product markets. This is particularly relevant for industries characterized by rapid innovation such as mobile devices, information communication technologies and prepackaged software. Such industries require firms to take into account the actions of other players when mapping out their future direction of innovation. In this respect, firms may see the presence of other companies as a competitive threat when racing to incorporate new technologies into products or as potential collaborators who may provide complementary knowledge through joint development and production.

The first empirical study (chapter 2) investigates the first case in the form of innovation timing and market entry. I provide a novel conceptual and empirical approach to illuminate attention patterns in the organization and link it to the introduction of new technologies following a pioneer. I highlight the benefits of both shared attention to rivals and dispersed attention across the vast variety of industry topics that are present. In doing so, I show that attentional similarity and dispersion both have important implications for technology strategy and that attention may

serve as a critical managerial cognitive capability to improve speed of technological commercialization as well as gain recovery from launching failed technologies. These findings offer a base for subsequent research such as examining how attentional patterns emerge within firms and understanding how such the direction of attention across firms may converge or diverge over time.

The second empirical study (chapter 5) addresses the latter case when firms view one another as potential collaborators. Acknowledging that firm technological characteristics and partner selection represent important topics in the alliance literature, I show that joint value creation and division are contingent upon them. I define technological interests of partners as a way to conceptualize each party's knowledge base where the firm is active in ongoing research. Next, I characterize firm technological domains as product areas where new knowledge generated within an alliance can be commercialized. Utilizing a new method (developed in chapters 3 & 4) that is appropriate for a setting where firms compete for limited collaboration opportunities, I provide evidence that partners with common technological interests and disparate technological domains generate higher alliance value from greater joint knowledge production and its wider application across various product areas.

Additionally, I decompose common interest into technologies that are patented by partners within a focal alliance and those that reside outside the alliance. Under this distinction, I discover that alliances generate more added value when common interests of partners reside in technologies that neither party owns, often requiring the combined efforts of both firms to realize its commercial value. Lastly, the nature of incomplete contracts combined with opportunity costs of partners enable relatively large firms and those with more distinct technological domains to appropriate a greater share of the joint alliance value. Since these firms have higher-valued exit options in the event of alliance dissolution, they are less reliant on partners, giving them relatively greater bargaining power for dividing up the joint value produced

within a collaboration.

This raises several questions for future research. For example, how do other technological characteristics such as the depth of a firm's specialization in certain areas affect its tendency for collaboration? This is not obvious since having a highly focused area of technological expertise may limit the firm's reception to outside knowledge or enable it to uncover greater potential complementarities with partners in other areas of specialization. Another promising body of work with methodological implications and future empirical applications can investigate how partner opportunity costs may change over time once a collaboration has been initiated but not completed. Under this condition, potential partners may initiate alliances with the intention of dissolving them prematurely since the firm's ability to generate value by itself depends on such prior engagements.

Appendix A

Formal Proofs of Propositions

This subsection provides a formal treatment of the theoretical results from chapter 5. The first part addresses the propositions from the alliance formation section (5.1.2). In the previous chapter, it is straightforward to verify that propositions 1 and 2 imply each other based on the model setup. Therefore, only the proof for proposition 2 is demonstrated. In the second part, the propositions corresponding to the hypotheses from section (5.2) are proven using the established model.

A.1   Matching Equilibrium of Alliance Formation

To conceptualize alliance formation within a matching framework, I decided to adopt a Non-Transferable Utility (NTU) setup. The idea is that since the value within alliances are not realized upfront, but may take months or years before such inter-firm agreements get completed, the utility or values that each partner obtains cannot be committed or shared across alliances at the time of formation. Letting V_{ij} denote the total value that an alliance between firm i and firm j would generate and v_i, v_j respectively denote the values that firm i and firm j would privately obtain in the

absence of an alliance, I define the match surplus as:

$$\pi_{ij} = V_{ij} - v_i - v_j$$

Next, in the event of alliance formation, let the values that each firm privately captures be defined as:

$$\pi_{ij}^i = v_i + \sigma_{ij}\pi_{ij}$$

$$\pi_{ij}^j = v_j + (1 - \sigma_{ij})\pi_{ij}$$

The above equations state that when firms i and j form an alliance, they privately capture the values that is equal to their opportunity cost plus a share of the surplus or added value from alliance formation. In accordance with the NTU Matching framework, I define an equilibrium based on two criteria.

I say that a matching outcome is in **equilibrium** if:

1. Individual Rationality: For an observed match between firms A and B, the following conditions must hold:

   - $\pi_{AB}^A > v_A$

   - $\pi_{AB}^B > v_B$

2. No Blocking: For two observed matches between firms {A,B} and firms {C,D}, each of the following conditions must not hold:

   - $\pi_{AC}^A > \pi_{AB}^A$ and $\pi_{AC}^C > \pi_{CD}^C$

   - $\pi_{AD}^A > \pi_{AB}^A$ and $\pi_{AD}^D > \pi_{CD}^D$

   - $\pi_{BC}^B > \pi_{AB}^B$ and $\pi_{BC}^C > \pi_{CD}^C$

   - $\pi_{BD}^B > \pi_{AB}^B$ and $\pi_{BD}^D > \pi_{CD}^D$

The Individual Rationality Condition states that it is in the best interest of each firm in an observed alliance to stay in the alliance than to exit the alliance. The No Blocking Condition states that for two observed alliances among 4 distinct firms, there is no incentive for any two firms to forgo their current partnership and form a new alliance.

Now I provide a necessary and sufficient assumption for Proposition 2 in the paper to hold.

Assumption (Surplus-Private Value Alignment). *The divisions of surplus within potential alliances are allocated in a way such that if $\pi_{AB} > \pi_{CD}$ then $\pi_{AB}^A - v_A, \pi_{AB}^B - v_B > \pi_{CD}^C - v_C, \pi_{CD}^D - v_D$*

This assumption states that if the total surplus from a potential alliance is greater than the total surplus from another potential alliance, then each partner in the higher surplus alliance will privately capture greater added value than the partners in the lower surplus alliance. Thus, even though utility is not transferable across alliances, it is divided within the alliances in such a way that firms in the alliance formation stage will select alliances with the highest total alliance level surplus. In the paper, a special case of the above assumption is adopted, namely that the surplus division between partners is 1/2, as determined by the Nash Bargaining solution. The following proposition now follows.

Proposition (Equilibrium Characterization). *Given the above assumption, when two alliances between firms {A,B} and firms {C,D} are observed. The following conditions must be true:*

1. *$max\{\pi_{AB}, \pi_{CD}\} > max\{\pi_{AC}, \pi_{BD}\}$*

2. *$max\{\pi_{AB}, \pi_{CD}\} > max\{\pi_{AD}, \pi_{BC}\}$*

Proof. It suffices to prove case 1, since case 2 will follow by the same argument. The proof is given by contradiction. Suppose the condition does not hold and without loss of generality, let:

$$\pi_{AC} = max\{\pi_{AC}, \pi_{BD}\}$$

Then, we have:

$$\pi_{AC} > \pi_{AB}, \pi_{CD}$$

From the assumption given above, this also implies that:

$$\pi_{AC}^A - v_A > \pi_{AB}^A - v_A$$

and

$$\pi_{AC}^C - v_C > \pi_{CD}^C - v_C$$

But this violates the no-blocking condition since: $\pi_{AC}^A > \pi_{AB}^A$ and $\pi_{AC}^C > \pi_{CD}^C$ □

A.2 Alliance Value Creation and Appropriation

Here, I state the propositions and provide proofs based on the established model of chapter-section (5.2). The propositions will differ slightly from the hypotheses to reflect more precise statements.

To simplify notation, I re-write the value of technological configurations as knowledge production functions:

$$K_{AB}(\tau) = L(\boldsymbol{\omega_A}, \boldsymbol{\omega'_A}, \boldsymbol{\omega_B}, \boldsymbol{\omega'_B})$$

$$K_A = max\{\theta(\boldsymbol{\omega'_A}), \theta(\boldsymbol{\omega_A})\}$$

$$K_B = max\{\theta(\boldsymbol{\omega'_B}), \theta(\boldsymbol{\omega_B})\}$$

The knowledge production terms $K_{AB}(\cdot)$, K_A, and K_B respectively represent the quality of technological development from an alliance between firms A and B, and separately for firm A and firm B in the absence of an alliance. The input term τ is

a natural number that indicates the number of common component configurations between both firms. I can further make an innocuous simplification by restricting the analysis to the case of no overlapping markets between firms A and B. The value functions can now be re-written as:

$$V_{AB} = [||\boldsymbol{\beta_A}||_1 + ||\boldsymbol{\beta_B}||_1]K_{AB}(\tau)$$

$$v_A = ||\boldsymbol{\beta_A}||_1 K_A + VONI(s_A)$$

$$v_B = ||\boldsymbol{\beta_B}||_1 K_B + VONI(s_B)$$

Proposition 3. *Given the alliance surplus* $\pi = V_{AB} - v_A - v_B$,

A. *If technological components are coupled, then* $\frac{d\pi}{d\tau} = 0$

B. *If technological components are modular, then* $\frac{d\pi}{d\tau} > 0$

Proof. Note that $\frac{dK_A}{d\tau} = \frac{dK_B}{d\tau} = 0$ by construction. When components are coupled, recall that:

$$K_{AB}(\tau) = Max\{\theta(\boldsymbol{\omega_A}), \theta(\boldsymbol{\omega'_A}), \theta(\boldsymbol{\omega_B}), \theta(\boldsymbol{\omega'_B})\}$$

Thus, $\frac{dK_{AB}}{d\tau} = 0$. On the other hand, when components are modular,

$$K_{AB}(\tau) = Max\{\theta(\boldsymbol{\omega_A}), \theta(\boldsymbol{\omega'_A}), \theta(\boldsymbol{\omega_B}), \theta(\boldsymbol{\omega'_B})\} \text{ for } \tau = 0$$

$$K_{AB}(\tau) = Max\{\theta(\boldsymbol{\omega_A}), \theta(\boldsymbol{\omega'_A}), \theta(\boldsymbol{\omega_B}), \theta(\boldsymbol{\omega'_B})\}$$

$$+ \frac{\tau(\tau - 1)}{N^2} E[max\{\theta(\boldsymbol{\omega'}), \theta(\boldsymbol{\omega})\} - \theta(\boldsymbol{\omega})] \text{ for } \tau \geqslant 1$$

The combined knowledge production given modular components is equal to the maximum of observed configuration values plus the chance that both firms selected different components within their common initial configured set times the expected learning from incremental component adjustment. The marginal effect of more common components is given by:

$$\frac{dK_{AB}}{d\tau} = \frac{2\tau - 1}{N^2} E[max\{\theta(\boldsymbol{\omega'}), \theta(\boldsymbol{\omega})\} - \theta(\boldsymbol{\omega})] > 0 \text{ for } \tau \geqslant 1$$

This leads to the following result:

$$\frac{d\pi}{d\tau} = [||\boldsymbol{\beta_A}||_1 + ||\boldsymbol{\beta_B}||_1]\frac{dK_{AB}}{d\tau} = 0 \text{ when components are coupled}$$

$$\frac{d\pi}{d\tau} = [||\boldsymbol{\beta_A}||_1 + ||\boldsymbol{\beta_B}||_1]\frac{dK_{AB}}{d\tau} > 0 \text{ when components are modular}$$

$\square$

Before stating the next proposition, I modify the alliance knowledge production function to take two inputs τ_L and τ_D that respectively represent the number of commonly configured local and distant components. In addition, for convenient simplification, I consider only the extreme cases where the configurations of firms A and B can have only common local components or only common distant components, allowing the joint knowledge production function to be written as:

$$K_{AB}(\tau_L, 0) = Max\{\theta(\boldsymbol{\omega_A}), \theta(\boldsymbol{\omega'_A}), \theta(\boldsymbol{\omega_B}), \theta(\boldsymbol{\omega'_B})\}$$

$$+ \frac{\tau_L(\tau_L - 1)}{N^2}E[max\{\theta(\boldsymbol{\omega'}), \theta(\boldsymbol{\omega})\} - \theta(\boldsymbol{\omega}) \mid j \in S_L] \text{ for } \tau_L \geqslant 1$$

$$K_{AB}(0, \tau_D) = Max\{\theta(\boldsymbol{\omega_A}), \theta(\boldsymbol{\omega'_A}), \theta(\boldsymbol{\omega_B}), \theta(\boldsymbol{\omega'_B})\}$$

$$+ \frac{\tau_D(\tau_D - 1)}{N^2}E[max\{\theta(\boldsymbol{\omega'}), \theta(\boldsymbol{\omega})\} - \theta(\boldsymbol{\omega}) \mid j \in S_D] \text{ for } \tau_D \geqslant 1$$

Proposition 4. *Given modular technological components, there exist values γ_L, γ_H such that:,*

A. $\frac{E_D[\theta(\boldsymbol{\omega'})-\theta(\boldsymbol{\omega}) \mid \theta(\boldsymbol{\omega'})>\theta(\boldsymbol{\omega})]}{E_L[\theta(\boldsymbol{\omega'})-\theta(\boldsymbol{\omega}) \mid \theta(\boldsymbol{\omega'})>\theta(\boldsymbol{\omega})]} > \gamma_H$ *implies* $\frac{d\pi}{d\tau_D} > \frac{d\pi}{d\tau_L}$ *for all values of $\gamma_L, \gamma_H \geqslant 1$*

B. $\frac{E_D[\theta(\boldsymbol{\omega'})-\theta(\boldsymbol{\omega}) \mid \theta(\boldsymbol{\omega'})>\theta(\boldsymbol{\omega})]}{E_L[\theta(\boldsymbol{\omega'})-\theta(\boldsymbol{\omega}) \mid \theta(\boldsymbol{\omega'})>\theta(\boldsymbol{\omega})]} < \gamma_L$ *implies* $\frac{d\pi}{d\tau_D} < \frac{d\pi}{d\tau_L}$ *for all values of $\gamma_L, \gamma_H \geqslant 1$*

Proof. From the prior result, we have:

$$\frac{d\pi}{d\tau_L} = [||\boldsymbol{\beta_A}||_1 + ||\boldsymbol{\beta_B}||_1]\frac{2\tau_L - 1}{N^2}E[max\{\theta(\boldsymbol{\omega'}), \theta(\boldsymbol{\omega})\} - \theta(\boldsymbol{\omega}) \mid j \in S_L]$$

$$\frac{d\pi}{d\tau_D} = [||\boldsymbol{\beta_A}||_1 + ||\boldsymbol{\beta_B}||_1]\frac{2\tau_D - 1}{N^2}E[max\{\theta(\boldsymbol{\omega'}), \theta(\boldsymbol{\omega})\} - \theta(\boldsymbol{\omega}) \mid j \in S_D]$$

Recall that:

$$E[max\{\theta(\boldsymbol{\omega'}), \theta(\boldsymbol{\omega})\} - \theta(\boldsymbol{\omega}) \mid j \in S_L]$$

$$= P_L(\theta(\boldsymbol{\omega'}) > \theta(\boldsymbol{\omega})) \times E_L[\theta(\boldsymbol{\omega'}) - \theta(\boldsymbol{\omega}) \mid \theta(\boldsymbol{\omega'}) > \theta(\boldsymbol{\omega})]$$

$$E[max\{\theta(\boldsymbol{\omega'}), \theta(\boldsymbol{\omega})\} - \theta(\boldsymbol{\omega}) \mid j \in S_D]$$

$$= P_D(\theta(\boldsymbol{\omega'}) > \theta(\boldsymbol{\omega})) \times E_D[\theta(\boldsymbol{\omega'}) - \theta(\boldsymbol{\omega}) \mid \theta(\boldsymbol{\omega'}) > \theta(\boldsymbol{\omega})]$$

Rearranging the terms, we see:

$$\frac{d\pi}{d\tau_D} \gtrless \frac{d\pi}{d\tau_L} \text{ if and only if}$$

$$\frac{E_D[\theta(\boldsymbol{\omega'}) - \theta(\boldsymbol{\omega}) \mid \theta(\boldsymbol{\omega'}) > \theta(\boldsymbol{\omega})]}{E_L[\theta(\boldsymbol{\omega'}) - \theta(\boldsymbol{\omega}) \mid \theta(\boldsymbol{\omega'}) > \theta(\boldsymbol{\omega})]} \gtrless \frac{P_L(\theta(\boldsymbol{\omega'}) > \theta(\boldsymbol{\omega}))}{P_D(\theta(\boldsymbol{\omega'}) > \theta(\boldsymbol{\omega}))}\left[\frac{2\tau_L - 1}{2\tau_D - 1}\right]$$

Since $1 \leqslant \tau_L, \tau_D < N$, the proposition follows by setting:

$$\gamma_H = \frac{P_L(\theta(\boldsymbol{\omega'}) > \theta(\boldsymbol{\omega}))}{P_D(\theta(\boldsymbol{\omega'}) > \theta(\boldsymbol{\omega}))}[2N - 1] \text{ and } \gamma_L = \frac{P_L(\theta(\boldsymbol{\omega'}) > \theta(\boldsymbol{\omega}))}{P_D(\theta(\boldsymbol{\omega'}) > \theta(\boldsymbol{\omega}))}\left[\frac{1}{2N - 1}\right]$$

$\square$

It should be mentioned that there are two simplifications made to illuminate the proofs and that without such simplifications, the propositions would still hold:

1. The alliance knowledge production function $K_{AB}(\cdot)$ only took into consideration the possibility of learning from partners experiments when both firms component adjustments are in the common configured set. However, there is also the possibility of learning when only 1 firm experiments with a component in the common configured set and finds an improvement, but still has inferior quality technology relative to its partner. Proposition 1 remains valid since more common configured components increases probability of learning within the alliance.

2. Firms may have a mixture of common local and common distant components. Proposition 2 would still hold since thresholds $0 < \gamma_L, \gamma_H < \inf$ would always exist. This is easy to see since the extreme case of $E_D[\theta(\boldsymbol{\omega'}) - \theta(\boldsymbol{\omega}) \mid \theta(\boldsymbol{\omega'}) \geqslant \theta(\boldsymbol{\omega})] = 0$ implies that $\frac{d\pi}{d\tau_D} = 0$.

Proposition 5. $\frac{d\pi}{d\|\boldsymbol{\beta_A}\|_1}, \frac{d\pi}{d\|\boldsymbol{\beta_B}\|_1} > 0$

Proof. Checking both cases:

1. $\frac{d\pi}{d\|\boldsymbol{\beta_A}\|_1} = K_{AB}(\tau) - K_A > 0$

2. $\frac{d\pi}{d\|\boldsymbol{\beta_B}\|_1} = K_{AB}(\tau) - K_B > 0$

$\square$

Proposition 6. $\frac{d\pi}{ds_A}, \frac{d\pi}{ds_B} < 0$

Proof. This results follow directly from the assumption in Section 3.5:

1. $\frac{d\pi}{ds_A} = -\frac{dVONI(s_A)}{ds_A} < 0$

2. $\frac{d\pi}{ds_B} = -\frac{dVONI(s_B)}{ds_B} < 0$

$\square$

Bibliography

Abdulkadiroglu, A. and Sonmez, T. (2013), "Matching Markets: Theory and Practice," *Advances In Economics and Econometrics*, 1, 3–47.

Abdulkadiroglu, A., Pathak, P., Roth, A., and Sonmez, T. (2005), "The Boston Public School Match," *American Economic Review*, 95, 368–371.

Abrahamson, E. and Park, C. (1994), "Concealment of Negative Organizational Outcomes: An Agency Theory Perspective." *Academy of Management Journal*, 37, 1302–1334.

Agarwal, N. (2015), "An Empirical Model of the Medical Match," *American Economic Review*, 105, 1939–1978.

Agarwal, R. and Bayus, B. L. (2004. Creating and Surviving in New Industries. Advances in Strategic Management, 21(1): 107130), "Creating and Surviving in New Industries," *Advances in Strategic Management*, 21, 107130.

Ahuja, G. (2000), "Collaboration Networks, Structural Holes, and Innovation: A Longitudinal Study," *Administrative Science Quarterly*, 45, 425455.

Ahuja, G., Polidoro, F., and Mitchell, W. (2009), "Structural Homophily or Social Asymmetry? The Formation of Alliances by Poorly Embedded Firms," *Strategic Management Journal*, 30, 941–958.

Alcacer, J., Cantwell, J., and Gittelman, M. (2009), "Are Licensing Markets Local? An Analysis of the Geography of Vertical Licensing Agreements in Biopharmaceuticals." *Tech. Report. Working Paper. Harvard Business School, Harvard University. Boston, MA.*

Alexopoulos, M. (2011), "Read all About it!! What Happens Following a Technology Shock?" *American Economic Review*, 101, 1144–1179.

Almeida, P. and Kogut, B. (1999), "Localization of Knowledge and the Mobility of Engineers in Regional Networks," *Management Science*, 45, 905–917.

Anand, J., Oriani, R., and Vassolo, R. (2010), "Alliance Activity As a Dynamic Capability in the Face of a Discontinuous Technological Change," *Organization Science*, 21, 1213–1232.

Andersson, T., Gudmundsson, J., Talman, D., and Yang, Z. (2013), "A Competitive Partnership Formation Process," *Games and Economic Behavior*, 86, 165–177.

Argote, L. and Ingram, P. (2000), "Knowledge Transfer: A Basis for Competitive Advantage in Firms," *Organizational Behavior and Human Decision Processes*, 82, 150–169.

Arora, A., Fosfuri, A., and Gambardella, A. (2001), "Markets for Technology and Their Implications for Corporate Strategy," *Industrial And Corporate Change*, 10, 419–451.

Arora, A., Belenzon, S., and Patacconi, A. (2016), "Knowledge Sharing in Alliances and Alliance Portfolios," *Working Paper*.

Artz, K. and Brush, T. (2000), "Asset Specificity, Uncertainty, and Relational Norms: An Examination of Coordination Costs in Collaborative Strategic Alliances," *Journal Of Economic Behavior & Organization*, 41, 337–362.

Banbury, C. M. and Mitchell, W. (1995), "The Effect of Introducing Important Incremental Innovations on Market Share and Business Survival," *Strategic Management Journal*, 16, 161182.

Barnett, M. L. (2008), "An Attention-based View of Real Options Reasoning," *Academy of Management Review*, 33, 606628.

Barney, J. (1991), "Firm Resources and Sustained Competitive Advantage," *Journal of Management*, 17, 99–120.

Barr, P. S., Stimpert, J., and Huff, A. (1992), "Cognitive Change, Strategic Action, and Organizational Renewal," *Strategic Management Journal*, 13, 1536.

Barreto, I. and Patient, D. (2013), "Toward a Theory of Intraorganizational Attention Based on Desirability and Feasibility Factors," *Strategic Management Journal*, 34, 687–703.

Becker, G. (1981), *A Treatise on the Family*, Harvard University Press, Cambridge, MA.

Beckman, C., Haunschild, P., and Phillips, D. (2004), "Friends or Strangers? Firm-Specific Uncertainty, Market Uncertainty, and Network Partner Selection," *Organization Science*, 15, 259–275.

Bettman, J. and Weitz, B. (1983), "Attributions in the Board Room: Causal Reasoning in Corporate Annual Reports," *Administrative Science Quarterly*, 28, 165–183.

Blei, D. M. (2012), "Probabilistic Topic Models," *Communications of the ACM*, 55, 7784.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003), "Latent Dirichlet Allocation," *The Journal of Machine Learning Research*, 3, 9931022.

Boyd, D., Hamilton, L., Loeb, S., and Wyckoff, J. (2013), "Analyzing the Determinants of the Matching of Public School Teachers to Jobs: Estimating Compensating Differentials in Imperfect Labor Markets." *Journal of Labor Economics*, 31, 83–117.

Bresnahan, T. and Reiss, P. (1990), "Entry in Monopoly Markets." *Review of Economic Studies*, 57, 531–553.

Bucklin, L. and Sengupta, S. (1993), "Organizing Successful Co-Marketing Alliances," *Journal of Marketing*, 57, 32–46.

Cannon, M. D. and Edmondson, A. C. (2005), "Failing to Learn and Learning to Fail (Intelligently): How Great Organizations Put Failure to Work to Innovate and Improve," *Long Range Planning*, 38, 299–319.

Capron, L. and Mitchell, W. (2012), *Build, Borrow, or Buy: Solving the Growth Dilemma*, Harvard Business Press, Boston, MA.

Chesbrough, H. and Rosenbloom, R. (2002), "The Role of the Business Model in Capturing Value from Innovation: Evidence from Xerox Corporations Technology Spin-off Companies," *Industrial and Corporate Change*, 11, 529–555.

Chiappori, P. and Salanie, B. (2016), "The Econometrics of Matching Models," *Journal of Economic Literature*, 54, 832–861.

Cho, T. S. and Hambrick, D. C. (2006), "Attention as the Mediator Between Top Management Team Characteristics and Strategic Change: The Case of Airline Deregulation," *Organization Science*, 17, 453469.

Choo, E. and Siow, A. (2006), "Who Marries Whom and Why," *Journal of Political Economy*, 114, 175–201.

Ciliberto, F. and Tamer, E. (2009), "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77, 1791–1828.

Cockburn, I. and MacGarvie, M. (2011), "Entry and Patenting in the Software Industry," *Management Science*, 57, 915–933.

Cohen, W. and Klepper, S. (1996), "A Reprise of Size and R&D," *The Economic Journal*, 106, 925–951.

Conti, R., Gambardella, A., and Novelli, E. (2013), "Research on Markets for Inventions and Implications for R&D Allocation Strategies," *Academy of Management Annals*, 7, 715772.

Cummings, J. and Kiesler, S. (2007), "Coordination Costs and Project Outcomes in Multi-University Collaborations," *Research Policy*, 36, 1620–1634.

Cyert, R. and March, J. (1963), *A Behavioral Theory of the Firm*, Wiley-Blackwell, Macklen, MA.

Dahlander, L., OMahony, S., and Gann, D. (2016), "One Foot In, One Foot Out: How Does Individuals External Search Breadth Affect Innovation Outcomes?" *Strategic Management Journal*, 37, 280–302.

De Figueiredo, J. and Kyle, M. (2006), "Surviving the Gales of Creative Destruction: The Determinants of Product Turnover," *Strategic Management Journal*, 27, 241–264.

Deeds, D., DeCarolis, D., and Coombs, J. (2000), "Dynamic Capabilities and New Product Development in High Technology Ventures: An Empirical Analysis of New Biotechnology Firms," *Journal of Business Venturing*, 15, 211–229.

Del Boca, D. and Flinn, C. (2014), "Household Behavior and the Marriage Market," *Journal of Economic Theory*, 150, 515–550.

Denrell, J. and March, J. (2001), "Adaptation as Information Restriction: The Hot Stove Effect," *Organization Science*, 12, 523–538.

Diestre, L. and Rajagopalan, N. (2012), "Are All Sharks Dangerous? New Biotechnology Ventures and Partner Selection in R&D Alliances," *Strategic Management Journal*, 33, 1115–1134.

DiMasi, J., Hansen, R., and Grabowski, H. (2003), "The Price of Innovation: New Estimates of Drug Development Costs," *Journal of Health Economics*, 22, 151–185.

Dosi, G. (1982), "Technological Paradigms and Technological Trajectories: A Suggested Interpretation of the Determinants and Directions of Technical Change," *Research Policy*, 11, 147–162.

Dosi, G. (1988), *Technical Change and Economic Theory*, chap. The Nature of Innovative Process, pp. 221–238, London: Pinter.

Dupuy, A. and Galichon, A. (2014), "Personality Traits and the Marriage Market," *Journal of Political Economy*, 122, 1271–1319.

Dushnitsky, G. and Shaver, J. (2009), "Limitations to Interorganizational Knowledge Acquisition: The Paradox of Corporate Venture Capital," *Strategic Management Journal*, 30, 1045–1064.

Dutton, J. E., Ashford, S. J., O'neill, R. M., Hayes, E., and Wierba, E. E. (1997), "Reading the Wind: How Middle Managers Assess the Context for Selling Issues to Top Managers," *Strategic Management Journal*, 18, 407–423.

Echenique, F. and Oviedo, J. (2006), "A Theory of Stability in Many to Many Matching Markets," *Theoretical Economics*, 1, 233–273.

Eggers, J. and Kaplan, S. (2009), "Cognition and Renewal: Comparing CEO and Organizational Effects on Incumbent Adaptation to Technical Change." *Organization Science*, 20, 461477.

Eggers, J. P. (2012a.), "All Experience is Not Created Equal: Learning, Adapting, and Focusing in Product Portfolio Management," *Strategic Management Journal*, 33, 315335.

Eggers, J. P. (2012b.), "Falling Flat: Failed Technologies and Investment Under Uncertainty," *Administrative Science Quarterly*, 57, 4780.

Fleming, L. and Sorenson, O. (2004), "Science as a Map in Technological Search," *Strategic Management Journal*, 25, 909–928.

Fosfuri, A., Lanzolla, G., and Suarez, F. F. (2013), "Entry-timing Strategies: The Road Ahead," *Long Range Planning*, 46, 300311.

Fountain, K. and Stovel, C. (2009), *The Oxford Handbook Of Analytical Sociology*, chap. Matching, Oxford University Press, New York.

Fox, J. (2010), "Identification in Matching Games," *Quantitative Economics*, 1, 203–254.

Fox, J. (2017), "Estimating Matching Games with Transfers," *Tech. Report. NBER Working Paper.*

Fox, J., Hsu, D., and Yang, C. (2012), "Unobserved Heterogeneity in Matching Games with an Application to Venture Capital," *Tech. Report. NBER Working Paper.*

Franco, A. M., Sarkar, M., Agarwal, R., and Echambadi, R. (2009), "Swift and Smart: The Moderating Effects of Technological Capabilities on the Market Pioneeringfirm Survival Relationship," *Management Science*, 55, 18421860.

Fuentelsaz, L., Gomez, J., and Polo, Y. (2002), "Followers' Entry Timing: Evidence from the Spanish Banking Sector After Deregulation," *Strategic Management Journal*, 23, 245264.

Gale, D. and Shapley, L. (1962), "College Admissions and the Stability of Marriage," *American Mathematical Monthly*, 69, 9–15.

Galichon, A. and Salanie, B. (2015), "Cupids Invisible Hand: Social Surplus and Identification in Matching Models," *Tech. Report. New York University Working Paper.*

134

Gerken, J. and Moehrle, M. (2012), "A New Instrument for Technology Monitoring: Novelty in Patents Measured by Semantic Patent Analysis," *Scientometrics*, 91, 645–670.

Giachetti, C. (2013), *Competitive Dynamics in the Mobile Phone Industry*, Basingstoke: Palgrave Macmillan.

Giachetti, C. and Lampel, J. (2010), "Keeping Both Eyes on the Competition: Strategic Adjustment to Multiple Targets in the UK Mobile Phone Industry." *Strategic Organization*, 8, 347376.

Golder, P. and Tellis, G. J. (1993), "Pioneer Advantage: Marketing Logic or Marketing Legend?" *Journal of Marketing Research*, 30, 158170.

Gomez, J. and Maicas, J. (2011), "Do Switching Costs Mediate the Relationship Between Entry Timing and Performance?" *Strategic Management Journal*, 32, 1251–1269.

Gordon, N. and Knight, B. (2009), "A Spatial Merger Estimator with an Application to School District Consolidation," *Journal of Public Economics*, 93, 752–765.

Graham, B. (2011), *Handbook of Social Economics*, vol. 1B, chap. Econometric Methods for the Analysis of Assignment Problems in the Presence of Complementarity and Social Spillovers, pp. 965–1052, Elsevier.

Greve, H., Mitsuhashi, H., and Baum, J. (2013), "Greener Pastures: Outside Options and Strategic Alliance Withdrawal," *Organization Science*, 24, 79–98.

Grun, B. and Hornik, K. (2011), "Topicmodels: An R Package for Fitting Topic Models," *Journal of Statistical Software*, 40, 130.

Gulati, R. and Singh, H. (1998), "The Architecture of Cooperation: Managing Coordination Costs and Appropriation Concerns in Strategic Alliances," *Administrative Science Quarterly*, 43, 781–814.

Guler, I. (2007), "Throwing Good Money After Bad? Political and Institutional Influences on Sequential Decision Making in the Venture Capital Industry," *Administrative Science Quarterly*, 52, 248285.

Hagedoorn, J. (1993), "Understanding the Rationale of Strategic Technology Partnering: Nterorganizational Modes of Cooperation and Sectoral Differences," *Strategic Management Journal*, 14, 371–385.

Hallen, B. (2008), "The Causes and Consequences of the Initial Network Positions of New Organizations: From Whom Do Entrepreneurs Receive Investments?" *Administrative Science Quarterly*, 53, 685–718.

Hamel, G. (1991), "Competition for Competence and Interpartner Learning Within International Strategic Alliances," *Strategic Management Journal*, 12, 83–103.

Hart, O. and Moore, J. (1988), "Incomplete Contracts and Renegotiation," *Econometrica*, 56, 755–785.

Hatfield, J. and Kominers, S. (2017), "Contract Design and Stability in Many to Many Matching," *Games and Economic Behavior*, 101, 78–97.

Hatfield, J. and Milgrom, P. (2005), "Matching with Contracts," *American Economic Review*, 95, 913–935.

Hawk, A., PachecoDeAlmeida, G., and Yeung, B. (2013), "Fastmover Advantages: Speed Capabilities and Entry into the Emerging Submarket of Atlantic Basin LNG." *Strategic Management Journal*, 34, 1531–1550.

Helfat, C. and Peteraf, M. (2015), "Managerial Cognitive Capabilities and the Microfoundations of Dynamic Capabilities," *Strategic Management Journal*, 36, 831–850.

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

Hitsch, G., Hortacsu, A., and Ariely, D. (2010), "Matching and Sorting in Online Dating," *American Economic Review*, 100, 130–163.

Hoffman, A. and Ocasio, W. (2001), "Not All Events are Attended Equally: Toward a Middle-Range Theory of Industry Attention to External Events," *Organization Science*, 12, 414–434.

Horowitz, J. (1992), "A Smoothed Mazimum Score Estimator for the Binary Response Model," *Econometrica*, 60, 505–531.

Hsieh, Y. (2012), "Understanding Mate Preferences from Two-sided Matching Markets: Identification, Estimation and Policy Analysis," *Tech. Report. Unpublished Manuscript.*

Irving, R. and Leather, P. (1986), "The Complexity of Counting Stable Marriages," *SIAM Journal of Computing*, 15, 655–667.

Joseph, J. and Ocasio, W. (2012), "Architecture, Attention, and Adaptation in the Multibusiness Firm: General Electric from 1951 to 2001," *Strategic Management Journal*, 33, 633660.

Kaplan, S. (2008), "Cognition, Capabilities, and Incentives: Assessing Firm Response to the Fiber-optic Revolution," *Academy of Management Journal*, 51, 672695.

Kaplan, S. and Vakili, K. (2015), "The Doubleedged Sword of Recombination in Breakthrough Innovation," *Strategic Management Journal*, 36, 14351457.

Katila, R., Rosenberge, J., and Eisenhardt, K. (2008), "Swimming With Sharks: Technology Ventures, Defense Mechanisms, and Corporate Relationships," *Administrative Science Quarterly*, 53, 295–332.

Kauffman, S., Lobo, J., and Macready, W. (2000), "Optimal Search on a Technological Landscape," *Journal of Economic Behavior & Organization*, 43, 141–166.

Kavusan, K., Noorderhaven, N., and Duysters, G. (2016), "Knowledge Acquisition and Complementary Specialization in Alliances: The Impact of Technological Overlap and Alliance Experience," *Research Policy*, 45, 2153–2165.

Keil, M. and Robey, D. (1999), "Turning Around Troubled Software Projects: An Exploratory Study of the De-escalation of Commitment to Failing Courses of Action," *Journal of Management Information Systems*, 15, 63–87.

Keil, T., McGrath, R., and Tukiainen, T. (2009), "Gems from the Ashes: Capability Creation and Transformation in Internal Corporate Venturing," *Organization Science*, 20, 601–620.

Kelso, A. and Crawford, V. (1982), "Job Matching, Coalition Formation, and Gross Substitutes," *Econometrica*, 50, 1483–1504.

Khanna, T., Gulati, R., and Nohria, N. (1998), "The Dynamics of Learning Alliances: Competition, Cooperation, and Relative Scope," *Strategic Management Journal*, 19, 193–210.

Kiesler, S. and Sproull, L. (1982), "Managerial Response to Changing Environments: Perspectives on Problem Sensing from Social Cognition," *Administrative Science Quarterly*, 27, 548–570.

Klepper, S. and Simons, K. L. (2000), "Dominance by Birthright: Entry of Prior Radio Producers and Competitive Ramifications in the U.S. Television Receiver Industry," *Strategic Management Journal*, 21, 9971016.

Klingebiel, R. (2017), "Risk-Type Preference Shifts in Response to Performance Feedback," *Strategic Organization, Published Online and Forthcoming.*

Klingebiel, R. and Joseph, J. (2016), "Entry Timing and Innovation Strategy in Feature Phones," *Strategic Management Journal*, 37, 1002–1020.

Klingebiel, R. and Rammer, C. (2014), "Resource Allocation Strategy for Innovation Portfolio Management," *Strategic Management Journal*, 35, 246268.

Koski, H. and Kretschmer, T. (2010), "New Product Development and Firm Value in Mobile Handset Production," *Information Economics and Policy*, 22, 4250.

Lant, T., Milliken, F., and Batra, B. (1992), "The Role of Managerial Learning and Interpretation in Strategic Persistence and Reorientation: An Empirical Exploration," *Strategic Management Journal*, 13, 585–608.

Laursen, K. and Salter, A. (2006), "Open for Innovation: The Role of Openness in Explaining Innovation Performance Among UK Manufacturing Firms," *Strategic Management Journal*, 27, 131150.

Lavie, D. (2007), "Alliance Portfolios and Firm Performance: A Study of Value Creation and Appropriation in the US Software Industry," *Strategic Management Journal*, 28, 1187–1212.

Lee, G. K. (2008), "Relevance of Organizational Capabilities and its Dynamics: What to Learn from Entrants' Product Portfolios about the Determinants of Entry Timing," *Strategic Management Journal*, 29, 12571280.

Leiponen, A. and Helfat, C. (2010), "Innovation Objectives, Knowledge Sources, and The Benefits of Breadth," *Strategic Management Journal*, 31, 224236.

LeonardBarton, D. (1992), "Core Capabilities and Core Rigidities: A Paradox in Managing New Product Development," *Strategic Management Journal*, 13, 111125.

Levinthal, D. and Rerup, C. (2006), "Crossing an Apparent Chasm: Bridging Mindful and Less-mindful Perspectives on Organizational Learning," *Organization Science*, 17, 502513.

Levy, D. (2005), "Offshoring in the New Global Political Economy," *Journal of Management Studies*, 42, 685–693.

Li, D., Eden, L., Hitt, M., and Ireland, R. (2008), "Friends, Acquaintances, or Strangers? Partner Selection in R&D Alliances," *Academy of Management Journal*, 51, 315–334.

Li, Q., Maggitti, P. G., Smith, K. G., Tesluk, P. E., and Katila, R. (2013), "Top Management Attention to Innovation: The Role of Search Selection and Intensity in New Product Introductions," *Academy of Management Journal*, 56, 893916.

Li, S. and Rowley, T. (2002), "Inertia and Evaluation Mechanisms in Interorganizational Partner Selection: Syndicate Formation Among U.S. Investment Banks," *Academy of Management Journal*, 45, 1104–1119.

Lieberman, M. B. and Montgomery, D. B. (1988), "First-Mover Advantages," *Strategic Management Journal*, 9, 4158.

MacDonald, G. and Ryall, M. (2004), "How Do Value Creation and Competition Determine Whether a Firm Appropriates Value?" *Management Science*, 50, 1319–1333.

Manski, C. (1983), "Closest Empirical Distribution Estimation," *Econometrica*, 51, 305–319.

Manski, C. (1985), "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator," *Journal of Econometrics*, 27, 313–333.

March, J. (1991), "Exploration and Exploitation in Organizational Learning," *Organization Science*, 2, 71–87.

Martin, X., Mitchell, W., and Swaminathan, A. (1995), "Recreating and Extending Japanese Automobile Buyer-Supplier Links in North America," *Strategic Management Journal*, 16, 580619.

Mason, R. and Drakeman, D. (2014), "Comment on Fishing For Sharks: Partner Selection in Biopharmaceutical R&D Alliances by Diestre and Rajagopalan," *Strategic Management Journal*, 35, 1564–1565.

Maula, M. V., Keil, T., and Zahra, S. A. (2013), "Top Managements Attention to Discontinuous Technological Change: Corporate Venture Capital as an Alert Mechanism," *Organization Science*, 24, 926947.

Mawdsley, J. and Somaya, D. (2016), "Employee Mobility and Organizational Outcomes: An Integrative Conceptual Framework and Research Agenda," *Journal of Management*, 42, 85–113.

McCardle, K. F. (1985), "Information Acquisition and The Adoption of New Technology," *Management Science*, 31, 13721389.

McGrath, R., Keil, T., and Tukiainen, T. (2006), "Extracting Value from Corporate Venturing," *MIT Sloan Management Review*, 48, 50.

Menzel, K. (2015), "Inference for Games with Many Players," *Tech. Report. New York University Working Paper.*

Miller, D. and Chen, M. (1994), "Sources and Consequences of Competitive Inertia: A Study of the U.S. Airline Industry," *Administrative Science Quarterly*, 39, 1–23.

Mindruta, D., Moen, M., and Agarwal, R. (2016.), "A Two-sided Matching Approach for Partner Selection and Assessing Complementarities in Partners Attributes in Inter-firm Alliances," *Strategic Management Journal*, 37, 206–231.

Mintzberg, H., Raisinghani, D., and Theoret, A. (1976), "The Structure of "Unstructured" Decision Processes," *Administrative Science Quarterly*, 21, 246275.

Mitchell, W. (1989), "Whether and When? Probability and Timing of Incumbents' Entry Into Emerging Industrial Subfields," *Administrative Science Quarterly*, 34, 208230.

Mitchell, W. (1991), "Dual Clocks: Entry Order Influences on Incumbent and Newcomer Market Share and Survival When Specialized Assets Retain their Value," *Strategic Management Journal*, 12, 85100.

Mitchell, W. and Singh, K. (1996), "Survival of Businesses Using Collaborative Relationships to Commercialize Complex Goods," *Strategic Management Journal*, 17, 169–195.

Mowery, D., Oxley, J., and Silverman, B. (1996), "Strategic Alliances and Interfirm Knowledge Transfer," *Strategic Management Journal*, 17, 77–91.

Mowery, D., Oxley, J., and Silverman, B. (1998), "Technological Overlap and Interfirm Cooperation: Implications for the Resource-based View of the Firm," *Research Policy*, 27, 507–523.

Nadkarni, S. and Barr, P. (2008), "Environmental Context, Managerial Cognition, and Strategic Action: An Integrated View," *Strategic Management Journal*, 29, 1395–1427.

Nadkarni, S. and Chen, J. (2014), "Bridging Yesterday, Today, and Tomorrow: CEO Temporal Focus, Environmental Dynamism, and Rate of New Product Introduction," *Academy of Management Journal*, 57, 1810–1833.

Nadkarni, S. and Narayanan, V. K. (2007), "Strategic Schemas, Strategic Flexibility, and Firm Performance: The Moderating Role of Industry Clockspeed," *Strategic Management Journal*, 28, 243270.

Nelson, R. R. and Winter, S. G. (1982), *An Evolutionary Theory of Economic Change*, Belknap Press, Cambridge, MA.

Nelson, R. R. and Winter, S. G. (2009), *An Evolutionary Theory of Economic Change*, Harvard University Press, Cambridge, MA.

Nickerson, J. A. and Zenger, T. R. (2004), "A Knowledge-Based Theory of the FirmThe Problem-solving Perspective," *Organization Science*, 15, 617632.

Nutt, P. C. (1984), "Types of Organizational Decision Processes," *Administrative Science Quarterly*, 29, 414450.

Ocasio, W. (1997), "Towards an AttentionBased View of the Firm," *Strategic Management Journal*, 18, 187206.

Ocasio, W. (2011), "Attention to Attention," *Organization Science*, 22, 12861296.

Ocasio, W. and Joseph, J. (2005), "An Attention-Based Theory of Strategy Formulation: Linking Micro-and Macroperspectives in Strategy Processes," *Advances in Strategic Management*, 22, 3961.

Oxley, J. (1997), "Appropriability Hazards and Governance in Strategic Alliances: A Transaction Cost Approach," *Journal of Law, Economics and Organization*, 13, 387–409.

Palmie, M., Lingens, B., and Gassmann, O. (2016), "Towards an Attention-Based View of Technology Decisions," *R&D Management*, 46, 781–796.

Paulson Gjerde, K. A., Slotnick, S. A., and Sobel, M. J. (2002), "New Product Innovation with Multiple Features and Technology Constraints," *Management Science*, 48, 12681284.

Penrose, E. (1959), *The Theory of the Growth of the Firm*, Oxford University Press, New York.

Porac, J. F., Thomas, H., and BadenFuller, C. (1989), "Competitive Groups as Cognitive Communities: The Case of Scottish Knitwear Manufacturers," *Journal of Management Studies*, 26, 397–416.

Powell, W., Koput, K., and Smith-Doerr, L. (1996), "Interorganizational Collaboration and the Locus of Innovation: Networks of Learning in Biotechnology." *Administrative Science Quarterly*, 41, 116–145.

Putsis Jr., W. P. and Bayus, B. L. (2001), "An Empirical Analysis of Firms Product Line Decisions," *Journal of Marketing Research*, 38, 110118.

Pycia, M. (2012), "Stability and Preference Alignment in Matching and Coalition Formation," *Econometrica*, 80, 323–362.

Rao, R. (1962), "Relations Between Weak and Uniform Convergence of Measures with Applications," *Annals of Mathematical Statistics*, 33, 659–680.

Rerup, C. (2009), "Attentional Triangulation: Learning from Unexpected Rare Crises," *Organization Science*, 20, 876893.

Robinson, W. T. and Chiang, J. (2002), "Product Development Strategies for Established Market Pioneers, Early Followers, and Late Entrants," *Strategic Management Journal*, 23, 855866.

Rogers, E. (2010), *Diffusion of Innovations*, Simon and Schuster, New York, NY.

Romano, J. and Shaikh, A. (2010), "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78, 168–211.

Rosenkopf, L. and Almeida, P. (2003), "Overcoming Local Search Through Alliances and Mobility," *Management Science*, 49, 751–766.

Roth, A. (1984), "Stability and Polarization of Interests in Job Matching," *Econometrica*, 52, 47–57.

Roth, A. (1985), "The College Admissions Problem is not Equivalent to the Marriage Problem," *Journal of Economic Theory*, 36, 277–288.

Roth, A. and Sotomayor, M. (1992), *Handbook of Game Theory with Economic Applications*, chap. Two-sided Matching, pp. 485–541, Elsevier, 1 edn.

Roth, A., Sonmez, T., and Unver, M. (2004), "Kidney Exchange," *Quarterly Journal of Economics*, 119, 457–488.

Rothaermel, F. (2002), "Technological Discontinuities and Interfirm Cooperation: What Determines a Startups Attractiveness as an Alliance Partner?" *IEEE Transactions on Engineering Management*, 49, 388–397.

Rothaermel, F. and Boeker, W. (2008), "Old Technology Meets New Technology: Complementarities, Similarities, and Alliance Formation," *Strategic Management Journal*, 29, 47–77.

Schilling, M. (2015), "Technology Shocks, Technological Collaboration, and Innovation Outcomes," *Organization Science*, 26, 668–686.

Schmidt, P. and Witte, A. D. (1989), "Predicting Criminal Recidivism Using Split Population Survival Time Models," *Journal of Econometrics*, 40, 141159.

Shapley, L. and Schubik, M. (1972), "The Assignment Game 1: The Core," *International Journal of Game Theory*, 1, 111–130.

Simon, H. A. (1947), *Administrative Behavior*, The Free Press, New York.

Song, J., Almeida, P., and Wu, G. (2003), "Learning-by-Hiring: When is Mobility More Likely to Facilitate Interfirm Knowledge Transfer?" *Management Science*, 49, 351–365.

Sorensen, M. (2005), "An Economic and Econometric Analysis of Market Sorting with an Application to Venture Capital," *Unpublished dissertation (Stanford University)*.

Sorensen, M. (2007.), "How Smart is Smart Money? A Two-sided Matching Model of Venture Capital," *Journal of Finance*, 62, 2725–2762.

Sorenson, O., McEvily, S., Ren, C. R., and Roy, R. (2006), "Niche Width Revisited: Organizational Scope, Behavior and Performance," *Strategic Management Journal*, 27, 915936.

Sotomayor, M. (2009), "Adjusting Prices in the Multiple-partners Assignment Game," *International Journal of Game Theory*, 38, 575–600.

Stembridge, B. (1999), "International Patent Classification in Derwent Databases," *World Patent Information*, 21, 169–177.

Stern, I., Dukerich, J., and Zajac, E. (2014), "Unmixed Signals: How Reputation and Status Affect Alliance Formation," *Strategic Management Journal*, 35, 512–531.

Stevens, R., Moray, N., Bruneel, J., and Clarysse, B. (2015), "Attention Allocation to Multiple Goals: The Case of For-Profit Social Enterprises," *Strategic Management Journal*, 36, 1006–1016.

Stuart, T. and Podolny, J. (1996), "Local Search and the Evolution of Technological Capabilities," *Strategic Management Journal*, 17, 21–38.

Stuart, T., Hoang, H., and Hybels, R. (1999), "Interorganizational Endorsements and the Performance of Entrepreneurial Ventures," *Administrative Science Quarterly*, 44, 315–349.

Suarez, F. F. and Lanzolla, G. (2007), "The Role of Environmental Dynamics in Building a First Mover Advantage Theory," *Academy of Management Review*, 32, 377392.

Teece, D. (1989), "Competition and Cooperation in Technology Strategy," *Business Review*, 36.

Teece, D., Pisano, G., and Shuen, A. (1997), "Dynamic Capabilities and Strategic Management," *Strategic Management Journal*, 18, 509–533.

Tuggle, C. S., Sirmon, D. G., Reutzel, C. R., and Bierman, L. (2010), "Commanding Board of Director Attention: Investigating How Organizational Performance and CEO Duality Affect Board Members' Attention to Monitoring," *Strategic Management Journal*, 31, 946968.

Van de Ven, A. (1986), "Central Problems in the Management of Innovation," *Management Science*, 32, 590–607.

VanderWerf, P. A. and Mahon, J. F. (1997), "Meta-analysis of the Impact of Research Methods on Findings of First-Mover Advantage," *Management Science*, 43, 15101519.

Vidal, E. and Mitchell, W. (2013), "When Do First Entrants Become First Survivors?" *Long Range Planning*, 46, 335347.

Walkley, T. M. and Ramsay, M. D. (2011), "Investor Report Wireless Telecommunications Equipment: Global Market Summary," *Toronto: Canaccord Genuity*.

Weick, K. E. and Sutcliffe, K. M. (2006), "Mindfulness And the Quality of Organizational Attention," *Organization Science*, 17, 514524.

Wernerfelt, B. and Karnani, A. (1987), "Competitive Strategy Under Uncertainty," *Strategic Management Journal*, 8, 187194.

Williams, C. and Mitchell, W. (2004), "Focusing Firm Evolution: The Impact of Information Infrastructure on Market Entry by US Telecommunications Companies, 19841998," *Management Science*, 50, 15611575.

Williamson, O. (1979), "Transaction-cost Economics: The Governance of Contractual Relations." *The Journal of Law and Economics*, 22, 233–261.

Wilson, A. J. and Joseph, J. (2015), "Organizational Attention and Technological Search in the Multibusiness Firm: Motorola from 1974 to 1997. Cognition and Strategy," *Advances in Strategic Management*, 32, 407435.

Yadav, M., Prabhu, J., and Chandy, R. (2007), "Managing the Future: CEO Attention and Innovation Outcomes," *Journal of Marketing*, 71, 84–101.

Yan, A. and Gray, B. (1994), "Bargaining Power, Management Control, and Performance in United States-China Joint Ventures: A Comparative Case Study," *Academy of Management Journal*, 37, 1478–1517.

Yang, Y., Shi, M., and Goldfarb, A. (2009), "Estimating the Value of Brand Alliances in Professional Team Sports," *Marketing Science*, 28, 1095–1111.

Zahra, S. and George, G. (2002), "Absorptive Capacity: A Review, Reconceptualization, and Extension," *Academy of Management Review*, 27, 185–203.

Biography

Kevin Kai Du was born in Shanghai, China on October 16, 1987. He immigrated to the United States at the age of 4 and received his formal education through the New York City public school system. After graduating from Stuyvesant High School in 2005, he attended Dartmouth College where he majored in Mathematics and Economics, receiving his Bachelor of Arts degree in June 2009. He also holds a Master of Arts degree with a concentration in Economics from New York University as of May 2011. He is a current student at Duke University, a recipient of the Fuqua School of Business Doctoral Fellowship (2012-2017), and in the process of completing his Ph.D. degree in Business Administration with a focus in strategic management. After his studies at Duke is completed, he will take up a faculty position at the University of Southern Denmark as an Assistant Professor of Strategic Organization Design.