Competing in the Era of Emergent Architecture: The Case of the Packaged Software Industry

Bala Iyer TOIM division Babson College Babson Hall Babson Park, MA 02457 biyer@babson.edu	Chi-Hyon Lee School of Management George Mason University Enterprise Hall 208 Fairfax, VA 22030 cle5@gmu.edu	David Dreyfus School of Management Boston University Boston, MA 02215 ddreyfus@bu.edu

Abstract

The firms that comprise the prepackaged software industry form a complex system engaged in systems-based competition. This complex system survives and grows because it follows emergent design principles notably articulated by Herbert Simon. In particular, complex systems form stable subsystems – clusters – that represent the underlying architecture of the industry. In the software industry this architecture is often referred to as a stack.

These stable subsystems can be statically or normatively defined a priori or computed dynamically based on firms introducing products that interact with other products in the marketplace ex ante. We refer to the statically defined architecture as the espoused stack and the dynamically computed architecture as the emergent stack.

In this research, we study the decomposition of the software industry using 13 years (1990-2002) of data on packaged software development firms selling into functional markets. The results suggest that firms that exploit complementarities perform better than those that do not, and that firms that exploit complementarities based on the emergent stack/architecture outperform competitors that use complementarities based on the espoused architecture/stack and competitors that use complementarities without regard for architecture.

1 Introduction

Technology Strategy Researchers are generally interested in the interplay between the evolution of technologies and their impact on firm and industry performance [1-5]. Technological evolution both influences and is influenced by industry structure. Firms that incorporate an understanding of industry structure into their decision making processes often outperform firms that don't [6]. One dimension of industry structure is the interdependencies between market segments. As a result, understanding the pattern of interdependencies within an industry – its architecture – is an important element in managerial decision making.

In general, architecture can be espoused – as viewed by a designer – or emergent – the architecture "as is." In complex, emergent systems lacking central control, the notion of an espoused design is tenuous. We substitute the "industry analyst" for the designer and his or her abstract description of market segments for the espoused design. Understanding the emergent architecture is far more difficult because it requires examining the individual components (i.e., firms) and the links between them (i.e., interdependencies) in order to infer the architecture.

The managerial question is three-fold: is there a non-random structure within the industry and, if so, should the decision maker adopt the emergent or espoused view of it. If the analyst's perspective is as valuable as an understanding of the emergent architecture, then the manager can dispense with the effort of detailed analysis to identify the emergent architecture and defer to the industry analysts. If, however, the emergent architecture is sufficiently more nuanced than the analyst's abstraction, and enables the identification of additional performance opportunities, its discovery may be worth the effort.

We are led to believe that any sufficiently complex industry will have an architecture worth understanding due to the insights of Simon [7] who argues that complex systems are built of stable subsystems. These stable subsystems are the modular clusters [8] that give rise to value opportunities [2]. It is these stable subsystems and the nature of their emergent interconnections, and associated behaviors, that form the structure that strategists are interested in and that are the subject of our interest.

We consider an industry to be an example of a complex system that has many subsystems and no single point of control - its leadership is often divided [9]. Industries, like many complex systems, have structure and emergent behaviors. In this work we explore one aspect of an industry's structure – its modular decomposition – and one of its behaviors – the strategic product offerings made by firms. We evaluate the quality of the imputed strategic product offering decision by measuring the firm's financial performance.

In this work, industry structure is a state description of the trading relationships among firms and their customers. The structure is maintained and changed as a result of firms' decisions to introduce and withdraw products from markets. The position of a firm within this structure is a result of the firm's specific product introduction decisions and reflects the firm's joint position in multiple markets. The unique position of the firm may influence its financial performance above and beyond the benefit conferred by the individual product markets. This study joins other studies that have looked at industry structure and a firm's position within it in order to predict the firm's financial performance [10-13].

In constructing our industry structure we started by examining the role of complementarities and network effects in strategy development [14-16]. We use the trading relationships among firms and their customers to deduce complementary product markets. The firm-specific question is when should the firm "stick to its knitting" and when should it diversify. If the firm diversifies, it must then decide whether or not to enter into complementary markets, and how it should evaluate potential complementary markets.

In this paper we explore the impact of product complementarities on firm performance in the software industry. Software products are often complementary because they are used together and form the basis of systems-based competition [9]. The complementarities lead to market interdependencies, which are the basis for the emergent industry architecture. We measure complementarity between two products as the degree to which they are purchased in tandem. We find that product complementarities based on the emergent industry architecture is a better predictor of firm performance than complementarities based on the espoused architecture, which is better than assuming no industry architecture (i.e., a random architecture). We find that while complementarity is good in general, it has a diminishing marginal return.

The rest of our paper is organized as follows. We begin with an overview of how complementarities have influenced the software industry. In section 3 and 4 we describe espoused and emergent architecture. Section 5 shows how we use market boundaries to identify architecture. Subsequently, we describe our methods – including data, measures and modeling specifications – and present our results.

2 Complementarities in software

In markets characterized by systems-based competition in which customers must purchase bundles of products, often from multiple vendors, value is derived from complementary products. In simple terms, a complementary product is one that enhances the value of another product when the two of them are used together by end-users [17]. For example, in the software industry database products and operating systems are complementary. A database product cannot even be used without an operating system; thus, the existence of operating systems increases the value of the database product. Similarly, the existence of database products drives the sales of hardware and operating systems.

In systems-based industries two or more components made by different manufacturers using different technologies may have to be interoperable. This need for interoperability leads to the creation of standards, which facilitates the creation of additional complementary products. Network effects across markets [14-16] result in higher valuation for products with larger complementary markets and create incentives for producers of a particular good to enter the markets for complements.

The desire to exploit complementarities to derive competitive advantages and create and appropriate value motivates a number of managerial decisions. These decisions include those that lead to mergers and acquisitions, alliance formation, standards creation, and product introductions. Companies that produce highly complementary components may want to merge or vertically integrate if customers value a more reliable systems integration supplied by a single provider [18] or if they want to quickly gain market share in the complementary market. Companies also make acquisitions in a complementary market with the purpose of foreclosing competitors in that market. The "winner takes all" nature of software economics gives firms that achieve major platform status massive profit pools from which to invest in adjacent software categories.

Companies can form alliances and standards committees to facilitate tighter integration at a strategic or technological level. Interoperability among products occurs when the products can utilize each others' published interfaces (APIs). The interfaces are the result of negotiations among companies. These negotiations are sometimes public and are conducted in standards committees and are sometimes private. Both public and private negotiations involve the sharing of varying levels of company confidential information, which leads to the formation of alliances.

Companies can also use either their installed base, or the installed base of complementary components, to leverage and promote growth through product introductions. Firms developing products can choose to participate in developing and marketing complementary products or they may allow third-party developers to provide them. Firms can actively engage in making sure that complementary products are interoperable, or they can rely upon their customers to do that. Historically, large firms have developed complementary products in-house to ensure that the product interfaces are properly utilized and incremental profits appropriated [19].

Based on the resource-based views of the firm [20, 21], the use of complementary factors of production across multiple business units should lead to production-side synergies, economies of scope and improved firm performance [22]. For example, software firms that reuse the same software code in multiple software products should gain economies of scope in software development and perform better than software firms that write new code for each new product. Moreover, the firm can leverage their complementary assets - sales force, customer support departments, installed base, and their understanding of customer requirements [23, 24].

Related products can also exploit consumption side synergies. When a set of products serve the needs of the same customer base, and the value of the set of products to the consumer is greater than the sum of the value of each product in isolation, the set is said to offer consumption-side synergies. There are three types of consumption-side synergies: shopping cart and search cost savings, demand variance reduction [25], and product value in-learning and in-use [2].

Search and shopping cart costs refer to the costs borne by the consumer in locating compatible products, selecting the right product, and making the software purchase. A firm offering more than one product can reduce these costs to the consumer and share in the cost savings. This type of synergy allows the producer to reduce customer acquisition costs by cross-selling products to the same customer base. However, the costs of developing, supporting, and coordinating the delivery of multiple products is high. Therefore, firms may utilize a more nuanced diversification strategy than one that is solely intended to capture search and shopping cart savings. By limiting the firm's product diversification to compatible, complementary products, the firm increases the likelihood that a consumer will select more than one product in order to realize the potential product synergies. This increases the value to the firm of offering a diversified product portfolio.

Demand variance reduction refers to the reduction in uncertainty that a specific second product, B, will be purchased if a first product, A, is purchased [25]. If two products are not complements, the estimated probability that product B will be purchased is unchanged with the knowledge that product A is purchased. However, if the two products are complements, then the knowledge of a purchase of product A increases the probability that product B will also be purchased. By providing complementary products, the firm not only increases the odds of the consumer making a subsequent purchase, but also increases the odds that the subsequent purchase will be one of a specific set of complementary products.

Product value in-learning and in-use refer to the additional value that the consumer receives in using complementary products together. Customers benefit by a set of interoperable products with familiar user interfaces [26], trained and supported by a single vendor that understands the task environment in which the set of compatible products will be deployed. Firms that provide a set of complementary products can appropriate some of the value created and better weather competitive price pressures in the markets for the individual products.

Our core assertion is that firms that have the necessary assets, resources and skills to develop and market complementary products will outperform those that don't. They will take advantage of production-side and consumption-side synergies to improve their performance. More formally:

Hypothesis-1a: Complementarity of a firm's products has a positive effect on firm performance.

Any pair of products can range from not complementary, in which case the demand for one product provides no information regarding demand for the other product, to perfectly complementary, in which demand for one product provides perfect information regarding demand for the other product. We argued previously that firms selling complementary products would outperform those that either sold only one product or sold multiple, non-complementary products. From the perspective of value appropriation; however, too much complementarity can be as bad as not enough.

Two products that are highly complementary are those in which the purchase of one always occurs with the purchase of the other. When the two products are always purchased in tandem, one product is just extension of the other. From a packaging perspective the two products may be separable, but from the consumer's purchasing perspective they are not. As a result, the ability of the firm to use one product to extend the market share of the other may be eroded. The ability of the firm to support higher price levels for the products independently may be diminished. The firm may be forced to expend energy in the production, sales, and support of both products without a commensurate increase in sales volume or revenue. The firm may not have a choice in supporting both products. Firms that sell only one will be at a competitive disadvantage.

The production-side and consumption-side synergies still exist for highly complementary products. Search and selling costs will be even lower for highly complementary products than for moderately complementary products. Demand uncertainty for the complementary products drops to zero. The two products may always be used together, increases the in-use synergies. The significant change to the firm as products become highly complementary is that the ability of the firm to appropriate the value created is reduced. Competitive pressures ensure that the value is appropriated by the consumer and not the producer. More formally:

Hypothesis-1b: Complementarity of a firm's products has a diminishing marginal impact on firm performance.

3 Espoused stack complementarity

The espoused industry architecture is often presented as an analog of the software stack. Industry analysts from firms such as IDC and Gartner publish reports describing the industry stack. In the early days of the software industry, a set of vertically integrated companies producing everything that a consumer needed (e,g., DEC, IBM and Wang) delivered the entire stack. As described by Andy Grove [27], somewhere around the late 80's a transition from vertical integration to horizontal layers occurred. As a result of this transition, the industry moved from single firms offering end-to-end services to modular clusters [2] or stacks populated by specialist firms.

The industry stack (see figure 1) divides activities into layers that are complementary to each other [18, 23, 28, 29]. Today, just as was the case during the era of vertical integration, firms can deliver products that support most (if not all) layers of the stack. For example, consumers can buy chipsets, assembled computers, operating systems (AIX), middleware (Websphere), applications (CRM) or services (Global consulting) from IBM.

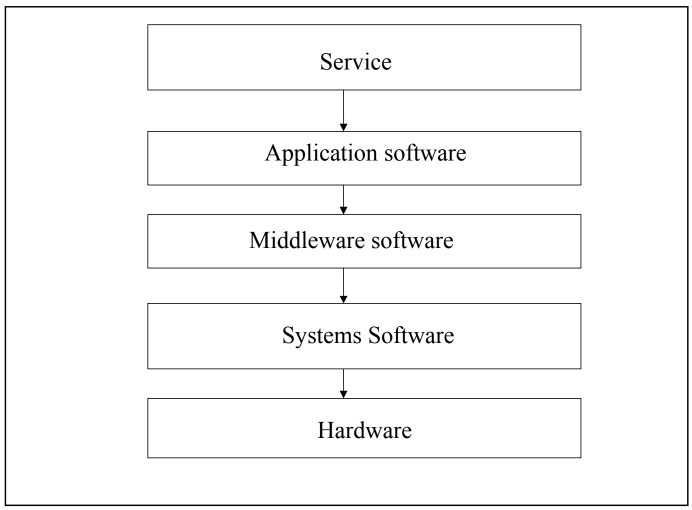


Figure 1: IT industry stack

The main difference in the era of stacks today is that IBM provides these products with looser coupling and with open interfaces between them. As a result, consumers of these services have the option to mix and match IBM's products with those provided by other vendors. In the earlier era this was not possible – a consumer had to pick a vendor and buy all required services from them.

According to Lou Gerstner, former CEO of IBM, most companies specialize in one or a few layers and rely on other companies to offer complementary components. Each of these components is layered above or below the other, and communicates through more or less standard interfaces, with closer layers being more related to each other than layers that are further apart in the stack. Each layer is dependent on the layer below to deliver the promised functionality. This arrangement works well for vendors as they have to simply focus on what they do best and leave the rest to other product vendors.

Lower layers and their components such as hardware and network services are often referred to as operating platforms and are fast becoming commodities. They have well defined interfaces with well defined terms of trade (prices). Firms build competencies on top of these lower layers by carefully selecting application packages and middleware packages and then launch business services on top of the application layer.

The idea of stacks is really just a simplification of the idea of modular systems. Modularity is a technique for managing complexity [2] long used by designers of many types of systems. However, modularity is not just a normative goal; Simon theorizes that modular systems are natures way of managing complexity by using nearly decomposable, stable subsystems [30]. This suggests that modularity in software systems was inevitable. It follows that, absent significant efforts by dominant firms, the industry would follow accordingly [1, 31]. The espoused industry stack as diagrammed above (Figure 1) is an abstraction of the modular system Simon anticipated.

The products within a cluster (i.e., module or stack layer) are either highly complementary or substitutes. When analysts consider the products that constitute the layers of the stack, they will group like things together (e.g., all database products) and often include those highly complementary products always purchased with the database product (e.g., database utilities). If we consider complementary products being linked together, we would anticipate no direct links between competing products within a cluster, but indirect links would embed all products within the cluster as complementary products interoperated with competing products.

The products in adjacent layers of the espoused architecture are interoperable and complementary, but less so than products within the same layer. The products are complementary by definition: it is the fact that the products are complementary that determines the ordering of the layers. Products more than one layer away from each other may still be complementary, but will be less so than those in adjacent layers. Database management systems and database applications are highly complementary. Database applications and operating systems are assumed to be less so.

Firms choosing to provide complementary products can provide products within a single stack layer, in adjacent stack layers, or in non-adjacent stack layers. This is a choice between providing highly complementary products, medium complementary products, or low or non complementary products. The logic explaining why the firm would choose to provide products in adjacent stack layers based upon product complementarity considerations is presented in the previous section. However, there are some additional issues the firm must wrestle with in making its choice.

Products are in the same stack layer because they are directly and indirectly associated (bound or linked) to the other products in the layer. Each product may be highly complementary to one or more other products. So, even if two products are not highly complementary to each other, they may belong together in the layer because their complements are highly complementary to each other. As a result, the firm selling multiple products within the same layer may not have pricing flexibility, due to competitive pressure, to appropriate a significant part of the value generated by the complementarities.

Firms selling products in complementary layers, however, have more flexibility to adjust pricing, product features, and the other elements of the product offering so as to appropriate this value than firms selling into a single layer. Complementary layers represent a combination of complementary products, which increases value, and fewer constraints, which lessens the difficulty in appropriating that value. Formally:

Hypothesis-2a: Espoused stack complementarity of a firm's products has a positive effect on firm performance.

Excessive complementarity can also have negative consequences. The argument against highly complementary products was presented for hypothesis-1b. To the extent that highly complementary stack layers consist of highly complementary products, the firms selling these highly complementary products may have diminished flexibility to set pricing levels and adjust packaging to appropriate the super-additive value produced by the complementary products. More formally:

Hypothesis-2b: Espoused stack complementarity of a firm's products has a diminishing marginal impact on firm performance.

4 Emergent stack complementarity

In the previous section we argued that the espoused stack – an abstraction of the underlying complexity of the industry – would help predict which pairs of products would be complementary and offer the possibility of superior returns. The espoused stack is a simpler, more aggregated view of the industry than the actual, emergent structure of the industry. In this section we continue with the actual complex system – the software industry – and suggest that understanding its emergent structure will give the firm a better understanding of the industry's behavior and enable it to select a more appropriate strategy than the one it would pick examining the espoused stack.

The espoused stack hides the imperfect modularity and interdependency between products that leads to interdependencies between firms as they cope with emerging technology. The espoused stack also hides the relationships between firms that may give rise to the specific organizational and industrial designs that are created [32]. There is a close relationship between the structure of designs and the economic structures – firms and markets – that emerge to realize them [2]. A set of firms and markets that support the evolution of a modular design is called a modular cluster [2]. The modular clusters are not, however, designed. They arise do to the interactions between the firms in the industry.

Each firm in the marketplace learns to huddle together as an industry and take cues from their fellow producers who face the same uncertainty of buyers and who need to offer differentiated products filling distinct niches. As these firms create and operate as actors, they socially construct their markets in response to uncertainty [33]. Since each firm within the industry will try to find a niche for itself by satisfying needs of particular sets of customers and differentiating from the rest, the architecture of the system will be reflected in the market structure of the industry [2, 33, 34]. We refer to this architecture as the emergent stack, even though a diagram of it would look like a network graph and not the linear abstraction presented earlier (see Figure 1).

The emergent stack is a dynamic abstraction that changes as new products are introduced and as product-use and purchase patterns shift. It is less a pedagogical abstraction and more an analytic artifact. Although the emergent and espoused stacks may differ considerably, the logic of complementarities, and their calculations, that drive our hypotheses is the same.

Hypothesis-3a: Emergent stack complementarity of a firm's products has a positive effect on firm performance

Hypothesis-3b: Emergent stack complementarity of a firm's products has a diminishing marginal impact on firm performance.

The emergent industry architecture is potentially more nuanced and intricate than the espoused, abstracted architecture. It reflects the actual behavior of producers and consumers. As a result, we expect greater predictive value when using the emergent stack in our analysis. More formally:

Hypothesis-4: Emergent stack complementarity of a firm's products is a better determinant of firm performance than the espoused stack complementarity of a firm's products.

5 Markets, Layers, and Market Boundaries

The task of identifying architecture in the software industry requires that we come up with a fine-grained definition of market segments (modular clusters) and the relationships between them. The challenge is coming up with a basis for deciding which transactions constitute one market segment and not another.

In order to meet this challenge we adopt the approach taken in highly related research on market definition and market boundaries. A market is a social structure resulting from complex and dynamic patterns of selling and buying among firms and customers, respectively [34-37]. Not only can markets and the boundaries that separate them be differentiated and determined by identifying differences in buying and selling patterns [36] but "to the extent that the producers of one commodity and producers of another have identical suppliers and identical consumers, they are competitors in the same market" [34] (p. 358). Competition among firms in a market exists because they have common consumers.

Researchers have since extended these concepts by emphasizing the bounded rationality of the actors that interact in a market. The search for competitive advantage on the part of a firm is the result of the actions of people in firms. Decision makers form models of the market system, with special attention to the actions of competitors and consumers, and (re)act accordingly [38, 39]. Simon [39, 40] has argued that:

"The bounded rationality of humans does not allow us to grasp the complex situations that provide the environments for our actions in their entirety. The first step in rational action is to focus attention on some specific (strategic) aspects of the total situation and to form a model of the situation in terms of the aspects that lie in that focus of attention (p. 37)"

In other words, markets should be defined with respect to the focus of attention of the actors and the models they use to compete [38, 39]. One result of incorporating bounded rationality into the definition of a market and determination of its boundary is that a market may span product, service, geographic, and other pertinent spaces [39]. Firms are not atomistic, single product actors that compete in a single niche. Firms may, in fact, compete in multiple markets. Moreover, as firms repeatedly interact with other firms and customers, a shared definition of what constitutes a market, among the markets' constituent parts and competitive relationships, emerges. It is this shared model of markets that defines the boundaries of a market and not an exogenous static definition per se [38].

For example, Burt and Carlton [36] conclude from their analysis of the network boundaries of 77 American markets that some of them can be combined into 7 classes. Brooks [39] analyzed patient flows to 97 hospitals from the zip codes contained within a 10 county area of San Francisco Bay area. Results indicate a firm's market did not coincide with their espoused markets - statutory boundaries such as cities or counties. Similar results are echoed by other researchers in a wide range of industries [41, 42].

Therefore, we base the identification of the software industry's emergent architecture on market segments that are defined by the pattern of company-specific product sales. Markets are not defined by the mere existence of products. Critical to market definition is knowledge of which companies sell which products and in what quantity. It is the shared attention by multiple actors (firms) on specific collections of products that defines markets. The specifics of how we did this in the software industry are the subject of the next section.

6 Methods

6.1 Sample and Data

The data for this study were collected and assembled by International Data Corporation (IDC). IDC data has been used extensively in academic research (e.g., [2, 43]), by industry analysts, and by the US Department of Justice's litigation with Microsoft. Consequently, IDC data are arguably one of the most exhaustive and complete data of the software industry. These data meet the general requirement in that they are at a finer level of granularity than the eventually defined markets [39].

The database tracks the sales of independent software vendors (ISV) that sell software products to consumers (i.e., people and other companies). Specifically, the IDC database contains over 260,000 unique ISV / operating system (OS) / product / geographic region/year/sales combinations spanning the years 1990 to 2002.

Overall, IDC tracks about 1200 ISVs, 10 operating systems (e.g., Windows and Unix), and 90 product categories (e.g., accounting, content management, database management), and 6 geographic regions (e.g., North America) as of 2002. Each ISV / OS / functional product offering / year / sales represents a unique transaction between ISV and customers. Each observation, with its associated sales, represents a competitive outcome stemming from the actions and reactions of firms to each other and with respect to consumers and their needs. Table 1 summarizes the number of firms, operating systems, and products in the sample.

--Insert Table 1 about here--

Within the subsequent sections, we refer to functional product categories simply as products. For example, two firms may sell different applications (e.g., Oracle and SQL Server), but for the purpose of this paper the firms sell the same product (i.e., database management). Because competitive applications are grouped together by IDC in the same product category, and we only examine product categories, the issue of competitive versus complementary products is avoided.

6.2 Dependent variable

We use sales growth to measure firm performance. Sales growth is a financial measure commonly used to assess firm performance in management studies [44]. Industry analysts also use sales growth as a key metric in valuing software firms [45]. We introduce a one-year lag in the performance measure to assess how the independent variables and controls in year t impact firm performance in year $t+1$. The range of the time variable t is 1991 to 2001 – we omit 1990 and 2002 from the sample because they are needed for control and dependent variables.

Firm sales growth$_{i,t+1}$. Firm i's sales growth is computed as the sum of the natural logarithm of the ratio of firm i's $t+1$ and t sales of product j weighted by the firm's percentage of sales in product j: $\sum_j p_{ij,t} \ln\left(x_{ij,t+1} / x_{ij,t}\right)$ where $x_{ij,t}$ is firm i's sales in product j for year t and p_{ij} is the proportion of firm i's sales from product j. Although firm growth can be computed differently, it is best denoted by the log of year $t+1$ and year t revenue ratio [46].

6.3 Independent variables

A key set of measures in this paper are the extent to which pairs of products are complementary. We argue that to the extent two products are highly complementary, firms that sell one product will also sell the other. To the extent that complementarity between a pair of products is low, a firm that sells one product may not necessarily sell the other. Table 2 summarizes the three complementarity operationalizations used in this paper. We also calculate two measures of stack complementarity in order to test for the existence of layers (modular clusters) and to test their predictive value on firm performance.

--Insert Table 2 about here--

Firm product complementarity$_{i,t}$. This variable measures the extent to which a firm offers complementary products. Firm product complementarity$_{i,t}$ considers all products as being distinct, as opposed to belonging to groups like the products in the espoused and emergent stacks. Computing this variable is a two step process. First, the complementarity for all product pairs is computed. Second, a specific firm's product complementarity is computed, based on the subset of the products offered by the firm, using the all-pairs product complementarities computed in the first step.

To compute complementarity for all product pairs, we reference recent studies that focus on within-industry diversification [47-49]. We compute similarity coefficients $r_{jj'}$ for every pair of products j and j', over all firms and all OS, in year t using the Sohn [13] similarity metric. Specifically, $r_{jj'}$ is $r_{jj',t} = \sum_k p_{jk,t}\left(\frac{\sum_i x_{ijk,t}\ \min\left(x_{ijk,t}, x_{ij'k,t}\right)}{\sum_i x_{ijk,t}^2}\right)$. where $x_{ijk,t}$ denotes firm i's sales of product j, on OS k, in year t. The term $p_{jk,t}$ represents the proportion of product j's sales on OS k in year t. The range of $r_{jj',t}$ is zero to one. The metric is zero if two products exhibit no complementarities and one if they are perfectly complementary.

This approach has been used to infer underlying resource similarities and complementarities of industries or products from the sales distributions of populations of firms across industries or products [34, 36, 39]. Lemelin [50] argues that complementary in use of products is one way to recognize the relatedness of products and markets. Full details regarding the computation of the similarity metric $r_{jj',t}$ can be found in Appendix A.

A specific firm's product complementarity is the normalized sum of the similarity between every pair of firm i's products weighted by the firm's proportion of sales in those products or $\frac{\sum_{jj'|j\neq j'} r_{jj',t}(p_{ij,t}+p_{ij',t})}{\left|N_{i,t}^{product}\right|-1}$, where $r_{jj',t}$ is described above, $p_{ij,t}$ is the proportion of firm i's sales from product j, and $N_{i,t}^{product}$ is the number of products offered by the firm in year t. We divide by $N_{i,t}^{product}$ to normalize and remove double-counting bias [21].

Firm emergent stack complementarity$_{i,t}$. This variable measures the extent to which a firm offers products across complementary layers (clusters) of the emergent stack. The previous variable, *Firm product complementarity*$_{i,t}$, ignores industry decomposition. However, our argument is that industries self-organize into non-random typologies in order to manage complexity. Through the collective behaviors of firms, industries form stacks – interdependent clusters or layers that contain highly complementary products. Computing this variable is a two step process. First, the products are grouped into layers by using a clustering algorithm. These clusters represent the emergent, actual layering of products (note, these layers may not be linearly organized as shown in Figure 1). This step also results in a set of values $\omega_{jj',t}$ that denotes the degree of emergent complementarity between layers. Second, a specific firm's emergent stack complementarity is computed, based on the subset of the products offered by the firm, using the all-pairs emergent complementarities computed in the first step.

To group products into layers and compute complementarity between layers, we create a graph formed from the set of products offered by all firms in year t (i.e., nodes) and the $r_{jj',t}$ defined previously as the arcs between the nodes. We transform this directed graph into an undirected graph by first summing the parallel arcs and then renormalizing edge weights.

A goal of grouping is to find a partition of the nodes that maximizes between subgroup differences and minimizes within subgroup differences. Grouping N elements into M (≤N) disjoint groups is *NP-hard* [51]. The computation to find the best partition quickly becomes intractable. Other methods involves spectral methods (e.g., [52]), semi-definite programming relaxations (e.g., [53]), and non-deterministic approximation methods such as genetic algorithms and neural networks.

To create the emergent stack layer definitions we use recent advances in graph partitioning to arrive at a dynamic clustering of products. Recently, interest has been renewed in flow-based cut clustering methods. These algorithms have enjoyed success at grouping similar search terms for use in internet search engines. These partitioning algorithms are defined for undirected graphs; hence the need to transform our weighted directed graph into a weighted undirected graph. The partitioning algorithm is based on Gomory-Hu [54] min-cut spanning trees and is described by Flake [55].

To summarize, an artificial sink node is added to the graph, the min-cut tree is computed, the artificial node and its associated edges are removed from the min-cut tree, and the remaining disconnected sub-graphs, if any, correspond to the distinct groups or layers [55]. This partitioning algorithm requires as input a weight α, that depending on its value, determines the number of distinct sub-groups that the graph will be partitioned into. The number of sub-groups is between 1 and N where N is the number of distinct products in year *t*. Our custom algorithm generates all the partition structures (e.g., {1,2 | 3, 4}, {1| 2,3 |4}, and {1,2,3 | 4}) that can be generated by varying *α*, and for the purposes of arriving at a partition for use in our computation, we chose the one that maximizes between subgroup differences and minimizes within subgroup differences. Our partitioning "goodness" criterion is similar to the one described in Freidman [56].

Once we have the groupings, the complementarity or $\omega_{jj',t}$ between layers *j* and *j′* in year *t* is calculated by computing the average of the edges between the products contained in the two layers. Table 3 presents the partitioning results for the sample years. Figure 2 depicts the product groupings for the products offered by in year 1991 and the complementarity between the layers. In the figure all the relational database management products and the object relational database management product are grouped with the third generation languages used to create applications. Since they represent databases and tools to create applications, the grouping makes sense. Similarly, operating systems and utilities are grouped together. According to the figure there is high complementarity (0.79) between the two groups. The high complementarity is expected because database management systems work directly with the operating systems. On the other hand, accounting systems, human resource systems and inventory management tools do not interact with the operating systems directly and hence have a low complementarity score (0.03) with the OS layer.

--Insert Table 3 and Figure 2 about here--

A firm's emergent stack complementarity is calculated in an analogous fashion to the *Firm product complementarity*$_{i,t}$ variable. However, we compute complementarity with respect to the layers, and the products contained within them, instead of discrete products. A firm's emergent stack complementarity is the normalized sum of the similarity between every pair of firm *i*'s supported layers weighted by the firm's proportion of sales from the layer or $\frac{\sum_{jj'|j\neq j'}\omega_{jj',t}(p_{ij,t}+p_{ij',t})}{\left|N_{i,t}^{layer}\right|-1}$. The variable $p_{ij,t}$ denotes the proportion of firm *i*'s sales from layer *j*, and $N_{i,t}^{layer}$ is the number of layers supported by the firm in year *t*. We divide by $N_{i,t}^{layer}$ to normalize and control for double counting bias. All of the firm's products that are identified as belonging to a stack layer are combined by summing their sales.

Firm espoused stack complementarity$_{i,t}$. This variable measures the extent to which a firm offers products across complementary layers of the espoused stack. This variable is computed in two steps. First, the products are grouped into layers by a panel of experts. As is evident, the key difference between this measure of stack complementarity and the previous measure is the method used to create disjoint product groups. In addition to the product groupings, this step results in a set of values $\tau_{jj',t}$ that denote the degree of espoused complementarity between layers. Second, a firm's specific espoused stack complementarity is computed, based on the subset of products offered by the firm, using the all-pairs espoused stacks complementarity computed in the first step.

To reiterate, the key difference between the two stack complementarity measures is the method used to identify the layers. For the espoused stack layers we used a panel of experts to put different products into layers. Each expert was provided with a list of products found in the IDC database. They were then asked to group products into layers. The experts grouped the products into nine distinct groups (layers). We then used these nine layers to compute the firm i's sales by layer j in time period t.

The complementarity or $\tau_{jj',t}$ between layers j and j' in year t is calculated by computing the average of the $r_{jj',t}$ between the products contained in the two layers. Table 4 shows the results of this espoused stack complementarity product grouping.

--Insert Table 4 about here--

A firm's espoused stack complementarity is the normalized sum of the similarity between every pair of firm i's supported layers weighted by the firm's proportion of sales from the layer or $\dfrac{\sum_{jj'|j\neq j'} \tau_{jj',t}(p_{ij,t}+p_{ij',t})}{\left|N_{i,t}^{layer}\right|-1}$. The variable $p_{ij,t}$ denotes the proportion of firm i's sales from layer j, and $N_{i,t}^{layer}$ is the number of layers supported by the firm in year t. We divide by $N_{i,t}^{layer}$ to normalize and control for double counting bias. All of the firm's products that are identified as belonging to a stack layer by the experts are combined by summing their sales.

6.4 Control variables

To address alternative explanations for our findings, we include key controls that are likely to have a bearing on the dependent variable. To minimize potential endogeneity concerns, we also include controls that are likely to influence both independent and dependent variables of the study.

The diversity of a firm's operations (i.e., the extent to which a firm's products are spread across distinct OS and geographic regions) may impact firm performance [57]. We control for diversity by computing the entropy measure of diversification across platforms and geographic regions. *Platform diversity*$_{i,t}$ for firm i in year t is $\sum_k p_{ik,t} ln(1/p_{ik,t})$ where $p_{ik,t}$ is the proportion of a firm i's sales in platform k in year t. *Geographic diversity*$_{i,t}$ for firm i in year t is $\sum_l p_{il,t}\ ln(1/p_{il,t})$ where $p_{il,t}$ is the proportion of a firm i's sales in geographic region l in year t.

A firm's dominance in products, OS, or geographic regions may impact its growth. Dominant firms may also have more resources to offer more products, support more stacks layers, or both. *Platform dominance*$_{i,t}$ for firm i in year t is the weighted sum of a firm's proportion of OS k or $\sum_k p_{ik,t}(x_{ik,t}/\sum_i x_{ik,t})$. The term $p_{ik,t}$ denotes the proportion of a firm i's sales in OS k in year t, $x_{ik,t}$ denotes the firm's total sales in OS k, and the sum $\sum_i x_{ik,t}$ denotes the total sales in OS k for all firms. *Product dominance*$_{i,t}$ for firm i in year t is the weighted sum of a firm's proportion of product j or $\sum_j p_{ij,t}(x_{ij,t}/\sum_i x_{ij,t})$. The term $p_{ij,t}$ denotes the proportion of a firm i's sales in product j in year t, $x_{ij,t}$ denotes the firm's total sales in product j, and the sum $\sum_i x_{ij,t}$ denotes the total sales in product j for all firms. *Geographic dominance*$_{i,t}$ for firm i in year t is the weighted sum of a firm's proportion its sales in region l or $\sum_l p_{lj,t}(x_{il,t}/\sum_i x_{il,t})$. The term $p_{il,t}$ denotes the proportion of a firm i's sales in region l in year t, $x_{il,t}$ denotes the firm's total sales in region l, and the sum $\sum_i x_{il,t}$ denotes the total sales in region l for all firms.

Positioning across growing or declining markets, platforms, and geographic regions can influence the sales growth of a firm. Thus, we control for sales growth in the platforms, markets, and geographic regions supported by the firm. *Platform growth*$_{i,t}$ for firm i in year t is the sum, over all platforms, of the growth of platform k (i.e., $\text{Growth}_{k,t}$) between year t-1 and t multiplied by the firm's proportion of sales in that platform (i.e., $\sum_k p_{ik,t}\ Growth_{k,t}$) where $p_{ik,t}$ is the proportion of firm i's sales in platform k in year t, $Growth_{k,t} = ln(x_{k,t} / x_{k,t-1})$, and $x_{k,t}$ is the total sales of platform k. Similarly, *Product Growth*$_{i,t}$ for firm i in year t is the sum, over all markets, of the growth of product market j (i.e., $\text{Growth}_{j,t}$) multiplied by the firm's proportion of sales in that product market (i.e., $\sum_j p_{ij,t}$ *Growth*$_{j,t}$). Finally, *Geographic Growth*$_{i,t}$ for firm *i* in year *t* is the sum, over all geographic regions, of the growth of region *l* (i.e., Growth$_{l,t}$) multiplied by the firm's proportion of sales in that region (i.e., $\sum_l p_{il,t}$ *Growth*$_{l,t}$).

Firm size can influence the performance of firms. Large firms offer more products [58], have more synergy exploitation opportunities, and suffer more from managerial diseconomies [59]. We control for firm *i*'s size in year *t* by taking the natural logarithm of the firm's total sales.

Firm age is associated with within-industry diversification strategies and the performance of firms [49]. Thus, we control firm *i*'s age in year *t*. We measure firm age by counting the number of years since the firm's founding or the first year in which the firm generated sales in the software industry. We compute the natural logarithm of this measure.

A firm's initial performance advantages can influence subsequent performance of the firm as well. Thus, we control for a firm's growth, as well as other unobserved firm heterogeneity, by including a lagged dependent variable, *Firm sales growth*$_{i,t}$, in the model. Finally, we include a set of indictor variables to control for industry and period specific effects. Summary statistics of our sample can be found in Table 5.

--Insert Table 5 about here--

6.5 Statistical Method and Analysis

We test the hypotheses using a cross sectional time series or panel design. The sample contains 4,392 distinct firm-years with approximately 5.0 observation years per firm. The design repeatedly measures firm performance and covariates, which includes a lagged performance measure. Under these conditions, ordinary least squares (OLS) may result in biased and inefficient estimates [60].

We thus use the generalized estimating equations (GEE) approach [61, 62] that has been used in prior studies [60, 63]. GEE is a flexible estimation procedure that addresses within firm correlation and heterogeneity and thus results in more efficient and unbiased parameters than OLS. Specifically, GEE estimators are asymptotically normal and consistent given an arbitrary correlation among observations [61, 62]. Because of its flexibility, a set of options must be specified prior to performing estimations. We use a Gaussian distribution for the dependent variable, an identity link function and an unstructured working correlation matrix. We also performed estimations using an exchangeable working correlation matrix, which denotes an equal correlation model and is equivalent to a random effects estimation with consistent results. We report the more conservative results obtained from the unstructured working matrix; we use a sandwich variance estimator for correcting standard errors.

7 Results

Tables 6 and 7 present Stata coefficient estimations from the GEE regression for sales growth. We estimate twelve models. Model 1 is the baseline model and only contains controls. Models 2, 3, and 4 build upon Model 1 and independently add the three main effects firm complementarity variables. Models 5, 6, and 7 add the curvilinear quadratic effects to Models 2 through 4. Models 8 and 9 add in tandem two of the three firm product complementarity variables to the baseline Model 1. Models 10 and 11 build upon Models 8 and 9 and add in tandem the two of the three complementarity main effects and associated quadratic effects. Model 12 in the final model estimated in this paper. It contains all the main and quadratic complementarity operationalizations.

--Insert Tables 6 and 7 about here--

All twelve models are individually significant. With the exception of Model 2, the additional main or quadratic effects variables significantly increase model fit. H_1 hypothesizes that a firm's product complementarity will positively impact firm performance. The coefficients in Model 5 for the direct effect (0.258, $p<0.01$) is significant, suggesting support for H_{1a}. H_{1b} hypothesizes a curvilinear relationship between a firm's product complementarity and firm performance. The directionality of the quadratic term in Model 5 (-0.317, $p<0.05$) supports H_{1b}. The significance of both coefficients is lost in Models 10 and 12, suggesting that understanding complementarities based on the emergent stack is superior to looking only at product complementarities and that understanding stack-based complementarities is not improved by also understanding product complementarities.

H_2 hypothesizes that a firm's espoused stack complementarity will positively impact firm performance. The coefficient in Model 3 (0.110, $p<.01$) is significant but only weakly significant in Model 6 (0.170, $p<.10$). Hence, the results suggest weak support for H_{2a}. H_{2b} hypothesizes a curvilinear relationship between a firm's espoused product complementarity and firm performance. The directionality of the quadratic term in Model 6 (-0.115) is appropriate but it is not significant. Furthermore, the quadratic effect coefficient is also not significant in Models 11 or 12. Hence, the results suggest no support for H_{2b}.

H_3 hypothesizes that a firm's emergent stack complementarity will positively impact firm performance. The coefficient in Model 4 (0.123, p<.01) is significant and positive. With the exception of Model 9, the coefficients for emergent stack complementarity remains similar in magnitude and constant in direction even in the presence of other main and quadratic effect variables. The addition of emergent stack complementarity also increases overall model fit at p<.10 or better. The results thus suggest support for H_{3a}. H_{3b} hypothesizes a curvilinear relationship between a firm's emergent product complementarity and firm performance. The quadratic term in Model 7 (-0.551) is significant (p<.01). The quadratic term is significant in all models and increases overall model fit. This coefficient remains similar in magnitude and constant in direction. Based on the pattern of the results among the twelve models, the joint impact of main and quadratic emergent stack complementarity on firm performance is robust and consistent. The results thus suggest support for H_{3b}.

We finally test the greater explanatory power of emergent stack complementarity (H_4) over espoused stack complementarity. In Model 12, emergent stack complementary is significant, similar in magnitude to the coefficients in Model 7, 10, and 11, and in the expected direction. In fact, the coefficients for product and espoused complementarity are not significant in Model 12. Overall, the addition of product and espoused complementarity covariates do not increase the fit (cf. Model 7 and Model 10, and Model 7 and Model 11) while the opposite assertion can be made. Furthermore, the main and quadratic emergent stack complementarity is consistently significant while the other measures of complementarity are not significant. These results suggest support for (H_4).

Figure 3 is a plot of emergent stack complementarity and firm sales growth using the estimated coefficients from Model 12. Firm sales growth is expressed as the ratio of sales in period $t+1$ to sales in period t. The baseline is equal to 1.0 or no change in firm sales for two consecutive years. The positive impact of firm emergent stack complementarity increases but after a critical value begins to decrease. In fact, for high degrees of emergent stack complementarity, the impact on firm growth is negative.

--Insert Figure 3 about here--

8 Contributions and Limitations

This study makes a contribution by synthesizing the economic theory of complementarities with the resource-based view of diversification and design science computational methods. While the resource-based view of diversification recognizes production and consumption-side synergies, it does not use the theory of complementarities [26]. In addition to introducing complementarities, we also distinguish between espoused and emergent complementarity and confirm the existence and importance of emergent industry architecture - stacks.

However there are limitations to this work. We have summarized these under two headings: complementarity and clustering techniques.

Operationalization of complementarity

Complementarity can be based upon espoused (e.g., design-time) or emergent (e.g., customer buying patterns or hedonic measures) characteristics. We operationalize complementarity based on customer buying behavior. If customers purchase a couple of products together we assume that they are part of a cluster of complementary products. This notion of complementarity is based on usage and not based on design or hedonic measures.

If complementarity were based on design, we would identify complementors for each product based on other products that work on top of it (using the stack analogy). For example, we would look at the database management system from Oracle and see which report writers actually work on top of it. If BusinessObjects has built in application program interfaces that work with Oracle, we would declare the two products to interoperate, which we take as a precondition for producing super-additive value. Going from a level of interoperation to a level of complementarity, however, is very difficult because it requires us to base a market valuation on a specific design feature. This data is also very hard to collect for two reasons. One, for every product we have to talk to the developer/engineering team to find out if they have specific hooks built for exchanging data and functionality with other products. Second, most companies would claim that their products could be made interoperable with any other product given sufficient time.

Hedonic measures might offer different insights than the analysis we performed. The result of such an analysis would help value interoperability among combinations of products. In order to calculate the value of interoperability one would collect data on the prices of different combinations of goods with differing levels of interoperability. One would then perform regressions in order to estimate the value of interoperability without having recourse to a market price.

Clustering algorithm

In our analysis, we treat the dependencies between firms as a network, with the nodes and links representing firms and dependencies, respectively. The problem of finding a grouping of nodes, such that all nodes within a group are powerfully connected (homogeneity) and weakly connected to others (separation), remains an NP-complete problem since Karp [51] identified it as such. Social network analysis methods have provided polynomial-like algorithms, mostly applicable to boolean graphs (e.g., LS sets: Lawler (1973), Lambda sets: [64]), that have been implemented in various software packages (e.g., Ucinet)[1]. Freeman [66], however, has argued that linkages are mostly valued although many studies dichotomize them. When a grouping needs to be performed on weighted graphs, much of the research to date has been directed towards finding appropriate relaxations of this NP-hard optimization problem [67]. For example, Goemans and Williamson [53] provide a relaxation and with the use of semidefinite programming solve the problem of findings two groupings of vertices (i.e., MAXCUT and MAX DICUT). Other related methods include spectral methods and algorithms, such as eigenvalue optimization (e.g., [68]) and approximations based on neural networks and its variants.

We use in this study the clustering algorithm proposed by Zhou [69, 70]. This method is applicable to the problem of identifying markets and market boundaries because the algorithm: (1) is decisive [65], permitting a hierarchical decomposition that is indicative of competitive markets [38, 39], (2) incorporates a decomposition stopping criteria such that clusters are not decomposed to individual vertices, (3) is applicable to directed weighted graphs, (4) is an appropriate relaxation, (5) is computationally tractable, and (6) provides acceptable groups with quantifiable behaviors [69, 70]. While the technique that we used is robust, it is still an approximation and not the optimal clustering.

---

[1] An introduction to clustering methods can be found in 65. Everitt, B.S., S. Landau, and M. Leese, *Cluster Analysis*. 2001: Arnold Publishers.

9 Discussion and Conclusions

The concept of complementarity has been used in the academic literature as a way to explain strategic initiatives and firm performance. However, very little has been done to measure it. In this paper we introduced recent developments in computational methods to measure complementarity.

We used the same computational methods to identify modular clusters, the structure of the software industry, and associate a value for doing so – firm performance. In doing so, we confirm Simon's theorizing [7] and provide statistical support for the insights of Baldwin and Clark [2]. Complex systems seem to self-organize into nearly decomposable, stable subsystems, which have emergent properties. Understanding this structure – the architecture – may not be required in order to succeed, but operating in accordance with it seems to be a worthwhile endeavor [1, 31].

We looked at the structure of the prepackaged software industry. We argued that because software is an example of systems based competition, with significant costs and benefits associated with providing complementary products, firms that offered complementary products would benefit over those firms that didn't. We also argued that offering complementary products across market cluster boundaries would enable the firm to more readily appropriate the value created by complementary products. The implication is that firms need to understand both the nature of complementary products and the emergent the structure of the industry.

We used consumer purchasing behavior to determine complementary products and two different mechanisms for describing the clustering of firms that form the software industry. The first mechanism used a panel of experts to create an espoused stack. The second mechanism used dynamic clustering to create an emergent stack.

We showed that the industry does cluster, that firms that sell complementary products perform better than those that don't, and those that sell to complementary stack layers (modular clusters) benefit even more. There is, however, a threshold to the benefits. Beyond a certain limit the returns from complementarity begin to show diminishing returns. This is because the benefits of complementarity can be competed away most easily when the products are within the same modular cluster or when two modular clusters are about to merge.

While the setting that was used in the study was the packaged software industry, we believe that our findings would apply to other settings where complementarities are present. Some examples are the video game, computer, and the automotive industries.

In our work we infer architecture through market boundaries. In future work we will collect data on investments that a firm makes in making sure that the different products interoperate. For example, do they invest in the development of application program interfaces (API), or allocate resources to manage technology alliance relationships.

8 References

1. Henderson, R.M. and K.B. Clark, *Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms.* Administrative Science Quarterly, 1990. **35**(1): p. 9-30.
2. Baldwin, C. and K. Clark, *Design Rules: The Power of Modularity*. 2000, Cambridge, MA: MIT Press.
3. Christensen, C.M. and J.L. Bower, *Customer power, strategic investment, and the failure of leading firms.* Strategic Management Journal, 1996. **17**(3): p. 197-218.
4. Tushman, M.L. and P. Anderson, *Technological Discontinuities and Organizational Environments.* Administrative Science Quarterly, 1986. **31**(3): p. 439.
5. Messerschmitt, D. and C. Szyperski, *Software Ecosystems: Understanding an Indispensable Technology and Industry*. 2003, Cambridge, MA: The MIT Press. 424.
6. Mintzberg, H., B.W. Ahlstrand, and J. Lampel, *Strategy safari : a guided tour through the wilds of strategic management*. 1998, New York: Free Press. x, 406 p.
7. Simon, H., *The architecture of complexity*, in *Managing in the modular age : architectures, networks, and organizations*, R. Garud, A. Kumaraswamy, and R.N. Langlois, Editors. 2003, Blackwell Publishing: Malden, MA. p. 411.
8. Schilling, M., *Toward a General Systems Theory and its Application to Inter-firm Product Modularity.* Academy of Management Review, 2000. **25**(2): p. 312-334.
9. Bresnahan, T. *New Modes of Competition: Implications for the Future Structure of the Computer Industry*. 1998.
10. Ahuja, G., *Collaboration Networks, Structural Holes, and Innovation: A Longitudinal Study.* Administrative Science Quarterly, 2000. **45**(3): p. 425-455.
11. Powell, W., K.W. Koput, and L. Smith-Doerr, *Interorganizational collaboration and the locus of innovation: networks of learning in biotechnology.* Administrative Science Quarterly, 1996. **42**(1): p. 116-145.
12. Powell, W.W., et al., *Network Position and Firm Performance: Organizational Returns to Collaboration in the Biotechnology Industry*, in *Networks In and Around Organizations*, S. Andrews and D. Knoke, Editors. 1999, JAI Press: Greenwich, CT. p. 129-160.
13. Sohn, M.-W., *Distance and cosine measures of niche overlap.* Social Networks, 2001. **23**(2): p. 141-165.
14. Economides, N. and S. Salop, *Competition and integration among complements and network market structure.* Journal of Industrial Economics, 1992. **40**(1): p. 105-123.
15. Farrell, J. and G. Saloner, *Standardization, compatibility, and innovation.* Rand Journal of Economics, 1985. **16**(1): p. 70-83.
16. Katz, M. and C. Shapiro, *Network externalities, competition and compatibility.* American Economic Review, 1985. **75**(3): p. 424-440.
17. Milgrom, P. and J. Roberts, *An Economic-Approach to Influence Activities in Organizations.* 1988. **94**: p. S154-S179.
18. Gao, L. and B. Iyer, *Using Software Stacks to Explain Complementarities: The Case of Mergers and Acquisitions in the Software Industry. (forthcoming).* Journal of Management Information Systems, 2006. **23**(2): p. 121-149.
19. Sengupta, S., *Some approaches to complementary product strategy.* Journal of Product Innovation Management, 1998. **15**(4): p. 352-367.
20. Farjoun, M., *Beyond industry boundaries: Human expertise, diversification and resource-related industry groups.* Organization Science, 1994. **5**(2): p. 185-199.
21. Robins, J. and M.F. Wiersema, *A resource-based approach to the multibusiness firm: Empirical analysis of portfolio interrelationships and corporate financial performance.* Strategic Management Journal, 1995. **16**(4): p. 277-299.

22. Davis, R. and L.G. Thomas, *Direct estimation of synergy: A new approach to the diversification-performance debate.* Management Science, 1993. **39**(11): p. 1334-1346.
23. Gawer, A. and R. Henderson, *Platform owner entry and innovation in complementary marekets: Evidence from Intel*. 2005, Nation Bureau of Economic Research: Cambridge, MA. p. 36.
24. Teece, D.J., *Profiting from Technological Innovation: Implications for Integration, Collaboration, Licensing and Public Policy.* Research Policy, 1986. **15**(6): p. 285-305.
25. Bakos, Y. and E. Brynjolfsson, *Bundling information goods: Pricing, profits, and efficiency.* Management Science, 1999. **45**(12): p. 1613-1630.
26. Tanriverdi, H. and C.-H. Lee, *Production-side and consumption-side synergies in the software industry: Implications for sales growth and market share*, in *Academy of Management*. 2005: Honolulu, HI.
27. Grove, A.S., *Only the paranoid survive : how to exploit the crisis points that challenge every company and career*. 1st ed. 1996, New York: Currency Doubleday. xii, 210 p.
28. Silva, L. and B. Iyer. *Using Software Stacks to Explain Complementarities: The Case of Mergers and Acquisitions in the Software Industry*. in *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS-39)*. 2006. Koloa, Kauai HI.
29. West, J. and J. Dedrick, *Innovation and Control in Standards Architectures: The Rise and Fall of Japan's PC-98.* Information Systems Research, 2000. **11**(2): p. 197-216.
30. Simon, H.A., *The Sciences of the Artificial*. Third ed. 1996 [1969], Cambridge, MA: The MIT Press. 231.
31. Alexander, C., *Notes on the synthesis of form*. 1964, Cambridge: Harvard University Press. 216.
32. Kidder, T., *The soul of a new machine*. Large print ed. 1981, Thorndike, Me.: Thorndike Press. 471 p.
33. White, H., *Markets from Networks: Socioeconomic Models of Production*. 2002, Princeton, NJ: Princeton University press.
34. Burt, R.S., *The stability of American markets.* American Journal of Sociology, 1988. **94**(2): p. 356-395.
35. White, H., *Where do markets come from?* American Journal of Sociology, 1981. **87**: p. 517-547.
36. Burt, R.S. and D.S. Carlton, *Another Look at the Network Boundaries of American Markets.* American Journal of Sociology, 1989. **95**(3): p. 723-753.
37. Burt, R.S., *Structural Holes*. 1992, Cambridge, MA: Harvard University Press.
38. Porac, J., et al., *Rivalry and the industry model of Scottish knitwear producers.* Administrative Science Quarterly, 1995. **40**(203-227).
39. Brooks, G., *Defining Market Boundaries.* Strategic Management Journal, 1995. **16**(7): p. 535-549.
40. Simon, H.E., *Organizations and markets.* Journal of Economic Perspectives, 1991. **5**: p. 25-44.
41. Porac, J.P., H. Thomas, and C. Baden-Fuller, *Competitive groups as cognitive communities: The case of Scottish knitwear manufacturers.* Journal of Management Studies, 1989. **26**(4): p. 397-416.
42. Cool, K., I. Dierickx, and D. Jemison, *Business strategy, market structure and risk-return relationships: A sturctural approach.* Strategic Management Journal, 1989. **10**(6): p. 507-522.
43. Bothner, M., *Competition and Social Influence: The Diffusion of Sixth-Generation Processor in the Global Computer Industry.* American Journal of Sociology, 2003. **108**(6): p. 1175-1210.
44. Ramanujam, V. and P. Varadarajan, *Research on Corporate Diversification: A Synthesis.* Strategic Management Journal, 1989. **10**: p. 523 – 551.
45. Cusumano, M., *The business of software: What every manager, programmer, and entrepreneur must know to thrive and survive in good times and bad*. 2004, New York: Free Press.
46. Carroll, G.R. and M.T. Hannan, *The Demography of Corporations and Industries*. 2000, Princeton, NJ: Princeton university Press.
47. Cottrell, T. and B.R. Nault, *Product Variety and Firm Survival in the Microcomputer Software Industry.* Strategic Management Journal, 2004. **25**(10): p. 1005-1026.
48. Li, S. and R. Greenwood, *The effect of within-industry diversification on firm performance: Synergy creation, multi-market contact and market structuration.* Strategic Management Journal, 2004. **25**(12): p. 1131-1153.

49. Stern, I. and A.D. Henderson, *Within-business diversification in technology-intensive industries.* Strategic Management Journal, 2004. **25**(5): p. 487-505.
50. Lemelin, A., *Relatedness in the patterns of interindustry diversification.* Review of Economics and Statistics, 1982. **64**(4): p. 646-657.
51. Karp, R.M., *Reducibility among combinatorial problems*, in *Complexity of Computer Computations*, M.a. Thatcher, Editor. 1972, Plenum Press. p. 85-103.
52. Kannan, R., S. Vempala, and A. Veta, *On clusterings: Good, bad and spectral.* Journal of the ACM 2004. **51**(3): p. 497 - 515.
53. Goemans, M.X. and D.P. Williamson, *Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming.* Journal of the ACM,, 1995. **42**: p. 1115--1145.
54. Gomory, R.E. and T.C. Hu, *Multi-terminal network flows.* SIAM Journal of Applied Mathematics, 1961. **9**(4): p. 551-570.
55. Flake, G.W., R.E. Tarjan, and K. Tsioutsiouliklis, *Graph Clustering and Minimum Cut Trees.* Internet Mathematics, 2004. **1**(3): p. 355-378.
56. Friedman, H.P. and J. Rubin, *On Some Invariant Criteria for Grouping Data.* Journal of the American Statistical Association, 1967. **62**(320): p. 1159-1178.
57. Palepu, K., *Diversification Strategy, Profit Performance and the Entropy Measure.* Strategic Management Journal, 1985. **6**(3): p. 239-255.
58. Sorenson, O., *Letting the market work for you: An evolutionary perspective on product strategy.* Strategic Management Journal, 2000. **21**(5): p. 577-592.
59. Nayyar, P.R., *Performance effects of information asymmetry and economies of scope in diversified service firms.* Academy of Management Journal, 1993. **36**(1): p. 28-57.
60. Bae, J. and M. Gargiulo, *Partner substitutability, alliance network structure, and firm profitability in the telecommunications industry.* Academy of Management Journal, 2004. **47**(6): p. 843-859.
61. Liang, K.Y. and S. Zeger, *Longitudinal data analysis using generalized linear models.* Biometrika, 1986. **73**: p. 13-22.
62. Zeger, S., K. Liang, and P. Albert, *Models for longitudinal data: a generalized estimating equation approach.* Biometrika, 1988. **44**(1049-1060).
63. Katila, R. and G. Ahuja, *Something Old, something new: A Longitudinal Study of search behavior and new product introduction.* Academy of Management Journal, 2002. **45**(6): p. 1183-1194.
64. Borgatti, S.P., M.G. Everett, and P. Shirey, *LS sets, lambda sets and other cohesive subsets.* Social Networks, 1990. **12**: p. 337-357.
65. Everitt, B.S., S. Landau, and M. Leese, *Cluster Analysis*. 2001: Arnold Publishers.
66. Freeman, L.C., *The Sociological Concept of "Group": An Empirical Test of Two Models.* The American Journal of Sociology, 1992. **98**(1): p. 152-166.
67. Ahuja, R.K., T.L. Magnanti, and J.B. Orlin, *Network Flows: Theory, Algorithms and Applications*. 1993, NJ: Prentice Hall.
68. Girvan, M. and M.E.J. Newman. *Community structure in social and biological networks*. in *National Academy of Science*. 2002.
69. Zhou, H., *Network landscape from a Brownian particle's perspective.* Physical Review E, 2003. **67**: p. 041908.
70. Zhou, H., *Distance, dissimilarity index, and network community structure.* Physical Review E, 2003. **67**: p. 061901.
71. Wasserman, S. and K. Faust, *Social Network Analysis: Methods and Applications*, ed. M. Granovetter. 1994, Cambridge: Cambridge University Press.
72. Pianka, E.R., *Evolutionary ecology*. 1973, New York,: Harper & Row. viii, 356 p.

TABLE 1: IDC descriptors

Year	Number (Firms)	Number (Platforms)	Number (Products)
1990	1.00	1.00	1.00
1991	1.13	1.00	1.00
1992	1.85	1.00	1.11
1993	2.55	1.20	1.95
1994	3.11	1.20	2.32
1995	4.04	1.20	2.79
1996	5.62	1.60	3.26
1997	7.85	1.60	3.74
1998	10.23	1.80	4.11
1999	12.61	1.80	4.26
2000	14.25	1.80	4.47
2001	16.00	1.80	4.47
2002	16.06	1.80	4.47

Scaled with respect to 1990 numbers.

TABLE 2: Complementarity operationalization

Type	Symbol	Key characteristics
Product	$r_{jj',t}$	Products are distinct. The complementarity for products j and j' ($j \neq j'$), computed over all firms i and all platforms k in year t, is $\sum_k p_{jk,t}\left(\frac{\sum_i x_{ijk,t}\min\left(x_{ijk,t},x_{ij'k,t}\right)}{\sum_i x_{ijk,t}^2}\right)$. High $r_{jj',t}$ denotes high complementarity between products.
Emergent	$\omega_{jj',t}$	Products are related and are grouped into disjoint clusters or layers. Layers for year t are formed by using a min-cut spanning tree algorithm on the graph based on the $r_{ll',t}$ (i.e., arcs for all products $l \neq l'$) between products (i.e., nodes). The complementarity for layers j and j' ($j \neq j'$) in year t is the average of $r_{ll',t}$ between the nodes in each layer. Like $r_{jj',t}$, high $\omega_{jj',t}$ denotes high complementarity.
Espoused	$\tau_{jj',t}$	Products are related and grouped into disjoint layers by an expert panel. The complementarity for layers j and j' ($j \neq j'$) in year t is the average of $r_{ll',t}$ between the products in each layer. High $\tau_{jj',t}$ denotes high complementarity.

TABLE 3 Emergent stack descriptors

Year	N (Stack Layers)	Average (Products per layer)	S.D. (N products / layer)	Max (products per layer)
1991	5	3.80	1.72	6
1992	5	4.20	1.38	6
1993	8	4.63	2.89	11
1994	11	4.00	1.67	6
1995	16	3.31	1.93	8
1996	20	3.10	2.38	8
1997	25	2.84	2.52	10
1998	23	3.39	4.29	14
1999	22	3.68	2.48	10
2000	28	3.04	1.77	7
2001	21	4.05	2.82	12

TABLE 4: Espoused stack descriptors

	Products			Operating Systems		
	Avg.	σ	Max	Avg.	σ	Max
1990	2.915	3.004	16	2.437	1.156	5
1991	2.863	2.954	16	2.375	1.129	5
1992	2.382	2.627	17	2.321	1.104	5
1993	3.260	4.308	28	2.680	1.389	6
1994	3.109	4.316	29	2.692	1.400	6
1995	3.272	5.094	41	2.937	1.398	6
1996	2.992	4.834	49	2.885	1.323	6
1997	2.634	4.551	57	2.715	1.319	7
1998	2.372	4.041	54	2.481	1.270	7
1999	2.245	3.726	54	2.356	1.257	8
2000	2.303	3.744	56	2.289	1.244	8
2001	2.289	3.649	58	2.246	1.250	8
2002	2.294	3.639	58	2.256	1.225	8

TABLE 5: Means, standard deviations, and zero-order correlations[a]

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1.Emergent stack complementarity$_{i,t}$	1.000													
2.Espoused stack complementarity$_{i,t}$	0.463	1.000												
3.Product complementarity$_{i,t}$	0.496	0.396	1.000											
4.Platform diversity$_{i,t}$	0.108	0.123	0.107	1.000										
5.Geographic diversity$_{i,t}$	0.171	0.267	0.134	0.185	1.000									
6.Platform dominance$_{i,t}$	0.099	0.175	0.057	*0.013*	0.165	1.000								
7.Product dominance$_{i,t}$	0.084	0.192	0.046	0.073	0.255	0.401	1.000							
8.Geographic dominance$_{i,t}$	0.098	0.161	0.048	0.093	0.150	0.413	0.354	1.000						
9.Platform growth$_{i,t}$	0.079	0.076	0.034	-0.036	0.098	0.118	0.085	0.032	1.000					
10.Product growth$_{i,t}$	0.077	0.095	0.088	-0.047	*0.008*	*0.024*	0.139	*0.010*	0.284	1.000				
11.Geographic growth$_{i,t}$	0.179	0.166	0.110	*-0.006*	0.124	0.103	0.081	0.099	0.451	0.424	1.000			
12.ln(Firm size$_{i,t}$)	0.268	0.337	0.264	0.249	0.395	0.394	0.498	0.434	*0.008*	0.030	0.113	1.000		
13.ln(Firm age$_{i,t}$)	0.260	0.312	0.240	0.248	0.334	0.248	0.172	0.297	-0.106	-0.079	0.066	0.516	1.000	
14.Firm sales growth$_{i,t}$	-0.040	*-0.016*	-0.044	-0.114	-0.108	*0.013*	0.061	*-0.014*	0.170	0.283	0.118	*0.005*	-0.237	1.000
Mean	0.070	0.101	0.116	0.581	0.685	0.006	0.035	0.002	0.083	0.072	0.138	3.243	1.628	0.221
S.D.	0.134	0.179	0.176	0.408	0.414	0.028	0.074	0.012	0.081	0.060	0.119	1.775	0.810	0.247

N=4392; 884 panels; [a]Due to large sample, italicized number denote p>0.05

TABLE 6: GEE regression results

	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6		Model 7	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Emergent stack complementarity$_{i,t}$							0.123**	0.046					0.408***	0.110
Emergent stack complementarity$_{i,t}{}^2$													-0.551**	0.192
Espoused stack complementarity$_{i,t}$					0.110**	0.033					0.170†	0.101		
Espoused stack complementarity$_{i,t}{}^2$											-0.115	0.177		
Product complementarity$_{i,t}$			0.080**	0.031					0.258**	0.084				
Product complementarity$_{i,t}{}^2$									-0.317*	0.127				
Platform diversity$_{i,t}$	-0.013	0.014	-0.014	0.014	-0.014	0.014	-0.014	0.014	-0.015	0.014	-0.014	0.014	-0.015	0.014
Geographic diversity$_{i,t}$	-0.032*	0.015	-0.032*	0.015	-0.037*	0.015	-0.033*	0.015	-0.032*	0.015	-0.037*	0.015	-0.032*	0.015
Platform dominance$_{i,t}$	0.280†	0.145	0.274*	0.140	0.244†	0.131	0.265†	0.142	0.258†	0.137	0.249†	0.134	0.237	0.145
Product dominance$_{i,t}$	0.165*	0.084	0.184*	0.084	0.158†	0.085	0.178*	0.083	0.194*	0.084	0.160†	0.085	0.189*	0.084
Geographic dominance$_{i,t}$	0.808*	0.361	0.920*	0.374	0.917**	0.338	0.900*	0.369	0.978**	0.377	0.875**	0.338	0.912*	0.370
Platform growth$_{i,t}$	0.098*	0.041	0.099*	0.041	0.100*	0.041	0.101*	0.041	0.098*	0.041	0.100*	0.041	0.101*	0.041
Product growth$_{i,t}$	0.104***	0.025	0.100***	0.025	0.100***	0.025	0.102***	0.025	0.100***	0.025	0.101***	0.025	0.101***	0.025
Geographic growth$_{i,t}$	0.718**	0.270	0.763**	0.270	0.721**	0.270	0.738**	0.269	0.802**	0.271	0.724**	0.270	0.758**	0.270
ln(Firm size$_{i,t}$)	-0.023***	0.005	-0.025***	0.005	-0.025***	0.005	-0.025***	0.005	-0.028***	0.005	-0.026***	0.005	-0.028***	0.005
ln(Firm age$_{i,t}$)	-0.024**	0.008	-0.026**	0.008	-0.028**	0.009	-0.026**	0.009	-0.028**	0.009	-0.028***	0.009	-0.029**	0.008
Firm sales growth$_{i,t}$	0.164***	0.020	0.165***	0.020	0.162***	0.020	0.164***	0.020	0.166***	0.020	0.162***	0.020	0.163***	0.020
Year 1991	0.052	0.045	0.041	0.045	0.041	0.045	0.046	0.045	0.035	0.045	0.041	0.045	0.050	0.045
Year 1992	-0.165	0.162	-0.192	0.162	-0.168	0.161	-0.184	0.161	-0.214	0.162	-0.169	0.161	-0.191	0.162
Year 1993	-0.384*	0.169	-0.415*	0.169	-0.396*	0.168	-0.410*	0.168	-0.440**	0.169	-0.397*	0.168	-0.419*	0.169
Year 1994	0.188***	0.046	0.176***	0.046	0.183***	0.046	0.177***	0.046	0.169***	0.046	0.182***	0.046	0.169***	0.046
Year 1995	0.020	0.064	0.006	0.064	0.013	0.064	0.007	0.064	-0.002	0.064	0.013	0.064	-0.001	0.064
Year 1996	0.094*	0.043	0.086*	0.043	0.090*	0.043	0.085†	0.043	0.079†	0.043	0.089*	0.043	0.081†	0.044
Year 1997	0.089†	0.048	0.080†	0.048	0.083†	0.048	0.081†	0.048	0.073	0.049	0.082†	0.048	0.076	0.049
Year 1998	0.059	0.043	0.053	0.043	0.056	0.043	0.058	0.043	0.046	0.043	0.056	0.043	0.055	0.043
Year 1999	0.106*	0.042	0.100*	0.042	0.102*	0.042	0.103*	0.042	0.095*	0.042	0.101*	0.042	0.101*	0.042
Year 2000	-0.078*	0.034	-0.083*	0.035	-0.078*	0.034	-0.082*	0.034	-0.088*	0.035	-0.079*	0.034	-0.085*	0.035
Constant	0.077***	0.019	0.079***	0.019	0.088***	0.019	0.081***	0.019	0.085***	0.019	0.089***	0.019	0.088***	0.019
df	21		22		22		22		23		23		23	
Wald χ^2	831.40		833.34		835.45		841.01		857.52		848.38		871.18	
Difference calculation			(2)-(1)		(3)-(1)		(4)-(1)		(5)-(2)		(6)-(3)		(7)-(4)	
$\Delta\chi^2$			1.94		4.05*		9.61**		24.18***		12.93***		30.17***	

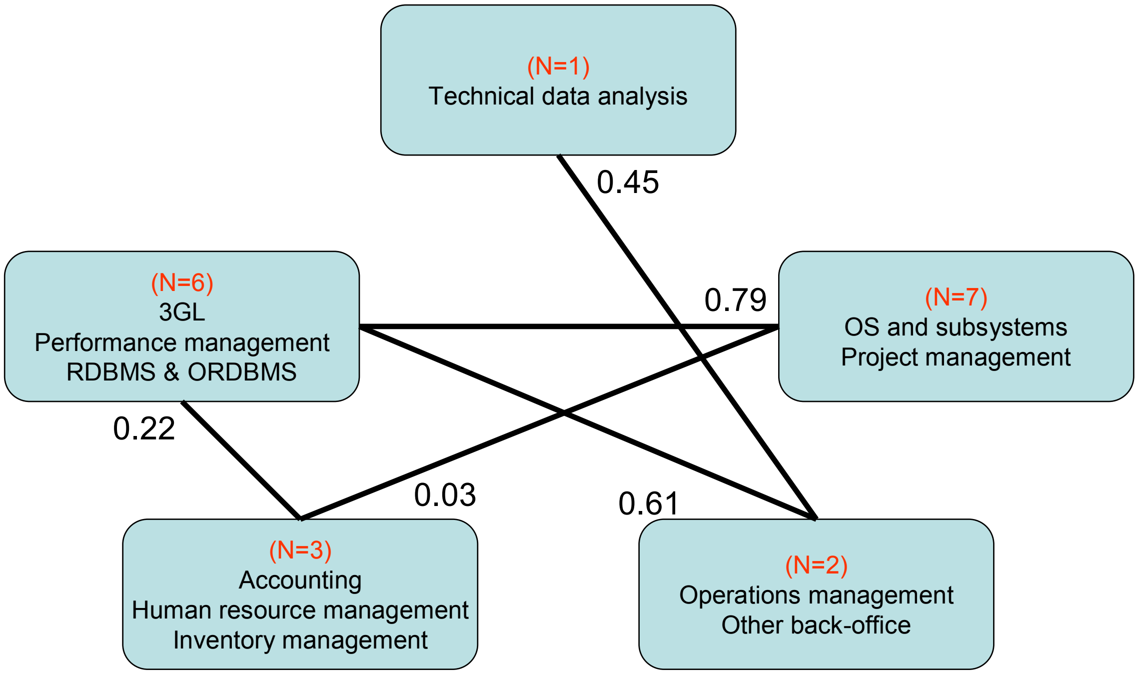
N=4392; 884 panels; Semi-robust standard errors; †<.10, *<.05, **<.01, ***<.001

TABLE 7: GEE regression results

	Model 8		Model 9		Model 10		Model 11		Model 12	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Emergent stack complementarity$_{i,t}$	$0.094^{\dagger}$	0.050	0.079	0.050	0.340^{**}	0.132	0.341^{**}	0.123	0.299^{*}	0.137
Emergent stack complementarity$_{i,t}^{2}$					-0.483^{*}	0.208	-0.486^{*}	0.202	-0.446^{*}	0.213
Espoused stack complementarity$_{i,t}$			0.086^{*}	0.035			0.063	0.111	0.054	0.113
Espoused stack complementarity$_{i,t}^{2}$							0.007	0.183	0.013	0.186
Product complementarity$_{i,t}$	0.049	0.033			0.098	0.105			0.070	0.108
Product complementarity$_{i,t}^{2}$					-0.101	0.149			-0.069	0.151
Platform diversity$_{i,t}$	-0.014	0.014	-0.015	0.014	-0.015	0.014	-0.015	0.014	-0.015	0.014
Geographic diversity$_{i,t}$	-0.033^{*}	0.015	-0.037^{*}	0.015	-0.032^{*}	0.015	-0.035^{*}	0.015	-0.035^{*}	0.015
Platform dominance$_{i,t}$	$0.265^{\dagger}$	0.140	$0.243^{\dagger}$	0.132	$0.236^{\dagger}$	0.142	0.223	0.136	$0.223^{\dagger}$	0.134
Product dominance$_{i,t}$	0.187^{*}	0.083	0.168^{*}	0.084	0.197^{*}	0.084	0.179^{*}	0.085	0.186^{*}	0.085
Geographic dominance$_{i,t}$	0.949^{*}	0.377	0.954^{**}	0.349	0.965^{*}	0.379	0.959^{**}	0.360	0.995^{**}	0.372
Platform growth$_{i,t}$	0.101^{*}	0.041	0.101^{*}	0.041	0.101^{*}	0.041	0.101^{*}	0.041	0.101^{*}	0.041
Product growth$_{i,t}$	0.100^{***}	0.025	0.100^{***}	0.025	0.100^{***}	0.025	0.100^{***}	0.025	0.099^{***}	0.025
Geographic growth$_{i,t}$	0.761^{**}	0.269	0.733^{**}	0.269	0.786^{**}	0.272	0.752^{**}	0.270	0.772^{**}	0.272
ln(Firm size$_{i,t}$)	-0.026^{***}	0.005	-0.026^{***}	0.005	-0.029^{***}	0.005	-0.028^{***}	0.005	-0.029^{***}	0.005
ln(Firm age$_{i,t}$)	-0.027^{**}	0.009	-0.029^{**}	0.009	-0.030^{***}	0.009	-0.031^{***}	0.009	-0.031^{***}	0.009
Firm sales growth$_{i,t}$	0.165^{***}	0.020	0.162^{***}	0.020	0.164^{***}	0.020	0.161^{***}	0.020	0.162^{***}	0.020
Year 1991	0.040	0.045	0.039	0.045	0.044	0.045	0.045	0.045	0.041	0.045
Year 1992	-0.196	0.161	-0.179	0.161	-0.206	0.163	-0.187	0.162	-0.198	0.163
Year 1993	-0.423^{*}	0.168	-0.410^{*}	0.168	-0.435^{**}	0.169	-0.418^{*}	0.169	-0.429^{*}	0.169
Year 1994	0.172^{***}	0.046	0.177^{***}	0.046	0.165^{***}	0.046	0.170^{***}	0.046	0.167^{***}	0.046
Year 1995	0.001	0.064	0.007	0.064	-0.006	0.064	0.000	0.064	-0.004	0.064
Year 1996	$0.082^{\dagger}$	0.043	$0.085^{\dagger}$	0.043	$0.078^{\dagger}$	0.043	$0.082^{\dagger}$	0.044	$0.079^{\dagger}$	0.043
Year 1997	0.078	0.048	$0.080^{\dagger}$	0.048	0.072	0.049	0.075	0.049	0.073	0.049
Year 1998	0.054	0.043	0.056	0.043	0.051	0.043	0.055	0.043	0.051	0.043
Year 1999	0.100^{*}	0.042	0.101^{*}	0.042	0.098^{*}	0.042	0.100^{*}	0.042	0.098^{*}	0.042
Year 2000	-0.084^{*}	0.034	-0.080^{*}	0.034	-0.088^{*}	0.035	-0.084^{*}	0.035	-0.086^{*}	0.035
Constant	0.081^{***}	0.019	0.088^{***}	0.019	0.089^{***}	0.019	0.093^{***}	0.020	0.093^{***}	0.020
df	23		23		25		25		27	
Wald χ^2	857.78		838.49		864.50		860.65		876.52	
Difference calculation	(8)-(2)		(9)-(3)		(10)-(5)		(11)-(6)		(12)-(7)	
$\Delta\chi^2$	24.44^{***}		$3.04^{\dagger}$		6.98^{*}		12.27^{**}		5.34	

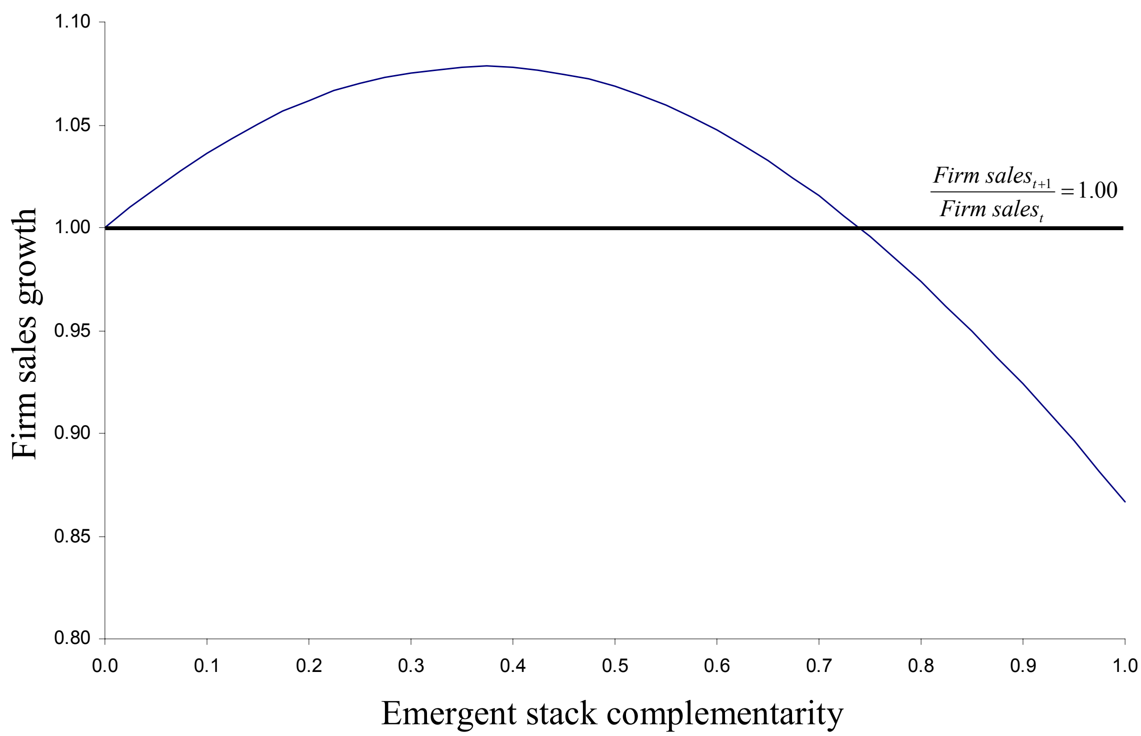
N=4392; 884 panels; Semi-robust standard errors; †<.10, *<.05, **<.01, ***<.001

FIGURE 2: Product, layers, selected complementarity for year 1991.



N equals the total number of products in a layer.

FIGURE 3: Impact of emergent stack complementarity on firm growth



APPENDIX A: Similarity metric computations

We use the structural equivalence approach [34, 36, 37], based on the sales patterns of ISV across products and operating systems, to determine similarity between two products. Two products are structurally equivalent "to the extent that they have identical relations with every other element [34]." Formally, let $x_{ijk,t}$ denote the sales that firm i generated from selling its product j for OS k in year t. The number of firms, products, and operating systems varies from year to year as the software industry evolves. The number of ISV changes as they enter or exit. The number of OS changes as new ones are introduced and demand by consumers, and ISV port their products to the new platforms. Finally, the number of products that ISV offer changes as consumer needs change.

The structural equivalence method requires using a "projecting" transformation of which the Euclidean distance, proportional [34] and marginal [36] are the most common. Another transformation, used in empirical studies (e.g., [38]), is to calculate the correlation coefficient between every pair of functions [71]. Most of the transformations used in prior studies have been symmetric in that the structural equivalence between products j and j' is equal to the structural equivalence between functions j and j'.

Recently, Sohn [13] has extended the specifications of Burt [34] and Burt and Carlton [36] by explicitly incorporating the ability to accommodate size differences. Sohn's [13] specification for structural equivalence between two functions j and j' is an asymmetric cosine metric first proposed by Pianka [72] and given by:

$$r_{jj',t} = 1.0 - \sum_k p_{jk,t} \left(\frac{\sum_i x_{ijk,t} \min\left(x_{ijk,t}, x_{ij'k,t}\right)}{\sum_i x_{ijk,t}^2} \right)$$

where $p_{jk,t}$ is the proportion or importance of operating system k for product j in year t. We subtract one from the Sohn [13] metric because in the original specification, the Sohn [13] metric is bounded between 1.0 and 0.0 with the former denoting perfect similarity or equivalence. In general, $r_{jj',t} \neq r_{j'j,t}$ but $r_{jj',t} > 0$ if $r_{j'j,t} > 0$. A very usefully property of this metric is that it is ratio. For example, twice the value of the metric denotes a two-fold increase in the distance of two products j and j'. Two products j and j' are structurally equivalent if there exist software firms that offered and generated from functions j and j'.

Figure A.1 represents four stylized situations, adapted from Sohn [13], between two products (j and j') and two ISV (i and i') that help to illustrate the properties of the similarity metric. The vertical axis in each figure denotes the sales of the ISV. For expositional clarity, this figure assumes a single OS permitting one to drop the outer summation over all OS k. Figure A.1a depicts a situation of no distance between two products. Firms i and i' generated sales form both products at equal levels 1 and ½, respectively resulting in $r_{jj',t} = r_{j'j,t} = 0$. Figure A.1b depicts a situation of maximum distance. Firm i generated sales from function j but did not generate sales from function j'. Firm i' generated sales from function j' but did not generate sales from function j. Because no ISV exists that generated sales from both functions $r_{jj',t} = r_{j'j,t} = 1$.

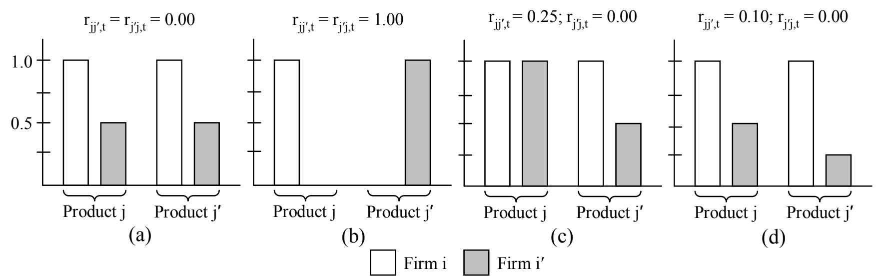


Figure A.1c depicts an asymmetric situation. Firms i and i' generated sales from both products. However, firm i generated equals levels of sales from both products while firm i' generated sales from function j and j' at levels 1 and ½, respectively. Function j's distance from function j' is $r_{jj',t}$ = ¼ because the total sales by the two ISV for markets j and j' are 2 and 1½, respectively. Because firm i' emphasizes function j' less than function j, the distance is asymmetric. Interestingly, function j''s distance from function j is $r_{j'j,t} = 0$. The asymmetry results because the ISV sales function j' is "included" in the ISV sales profile for function j. Figure A.1c illustrates that the distance metric used in this study is sensitive to the sales distributions of ISV. Figure A.1d also illustrates an asymmetrical situation but further illustrates the impact of overall market size. Note that Figure A.1d is identical to Figure A.1a except that firm i' generates sales from market j' at level ¼ instead of ½. As in Figure A.1c, the ISV sales profile in Figure A.1d for market j' is "included" in the ISV sales profile for market j. More interesting is the observation that for function j, ISV i's sales is twice as "important" as ISV i' because the latter generated sales from market j at twice the level. Hence, any decrease in ISV i'

sales from market j' in Figure A.1d from Figure A.1c should be attenuated even though the sales dropped by 50%. However, when the entire sales profile is considered and given the ratio scale property of the metric, one indeed finds that the function j's distance with function j' increases by 10% to 0.10 given a 50% decrease in ISV i' sales from function j' thus illustrating the sensitively of the structural equivalence measure to overall size.