Inventing by Combining Pre-Existing Technologies: Patent Evidence on Learning and Fishing Out

Matt Clancy*

Preliminary Working Paper

Abstract

I test a model of innovation where new technologies are combinations of pre-existing technologies. The model captures two opposing forces: the best ideas are used up (knowledge is exhaustible), but as firms learn which technologies can be combined, new ideas become feasible (knowledge accumulates). My dataset consists of more than 80 years of data on the patenting of 10,000 sets of technological components, themselves proxied by 13,517 US patent office technology classifications. I show the number of patent applications using a given technology combination increases as firms learn pairs of constituent technologies are compatible, but decreases as the exact set of combinations is patented (i.e., fished out). My results suggest the fishing out effect is dominant at the level of an individual combination.

Keywords: Innovation, Patents, Combinatorial growth, Spillovers, R&D

JEL Codes: O31

* Matt Clancy (matthew.clancy@ers.usda.gov) is an economist at the USDA Economic Research Service. The views expressed herein do not necessarily represent the views of the USDA.

Acknowledgments: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

This is a paper about where ideas come from. There is a large and fruitful literature on the economics of innovation, but the source of ideas is largely treated as a black box. As Weitzman (1998) expressed it, "[In endogenous growth models,] 'new ideas' are simply taken to be some exogenously determined function of 'research effort' in the spirit of a humdrum conventional relationship between inputs and outputs." In this paper, I help to open the black box by presenting evidence that innovation is a combinatorial process subject to opposing forces. On the one hand, knowledge spills over and accumulates via a process where researchers learn from the innovations that succeed. On the other hand, the set of ideas if large but finite and can be exhausted by the research process. The first effect expands the set of research possibilities, and the second contracts it.

Consider the internal combustion engine as an example. The internal combustion engine is a single idea created by inventors, but is also a combination of pistons, crankshafts, flywheels, valves, and combustible chemicals.[1] All of these ingredients existed (in one shape or another) before the engine, but the internal combustion engine itself did not exist until they were brought together. At the same time, an internal combustion engine is more than the combination of these building blocks: piling pistons, crankshafts, flywheels, valves, and combustible chemicals in a heap would not yield up an engine. These elements must work together in harmony to perform useful tasks. The combustible chemicals drive the pistons, the crankshaft converts this back-and-forth motion into jerky rotational motion, the flywheel converts the crankshaft's jerky motion into comparatively smooth rotational motion, and so on.

When conducting the research and design (R&D) that would eventually culminate in an internal combustion engine researchers began with some pieces of information. They knew how the components they intended to use were likely to interact because the components had been paired before in other contexts. Waterwheels also use crankshafts and pistons, and a manual potter's wheel

[1] The following example draws on Dartnell (2014), pgs. 201-206.

couples a flywheel to a jerky source of motion. The likely interaction of these pairs of technologies (crankshaft and piston, flywheel and piston) could be inferred from these other contexts. Knowledge spills over in both the recycling of components and in the understanding of how components may be used together. The combustion engine was possible, in principle, once its components existed, but researchers did not try to invent it until they knew enough about how its components would interact.

Just as the designers of the first internal combustion engines drew on knowledge from other contexts, designers of the second-generation of engines relied heavily on what worked for the first generation. Engines became better and better but not without limit. There are ultimately a finite number of ways to combine and recombine the components making up the internal combustion engine, and technologies can reach a point where no further improvements can be wrung out of the same set of components. The set of ideas – from a given set of components – is exhaustible.

These observations illustrate a more general theory of innovation. First, innovation is **combinatorial**. New ideas are built by combining older, pre-existing ideas and concepts in novel ways. Second, innovation relies on **learning**. Researchers learn from past combinations, observing what sets of building blocks can be usefully combined and which cannot. Third, the pool of ideas can be **fished out**. The number of ways to combine any finite set of building blocks is limited, and every time a new idea is discovered, there is one fewer idea left to discover for followers.

The main contribution of this paper is provide evidence for learning and fishing out in a combinatorial setting, with historical patent data. Section 1 briefly reviews related literature, and section 2 presents a short expositional model to motivate four hypotheses that will be empirically tested. In section 3, I describe how I take these hypotheses to data, using patent technology classifications as proxies for the technological building blocks that are combined to yield ideas. Section 4 presents my estimation approach and section 5 presents the results. Using the results of section 5, section 6 discusses the relative magnitude of the learning and fishing out effects, concluding that the fishing out effect dominates at the level of a particular combination, but that more research is required to determine which effect dominates at the industry level. Section 6 summarizes the paper and points out areas for future research.

1 – Background

The internal combustion engine is just one example of the ways technology and ideas themselves can be viewed as fundamentally combinatorial. Any physical technology can be broken down into component parts, while invented procedures can be broken down into steps and actions. Works of art draw on a common set of themes, styles, symbols, conventions, and other tropes. Indeed, even abstract ideas can be understood as combinations of concepts, arguments, mathematical tools, facts, and so forth. Weitzman (1998) was the first to incorporate combination into the economist's knowledge-production function (although it has not been widely adopted since). Later, Arthur (2009) echoes Weitzman's model and views all technologies as hierarchical combinations of sub-components. Similarly, Ridley (2010) argues the best innovations emerge when "ideas have sex." Others, such as Jovanovic and Rob (1990), Kauffman, Lobo, and Macready (2000) and Auerswald et al. (2000) have drawn models where innovators learn about the likely outcome of R&D from observing the outcomes of other R&D projects, but not in a combinatorial setting.

Fleming (2001) is an important empirical investigation of combinatorial innovation. Fleming explicitly considers a framework where innovation is combinatorial and where researchers learn which components tend to work together. Using a sample of 17,264 patents, Fleming (2001) exploits the fact that most patents are assigned to more than one technological subclass, interpreting each sub-class assignment as a component. He shows the number of times a combination of sub-classes have been assigned to a patent is negatively correlated with the number of citations. Fleming also computes a measure of combination familiarity which Fleming interprets as accounting for the distance of search, but which could also be viewed as a measure of what firms have learned about the components. This metric is also associated with more citations. Other papers exploring the combinatorial properties of patents and academic papers and their relationship to future citations include Keijl et. al (2016), Nemet (2012), Nemet and Johnson (2012), Schilling (2011), and Shoenmakers (2010). Preliminary work by Akcigit, Kerr, and Nicholas (2013) explores similar territory as this paper with US patent data, but focuses on explaining a different set of stylized facts from this paper.

This paper differs from other empirical papers in a number of respects. First, my objective is to highlight the twin forces of learning and fishing out rather than citation behavior. My research design allows me to see how combinatorial factors are correlated with the propensity to develop new ideas, not whether ideas are cited heavily or not. Second, I decompose innovators' knowledge into knowledge about pairwise interactions between distinct technologies. This allows me to separately identify the effects of fishing out and learning and to illustrate how knowledge spills across different contexts. Fleming (2001) models innovation as drawing only on combinatorial information from technologies with the exact same combination of technology subclasses. This paper's model implicitly assumes the most important characteristics of an idea can be decomposed into a function of the set of pair-wise interactions between all its building blocks. Third, my use of technology subclasses improves on earlier work by aggregating up to the mainline class, so that technology classifications are non-nested, exhaustive, and comparable (a similar approach as applied to maps of the technological landscape has recently been explored by Aharonson and Schiling 2016).

2 – An Expositional Model of Combinatorial Innovation

The following model is used to construct several hypotheses and to more precisely describe the learning and fishing out effect.

2.1 – Model

New ideas are created by myopic profit-seeking firms. Firms may conduct R&D on one idea (denoted i) per period. Innovation is stochastic, and not all ideas turn out to be viable. Research on i costs k_i, and at the end of the period, researching firms learn if i is viable or not. Viable ideas are patented, and the inventing firm obtains value π_i. Ideas that are not viable have zero value: they are fatally flawed and not useful to any buyer. Firms know π_i and k_i when deciding whether to conduct R&D, but not whether an idea is viable. Once an idea has been attempted, no other firm attempts it: it is either patented or discovered to be unviable, and this is public information.

Ideas are assembled out of combinations of pre-existing technological elements. There exists a set Q of such elements and any idea is a unique subset of Q with at least two elements. Some elements of Q tend to be mutually compatible with each other; others not. Ideas are most likely to be viable if they are composed of elements that tend to be compatible.

To measure the degree of compatibililty, every element-pair j has a scalar measure called *affinity*, denoted a_j^*. Higher values of a_j^* mean the element-pair j is highly compatible. Let $\vec{a}_i$ denote the vector of affinities of all pairs between elements in idea i (there will be $n(n-1)/2$ pairs of elements if an idea has n elements). The probability an idea is viable is a function of the affinities between its constituent elements:

$$\Pr(i \text{ viable}) = f(\vec{a}_i) \tag{1}$$

where:

$$\frac{\partial f}{\partial a_j^*} \geq 0 \forall a_j^* \in \vec{a}_i \tag{2}$$

That is, the probability an idea is viable, and therefore patentable, is increasing in the affinity between the elements that comprise the idea.

While I assume the probability an idea is viable is increasing in a_j^*, I do not make assumptions about how affinities interact with each other. In the empirical section, I consider two extreme cases:

Perfect substitutes: $$f\left(\vec{a}^i\right) = \tilde{f}\left(\sum_{a_j^* \in \vec{a}_i} a_j^*\right)$$

Perfect complements: $$f\left(\vec{a}_i\right) = \tilde{f}\left(\min\left(\vec{a}_i\right)\right)$$

where $\tilde{f}' > 0$. If affinity operates as a substitute, then it is possible to compensate for low-affinity elements with high affinity ones. This may serve as a metaphor for how some technological components might regulate and buffer the interaction of technological components that otherwise interact poorly. If affinity is complementary, it would imply *all* the elements in an idea must operate harmoniously with each other, and a technology is only as strong as its weakest link. When combinatorial knowledge is complementary, knowing, for instance, that *A* and *B* are highly compatible technological elements increases the returns to knowing that *A* and *C* or *B* and *C* are highly compatible.

Firms do not observe the affinity of an element-pair but are Bayesians whose beliefs about the likely values of a_j^* are described in period t by the cumulative probability function $G_t\left(a_j\right)$. As a consequence of Bayes law and equation (2), the expected value of a_j is increasing in the number of patented ideas that include element-pair j. Firms believe ideas are more likely to be viable if they are comprised of elements that have been successfully combined in the past.

Every period, new values of k_i and π_i are drawn, representing changes associated with the cost of R&D and the value of individual inventions. There is free entry into the innovation game, and in each period a single[2] firm attempts each idea satisfying:

$$\pi_{it} E\left[f\left(\vec{a}_{it}\right)\right] - k_{it} > 0 \tag{3}$$

The probability the untried idea i is patented in a given period is:

$$\Pr\left(i \text{ patented}\right) = \Pr\left(E_t\left[f\left(\vec{a}_{it}\right)\right] > k_{it} / \pi_{it}\right) \times E_t\left[f\left(\vec{a}_{it}\right)\right] \tag{4}$$

The first term here is the probability the idea will be attempted, and the second term is the probability it will be viable (and therefore patented), if attempted.

As equation (4) makes explicit, a long-neglected idea might suddenly be patented for one of several reasons. The cost of research k_{it} might decline (perhaps because of an inflow of human capital). The value of the patented invention π_{it} might rise (perhaps because of the development of complementary products or shifting consumer demands). Or, most relevant to this paper, firms may update their beliefs about the affinity between elements comprising the idea. If other inventions using some of the same elements are patented, this generates a knowledge spillover, leading firms to revise up their estimates of a_j^*, and hence their belief that the idea is viable.

Note that equation (4) applies to untried ideas. I assume a specific combination can only be tried once: while there may be many different ways to make an internal combustion engine from slightly different sets of components, at this point I assume there is only one way to try a specific set of

[2] We may assume firms enter each period in a random ordered sequence and may claim any unclaimed idea upon entry. Claims are public knowledge and firms complete their R&D in the same sequence, and so no two firms will ever attempt the same idea.

components. Once it has been attempted, the probability it will be tried again falls to zero. It has either been patented or found unviable.

The focus of this paper is on these last two effects, which I call the learning and fishing out effect. Whenever elements are combined successfully, this conveys useful information to rivals who may pursue other ideas that draw on some of the same elements. This is the learning effect. At the same time, the number of ideas is finite (equal to the number of subsets of $\mathbb{Q}$ with at least two elements). Every time an idea is attempted, there is one less idea available for future firms to try. This is the "fishing out" effect. These elements are in tension. If ideas are like fish in a pond, then innovation both pulls fish out and restocks the pond at the same time, although not necessarily at equal rates.

2.2 – Empirical Application

We do not actually observe the elements of Q. Instead, we observe their proxies in the form of the patent office's technology classifications (discussed in section 3), which are called **mainlines**. In Section 2.1, we assumed an idea i designates both a subset of Q and a single idea. Moving forward, I relax this assumption. In the empirical component of this paper let i designate a particular set of mainlines. Specifically, I will restrict my attention to sets of three mainlines. I assume there are λ_{it} untried ideas that use this set of mainlines, where $\lambda_{it} \geq 1$ initially. For example, multiple patents might be assigned the same three mainlines relating to an internal combustion engine regulator, transmission control, and vehicle guidance software, because mainlines do not precisely identify a technological component.

The number of patents assigned the mainline-set i and applied for in period t is denoted y_{it}. Returning to equation (4), given our imperfect proxy for the elements combined, the expected number of patent applications in period t assigned mainline-set i is:

$$E\left[y_{it}\right]=\lambda_{it}\times\Pr\left(E_t\left[f\left(\vec{a}_{it}\right)\right]>k_i/\pi_i\right)\times E_t\left[f\left(\vec{a}_{it}\right)\right] \tag{5}$$

Taking equation (5) from left-to-right, the expected number of patents assigned mainline-set i is equal to (1) the number of untried ideas using mainline-set i, multiplied by (2) the probability each of these ideas is attempted, multiplied by (3) the probability each attempted idea is patentable.

We do not actually observe λ_{it} or $\vec{a}_{it}$ and so my empirical exercise is based on the following reduced form equations:

Perfect Substitutes: $$E\left[y_{it}\right]=h\left(\sum_{j\in i}\tilde{a}_{jt},\tilde{\lambda}_{it},X_{it}\right) \tag{6}$$

Perfect Complements: $$E\left[y_{it}\right]=h\left(\min_{j\in i}\left(\tilde{a}_{jt}\right),\tilde{\lambda}_{it},X_{it}\right) \tag{7}$$

where $\tilde{\vec{a}}_{it}$ is a proxy for $\vec{a}_{it}$, $\tilde{\lambda}_{it}$ is a proxy for λ_{it}, and X_{it} is a set of controls and a stochastic error term. Based on equation (5), we assume $h_1\geq 0$ and $h_2\leq 0$.

Define n_{jt} to be the number of patents assigned mainlines that include mainline-pair j and granted prior to or in period t. Define N_{it} to be the number of patents assigned mainline-set i granted prior to or in period t. For example, suppose mainline-set i consists of mainlines *ABC*. Three patents have been granted through period t, with mainline sets *ABD, AB,* and *ABC*. In this example, n_{jt} for the pair *AB* is 3, and n_{jt} for the pairs *AC* and *BC* is 1, while N_{it} for the set *ABC* is equal to 1.

Define our proxies for affinity and the number of untried ideas using mainline-set i as follows:

$$\tilde{a}_j=\beta_1 n_{jt}+\beta_2 n_{jt}^2 \tag{8}$$

$$\tilde{N}_t^i=\gamma+\phi_1 N_{it}+\phi_2 N_{it}^2+\alpha_1 t+\alpha_2 t^2 \tag{9}$$

With these functional forms, we can compute the Perfect Substitutes and Perfect Complements measures as follows:

Perfect Substitutes: $$\sum_{j\in i}\tilde{a}_{jt} = \beta_1 \sum_{j\in i} n_{jt} + \beta_2 \sum_{j\in i} n_{jt}^2 \tag{10}$$

Perfect Complements: $$\min_{j\in i}\left(\tilde{a}_{jt}\right) = \beta_1 \min_{j\in i}\left(n_{jt}\right) + \beta_2\left[\min_{j\in i}\left(n_{ji}\right)\right]^2 \tag{11}$$

I use equations (6) and (7) to test four hypotheses:

> **Hypothesis 1 (learning shape):** $\beta_1 > 0$ and $\beta_2 < 0$. $E\left[y_{it}\right]$ is increasing in n_{jt} and is bounded from above.

Based on equation (5), $E\left[y_{it}\right]$ should be increasing in $n_{j,t}$ for two reasons. As noted earlier, the expected value of a_j^* is increasing in n_{jt}, and $\partial f / \partial a_j^* \geq 0$. An increase in the number of patents using pair j leads firms to increase their belief the remaining ideas using pair j are also viable. This leads to more ideas being perceived as worth attempting and also more attempted ideas being found viable (in expectation). $E\left[y_{it}\right]$ should have an upper bound in $n_{j,t}$ because a_j^* acts through probabilities, which have an upper bound of 1. As this upper bound is approached, subsequent increases in $n_{j,t}$ cannot move $f\left(\vec{a}^i\right)$ as much. At this stage, researchers are already confident the pair will be compatible and the decision to attempt research on the idea is driven by other factors (for example, the draw of k_i / π_i). Indeed, the fact that equation (8) is proxying for an upper bound implies the following additional hypothesis.

> **Hypothesis 2 (learning range):** A large majority of observations have $n_{jt} < -\beta_1 / 2\beta_2$. Most observations have positive marginal contributions to $E\left[y_{it}\right]$.

Note the peak of a parabola is given by $-\beta_1 / 2\beta_2$. We anticipate observations lying to the left of this peak. We next turn to our hypotheses about the fishing out effect.

> **Hypothesis 4 (fishing out – Mainline-sets):** The estimated coefficients ϕ_1 and ϕ_2 are such that the marginal contribution of N_{it} is negative for a large majority of observations. $E[y_{it}]$ is decreasing and convex in the number of patents using mainline-set i.

> **Hypothesis 3 (fishing out - time):** The estimated coefficients α_1 and α_2 are such that the marginal contribution of t is negative for a large majority of observations. $E[y_{it}]$ is decreasing and convex in time.

Again returning to equation (5), λ_{it} decreases as more ideas are attempted. In every period, we observe the number of patents assigned mainline-set i, and so we observe the number of attempts that turn out to be patentable. The cumulative number of these patents is N_{it} and they subtract directly from λ_{it}, leading to a decrease in $E[y_{it}]$ once we control for the learning effect. We do not observe attempts of mainline-set i that are not patentable, but overtime they will invisibly accumulate. Therefore, $E[y_{it}]$ should be decreasing over time.

In either case, λ_{it} is bounded from below by 0, and so the marginal contribution of time or patents with mainline-set i should also go to zero at sufficiently high levels. It will turn out that our estimation methodology requires this anyway, but I allow t and N_{it} to be quadratic for maximum flexibility.

It worth pausing to expand on a few points before turning to data.

First, it is only possible to separately identify the learning and fishing out effects because I have assumed beliefs about the viability of ideas can be disaggregated into information about pairs of mainlines. For example, if an idea uses mainlines *ABC*, then a patent using mainlines *ABX* or *ABY* conveys useful information about the viability of *ABC* (specifically, it suggests *A* and *B* can be usefully combined), without fishing out one of the finite number of ideas that use *ABC*. If we assumed firms only learn from exact matches, then *ABX* and *ABY* would convey no useful information about *ABC*: only other patents with *ABC* would, which would mix together the learning and fishing out effects.

Second, it is important to consider what information is in firms' information set at the time of a patent application. In all cases, we assume patent information is disclosed by patent publication. All explanatory variables are based on the date of a patent grant, while dependent variables are based on the date of patent application. (typically it takes 3 years to move a patent from application to grant).

3 – Data

3.1 – The Patent Classification System

My data draws on the full set of US utility patents granted between 1836 and 2009: 7.6 million patents. The US Patent and Trademark Office (USPTO) has developed the US Patent Classification System (USPCS) to organize patent and other technical documents by common subject matter. Subject matter can be divided into a major component, called a class, and a minor component, called a subclass. USPTO (2012) states, "A class generally delineates one technology from another. Subclasses delineate processes, structural features, and functional features of the subject matter encompassed within the scope of a class." Subclasses are a natural candidate for the building blocks of combination, out of which researchers build new ideas.

Using patent subclasses as proxies for the building blocks of ideas has many advantages over plausible alternatives, such as the words used in a patent document or citations to prior art. Unlike text or citations, patent classifications are chosen by an ostensibly disinterested party, namely the patent examiner. Classifications have no special legal standing and are not generally of interest to patent applicants (and therefore not chosen strategically). Instead, they are chosen to facilitate searches by future parties who wish to verify that new applications are, in fact, novel. Furthermore, the classification system is updated over time, with older patents assigned updated classifications as the system changes, so that searches of the patent record remain feasible. In contrast, the words used to describe common features may change with legal and aesthetic fashion but are not retroactively updated as the nomenclature changes.

There are more than 450 classes and more than 150,000 subclasses in the USPCS. To take two examples, class 014 corresponds to "bridges," and class 706 corresponds to "data processing (artificial intelligence)." A complete list of the current classes can be found on the USPTO website.[3] The subclasses are nested within each class and correspond to more fine-grained technological characteristics. For example, subclass 014/8 corresponds to "bridge; truss; arrangement; cantilever; suspension," while the subclass 706/29 corresponds to "data processing (artificial intelligence); neural network; structure; architecture; lattice."

Simply using the technology subclasses as elements to be combined is problematic because the categories may not correspond to the same level of specificity. For example, consider three subclasses, that all belong to class 706, "data processing (artificial intelligence)":

706/29: Data processing (artificial intelligence); neural network; structure; architecture; lattice.

706/15: Data processing (artificial intelligence); neural network.

706/45: Data processing (artificial intelligence); knowledge processing system.

[3] http://www.uspto.gov/web/patents/classification/selectnumwithtitle.htm

Classes 706/29 and 706/15 are both associated with neural networks, but at different levels of specificity, while 706/45 is not associated with neural networks at all. Without looking at the USPC index, it is impossible to know there is a relationship between some of the subclasses, but not others.

The uppermost subclass is called a mainline subclass, hereinafter "**mainline**." For example, the subclasses "bridge; truss," and "data processing (artificial intelligence); neural network," are both mainlines. The subclass nested one level down is said to be "one indent" in from the mainline. Within these one-indent subclasses will be still further subclasses, called two indent subclasses, and so on. I use technology mainlines as my primary elements of combination. This identifies a set comprising 13,517 elements, designed to be exhaustive and nonoverlapping.

3.2 – Assigning Each Patent A Combination of Mainlines

I observe the subclasses assigned to every patent[4] and for the reasons discussed above, I next collapse each technology subclass down to the mainline to which it belongs. For example, any instance of subclass 706/29, discussed above, is recoded as the mainline 706/15, since subclass 706/29 is a more specific description of the broader technology type described by mainline 706/15.

Out of 91.3 million possible mainline pairs, 1.75 million pairs are actually assigned to at least one patent over the period 1926-2009. Viewed through a combinatorial innovation lens, of the 91.3 million possible pairs, researchers have discovered how to usefully combine only 2%. Over the same period, the mean number of patents each pair belongs to over the entire period is 10.1, but the distribution is highly skewed: 51.2% of observed pairs are only ever assigned to one patent, but 46.1% of all pair assignments belong to 1% of pairs.

---

[4] U.S. Patent and Trademark Office (2014c). Technology classifications can be downloaded for free from http://patents.reedtech.com/classdata.php. I downloaded it in August 2014.

3.3 – Dates

Patents are sequentially numbered as they are granted, so that the year any patent is granted can be inferred from the patent number.[5] Once a patent is granted, the document becomes publicly available, and its content is disclosed. I assume the information in a patent is known to other researchers beginning in the year the patent is granted.

I use the year of a patent's application to denote the year researchers develop an idea. This information is not available for all patents, but Kogan et al (2015) extracts patent application years for every US patent from 1926-2009. Thus, although I use patent data from 1836 to construct measures of researcher knowledge, I only examine patenting behavior for the period 1926-2009. There were 6.0 million patents granted during this period.

4 – Methodology

4.1 – Sample

To test hypotheses 1-4, I would like to estimate equation (6), which predicts the expected number of patent applications (per year) that are assigned a particular set of three mainlines, as a function of three explanatory variables of interest: (1) the number of times each pair of mainlines in the set of three has been assigned to other patents, (2) the number of years the set of three mainlines has been available, and (3) the number of times the set of three mainlines together has been assigned to a patent. A straightforward way to achieve this is to compute these variables for every combination of three mainlines in the dataset and run a count-model regression.

This straightforward strategy is computationally infeasible. With 13,517 mainlines in my dataset, there are 4.1×10^{11} unique sets of three mainlines, each of which has annual observations for the

---

[5] This can be inferred from US Patent and Trademark Office (2014a).

period 1926-2009, totaling more than a trillion data points. To obtain a more manageable dataset I restrict my attention to mainline-sets that are assigned to a patent at least once during the period 1926-2009. I draw a sample encompassing yearly observations on 10,000 randomly selected mainline-sets (used at least once). This gives me an unbalanced panel of 800,576 mainline-set/year observations.

Because I am restricting my attention to mainline-sets used at least once, my results are conditional and do not apply to a randomly selected set of three mainlines, but only to a set that is assigned to a patent between 1926 and 2009. Estimating an unconditional model is made very difficult by the extreme rarity of a set of three mainlines actually being assigned to a patent. From the 4.1×10^{11} possible sets of three mainlines, only 495,369 are ever actually used at any point (less than 1 in 800,000). In contrast, once we select a mainline-set used at some point, it is actually assigned to a patent in 2.2% of years. Although the empirical exercise is restricted to this conditional dataset, in the next section I provide summary statistics for a complementary sample of 10,000 mainline-sets never assigned to any patent.

4.2 – Measures

Measures of interest are described in Table 1.

Table 1 here

Observations lie in the interval 1926-2009, but are constructed from data stretching back to 1836. These measures are best expressed with an example. Consider the following set of three mainlines:

123/319: Internal Combustion Engine; Engine Speed Regulator

477/34 : Interrelated Power Deliver Controls, Including Engine Control; Transmission Control

701/1: Data Processing: Vehicles, Navigation, and Relative Location; Vehicle Control, Guidance, Operation, or Indication

In 1998, 4 patents were applied for that were assigned these three mainlines, so that $1(y_{it} > 0) = 1$ and $y_{it} = 4$. The following year, no patent applications using these three mainlines occurred, so that $1(y_{it} > 0) = y_{it} = 0$.

Mainline 123/319 was first assigned to a patent in 1860, mainline 477/34 in 1887, and 701/1 in 1923. This last example is an illustration of how the patent office updates technology classifications over time: 701/1 was first assigned to patent 1,459,106 – "Gasoline-consumption indicator for motor vehicles" which was granted in 1923. I assume that since 701/1 was first assigned in 1923, it was only feasible to combine these three technologies beginning in that year, and the "Availability Age" in 1998 is 75 years ($age_{it} = 75$). Between 1923 and 1998, 5 other patents had already been granted that were assigned the exact same set of mainlines, so that $N_{it} = 5$.

By 1998, each of the pairs of mainlines had been used a large number of times. As of 1998, 178 patents had been granted that were assigned mainlines 123/319 and 477/34 (although not necessarily just these two mainlines). In the same year, mainlines 123/319 and 701/1 had been assigned together to patents 320 times and mainlines 477/34 and 701/1 had been jointly assigned 808 times. This data is used to estimate the probability of patentability for this mainline-set as follows:

Perfect substitutes: $$\sum_{j \in i} \tilde{a}_{jt} = \beta_1 (178 + 320 + 808) + \beta_2 (178^2 + 320^2 + 808^2)$$

Perfect complements: $$\min_{j \in i} (\tilde{a}_{jt}) = \beta_1 \times 178 + \beta_2 \times 178^2$$

Table 2 presents some summary statistics for my dataset.

Table 2 here

As noted earlier, in most years, no patent applications are made that are assigned a given set of three mainlines, so that $1(y_{it} > 0) = 0$ in most cases. Conditional on $y_{it} > 0$, the average value of y_{it} is a little over 1, with a maximum of 51.

The median Availability Age of a given set of three mainlines is slightly under 90 years: most mainlines in my dataset have been available as a combination for a long time. For comparison, if each set was available in 1836, then the mean age for observations in 1926-2009 would be 131.5. The mean value of Mainline-Set Count is just under 1, indicating most are used once and never again. Turning to data on pairs, we see pairs of mainlines are used together much more commonly. The minimum Pair Count in a set of three still has a mean value of 15.6.

It is important to note that this data is not representative of a randomly chosen set of three mainlines. Rather, it reflects the characteristics of mainlines eventually assigned to one patent over the period 1926-2009. For comparison, Table 2 provides statistics on a random set of 10,000 mainline-sets that are not assigned to a patent over the same period. These sets of mainlines make up the vast majority of possible combinations.

Table 3 here

Table 3 does not include data on y_{it} or Tr_{it} because these are all zero when restricting attention to mainline-sets that were never assigned to a patent.

The first thing to note is that unused sets of three mainlines tend to have a much lower Availability Age than those that are used at some point. The biggest difference, however, is in the number of times pairs making up a mainline-set are used. Note that $\sum_{j \in i} n_{jt}$ has a median of 0 and a mean of 0.293 for mainline-sets that are never used, compared to 108 and 364.4 for sets that are used. Sets of mainlines that are eventually used tend to have far more history of using their components together.

Tables 4 and 5 presents evidence on the correlation across these measures.

Table 4 and 5 here

In all cases, measures are positively correlated, ranging from a minimum of 0.009 (between age and y_{it} when we restrict attention to observations with $y_{it} > 0$) to a maximum of 0.736 (between N_{it} and $\min_{j \in i}\left(n_{jt}\right)$ when we restrict attention to observations with $y_{it} > 0$). It is notable that in a simple correlation framework, $1\left(y_{it} > 0\right)$ and $y_{it} \mid y_{it} > 0$ are more highly correlated with $\min_{j \in i}\left(n_{jt}\right)$ than $\sum_{j \in i} n_{jt}$.

4.3 - Estimation

Because it is relatively rare that a mainline-set is assigned to a patent in a given year, I use a two-stage estimation model:

$$E\left[y_{it}\right] = E\left[y_{it} \mid y_{it} > 0\right] \times \Pr\left(y_{it} > 0\right) \tag{12}$$

To estimate $\Pr\left(y_{it} > 0\right)$, my baseline model is a logit model. I include as explanatory variables my metric for perfect substitutes or complements (see equations (10) and (11)), and quadratic terms for Availability Age and Mainline-Set Count. I also include a time trend in the baseline model. In the baseline, I estimate clustered standard errors by resampling with replacement on mainline-sets and re-estimating coefficients.

There are potential omitted variables in this approach. One potential source of bias is the persistent differences among technologies. In the framework of this paper's expositional model, it might be that the cost of research on ideas using mainline-set i (k_i) or the value of these ideas (π_i) varies systematically across i, or that the imperfect nature of my measure of technological components means λ_{it} varies systematically with i. This would tend to bias the coefficient on N_{it}, since N_{it} would be proxying for the underlying tendencies of different mainline sets to be implemented, in addition to the fishing out effect we are interested in.

A standard way to control for persistent technological differences is to assign patents to different technological categories based on the single *primary class* assigned to each patent by the USPTO. However, because my unit of observation is a set of three mainlines rather than an individual patent, this is not feasible. In many cases, patents assigned the same set of three mainlines are given differing primary classes. Instead, in some specifications I use the Chamberlain estimator to strip out fixed effects from each mainline set i.

A second potential source of bias is variation in the propensity to innovate or patent over time. To address the potential for non-linear variation in the underlying propensity to innovate and patent, in some specifications we substitute time fixed effects for a linear time trend.

To estimate $E\left[y_{it} \mid y_{it} > 0\right]$ I run count models using either a poisson or negative binomial model, truncated below 1. In both cases, I include only observations where $y_{it} > 0$. This dramatically reduces the number of observations to just 2.2% of the original, as indicated in Table 1. I estimate clustered standard errors by resampling with replacement on mainline-sets and re-estimating coefficients.

5 – Results

My results are presented in two tables, one of which uses the Perfect Substitutes framework (Table 6 and equation (10)) and the other uses the Perfect Complements framework (Table 7 and equation (11)).

Tables 6 and 7 here

For the logit models (columns 1-3), estimated coefficient are different from zero with a *p*-value of under 0.001. For the poisson and negative binomial models (columns 4-5), we cannot reject the null that the coefficient on Availability Age is equal to zero, but the rest of the coefficients are different from zero with a *p-value* of at least 0.05.

In Figure 1, we combine the models from column (1) and column (4) as in equation (12) to diagram the relationship between our measure of affinity and the expected number of patent applications using a particular set of mainlines.

Figure 1 here

To compute Figure 1, I use the estimated coefficients in Columns 1 and 4 of Tables 4 and 5 and compute:

$$\Pr\left(y_{it} > 0\right) = \exp\left(X'\beta\right) / \left(1 + \exp\left(X'\beta\right)\right) \tag{13}$$

$$E\left[y_{it} \mid y_{it} > 0\right] = \exp\left(X'\beta\right) / \left(1 - \exp\left(-\exp\left(X'\beta\right)\right)\right) \tag{14}$$

Marginal and direct effects in non-linear models cannot be separated from other explanatory variables, and so I illustrate the impact of changing Pair Use for three different hypothetical examples. I variously assign all other explanatory variables to be (1) the mean values given in Table 1, (2) mean values plus one standard deviation, and (3) mean values minus one standard deviation. For Figure 1 Left (Perfect Substitutes), the horizontal axis corresponds to $\sum_{j \in i} n_{jt}$ and a there is no direct way to transform $\sum_{j \in i} n_{jt}$ into the $\sum_{j \in i} n_{jt}^2$ required of the model. However, a simple approximation:

$$\sum_{j \in i} n_{jt}^2 = 0.8752 \times \left[\sum_{j \in i} n_{jt}\right]^2 + \varepsilon \tag{15}$$

fits the data very well, with an $R^2 = 0.96$. Therefore, I approximate $\sum_{j\in i} n_{jt}^2$ with $0.8752 \times \left[\sum_{j\in i} n_{jt}\right]^2$ purely for the sake of illustration in Figure 1. For Figure 1 right (Perfect Complements), the horizontal axis corresponds to $\min_{j\in i}\left(n_{jt}\right)$ and it is straightforward to compute the additional explanatory variable $\left[\min_{j\in i}\left(n_{jt}\right)\right]^2$.

These figures provide solid support for Hypothesis 1: there is an increasing but bounded contribution of n_{jt} to $E\left[y_{it}\right]$. This is consistent with a learning story. The results of other models provide further evidence that this relationship is robust.

The empirical exercise also support hypothesis 2. In Table 4, the model with the minimum value of $-\beta_1 / 2\beta_2$ is column 4, where $-\beta_1 / 2\beta_2 = 4,551$. In my dataset, 99.74% of all observations of n_{jt} lie below 4,551, so that their marginal contribution is positive. In Table 5, the model with the minimum value of $-\beta_1 / 2\beta_2$ is column 1, with $-\beta_1 / 2\beta_2 = 1,259$. In my dataset, 99.97% of observations of $\min_{j\in i}\left(n_{jt}\right)$ lie below this range.

Hypotheses 3 and 4 find equivocal support. The coefficients attached to N_{it} and N_{it}^2 take on the expected sign when I include fixed effects for each set of mainlines (column 2), but otherwise do not. This is consistent with the presence of fixed effects at the level of a mainline-set, which would tend to bias estimates upwards as anticipated. For Table 4, Column 2, $-\phi_1 / 2\phi_2 = 205.7$ while in Table 5, Column 2, $-\phi_1 / 2\phi_2 = 184.9$. In my dataset, 99.99% of observations of N_{it} lie below 184.9. So long as we include mainline-set level fixed effects, each patent assigned mainline-set i reduces the expected number of further such applications for the vast majority of observations.

The coefficients on Available Age offer a more complicated story, but are in line with hypothesis 4 over a large range of observations. Unlike the other estimated coefficients, the relationship between patent applications and Availability Age is non-monotonic. Depending on the specification chosen from Tables 4 and 5, $-\alpha_1 / 2\alpha_2 \in [41.6, 123.7]$ with a mean of 68.7 years and a median of 63.5 years. In the early years after a set of mainlines becomes available, each passing year increases the expected number of patents applications making use of the mainlines. After approximately 65 years, additional years begin to subtract from the expected number of patent applications that will be made using the mainlines. In my dataset, 18.6% of observations have an age above 123.7, but 69.5% of observations have an age above 68.7 years, so that the marginal contribution of an extra year is negative for a majority of observations in most specifications (but not in the one that best fits hypothesis 3). In Figure 2, we combine the models from column (1) and column (4) of Table 4, as in equation (12), to diagram the relationship between the availability age of a mainline-set and the expected number of patent applications using a particular set of mainlines. For the values of the other explanatory variables, I again use mean values plus or minus one standard deviation.

Figure 2 here

Because the Logit is bounded from below by 0 and a positive poisson is bounded from below by 1, the expected convex shape emerges (above 65 years), even though the quadratic $\alpha_1 \times age_{it} + \alpha_2 \times age_{it}^2$ is concave.

Both the Perfect Substitutes and Perfect Complements frameworks do well in predicting the number of patent applications in a given year. Of the two, the Perfect Complements approach performs slightly better in terms of the log likelihood of the model, and has a less flat relationship between the expected number of patent applications and the value of the explanatory variable (a point made well by the differences between the y-axes of Figure 1's left and right sides). However, in columns 4 and 5, the coefficients on the perfect substitutes value have slightly better *p* values.

6 – Discussion

Whenever a viable idea is discovered, there are two opposing effects. The successful combination of technological components has a positive learning effect, because it raises beliefs about the affinity between components (firms believe the components will be compatible in more settings). It also has a negative effect, because it uses up one possible combination of technological components. Which effect dominates depends on how much firms already know about the affinity of the pairs in question. Because we were able to separately identify the learning and fishing out effects, we can use our reduced form model estimates to shed some light on when each effect dominates.

First, consider these two effects on a particular set of mainlines. Whenever a mainline set is patented, it increases both N_{it} and $\min_{j\in i}\left(n_{jt}\right)$ by one. When using the Perfect Complements framework and defining $\underline{n}_{jt} \equiv \min_{j\in i}\left(n_{jt}\right)$ to economize on space, the fishing out effect dominates the learning effect at the level of a mainline-set if the following condition holds:

$$\beta_1\left(\underline{n}_{jt}+1\right)+\beta_2\left(\underline{n}_{jt}+1\right)^2+\phi_1\left(N_{it}+1\right)+\phi_2\left(N_{it}+1\right)^2<\beta_1\underline{n}_{jt}+\beta_2\underline{n}_{jt}^2+\phi_1 N_{it}+\phi_2 N_{it}^2 \qquad (16)$$

Which can be expressed as:

$$-\frac{1}{2\beta_2}\left\{\beta_1+\beta_2+\phi_1+\phi_2+2\phi_2 N_{it}\right\}<\underline{n}_{jt} \qquad (17)$$

Using the coefficients from Table 7, column 2 (the fixed effect model) and converting into consistent units (remembering that we measured n_{jt} in 1,000s and N_{it} in 100s to facilitate display in the table), the condition is:

$$120.6N_{it}-20,355<\underline{n}_{jt} \qquad (18)$$

Note the minimum value of $\underline{n}_{jt}$ is zero, and $\underline{n}_{jt}\geq N_{it}$ be definition, implying the fishing out effect always dominates for $N_{it}<170.2$ (corresponding to nearly all observations). Taking the reduced form model literally, it is possible for the learning effect to dominate above $N_{it} = 170.2$ if n_{jt} is sufficiently small (implying there are few other patents using the pair), because by this point the negative effect of fishing out has been dissipated. However, the underlying theory justifying the reduced form model implies the fishing out effect only disappears when there are no ideas left to try, which would mean there are no untried ideas left to apply better information about affinity towards.

When considering a particular set of mainlines, the fishing out effect dominates the learning effect: every time that set of mainlines is combined, the expected number of patents that will use this set in the future declines. This does not imply the fishing out effect dominates the learning effect on the whole though, because every successful combination has positive spillovers for a large number of other ideas.

To investigate the magnitude of the effect, we use the coefficients in columns 1 and 4 of Table 7 to estimate the change in the expected number of patent applications, per year, for each observation in the dataset if we increased $\min_{j \in i}\left(n_{jt}\right)$ by 1. Across all observation, the average increase in the expected number of patent applications per year, if we increase $\min_{j \in i}\left(n_{jt}\right)$ by 1, is 0.00013.

Restricting attention to the set of 495,369 mainline-sets that are combined at some point in 1926-2009, a typical set has a pair of mainlines in common with 50 other mainline-sets. If we assume there is a 1 in 3 chance that the shared mainline is the pair with the minimum n_{jt} (while this is probably an overestimate, neither is it true that pairs are perfect complements and only the minimum n_{jt} contributes), then each time a mainline-set is patented, it increases the expected number of patent applications for other combinations by $0.00013 \times 50 / 3 = 0.0022$ per year, or approximately 0.11 patent applications in a 50 year period.

If we assume, pessimistically, that every set of three mainlines can be used only once, then the learning effect, even with spillovers, is insufficient to overcome the fishing out effect. Each patent reduces the number of future patents by 1 (via the fishing out effect), but increases it by just 0.11 (via the learning effect) over 50 years. Taking our model seriously, unless every new patent increases the expected affinity of its constituent pairs enough to induce 1 or more additional patents in expectation, the number of patents will fall to zero over time.

That said, our rough estimate that each new patent induces 0.11 additional patent applications in 50 years is an underestimate for two reasons. First, we have restricted our attention to sets of mainlines that are used at some point. Each pair of mainlines in a set of three can be combined with one of the remaining 13,515 mainlines to form a new set of three, implying the set of potential spillovers is far higher than 50. Second, this analysis has restricted its attention to sets of three mainlines, in order to separately identify the learning and fishing out effects. Although this is beyond the scope of this paper, spillovers may also occur to combinations of two, four, or more mainlines. If, between these two omissions, the number of patent applications induced by a successful combination of patents is increased by a factor of just over 9, then innovation becomes self-perpetuating in the sense that every patent generates enough new knowledge to facilitate development of yet another patent to replace the fished out patent. This effect grows with the number of technological components available for recombination.

7 – Conclusions

This paper has provided novel evidence for a combinatorial model of innovation. I describe a simple model where ideas are created by combining pre-existing technological elements. The probability a given combination of technological elements yields a viable idea is a function of the affinity elements have for each other, a measure of how frequently the elements prove to be compatible. Firms learn the affinity between elements by observing which elements have been successfully combined in the past.

To test this model, I use the US patent office's technology classification system to proxy for unobserved technological elements that are combined to create ideas. Using a sample of 10,000 different combinations of technologies, I estimate a model linking the expected number of patents using a given technology combination to various explanatory variables. Each of the technology combinations considered is comprised of multiple *pairs* of components. I show the expected number of patents using a technology combination rises when these pairs have been shown to be compatible, but only up to an upper bound (which I argue corresponds to a belief that the components are certain to yield a viable idea, though not necessarily one valuable enough to justify R&D costs). I call this the learning effect.

This effect, however, is offset by another. I also show the expected number of patents using a technology combination falls when that exact combination has been used more often (but only when I include individual fixed effects). I argue this represents a fishing out effect. There are a finite number of ways to combine technologies, and innovation extracts these like any other finite resource. I find some evidence that the expected number of patents using a technology combination also falls with time, which I interpret as proxying for unobserved attempts to develop ideas.

Which effect is stronger? In my model, the fishing out effect appears to dominate. For a given technology combination *ABC*, firms do not learn enough from successful versions of *ABC* to offset the exhaustion of the combination. At the industry-wide level, the learning effect is stronger, because discovering *ABC* is a viable combination provides useful information about many other ideas (*ABD, ABE, ABF…*) but only fishes out the combination *ABC*. My results suggest the learning effect, even when we take into account spillovers, is not large enough to overcome the fishing out effect. However, this result is almost certainly a lower bound, and should be interpreted with caution. It is based on a relatively small sample and the learning effect grows with the number of ideas available for spillover.

Future research could clarify this issue by expanding analysis to the full set of patents and potential ideas, to determine whether the fishing out effect will dominate the learning effect over long time horizons (it certainly hasn't yet). Other limitations of the research include the simplistic modeling of R&D decisions as myopic and binary, rather than forward-looking and continuous. Furthermore, this model has taken the set of technological elements available for combination as fixed and given. In fact, the set of elements grows over time. Incorporating this dynamic endogenously into the framework would be another possible avenue for future research.

References

Arthur, W. Brian. 2009. *The Nature of Technology: What It Is and How It Evolves.* New York: Free Press.

Akcigit, Ufuk, William R. Kerr, and Tom Nicholas. (2013). The Mechanics of Endogenous Innovation and Growth: Evidence from Historial U.S. Patents. Working Paper.

Auerswald, Philip, Stuart Kauffman, Jose Lobo, and Karl Shell. 2000. "The Production Recipes Approach to Modeling Technological Innovation: An Application to Learning By Doing." *Journal of Economic Dynamics and Control 24* 389-450.

Dartnell, Lewis. (2014). *The Knowledge: How to Rebuild Our World From Scratch*. New York, NY: Penguin Press.

Fleming, Lee. 2001. "Recombinant Uncertainty in Technological Search." *Management Science 47(1)* 117-132.

Jovanovic, Boyan and Rafael Rob. 1990. "Long Waves and Short Waves: Growth Through Intensive and Extensive Search." *Econometrica* 1391-1409.

Kauffman, Stuart, Jose Lobo, and William G. Macready. 2000. "Optimal Search on a Technology Landscape." *Journal of Economic Behaviour and Organization 43* 141-166.

Keijl, S., V.A. Gilsing, J. Knoben, and G. Duysters. 2016. “The Two Faces of Inventions: The Relationship Between Recombination and Impact in Pharmaceutical Biotechnology.” Research Policy 45(5): 1061-1074.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman. (2015). Technological Innovation, Resource Allocation and Growth. http://ssrn.com/abstract=2193068.

Nemet, Gregory F. 2012. "Inter-technology Knowledge Spillovers for Energy Technologies." *Energy Economics 34* 1259-1270.

Nemet, Gregory F., and Evan Johnson. 2012. "Do Important Inventions Benefit from Knowledge Originating in Other Technological Domains?" *Research Policy 41* 190-200.

Ridley, Matt. 2010. *The Rational Optimist.* New York, NY: HarperCollins Publishers.

Schilling, Melissa A., and Elad Green. 2011. "Recombinant Search and Breakthrough Idea Generation: An Analysis of High Impact Papers in the Social Sciences." *Research Policy 40* 1321-1331.

Schoenmakers, Wilfred and Geert Duysters. 2010. "The Technological Origins of Radical Inventions." *Research Policy 39* 1051-1059.

US Patent and Trademark Office. 2012. "Overview of the U.S. Patent Classification System (USPC)." *The United States Patent and Trademark Office.* 12. Accessed 10 15, 2014. http://www.uspto.gov/patents/resources/classification/overview.pdf.

—. 2014a. *Table of Issue Years and Patent Numbers, for Selected Document Types Issued Since 1836.* March 26. Accessed October 16, 2014. http://www.uspto.gov/web/offices/ac/ido/oeip/taf/issuyear.htm.

—. 2014b. "U.S. Manual of Classification File (CTAF)." *USPTO Data Sets.* Accessed August 2014. http://patents.reedtech.com/classdata.php.

—. 2014c. "U.S. Patent Grant Master Classification File (MCF)." *USPTO Data Sets.* Accessed August 2014. http://patents.reedtech.com/classdata.php.

Weitzman, Martin L. 1998. "Recombinant Growth." *Quarterly Journal of Economics 2* 331-360.

Table 1. Measures

Variable	Name	Description	Intuition
y_{it}	Application Count	The number of patent applications in year t assigned mainline-set i	This is a dependent variable in some specifications.
$1(y_{it} > 0)$	Application Dummy	Dummy variable equal to 1 when $y_{it} > 0$, and 0 otherwise.	This is a dependent variable in some specifications.
age_{it}	Availability Age	The number of years all three mainlines in mainline set i have each been in use (each has been assigned to at least one patent)	This is a proxy for t in equation (9) and helps capture fishing out effects
N_{it}	Mainline-Set Count	The cumulative number of patents granted up through the current period and assigned the particular mainline-set i, and only this set.	This directly captures the fishing out of feasible ideas.
n_{jt}	Pair Count	The cumulative number of patents granted up through the current period and assigned mainline-pair *j*, possibly in addition to other mainlines.	These are an input into my measure of affinity.
t	Time	A time-trend rescaled to 1926=0.	Used as a control.

Table 2. Regression Data Summary Statistics

	Min	Median	Mean	Max	St. Dev.
$1\left(y_{it} > 0\right)$	0	0	0.022	1	0.146
$y_{it} \mid y_{it} > 0$	1	1	1.265	51	1.658
age_{it}	0	89	86.97	173	37.80
N_{it}	0	0	0.935	647	5.115
n_{jt}	0	13	121.5	19,230	515.3
$\sum_{j \in i} n_{jt}$	0	108	364.4	23,010	930.9
$\min_{j \in i}\left(n_{jt}\right)$	0	2	15.61	4,124	61.66
t	1926	1969	1969	2009	24.06

Table 3. Summary Stats on a Sample of Unassigned Mainline Pairs

	Min	Median	Mean	Max	St. Dev.
age_{it}	0	52	55.60	171	34.82
n_{jt}	0	0	0.098	684	3.210
$\sum_{j \in i} n_{jt}$	0	0	0.293	684	5.582
$\min_{j \in i}\left(n_{jt}\right)$	0	0	0.001	6	0.049
t	1926	1974	1972	2009	23.36

Table 4. Regression Data Correlations, all data

	age_{it}	Tr_{it}	$\sum_{j\in i} n_{jt}$	$\min_{j\in i}\left(n_{jt}\right)$	t
$1\left(y_{it} > 0\right)$	0.017	0.135	0.084	0.138	0.050
age_{it}		0.110	0.218	0.175	0.538
N_{it}			0.250	0.587	0.115
$\sum_{j\in i} n_{jt}$				0.453	0.265
$\min_{j\in i}\left(n_{jt}\right)$					0.172

Table 5. Regression Data Correlations, all data, $y_{it} > 0$

	age_{it}	Tr_{it}	$\sum_{j \in i} n_{jt}$	$\min_{j \in i}(n_{jt})$	t
$y_{it} \mid y_{it} > 0$	0.009	0.480	0.283	0.414	0.079
age_{it}		0.103	0.217	0.176	0.298
N_{it}			0.445	0.736	0.075
$\sum_{j \in i} n_{jt}$				0.670	0.244
$\min_{j \in i}(n_{jt})$					0.143

Table 6. Regression Results; Perfect Substitutes Framework

	Dependent Variable				
	$\Pr(y_{it} > 0)$	$\Pr(y_{it} > 0)$	$\Pr(y_{it} > 0)$	$E[y_{it} \mid y_{it} > 0]$	$E[y_{it} \mid y_{it} > 0]$
Time	0.014***			0.018***	0.019***
	(0.0006)			(0.005)	(0.004)
$\sum_{j \in i} n_{jt}$	4.730***	4.121***	4.893***	4.036***	4.241***
	(0.464)	(0.203)	(0.124)	(1.460)	(1.230)
$\sum_{j \in i} n_{jt}^2$	-4.426***	-2.793***	-4.518***	-4.434***	-4.354***
	(0.960)	(0.192)	(0.184)	(2.588)	(2.330)
age_{it}	1.507***	5.074***	1.525***	1.789	1.156
	(0.138)	(0.131)	(0.094)	(1.058)	(0.782)
age_{it}^2	-1.178***	-2.104***	-1.206***	-1.642*	-1.342**
	(0.081)	(0.066)	(0.055)	(0.696)	(0.474)
N_{it}	6.185***	-3.868***	6.598***	1.686***	7.343***
	(1.088)	(0.264)	(0.150)	(1.307)	(1.728)
N_{it}^2	-1.017***	0.938***	-1.073***	-0.270***	-1.154***
	(1.155)	(0.107)	(0.038)	(0.630)	(0.901)
Constant	-4.967***			-2.605***	-7.770***
	(0.060)			(0.393)	(0.790)
Observations	800,576	800,576	800,576	17,472	17,472
Distribution	Logit	Logit	Logit	Truncated Poisson	Truncated Neg. Bin.
Fixed Effects	None	Mainline-set	Time	None	None
Log Likelihood	-79,659.94	-62,887.78	-78,486.69	-9,333.157	-7,136.89

Notes: To make coefficients more readable, n_{jt} is measured in 10,000s, N_{it} is measured in 100s, and age_{it} is measured in centuries. Standard errors are in parentheses (clustered by bootstrapping in columns 1, 4, 5).
* = *p-value*<0.05, ** = *p-value*<0.01, *** = *p-value*<0.001

Table 7. Regression Results; Perfect Complements Framework

	Dependent Variable				
	$\Pr(y_{it}>0)$	$\Pr(y_{it}>0)$	$\Pr(y_{it}>0)$	$E[y_{it} \mid y_{it}>0]$	$E[y_{it} \mid y_{it}>0]$
Time	0.016***			0.024***	0.020***
	(0.0006)			(0.004)	(0.004)
$\min_{j\in i}(n_{jt})$	5.281***	4.388***	5.416***	3.074**	4.051**
	(0.957)	(0.221)	(0.121)	(0.918)	(1.024)
$\left[\min_{j\in i}(n_{jt})\right]^2$	-2.098***	-1.172***	-2.047***	-0.953*	-1.568*
	(1.222)	(0.083)	(0.078)	(0.427)	(0.902)
age_{it}	1.599***	5.443***	1.611***	1.395	1.120
	(0.143)	(0.131)	(0.094)	(0.868)	(0.781)
age_{it}^2	-1.239***	-2.200***	-1.262***	-1.465*	-1.347**
	(0.084)	(0.066)	(0.055)	(0.589)	(0.476)
N_{it}	4.747***	-5.224***	5.096***	1.384***	6.637***
	(1.090)	(0.294)	(0.163)	(1.365)	(1.776)
N_{it}^2	-0.696*	1.413***	-0.746***	-0.213***	-0.993***
	(1.099)	(0.113)	(0.041)	(0.009)	(0.054)
Constant	-4.993***			-2.698**	-7.745*
	(0.061)			(0.641)	(0.900)
Observations	800,576	800,576	800,576	17,472	17,472
Distribution	Logit	Logit	Logit	Truncated Poisson	Truncated Neg. Bin.
Fixed Effects	None	Mainline-set	Time	None	None
Log Likelihood	-79,659.94	-62,887.78	-78,3783.52	-9,169.848	-7,098.21

Notes: To make coefficients more readable, n_{jt} is measured in 1,000s, N_{it} is measured in 100s, and age_{it} is measured in centuries. Standard errors are in parentheses (clustered by bootstrapping in columns 1, 4, 5).
* = *p-value*<0.05, ** = *p-value*<0.01, *** = *p-value*<0.001

Figure 1. $E[y_{it}] = E[y_{it} \mid y_{it} > 0] \times \Pr(y_{it} > 0)$ as a function of $\sum_{j \in i} n_{jt}$ (left) and $\min_{j \in i}(n_{jt})$ (right)

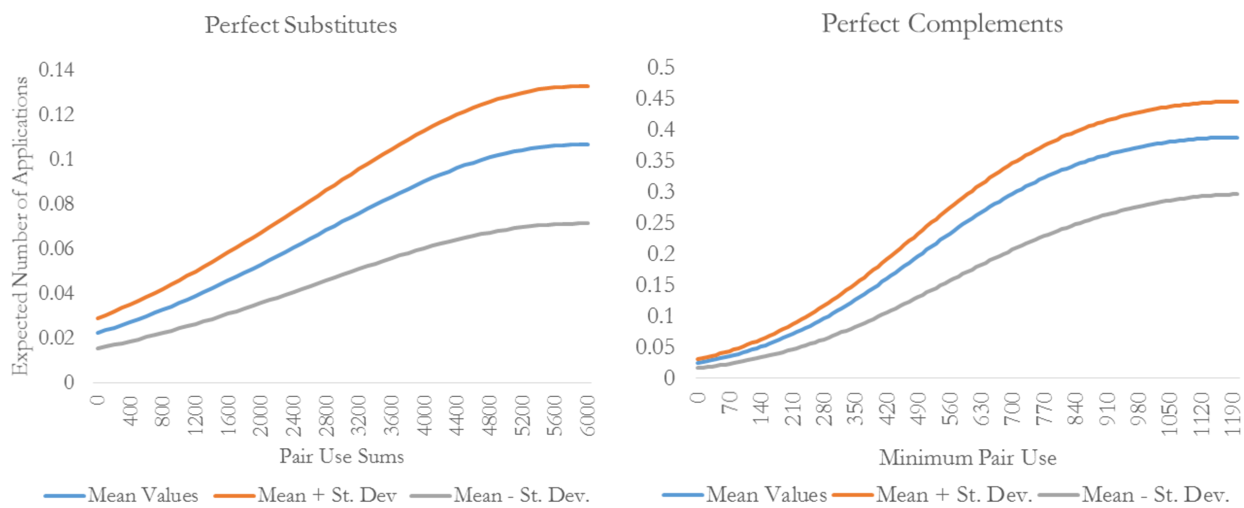

Figure 2. $E[y_{it}] = E[y_{it} \mid y_{it} > 0] \times \Pr(y_{it} > 0)$ as a function of Availability Age