Combinatorial Innovation: Evidence from Two Centuries of Patent Data

Matthew S. Clancy*

Abstract

This paper presents an original model of knowledge production, and tests several predictions of the model using a novel dataset built from 8.3 million US patents. In this model, new ideas are built by combining pre-existing technological building blocks into new combinations. The outcome of research is always stochastic, but firms are Bayesians who learn which sets of technological building blocks tend to yield useful discoveries and which do not. Consistent with this model's prediction, I show that the number of patents granted in a particular technology class increases in the years after new useful combinations of technology first appear in the class. Moreover, after new combinations first appear, I show subsequent patents are more likely to draw on the same combination of technology, consistent with firms learning the technologies can be fruitfully combined. Patents are also more likely to combine technologies that have already been combined with many of the same (other) technologies, even if they have never been combined with each other. Finally, I show that the probability of using a combination declines over time, and that the total number of patents granted in a technology class also declines over time, if there are not new connections between technologies continuously discovered. This is consistent with the model's predictions about firms using up all the useful ideas that can be built from a fixed set of technological building blocks.

Key Words: R&D, Research, Innovation, Patents, Knowledge, Bayesian

JEL codes: O31, O34

This Version: May 19, 2015.

* Matthew S. Clancy is a Ph.D. candidate, Department of Economics, Iowa State University, Ames, IA 50011. He was supported by a USDA NIFA National Needs Fellowship grant.

0 – Introduction

This is a paper about where ideas come from. There is a large and fruitful literature on the economics of innovation, but the nature by which ideas are found 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." We may add to Weitzman's "research effort" additional inputs such as human capital or measures of knowledge itself (e.g., the number of varieties of machines, or their quality), but the birth of ideas is usually not given strong microfoundations.

Of course, given the fecundity of the literature, this approach is justified in many applications. Nevertheless, there are questions that are difficult to answer without a better theory of how ideas are created. Gordon (2012), for example, argues the supply of good ideas has run out, a theme also discussed in Cowen (2011). Conversely, Brynjolfsson and McAffee (2014), expressing a view common to technology companies today, argue we are entering a period of brilliant technological progress. We need a theory of where ideas come from to approach this controversy. Such a theory would also be useful in addressing questions related to science policy: tightening budgets increasingly require agencies to make difficult decisions in funding different branches of knowledge, at different stages of exploratory and applied research.

This paper proposes a microfounded model of innovation, and tests some predictions of the model using with a novel dataset constructed from US patent data. It models the creative act as drawing on a pre-existing set of technological building blocks. These "components" must be assembled into a novel combination by an innovating firm to develop a new technology. The way these components interact with each other determines whether or not the new technology is useful. Moreover, I explicitly model the way firms *learn* how components tend to interact.

Consider the internal combustion engine as in illustrative 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 some shape or another, before the invention of the internal combustion, but the internal combustion engine 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, for example, would not yield up an engine. Instead, the connections between the elements matters, with all the elements working in harmony, rather than in opposition, to perform useful tasks. The combustible chemicals drive the pistons, the crankshaft converts this back and forth motion into jerky rotational motion, and 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 information. They knew how the components they intended to use were likely to interact, because the components had been used before in other contexts. Waterwheels also use crankshafts and pistons, and a potter's wheel couples a flywheel to a

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

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. In this framework for thinking about innovation, knowledge spills over in both the recycling of component parts, and in the understanding of how parts may be used together.

This framework yields a number of predictions. For example, firms will be more likely to combine building blocks that they know tend to work well together, or at least belong to a set of components that firms know to be mutually compatible. Moreover, the more technological building blocks that researchers know can be fruitfully combined with each other, the more ideas that will be worth developing. At the same time, the number of combinations that can be built from a finite set of building blocks is also finite, and over time, firms will run out of things to build.

I test these predictions using a novel dataset on US patents. I exploit the US Patent Office's Classification system to assign each patent a set of technological building blocks that comprise it. Each such building block is proxied by one of the patent office's technology *subclasses*, a designation considerably more fine-grained then the more typical technology classes that are used in most empirical patent papers. I show that aggregate patenting activity rises in the year after patents successfully combine subclasses that are otherwise rarely used together, and that new patents are more likely to draw on these newly combined classes. Moreover, I show that, in the absence of these new connections, patenting activity falls off. This is consistent with the predictions of my model.

I begin by providing some further background (section 1) before a formal model of combinatorial knowledge production is introduced (section 2). I then set up a model of a single inventor's problem (section 3). To motivate my empirical application, I make a series of predictions when there are multiple firms operating within the same technological sector (section 4). This model predicts the number of ideas produced declines over time, unless useful connections between elements are continually discovered. It also predicts new connections will be made where they are most complementary to existing connections, and that the relationship between a connection's past and future use is increasing. I construct a novel dataset from US patent data (section 5), which I use to test these predictions (sections 6 and 7). After some discussion of the results (section 8), I consider some directions for future research (section 9).

1 – Background

The internal combustion 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. Non-physical creations can also be understood as combinations. Works of fiction draw on a common set of themes, styles, character archetypes, and other tropes; musical compositions rely on combinations of instruments, playing styles, and other conventions; and paintings deploy common techniques, symbols, and conventions. Indeed, even abstract ideas can be understood as combinations of concepts, arguments, mathematical tools, facts, and so forth.

Weitzman (1998) is the first to incorporate this feature of knowledge creation into the economist's knowledge production function. In Weitzman's model, innovation consists of pairing "idea-cultivars"[2] to see if they yield a fruitful innovation (a new idea-cultivar), where the probability an idea-pair will bear fruit is an increasing function of research effort. If successful, the new idea-cultivar is included in the set of possible idea-cultivars that can be paired in the next period. Weitzman's main contribution is to show that combinatorial processes eventually grow at a rate faster than exponential growth processes, so that, absent some extreme assumptions about the cost of research, in the limit growth eventually becomes constrained by the share of income devoted to R&D rather than the supply of ideas. Simply put, combinatorial processes are so fecund that we will never run out of ideas, only the time needed to explore them all.

Weitzman's model is echoed in Arthur (2009), who views all technologies as hierarchical combinations of sub-components. Arthur agrees that the internal combustion engine is composed of pistons, crankshafts, flywheels, and so on, but goes further, pointing out that, say, the piston, is itself a combination of two shaped metal components, as well as lubrication, and so forth. These sub-components themselves are combinations of still further subcomponents (metal alloys, for instance). Ridley (2010) also proposes a model akin to Weitzman's, arguing the best innovations emerge when "ideas have sex."

These models display *re*combinant growth, wherein new combinations subsequently become the raw ingredients of future combinations. In contrast, this model is concerned primarily with the ways a *fixed* set of components can be reconfigured in many ways, and the interaction between two elements is used again and again to achieve different purposes. I have no doubt that Weitzman's model is appropriate for understanding where the elements of combination themselves come from, in the long run. This model is concerned with the shorter run, where the set of elements available for combination is relatively static and new elements are rare and unanticipated shocks to the set of available technological building blocks.

Several other papers have modeled the innovation process as one of learning about an underlying space of potential ideas. Jovanovic and Rob (1990) represents a technology by an infinite vector, each element of which ranges between 0 and 1. Technologies are production functions and agents learn the mapping from technology vectors to productivity via Bayesian updating. Research consists in changing the values of the elements in a vector and observing the labor productivity associated with the new vector. Kauffman, Lobo and Macready (2000) and Auerswald et al. (2000) follow Jovanovic and Rob (1990) in thinking of technologies as a large combination of distinct operations, although here the length of a technology vector is finite and each element can take on one of a finite number of states (rather than ranging over a continuous interval). The mapping between each technology vector and its productivity level is called a fitness landscape. When states are interdependent, the authors show this landscape is characterized by many local maxima. Innovation in such a model consists of exploring the fitness landscape by changing different operations.

The model presented in this paper differs from the above papers by modeling the learning process as focused on the relationships between components in an idea directly. To illustrate the difference, consider again the internal combustion engine. In my model, I asserted that an inventor would know how a piston and crankshaft were likely to interact in an internal combustion engine by observing how they interact in a waterwheel. Although the engine and the waterwheel are otherwise very

[2] So-called because the hybridization of ideas in the model is analogous to the hybridization of plant cultivars.

different, useful information can be extracted from the waterwheel that can be applied to the engine. In the models of Kauffman, Lobo and Macready (2000) or Jovanovic and Rob (1990), the usefulness of the waterwheel would instead be defined by the technological distance between the waterwheel and the engine, which is likely to be quite large. 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. The other papers instead consider the entire idea as a fixed point, and new possible ideas that are near the fixed point will have performance more highly correlated with it. This emphasis on the decomposability of technology into pair-wise interactions greatly simplifies empirical analysis.

Another line of literature uses a combinatorial framework in empirical applications. Such papers invariably use patents or academic papers as measures of innovation. There tend to be two ways of measuring the "components" of an idea. One approach is to use the citations of a patent or paper, generally grouped into technology or discipline categories, and another is to look at the technology or discipline categories assigned directly to a patent or paper, as this paper does. Meanwhile, a patent or paper's value tends to be measured by the number of citations it has received, in line with work by Trajtenberg (2002) and Harhoff et al. (1999) that shows such measures are correlated with independent measures of patent value. Studies using a combinatorial framework usually attempt to predict the value of patents (or papers) by using traits of the combination, such as the extent of unusual combinations or frequency with which ideas have been used. As far as I have been able to determine, this is the first paper that attempts to predict changes in aggregate patenting activity by employing a combinatorial innovation framework.

The paper most similar in spirit to this one is Fleming (2001), which examines a sample of 17,264 patents through a similar combinatorial lens. Fleming (2001) exploits the fact that most patents are assigned to more than one technological subclass, interpreting each sub-class assignment as a component. Fleming then simply counts the number of times such a combination of sub-classes has resulted in a patent, finding that less commonly used combinations obtain more citations. Fleming also computes a measure of combination familiarity, which is a weighted sum of the number of times a combination of sub-classes has been patented, with more recent combinations weighted more heavily. This familiarity metric, which Fleming interprets as accounting for the distance of search, is also associated with more citations. Other papers using patent data include Nemet (2012), Nemet and Johnson (2012), and Shoenmakers (2010). These papers all rely on citation data to proxy for the technological building blocks used by an idea, and tend to show unusual connections between technological fields in citations results in a higher citation rate. Schilling (2011), obtains a similar result for a sample of academic papers.

This paper differs from these other empirical papers in a number of respects. First, my dataset is considerably larger than previous studies, encompassing all 8.3 million US utility patents granted through 2012. Second, this paper does not use patent citation data. To measure the technological building blocks that comprise an idea, it uses patent subclasses, albeit in a novel way (discussed in section 5). One virtue of this approach is that a patent's technology classifications are supplied by a single ostensibly neutral arbiter, namely the patent office, while citations are made by both applicants and the patent office. Third, this paper develops an explicit theoretical model, from which predictions are derived (sections 2-4). These predictions do not require me to have any theory of citations. Instead, they pertain to the *number* of patents that will be granted in future period, as well as the probability a pair of technological components will be assigned to a patent. As far as I have been able to determine, mine is the first paper to show combinatorial factors have a measurable impact on the aggregate rate of future patent activity. I turn now to the presentation of this model.

2 – Model Basics

2.1 – The Knowledge Production Function

I now describe formally how ideas are created in this model.

> **Definition 1: Primitive Elements.** Let Q denote the set of primitive elements of knowledge q that can be combined with other elements to produce ideas, where $q \in Q$.

> **Definition 2: Pairs.** Let p denote a two-element subset of Q, or "pair," and P denote the set of two-element subsets of Q, where $p \in P$.

> **Definition 3: Ideas.** An *idea* d is a set of pairs p, satisfying the condition that if $p_0 \in d$ and $p_1 \in d$, then $p \in d$ for any $p \subseteq p_0 \cup p_1$.

A fixed set Q of technological building blocks can be assembled into ideas, where any idea must contain at least two q from the set Q. For convenience, I define ideas in terms of the *pairs* of elements contained therein. For example, an idea combining elements q_1, q_2, and q_3 is represented as the set of subsets $\left((q_1,q_2),(q_1,q_3),(q_2,q_3)\right)$. The condition attached to Definition 3 merely insures the pairs between all elements in the idea are included in the idea, so that we do not have ideas such as $\left((q_1,q_2),(q_1,q_3)\right)$, which uses elements q_1, q_2, and q_3 but does not include the pair corresponding to (q_2,q_3).

There are three important concepts in this model.

> **Definition 4: Compatibility.** The *compatibility* of pair p in idea d is $c(p,d) \in \{0,1\}$. When $c(p,d)=1$ then the pair p is compatible in d. When $c(p,d)=0$ then the pair p is incompatible in d.

Note that $c(p,d)=c(p,d')$ is not generally true. The compatibility of a pair may be equal to 1 in one idea and 0 in another.

> **Definition 5: Affinity.** The probability a pair p is compatible defines its *affinity*
>
> $a(p) \in [0,1]$.

The notions of compatibility and affinity are related as follows:

$$c(p,d) = \begin{cases} 1 & \text{with probability } a(p) \\ 0 & \text{with probability } 1-a(p) \end{cases} \tag{1}$$

Essentially, this model assumes pairs of elements have an underlying tendency to be compatible or incompatible, and this tendency is described by the affinity of the pair.

Lastly, ideas are either *effective* or *ineffective*, where an idea is effective if and only if all the pairs of its constituent elements are compatible.

> **Definition 6: Efficacy.** An idea d is *effective*, represented by $e(d)=1$, iff $c(p,d)=1\ \forall p \in d$. In all other cases, represented by $e(d)=0$, idea d is *ineffective*.

Restated, affinity determines the probability a pair is compatible, and when all pairs in an idea are compatible, the idea is effective. We may imagine ideas as sets of interacting elements that must be mutually compatible for the idea to prove useful. If any two elements are incompatible, I assume the idea suffers a catastrophic failure that renders it unfit for use. Note the probability an idea is effective can be written as:

$$\Pr(e(d)=1)=E[e(d)]=\prod_{p \in d} a(p) \tag{2}$$

This is the joint probability that every pair in the idea is compatible. Ideas are most likely to be effective when they are composed exclusively of elements that have a high affinity for each other, and least likely to be effective when composed of elements with a low affinity for each other.

2.2 – Beliefs

The affinity $a(p)$ between a pair of elements is *ex ante* unknown to researchers. Instead, researchers are Bayesians with prior beliefs over the possible distribution of $a(p)$. Though the researcher does not observe $a(p)$ directly, she can make educated guesses based on the tendency of p to be compatible or incompatible. Using her beliefs about the affinities between all pairs in an idea, she can compute the probability an idea will be effective. In more formal terms, a crucial part of the discovery process is the inference of likely affinity values from the compatibility or incompatibility of component-pair interactions.

I impose one assumption on the researcher's beliefs:

> **Assumption 1: Independence of Affinity.** The researcher believes $a(p)$ is independently distributed for all p.

As long as this assumption stands, the updating of beliefs about any $a(p)$ depends only on observations on the pair p alone. If $a(p)$ were not independently distributed, it would be necessary to also take into account the observations on correlated pairs, greatly complicating the problem.

Each observation of compatibility is the outcome of a Bernoulli trial governed by the pair's true affinity, with the two possible states being compatibility (probability $a(p)$) or incompatibility (probability $1-a(p)$). Given s instances of compatibility ("success") and f instance of incompatibility ("failure"), the researcher updates her beliefs according to Bayes law under the Bernoulli distribution:

$$\Pr\left(a(p)=\tilde{a}\mid s,f\right)=\binom{s+f}{s}\tilde{a}^{s}\left(1-\tilde{a}\right)^{f}\frac{\Pr\left(a(p)=\tilde{a}\right)}{\int_{0}^{1}\binom{s+f}{s}a^{s}\left(1-a\right)^{f}\Pr\left(a(p)=a\right)da} \tag{3}$$

where $\binom{s+f}{s}\tilde{a}^{s}\left(1-\tilde{a}\right)^{f}=\Pr\left(s,f\mid a(p)=\tilde{a}\right)$.

The expected value of $a(p)$ is given by:

$$E\left[a(p)\mid s,f\right]=\frac{\int_{0}^{1}a^{s+1}\left(1-a\right)^{f}\Pr\left(a(p)=a\right)da}{\int_{0}^{1}a^{s}\left(1-a\right)^{f}\Pr\left(a(p)=a\right)da} \tag{4}$$

Equation (4) says the expected value of $a(p)$ is an integral over all possible values of $a(p)$, where every a is weighted by $a^{s}\left(1-a\right)^{f}$ and a factor that normalizes the sum of probabilities to 1. As s increases, the relative weight attached to higher values of a increases more than the weight attached to low values, and $E\left[a(p)\mid s,f\right]$ increases. In the limit, $E\left[a(p)\mid s,f\right]$ converges to 1, as $a^{s}\left(1-a\right)^{f}$ goes to zero for all $a\neq 1$. Conversely, $E\left[a(p)\mid s,f\right]$ decreases as f increases, and converges to 0 as f grows large relative to s. In other words, researchers believe a pair is more likely to be compatible in the future if it has been compatible in the past, and vice-versa.

The term $a^{s}\left(1-a\right)^{f}$ can also be written as $\left\{a^{x}\left(1-a\right)^{1-x}\right\}^{n}$ where $x=s/n$ and $n=s+f$. The term $a^{x}\left(1-a\right)^{1-x}$ attains its maximum when $a=x$, so that as n increases, the expected value of $a(p)$ converges to s/n. In other words, as the number of observations increases, researchers come to believe the affinity of a pair is just equal to the proportion of times the pair has been observed compatible.

Given equation (2) and Assumption 1, the expected efficacy of an idea when researcher beliefs are uncertain can be written as:

$$E\left[e(d)\right]=\prod_{p\in d}E\left[a(p)\mid s(p),f(p)\right] \tag{5}$$

where $s(p)$ and $f(p)$ denote the number of observations of pair p's compatibility or incompatibility respectively. This expression captures the core notion of this model: ideas are more likely to be effective ("useful") when composed of elements that have worked well together frequently in the past.

3 – The Firm's Problem

Suppose this production function is used by a firm trying to discover effective ideas. The firm knows every element in set Q, and in each period may choose to conduct a research project on some idea d built from the elements in Q.

> **Definition 7: Possible Ideas.** The set D_P is the set of all possible ideas that can be made from elements in Q. It contains all subsets of P that satisfy the condition in Definition 3.

> **Definition 8: Eligible Ideas.** A set of eligible ideas $\tilde{D}$ is a subset of D_P. It is only sensible to conduct research projects on eligible ideas, and when a research project is attempted, the idea is removed from $\tilde{D}$ at the end of the period.

The set $\tilde{D}$ is primarily intended to indicate the set of *untried* ideas, and so it shrinks as research proceeds. I add to this set an additional element, the null set $d_0 \equiv \{\varnothing\}$, which represents the option not to conduct research in a period.

> **Definition 9: Available Actions.** The firm's set of available actions is $D \equiv d_0 \cup \tilde{D}$.

Note that because $d_0 \notin \tilde{D}$, if the firm chooses not to conduct research, then this option is not removed from its action set in the next period.

In principle, the firm "knows" every idea that can be built from elements in Q, in the same sense that I "know" every economics article that can be written with words and symbols in my repertoire. However, just as I do not know whether any of these articles are good until I think more about them, or actually write them out, the firm does not learn if an idea is effective until it decides to conduct research on it.[3] Indeed, research is costly, requiring investments of time and other resources. I assume that research on any idea has cost $k(d)$, known to the firm, and that the option d_0, to do nothing, has $k(d_0) = 0$.

The reward from conducting research may take the form of a patent that pays $\pi(d)$ to the patent holder at the end of the period. Patents may only be obtained for ideas that are eligible and which have been shown to be effective (patents are only issued for useful inventions). Because each chosen idea is removed from the set of eligible ideas D at the end of a period, firms cannot patent the same idea multiple times. I assume the patent value of the outside option d_0 is always zero.

Hence, a firm that chooses to conduct research on idea d expects to receive a net value of:

---

[3] Jorge Luis Borges tells a parable of an infinite library containing books with every combination of letter and punctuation mark. In this library, there is a book resolving the basic mysteries of humanity, since every possible book exists, but finding the book and verifying it is true amongst all the gibberish and babel is a daunting task for the library inhabitants. See Borges (1962).

$$\pi(d)E[e(d)]-k(d) \tag{6}$$

The idea is successful with probability equal to the expected efficacy $E[e(d)]$, in which case the researcher obtains the patent value $\pi(d)$. Whether the idea succeeds or not, the researcher pays up front research costs $k(d)$. This formulation of the innovator's problem is not unusual, except for the term $E[e(d)]$, which is determined by the knowledge production function described earlier and the beliefs of the researcher.

I summarize the firm's information about the prior number of compatibilities and incompatibilities of each pair by the vector I where:

$$I=((s(p_0),f(p_0)),...,(s(p_i),f(p_i)),...,(s(p_n),f(p_n))) \tag{7}$$

and where n is the number of pairs in P. I now assume that research on an idea d also reveals information on which pairs are compatible and which are not, in the form of the vector $\omega(d)$.[4] This vector has the same number of elements as B and is defined so that after conducting a research project on d, the updated information vector I' is given by:

$$I'=I+\omega(d) \tag{8}$$

For example, if a research project on some d' reveals pair p_0 is compatible and pair p_n is incompatible, and does not reveal any other information, then $\omega(d')$ takes the form:

$$\omega(d')=((1,0),(0,0),\ldots,(0,0),(0,1)) \tag{9}$$

Of course, $\omega(d_0)=((0,0),\ldots,(0,0))$ by assumption: agents learn nothing when they choose not to do a research project.

4 – Predictions

Consider a technology sector composed of many firms, all engaged in R&D that draws upon a common pool of elements Q. Suppose these firms behave more or less myopically.[5]

A research project is worth pursuing so long as its expected value is positive:

$$E[e(d)]\pi(d)-k(d)\geq 0 \tag{10}$$

---

[4] It may seem to be a natural assumption that research on d reveals the compatibility or incompatibility of *every* $p\in d$. I do not believe research is so straightforward. For the purposes of this paper, this assumption simplifies the exposition, without much cost. But see Clancy (2015) for a more complete discussion of these issues.

[5] In addition to the usual factors that may induce competitive firms to focus on the short-term, innovating firms will behave myopically if knowledge rapidly spills over. When this is the case, competitors can exploit research that yields a return in future periods (for example, by enabling firms to learn currently unprofitable projects are in fact profitable in expectation).

This can be rewritten as:

$$\prod_{p \in d} E\left[a(p)\right] \geq k(d) / \pi(d) \tag{11}$$

We do not typically observe the right-hand side of equation (11), but I assume it is known to firms. Equation (11) implies that any given idea is more likely to be pursued when the left-hand side is larger, or when the idea's expected affinities are higher. More broadly, within an industry engaged in R&D that draws upon a common pool of elements Q, there will be more ideas worth pursuing, and therefore more patents granted, when the expected affinity of pairs in P_Q is high. This is our first prediction:

> **Prediction 1: Number of Grants and Affinity.** The annual number of patents granted in a technology sector is positively correlated with the average expected affinity of pairs used by the sector.

This prediction will hold in any given year and does not require that the set of elements Q be fixed over time. Indeed, it seems likely to me that new elements are occasionally added to the set of technological building blocks used by a sector.[6] Over time, however, all ideas satisfying equation (11) are used up (since ideas can only be patented once). This is our second prediction:

> **Prediction 2: Number of Grants and Time.** The annual number of patents granted in a technology sector is negatively correlated with time.

Together, these two predictions imply the stock of available ideas is subjected to two opposing forces. First, a learning effect has the tendency to expand the set of research projects that are *ex ante* profitable to pursue. Second, a fishing out effect uses up profitable ideas.

Predictions 1 and 2 correspond to the aggregate activities of a technology sector, but predictions about individual pairs of elements are also possible. Define the term "use" as follows:

> **Definition 10: Pair Use.** A pair p is used in year t if there is any research project d attempted in year t such that $p \in d$.

In other words, a pair has been used if any research project is attempted on an idea that includes the pair as one of its constituents. For some idea that includes the pair p', equation (11) can be rearranged to yield:

$$E\left[a(p')\right] \geq \frac{k(d) / \pi(d)}{\prod_{p \in d \setminus p'} E\left[a(p)\right]} \tag{12}$$

[6] This could be due to recombinant growth as in Weitzman (1998), or by the discovery of entirely new physical processes that can be co-opted, as in Arthur (2009).

Just as an idea is worth attempting if it satisfies equation (11), the pair p' is worth *using* if there is some eligible idea that satisfies equation (12). Because this condition is more likely to be satisfied when the left-hand side of equation (12) is large, I can make the following prediction.

> **Prediction 3: Expected Affinity and Use.** In any given year, the higher the expected affinity of a pair, the more likely it is to be used.

As discussed above, however, over time research projects satisfying equation (12) are used up.

> **Prediction 4: Time and Use.** The probability a pair will be used decreases over time.

The numerator of equation (12) also highlights the complementarity between pair affinities. If $\prod_{p \in d \setminus p'} E[a(p)] = 0$ then there is no $E[a(p')]$ sufficiently high for the research project to be profitable in expectation. More generally, the higher is $\prod_{p \in d \setminus p'} E[a(p)]$, the more likely equation (12) is true. Furthermore, for the pair p' to be used, there just needs to be one idea that satisfies (12).

> **Prediction 5: Complementarity and Use.** The more pairs with high expected affinity that can be used in conjunction with some pair p' to form eligible ideas, the more likely the pair p' is to be used.

I now turn to the testing of these predictions with data on patents.

5 –Data

5.1 – Patent Data

To assess these predictions, I use data on patents granted by the US Patent and Trademark Office (USPTO). My data set includes the full set of US utility patents granted between 1836 and 2012, which amounts to 8.3 million patents. The year 1836 marks the beginning of the current patent numbering system, so that my dataset includes patent #1.[7]

Patents represent an imperfect but widespread and voluminous record of innovation outputs. To secure a patent grant, an idea is assessed on three criteria by a patent examiner. Patents must be novel relative to the prior art, nonobvious to someone with ordinary skills in the field, and useful, in the sense that it solves some problem.[8] These hurdles make patent grants a useful proxy for effective ideas, in the sense used by this paper.

To be sure, patents are not a perfect measure of ideas. Not all ideas are patented, or even patentable. Abstract ideas, for instance, cannot be patented. Moreover, ideas that are patent*able* may not be patent*ed*, even if they are novel, nonobvious, and useful, because patenting is not costless and requires divulging the details of the innovation.[9] Furthermore, patents can also be inappropriately

---

[7] Prior to the 1836 numbering scheme, an additional 9,957 patents were issued, which are not in my dataset. See US Patent and Trademark Office (2014a).

[8] See Clancy and Moschini (2013).

[9] See Clancy and Moschini (2013).

granted if the patent office is too lenient.[10] That said, no alternatives are clearly superior and patents provide an extensive source of microdata on individual innovations across a huge array of industries and years. No other single source provides coverage across hundreds of sectors and more than 100 years.

5.2 – Technology Mainlines as Elements

With patents proxying for the realization of an effective idea, a decision to be confronted is what to use as a proxy for the elements that are combined to build ideas. I choose to use the technology classifications assigned to each patent. The 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. The USPTO 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."[11] Subclasses are therefore a natural candidate for the elements of combination, out of which are built new ideas.

Patent subclasses as proxies for elements of combination have 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. These classifications are also meant to be exhaustive and non-overlapping, two desirable characteristics for our proxy. Finally, 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.[12] 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."

Subclasses are nested and hierarchical. The uppermost subclass is called a mainline subclass, hereafter simply "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. For example, the subclass "bridge; truss; arrangement"

---

10 See Bessen and Meurer (2008) for a study related to software patents.
11 US Patent and Trademark Office (2012b), pg I-1.
12 http://www.uspto.gov/web/patents/classification/selectnumwithtitle.htm

is one indent in from the mainline "bridge; truss." Within these one indent subclasses will be still further subclasses, called two indent subclasses, and so on.

Definition 11: Mainline. The USPCS subclass one indent in from a class.

A classification is assigned to a patent by the patent office with the following methodology.[13] The examiner has some portion of the patent, called the subject matter, he would like to assign to a subclass. Scanning through the list of mainlines in a class, the examiner stops when he finds a mainline that corresponds to the subject matter. The examiner then scans through the list of subclasses one indent in from the mainline. If none of the one indent subclasses apply to the subject matter, the examiner assigns the mainline to the patent. If one of the subclasses does apply, then the examiner repeats this process for the two indent subclasses that lie within the one indent subclass. The examiner then repeats the above process for three indent subclasses and so forth, until he arrives at a point where no deeper subclasses apply to the subject matter. At this point, the highest indent subclass (which will correspond to the most specific and narrow definition) found to be applicable is assigned to the patent. The USPCS is continually updated to reflect new technological categories, and patent classifications are updated as part of this process.

For every patent in my dataset, I observe both the year it was granted and the technology subclasses to which it is assigned. However, simply using the technology subclasses as elements to be combined is problematic because the categories may not correspond to the same level of specificity, since they are nested. 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.

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, we would not know there is a relationship between some of the subclasses, but not others.

Instead, I use technology mainlines as my primary elements of combination. This identifies a set comprising approximately 17,000 elements. Of these, approximately 13,000 are assigned to utility patents in my dataset (from here on, I restrict attention to the set of mainlines actually assigned). These mainlines are meant to be exhaustive and nonoverlapping. The mean number of mainlines used per class is 29.6, with a median of 21.

The number of mainlines per class varies widely. The maximum is 246 mainlines in one class, while 18 classes have just 1 mainline each. If each mainline is assumed to proxy for a specific technological building block, then ideally each mainline would cover roughly the same scope of technologies. However, we must remember that the technology classification system itself differs in how many technologies are encompassed in one class. For instance, the class 002, which corresponds to "apparel," has 33 mainlines in it. Class 004, which corresponds to the group "baths, closets, sinks,

[13] US Patent and Trademark Office (2012b), page I-13.

and spittoons," has 70. Because class 004 appears to tie together a more disparate set of technologies, more mainlines does not necessarily indicate that a mainline from class 002 covers a wider set of technologies than a mainline from class 004. Moreover, in some cases, the reverse happens and one type of technology is split into many classes. For example, classes 532-570 all correspond to organic compounds, and classes 520-528 all correspond to synthetic resins or natural rubbers. Of the 18 classes with one mainline each, 13 belong to one of these two series. In these cases, classes already divide up the space of technologies very finely, so that additional division into many mainlines is not necessary. Because defining the scope of what constitutes an element across different technologies is bound to be somewhat arbitrary, using mainlines as a proxy seems to me an appropriate first step.

5.3 – Assigning Each Patent A Combination of Mainlines

The USPTO makes available a large text document (U.S. Patent and Trademark Office 2014c) listing the technology subclass assigned to each patent, under the most recent classification scheme.[14] Each line of this text document contains a patent number, a subclass code, and an indicator for whether the subclass is the primary subclass (discussed more in Section 10). I extract from this document the subclasses assigned to each patent, as well as the identity of the primary subclass. I can then use the patent number to infer the year of the patent's grant.[15] For the reasons discussed above, I next collapse each technology subclass down to the mainline to which it belongs. For example, US patent 7,640,683 is titled "Method and apparatus for satellite positioning of earth-moving equipment" and describes a method of attaching antenna to the arm of an earthmoving machine in such a way that using satellite positioning systems is possible. This patent was assigned to four technological subcategories:

1. 37/348 – Excavating; ditcher; condition-responsive.
2. 414/699 – Material or article handling; vertically swinging load support; shovel or fork type; tilting; control means responsive to sensed condition
3. 701/50 – Data processing: vehicles, navigation and relative location; vehicle control guidance, operation, or indication; construction or agricultural vehicle type
4. 37/382 – Excavating; road grader-type; condition responsive.

Using the USPC index[16] I coded a program to reassign each subclass to its associated mainline. Applying this program to the above patent, I reclassify it as consisting of the following elements:

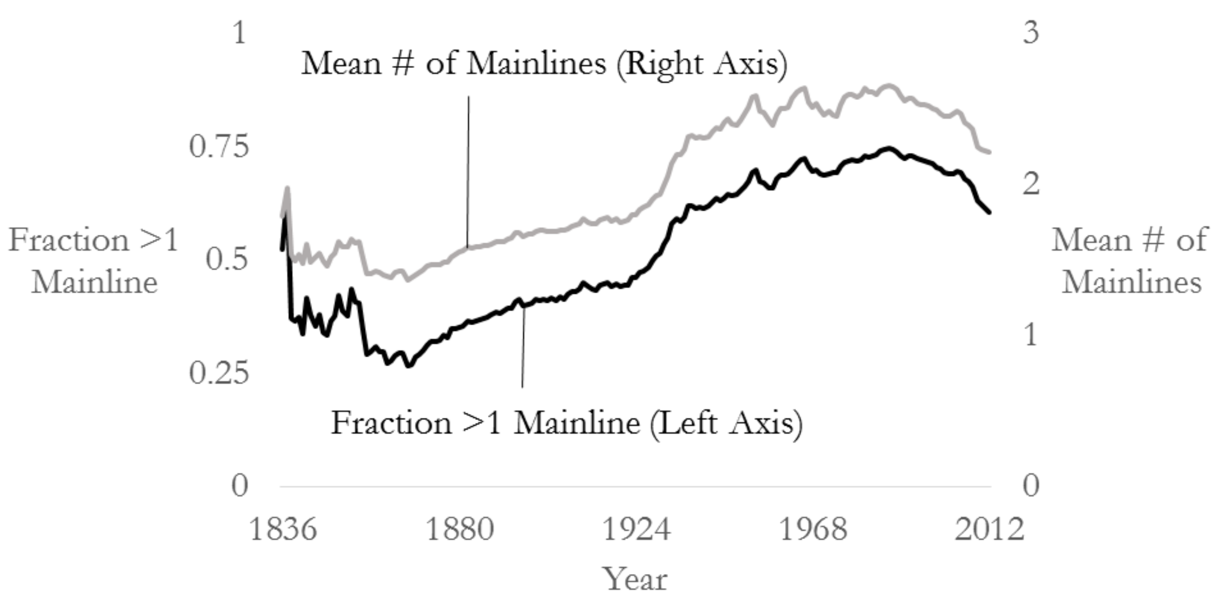
1. 37/347: Excavating; ditcher
2. 414/680: Material or article handling; vertically swinging load support
3. 701/1: Data processing: vehicles, navigation and relative location
4. 37/381: Excavating; road grader-type

[14] Technology classifications can be freely downloaded from http://patents.reedtech.com/classdata.php. I downloaded it in August 2014.
[15] This can be inferred from US Patent and Trademark Office (2014a).
[16] Specifically, I download the US Manual of Classification file from http://patents.reedtech.com/classdata.php, in August 2014.

Ideally, every patent would be comprised of two or more mainlines, because the model is premised on ideas as sets of pairs. In practice, after collapsing all patent assignments to mainlines, only 62.5% of patents are assigned more than one mainline. This share varies over time, averaging 40% over the period 1836-1935 and 69% between 1936 and 2012. A potential explanation is that new classification schemes consolidate commonly used pairs of mainlines into a single technology subclass. If this is the case, then we do not observe as many combinations of elements for older technologies, because combinations frequently used are re-classified as a single technology. I hope to explore this potential explanation in future work.

Over the entire period, the mean number of mainlines per patent is 2.29. The share of patents in a given year assigned more than one mainline, as well as the mean number of mainlines per patent, are each plotted in Figure 1. Over time, the number of elements in a patent has grown. While this may be the consequence of consolidation of subclasses, as discussed above, an alternative interpretation is that the complexity of patents has risen over time. The behavior of complexity in the context of this model is the subject of another research project.

Figure 1: Mainlines per Patent



There is also a converse problem of more than one patent being assigned the exact same set of mainlines. In the model, an idea is defined by the combination of elements that comprise it, so that no two ideas can share the same set of elements without being identical. However, mainlines are aggregations of multiple distinct technological features, since I have collapsed subclasses into mainlines. An exact analogue for an element, in the sense used by the model, is infeasible. However, so long as firms learn from observing combination of mainlines, and so long as the number of distinct technologies drawing on the same mainlines is finite (and small), then our predictions remain apt. I believe these are reasonable assumptions.

Out of approximately $13,000 \cdot 13,000 / 2 = 84.5$ million possible mainline pairs, 1.98 million pairs are actually assigned to at least one patent over the period 1836-2012. Put another way, of the 84.5 million possible pairs, I observe about 2% of the pairs as belonging to effective ideas. The mean number of patents each pair belongs to over the entire period is 10.9, but the distribution is highly skewed. Some 50.6% of observed pairs are only ever assigned to one patent, but the maximum patents assigned to a pair is 22,113.

To summarize, my dataset comprises 8.3 million patents granted between 1836 and 2012. Each of these patents is represented as a combination of mainlines, with 62.5% of patents being assigned more than one mainline. Since patents proxy for the realization of an effective idea, when two mainlines are assigned to the same patent, this proxies for one observation of compatibility between the mainlines.

5.4 – Timing Issues

Another mismatch between the model and the dataset involves timing issues. Our data includes the year a patent was granted, but has no information on the year the patent applicant decided to initiate research on the idea that was subsequently patented. I want to test our predictions using only information that would plausibly have been available at the time the researcher decided to begin research, and so will construct the independent variables from lagged data.

The choice of lag length depends on the time required to get a patent granted, the time needed to conduct research, and the speed of diffusion about the outcomes of other researcher projects. Data on patents from 1967-2000 from Hall, Jaffe, and Trajtenberg (2001) suggests the typical lag between a patent application and grant is 2 years. Suppose we assume, optimistically, that research takes 1 year to complete and the results of research are instantly made available to all rival firms. In this case, a patent granted in year t was applied for in year $t-2$, and the decision to initiate research began in year $t-3$. Such a researcher would be able to base his decision on any information available in year $t-3$. If the results of rival firm projects are instantly diffused, then this includes the outcomes of all research projects initiated before year $t-4$. Projects initiated in year $t-4$ are finished in year $t-3$ and patents are granted in year $t-1$. So patents granted in year t can draw on information derived from patents granted in year $t-1$. Hence, we might construct our independent variables from data lagged by one year.

Conversely, suppose, pessimistically that research takes 6 years to complete (the time needed to finish a long Ph.D.) and research is only made public by patent grants themselves. In this case, a patent granted in year t is applied for in year $t-2$ and research is initiated in year $t-8$. If the outcome of rival firm research is only revealed by patent grants, then patents granted in year t can draw on information derived from patents granted in year $t-8$. Hence, we might construct our independent variables from data lagged by eight years.

These arguments provide some plausible bounds for lagged variables, and I experiment with a few variations. In practice, I find lags of 3-5 years provide the best fit.

6 – Probability of Use

In Section 4, Predictions 3-5 pertain to the probability a given pair is used, where use means a research project is conducted over an idea that includes the pair. To test these predictions, I draw a sample of data from the patent dataset and use a logit model to test the probability a pair of mainlines is used in any given year. Section 6.1 will present the function form and variables used to assess these predictions, section 6.2 will discuss the results, and section 6.3 will explore the robustness of these results.

6.1 – Functional Form and Explanatory Variables

To test prediction 3-5, I draw a sample of 10,000 pairs of mainlines from the 1.84 million that are assigned to at least one patent (section 6.3 will broaden the sample to pairs never used). Using this sample, I estimate the following logit model:

$$\Pr\left(u_{p,t}=1\right)=\operatorname{logit}\left(\sum_i \phi_i\left(\text{Compatibility Dummy}\right)_{i,p,t-l}+\beta_1\cdot\text{Age}_{p,t-l}+\beta_2\cdot\text{Complements}_{p,t-l}+X'\beta\right) \tag{13}$$

where each observation corresponds to a pair-year and the dependent variable $u_{p,t}$ is a dummy variable equal to 1 if the pair of mainlines p is assigned to at least one patent in year t. Notice that all the explanatory variables are lagged by l years. I discuss the explanatory variables below, and their relationship to Prediction 3-5.

Prediction 3 says researchers are more likely to use pairs that have a high expected affinity. Unfortunately, I do not observe the beliefs of researchers, and so I do not know the expected affinity of a given pair. However, expected affinity is increasing in the number of prior observations of compatibility and this can be proxied by the number of patent grants that have been assigned the pair. This is captured by our first set of explanatory variables, which I have called Compatibility Dummy.

The relationship between expected affinity and the number of prior uses of a pair is not necessarily linear though. Indeed, Bayesian updating predicts expected affinity should asymptote at a value of 1, as the number of prior instances of success grow very large. To allow for this, I break prior instances of compatibility into five bins: [1], [2], [3,4], [5,10], and $[11,\infty)$. With the exception of the [1] bin, approximately 10% of pair-year observations fall into each of these bins. I then define a set of dummy variables, $\left(\text{Compatibility Dummy}\right)_{i,p,t}$, equal to 1 if the number of prior uses of pair p before or up to year t falls into bin i. I predict the coefficients on these bins is increasing and asymptotes for high levels.

Prediction 4 says researchers are less likely to use pairs over time, as they use up all the profitable ideas. To assess this prediction, I define the explanatory variable $\text{Age}_{p,t}$ which counts the number of years that have elapsed between year t and the year in which pair p was first used. I predict the coefficient on $\text{Age}_{p,t}$ will be negative.

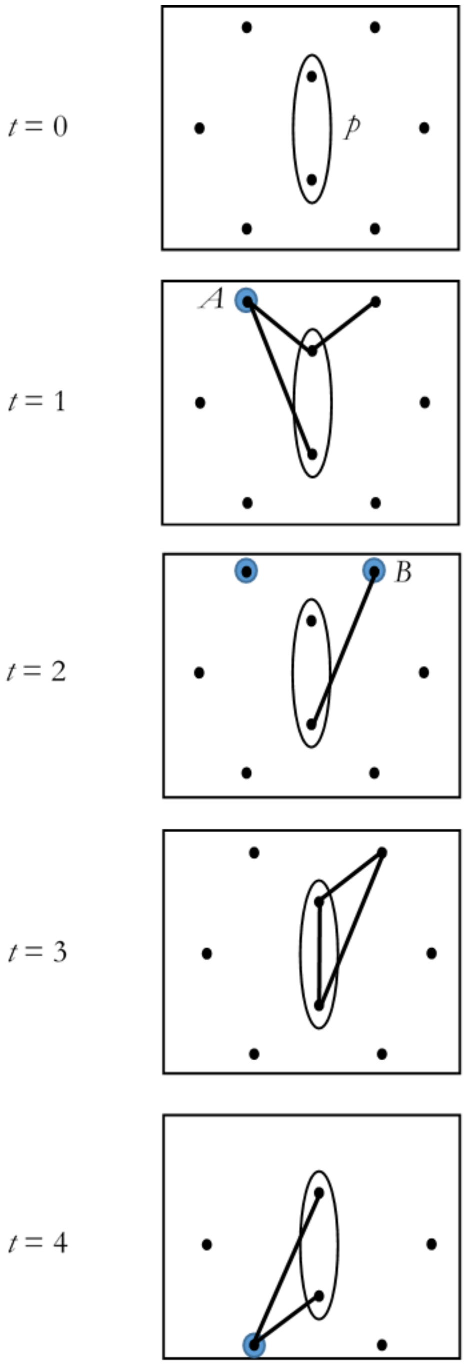
Prediction 5 says researchers are more likely to use pairs that are complementary with many other eligible ideas. To assess this prediction, I construct the variable $\text{Complements}_{p,t}$ which proxies for the number of pairs with a comparatively high expected affinity that might belong to eligible ideas. The construction of this variable is more complex than the other two variables.

If I knew the true $E[a(p)]$ of every pair, I could compute $\prod_{p \in d \setminus p'} E[a(p)]$ for every eligible idea, and find some way of aggregating these into a summary statistic. I cannot actually perform such a construction, for practical and computational reasons. To begin, the number of potentially eligible ideas is literally astronomical (with 13,000 mainlines in use, the number of potential combinations of mainlines is $2^{13,000}$). To get the number of potential ideas down to manageable levels, I restrict my attention to combinations of just 3 mainlines. If each pair can be matched with one of 13,000 other mainlines to form a combination of 3 mainlines, then there are 13,000 such ideas to assess for each pair.

For each pair p, there are approximately 13,000 potential 3-mainline ideas, each of which is defined by one mainline in addition to the pair p. For clarity of exposition, define some idea d which is composed of the mainlines p_1 and p_2, which jointly form the pair p', and an additional mainline m. For the idea composed of p_1, p_2, and m, $\prod_{p \in d \setminus p'} E[a(p)]$ corresponds to the product of the expected affinities of the pairs corresponding to (p_1, m) and (p_2, m). I do not observe $E[a(p)]$, but I do observe any patents that have been assigned the mainline pairs (p_1, m) or (p_2, m). To begin the construction of $\text{Complements}_{p',t}$, I count the number of mainlines m for which there has been at least one previous patent assigned the pair (p_1, m) *and* a patent assigned the pair (p_2, m) in any year up through year t. As time goes by, this gives an increasing count of the number of mainlines that the pair p' has been connected with from both elements of the pair.

However, simply counting the number of mainlines that have been previously connected with each mainline in a pair will tend to overstate the number of complementary pairs, since eligible ideas are used up over time. My dataset does not allow us to observe ideas that are attempted but which prove ineffective (and are never patented). Instead, I assume that whenever a pair is used, all other eligible ideas that are profitable to attempt are also attempted. I therefore reset the count of complementary pairs to zero each time the pair is used. In section 6.3, I will drop this assumption as a robustness check, and construct a second explanatory variable called $\text{Cumulative Complements}_{p',t}$. To summarize, the explanatory variable $\text{Complements}_{p',t}$ counts the number of *new* mainlines m that have been used in conjunction with each element in the pair p' since the last time the pair was used. Figure 2 presents the evolution of $\text{Complements}_{p',t}$ for one sample. I predict that the coefficient on $\text{Complements}_{p',t}$ will be positive.

Figure 2: Construction of Complements Variable



In period 0, the circled pair p has $\text{Complements}_{p,0} = 0$.

Three pairs are used. Mainline A has now been used in conjunction with both ends of pair p, so $\text{Complements}_{p,1} = 1$.

One pair is used. Between periods 1 and 2, mainline B has now been used in conjunction with both ends of pair p, so $\text{Complements}_{p,2} = 2$.

An idea drawing on mainline B and pair p is patented. This idea is no longer eligible for further patenting, and we assume the mainline A was also used. We reset $\text{Complements}_{p,3} = 0$.

Two pairs are used, linking both ends of pair p with mainline C, so $\text{Complements}_{p,4} = 1$, and $\text{Cumulative Complements}_{p,4} = 3$. Note mainlines A and B are no longer counted.

Finally X denotes some control variables that I will use in my robustness checks, discussed in section 6.3.

Table 1 presents some summary statistics for the data sample used.

Table 1: Summary of Pair Sample Data

	Min	Median	Mean	Max	St. Dev
$u_{p,t}$	0	0	0.085	1	0.279
Compatibility Dummy					
[1]	0	1	0.556	1	0.497
[2]	0	0	0.143	1	0.350
[3,4]	0	0	0.115	1	0.319
[5,10]	0	0	0.092	1	0.289
[11,∞)	0	0	0.094	1	0.292
Age	0	33	39.67	176	31.22
Complements	0	17	32.42	762	44.02
Cumulative Complements	0	74	111.5	2,113	124.5
$\max\{n_t(p_1), n_t(p_2)\}$	0	43	124.2	6,654	269.5
$\min\{n_t(p_1), n_t(p_2)\}$	0	5	15.89	3,403	42.88
$n_t(p_1) \times n_t(p_2)$	0	184	5,965	16,010,995	77,078

6.2 – Results

Table 2 presents the regression results for equation (13).

Column (1) is our baseline result, and the simplest test of Prediction 3-5. Explanatory variables are lagged by 3 years (results are not sensitive to the choice of lag). All the coefficients are statistically significant, and in the directions expected. The probability a pair will be used in any given year is increasing in the number of prior uses, which is consistent with firms learning the expected affinity of the pair is high. The probability of use declines with time, consistent with firms exhausting the best ideas that use any given pair. And pairs are more likely to be used if they have many complements, consistent with this combinatorial model of innovation.

Furthermore, the coefficients on our compatibility dummies strongly support a non-linear relationship between prior uses and probability of use, consistent with a Bayesian learning framework. This is most easily seen in Figure 3, which plots the estimated coefficient for the Compatibility Dummies, against the median number of prior uses for observations that lie within dummy's bin.

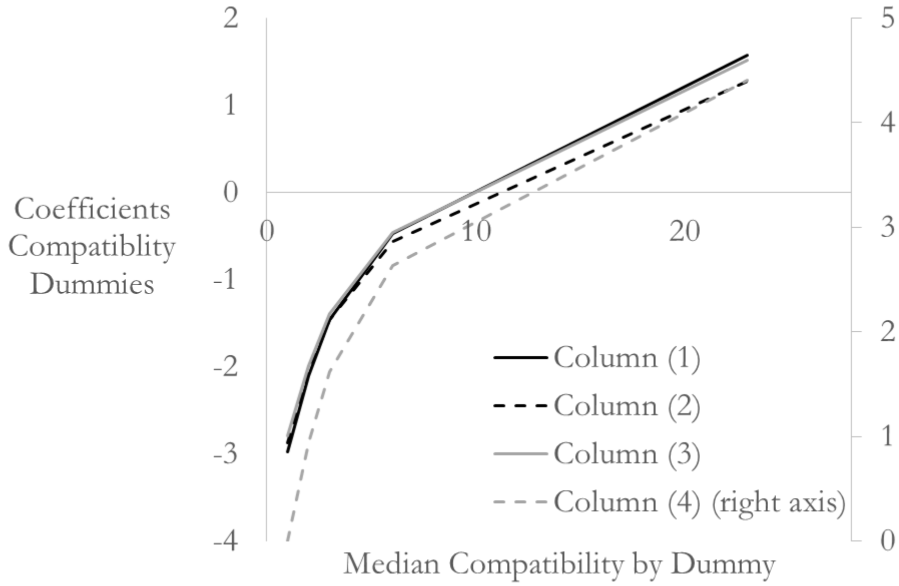
This non-linear shape implies researchers learn disproportionately more from their first observations then from later ones.

Table 2: Probability of Use Logistic Regression Results

Dependent Variable	$u_{p,t}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Compatibility Dummy						
[1]	-2.982*** (0.012)	-2.869*** (0.011)	-2.796*** (0.012)	-	-	-
[2]	-2.109*** (0.017)	-2.055*** (0.017)	-1.993*** (0.017)	0.939*** (0.028)	-	-
[3,4]	-1.461*** (0.016)	-1.462*** (0.016)	-1.389*** (0.017)	1.614*** (0.030)	-	-
[5,10]	-0.471*** (0.016)	-0.561*** (0.016)	-0.460*** (0.017)	2.629*** (0.033)	-	-
[11, ∞)	1.579*** (0.018)	1.278*** (0.019)	1.513*** (0.020)	4.410*** (0.043)	-	-
Age	-0.031*** (0.0002)	-0.030*** (0.0002)	-0.029*** (0.0002)	-0.046*** (0.0004)	-	-
Years Feasible	-	-	-	-	-0.0004 (0.002)	-0.009*** (0.002)
Complements	0.007*** (0.0001)	-	0.004*** (0.0001)	0.021*** (0.0002)	0.015*** (0.001)	0.017*** (0.001)
Complements (cumulative)	-	0.001*** (0.00004)	-	-	-	-
Controls	N	N	Y	Y	N	Y
Fixed Effects	N	N	N	Y	N	N
Sample – Pair Use	1 ≥	1 ≥	1 ≥	1 ≥	0 ≥	0 ≥
Observations	564,531	564,531	462,093	462,093	773,233	328,428
Log Likelihood	-126,122	-125,439	-118,338	-88,826	-1,580	-1442

**** signifies significance at p = 1%. Standard errors in parentheses.*

Figure 3: Coefficients of Compatibility Dummy



6.3 – Robustness

Columns 2-6 report various robustness checks. In all cases, the results support Prediction 3-5, where applicable. For clarity, I will continue to let p_1 and p_2 denote the mainlines that make up the pair p'.

Column 2 tests a simplified version of the explanatory variable $\text{Complements}_{p',t}$. In the baseline, I reset the count of mainlines that have been used with p_1 and p_2 every time the pair p' is used. This resetting is meant to control for the gradual using up of eligible ideas of which p' form a part. This is a strong assumption, and to check the robustness of the results to its weakening, I construct the explanatory variable $\text{Cumulative Complements}_{p',t}$ which does not perform this resetting. Instead, this variable is a count of all mainlines that have been used with p_1 and p_2 up through period t. Its cumulative nature means it can never decrease. Using this explanatory variable instead of the baseline's has little overall effect on our results. Note the coefficient on $\text{Cumulative Complements}_{p',t}$ is below the coefficient on $\text{Complements}_{p',t}$, though, which suggests the models fit decreases if I do not attempt to control for the gradual exhausting of eligible ideas.

Columns 3 and 4 attempt to address omitted variable bias by inclusion of additional controls. Let $n_t(p_i)$ denote the number of patents granted in year t that include the mainline p_i. In columns 3 and 4, I add to equation (13) the variables $\max\{n_t(p_1), n_t(p_2)\}$, $\min\{n_t(p_1), n_t(p_2)\}$ and $n_t(p_1) \times n_t(p_2)$. These are meant to control for any time-varying propensities to develop patents that use mainlines p_1 or p_2 outside of this model, and to therefore use the two mainlines together simply by chance. Note that when $n_t(p_i)=0$, then there are no patents that are assigned p_i and so $n_t(p_1)\times n_t(p_2)$ and $u_{p,t}$ must each be equal to zero. Since I am only interested in predicting probabilities when use is in principle possible, I omit these observations, which reduces the number of observations from 564,531 to 462,093. Introducing these controls reduces the size of the estimated coefficients, but results remain significant and in the directions predicted.

Column 4 adds to this specification fixed effects, which should control for time-invariant propensities to use a given pair. In a non-linear setting, computing fixed effects by de-meaning the data is not appropriate. Moreover, simply estimating a coefficient for dummy variables associated with each pair leads to the incidental parameters problem, which can bias coefficients.[17] Fortunately, the Chamberlain estimator is an alternative to demeaning data that can be used in a logistic regression framework. For every pair p, the Chamberlain estimator conditions estimation on the sum of $u_{p,t}$ over the pair's lifecycle, and it can be shown that this approach strips out fixed effects, just as de-meaning the data does in a linear model.[18] Adding these fixed effects strengthens our results.

Columns 5 and 6 consider sample selection issues. The preceding columns rely on a dataset that is drawn from pairs that are used at least once. This should not bias our results, because I only include observations that begin *after* the initial use of the pair (beginning after the first use allows us to construct all our explanatory variables). However, this sampling methodology implies our results pertain only to the probability of use, after a first initial use. In columns 5 and 6, I see if our model is also applicable to the first time a pair is used.

To assess this, I draw 8,003 pairs of mainlines from the set of 84.5 million possible pairs. Most pairs are never used. From the 8,003 pairs selected, only 183 pairs are ever used by 2012 (2.2%). Since I have an observation for every year and pair, there are 773,233 total observations. I am only trying to predict the first time a pair is used now, and so the dummy variables associated with prior uses are all equal to zero, and $\mathrm{Age}_{p,t}$ is undefined. Instead, I find the first year that both mainlines have been assigned to a patent, and set this year as the pair's first "available" year. I interpret this year as an indicator of the earliest time the pair could possibly have been used by innovators. For example, if the mainline 14/3 ("Bridge; Truss") was first assigned to a patent in 1836 and the mainline 706/15 ("Data processing (artificial intelligence); neural network") was first assigned to a patent in 1980, then the pair (14/3, 706/15) is first available in 1980, rather than 1836. I then measure the years the patent has been available. It is worth including this metric, since not using a pair shortly after it is available is a signal that the ideas using the pair do not have $\pi(d)$ high enough or $k(d)$ low enough to warrant a research project even when expected efficacy of the ideas is low.

Without information on prior uses, only Prediction 5 is relevant. For each pair-year, I compute the variable $\mathrm{Complements}_{p',t}$, as before. In column 5, this variable is a strong predictor of first use,

[17] See Greene (2008), pg 800-801.
[18] See Greene (2008), pgs. 803-805.

which is again consistent with theory. In column 6, I include as additional controls the same ones constructed for column 3. In this instance the results are stronger when these controls are included.

These robustness checks provide further support for the veracity of predictions 3-5. These predictions pertained to the use of specific pieces of knowledge, in this case pairs of mainlines. In the next section, I show that this model's predictions also apply to entire technology sectors.

7 - Patents Issued Per Year

Section 4 argued that more patents will be granted when the expected affinity of pairs in a sector are high, but that the number of patents granted will decrease over time. There are a number of data challenges that must be surmounted in order to test these predictions. I discuss these data issues, as well as some functional form assumptions, in Section 7.1. In Section 7.2, I present my baseline results. Section 7.3 provides some robustness checks.

7.1 – Data and Functional Form Issues

7.1.1 – Units of Observation

Predictions 1 and 2 pertain to the number of patents granted within a given technology sector, which was understood to be a group of firms drawing upon the same set of technological building blocks. In this section, I use 429 technology classes in the USPCS as proxies for different technological sectors. This gives us a panel of 429 classes, each with up to 176 years of innovative activity.

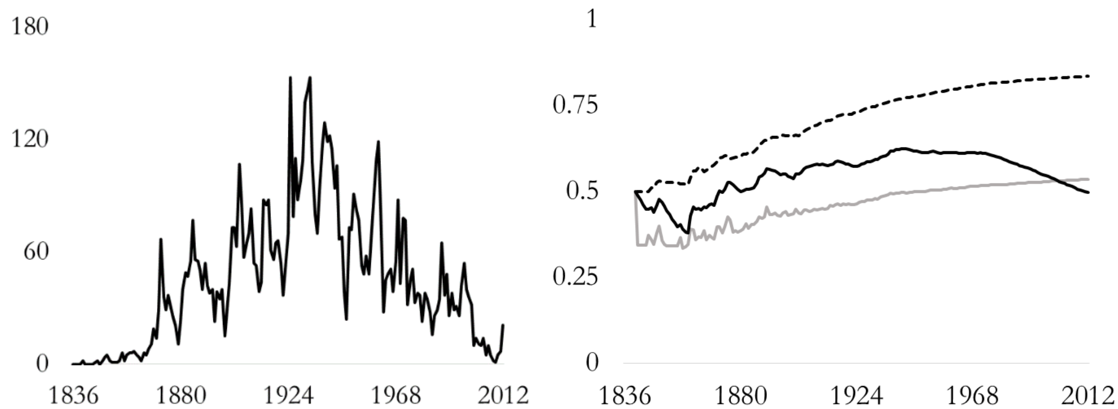
To determine the number of patents granted to a class in a year, I use the primary classification assigned to each patent. Each patent is assigned one, and only one, primary classification, which is based on the main inventive concept. The primary classification is generally used in economics to assign patents to different technology classes (see for example, Hall, Jaffe, and Trajtenberg 2001). The dependent variable $n_{cl,t}$ corresponds to the number of patents with primary class cl granted in year t . The path of three $n_{cl,t}$ variables over time is plotted in the left column of Figure 4.

Testing predictions 1 and 2 also requires information on the pairs used by each technology class. To assemble a list of the pairs used by each class, I tally the mainlines assigned to each patent in a class. For example, recall that patent 7,640,683, for "Method and apparatus for satellite positioning of earth-moving equipment" was assigned the mainlines:

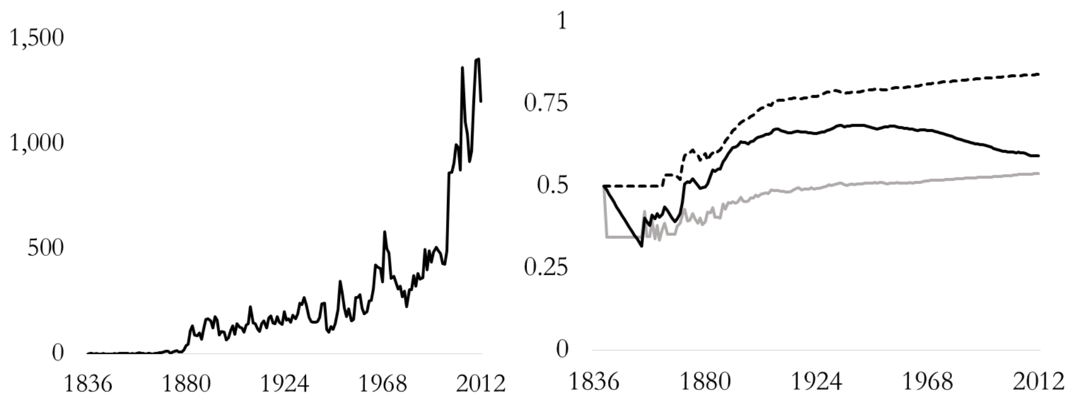
5. 37/347: Excavating; ditcher
6. 414/680: Material or article handling; vertically swinging load support
7. 701/1: Data processing: vehicles, navigation and relative location
8. 37/381: Excavating; road grader-type

Figure 4: Patent Class Trends

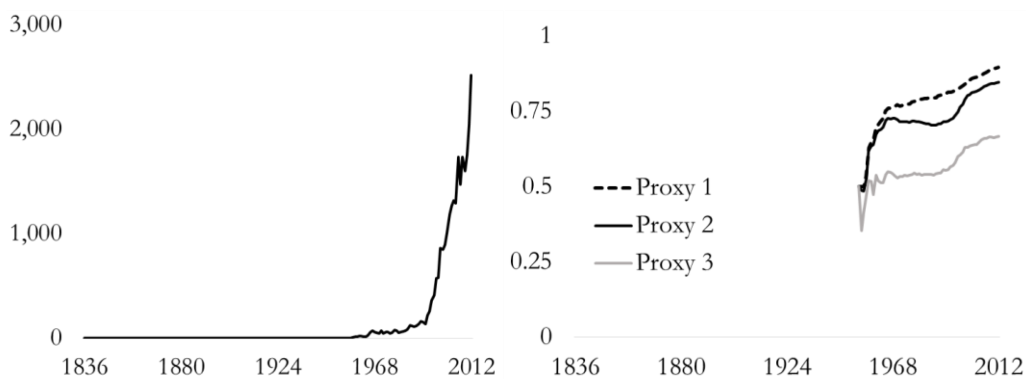
Left Column: Patents Granted per Year **Right Column:** Average Affinity Proxies



Class 178: Telegraphy



Class 310: Electrical Generator or Motor Structure



Class 711: Electrical Computers and Digital Processing Systems: Memory

The primary class for this patent is class 37, excavating, and it was granted in year 2010. I therefore assign mainline-pairs between 37/347, 414/680, 701/1, and 37/381 as belonging to class 37 from

2010 onwards. After doing this for every patent in the class, I obtain the list of all mainline-pairs used by any patent in the class, in each year.

The number of mainlines used by a class is different from the number of mainlines that are *nested* under a class by the USPCS, because most patents assigned to a class draw on mainlines from outside the class (just as the above patent is assigned to class 37 but is also assigned mainlines from classes 414 and 701). The mean number of mainlines nested under one class is 29.6, but the mean number of mainlines used by patents in a class is 1,161. Moreover, the minimum number of mainlines nested under one class was 1, while the minimum used by patents in a class is 5. The maximum number of mainlines nested under a class was 246, while the maximum used by a class is 5,126.

7.1.2 – Proxying for Expected Affinity

Predictions 1 and 3 both relate to the expected affinity of a pair. In section 6, I did not attempt to construct a proxy for expected affinity, but instead took a relatively non-parametric approach to the question. The results were consistent with a Bayesian learning framework, where more observations of success are correlated with a higher probability of use, and the first few such observations are more important than later ones. In this section, I will be a bit more ambitious, and explicitly construct three different proxies for the expected affinity of a pair. Given that we do not observe the beliefs of researchers, nor research projects that do not yield patents, these proxies are bound to be noisy. Nevertheless, as we will see, they do carry useful information.

Each of these proxies is 0 when there are no observations of prior success, and therefore assumes the initial prior of research firms assigns a very low probability any given pair of elements will be compatible. Each of these proxies then draws on past observations of $s_t(p)$, where $s_t(p)$ is the number of patents pair p is assigned to in year t.
The first proxy for expected affinity is:

$$\tilde{E}_t^1\left[a(p)\right] \equiv \frac{\sum_{\tau=0}^{t} s_\tau(p)}{1+\sum_{\tau=0}^{t} s_\tau(p)} \tag{14}$$

This is the simplest proxy, and it exploits only information on the cumulative number of past observations of a pair being assigned to a patent. It approaches 1 as the total number of prior uses of a pair goes to infinity.

The second proxy for expected affinity is:

$$\tilde{E}_t^2\left[a(p)\right] \equiv \frac{\sum_{\tau=0}^{t} 0.95^{t-\tau} s_\tau(p)}{1+\sum_{\tau=0}^{t} 0.95^{t-\tau} s_\tau(p)} \tag{15}$$

This formulation is equivalent to the first proxy, but where more recent observations of a pair are accorded more weight. It implicitly assumes mainlines are an imperfect proxy for the true building blocks that make up an idea, and assumes more recent observations may be more relevant for estimating expected affinity. In practice, this proxy builds in a tendency for $E\left[a(p)\right]$ to decay over time if a pair is not continuously assigned to new patents.

The third proxy for expected affinity is:

$$\tilde{E}_t^3\left[a(p)\right] \equiv \frac{\sum_{\tau=0}^{t} s_\tau(p)}{1+\sum_{\tau=0}^{t} n_\tau(p)} \tag{16}$$

where

$$n_t(p) = \max\left\{s_t(p),(1+g_t)s_{t-1}(p)\right\} \tag{17}$$

and

$$1+g_t = \sum_p s_t(p) / \sum_p s_{t-1}(p) \tag{18}$$

This formulation attempts to mechanically infer the number of times researchers tried to use a pair with a simple rule. In each period, researchers expect the number of attempts to use pair p is equal to the number of times the pair was assigned to patents in the last period, multiplied by an aggregate growth term $1+g_t$, or the number of patents it was actually assigned to this period, whichever is larger. The growth term $1+g_t$ is the overall growth rate of patent pair assignments. In practice, this proxy penalizes pairs that fail to "keep up" with the aggregate growth rate of all pair assignments, by assuming their failure to keep up reflects a failure of research ideas to be effective, rather than an absence of research attempts.

As a simple test of these three proxies, I use them in place of the Compatibility Dummies on the same sample of pairs drawn in Section 6 to estimate the following:

$$\Pr\left(u_{p,t}=1\right) = \text{logit}\left(\beta_0 \cdot \tilde{E}_{t-l}^i\left[a(p)\right] + \beta_1 \cdot \text{Age}_{p,t-l} + \beta_2 \cdot \text{Complements}_{p,t-l} + \beta_3\right) \tag{19}$$

The results to regression (19) are presented in Table 3.

Note that the log-likelihood using the first proxy is no better than the non-parametric baseline, which is plausible, since both embody a concave function of cumulative past observations. The log-likelihood improves significantly when I begin to weight more recent observations more than distant ones, as in the second proxy, and improves again if I exploit the growth rate of pair uses, relative to all pair uses, as in the third proxy.

Table 3: Probability of Use Logistic Regression Results, With Expected Affinity Proxies

Dependent Variable	$u_{p,t}$		
	Proxy 1	Proxy 2	Proxy 3
$\tilde{E}_{t-l}^{i}\left[a(p)\right]$	9.443***	10.240***	24.982***
	(0.041)	(0.041)	(0.094)
$\text{Age}_{p,t-l}$	-0.027***	-0.020***	-0.030***
	(0.0002)	(0.0002)	(0.0002)
$\text{Complements}_{p,t-l}$	0.006***	0.019***	0.009***
	(0.0001)	(0.0002)	(0.0002)
Observations	564,531	564,531	564,531
Log-Likelihood	-127,581	-95,550	-79,152

**** signifies significance at p = 1%. Standard errors in parentheses.*

Given my proxies for $E\left[a(p)\right]$, and the assignment of mainline-pairs to classes, I can now compute a measure for the average expected affinity for each class, in each year:

$$\left(\text{Average Affinity Proxy}\right)_{cl,t-l} = \frac{1}{N_{cl,t}} \sum_{p \in cl,t} \tilde{E}_{t}^{i}\left[a(p)\right] \tag{20}$$

I plot the average affinity proxy in the right column of Figure 4, for three example classes. As can be seen, my measure for average affinity proxy is only defined from the point at which the first patent in the class is granted. Moreover, it exhibits variation over time. Finally, although the three proxies do not move in sync with each other, they are correlated.

7.1.3 – Functional Form

To test predictions 1 and 2, a simple but naïve approach would be to estimate the following regression:

$$n_{cl,t} = \tilde{\beta}_0 + \tilde{\beta}_1 \cdot \left(\text{Average Affinity Proxy}\right)_{cl,t-l} + \tilde{\beta}_2 \cdot \text{Age}_{p,t-l} + \varepsilon_{cl,t} \tag{21}$$

Where $\text{Age}_{cl,t}$ denotes the number of years that have elapsed since a patent was first given class cl as its primary assignment. The unit of observation in equation (21) is a class-year, and I would expect the coefficient on the Average Affinity Proxy to be positive and the coefficient on Age to be negative.

There are a number of problems with this naïve specification. First, classes may vary widely in the number of firms, the inherent profitability of research in the sector, or the number of building blocks available. Indeed, as Figure 4 makes clear, the number of patents granted across different classes is highly variable. These differences suggest a log model is more appropriate to estimate. To estimate a log model, I drop the 6.3% of observations where $n_{cl,t} = 0$, so that the results should be viewed as being conditional on positive research activity taking place. In my robustness checks, I add 1 to the observations so that these dropped observations can be included.

Second, it is clear from Figure 4 that the number of patents granted per year and Average Affinity Proxy are trending variables, so that a simple regression of one onto the other would mostly pick up trends. Since these trends may well vary by class, I estimate a fixed effect model on differenced data:

$$\Delta \ln n_{cl,t} = \alpha_{cl} + \beta_1 \cdot \Delta \ln\left(\text{Average Affinity Proxy}\right)_{cl,t-l} + \beta_2 \cdot \text{Age}_{cl,t-l} + X'_{cl,t}\beta + \varepsilon_{cl,t} \tag{22}$$

Note the term α_{cl} can account for class-specific growth rates.

Third, to account for higher order trends and omitted variables, I include as controls several lags of the dependent variable $\Delta \ln n_{cl,t}$. Moreover, to control for omitted variables that influence aggregate patenting, I include lags of changes in log-aggregate patenting $\Delta \ln N_t$ where $N_t \equiv \sum_{cl} n_{cl,t}$.

To determine the number of lags to include, I estimate the following baseline model, which does not include the variable $\Delta \ln\left(\text{Average Affinity Proxy}\right)_{cl,t-l}$:

$$\Delta \ln n_{cl,t} = \alpha_{cl} + \sum_{i=1}^{8} \varphi_i L^i \Delta \ln n_{cl,t} + \sum_{i=1}^{3} \tilde{\varphi}_i L^i \Delta \ln N_t + \beta_3 \cdot Age_{cl,t} + \varepsilon_{cl,t} \tag{23}$$

This model forecasts changes in class patenting activity with a class-specific fixed effect, lags of changes in class patents, lags of the change in total patents, and the age of the class. The estimated coefficients for this model are presented in Table 4.

Table 4: Coefficients of Baseline Model

Dependent Variable:					$\Delta \ln n_{cl,t}$				
Lags	0	1	2	3	4	5	6	7	8
$\Delta \ln n_{cl,t}$		-0.421 (0.009)	-0.246 (0.008)	-0.178 (0.008)	-0.120 (0.007)	-0.099 (0.005)	-0.066 (0.006)	-0.033 (0.006)	-0.021 (0.005)
$\Delta \ln N_t$		0.446 (0.016)	0.248 (0.016)	0.130 (0.016)					
$\text{Age}_{cl,t}$	-0.001 (0.00004)								
Class Fixed Effects?	Yes								
Observations	53,930								
Adjusted R²	0.145								

Note: All coefficients are significant at p = 0.1%. White standard errors clustered by class in parentheses.

I chose to include 8 lags of $\Delta \ln n_{cl,t}$ and 3 lags of $\Delta \ln N_t$ since the statistical significance levels of further lags is lower, and also reduces the adjusted R^2. The baseline model has a few notable features. First, there is a tendency for the growth rate of patents in a class to converge to the aggregate growth rate of patents, which can be seen from the approximately equal but opposite coefficients on $\Delta \ln n_{cl,t}$ and $\Delta \ln N_t$. Second, the growth rate of patents slows over time, as indicated by the negative coefficient on $Age_{cl,t}$. Finally, a Hausman test strongly rejects the hypothesis that class fixed effects can be ignored. Different classes have different growth rates. Unless explicitly stated, all the explanatory variables in equation (23) and Table 4 are included in all the following regressions, although I do not report their estimated coefficients to save space.

Lastly, the Average Affinity Proxy conflates two sources of variation, namely the sample of pairs and $\tilde{E}_t^i\left[a(p)\right]$ for each pair. In period t, the set of pairs used by a class cl is defined as the set of pairs used by all patents awarded to class cl in any period prior to or including period t. However, defining classes in this manner means the set of pairs used is growing over time, rather than constant. To make sure the change in expected affinity is due to changes in affinity, rather than the definition of the set, in each period, I compute $\Delta \ln\left(\text{Average Affinity Proxy}\right)_{cl,t-l}$ using the same set of pairs. Specifically, when comparing Average Affinity Proxy in period t and period $t-1$, I always use the set of pairs defined for period t.

7.1.4 – Data Description

After computing the above metrics, I have an unbalanced panel of 429 classes with up to 176 years of data per class (the panel is unbalanced, because I exclude years before a class has its first granted patent), for a total of 59,592 observations. Some summary statistics are presented below in Table 5.

Table 5: Summary Patent Class Statistics

	Min	Median	Mean	Max	Standard Deviation
$\text{Age}_{cl,t}$	1	81	82.8	176	47.39
Patents Granted ($n_{cl,t}$)	0	57	132.9	8,348	276.1
Average Affinity Proxy 1	0.071	0.722	0.705	0.999	0.092
Average Affinity Proxy 2	0.064	0.543	0.552	0.996	0.090
Average Affinity Proxy 3	0.053	0.455	0.454	0.805	0.060
$\Delta \ln n_{cl,t}$	-2.984	0.001	0.031	4.414	0.426
$\Delta \ln\left(\text{Average Affinity Proxy 1}\right)_{cl,t-l}$	0.000	0.012	0.034	2.881	0.090
$\Delta \ln\left(\text{Average Affinity Proxy 2}\right)_{cl,t-l}$	-0.044	0.008	0.030	2.877	0.097
$\Delta \ln\left(\text{Average Affinity Proxy 3}\right)_{cl,t-l}$	-1.216	0.010	0.030	3.105	0.110

Note that the mean values for the change in log-transformed variables are all of a similar magnitude, although $\Delta \ln n_{cl,t}$ has more variance than any of the $\Delta \ln(\text{Average Affinity Proxy})_{cl,t-l}$ metrics. Some correlations are presented in Table 6:

Table 6: Correlations Among Proxies

	Average Affinity Proxy 1	Average Affinity Proxy 2	Average Affinity Proxy 3
Average Affinity Proxy 1	1		
Average Affinity Proxy 3	0.919	1	
Average Affinity Proxy 2	0.661	0.786	1

	$\Delta \ln(\text{Average Affinity Proxy})_{cl,t-l}$ 1	$\Delta \ln(\text{Average Affinity Proxy})_{cl,t-l}$ 2	$\Delta \ln(\text{Average Affinity Proxy})_{cl,t-l}$ 3
$\Delta \ln(\text{Average Affinity Proxy 1})_{cl,t-l}$	1		
$\Delta \ln(\text{Average Affinity Proxy 2})_{cl,t-l}$	0.919	1	
$\Delta \ln(\text{Average Affinity Proxy 3})_{cl,t-l}$	0.994	0.931	1

Note that these three measures are much more correlated in their differences than in their levels. This stems from the fact that the difference between Average Affinity Proxy in one period to another is most prominently driven by the pairs which go from 0 to 1 assigned patents, which all three proxies measure as a change from 0 to 0.5.

7.2 – Baseline Results

In Section 6, my preferred specification was a lag of 3 years, but I found the choice of lag had little impact on the estimated results. In this model, the choice of lag is not inconsequential. My first investigation examines all three proxies for $\Delta \ln(\text{Average Affinity Proxy})_{cl,t-l}$ across two lag specifications. The results are displayed in Table 7.

As discussed earlier, the significance of explanatory variables with a lag greater than two is consistent with patent grants as a primary vehicle for knowledge diffusion, while the significance of variables with smaller lags indicates knowledge can spread before the patent is granted. I find evidence for both effects, with $\Delta \ln(\text{Average Affinity Proxy})_{cl,t-l}$ usually positive and statistically significant for lags of 1 year and 3-5 years, although significance is stronger for the 3-5 year period. Since research may take different lengths of time for different projects, it is not surprising that multiple years of explanatory variable are statistically significant.

The coefficients on all three proxies are similar in magnitudes, but as found in Table 7, Proxy 3 has the strongest predictive power. It is the only proxy with statistical significance for a lag of 5 years, and the adjusted R-squared is usually highest for this proxy. I also conduct an F-test for the joint hypothesis that all coefficients on lagged values of $\Delta \ln(\text{Average Affinity Proxy})_{cl,t-l}$ are statistically indistinguishable from zero, but this is rejected at p = 0.1% in every specification. Again, the F-statistic is largest for Proxy 3.

Table 7: Different Lag and Proxy Specifications

Dependent Variable:	$\Delta \ln n_{cl,t}$					
	(1)	(2)	(3)	(4)	(5)	(6)
	Proxy 1	Proxy 2	Proxy 3	Proxy 1	Proxy 2	Proxy 3
Age	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.0001)	(0.0001)	(0.0001)	(0.00005)	(0.00005)	(0.00005)
$L\Delta \ln(\text{Average Affinity Proxy})$	0.167**	0.120*	0.127**			
	(0.077)	(0.064)	(0.053)			
$L^2\Delta \ln(\text{Average Affinity Proxy})$	-0.058	-0.054	0.016			
	(0.079)	(0.066)	(0.053)			
$L^3\Delta \ln(\text{Average Affinity Proxy})$	0.191***	0.152***	0.142***	0.202***	0.156***	0.149***
	(0.065)	(0.054)	(0.043)	(0.061)	(0.052)	(0.043)
$L^4\Delta \ln(\text{Average Affinity Proxy})$	0.173***	0.152***	0.167***	0.186***	0.158***	0.179***
	(0.059)	(0.050)	(0.041)	(0.059)	(0.050)	(0.041)
$L^5\Delta \ln(\text{Average Affinity Proxy})$	0.082	0.077	0.116***	0.088	0.078	0.122***
	(0.060)	(0.051)	(0.041)	(0.056)	(0.049)	(0.040)
$L^6\Delta \ln(\text{Average Affinity Proxy})$	-0.040	-0.036	0.011			
	(0.049)	(0.042)	(0.034)			
$L^7\Delta \ln(\text{Average Affinity Proxy})$	0.040	0.029	0.015			
	(0.050)	(0.043)	(0.034)			
$L^8\Delta \ln(\text{Average Affinity Proxy})$	0.002	0.004	-0.022			
	(0.044)	(0.038)	(0.032)			
Controls						
Lags of $\Delta \ln n_{cl,t}$	8	8	8	8	8	8
Lags of $\Delta \ln N_t$	3	3	3	3	3	3
Class Fixed Effects	Y	Y	Y	Y	Y	Y
F-Test†	4.925	4.000	5.346	11.086	9.340	11.952
Observations	53,930	53,930	53,930	53,930	53,930	53,930
Adjusted R^2	0.147	0.146	0.147	0.146	0.146	0.147

Note: † *The null hypothesis is the joint insignificance of all coefficients on lags of* $\Delta \ln Average.affinity.proxy$.
White standard errors clustered by class are in parentheses. *p<0.1; **p<0.05; ***p<0.01

Table 7 supports Prediction 1 and 2. The average expected affinity between elements in a technology field is a positive predictor of the change in patents granted in the future, with an elasticity between 0.1 and 0.2, but applying over several periods. This occurs even though I control for lagged behavior, and despite the imprecision in the proxies for elements, ideas, and expected affinity. Moreover, there is a general tendency for fewer patents to be granted as time goes on.

7.3 – Robustness Checks

I next perform a set of robustness checks. A more detailed discussion is available in the appendix. A primary result is the variable strength of the above-noted relationship over time. Generally speaking, the coefficients on $\Delta\ln\left(\text{Average Affinity Proxy}\right)_{cl,t-l}$ are most likely to be positive and statistically significant for the period of 1936-2012. In the robustness checks, I will sometimes find the coefficients on $\Delta\ln\left(\text{Average Affinity Proxy}\right)_{cl,t-l}$ lose significance or change sign, but this result is invariably driven by the 1836-1936 period. The 1936-2012 period generally conforms to the patterns described in section 7.2, and is often stronger.

There are two reasons why this result is not surprising. As shown in Figure 1, the fraction of patents with more than one mainline varies substantially over the period 1836-2012. Since the model is premised on ideas consisting of combinations of at least 2 elements, the model is a better fit for the data after 1936, when 69% of patents were assigned 2 or more mainlines, as opposed to 40% before 1936. This suggests the pre-1936 data may be subject to greater measurement error, which would lead to attenuation bias.

In addition to measurement error, a second potential source of bias may stem from the USPTO's ongoing updating of its classification system. If commonly used pairs of mainlines are eventually consolidated into single mainlines, then the only pairs we will observe in earlier periods will be pairs of mainlines that were *not* subsequently combined many times. This would tend to bias the results, with the bias more severe for earlier periods. Indeed, if I break the period 1936-2012 into two more periods from 1936-1986 and 1986-2012, the results for the earlier period remain very strong, but the size of estimated coefficients for 1986-2012 grows substantially (the coefficient on $L^3\Delta\ln\left(\text{Average Affinity Proxy}\right)_{cl,t-l}$ is over 2).

I find a similar effect when I measure $\Delta\ln\left(\text{Average Affinity Proxy}\right)_{cl,t-l}$ with reference to the pairs used in the previous period, rather than the current period. That is, when comparing the Average Affinity Proxy in period t and period $t-1$, I now use the set of pairs defined in $t-1$. This has the effect of omitting all pairs that were used for the first time in period t, and restricts attention to pairs that have already been used at least once. As reported in the appendix, this leads to a *negative* and statistically significant coefficient for the 1836-1935 data, but generally familiar (and strong) positive and significant coefficients for the 1936-2012 period.

Another potential source of bias in my estimates is the construction of proxies. Since these rely on cumulative counts of patent grants, they will be correlated with the growth rate of patents over time, which is the dependent variable. While I have attempted to control for this by adding 8 lags of

$\Delta \ln n_{cl,t}$, it may be that the proxy is picking up the effect of lags beyond 8. To check for this, I double the number of lags. This leads to a strengthening of the results, compared to those found in Table 7, column (4)-(6).

In another specification, I restore observations where $n_{cl,t} = 0$ by transforming the dependent variable to $\Delta \ln\left(1 + n_{cl,t}\right)$. All variables remain positive and significant in this specification. Lastly, when I omit all classes that contain less than two mainlines, results are largely unaffected.

8 – Discussion

The preceding empirical analysis shows that accounting for the combinations of mainlines used in a class of patents does help predict the growth rate of patents by class, as compared to a model composed of lags, fixed effects, and class age. Patenting increases are correlated with new combinations made before the patent was granted. Where the data better fits the model, results are stronger, and modifications to the proxy and data used do not dramatically impact the results. The implied elasticity is on the order of 0.1-0.2 for several years, although this may attenuated by measurement error. If I rely only on data from 1936 onwards, the elasticity rises.

According to the model presented in this paper, a rise in expected affinity for some pairs in a class increases the expected number of patent grants, because ideas using these pairs are more likely to be effective. The rise in class patent output should come disproportionately from patents that include pairs whose increase led to an increase in Average Affinity Proxy. While I have not tested for this effect directly, this story is consistent with the results from section 6, which showed the probability a pair of mainlines is assigned to a patent is also increasing in the expected affinity of the pair, and of the pairs in related ideas.

Across all specifications, I also find the age of a class has negative and significant impact on the growth rate of patents granted. This is also consistent with the presence of fishing out effects, implying a long-run decline in the growth rate of any one class. As a back of the envelope calculation, the growth rate of 2.7% for all patents over the last 100 years falls to zero in 27 years all else equal, when the coefficient on age is -0.001 (a common result). I find a similar time-scale at work when estimating the probability a pair will be used in any given year. In both cases, I find a story consistent with the model, where technological progress in any given class must constantly reinvigorate itself by discovering new pairs have a high affinity, or else research productivity falls to zero.

9 – Conclusions

This paper has presented a new model of knowledge production, and shown some of its predictions are consistent with tests using U.S. patent data. This approach helps clarify some questions about the production of knowledge.

Have we run out of things to invent, or are we on the cusp of a new era of technological advance? This paper argues these two forces are in a constant state of ebb and flow. In the short and medium run, firms fish out the best ideas that can be constructed from a given set of technological building blocks. At the same time, this process helps firms learn which types of combinations are feasible, which serves to expand the set of ideas that are worth their research costs. The future for technological innovation is brightest after a set of technological building blocks has been discovered (possibly via recombinant growth, as in Weitzman 1998, or via other processes) *and* some initial work has established many elements in this set have a high affinity, but *before* the best ideas have been fished out. If a successful sector keeps recycling the same combinations however, it will learn less and less useful information from each invention, and eventually exhaust the finite set of possible discoveries.

The empirical methodology used in this paper could be used as a method of establishing a measure for the outlook for many technological sectors, or academic disciplines. This could then be used as a way to control for technological opportunity in many other applications, or even as a method of forecasting. Such an approach might give some guidance on the question of whether US and global innovation is stagnating or surging.

The approach used in this paper could also serve as one input into funding decisions for agencies conducting R&D across many disciplines and sectors. This paper's model also suggests public funders of science can have the most positive spillovers by doing the exploratory work of bringing together untried combinations of elements.

References

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

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.

Bessen, James, and Michael J. Meurer. 2008. *Patent Failure: How Judges, Bureaucrats and Lawyers Put Innovators at Risk.* Princeton, NJ: Princeton University Press.

Borges, Jorge Luis. 1962. *Ficciones (English Translation by Anthony Kerrigan).* New York, NY: Grove Press.

Brynjolfsson, Erik and Andrew McAffee. 2014. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies.* New York, NY: W.W. Norton & Company, Inc.

Clancy, Matthew. 2015. "Optimal Research Strategies When Innovation is Combinatorial." *Dissertation.*

Clancy, Matthew and GianCarlo Moschini. 2013. "Incentives for Innovation: Patents, Prizes, and Research Contracts." *Applied Economic Perspectives and Policy 35 (2)* 206-241.

Cowan, Tyler. 2011. *The Great Stagnation: How America Ate All The Low-Hanging Fruit of Modern History, Got Sick, and Will (Eventually) Feel Better.* New York, NY: Penguin Group.

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.

Gordon, Robert J. 2012. "Is U.S. Economic Growth Over? Faltering Innovation Confronts the Six Headwinds." *NBER Working Paper 18315.*

Greene, William H. 2008. *Econometric Analysis: Sixth Edition.* Upper Saddle River, New Jersey: Pearson Prentice Hall.

Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg. 2001. "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools." *NBER Working Paper 8498.*

Harhoff, Dietmar, Francis Narin, F.M. Scherer and Katrin Vopel. "Citation Frequency and the Value of Patented Inventions." *The Review of Economics and Statistics*, 1999: 511-515.

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.

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.

Trajtenberg, Manuel. "A Penny for your Quotes: Patent Citations and the Value of Innovations." In *Patents, Citations and Innovation*, by Adam B. Jaffe and Manuel Trajtenberg, 25-50. Cambridge, MA: MIT Press, 2002.

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.

Appendix – Robustness Checks for Section 7

This appendix contains the results from a series of robustness checks on the regressions discussed in section 7.3.

A1 – Different Time Frames

As shown in Figure 1 in the main text, the fraction of patents with more than one mainline varies substantially over the period 1836-2012. Since the model is premised on ideas consisting of combinations of at least 2 elements, the model is a better fit for the data after 1936, when 69% of patents were assigned 2 or more mainlines, as opposed to 40% before 1936. This suggests the pre-1936 data may be subject to greater measurement error, which would lead to attenuation bias. Accordingly, I re-estimate my model for two time periods, 1836-1935 and 1936-2012. The results are presented in Table A1.

When I restrict attention to the period 1836-1935, the results are no longer statistically significant. However, for the second period, encompassing much of the 20th century, the results are significantly strengthened relative to the complete sample (compare to Table 7, column 4-6). This supports the argument that attenuation bias due to measurement error has likely reduced the size of the estimated coefficients.

As discussed in section 7.3, another potential source of bias in the estimates is the construction of the proxies. To check for the sensitivity of the results to the number of lags of the dependent variable, I double the number of lags in the last three columns of Table A1. This leads to a strengthening of the results, compared to those found in Table 7, column (4)-(6).

A2 – Measuring Changes from the Prior Set

In Table A2, I measure the change in expected affinity for a class with reference to the pairs used in the previous period, rather than the current period. That is, when comparing Average Affinity Proxy in period t and period $t-1$, I now use the set of pairs defined in $t-1$. This has the effect of omitting all pairs that were used for the first time in period t, and restricts attention to pairs that have already been used at least once. As can be seen in the first three columns of Table A2, this has a significant impact on the coefficients attached to $\Delta \ln\left(\text{Average Affinity Proxy}\right)_{cl,t-l}$. For all three proxies, some of the lagged changes now have a negative and statistically significant coefficient, in violation of Prediction 1.

Table A1: Robustness over time periods

	$\Delta \ln n_{cl,t}$								
Years	1844-1935			1944-2012			1852-2012		
Proxy	1	2	3	1	2	3	1	2	3
Age	-0.002***	-0.002***	-0.002***	-0.0005***	-0.0005***	-0.0005***	-0.001***	-0.001***	-0.001***
	(0.0002)	(0.0002)	(0.0002)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.00005)	(0.00005)
$L^3 \Delta \ln(\text{Average Affinity Proxy})$	0.034	0.020	0.023	1.132***	0.899***	0.783***	0.282***	0.208***	0.214***
	(0.067)	(0.057)	(0.047)	(0.354)	(0.273)	(0.214)	(0.076)	(0.062)	(0.050)
$L^4 \Delta \ln(\text{Average Affinity Proxy})$	0.061	0.055	0.055	0.168	0.169	0.434**	0.129	0.111	0.165***
	(0.061)	(0.051)	(0.044)	(0.380)	(0.309)	(0.211)	(0.089)	(0.074)	(0.058)
$L^5 \Delta \ln(\text{Average Affinity Proxy})$	-0.021	-0.013	0.030	0.145	0.114	0.361**	0.245***	0.191***	0.206***
	(0.060)	(0.052)	(0.043)	(0.249)	(0.208)	(0.163)	(0.079)	(0.065)	(0.050)
Controls									
Lags of $\Delta \ln n_{cl,t}$	8	8	8	8	8	8	16	16	16
Lags of $\Delta \ln N_t$	3	3	3	3	3	3	6	6	6
Class Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	23,766	23,766	23,766	27,173	27,173	27,173	50,155	50,155	50,155
Adjusted R^2	0.181	0.181	0.181	0.126	0.126	0.126	0.140	0.140	0.140

Note: *p<0.1; **p<0.05; ***p<0.01

White standard errors clustered by class in parentheses.

Table A2: Set of Pairs Defined by Earlier Period

	$\Delta \ln n_{cl,t}$								
Years	1844-2012			1844-1935			1944-2012		
Proxy	1	2	3	1	2	3	1	2	3
Age	-0.001***	-0.001***	-0.001***	-0.003***	-0.002***	-0.002***	-0.0005***	-0.001***	-0.001***
	(0.0001)	(0.00004)	(0.00005)	(0.0002)	(0.0002)	(0.0002)	(0.0001)	(0.0001)	(0.0001)
$L^3 \Delta \ln($ Average Affinity Proxy $)$	-0.264	-0.841***	-0.463***	-1.647***	-1.184***	-0.250*	7.178***	2.947***	0.532
	(0.466)	(0.292)	(0.121)	(0.514)	(0.322)	(0.128)	(1.839)	(0.996)	(0.565)
$L^4 \Delta \ln($ Average Affinity Proxy $)$	0.655	-0.200	-0.274**	-0.361	-0.370	-0.053	0.179	0.042	0.221
	(0.477)	(0.320)	(0.124)	(0.522)	(0.336)	(0.133)	(2.492)	(1.243)	(0.441)
$L^5 \Delta \ln($ Average Affinity Proxy $)$	-1.031**	-1.092***	-0.068	-1.912***	-1.309***	0.126	-1.137	-0.446	0.592
	(0.465)	(0.302)	(0.108)	(0.517)	(0.333)	(0.118)	(1.850)	(0.921)	(0.522)
Controls									
Lags of $\Delta \ln n_{cl,t}$	8	8	8	8	8	8	8	8	8
Lags of $\Delta \ln N_t$	3	3	3	3	3	3	3	3	3
Class Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	53,930	53,930	53,930	23,766	23,766	-0.002***	27,173	27,173	27,173
Adjusted R^2	0.145	0.146	0.146	0.183	0.184	(0.0002)	0.126	0.125	0.125

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
White standard errors clustered by class in parentheses.

Further investigation reveals this result is not consistent over time. When I restrict our data to the first century, the negative coefficients remain statistically significant, and generally increase in magnitude (see the middle three columns in Table A2). However, when I restrict our attention to the 1936-2012 period (see the last three columns in Table A2), the sign on two of the proxies flips from negative to positive and statistically significant, and the coefficients associated with Proxy 3 are positive but insignificant. As noted above, I believe the results from the later period are more reliable, since the data better fits the model during this period.

Turning first to the later period, the change in coefficients is primarily due to the decrease in variation in $\Delta \ln\left(\text{Average Affinity Proxy}\right)_{cl,t-l}$ when measured with the prior set of pairs. For example, $\Delta \ln\left(\text{Average Affinity Proxy 1}\right)_{cl,t-l}$ has a mean of 0.005 when measured with the prior set of pairs, compared to 0.034 when measured with the most recent set of pairs. The means of $\Delta \ln\left(\text{Average Affinity Proxy 2}\right)_{cl,t-l}$ and $\Delta \ln\left(\text{Average Affinity Proxy 3}\right)_{cl,t-l}$ actually become negative, as well as closer to zero, when measured with the prior period's pair set. This is primarily because about half of pairs are only assigned to a patent once, and therefore much of the period-to-period change in Average Affinity Proxy comes from pairs being assigned their first and only patent. Such changes are only picked up by measuring $\Delta \ln\left(\text{Average Affinity Proxy}\right)_{cl,t-l}$ with respect to the second period. Thereafter, these pairs either do not change (in Proxy 1), decline for one period (in Proxy 3), or decline by a small amount every period (in Proxy 3). While Proxy 3 appears incapable of picking up the positive impact of Average Affinity Proxy for the 1944-2012 period, Proxies 1 and 2 do, in support of the model.

As discussed in Section 8.3, a potential explanation for the negative coefficients in the earlier sample may be that the USPTO's updated classification has induced bias into the definition of mainlines.

A3 – Alternative Sample Selection

In the first three columns of Table A3, I restore the $n_{cl,t} = 0$ observations by transforming the dependent variable to $\Delta \ln\left(1 + n_{cl,t}\right)$. This leads to a weakening of the coefficients, but they remain positive and statistically significant. In the last three columns of Table A3, I include only classes with more than 1 mainline nested under the class, in case such classes are unusual or the mainlines they draw on are poor proxies. This does not have a significant impact on estimated coefficients, since there appear to be few patents assigned to such classes.

Table A3: Alternative Samples

Dependent Variable:	$\Delta\ln(1+n_{cl,t})$			$\Delta\ln n_{cl,t}$		
Omitted Classes	-			<2 Mainlines		
Proxy	1	2	3	1	2	3
Age	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.00004)	(0.00004)	(0.00004)	(0.00005)	(0.00005)	(0.00005)
$L^3\Delta\ln(\text{Average Affinity Proxy})$	0.080**	0.065**	0.058**	0.210***	0.163***	0.155***
	(0.035)	(0.030)	(0.026)	(0.062)	(0.052)	(0.043)
$L^4\Delta\ln(\text{Average Affinity Proxy})$	0.163***	0.141***	0.136***	0.198***	0.168***	0.191***
	(0.033)	(0.028)	(0.025)	(0.060)	(0.050)	(0.041)
$L^5\Delta\ln(\text{Average Affinity Proxy})$	0.105***	0.100***	0.113***	0.078	0.067	0.113***
	(0.032)	(0.027)	(0.024)	(0.056)	(0.049)	(0.041)
Controls						
Lags of $\Delta\ln n_{cl,t}$	8†	8†	8†	8	8	8
Lags of $\Delta\ln N_t$	3	3	3	3	3	3
Class Fixed Effects	Y	Y	Y	Y	Y	Y
Observations	61,678	61,678	61,678	52,358	52,358	52,358
Adjusted R^2	0.168	0.168	0.168	0.145	0.145	0.145

Note: † - Controls are lags of $\Delta\ln(1+n_{cl,t})$. *White standard errors clustered by class in parentheses.*

$^*p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$