INNOVATION α: WHAT DO STOCK PRICE INDEXES OF IP-INTENSIVE COMPANIES TELL US ABOUT INNOVATION?

By CAROL CORRADO, DAVID MARTIN, AND QIANFAN WU

*Corrado: The Conference Board, 845 Third Ave, New York, NY, 10022 (email: carol.corrado@tcb.org); Martin and Wu: M•CAM International, 513 E Main Street#2014, Charlottesville, VA 22902 (email: dem@m-cam.com and qw@m-cam.com).

"Innovation α" is a moniker for two stock price indexes comprised of companies whose intellectual property (IP) "network" is expected to confer above-average financial returns. The innovation-driven indexes are constructed using quantitative algorithms applied to an extensive global patent database.[1]

This paper outlines the construction of the innovation α indexes, focusing on a patent network analysis tool ("morphogenetics") used to determine the likely uniqueness and novelty of a subject patent. The tool helps pinpoint firms that are most likely to generate value from their intangible assets; it also yields a "commercial score" for patents, the utility of which is demonstrated in the appendix.

Our primary finding is that the innovation α indexes outperform market benchmark indexes by 5-7 percentage points annually since 2013, underscoring the growing importance of intangible capital in the market capitalizations of public firms (e.g., Corrado and Hulten 2010, Corrado et al., 2013). The outperformance also suggests that firms' ability to deploy intangible assets is a likely driver of the much-discussed increasing productivity divergence within industries (e.g., Andrews et al., 2016).

II. Index Methodology

A. Morphogenetics

Morphogenetics (Martin, 2001; Winer et al., 2003, Luse and Martin 2014) utilizes extrapolated patent citation networks to extract both the existing (direct) and hidden (indirect) patent similarities to measure the relationship between patents. Morphogenetics is a statistical method in which models derived from genomics are married with linguistic tools to measure associations in unstructured text corpus.

Unlike canonical natural language processing algorithms that are effective at identifying keywords and themes, our tool identifies associations that evade keyword or

[1] The indexes are the Innovation α® United States index and the Innovation α® Global index. They are part of a trio of indexes (the third is the Martin Global Innovation Equity Trade War Index) developed by M•CAM International. They are listed on the Solactive Index Calculation Platform as INAUP, INAGP, and TWARP, respectively. The indexes are published by The Conference Board; see https://www.conference-board.org/data/bcicountry.cfm?cid=18.

traditional machine learning linguistic techniques by leveraging "crowd-sourced" associations made by patent applicants and patent examiners. Many patents are procured as part of a defensive effort, in which multiple parties allege innovation though seeking a patent to cover and an actual or anticipated business development (Jaffee and Lerner, 2004). Frequently these patents are written with keyword substitution and use classification codes that, intentionally or unintentionally, obfuscate meaning and render detection improbable.

A schematic depiction of the risk exposure due to prior art detected by the morphogenetic patent analysis is given in the appendix. The tool further implies that for any given subject patent, we can reduce its network to two categories based on the attributes of a given patent and its implied impact on a subject patent. By representing the n-dimensional patent network in a two-dimensional icon, the morphogenetic analysis reveals both the direct and indirect networks of patent owners' innovation activities in an easily readable format. These connections and competitive inference cannot be easily discovered by standard methods for scanning patent contents and citations. Unlike citations of academic or artistic works, patent citations often delimit the options of prior innovation, i.e., they may define a narrowing of market options construed from the referenced patent.

The description and identification of the two categories, "Likely Threat" and "Likely Opportunity", are described in detail in the appendix, the upshot of which is as follows: A patent will be found to have limited novelty and uniqueness if there are several other patents that are categorized as a "Likely Threat" with respect to the subject patent; patents in this category have the highest potential prior or concurrent innovation relevance to the subject patent.[2] On the other hand, the originality and commercial value of the subject patent are likely to be higher with the existence of other patents categorized as "Likely Opportunity" with respect to the subject patent. All told, we expect the novelty and the commercial value of the subject patent to decrease due to the existence of one or more patents "X" falling in the likely threat impact category. On the other hand, a patent ("Y") will be categorized as a likely opportunity relative to the subject patent if the morphogenetic analysis finds that the subject patent may increase its commercial value by utilizing and enforcing "Y" as a potential licensing opportunity.

[2] The category also singles out cases of probably "reverse engineering", the practice of creating a patent based largely on the innovative steps of an ancestral patent while taking steps during the prosecution of a patent to obscure a citation connection.

B. Innovation Cohort Creation

To construct our stock price indexes, we group similar public companies by their innovation abilities, as derived from the morphogenetic analysis. For a subject public company, we assert that a group of other public companies are very likely to share similar R&D focuses, have similar intangible assets distributions, and compete in the same innovation space with the subject company if these companies hold a large number of patents that are in the likely threat category in relation to the patents owned by the subject company.

For every subject public company i, let T_{ij} denote the "threat count" of company i's patent portfolio with respect to another public company j. Let M and N denote the total number patents for company i and j. The indicator function $\mathbb{1}_{\{P_{ij} \in \{Threats\}\}}$ returns 1 if the morphogenetic attributes of patent j relative to subject patent i belongs to the "Likely Threat" category, and returns 0 otherwise:

$$(1) \qquad T_{ij} = \sum_{i=1}^{M} \sum_{j=1}^{N} \mathbb{1}_{\{P_{ij} \in \{Threats\}\}}$$

In order to identify the companies with similar intangible assets that are likely to compete closely with the subject company i, we rank T_{ij} in descending order for all companies j in our universe. Then we choose the top N companies based on the ranking to form a cohort for each subject company i. We selected $N = 5$ for our United States index and $N = 10$ for our Global Index as cohort sizes.

A cohort can be interpreted as a peer group of companies that share a common amount of business resources, follow each other's innovation activities very closely, and/or act as active members within each other's supply chain. We define the subject company i as a "sentinel company" and all other companies in the cohort as "innovation cohort members". We used the "threat count" rather than the "opportunity count" to group companies because the former allows us to infer the closest innovation *overlap* between companies.

The innovation cohort members of two companies: Apple Inc. and Procter & Gamble Company as of September 1, 2019 are shown in appendix tables A.1 and A.2. The cohorts demonstrate that Apple Inc. is competing most closely with companies such as Microsoft and IBM in its innovation activities while Procter & Gamble's innovation competitive space includes Kimberly-Clark, 3M, etc.

C. Price Dynamics Prediction

To capture the competition advantages reflected in company stock prices and build a statistical model with the capacity to predict future stock price dynamics between cohort members, we assume the following: First, consistent with efficient markets in its weak form, we assume that the stock price dynamics of a company reflects information on its market advantage (including the value of individual patents per Kogan et al. 2016), but that the full extent of its competitive network (especially connections obscured by language, misclassification, and concurrent filings) requires inference by the morphogenetics tool.

Second, we assume that the total amount of business and innovation resources among companies in any competition domain is fixed in the short-run, which suggests the relative advantage a company has in exerting pricing power and controlling costs across their value chain is likely to vary across their competition peer group over time (Lee et al., 2000). Consequently, the market price dynamics of companies within the cohorts are a significant factor in formulating innovation-driven stock price indexes.

Inspired by the advancements in machine learning (ML) methods in processing complex time series data (Liaw et al., 2002; Deng et al., 2013), we built a random forest prediction model that processes, extracts, and selects input features based on their statistical significance and calculates the predicted return differences of two companies in the upcoming quarter. A range of time series rolling features are constructed using the historical stock price time series of the 2 companies in the last 5 years. In our index construction methodology, within a cohort, we run the return prediction model for each cohort member versus the sentinel company. If the cohort number is 5, there are 5 quarterly predictions.

D. Index Construction

In order to measure the investment opportunities both across the U.S. and worldwide cohorts, we draw from companies that are constituents of widely recognized indexes. We also limit our selection to large-cap companies in developed markets because the patent and stock price data, stock trading liquidity, and corporate governance of large-cap companies (market capitalization ≥ \$10 billion) are more transparent. The companies from the U.S. index were selected from the Russell 1000 index, and the companies in the global index are drawn from the MSCI ACWI index (developed countries only).[3]

[3] A common challenge to grouping patents by the organizations owning them (the patent assignee) is that assignees are often written in ambiguous forms. In addition, if a patent is owned by one of the subsidiaries of a public company, usually only the subsidiary name will be shown within the patent context. To correctly identify the patent portfolio for companies in our universe, we use company filling information available on the SEC website to conduct a de-aliasing procedure that matches the ambiguous company names and subsidiary names with the original companies to which they associate. Corporate mergers, acquisitions and other activities are monitored on a regular basis to ensure that company ownership of IP is kept up to date.

The initial selection of the companies involves scoring every patent held by every company in our universe, updating innovation cohort, and calculating a return prediction as of the beginning of each quarter for 5 years prior to the commencement of the indexes.

A ranking score for each company *i* over the 5-year period is then calculated as an average of the number of times the company appears among all cohorts over the past 5 years; the number of times the company has the highest quarterly return prediction within its cohort over the past 5 years; and the company's average normalized quarterly return prediction within its cohort over the past 5 years.

The first factor (the appearance count) represents how active a company is in competing with its peers in innovation activities, i.e., the company's competitive innovation intensity. The other two factors signal the market performance of companies. Gaining a high score in the second and third factor indicates that the company is predicted to have a more advantageous market value outlook, which also implies it is likely to be more successful in commercializing its innovation and turning its patent resources into market value and profits (e.g., via strong brands).

Our index period commences in January 2013 and continues through October 2019. Cohort creation and quarterly return model training were conducted as per the methodology described above for each quarter between January 2008 to January 2013. We selected the top 100 and 120 companies with the highest weighted average scores for our U.S and global indexes respectively. The patent-based morphogenetics and cohort data are updated every quarter. Attempting to fully capture the market advantage shifts for each quarter starting from January 2013, our methodology weights companies based on its ranking scores for each quarter. To ensure that new companies with high innovation ability and potential market advantages can be added to the index, we replace 10 percent of the index components every year. A sentinel company will be replaced if one of its cohort members falls in the top 10 percent of all quarterly return prediction scores against its sentinel.

II. Results: Index Performance

The investment performance of the U.S. and global stock-price indexes are evaluated from January 2013 to September 2019. The modeled investment growth of $1000, descriptive statistics of the indexes as well as their benchmarks are demonstrated in Table 1. It can be observed that despite a slight increase in annualized volatility for both indexes, the indexes significantly outperformed their benchmarks (Russell 1000 Index and the MSCI ACWI Index) with respect to total investment growth and annualized return. See the appendix for charts displaying the time series on which these results are based.

TABLE 1: PERFORMANCE STATISTICS FOR THE INNOVATION α INDEXES

percent	U.S. Index	Global Index	Russell 1000	MSCI ACWI
Return (%)	16.12	14.97	11.11	6.03
Volatility (%)	14.33	12.79	12.95	10.78

The table suggests that our methodology that combines patent innovation inference networks and stock price dynamic prediction can successfully capture the market advantages of companies that are active in innovation activities. The U.S. and Global indexes are thus not only investment tools, but also indicators of the top innovative companies with the strongest ability to achieve market advantages in competition with their "innovation peers."

III: Conclusion

In this paper we introduced the morphogenetics patent innovation inference networks system to examine the direct and indirect relationship between individual patents. Additionally, we constructed two innovation-based stock price indexes to evaluate the investment advantages based on the system and analysis of stock price dynamics. Our methods for exploiting patent data constructing a stock price index of IP-intensive companies is, we believe, a novel addition to the innovation literature. The morphogenetics tool is designed to better capture the hidden competition interactions between patents than previous citation network analysis systems. The innovation α indexes were found to outperform their benchmarks without sacrificing risk dynamics.

The appendix to this paper reports statistics on "commercializable patents" for companies in two large technology-driven industry sectors, healthcare and telecommunications. The "commercial score" statistics reveal that non-U.S. companies generally have a higher percentage of commercializable patents than their U.S. counterparts.

The relative performance of our stock price indexes raises interesting questions in light of the commercial score analysis. As suggested by table 1 and shown in appendix figure A.4, the U.S. companies subindex of the Global index outperforms its non-U.S. companies subindex—despite the latter's apparent edge in commercializable patent holdings (in the two sectors we analyze). This may suggest that U.S. global firms are more successful at, e.g., exploiting synergies between their patent portfolios and other intangible assets. But many factors influence the disparate pattern in financial performance and commercial scores, and a fruitful line for future research is to (a) conduct deeper examinations of patent portfolio quality across sectors, industries, and geographies, and (b) further explain the market value variance of companies due to intangible assets using the patent innovation network system introduced in this paper.

REFERENCES

Andrews, Dan, Chiara Criscuolo, and Peter N. Gal. 2016. "The Global Productivity Slowdown, Technology Divergence and Public Policy: A Firm Level Perspective" OECD Productivity Working Papers No. 5, OECD Publishing, Paris.

Corrado, Carol, Jonathan Haskel, Cecilia Jona-Lasinio, and Massimiliano Iommi. 2013. "Innovation and Intangible Investment in Europe, Japan, and the United States." *Oxford Review of Economic Policy*, 29(2). 261–286.

Corrado, Carol, and Charles Hulten. 2010. "How Do You Measure a 'Technological Revolution'? *American Economic Review*, 100(2). 99–104.

Deng, H., Runger, G., Tuv, E. and Vladimir, M., 2013. "A time series forest for classification and feature extraction." *Information Sciences*, 239. 142–153.

Jaffe, Adam, and Josh Lerner (2004). *Innovation and its Discontents.* Princeton University Press.

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, Noah Stoffman. 2016. "Technological Innovation, Resource Allocation, and Growth," *Quarterly Journal of Economics*, 132(2). 665–712.

Lee, H., Smith, K.G., Grimm, C.M. and Schomburg, A., 2000. "Timing, Order and Durability of New Product Advantages with Imitation." Strategic Management Journal, 21(1). 23–30.

Liaw, Andy and Matthew Wiener. 2002. "Classification and Regression by randomForest." *RNews*, 2(3). 18–22.

Luse, Hayden T. and David E. Martin (2014) "Text-Based High Latency, Low Frequency Signal Analysis and Cohort Equity," CISML Conference, University of Tennessee and Oak Ridge National Labs; available as Conference Board EPWP #19 – 01.

Martin, David E, Mosaic Technologies Inc, 2001. *Method and apparatus for establishing and enhancing the creditworthiness of intellectual property*. U.S. Patent 6,330,547.

Winer, D.S., Martin, David E., Watson, J.O. and Pratt, D.J., M CAM Inc, 2003. Method and system for graphical representation of multitemporal, multidimensional data relationships. U.S. Patent 6,665,670.

Innovation α: What Do Stock Price Indexes of IP-intensive Companies Tell Us About Innovation?

By CAROL CORRADO, DAVID MARTIN, AND QIANFAN WU*

Appendix

Contents:

A. Additional Background

As one of the most significant forms of intellectual property, patents may not only be valuable to the individuals and organizations that possess them but may also be used to infer the dynamics and context of the innovations they disclose (Kline et al., 2019). Analyzing patents devoid of their contexts, including the prevailing legal and patent office approval practices, does not yield a complete picture of the value of innovation (Martin, 2001). Consequently, some studies (Yoon et al., 2004; Erdi et al., 2013; Mariani et al, 2017) have adopted network citation analysis to discover the complex hidden values within patent citation networks. These studies, as well as the traditional approach in the economics literature, generally take a simple quantitative approach, assuming that the value of an organization's patent portfolio increases if they have greater patent numbers and patent citation counts. Here, we take the position that patent and citation counts are

* We thank Christos Makridis and Nicholas Wales for helpful comments on an earlier draft.

generally poor measures of innovation activity due to the limited scope and potential information they can convey. As indicated in the main text, unlike academic or artistic works, patent citations can delimit the options of prior innovation, often defining a narrowing of market options construed from the referenced patent.

Additionally, patents vary in purpose (e.g., defensive vs offensive, Pavitt 1988) and quality (e.g., Griliches, 1990). From a commercial standpoint, a higher citation count may indicate reduced patent novelty due to the existence of competitors in the marketplace, as well as a narrowing of the inferred scope of the invention claimed, factor not generally acknowledged in the economics literature (e.g., for a review of commonly used metrics, see Squicciarini et al., 2013). Previous citation *network* research seeks to interleave citations and keywords to define quality or importance (Yoon et al., 2004; Tseing et al., 2007). Unfortunately, as patent applicants can be their own 'lexicographer' under the statute and under judicial review, keywords fail to acknowledge the use of metaphor (both overt and covert) that masks the ordinary meaning of words or concepts. All these challenges present the need for developing a methodology that effectively analyses patent novelty and related commercial value to fully understand the environment in which innovation drives market value.

B. Introduction to Morphogenetics

Our work introduces a patent innovation inference network analysis tool ("morphogenetics") that has been developed to determine the likely uniqueness and novelty of a subject patent. An explanation of the morphogenetic tool is shown in Figure 1. The figure illustrates a patent citation network between a patent, call it "X", and a subject patent (Label: 110). The arrows in the figure indicate the direction of a citation, with the triangle icons represent first order relationship between "X" and the subject patent, and the circle icons represent second order relationship between "X" and the subject patent. Icons labelled 120 – 170 demonstrate the statuses patent "X" can have with respect to the subject 110. Patent "X" can be in one or multiple statuses with respect to 110. We call Patent "X" the prior art of subject 110 if "X" is in status 120, meaning "X" is cited by 110. We call "X" the subsequent art of 110 if "X" is in status 150, in which case "X" cites 110. If "X" is in status 160, it means "X" is the prior art of patents citing 110. If "X" is in status 140, it means "X" is the subsequent art of the patents cited by 110. In addition, if "X" is both in status 140 and 160, it means "X" is the both the subsequent art of the patents cited by 110, as well as the prior art of patents citing 110. Patent “X” doesn’t cite or is cited by the subject 110 directly except in the status 120 or 150.

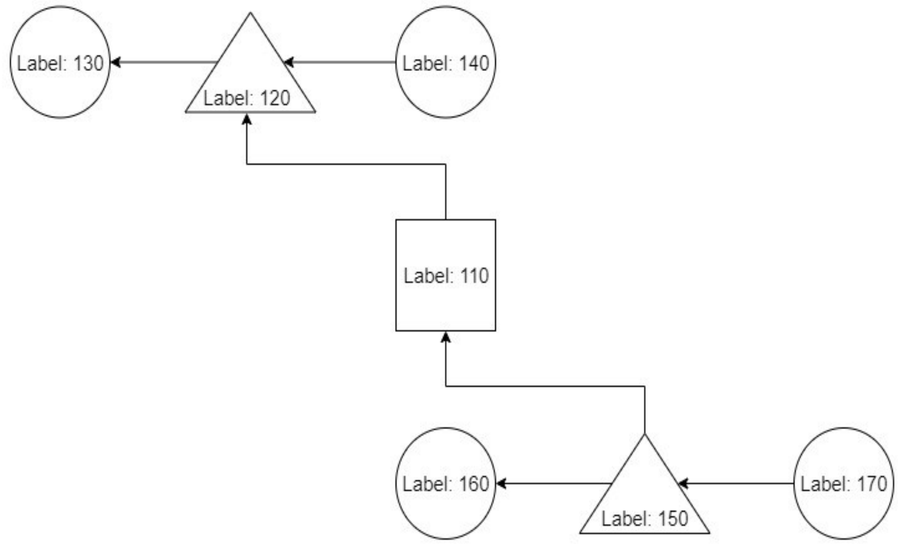

FIGURE A.1: AN ILLUSTRATION OF THE RELATIONSHIP BETWEEN PATENT “X” AND THE SUBJECT PATENT LABELLED 110

The morphogenetic tool seeks to uncover “thesaurus” patents by detecting work that other related patents and patent examiners have identified as conforming to the legal standard of relevance, but which were not cited by the subject patent. The more there are relevant works identified in the network analysis that were not cited, the higher the likelihood that the target patent is a functional redundancy. From this we can conclude that a patent is moderately to severely impaired and of limited commercial value.

For any given subject patent, by using the morphogenetic analysis, we can further identify 2 categories based on the attributes of patent “X” and its implied impact on the subject patent. These two categories are illustrated in Table A.1. A patent tends to have limited novelty and uniqueness if there are several other patents which are categorized as “Likely Threat” with respect to the subject patent. In contrast, originality and commercial value of the subject patent are likely to be higher with the existence of other patents categorized as “Likely Opportunity”. For example, a patent “X” will be categorized as “Likely Threat” to the subject patent if it has statuses 140, 160, and 170. We expect the novelty of commercial value of the subject patent to decrease due to the existence of several patents similar to “X”. On the other hand, patent “Y” will be categorized as “Likely Opportunity” to the subject patent if it has statuses 160 and 170, indicating the subject patent may increase its commercial value by utilizing and enforcing potential licensing opportunity.

Table A.1—Morphogenetic Categories

CATEGORY	PATENT ATTRIBUTES AND EXPLANATION
Likely Threat	*Patents in this group:* --Have a priority date preceding the subject patent, identified by the system as having claims language sufficiently consistent with that of the subject patent to be included in the innovation space of the subject patent yet were not cited by subject patent. --Or, were undergoing office action at the same time as the subject patent during some period of the subject patent's prosecution history, but neither cite nor are cited by the subject patent. This group contains patents with highest potential prior or concurrent innovation relevance to the subject patent. It also singles out cases of probable "reverse engineering," the practice of creating a patent based largely on the innovative steps of an ancestral patent, while taking steps during the prosecution of a patent to obscure a citation connection.
Likely Opportunity	*Patents in this group:* --Have been filed after the subject patent, identified by the system as having claims language sufficiently consistent with that of the subject patent to be included in the innovation space of the subject patent yet did not cite the subject patent. In this group one may identify potential licensing opportunities as the enforceability or commercial validity of patents in this group could be limited by the claims of the subject patent.

C. Innovation Cohorts: Examples

An innovation cohort consists of companies against which the sentinel company competes most closely with regard to innovation, i.e., companies whose morphogenetic threat scores vis a vis the sentinel company are the highest. Tables A.2 and A.3 show the cohort members of two sentinel companies, Apple, Inc. and Proctor & Gamble Company, as of September 1, 2019.

Table A.2: Innovation Cohort Members for Apple Inc. on 9/1/2019

	Company Name	*Sector*
Sentinel Company:	Apple Inc.	Electronic Technology

Cohort Members:	Microsoft Corporation	Technology Services
	International Business Machines Corporation	Technology Services
	Sony Corporation Sponsored ADR	Consumer Durables
	Nokia Oyj Sponsored ADR	Electronic Technology
	HP Inc.	Electronic Technology Technology Services

Note: FactSet Industries and Economic Sectors Classification System is used to determine the company sectors. FactSet Online Assistant®, Page ID 6739.25293. See Appendix E for more details.

Table A.3: Innovation Cohort Members for Procter & Gamble Company on 9/1/2019

	Company Name	*Sector*
Sentinel Company:	Proctor & Gamble Company	Consumer Non-Durables
Cohort Members:	Kimberly Clark Corporation	Consumer Non-Durables
	3M Company	Producer Manufacturing
	Johnson & Johnson	Health Technology
	Colgate-Palmolive Company	Consumer Non-Durables
	Pfizer Inc.	Health Technology

Note: FactSet Industries and Economic Sectors Classification System is used to determine the company sectors. FactSet Online Assistant®, Page ID 6739.25293. See Appendix E for more details.

D. Index Performance: Charts

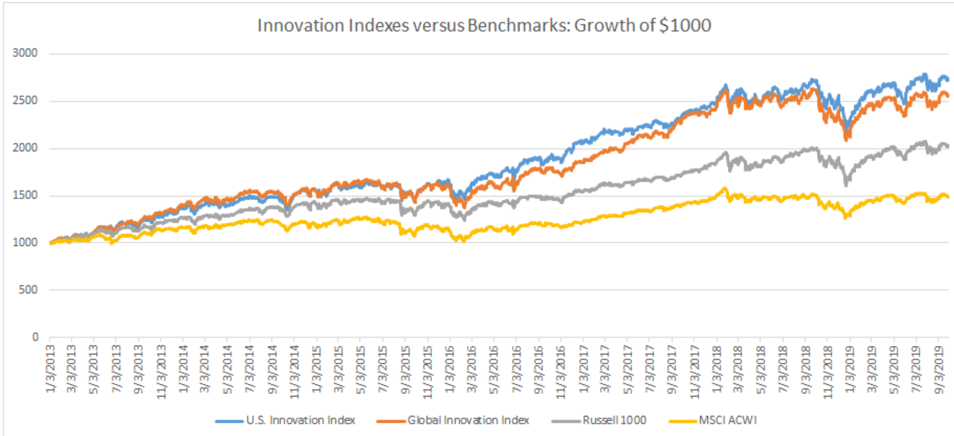

FIGURE A.2: INNOVATION α INDEXES PERFORMANCE VERSUS THEIR BENCHMARKS.

Note: The Morgan Stanley Capital International (MSCI) all country world index (ACWI) is a market capitalization weighted index designed to provide a broad measure of world equity-market performance.

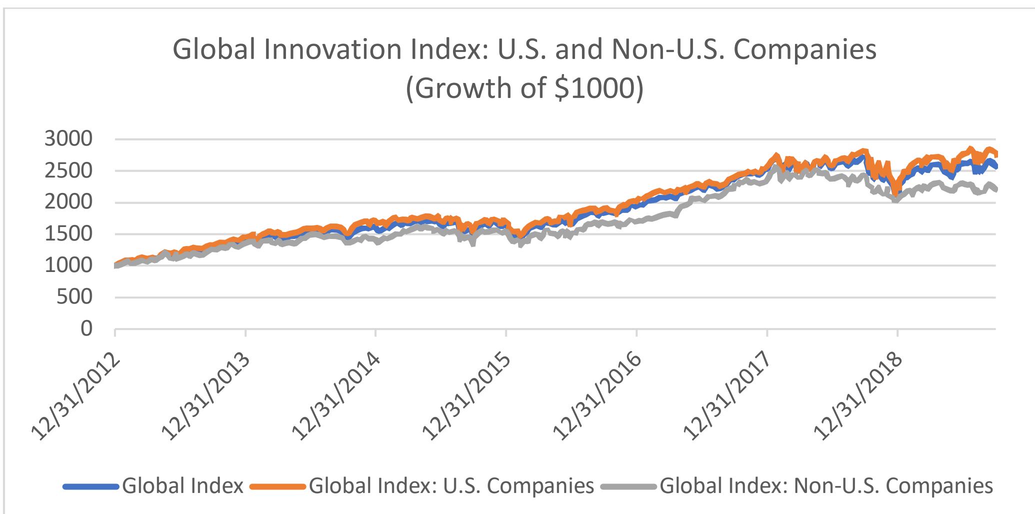

FIGURE A.3: U.S. AND NON-U.S. COMPANIES: COMPONENTS OF THE INNOVATION α GLOBAL INDEX.

Note: Based on the selection algorithm, the Global Index consists of 74 U.S. companies and 46 non-U.S. companies as of September 2019.

E. Patent Counts and Commercial Scores

In addition to the index methodology introduced in the main text, we propose a commercial score measure to evaluate the commercial value of a company's patent portfolio.

(a) Method

If a patent is categorized as "Likely Opportunity" through the morphogenetics, one may identify commercialization opportunities against the patent due to the limited innovation scope defined by the subject patent. We define a subject patent as a "commercializable patent" if another patent owned by a different organization exists in the "Likely Opportunity" category subject to the patent under review. Note also that a patent is categorized as "Likely Opportunity" only if it was filed after the subject patent. Data for some patents, especially the international equivalents for the primary patent, may be incomplete due to the variation in filing standards between different patent offices. Consequently, we only consider the patents with complete data in our commercial score calculation and define them as "scorable patents".

We define the commercial score CS_i as the percentage of commercializable patents among the patent portfolio:

$$CS_i = {}^{C_i}/_{N_i} * 100\% \tag{A.1}$$

where C_i is the number of commercializable patents and N_i is the total number of scorable patents. The commercial score evaluates the overall potential quality of a patent in the marketplace. It is also an indicator of the commercial quality of a company's patent portfolio when aggregated.

Commercial scores for two economic sectors: healthcare technology (pharma, biotech, medical devices, etc.) and telecommunications (equipment and services). The FactSet Industries and Economic Sectors Classification System was used to determine the sectors and industries for companies; details on this classification system are provided below. We selected 355 companies for the health technology sector and 43 for the telecommunications sector. The healthcare companies were selected by reviewing and combining the holdings of several well-recognized healthcare indexes and Exchange Traded Funds (ETFs), such as the Health Care Select Sector SPDR® and the MSCI World Healthcare Index. For telecommunications, we selected companies falling into these sectors from a universe consisting of all sentinel companies and cohort members in our U.S. and Global indexes as of July 1, 2019. Descriptive data for the companies are shown in Table A.4.

Table A.4: Descriptive data of companies selected for commercial score calculation

	No. of companies	Average market cap (billion $)	Average No. of scorable patents	Std. Dev. of scorable patents
Healthcare technology				
U.S. Companies	199	19.26	915.39	2704.70
Non-U.S. Companies	156	13.14	823.32	2356.07
Telecommunications				
U.S. Companies	22	24.12	6709.14	13939.50
Non-U.S. Companies	21	111.61	2535.38	5358.29

Note: For the healthcare technology sector, the number of scorable patents account for 44.37% of the total granted patents for U.S. companies, 63.54% for Non-U.S. companies; For the telecommunications sector, the number of scorable patents account for 49.18% for U.S. companies, 68.97% for Non-U.S. companies.

(b) Results

We calculated the commercial scores for companies in the two sectors and then grouped them by U.S. companies and non-U.S. companies. The results are presented in Figure A.4. In general, the percentage of commercializable patents held in the telecommunications sector is higher than the percentage of those held in the healthcare technology sector, by approximately 20 percent. Figure A.3 further shows that non-U.S. companies in both sectors have a higher percentage of commercializable patents. The differences in commercial scores between U.S. and Non-U.S. companies are possibly the consequences of the differences in patent law, competition intensity, and emphasis on R&D investments. Additionally, it is potentially the case that international companies have lower dependency on patent thickets for litigation risk management when compared to their U.S. counterparts. The data we calculate also indicates that U.S. companies are more likely to file "defensive" patents. On the other hand, non-U.S. companies are more inclined to file forward-looking patents purely for innovation purposes.

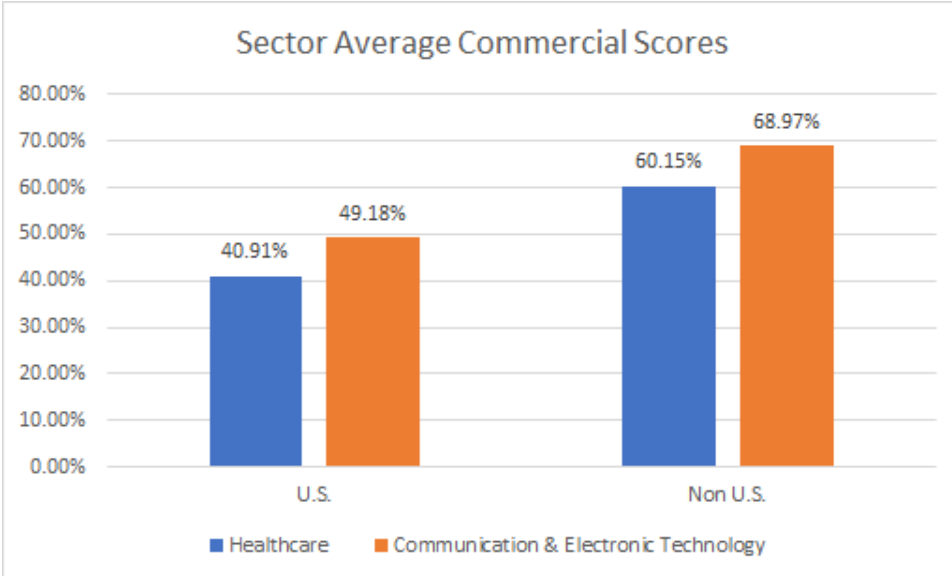

FIGURE A.4: COMMERCIAL SCORES FOR U.S AND NON-U.S. COMPANIES FOR THE HEALTH TECHNOLOGY AND TELECOMMUNICATIONS SECTORS

A further examination of the commercial scores involves breaking down the two sectors into several sub-industries. Because the industry distribution is highly skewed, we grouped only the top 5 sub-industries by company counts for the health tech sector and the telecommunication sector to avoid sub-population bias. The result for the two sectors is shown in Figures A.5 and A.6. Within the health technology sector, companies in the pharmaceutical industry tend to have higher commercial scores whereas scores for companies in medical specialties are slightly lower. This result is not surprising given that the pharmaceutical industry is known to rely heavily on innovation activities, with R&D expenditures higher than most of other industries (OECD 2017). Our data reinforce this result by demonstrating that the patent portfolios of pharmaceutical companies tend to have higher market values.

Within telecommunications, the wireless telecommunications industry achieved the highest level of commercial scores. Several companies in this industry have large patent portfolios and leading individual commercial scores as well. For example, Japanese company NTT DOCOMO INC has 6952 scorable granted patents and 79.47% of them are considered commercializable. On the other hand, the Computer Communications industry possessed the least commercial percentage among the top 5. However, since the population size within this industry is limited, further examination across a larger population size over this sector is expected in the future.

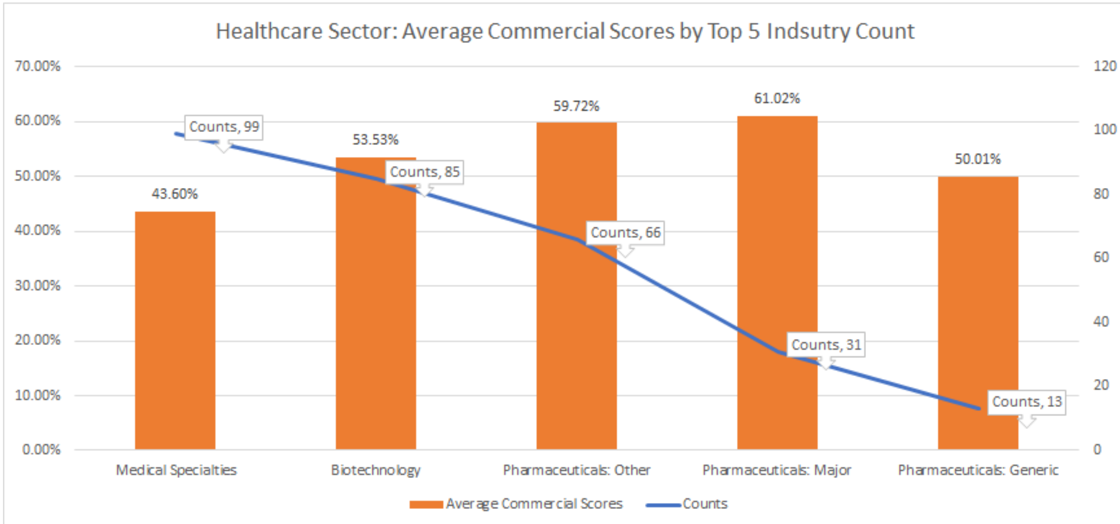

FIGURE A.5: AVERAGE COMMERCIAL SCORES BY TOP 5 INDUSTRY COUNT WITHIN THE HEALTH TECHNOLOGY SECTOR

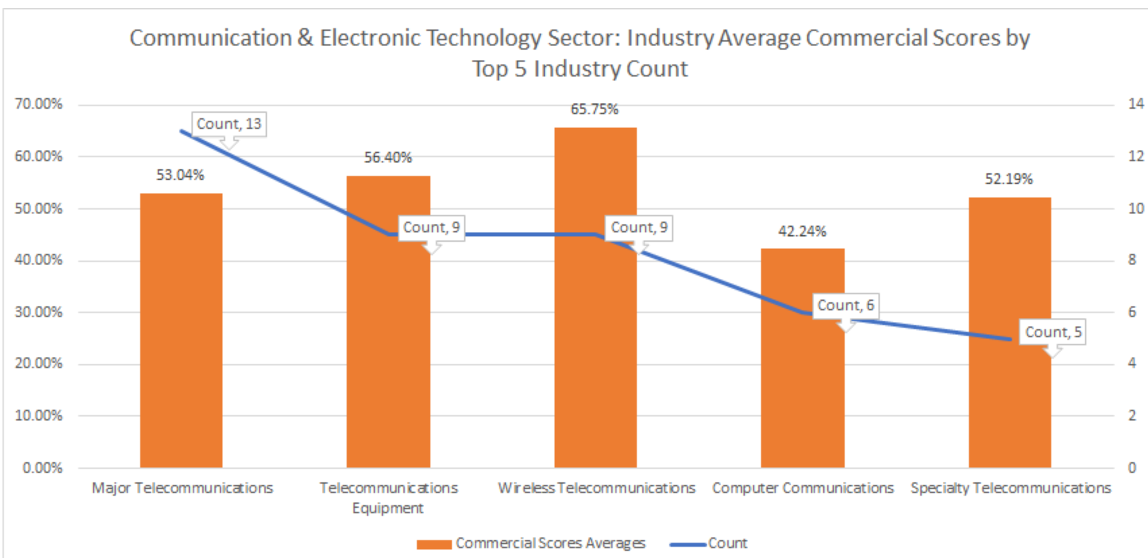

FIGURE A.6: THE AVERAGE COMMERCIAL SCORES FOR THE 5 INDUSTRIES WITHIN THE TELECOMMUNICATIONS SECTOR

The commercial score calculation for the two sectors under review suggests that non-U.S. companies generally have a higher percentage of commercializable patents than their U.S. counterparts. Yet, as also shown in figure A.4, the US companies sub-index of the Global index outperform non-U.S. companies, suggesting that U.S. firms exploit synergies with other intangible assets, e.g., organizational capital and/or the deployment of IT (e.g., Bloom et al., 2012). In the future, more research is expected on developing more efficient and effective analytical methods to explain the market value variance of companies by using patent and other intangible capital statistics and conducting deeper examinations on the patent portfolio quality across sectors, industries, and geographical areas.

(c) The FactSet Company Classification System

FactSet maintains a proprietary industry classification system in which every company carried in any of the fundamental databases on FactSet is assigned to a FactSet industry. The industries are organized into four general economic categories:

- Durables
- Non-durables
- Services
- Infrastructure

The first three of these (Durables, Non-durables, and Services) are cross-referenced into four economic sectors:

- Materials
- Producer
- Technology
- Consumer

The fourth general category, Infrastructure, incorporates services that are pervasive throughout the economy: transportation, utilities, finance, and communication services.

The resulting matrix generates a coherent and relevant organization of potentially investable corporations. One of the primary goals in this enterprise is to try to identify patterns of economic and industrial change that may not be as readily discernible elsewhere. The contact information for more details on the system is provided on the following website. https://www.factset.com/contact-us.

In our analysis, the healthcare sector includes two healthcare related FactSet economic sectors drawn from the classification system: Health Technology and Health Service. The communication & electronic technology sectors include two related FactSet economic sectors: communication and electronic technology. These FactSet sectors were combined to form the two "broader" sectors in our analysis. Their descriptions of the sub-industries shown in figures A.4

and A.5 are described below, and a table showing examples of companies included in groups follows.

Healthcare sector:

- Medical Specialties: This industry group consists of companies that develop and manufacture equipment and technologies designed specifically for the healthcare industry. Examples: Dialysis equipment, bacterial and viral identification kits, hygiene control equipment, blood analysis equipment, precision equipment, laboratory automation equipment, and patient monitoring systems, disposable bed pads, catheters, surgical equipment, and other medical diagnostic equipment.
- Biotechnology: This industry group consists of companies involved in the application of genetic engineering (genomics) and/or protein engineering (proteomics) to produce therapeutic and preventive medicine. and medical diagnostic products. Companies that manufacture biotechnology equipment and provide services to the biotech industry are also included in this industry.
- Pharmaceuticals: Major: This industry group consists of companies that discover, develop, and manufacture chemically based therapeutic and preventive medicine, and other medicinal products. Companies included in this industry generally oversee the entire drug development process (research, testing, production, distribution, etc.). Examples: Cardiovascular and anti-cancer drugs, antibiotics, vaccines, contraceptives, mental health products, and cough and cold medicine.
- Pharmaceuticals: Other: This industry group consists of companies that either discovers, develop, or manufacture chemically based therapeutic and preventive medicine, and medicinal products. Companies included in this industry tend to collaborate with other pharmaceutical companies.
- Pharmaceuticals: Generic: This industry group consists of companies engaged in the manufacture of generic therapeutic and preventive medicine, and other medicinal products.

Communication & electronic technology sector:

- Computer Communications: This industry consists of companies engaged in the manufacturing of computer connectivity and network products. Example: Routers, remote access servers, applications for Token Ring, Ethernet, and other high-speed networks, shared media hubs, local area networks (LAN) and wide-area networks (WAN), and Internet protocol (IP) products.
- Major Telecommunications: This industry group consists of companies that operate, maintain, and/or provide voice and data transport services based on multiple transmission (fixed line, digital subscriber line (DSL) technology, competitive local exchange carriers (CLEC), Internet-based communication services, wireless, etc.) technologies. Example: Local and long distance telephone services, message telecommunication services,

wireless services, Internet access (both cable and integrated services digital network (ISDN)), and directory and calling card services.

- Wireless Telecommunications: This industry group consists of companies that provide wireless antenna- or satellite-based telecommunication services. Example, Beeper and paging services, specialized mobile radio (SMR) services, and other wireless communication services.
- Telecommunications Equipment: This industry group consists of companies engaged in the manufacturing of voice and data communications equipment. Example: Fiber optic delivery products, digital signal processing (DSP), high speed voice, data, and video delivery and access platforms, global positioning systems (GPS), satellite systems, paging and wireless data systems, personal communication equipment and systems, two-way land mobile systems, wireless microcell systems, private branch exchange switches (PBX), telephone handsets, residential systems, and payload equipment for satellites.
- Specialty Telecommunications: This industry group consists of companies that operate, maintain, and/or provide voice and data transport services based on a single transmission (fixed-line, digital subscriber line (DSL) technology, Internet-based communication services, etc.) technology, and/or cover a specific market (facilities-based, competitive local exchange carriers (CLEC), etc.). Companies that provide services to the telecommunication industry are also included in this industry.

Table A.5: Examples of companies included in the commercial score calculation

Sub-Industry	**Sector**	**Examples of Companies included in Commercial Score Computation**
Medical Specialties	Health Technology	Abbott Laboratories; Medtronic, Inc; Thermo Fisher Scientific, Inc.
Biotechnology	Health Technology	Amgen, Inc; Gilead Sciences, Inc; CSL Limited
Pharmaceuticals: Major	Health Technology	Johnson & Johnson; Pfizer, Inc; Novartis AG
Pharmaceuticals: Other	Health Technology	Bayer AG; Daiichi Sankyo Co., Ltd.; Chugai Pharmaceutical Co., Ltd.
Pharmaceuticals: Generic	Health Technology	Zoetis, Inc.; Allergan, Inc.; Mylan, Inc.
Computer Communications	Electronic Technology	Fortinet, Inc.; Arista Networks, Inc.; Cisco Systems, Inc.
Major Telecommunications	Communications	BCE Inc.; Deutsche Telekom AG; Royal KPN NV
Wireless Telecommunications	Communications	Rogers Communications Inc; Sprint Corp.; T-Mobile US, Inc.
Telecommunications Equipment	Electronic Technology	Nokia Oyj; Garmin Ltd.; QUALCOMM Incorporated
Specialty Telecommunications	Communications	SoftBank Group Corp.; SES SA; CenturyLink, Inc.

ADDITIONAL REFERENCES

Balsmeier, B., Assaf, M., Chesebro, T., Fierro, G., Johnson, K., Johnson, S., Li, G.C., Lück, S., O'Reagan, D., Yeh, B. and Zang, G., 2018. Machine learning and natural language processing on the patent corpus: Data, tools, and new measures. *Journal of Economics & Management Strategy*, 27(3), pp.535-553.

Cockburn, I.M. and Griliches, Z., 1987. Industry effects and appropriability measures in the stock markets valuation of R&D and patents.

De Rassenfosse, G. and de la Potterie, B.V.P., 2009. A policy insight into the R&D–patent relationship. *Research Policy*, 38(5), pp.779-792.

Érdi, P., Makovi, K., Somogyvári, Z., Strandburg, K., Tobochnik, J., Volf, P. and Zalányi, L., 2013. Prediction of emerging technologies based on analysis of the US patent citation network. *Scientometrics*, 95(1), pp.225-242.

Griliches, Z., 1990. Patent statistics as economic indicators: A survey. *Journal of Economic Literature* 28:4 (December), pp 1661-1707.

Hall, B.H., Jaffe, A. and Trajtenberg, M., 2005. Market value and patent citations. *RAND Journal of Economics*, pp.16-38.

Hu, J., Li, S., Yao, Y., Yu, L., Yang, G. and Hu, J., 2018. Patent keyword extraction algorithm based on distributed representation for patent classification. *Entropy*, 20(2), p.104.

Jaffe, A.B. and Trajtenberg, M., 2002. *Patents, citations, and innovations: A window on the knowledge economy*. MIT press.

Jaffe, A.B. and Lerner, J., 2004b. Patent Prescription: A radical cure for the ailing [US patent policy]. *IEEE Spectrum*, 41(12), pp.38-43.

Kline, P., Petkova, N., Williams, H. and Zidar, O., 2019. Who profits from patents? Rent-sharing at innovative firms. *The Quarterly Journal of Economics*, 134(3), pp.1343-1404.

Lanjouw, J.O., Pakes, A. and Putnam, J., 1998. How to count patents and value intellectual property: The uses of patent renewal and application data. *The Journal of Industrial Economics*, 46(4), pp.405-432.

Mariani, M.S., Medo, M. and Lafond, F., 2017. *Early identification of important patents through network centrality*. arXiv preprint arXiv:1710.09182.

Martin, David E. 2001. Patents: Improving Quality and Curing Defects: Hearing Before the Subcommittee on Courts, the Internet, and Intellectual Property of the Committee on the

Judiciary House of Representatives,107th Congress, First Session (May 10). http://commdocs.house.gov/committees/judiciary/hju72305.000/hju72305_0f.htm

OECD (2017), "Research and development in the pharmaceutical sector", in Health at a Glance 2017: OECD Indicators, OECD Publishing, Paris, https://doi.org/10.1787/health_glance-2017-72-en.

Pavitt, K., 1988. Uses and abuses of patent statistics. In *Handbook of quantitative studies of science and technology* (pp. 509-536). Elsevier.

Squicciarini, Mariagrazia, Hélène Dernis, and Chiara Criscuolo. 2013. "Measuring Patent Quality: Indicators of Technological and Economic Value." OECD DSTI Working Papers, 2013/03, OECD Publishing.

Trappey, A.J., Trappey, C.V., Wu, C.Y. and Lin, C.W., 2012. A patent quality analysis for innovative technology and product development. Advanced Engineering Informatics, 26(1), pp. 26-34.

Tseng, Y.H., Lin, C.J. and Lin, Y.I., 2007. Text mining techniques for patent analysis. Information Processing & Management, 43(5), pp. 1216-1247.

Wu, J.L., Chang, P.C., Tsao, C.C. and Fan, C.Y., 2016. A patent quality analysis and classification system using self-organizing maps with support vector machine. *Applied soft computing*, 41, pp. 305-316.

Yoon, B. and Park, Y., 2004. A text-mining-based patent network: Analytical tool for high-technology trend. *The Journal of High Technology Management Research*, 15(1), pp.37-50.