Growth through Inter-sectoral Knowledge Linkages

Cai, Jie and Li, Nan

Shanghai University of Finance and Economics, International Monetary Fund

5 November 2012

Growth Through Inter-sectoral Knowledge Linkages

Jie Cai and Nan Li*

This version: May 2018

Abstract

The majority of innovations are developed by multi-sector firms. The knowledge needed to invent new products is more easily adapted from some sectors than from others. We study this network of knowledge linkages between sectors and its impact on firm innovation and aggregate growth. We first document a set of sectoral-level and firm-level observations on knowledge applicability and firms' multi-sector patenting behavior. We then develop a general equilibrium model of firm innovation in which inter-sectoral knowledge linkages determine the set of sectors a firm chooses to innovate in and how much R&D to invest in each sector. It captures how firms evolve in the technology space, accounts for cross-sector differences in R&D intensity, and describes an aggregate model of technological change. The model matches new observations as demonstrated by simulation. It also yields new insights regarding the mechanism through which sectoral fixed costs of R&D affect growth.

Keywords: Endogenous growth; R&D; Inter-sectoral knowledge spillovers; Firm innovation; Multiple sectors; Resource allocation

JEL Classification: O30, O31, O33, O40, O41

*Cai: Shanghai University of Finance and Economics, address: School of Economics, 111 Wuchuan Road, Yangpu, Shanghai, China 200433, Email: aprilcai@gmail.com. Li, International Monetary Fund, address: 700 19th Street NW, Washington DC 20431, email: nanli1@gmail.com. Acknowledgment: We thank Ufuk Akcigit, Paul Beaudry, Jess Benhabib, Chris Edmond, Sotirios Georganas, Joe Kaboski, Aubhik Khan, Sam Kortum, Amartya Lahiri, Roberto Samiengo, Julia Thomas, Shenghao Zhu, as well as seminar participants at Brown University, Georgetown University, London School of Economics, University of British Columbia, University of Melbourne, Ohio State University, the UNSW-University of Sydney joint macro seminar, NBER EFJK group meeting, Econometric Society Winter Meeting, ASSA, Reserve Bank of Australia, Chinese University of Hong Kong, York University and the IMF institute, for many helpful discussions and comments. Special thanks to Roberto Samaniego for sharing the firm-level R&D intensity data with us. The views expressed herein are those of the authors and should not be attributed to the IMF, its Executive Board, or its management.

1 Introduction

Innovation hardly ever takes place in isolation. Technologies depend upon one another, yet vary substantially in their applicability. Some innovations, such as the electric motor, create applicable knowledge that can be easily adapted to develop new products in a vast range of sectors; while others introduce knowledge that is limited in its scope of application. The *interconnections* between different technologies and the stark contrasts in their future impact have long been recognized by economic historians (e.g. David, 1991; Rosenberg, 1982; Landes, 1969). The majority of theoretical works on endogenous growth, however, tend to treat innovations in different technologies as isolated and equally influential.[1]

Empirical evidence based on patent citations suggests that knowledge spillovers vary substantially across sectors and are highly significant. More than half of patent citations are made between distinct technology categories, with some technologies contributing more knowledge to innovations in the entire economy than others.[2] In addition, inspecting the firm patenting data reveals the importance of *multi-sector firm innovations*: 42% of patenting firms innovate in more than one technological area, accounting for 96% of patents in the economy. These are the firms which are able to internalize knowledge spillovers across sectors.

The questions are: How do firms decide on what kinds of technologies to develop, and in which sectors to apply their existing knowledge and grow their business? How do technologies progress from one sector to another? And ultimately, what are the aggregate growth implications of technological diversification of firms? The efficacy of government policies directed at stimulating innovations in certain sectors hinges on a better understanding of the above questions. Addressing these questions requires a structural framework that integrates micro empirical evidence into a macro-growth model with important heterogeneities across firms and sectors.

This paper therefore pursues two goals. First, we document several novel observations that motivate our research. Technology interconnections are conceptual and difficult to measure. We handle this empirical challenge by first constructing a "technology network", which builds on the patent citation network linking the knowledge receiving and contributing sectors. We then propose a sector-specific measure of *technology applicability* using the method developed in the network lit-

[1] Notable exceptions include a body of work on General Purpose Technologies (GPTs) (e.g. Jovanovic and Rousseau, 2005; Helpman, 1998; Bresnahan and Trajtenberg, 1995). Differently from these studies, our paper focuses on the impact of technology linkages on firm innovation and aggregate growth. The associated notion of technology applicability is related to, but distinct from, the concept of generality of purpose of technologies.

[2] This is based on 428 technology classes (U.S. Patent Classification System) provided by U.S. Patent and Trade Office for the period 1976-2006. The share becomes even higher when using more disaggregated classifications. Previous empirical studies using other types of data also point to the importance of cross-sector knowledge spillovers. For example, using R&D investment data Bernstein and Nadiri (1988) find that knowledge spillovers across five high-tech industries are substantial and highly heterogeneous. The survey study by Wieser (2005) finds that spillovers between sectors are more important than those within sectors when evaluating both the social and private return of R&D.

erature.[3] The method establishes a particular hierarchy in the technology network that is amenable to empirical explorations. Combining this measure with firm R&D/patenting data, we document in Section 2 the following observations: (1) at the sector level technology applicability helps to explain the persistent variations in R&D intensity across sectors; (2) at the firm level more innovative firms—with larger patent stock and patent scope—concentrate more in highly applicable technologies; (3) as firms grow, they gradually enter less applicable, less connected technologies; and (4) firms with a larger share of highly applicable knowledge subsequently innovate faster.

The second objective is to develop a general equilibrium model of multi-sector firm innovation to explain these observations and to draw aggregate implications. The framework extends the leading growth models of firms R&D and patenting (such as Klette and Kortum, 2004) into a multi-sector environment. Relative to the existing studies, our framework emphasizes two new features: *heterogeneous intersectoral knowledge linkages* which affect firms' cross-sector R&D allocation; and *idiosyncratic fixed costs of innovation* which act as barriers to diversification and induce sequential entry of firms into different sectors.[4] Despite the multiple degrees of heterogeneity (at the sector-level, sector-pair-level and firm-level), the model is tractable and allows for closed-form equilibrium characterizations. The model captures how firms evolve in the technology space, and describes how knowledge accumulates in different sectors and in the aggregate economy. It relates growth to cross-sector knowledge circulation and R&D allocation, and yields new insight into the effects of barriers to diversification (sectoral fixed costs) on growth. When simulated using a large panel of firms innovating in different sectors, our model is able to reproduce each of the new facts above.

In the model, firms invent new products by adapting prior knowledge in various sectors through R&D. Applicable technologies enhance the innovational productivity of R&D and contribute to a sequence of innovations in many sectors. In adapting prior knowledge, firms can utilize their own private knowledge, public knowledge or obtain licenses to use other firms' private knowledge in various sectors which is subject to an absorption cost. The latter takes place in an efficient and competitive licensing market. Specifically, should any firm decide not to innovate in sector i in one period, it can—and finds it optimal in equilibrium to—license the application rights of its prior knowledge of sector j to other innovating firms in sector i during that period, assuming perfect intellectual property rights protection.

In order to conduct research in any given sector, a firm has to pay a period-by-period idiosyncratic fixed cost. The fixed cost of innovation leads to increasing return to knowledge capital, generating demand from innovating firms to acquire additional related knowledge in the licensing market. The equilibrium licensing fees that clear the market thus reflect the "application value"

[3]We focus on the "deep" knowledge linkages between technologies which are due to intrinsic characteristics of technologies and do not vary over time. In some sense, it takes the view of Nelson and Winter (1977) that "innovations follow 'natural trajectories' that have a technological or scientific rationale rather than being fine tuned to changes in demand and cost conditions." For this reason, we summarize citations made to (and from) patents that belong to the same technology class over thirty years to form the technology network.

[4]We note that throughout the paper, entry and exit refer to innovating or not in a particular sector.

of the source knowledge j in innovating in sector i. The existence of knowledge licensing market thus allows all knowledge to be utilized in equilibrium, either by its original inventor or by other firms that have acquired its application rights. Therefore, the equilibrium value associated with knowledge capital in sector j is no longer just determined by the profit it generates in its own sector as in conventional models, but also depends on its application value in all sectors. Higher application value attracts firms to invest in R&D in that sector. This explains why technology applicability helps to understand cross-sector differences in R&D intensity as documented in Section 2 (Observation 1).

The sectoral fixed costs also make research in multiple sectors a *self-selection* process: a firm develops new products in sectors where it can most efficiently utilize its existing range of knowledge. This explains the empirical observations that firms conducting research in multiple areas are more likely to concentrate in highly applicable technologies (Observation 2), because they are better at internalizing inter-sectoral knowledge spillovers and thus have stronger incentive to innovate in these sectors.

Although high applicability attracts firms to invest intensively in R&D in the "central" sectors, the model also suggests a counteracting force: the fierce competition in these sectors, as the composition of firms in different sectors is endogenous and ultimately determined by knowledge linkages. A firm would only conduct research in a sector if its knowledge is applicable enough to generate a larger expected value than the fixed cost. Therefore, as firms grow and accumulate more private knowledge in related sectors, they can afford to expand into "peripheral" technologies with lower applicability but allowing them larger market shares (Observation 3). The trade-off between *innovational applicability* and *product market competition*—which is at the heart of the R&D resource allocation mechanism in the economy—leads to a stable distribution of firms across sectors and a stable relative sector size on the balanced growth path.

Innovation by its nature is highly uncertain. In the model we assume that firms face two types of uncertainty every period: idiosyncratic risks to the success of R&D and idiosyncratic risks to its fixed costs of research in individual sectors. Therefore, although the underlying inter-sectoral linkages dictate that firms generally start from central sectors and gradually venture into periphery, not all firms follow the same sequence of sectoral entry. In any given sector, incumbents innovate, expanding their sizes as they create new varieties and knowledge, and pause or stop innovating after experiencing a sequence of adverse R&D shocks or high fixed costs. In addition, potential innovators enter if they have accumulated enough knowledge capital—either by creating its own knowledge or by acquiring external knowledge—in related sectors. This process *endogenously* generates a distribution of firm size in each sector, converging to a Pareto distribution in the upper tail, in line with existing empirical findings of firm size distribution.[5]

[5] Firm or establishment-level data show that firm size distributions within narrowly defined sectors and within the overall economy are widely dispersed and follow a Pareto distribution, as documented in Axtell (2001), Rossi-Hansberg and Wright (2007) and Luttmer (2007).

Not only a firm's R&D allocation across sectors but also its future growth is *path-dependent*. As the firm moves through the technology space, the scope and applicability of its knowledge change, and so do the opportunities to innovate, profit and grow in related sectors. The model predicts that conditional on the size and scope of knowledge stock firms with a larger share of applicable technologies tend to innovate faster, particularly by expanding into connected new sectors (Observation 4).

Lastly, at the aggregate level the model yields new insights regarding the mechanism through which sectoral fixed costs reduce growth in the presence of inter-sectoral knowledge linkages. As mentioned earlier, in the process of adapting the acquired external knowledge the firm faces an absorption cost such that only a fraction of the external knowledge is effectively utilized. Therefore, the market application value of any given knowledge is always lower than its internal application value. Higher absorption costs thus decrease the equilibrium value of knowledge and lower firms' incentive to invest in R&D. In addition, we assume that the higher the ratio of external knowledge to in-house knowledge, the lower the absorption rate. Therefore, raising sectoral fixed costs decreases the fraction of firms that innovate in multiple sectors and internalize cross-sector spillovers by themselves, increasing the external-to-own knowledge ratio in the economy. Consequently, less knowledge would be effectively absorbed and utilized in the economy, generating a negative "knowledge underutilization effect" on growth. Moreover, increasing the idiosyncratic uncertainty to the fixed costs leads to more randomness in allocation of R&D resources across sectors, as opposed to allocation according to fundamental knowledge linkages and firms' prior knowledge. This generates an additional negative "R&D misallocation effect" on growth.

Related Literature Our paper builds on Klette and Kortum (2004) (henceforth, KK) type of models, which connect growth theories with findings from firm-level and sectoral-level studies of innovation. In the past, most theoretical works on endogenous growth (e.g. Romer, 1986, 1990; Lucas, 1988 ; Segerstrom, Anant and Dinopoulos, 1990; Aghion and Howitt, 1992; Grossman and Helpman, 1991a, 1991b; and Jones 1995) and research on innovation and firm dynamics (e.g., KK; Luttmer, 2007, 2012; and Atkeson and Burstein, 2010) have not considered path-dependence in firm innovation behavior across multiple sectors, as these papers typically assume a single type of technological change or implicitly assume a homogeneous technology space in which innovation takes place in any sector with equal probability.

Empirical work by Jaffe (1986), on the other hand, suggests that firms' technological position provides different technological opportunities that matter for firms' innovative success. In that paper, however, firms' technology position is exogenous. Our study advances Jaffe's work by constructing a structural model which allows for the *endogenous sorting* of firms across technology classes, providing further understanding of the relationship between technological opportunities and firms' dynamic R&D decisions. Other empirical works by Bernard, Redding and Schott (2009,

2010) document that most firms switch their products frequently, and that endogenous product selection has important implications on firm and aggregate productivity. Obviously, our focus is entirely different: we examine firm innovation behavior instead of production performance. The more interesting difference is that the presence of inter-sectoral knowledge linkages fundamentally affect firms' R&D/patent allocation and their sectoral entry decisions.

Distinguishing between different types of research and their impact is currently being pursued in a number of papers. Akcigit, Hanley and Serrano-Velarde (2016) analyses the impact of appropriability on firms' incentives to conduct basic research relative to applied research. Akcigit and Kerr (2016) studies how exploration versus exploitation innovations affect growth. Akin to this notion, Acemoglu and Cao (2015) considers incremental R&D engaged in by incumbents and radical R&D undertaken by potential entrants. Different from these studies, we consider a richer structure of technological interdependence, and integrate it into the endogenous growth models.

Our work also builds on the earlier literature in development economics that emphasizes the role of sectoral linkages and complementarity in explaining growth (see Leontief, 1936 and Hirschman, 1958). Previous work in this area typically focuses on vertical input-output relationships in production between sectors—as in Jones (2011) and Bartelme and Gorodnichenko (2015), and export-based measures of product relatedness—as in Hidalgo, Klinger and Hausmann (2007) and Hausmann, Hwang and Rodrik (2007).

Finally, this paper also adds to previous works studying the determinants of persistent cross-sector differences in R&D intensity (e.g. Ngai and Samiengo, 2011; Klenow, 1996). Empirical evidence and the model developed in this paper both suggest that these differences can be attributed to technology applicability. We relegate the detailed discussions to Section 2.2.

The paper begins by presenting some new sector-level and firm-level findings which motivated our modeling approach. The model itself is developed and stationary balanced growth path equilibrium is characterized in Section 3. We then discuss firm, sectoral and aggregate implications generated by the model in Section 4. Section 5 discusses estimation and parameterization of the model, the ability of the model to replicate key observations and the results from counterfactual simulations. Section 6 concludes and discusses policy implications and future works.

2 Empirical Underpinning

This section starts by describing the algorithm for constructing our measure of "applicability". It then documents several novel empirical observations that motivate our model using patent citations, firm patenting and R&D investment data.

Data Description Our main data source is the 2006 edition U.S. Patent and Trade Office (USPTO) data from 1976 to 2006 (see Hall, Jaffe and Trajtenberg 2001 for detailed description of the data). We focus on firm patenting activities in this paper, as the model is designed to mainly understand firm innovation behavior. We observe the set of technological classes in which each firm applied for patent in each year and the citations associated with each patent application. Patent applications serve as proxies of firms' innovative output, and their citations are used to trace the direction and intensity of knowledge flows within and across technological classes.[6] In the dataset, each patent is assigned to one of the 428 three-digit United States Patent Classification System (USPCS) technological fields (NClass) and belongs to one to seven out of the 42 two-to-four-digit Standard Industrial Classification (SIC) categories.[7] The latter classification is used when we examine R&D at the sector level, because other sources of sector-specific characteristics are only available at the SIC level. Firm-level evidence, however, is reported based on more disaggregated NClass classification. Another data source is U.S. Compustat (1970-2000) which contains firm-level R&D expenditure and sales data associated with each sector. We use this information to obtain sector-specific R&D intensity.

2.1 The Measure of Technology Applicability

The Network of Inter-sectoral Knowledge Linkages We sum up patent citations connecting different technology classes to form the inter-sectoral knowledge diffusion network. Since we are interested in studying the deep, long-run characteristics between different technologies, we use patent citation data spanning the 1976-2006 period to form this network. Pooled citations for 30 years also help to smooth out noises in the annual data. We also test the sensitivity of our results to the use of time-variant knowledge linkages network based on rolling-window subsamples. The results, available in Appendix A.2, are robust to this alternative approach.

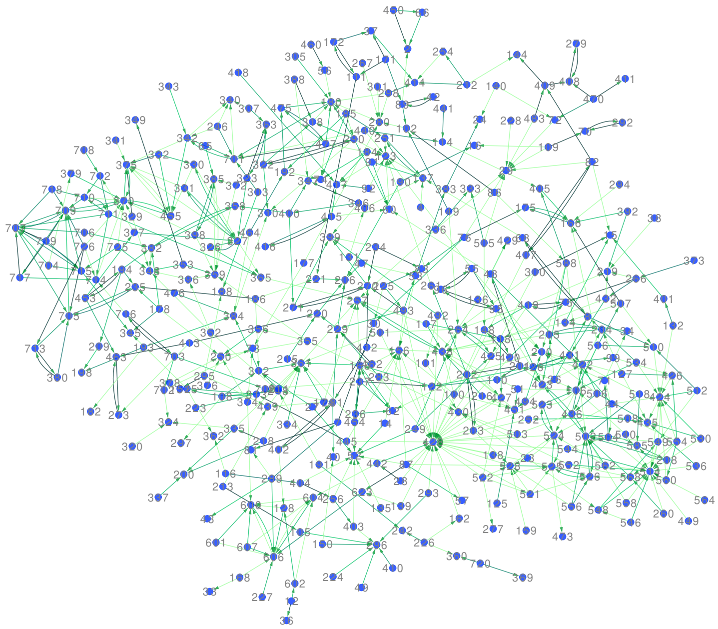
Figure I presents the network of inter-sectoral knowledge linkages, based on citations made between 428 3-digit technology classes. Each vertex corresponds to one type of technology, and every arrow indicates the direction of the knowledge flow. The darker color of the arrows signals a larger number of citations. The network exhibits strong heterogeneity in technology interconnections: not all technologies cite each other and some sectors are heavily cited while others are not. There are a few clusters of closely connected technologies, suggesting that they have a disproportionately important effect of knowledge spillovers.

Calculating Sector-Specific Technology Applicability The relationships of knowledge complementarity, especially the higher-order interconnections, make it difficult to evaluate the contribution of any innovation to the entire technology space. Hence, the first challenge is to construct

[6] Although patent statistics have been widely used in studies of firm innovations, not all innovations are patented, especially process innovations, which are often protected in other ways such as copyright, trademarks and secrecy (see Levin Klevorick, Nelson and Winter (1987)). Our measure implicitly assumes that for any sector, the unpatented and patented knowledge utilizes knowledge (patented or unpatented) from other sectors in the same manner, with the same likelihood and intensity.

[7] We use the probability mapping provided by USPTO to assign patents into different SIC categories. Details of the concordance are available at http://www.uspto.gov/web/offices/ac/ido/oeip/taf/data/sic_conc.

Figure I: Intersectoral Network Corresponding to Patent Citations between 428 Technology Classes



Data Source: NBER patent citation data, 428 technological categories (NClasses).
Notes: A (directed) link is drawn for every citation link that counts more than 5% of the total citations made by the citing sector.

such a sector-specific measure that characterizes the importance of different sectors as knowledge suppliers to their immediate application sectors as well as their role as indirect contributors to chains of downstream sectors.

To handle this issue, we apply Kleinberg's (1999) algorithm to the citation network and construct a measure quantifying the applicability of each technology. This algorithm generates two inter-dependent indices for each node in the network: the authority weight (aw^i)—the ability of contributing knowledge to the entire network; and the hub weight (hw^i)—the ability of absorbing knowledge. We use the authority weight as our measure of *technology applicability*, $app^i \equiv aw^i$.

Formally, let J be a set of technology categories. A citation matrix for J is a $|J|\times|J|$ nonnegative matrix $(c^{ji})_{(i,j)\in J\times J}$. For each $i, j \in J$, c^{ji} denotes the number of citations to sector i made by j (indicating knowledge flow from i to j). Then, the authority weight is calculated according to:

$$\begin{aligned} aw^i &= \lambda \sum_{j\in J} W^{ji} hw^j, \\ hw^i &= \mu \sum_{j\in J} W^{ij} aw^j, \end{aligned} \tag{1}$$

where λ and μ are the inverse of the Euclidean norms of vectors $(aw^i)_{i\in J}$ and $(hw^i)_{i\in J}$, respectively. W^{ji} denotes the weight of the link, corresponding to the strength of knowledge contribution by

i to j and is set to c^{ji}.[8] Intuitively, the technology with high authority weight provides large knowledge flows to sectors with highly ranked hub weights, and the technology with high hub weight largely utilizes knowledge flows from sectors with highly ranked authority weights. Kleinberg (1999) shows that this algorithm is more efficient at extracting information from a highly linked network environment compared to other quantitative indicators such as Garfield's "impact factor" and Pinski and Narin's "influence weight".[9]

A list of the ten most and ten least applicable technologies based on aw^i is provided in Table I. The ranking of technologies appears sensible. The ten least applicable technologies tend to be less sophisticated ones which have little application to innovations in other sectors. The technologies listed as the most applicable also seem reasonable.

Table I: The Ten Most and Ten Least Applicable Technologies (NClass-based)

Most applicable		Least applicable	
NClass	Technology description	NClass	Technology description
438	Semiconductor Device Manufacturing: Process	258	Railway Mail Delivery
257	Active Solid-State Devices	276	Typesetting
365	Static Information Storage and Retrieval	147	Coopering
361	Electricity: Electrical Systems and Devices	278	Land Vehicles: Animal Draft Appliances
428	Stock Material or Miscellaneous Articles	199	Type Casting
427	Coating Processes	314	Electric Lamp and Discharge Devices
430	Radiation Imagery Chemistry	79	Button Making
29	Metal Working	520	Synthetic Resins or Natural Rubbers
216	Etching a Substrate: Processes	295	Railway Wheels and Axles
324	Electricity: Measuring and Testing	231	Whips and Whip Apparatus

To distinguish our notion of knowledge applicability from other characterization of technologies—especially to emphasize the role of indirect knowledge linkages—we calculate the following measures for comparison. First, to differentiate the applicability across sectors from that within the sector, we construct a *self-applicability* measure using the number of citations received from the same sector per patent. Second, we consider an indicator that captures the importance of different sectors as a *direct* knowledge contributor: the *weighted (in)degree*, or *degree*. $degree^i \equiv \sum_j s^{ji}$, where the weight $s^{ji} (= c^{ji} / \sum_k c^{jk})$ is the fraction of citations made by j that is attributed to i.[10] Third,

[8] In the previous version of the paper, we also investigated results based on binomial weight: $W^{ji} = 1$ if j cites i and zero otherwise. That is, the weight is independent of the relative size between i and j. All the results still hold.

[9] Garfield's impact factor is the average number of citations received by a sector (pure in-degree counting), and hence is too crude a measure, as not all citations are equally important. Pinski and Narin's influence weight is a one-level iterative algorithm. The influence of i is the weighted sum of the influences of all sectors citing i. That is $w^i = \sum_j s^{ji} w^j$, where s^{ji} denotes the fraction of the citations from j that go to i. This method does not make a distinction between the importance as a contributor and the importance as a learner. Another advantage of Kleinberg's two-level pattern of linkages is that it exposes structure among both the set of hubs who may not know of one another's existence, and the set of authorities who may not wish to acknowledge each other's existence. Thus, it is more efficient at extracting information about the potential, as opposed to realized, knowledge contribution of each node.

[10] This measure is often applied to production Input-Output matrix (e.g. Acemoglu, Carvalho, Ozdaglar and Tahbaz-Salehi (2012). It is similar to Garfield (1972)'s "impact factor" or pure counting of the in-degrees of citations links, which only captures the sector's importance as knowledge supplier to its immediate application sectors.

we compare it to the *generality index* originally proposed by Trajtenberg, Jaffe and Henderson (1997), which in our context corresponds to $generality^i = 1 - \sum_j (\tilde{s}^{ji})^2$, where $\tilde{s}^{ji}(= c^{ji}/\sum_h c^{hi})$ is the fraction of citations made by j to i out of total citations received. Conceptually, generality captures a different notion from applicability. A sector directly cited by a wide range of sectors provides more general knowledge, but does not necessarily have a large overall knowledge impact, as the citing sectors themselves may not be important. Especially, Table II shows that our measures of applicability and generality are almost uncorrelated and even negatively correlated at the less disaggregated level (SIC). In addition, although the correlations between knowledge applicability and other indicators are positive, they are well below unity.

Table II: Correlations Between Applicability, Direct Spillovers, Generality and Self-Applicability

	NClass				SIC			
(all in log)	*applicability*	*degree*	*generality*	*self*	*applicability*	*degree*	*generality*	*self*
applicability	1				1			
degree	0.330**	1			0.549**	1		
generality	0.088	-0.121	1		-0.324*	-0.436**	1	
self	0.449**	0.643**	-0.506	1	0.670**	0.766**	-0.148	1

Notes: Correlation coefficients are reported. ** and * indicate significance at the 1 percent and 5 percent level, respectively.

2.2 Sector-level Observations

Observation 1: *Sectoral R&D intensity increases with its technology applicability.*

It has been documented previously in the literature that there are large and persistent cross-sector differences in R&D intensity. The literature has pointed to "technological opportunities" as one of the key explanations for these variations. Conceptually, technological opportunity reflects factors that allow research in some sectors to be more productive than others, such as sectoral future TFP growth as in Klenow (1996) or the opportunity in terms of knowledge spillovers from various sources as in Nelson (1988). Our measure of technology applicability provides a natural interpretation of Nelson's (1988) notion of technological opportunity and allows us to empirically investigate its relationship with sector-specific R&D intensity.

Table III reports determinants of long-run sectoral R&D intensity (averaged over 30 years for each SIC sector) based on different regression specifications. The sectoral R&D intensity is measured in three ways. In Column (1)–(4), sectoral R&D is measured by total R&D expenditure by all firms in a given sector divided by its sales value. Column (5) and (6) use the median ratio and the mean ratio of R&D expenditures to sales among firms in the same sector, respectively. All regressions control for sectoral market size (measured by sales) and profitability (measured by value of shipment, excluding material cost, divided by labor compensation). The former is motivated by prior empirical studies which suggest that a larger market size, indicating demand pull factor, creates an incentive for firms to invest in R&D.[11] Including profitability as a regressor is motivated by our theoretical model in Section 3. We also control for self-applicability in Column (2)-(6), future (scaled) TFP growth following Klenow (1996) in Column (3)-(6),[12] and cross-sector variations in direct knowledge spillovers (using the "degree" index) and knowledge "generality" in Column (4)-(6).

Table III: The Determinants of Sectoral R&D Intensity

	Sectoral R&D/ Sectoral sales				Median Intensity	Mean Intensity
	(1)	(2)	(3)	(4)	(5)	(6)
log *app*	0.199	0.289	0.296	0.292	0.461	0.377
	(0.048)**	(0.060)**	(0.060)**	(0.062)**	(0.123)**	(0.139)*
log $sales$	-0.059	-0.040	-0.033	-0.044	-0.294	-0.008
	(0.069)	(0.069)	(0.074)	(0.080)	(0.114)*	(0.121)
$profitability$	0.051	0.059	0.058	0.050	0.242	0.243
	(0.041)	(0.041)	(0.041)	(0.039)	(0.090)*	(0.068)**
log $self$-app		-0.085	-0.088	-0.048	-0.107	0.087
		(0.035)*	(0.035)*	(0.052)	(0.085)	(0.094)
$\Delta(scaled)TFP$			-0.014	-0.022	0.101	0.065
			(0.033)	(0.022)	(0.114)	(0.091)
log $degree$				-0.367	0.194	0.111
				(0.392)	(1.123)	(0.931)
log $generality$				-0.922	0.215	-1.633
				(0.524)	(1.477)	(1.395)
No. of observations	42	42	42	42	42	42
R^2	0.30	0.36	0.36	0.41	0.43	0.61

Notes: The dependent variables are sectoral R&D expenditure divided by sectoral sales, or median R&D intensity (RI) or mean R&D intensity among firms in the same sector, taking average over 1970-2000. Regression coefficients are reported, with robust standard errors in brackets. ** and * indicate significance at the 1 percent and 5 percent level, respectively. The constant terms are omitted to save space.

Across all specifications, technology applicability has a statistically significant positive association with R&D intensity across sectors, even when allowing for other technology characteristics to play a role simultaneously. Self-applicability, whenever significant, in fact is negatively associated with R&D intensity. In addition, similarly to previous studies, sales and research productivity

[11]The previous literature using survey data (e.g. Cohen, Levin and Mowery, 1987) also suggests that appropriability (the extent to which R&D benefits the inventor) might play a role in understanding cross-sector variations in R&D intensity. However, as pointed out by Ngai and Samaniego (2011) the particular survey question was designed in a way that cannot distinguish appropriability from opportunity. They also find that appropriability does not vary much across sectors, and hence cannot explain the persistent differences in sectoral R&D.

[12]The TFP growth two years ahead is scaled by the average R&D intensity. The sector-specific profitability and TFP data are constructed using NBER-CES Manufacturing Industry Database. We first map all 4-digit SIC87 industries in the dataset into 4-digit SIC72 industries using the concordance provided by the database. The 4-digit SIC72 industries are then mapped into 42 technology fields using the concordance provided by USPTO. NBER-CES manufacturing industry database provides information on value of shipment, payroll, employment, material cost, total factor productivity for each individual manufacturing sectors, which can be used to construct profitability and TFP for the more aggregated 42 sectors. We consider both average profitability for the current period and average profitability two years ahead. The results are virtually the same.

(TFP growth scaled by R&D intensity) are not significantly related to R&D intensity at the sector level. However, at the firm level, median firm R&D intensity is found to increase with the sector-specific profitability but decrease with the sector's market size (Column (5)), perhaps reflecting the negative impact of within-sector competition on individual firm's R&D.

These results suggest that overall knowledge spillovers, not just the direct spillovers, to downstream knowledge application sectors matter for understanding the cross-sector variations in R&D intensity. Forward-looking innovating firms allocate their R&D resources not only according to profitability in their current sectors but also the potential applicability of the knowledge in fostering future innovations in other sectors. Section 3 develops a model to conceptualize this intuition.

2.3 Firm-level Observations

In the dataset, at any given period t, each firm is identified by its history of patent applications, $\{(P^1_{f,\tau}, P^2_{f,\tau}, ..., P^{428}_{f,\tau})\}_{\tau=1,2,...,t}$, where $P^i_{f,\tau}$ is the number of patents firm f applied for in period τ in technology class i. Let $S^i_{f,t}$ denote firm f's patent stock in t. For simplicity, we assume that there is no physical depreciation of knowledge.[13] Hence, $S^i_{f,t} = S^i_{f,t-1} + P^i_{f,t}$, and its total patent stock is $S_{f,t} = \sum_{i \in J} S^i_{f,t}$. To measure a firm's multi-technology patenting (or knowledge scope), we count the number of distinct technology classes in which firm has patented and denote it by $N_{f,t}$. We find that firms with larger patent stock also tend to innovate in a wider range of technology classes, with the correlation between $S_{f,t}$ and $N_{f,t}$ greater than 95 percent for most years.

In order to characterize the applicability of a firm's knowledge, it is convenient to first define the firm's technological position by the *distribution* of the firm's patents over all patent classes, as in Jaffe (1986). Let vector $T_{f,t} = (T^1_{f,t}, T^2_{f,t}, ..., T^{428}_{f,t})$, where $T^i_{f,t} = S^i_{f,t}/S_{f,t}$, stand for firm f's "technological position" in t. A firm's overall technology applicability measure, TA_f, is then calculated as the (weighted) average applicability of its technologies: $TA_{f,t} = \sum_{i \in J} T^i_{f,t} \log(app^i)$. Thus, a firm's knowledge applicability is constructed independent of its knowledge stock. Similarly, the applicability of firm f's new technology classes—the new sectors that the firm entered in t—is calculated as $TA^{newsec}_{f,t} = \sum_{i \in J} \frac{P^{i,newsec}_{f,t}}{P^{newsec}_{f,t}} \log(app^i)$, where the superscript "newsec" signals that sector i is new to firm f at t. Using all these firm-level measures, we then document observations as follows.

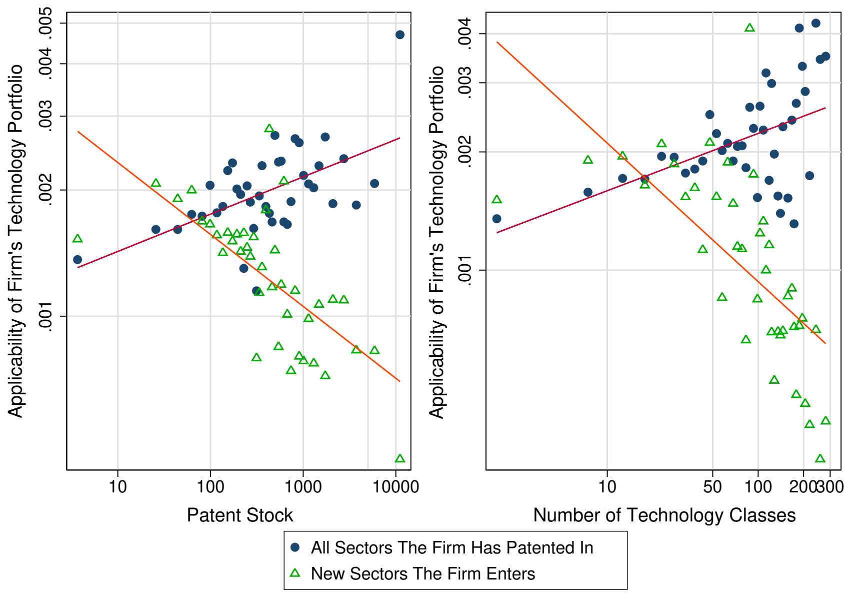
Observation 2 (Sectoral Composition): *Firms with more patents (or more technological classes) are more concentrated in highly applicable technologies.*

[13] Note that knowledge capital is different from R&D capital, which can literally depreciate over time as research labs are physical investment. For knowledge capital to depreciate, it means some idea is lost. In the literature there is a distinction between physical depreciation and economic depreciation of knowledge capital. Here we assume no physical depreciation, but make no assumption about economic depreciation. As shown in the Model section, knowledge capital in fact depreciates economically when newer knowledge accumulates in the same sector, and the depreciation rate is endogenous.

Observation 3 (Sectoral Entry): *As firms accumulate more patent in more technological glasses, they gradually enter sectors with lower technology applicability.*

Figure II illustrates the scale dependence in firms' patent allocation and entry pattern using the year with the highest number of firms (1997) as an example year. Results are similar in other years. All firms are divided into 40 bins according to their patent stocks (left panel) or their numbers of technology classes (right panel). The average firm in each bin constitutes one observation. The left panel plots firms' technology applicability, TA_f, against their patent stock, S_f, distinguishing the applicability of new sectors the firm entered in 1997, TA_f^{newsec} (the hollow triangles with the downward sloping fitted line) from its overall applicability (the solid dots the upward sloping line).[14] The right panel plots firms' technology applicability against numbers of technology classes in which the firms are engaged in patenting, N_f.

Figure II: Firm's Technology Applicability, Patent Stock and Multi-Technology Patenting



Notes: Y-axis measures the (weighted) average applicability of the firm's patent portfolio, TA_f. Firms are divided into 40 bins according to their patent stocks (left panel) or their numbers of technology classes (right panel). Each observation corresponds to an average firm in the same size bin. Both x- and y-axes are in log scale. The underlying sectors correspond to the Nclass technology fields categorized by USPTO. Data source: NBER Patent Data, 2006 edition.

Two observations stand out. First, firms with more knowledge capital (left panel) or broader knowledge scope (right panel) tend to innovate more in highly applicable technologies. This observation, however, is sharply reversed when focusing on the new technology classes firms just entered: TA_f^{newsec} is negatively related to both patent stock and the number of classes. Second, across firms of various sizes, the new sectors entered by a given firm tend to be less applicable relative to the

[14] A sector is new to a firm if the firm has not innovated in that sector before. The full data set expands from 1901 to 2006, thus, providing a good sample for identifying new sectors for each individual firm.

existing sectors (i.e. the observations that identify new sectors lie below the observations of all sectors), except for the very small firms.

Next, using firm-year observations, we explore how a firm's technology applicability is related to its knowledge stock (S_f) and scope (N_f) based on fixed-effects panel regressions. The dependent variable is TA_{ft} in Column (1) and (2) of Table IV, and $TA_{f,t}^{newsec}$ in Column (3) and (4).[15] The full set of year dummies is included to control for the level and change of any year-specific characteristic that influences the applicability of firm's technology. Firm-fixed effects control for any constant firm-specific characters. Both fixed effects deal with unobserved heterogeneity and error terms are allowed to be heteroskedastic and serially correlated.

As shown in Table IV, firms' technological position and sectoral entry are systematically related to their knowledge stock and scope. When firms become larger and have more knowledge in more areas, they become increasingly concentrated in highly applicable technologies. At the same time, this allows them to enter less occupied, less applicable technology classes.

Table IV: Firm's Patent Allocation, Knowledge Stock and Knowledge Applicability

	$TA_{f,t}$		$TA_{f,t}^{newsec}$	
	(1)	(2)	(3)	(4)
$\log(S_{f,t-1})$	0.019		-0.225	
	(0.004)**		(0.003)**	
$\log(N_{f,t-1})$		0.040		-0.330
		(0.006)**		(0.004)**
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
No. of obs	848593	848593	382968	382968

Notes: The dependent variables are the applicability of the firm's existing technology portfolio at time t for Column (1) and (2) and the applicability of the new sectors the firm entered at time t for Column (3) and (4). Regressions include firm and year fixed effects. Regression coefficients are reported, with robust standard errors adjusted for clustering by firms in brackets. Sample covers every year between 1976 and 2006. ** indicates significance at the 1 percent level. The constant terms are omitted to save space.

To further investigate how firms expand across different technology classes over time, we zero in on the new patent applications firms filed in each period. Note that the new patent applications are not necessarily in new technology classes. We adopt the following regressions using firm-sector-year observations, controlling for firm-fixed effects (η_f) and year-fixed effects (μ_t):

$$\log(app_{f,t}^i) = \beta_1 Newsec_{f,t}^i + \beta_2 \log(S_{f,t-1}) + \beta_3 Newsec_{f,t}^i \times \log(S_{f,t-1}) + \eta_f + \mu_t + \upsilon_{f,t}^i, \qquad (2)$$

where $app_{f,t}^i$ is the applicability of technology i in which firm f filed at least one patent at time t, and $Newsec_{f,t}^i$ is a dummy indicating that i is new to the firm at time t.

Table V shows evidence that is consistent with the previous firm-level observations. Column

[15] Since these two variables are highly correlated (correlation equals 0.93), we cannot include them in the same regression as that will cause multicollinearity issue.

(1) shows results based on Equation (2) while Column (2) substitutes $S_{f,t-1}$ with $N_{f,t-1}$. $\beta_2 > 0$ for both cases implying that in the sectors that a firm has previously entered, it tends to innovate more in the highly applicable technology classes as it grows larger. This is because the firm can now internalize this highly applicable knowledge in more sectors and thus has more incentive to do so. However, when it grows larger, the new technologies that a firm enters are farther away from the centre of the technology space than its existing technologies ($\beta_3 < 0$). As a firm accumulates more knowledge capital and in more categories, it can now apply this knowledge to enter sectors which are less connected with its existing knowledge portfolio, and enjoy less competition and higher market share.

Table V: Firm's Sectoral Entry Selection, Knowledge Stock and Knowledge Scope

Dependent Variable: $\log(app^i_{f,t})$	(1)	(2)
$Newsec^i_{f,t}$	0.087	0.103
	(0.006)**	(0.006)**
$\log(S_{f,t-1})$	0.015	
	(0.006)**	
$Newsec^i_{f,t} \times \log(S_{f,t-1})$	-0.137	
	(0.005)**	
$\log(N_{f,t-1})$		0.080
		(0.009)**
$Newsec^i_{f,t} \times \log(N_{f,t-1})$		-0.208
		(0.007)**
Firm FE	Yes	Yes
Year FE	Yes	Yes
No. of obs	995,244	995,244

Notes: The dependent variables are the natural log of applicability of the technology class in which the firm applied for patent at time t. Regressions include firm and year fixed effects. Regression coefficients are reported, with robust standard errors adjusted for clustering by firms in brackets. Sample covers every year between 1976 and 2006. ** and * indicate significance at the 1 percent level and 5 percent level respectively.

Observation 4 (Innovation Rate): *Controlling for the initial patent stock and patent scope, firms whose initial technologies are more applicable innovate faster.*

We adopt a firm growth regression by regressing firms' subsequent innovation rate on their previous knowledge applicability and patent stock, controlling for firm-fixed effects and year-fixed effect:[16]

$$g_{f,t} = \gamma_1 \log(S_{f,t-1}) + \gamma_2 \log(N_{f,t-1}) + \gamma_3 TA_{f,t-1} + \eta_f + \mu_t + \upsilon_{f,t}, \tag{3}$$

where the outcome variable innovation rate, $g_{f,t} = P_{f,t}/S_{f,t-1}$, is firm f's number of patent applications in t as a percentage of its previous patent stock. Furthermore, we differentiate a firm's

[16] We also investigate quality-adjusted innovation rates, which are measured by the growth rates of the forward-citation-weighted number of patents. When adjusted by the number of inward citations, the results are largely unchanged although larger firms' growth rates drop even faster.

growth in its existing sectors from its growth into new sectors. Define the innovation rate g_t^{in} as the intensive innovation rate as a result of patent applications in existing classes, and g_t^{ex} as the extensive innovation rate associated with patent applications in new technological classes. That is, $g_t^{ext} = P_{f,t}^{Newsec}/S_{f,t-1}$, $g_t^{int} = (P_{f,t} - P_{f,t}^{Newsec})/S_{f,t-1}$. In addition, innovation usually takes several years to occur. Hence there are often large time gaps between a firm's current patent application and its next one in the data. Therefore, $g_{f,t}$ is set to 0 during the years when firms did not apply for patent, and the results based on this are presented in Column (1)-(3). However, when we do not observe firm patenting, we have no information whether the firm has exited. Therefore, as an alternative method, we apply Heckman two-step processor to our regression to correct for selection bias, using firm's age as an instrument of exclusion restriction (Column (4)).

As shown in Table VI, the positive coefficients on the term $TA_{f,t-1}$ across all specifications indicate that firms whose initial technology applicability is greater, innovate faster subsequently, after controlling for knowledge stock and knowledge scope. Although not the focus of our paper, the result also shows that firms with larger initial knowledge stock tend to experience lower innovation rate in subsequent periods (i.e. the coefficient on $\log(S_{f,t-1})$ is negative). This could reflect the decreasing return of learning to scale: The more private knowledge a firm accumulates, the less is there to learn from others in relative terms. Broader scope of knowledge, on the other hand, allows firms to innovate faster, again pointing to the importance of inter-sectoral knowledge spillovers.

Table VI: Firm Innovation Rate, Knowledge Applicability, Stock and Scope

	(1)	(2)	(3)	(4)		(5)	(6)	(7)
	overall	intensive	extensive	Heckman Selection		Including self-applicability		
	g	g^{int}	g^{ext}	Main	Selection	g	g^{int}	g^{ext}
$\boldsymbol{TA_{f,t-1}}$	**0.028**	**0.008**	**0.020**	**0.081**	**0.051**	**0.025**	**−0.000**	**0.025**
	(0.007)**	(0.003)*	(0.006)**	(0.002)**	(0.003)**	(0.007)**	(0.003)	(0.006)**
$\log(S_{f,t-1})$	-1.173	-0.824	-0.349	-0.278	1.178	-1.175	-0.829	-0.346
	(0.013)**	(0.009)**	(0.008)**	(0.006)**	(0.009)**	(0.013)**	(0.009)**	(0.008)**
$\log(N_{f,t-1})$	0.144	0.789	-0.646	0.120	0.201	0.146	0.796	-0.650
	(0.014)**	(0.010)**	(0.010)**	(0.007)**	(0.011)**	(0.015)**	(0.010)**	(0.010)**
age					-0.042			
					(0.000)**			
$SA_{f,t-1}$						0.085	0.224	-0.139
						(0.033)*	(0.016)**	(0.027)**
Year FE	Yes	Yes	Yes	Yes		Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes		Yes	Yes	Yes
No. of obs	533,740	533,740	533,740	533,740		533,740	533,740	533,740

Notes: The dependent variables are the innovation rate ($g_{f,t}$) for Column (1), (4) and (5), and the decomposition of the innovation growth rate in the existing sectors (g_{ft}^{in}) for Column (2) and (6), and innovation rate in the new sectors (g_{ft}^{ex}) for Column (3) and (7). Regressions include firm and year fixed effects. Regression coefficients are reported, with robust standard errors adjusted for clustering by firms in brackets. Sample covers every year between 1976 and 2006. ** and * indicate significance at the 1 percent level and 5 percent level respectively.

In addition, both intensive and extensive firm innovation rates increase with firm's initial knowledge applicability, although the effect is larger on the extensive margin. This suggests that a central position on the technology space promotes firm innovation mainly through providing prerequisite knowledge while the firm expands into new sectors. Not surprisingly, while firm's scope enhances intensive innovation rate, it discourages extensive innovation. When inspecting the first stage of selection estimation of the Heckman procedure, higher knowledge applicability, larger knowledge stock and scope all increase the firm's survival probability, whereas firms' age significantly decreases the probability. In addition, Columns (5)-(7) show that it is the self-applicability that explains future innovation growth on the intensive margin, and once it is controlled for, TA_f no longer plays a significant role in explaining future intensive growth. However, self-applicability plays a negative role in predicting firms' future innovation on the extensive margin, as firms have less incentive to expand across the technology space when its knowledge can be easily applied to develop new products within the same sector.

Discussions We have presented a set of new observations which points to the importance of inter-sectoral knowledge linkages in understanding cross-sector differences in R&D intensity and firms' multi-technology innovation decisions. In particular, we establish that indirect higher-order knowledge linkages with other sectors matter and help to direct firms' R&D allocation—in addition to within-sector applicability and direct knowledge spillover to immediate downstream sectors. In Appendix A.2 we show that all the findings are robust to alternative measures of knowledge applicability (e.g. time-variant technology applicability or quality-adjusted measure of applicability) or allowing firm's patent stock to depreciate.

Existing theories without multiple sectors and/or knowledge interconnections between sectors cannot explain the observed relationship between firms' innovational activities (including sectoral entry, R&D allocation and innovation rate) and technology applicability/knowledge linkages. Motivated by these reduced-form analyses, in the following section we develop a general equilibrium multi-sector framework that helps us to interpret these observations.

3 The Model

Our model extends the previous literature on firm innovation and growth (especially, Klette and Kortum (2004)) to a multi-sector environment and is built on the tradition of variety expanding models (e.g., Romer 1990; Grossman and Helpman 1991a; Jones 1995).[17] The novel element is that sectors are connected by their knowledge linkages. It regards innovation as a process of generating new varieties in different sectors by applying existing knowledge in *all* related sectors. The existing knowledge includes both in-house knowledge, public knowledge and the external knowledge obtained from the licensing market which will be specified later. Therefore, allowing for external knowledge

[17] Recently, Balasubramanian and Sivadasan (2011) provides strong empirical evidence showing that firm patenting is associated with firm growth through the introduction of new products.

in the process of innovation is another notable difference from KK. In linking the model to the data, we interpret our *sector* as corresponding to different *technology classes* in the patent data, while *varieties/blueprints* within a sector map into *patents* granted in the respective technological class.[18]

We present the model in steps starting with goods demand and firm's static production decision, which follow the standard setup in the variety-expanding literature. We then introduce the dynamic multi-sector R&D, entry/exit decisions of firms which constitutes the main departure of our model from the existing literature. Industry behavior (including firm size distribution, mass of firms and R&D allocation across sectors) is then analyzed and the model is solved for aggregate implications.

We are interested in the long-run properties of our model and thus will focus on the stationary Balanced Growth Path Equilibrium (hereafter BGP) in which output, consumption and innovation grow at constant rates, and firm size distribution is stationary. The only source of uncertainty in the model are firm-sector-specific shocks to the success of R&D and to the fixed costs of research, and there are no shocks to goods production or shocks at the sectoral/economy-wide level.

3.1 Goods Demand and Production

Demand The economy is populated by a unit measure of identical infinitely-lived households. Households order their preferences over a lifetime stream of consumption $\{C_t\}$ of the final good according to

$$U = \sum_{t=0}^{\infty} \beta^t \frac{C_t^{1-\eta}}{1-\eta}, \tag{4}$$

where β is the discount factor and η is the risk-aversion coefficient. A typical household inelastically supplies a fixed unit of labor, L, which is allocated to produce goods, to conduct research or to maintain research labs (fixed costs of research). Households have access to a one-period risk-free bond with interest rate r_t and in zero aggregate supply. Optimal inter-temporal substitution of consumption implies

$$\beta(\frac{C_{t+1}}{C_t})^{-\eta} \frac{P_t}{P_{t+1}}(1+r_t) = 1. \tag{5}$$

The final good is produced by combining all types of sectoral intermediate goods $\{Q_t^i\}$ according to a Cobb-Douglas production function

$$\log Y_t = \sum_{i \in \mathcal{J}} s^i \log\left(Q_t^i\right), \tag{6}$$

where s^i captures the share of each sector in production of the final good. Without physical capital

[18] We refer to the terms technologies, technology classes and sectors interchangeably in the paper, as in the model one sector embodies one specific type of technology. Although distinguishing technology classes from industry classes can be interesting for certain issues (e.g. Bloom, Schankerman and Van Reenen (2010), it is not the focus of this paper.

in this closed-economy, the final good is only used for consumption: $C_t = Y_t$. Let $\mathcal{J}$ be the set of all sectors and K be the total number of sectors, i.e. $|\mathcal{J}| = K$.

At any moment, any sector $i \in \mathcal{J}$ contains a set of varieties that were invented before time t, indexed by $k \in [0, n_t^i]$, where n_t^i is the number (measure) of differentiated goods that are produced by individual monopolistically competitive firms.

$$Q_t^i = \left[\int_0^{n_t^i} \left(x_{k,t}^i \right)^{\frac{\sigma^i - 1}{\sigma^i}} dk \right]^{\frac{\sigma^i}{\sigma^i - 1}}, \quad \forall i \in \mathcal{J}, \tag{7}$$

where $x_{k,t}^i$ is the consumption of variety k in sector i and $\sigma^i > 1$ is the elasticity of substitution between differentiated goods in the same sector i.

The associated final good price is $P_t = B \prod_{i \in \mathcal{J}} (P_t^i)^{s^i}$, where B is some constant consistent with the Cobb-Douglas specification in (6) and the sectoral price index is given by $P_t^i = \left[\int_0^{n_t^i} p_{k,t}^{1-\sigma^i} dk \right]^{\frac{1}{1-\sigma^i}}$. These aggregates can then be used to derive the optimal consumption for sector-i goods and for individual variety k in sector i using $x_{k,t}^i = \left(\frac{p_{k,t}^i}{P_t^i} \right)^{-\sigma^i} Q_t^i$, where $Q_t^i = s^i \frac{P_t Y_t}{P_t^i}$.

Production Firms undertake two distinct activities: they create blueprints for new varieties of differentiated products and manufacture the products that have been invented. The firm inventing a new variety is the sole supplier of that variety. We assume that each differentiated good is manufactured according to a common technology: to produce one unit of any variety requires one unit of labor.

Without heterogeneity in production and demand, all varieties in the same sector are completely symmetric: they charge the same price and are sold in the same quantity. The firm producing variety k in sector i faces a residual demand curve with constant elasticity σ^i.[19] Wage is normalized to one: $w_t = 1$. This yields a constant pricing rule:

$$p_{k,t}^i = \frac{\sigma^i}{\sigma^i - 1}, \quad \forall k, i, t. \tag{8}$$

Thus the sectoral price, $P_t^i = \frac{\sigma^i}{\sigma^i - 1} (n_t^i)^{\frac{1}{1-\sigma}}$, decreases with the number of varieties in that sector.

Combining the pricing rule with the demand equation for individual variety, we derive the *total* profit in sector i from production as a constant share of GDP. As will be clear later, without population growth, the nominal GDP is constant: $P_t Y_t = PY$. Thus, sectoral profit is constant:

$$\pi_t^i = \int_0^{n_t^i} \frac{p_{k,t}^i x_{k,t}^i}{\sigma^i} dk = \frac{s^i PY}{\sigma^i}. \tag{9}$$

[19]To make the analysis more tractable, we follow Hopenhayn (1992) and Klette and Kortum (2004) by assuming that each firm is relatively small compared to the entire sector.

The demand for production labor in sector i is

$$L_{p,t}^{i} = \int_{0}^{n^{i}} x_{k,t}^{i} dk = \frac{\sigma^{i} - 1}{\sigma^{i}} s^{i} PY. \tag{10}$$

3.2 Innovation

3.2.1 Knowledge Creation

There is a continuum of firms, each developing new varieties and producing in a set of sectors. A firm f at time t is defined by a vector of its blueprints in all sectors, $\boldsymbol{z}_{f,t} = (z_{f,t}^{i})_{i \in \mathcal{J}}$, where $z_{f,t}^{i} \geq 0$ is the number of blueprints (or the amount of knowledge capital) of sector-i goods produced by firm f at time t. Let $\mathcal{S}_{f,t} = \{i : s.t. \;\; z_{f,t}^{i} > 0\} \subseteq \mathcal{J}$ denote the set of sectors in which firm f produces at time t and $\mathcal{F}_{t}^{i} = \{f : s.t. \;\; z_{f,t}^{i} > 0\}$ denote the set of firms that produce in sector i. Then the total number of varieties in sector i, $n_{t}^{i} = \int_{f \in \mathcal{F}_{t}^{i}} z_{f,t}^{i} df$.

Firm f's knowledge capital in i accumulates according to

$$z_{f,t+1}^{i} = z_{f,t}^{i} + \Delta z_{f,t}^{i}, \qquad i \in \mathcal{S}_{f,t+1}, \tag{11}$$

where the new knowledge, $\Delta z_{f,t}^{i}$, is created by conducting R&D to adapt prior related knowledge in *all* sectors. The prior knowledge of sector j comprises private knowledge and public knowledge. For convenience we call the innovative activities associated with the former "invention" and the activity associated with the latter "imitation". Since knowledge linkages are heterogeneous across sectors, we index firm's R&D investment by its *knowledge source sector* and *knowledge application sector*. Specifically, R_{f}^{ij} denotes a firm's investment in R&D associated with applying its knowledge in sector j (source sector) to innovation in i (application sector). The productivity of this R&D activity depends crucially on the knowledge applicability from j to i, A^{ij}. In the process of creating knowledge in various sectors, firms take the knowledge diffusion matrix, $\mathcal{A} = [A^{ij}]_{(i,j) \in \mathcal{J} \times \mathcal{J}}$, as exogenous.[20]

Formally, new knowledge in sector i is created based on the knowledge creation function:[21]

$$\Delta z_{f,t}^{i} = \sum_{j \in \mathcal{J}} \left[\underbrace{A^{ij} \left(\bar{z}_{t}^{i} R_{v,ft}^{ij} \right)^{\alpha} \left(z_{f,t}^{j} + \kappa(l_{f,t}^{ij}/z_{f,t}^{i}) l_{f,t}^{ij} \right)^{1-\alpha} \varepsilon_{f,t}^{ij}}_{\text{Invention}} + \underbrace{A^{ij} \left(\bar{z}_{t}^{i} R_{m,ft}^{ij} \right)^{\alpha} \left(\theta \bar{z}_{t}^{j} \right)^{1-\alpha}}_{\text{Imitation}} \right] \tag{12}$$

[20] It might be true that as technologies advance over time the interactions between them evolve, forming a dynamic network instead of a static one. Also, these relationships of complementarity may be hard to predict and not necessarily visible or well understood by innovators. Here, we intentionally choose to concentrate on the implications of "deep", time-invariant characteristics of technological linkages on firm's innovation and leave the study of dynamic knowledge network formation to future work, as we view the former as a necessary first step.

[21] The advantage of using additive instead of multiplicative function to combine the blueprints created using different source knowledge is that the separability of additive function allows for linear function of firms' value, which makes the model more tractable. In addition, it allows for Pareto firm size distribution in each sector (as shown in Sections 3.2.3) and B.3.

where $R^{ij}_{v,ft}$ is the number of researchers adapting private knowledge to *invent*, while $R^{ij}_{m,ft}$ is the number of researchers adopting public knowledge to *imitate*. There are two kinds of private knowledge: own knowledge ($z^{j}_{f,t}$) and external knowledge acquired from other firms ($l^{ij}_{f,t}$) which can be used for only one period. We will specify the acquisition of external knowledge later. The absorption capacity is given by the function $\kappa(\cdot) \in (0,1)$, which governs the fraction of acquired knowledge from other firms that is ultimately absorbed and utilized by the licensee. $\bar{z}^{i}_{t} R^{ij}_{ft}$ is the effective R&D input with $\bar{z}^{i}_{t} = n^{i}_{t}/M^{i}_{t}$ being the average knowledge capital per firm in sector i and M^{i}_{t} being the number of innovating firms in sector i at time t. α denotes the share of effective R&D input. θ governs the adaptability of the public knowledge relative to the own private knowledge. $\varepsilon^{ij}_{f,t}$ is the shock to invention.

We explain various features of (12) in detail as follows. First, similarly to KK, we assume that the knowledge creation function is constant returns to scale. In addition, the researchers' efficiency is proportional to the average knowledge capital per firm in the innovating sector, $\bar{z}^{i}_{t}$. This assumption keeps the number of R&D workers constant in the BGP equilibrium while the number of varieties increases. As shown in Section 4.3 and discussed in Peretto (1998), it removes the "scale effect" from the model—that is, the endogenous growth rate of the economy is independent of its population size.

Second, in the process of developing new knowledge in sector i, a firm utilizes all existing knowledge available—its own knowledge, acquired external knowledge and public knowledge from all sectors. The absorption of either acquired knowledge or public knowledge requires R&D input (in the spirit of Cohen and Levinthal (1990)), as opposed to other models where learning from others might be effortless. Importantly, we assume $\kappa(\cdot)$ is a decreasing function of external-to-own knowledge ratio $(l^{ij}_{ft}/z^{i}_{f,t})$: $\kappa'(l^{ij}_{f,t}/z^{i}_{f,t}) < 0$. That is, more own knowledge compared to external knowledge allows the "learner" to internalize the knowledge more effectively. We will show later that this assumption is necessary for the model to generate endogenous sorting of firms in different sectors.

Third, the size of the public knowledge pool is assumed to be proportional to $\bar{z}^{j}_{t}$. As firms randomly meet and exchange ideas with a limited number of peers, the average knowledge capital is a reasonable proxy for the size of the accessible public knowledge.[22] As will be clear later, imitation is allowed in the model such that new firms with no prior knowledge of any sorts can imitate to enter. The different value of public knowledge across sectors also helps to explain sequential sectoral entry. In addition, imitation helps to mitigate the dispersion of firm size distribution by preventing firms from becoming too small.

Lastly, innovation by its nature involves the discovery of the unknown and the success of a research project can be uncertain. We assume that invention is subject to idiosyncratic shocks $\varepsilon^{ij}_{f,t}$

[22] The similar assumption can be found in the knowledge diffusion literature, such as Monge-Naranjo (2012), Alverez, Buera and Lucas (2013). This assumption also helps to ensure that on BGP, the average knowledge capital per firm is a constant and the growth rate is independent of the number of firms and the total population.

that are *i.i.d.* across firms, sector-pairs and time. Its cumulative density function is given by $G(\varepsilon)$ with bounded support over $(0, \bar{\varepsilon}]$, $E(\varepsilon_{f,t}^{ij}) = 1$ and variance σ_{ε}^2.[23] Firms know the distribution of shocks but not their actual realizations before deciding on the optimal R&D input. A series of large adverse shocks leads to downsizing and a series of favorable ones causes further expansion.

3.2.2 Firm's R&D, Sectoral Entry and Exit Decisions

In order to develop blueprints in any sector $i \in \mathcal{J}$, the firm f must pay a per-period sector-specific fixed cost of $\zeta_{f,t}^i > 0$, measured in units of labor. The firm-sector specific idiosyncratic component, $\zeta_{f,t}^i \sim i.i.d.\ H^i(\zeta)$ with support over $(0, \infty)$, mean F^i and variance σ_{ζ}^i. This sector-specific cost of innovation can be interpreted as legal barrier or the cost of maintaining a research lab. Since firms have to pay the cost in every sector they innovate, the fixed costs act as barriers to diversification in the model.

Knowledge Licensing Market To price the knowledge, we assume that there exists an efficient and competitive licensing market for the application rights of sector-j knowledge in sector i, $\forall i, j \in \mathcal{J}$. Should a firm decide not to innovate in sector i in a given period, it can—and find it optimal to— license the application right of its knowledge in related sector j to another firm for adaption during that period, assuming perfect intellectual property rights protection. In equilibrium, an innovating firm in i would optimally obtain such rights (as the fixed costs of innovation imply increasing returns to knowledge capital) and a non-innovating firm would optimally license such rights. As will be shown later, the equilibrium licensing fee that clears the market of a given using (i)-used (j) sector-pair reflects the per-period application value of sector-j knowledge to sector-i innovation, ω^{ij}. As in standard models with competitive markets, firms in this economy take this fee as given when making their R&D decisions. As long as such a market exists, all knowledge is utilized in the economy—either completely utilized by its original inventor or incompletely utilized by licensees ("incompletely" due to the aforementioned absorption cost, $\kappa(l_{f,t}^{ij}/z_{f,t}^i)$) in (12).

Introducing the absorption costs is necessary, because otherwise, all firms would strictly prefer to license its knowledge application rights to others to avoid the fixed costs of innovation. In this case, firms' sectoral innovation decisions would be completely random (driven solely by the random draw of fixed costs)—as opposed to depending on its current knowledge portfolio. In addition, The absorption friction implies that the application value for its original inventor is larger than the application value for a licensee and hence larger than its value in the licensing market. Thus, as long as its current knowledge portfolio generates enough expected payoff in sector i, the firm would still prefer to continue innovating in i even though innovation incurs a fixed cost.

[23] A firm's market share in a given sector may shrink, however, if its innovation rate is lower than the average innovation rate in the sector. If a firm stops R&D in the sector, its market share will reduce to close to zero eventually and the firm may stop innovating. In this way 'creative destruction' is embodied in the model.

We acknowledge that the assumption of the existence of such licensing markets is certainly strong, but it is an important model element for providing the basic tractability of the model. In the presence of the complicated heterogeneities in multiple dimensions in the model, this assumption allows all heterogeneous firms to take the sector-specific knowledge value as given in equilibrium—as opposed to different firms bidding a different price. This allows us to focus on the interesting questions of path-dependence of firm's R&D decisions.

Timing A firm f begins period t with a knowledge portfolio $\boldsymbol{z}_{f,t}$. At the beginning of t, the firm draws a set of the idiosyncratic shocks to fixed costs $(\zeta^i_{f,t})_{i\in\mathcal{J}}$. It then decides whether to conduct R&D in each sector $i \in \mathcal{J}$. If it innovates in i, the firm would optimally acquire more related knowledge and decide its optimal R&D, financed by issuing equity. After that, the firm draws invention shocks $(\varepsilon^{ij}_{f,t})_{i,j\in\mathcal{J}}$ from $G(\varepsilon)$. $\Delta z^i_{f,t}$ new blueprints are then created which will generate profit in the next period and the firm updates its knowledge capital in sector i to $z^i_{f,t+1}$. If the firm pauses its R&D activity in sector i at t, it continues producing and making profit using its existing knowledge, and at the same time licenses the application rights of all its related knowledge to firms which innovate in i. A similar process takes place in every sector, and the firm enters period $t+1$ with a knowledge portfolio $\boldsymbol{z}_{f,t+1}$. Each firm compares the value of these two options to make innovation decisions in each sector based on the realized fixed costs and its existing $\boldsymbol{z}_{ft}$.

Knowledge Pricing Let v^i_t denote the total market value of knowledge capital in sector i and ω^{ji}_t denote the per-period licensing fee of using sector i's knowledge to innovate in sector j. Since all existing knowledge is adapted to create new knowledge in this economy in every period (although not necessarily by its original inventor), the market value of a given unit of knowledge capital is not only given by the present discounted value of its future profits in its own sector but also by its application values (the equilibrium licensing fees) in all sectors. Therefore,

$$\frac{v^i_t}{n^i_t} = \sum_{\gamma=0}^{\infty} \frac{1}{(1+r)^\gamma} \frac{\pi^i_{t+\gamma} + \sum_{j\in\mathcal{J}} \omega^{ji}_{t+\gamma}}{n^i_{t+\gamma}}. \tag{13}$$

On the BGP, $\pi^i_t = \pi^i$ according to (9) and $\omega^{ji}_t = \omega^{ji}$. Appendix B.1 proves that the number of varieties in different sectors grow at the same constant rate $n^j_{t+1}/n^j_t = g$. Let$\rho = [(1+r)g]^{-1}$. We can then rewrite (13) as

$$v^i = \frac{1}{1-\rho}(\pi^i + \sum_{j\in\mathcal{J}} \omega^{ji}). \tag{14}$$

Firm's R&D, Knowledge Acquisition and Sectoral Selection Decisions A firm, given its existing knowledge portfolio $(\boldsymbol{z}_{f,t})$ and expected future value per blueprint $(\frac{v^i_{t+1}}{n^i_{t+1}})$, makes three decisions—optimal R&D investment $(R^{ij}_{vf,t}, R^{ij}_{mf,t})_{j\in\mathcal{J}}$, optimal knowledge acquisition $(l^{ij}_{f,t})_{j\in\mathcal{J}}$ and innovation decisions ($I^i_{f,t} = \{1$ if innovates in i; 0 otherwise$\}$) in every sector i. The decisions are made by weighing the tradeoff between the expected gain from conducting research in that sector and the value of licensing rights in sector i. Therefore, given the prices v^i and ω^{ij}, the firm solves the following maximization problem in i:

$$\max_{(R^{ij}_{vf,t}, R^{ij}_{mf,t}, l^{ij}_{f,t})_{j\in\mathcal{J}}, I^i_{ft}} \left\{ \frac{1}{1+r} E_t\left(\frac{v^i \Delta z^i_{f,t}}{n^i_{t+1}}\right) - \sum_{j\in\mathcal{J}} (R^{ij}_{vf,t} + R^{ij}_{mf,t}) - \sum_{j\in\mathcal{J}} \omega^{ij} \frac{l^{ij}_{f,t}}{n^j_t} - \zeta^i_{f,t}, \quad \sum_{j\in\mathcal{J}} \omega^{ij} \frac{z^j_{f,t}}{n^j_t} \right\} \tag{15}$$

subject to the knowledge creation function (12). Conducting research in sector i entails an researcher cost ($-\sum_{j\in\mathcal{J}} R^{ij}_{vf,t} + R^{ij}_{mf,t}$), a cost of licensing fees ($\sum_{j\in\mathcal{J}} \omega^{ij} \frac{l^{ij}_{f,t}}{n^j_t}$) and a fixed cost ($\zeta^i_{f,t}$). But the effort creates additional blueprints of $\Delta z^i_{f,t}$, which generates a present value of $\frac{1}{1+r} E_t(\frac{v^i \Delta z^i_{f,t}}{n^i_{t+1}})$ in expectation. On the other hand, not innovating in sector i allows the firm to collect licensing fees from other firms who decide to conduct research in i with a value of $\sum_{j\in\mathcal{J}} \omega^{ij} \frac{z^j_{f,t}}{n^j_t}$. The firm would innovate in this sector if the value of the former is larger than that of the latter.

Solving the optimization problem in (15) involves three steps. First, suppose it innovates in i, given its private knowledge portfolio ($\boldsymbol{z}_{f,t}, \boldsymbol{l}_{f,t}$) and market prices for knowledge (v^i, ω^{ij}), the firm chooses the optimal R&D investment to maximize the expected value of innovation:

$$\lambda^i_{f,t} = \max_{(R^{ij}_{v,ft}, R^{ij}_{m,ft})_j} \frac{1}{1+r} E_t\left(\frac{v^i_t \Delta z^i_{f,t}}{n^i_{t+1}}\right) - \sum_{j\in\mathcal{J}} (R^{ij}_{v,ft} + R^{ij}_{m,ft}) - \sum_{j\in\mathcal{J}} \omega^{ij}_t \frac{l^{ij}_{f,t}}{n^j_t} - \zeta^i_{f,t}. \tag{16}$$

Solving (16) generates the optimal R&D associated with applying sector-j knowledge to sector i as

$$R^{ij}_{f,t} \equiv R^{ij}_{v,ft} + R^{ij}_{m,ft} = \frac{\alpha}{1-\alpha} \hat{\omega}^{ij} \left[\tilde{z}^j_{f,t} + \kappa(l^{ij}_{f,t}/z^i_{f,t}) \tilde{l}^{ij}_{f,t} + \theta \tilde{\bar{z}}^j_t \right], \tag{17}$$

where $\tilde{z}^j_{f,t} = \frac{z^j_{ft}}{n^j_t}$, $\tilde{l}^{ij}_{f,t} = \frac{l^{ij}_{f,t}}{n^j_t}$, $\tilde{\bar{z}}^j_t = \frac{\bar{z}^j_t}{n^j_t}$ are the normalized own knowledge capital, acquired knowledge and public knowledge, where

$$\hat{\omega}^{ij} = \frac{1-\alpha}{\alpha} \frac{n^j}{n^i} \left(A^{ij} \alpha \rho v^i\right)^{\frac{1}{1-\alpha}} \left(M^i\right)^{\frac{\alpha}{\alpha-1}}. \tag{18}$$

Therefore, according to (17) the firm scales up its R&D in proportion to its absorbed knowledge capital ($\tilde{z}^j_{f,t} + \kappa(l^{ij}_{f,t}/z^i_{f,t}) \tilde{l}^{ij}_{f,t} + \theta \tilde{\bar{z}}^j_t$), and the proportion is governed by $\hat{\omega}^{ij}$, which captures the *internal* application value of sector j's knowledge to innovation in sector i. $\hat{\omega}^{ij}$ increases with the knowledge applicability from j to i (A^{ij}) and the market value of knowledge in sector i (v^i), and decreases with competition—the number of firms that are innovating in sector i (M^i).

Second, substituting (14) and (17) to (16) and solving for the optimal amount of acquired knowledge ($l^{ij}_{f,t}$) lead to the demand function of licensing rights:

$$\omega^{ij} = \left[\kappa(\frac{l^{ij}_{f,t}}{z^j_{f,t}}) + \kappa'(\frac{l^{ij}_{f,t}}{z^j_{f,t}})\frac{l^{ij}_{f,t}}{z^j_{f,t}} \right] \hat{\omega}^{ij}. \tag{19}$$

Under the assumptions that $\kappa(\cdot) < 1$ and $\kappa'(\cdot) < 0$, (19) implies that (i) the market value of sector-j knowledge in sector i is strictly less than the internal application value, *i.e.* $\omega^{ij} < \hat{\omega}^{ij}$; and that (ii) all firms acquire external knowledge in the exact same proportion to its own knowledge and on the BGP this proportion is constant for any given sector pair ij. Let τ^{ij} denote the proportion, then

$$l^{ij}_{ft} = \tau^{ij} z^j_{f,t}. \quad \forall f \tag{20}$$

Therefore, (19) becomes

$$\omega^{ij} = \left[\kappa(\tau^{ij}) + \kappa'(\tau^{ij})\tau^{ij} \right] \hat{\omega}^{ij}. \tag{21}$$

Substituting (14), (17) and (21) back into (16), we can rewrite the expected value of innovation, $\lambda^i_{f,t}$, as a function of τ^{ij}:

$$\lambda^i_{f,t} = \sum_{j \in \mathcal{J}} \left(1 - \kappa'(\tau^{ij})(\tau^{ij})^2\right) \hat{\omega}^{ij} \tilde{z}^j_{f,t} + u^i - \zeta^i_{f,t}, \tag{22}$$

where u^i stands for the value of *public knowledge* per firm, measured by the total application value generated by public knowledge from all sectors to i:

$$u^i = \sum_{j \in \mathcal{J}} \hat{\omega}^{ij} \theta \bar{\tilde{z}}^j_t = \sum_{j \in \mathcal{J}} \frac{\theta \hat{\omega}^{ij}}{M^j}. \tag{23}$$

(23) says that when public knowledge is easier to adapt (higher θ), or abundant (low M^j), or more applicable (higher $\hat{\omega}^{ij}$), the value of public knowledge is higher.

The first term of the right-hand side of (22) implies that the value of innovating in sector i is high when a firm's private knowledge portfolio is highly applicable to innovation in i. Thus, the sectoral innovation decision is *path-dependent* as some firms' existing knowledge can be more easily adapted to make new innovation than others'.

Lastly, in deciding whether to innovate in sector i, the firm compares the value of innovating in i by itself ($\lambda^i_{f,t}$) against its outside option of collecting licensing fees, $\sum_{j \in \mathcal{J}} \omega^{ij} \tilde{z}^j_{f,t}$. Combining (21), (22) and (15), the firm would innovate in i at t if and only if

$$\zeta^i_{f,t} \leq \sum_{j \in \mathcal{J}} [1 - \kappa(\tau^{ij}) - \kappa'(\tau^{ij})\tau^{ij}(\tau^{ij} + 1)] \hat{\omega}^{ij} \tilde{z}^j_{f,t} + u^i \equiv \varphi^i_{f,t}, \tag{24}$$

where $\varphi^i_{f,t}$ is defined as the firm's expected value of innovating in sector i. Given the property of the function $\kappa(\cdot)$, $1-\kappa(\tau^{ij})-\kappa'(\tau^{ij})\tau^{ij}(\tau^{ij}+1)>0$ for all $\tau^{ij}>0$. Both sides of (24) contains firm-specific variables: a firm's existing knowledge portfolio $\{\tilde{z}^j_{f,t}\}_{j\in\mathcal{S}_{f,t}}$ and its idiosyncratic fixed costs $\zeta^i_{f,t}$. Even with the same draw of $\{\zeta^i_{f,t}\}_i$, firms would still choose to innovate in different sectors in order to best apply their existing knowledge. Other things equal, sectors with higher u^i attract more entry.

It becomes clear that without the absorption cost (i.e. $\kappa(\tau^{ij})=1$), $\hat{\omega}^{ij}=\omega^{ij}$ and $1-\kappa(\tau^{ij})-\kappa'(\tau^{ij})\tau^{ij}(\tau^{ij}+1)=0$. In this case, a firm would innovate in i if and only if it has a lucky draw of fixed cost in i (i.e. $\zeta^i_{f,t}\le u^i$). Then there would be no endogenous sorting of firms in different sectors or path-dependent sectoral selection, as firm's sectoral selections are guided solely by the random draws of fixed costs, and not by its existing knowledge portfolio.

Knowledge Market Clearing The knowledge market clears for each source-application sector pair. Given (24), we can express the fraction of sector j knowledge that is utilized by its original inventor in i as $\tilde{s}^{ij}\equiv\int_{f\in\mathcal{F}^j}H^i(\varphi^i_{f,t})\tilde{z}^j_{ft}df$, where $H^i(\cdot)$ is the *c.d.f.* of fixed costs shocks ζ^i_{ft}. The supply of application rights is thus $1-\tilde{s}^{ij}$, as all knowledge that is not applied by its own inventor in i is licensed to another innovator for one period in equilibrium. Given that all innovating firms demand the same external-to-own knowledge ratio (τ^{ij}), the demand for private external knowledge amounts to $\tau^{ij}\tilde{s}^{ij}$. Therefore, market clearing implies:

$$1-\tilde{s}^{ij}=\tau^{ij}\tilde{s}^{ij}. \tag{25}$$

In the end, the price ω^{ij} and quantity τ^{ij} of knowledge licensing rights are jointly determined by (21) and (25) in equilibrium.

Newborn Firms There is a mass of prospective newborn firms in the economy, which have never invented in *any* sector, i.e. $\tilde{z}^i_{f,t}=0,\ \forall i$. Therefore, in addition to sectoral entry discussed earlier (by innovating firms in other sectors), there is also entry by these newborn firms (startups). These firms also make a draw of $\zeta^i_{f,t}$ from the *c.d.f.* $H^i(\zeta^i_{f,t})$ each period and decide whether to enter any sectors. The same condition (24) applies to these firms, which implies that a newborn firm enters the economy by starting from the sector where the fixed cost can be covered by absorbing the public knowledge. Since firms have different random draws of fixed costs $\zeta^i_{f,t}$, different firms may initially enter different sets of sectors $S_{0,f}$:

$$S_{0,ft}=\left\{i\in\mathcal{J}\ |u^i-\zeta^i_{f,t}>0\right\}. \tag{26}$$

Thus, the probability that a newborn firm would enter sector i is given by $H^i(u^i)$. The fraction of newborn firms that enter sector i is the same as the probability of entering i conditional on entering at least one sector of the economy; hence,

$$h_0^i = \frac{H^i(u^i)}{1 - \Pi_{i \in \mathcal{J}}(1 - H^i(u^i))}, \tag{27}$$

where the denominator $1 - \Pi_i(1 - H^i(u^i))$ is the probability that a newborn firm enters at least one sector. Combining (12) and (17), we derive the average (normalized) size of a newborn firm in sector i as:

$$\bar{\tilde{z}}_0^i = \frac{(1+r)u^i}{(1-\alpha)v^i}. \tag{28}$$

3.2.3 Sectoral Behavior

Firm Size Distribution Since varieties in the same sector are produced at the same quantity, the normalized firm size in sector i for firm f is the same as its market share $\tilde{z}^j_{f,t}$. According to the equilibrium knowledge accumulation in (11), the knowledge creation in (12) and the optimal R&D investment in (17), firm size dynamics can be derived as follows (see Appendix B.3 for details):

$$\tilde{\boldsymbol{z}}_{f,t+1} = \boldsymbol{\Phi}_{f,t+1}\tilde{\boldsymbol{z}}_{f,t} + \boldsymbol{\Psi}_{f,t+1}\mathbf{b}, \tag{29}$$

where the K-dimensional vector $\tilde{\boldsymbol{z}}_{f,t} \equiv (\tilde{z}^1_{f,t}, ..., \tilde{z}^K_{f,t})'$, the constant vector $\mathbf{b} \equiv (\theta/M^1, ..., \theta/M^K)'$, and $\boldsymbol{\Phi}_{f,t}$ and $\boldsymbol{\Psi}_{f,t}$ are $K \times K$ matrices with the $(i,j)^{th}$ elements given by $\phi^{ij}_{f,t}$ and $\psi^{ij}_{f,t}$ respectively. Specifically,

$$\phi^{ij}_{f,t+1} = \left[\frac{1_{\{\text{if } i=j\}}}{g} + I^i_{f,t}\left(1 + \kappa(\tau^{ij})\tau^{ij}\right)\xi^{ij}\varepsilon^{ij}_{f,t}\right] \tag{30}$$

and

$$\psi^{ij}_{f,t+1} = \xi^{ij} I^i_{f,t} \tag{31}$$

where $\xi^{ij} = \frac{(1+r)\hat{\omega}^{ij}}{(1-\alpha)v^i}$, $1_{\{\text{if } i=j\}}$ is the indicator function for $i = j$ and the indicator variable $I^i_{f,t} = 1$ if firm f innovates in sector i at time t and zero otherwise. $I^i_{f,t} = 1$ with probability $H^i(\varphi^i_{f,t})$.

Note that $\{\phi^{ij}_{f,t+1}\}$ and $\{\psi^{ij}_{f,t+1}\}$ are Markov stochastic processes representing, respectively, the effective rate of returns (innovation rate) when applying sector-j private knowledge capital and public knowledge capital in sector i, respectively. If $\{\phi^{ij}_{f,t+1}\}$ and $\{\psi^{ij}_{f,t+1}\}$ were i.i.d. processes and independent of $\tilde{\boldsymbol{z}}_{f,t}$, this dynamics of firm size would converge to a stationary Pareto distribution according to Kesten (1973). However, $\{\phi^{ij}_{f,t+1}\}$ and $\{\psi^{ij}_{f,t+1}\}$, are endogenously determined by the knowledge accumulation condition and by firm's optimal R&D decisions. They depend not only on the i.i.d. innovation shocks ($\varepsilon^{ij}_{f,t}$), but also on firm's binary decision of whether or not to innovate in sector i ($I^i_{f,t}$), which in turn is determined by firm's current knowledge portfolio $\tilde{\boldsymbol{z}}_{f,t}$ and i.i.d. fixed cost shocks $\zeta^i_{f,t}$ (see Equation (24)). The latter component ($I^i_{f,t}$) implies autocorrelations in $\{\phi^{ij}_{f,t+1}\}$ which captures the interesting aspect of path-dependence in firms' innovation decisions. It also implies that our firm dynamics in (29) is not linear in $\tilde{\boldsymbol{z}}_{f,t}$.

Appendix B.3 shows that although it is not exactly linear, the firm dynamics (29) are asymptotically linear, a result which allows us to apply new findings in the mathematics of stochastic recursive processes (Mirek, 2011) under appropriate assumptions. We are able to show that the stationary firm distribution exists in the steady state and has a Pareto tail:[24]

$$\Pr(\tilde{z}_f^i > z) \sim (\frac{z}{k^i})^{-\mu^i}, \tag{32}$$

where k^i is the scale parameter and μ^i is the shape parameter, and Pr stands for the steady state probability. The economics of this result is intuitive. When knowledge creation is stationary, it accumulates additively to knowledge stock. However, since existing knowledge stock is also used in knowledge creation process as in (12), the multiplicative process of knowledge accumulation then leads to the Pareto tail distribution of firm size. Firms which have realized an extended series of high positive innovation shocks and low fixed costs populate the tails in the firm size distribution.

To gain further insights on the shape of the Pareto distribution, note that the lower bound of this distribution is associated with imitated new varieties, $\sum_j \xi^{ij} \frac{\theta}{M^j}$. The existence of public knowledge ($\theta > 0$) plays an important role in attenuating the size dispersion generated by idiosyncratic innovation shocks such that the minimum firm size would not become too small. Suppose the Pareto distribution also depicts the distribution of the small firms relatively well, the shape parameter μ^i can then be expressed as a function of public-private knowledge value ratio:

$$\mu^i = (1 - \frac{1+r}{1-\alpha} \frac{u^i M^i}{v^i})^{-1}. \tag{33}$$

It strictly increases with the ratio between the public knowledge value and private knowledge value in that sector, $u^i M^i / v^i$. Intuitively, a sector with a higher value of public knowledge relative to private knowledge has a more homogeneous distribution of firm sizes, as small firms disproportionately benefit more from the public knowledge.

The mass of firms innovating in sector i is the mass of firms which satisfies (24):

$$M^i = M \Pr \left\{ \varphi_{f,t}^i \geq \zeta_{f,t}^i \right\}. \tag{34}$$

In the stationary BGP, the mass of innovating firms is constant.

[24]For the stationary solution of the stochastic difference equation $Y_{n+1} = a_n Y_n + b_n$ to exhibit power law behavior, Kesten (1973) requires the multiplicative coefficients (a_n) to be a sequence of i.i.d. random variables. Recent works extending Kesten (1973) include Saporta (2005) and Roitershtein (2007) which allow the stochastic coefficients to be a finite state space Markov chain, and Hay, Rastegar and Roitershtein (2011) which allows Y_n to be a multi-dimensional variable. Stelzer (2008) proves that under certain conditions the Pareto law distribution holds with a_n being a continuous state space Markov process and with multi-dimensional variables. Mirek (2011) further shows that, for non-linear recursive processes that are approximately linear, the size distribution converges to a Pareto distribution at the tail. For more discussions and applications of these results, see Benhabib, Bisin and Zhu (2011, 2015) in the context of wealth distribution and Luttmer (2007), Gabaix (2009) and Cai (2012) in the context of firm size distribution in one-sector models.

3.3 Aggregate Conditions

The population supplies L units of labor in every period which are allocated in three areas: production workers, researchers, and workers who maintain the research labs (the fixed costs of R&D): $L = \sum_{i\in\mathcal{J}} L^i_{p,t} + \sum_{i\in\mathcal{J}}\sum_{j\in\mathcal{J}}\int_{f\in\mathcal{F}^i\cap\mathcal{F}^j} R^{ij}_{f,t}df + \sum_{i\in\mathcal{J}}\int_{f\in\mathcal{F}^i}\zeta^i_{f,t}df$. Using (10) and (17) we can rewrite the above equation as:

$$L = \frac{\sigma^i - 1}{\sigma^i}PY + \sum_{i\in\mathcal{J}}[\alpha\rho(g-1)v^i + F^i M^i]. \tag{35}$$

Therefore, the division of labor is also time-invariant in the BGP.

In this closed economy without physical capital, goods market clearing implies $C_t = Y_t$. In addition, the household owns all the firms and finances all the potential entrants. Given an interest rate r, every period the household gets net income $r\sum_i v^i$ from investing in firms.[25] The household's total income is

$$PY = L + r\sum_{i\in\mathcal{J}} v^i. \tag{36}$$

Thus, according to (9) the sectoral profit π^i in the BGP is indeed a constant. Following (5), the equilibrium interest rate is determined by

$$1 = \beta(1+r)g^{(\eta-1)\sum_i \frac{s^i}{1-\sigma^i}}. \tag{37}$$

3.4 Equilibrium Definition

Definition 1 *A stationary balanced growth path (BGP) is an equilibrium path in which output, consumption and innovation grow at constant rates and firm size distribution is stationary in every sector. It is given by: time paths of aggregate quantities and prices* $[C_t, Y_t, P_t, w_t, r_t]^\infty_{t=0}$*; time paths of sectoral numbers of varieties, numbers of firms, quantity and price of rental application rights and knowledge value* $[n^i_t, M^i_t, \tau^{ij}_t, \omega^{ij}_t, v^i_t]^\infty_{i,j\in\mathcal{J},t=0}$*; time paths of firms' R&D investment* $[R^{ij}_{f,t}]^\infty_{i,j\in\mathcal{J}\times\mathcal{J},f\in\mathcal{F}^j_t,t=0}$*, number of blueprints,* $[z^i_{f,t}]^\infty_{i\in\mathcal{J},f\in\mathcal{F}^i_t,t=0}$*; and time paths of firm's sectoral entry and exit decisions* $[I^i_{f,t}]^\infty_{i\in\mathcal{J},f\in\mathcal{F}^i_t,t=0}$*, such that:*

1. *Given w_t, r_t and P_t, the representative household maximizes life-time utility subject to an inter-temporal budget constraint. That is, (36) and (37) are satisfied.*

2. *Given w_t, r_t and P_t, the individual firm decides on the quantity and prices of goods produced and production labor needed. That is, (8) and (10) are satisfied.*

3. *Given w_t, r_t, P_t, $\{\omega^{ij}_t\}_{i,j\in\mathcal{J}}$, and $\{v^i_t\}_{i\in\mathcal{J}}$, firms decide on optimal R&D investment, optimal acquisition of external knowledge, and sectoral entry and exit decision, That is, (11), (12),*

[25] It is equivalent to receiving dividends as profit and capital gains.

(14), (18), (17), (21) and (23) are satisfied. A firm's sectoral entry and exit decision is based on (24) which also determines the number of firms in each sector.

4. *Labor markets clear as in (35).*

5. *Goods markets clear such that $C_t = Y_t$.*

6. *Knowledge markets clear as in (25).*

4 Model Implications

We have commented along the way on the intuitions and insights provided by the model. With our theory in hand, we now ask how our model can potentially fit the sector-level and firm-level observations previously documented in Section 2 and what are the aggregate implications for long-run growth.

4.1 Heterogeneous R&D intensity Across Sectors

We first turn to the observed positive relationship between sectoral R&D intensity and its knowledge applicability documented in Section 2.2. According to the model, R&D intensity (R&D expenditure as a fraction of sales) in sector i is given by $RI^i \equiv \frac{1}{s^i PY} \sum_{j \in \mathcal{J}} \int_{f \in \mathcal{F}^i} R_f^{ij} df$. Aggregating the optimal R&D investment in (17) over all firms in the same sector, we obtain (see Appendix B.2 for detailed derivation):

$$RI^i = \alpha \rho (g-1) \frac{v^i}{s^i PY}, \tag{38}$$

which implies

$$\frac{RI^i}{RI^k} = \frac{v^i / s^i}{v^k / s^k}. \tag{39}$$

Thus, the model predicts that R&D resources at the aggregate are allocated across sectors proportional to the knowledge value per output unit.

While v^i is not directly observed in the data, a central prediction of our model is that v^i depends on its knowledge market application value ω^{ji}, which is ultimately determined by $[A^{ij}]_{(i,j) \in \mathcal{J} \times \mathcal{J}}$. Therefore, v^i captures a similar notion to the empirical measure of applicability—and we will show by simulation in Section 5.3 that v^i and app^i are indeed highly positively correlated. According to (39) the model thereby predicts that persistent cross-sector variation in R&D intensity is an outcome of fundamental differences in knowledge applicability, in line with the **Observation 1**.

4.2 Heterogenous Firm Innovation

We have shown in Section 2 that the applicability of a firm's knowledge differs greatly, depending on where the firm positions itself in the technology space. In addition the firm's knowledge applicability is found to be systematically related to its patent stock and patent scope; more importantly, it affects its future innovation rate. If we equate patents with innovation (knowledge creation) we can use our model to interpret these findings.

R&D and Innovation Allocations Across Sectors To evaluate sectoral allocation of R&D and innovation within the firm, we summarize a firm's research effort in sector i by $R^i_{f,t} = \sum_j R^{ij}_{f,t}$ and its (expected) innovation in that sector by $E\Delta\tilde{z}^i_{f,t}$:

$$
\begin{aligned}
R^i_{f,t} &= \frac{\alpha}{1-\alpha} \sum_{j \in \mathcal{S}_{f,t}} \hat{\omega}^{ij} [\left(1 + \kappa(\tau^{ij})\tau^{ij}\right) \tilde{z}^j_{f,t} + \theta \tilde{\bar{z}}^j], \qquad (40) \\
E\Delta\tilde{z}^i_{f,t} &= \frac{1}{(1-\alpha)\rho} \sum_{j \in \mathcal{S}_{f,t}} \frac{\hat{\omega}^{ij}}{v^i} [\left(1 + \kappa(\tau^{ij})\tau^{ij}\right) \tilde{z}^j_{f,t} + \theta \tilde{\bar{z}}^j]. \qquad (41)
\end{aligned}
$$

Given (14), it is clear that $\frac{\hat{\omega}^{ij}}{v^i}$ strictly increases with $\hat{\omega}^{ij}$. At any given t, firms in the economy differ in their existing technology portfolios (captured by $\{\tilde{z}^j_{f,t}\}_j$) which is also reflected in their technology scope ($\mathcal{S}_{f,t}$). (40) and (41) imply that the firm allocates its innovation effort across sectors according to the application value generated by its existing set of knowledge.

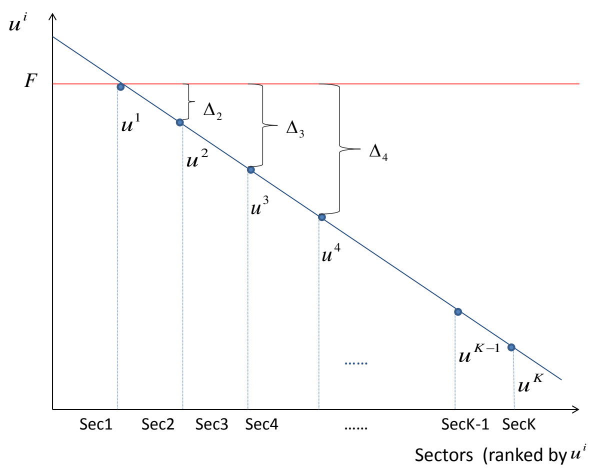
These equations help to explain Observation 2 (i.e. firms with larger patent stock or patent scope tend to concentrate more in the applicable central sectors). To see this, consider two representative sectors: a central sector (denoted by c) and a peripheral sector (denoted by p). A central sector is highly connected to other sectors with many large positive values of $\hat{\omega}^{cj}$, while the peripheral sector has few sectors linked to it. When the firm's knowledge stock increases (larger $\tilde{z}^j_{f,t}$ in general), the application value of its knowledge portfolio rises more for central sectors than it does for the peripheral ones, because the coefficients $\hat{\omega}^{cj} > \hat{\omega}^{pj}$. Similarly, when a firm's knowledge scope expands (larger $|\mathcal{S}_{f,t}|$), many large elements of $\hat{\omega}^{cj}\tilde{z}^j_{f,t}$ are added to R^c_f, but only a few elements of $\hat{\omega}^{pj}\tilde{z}^j_{f,t}$ are added to R^p_f. Therefore, R^c_f/R^p_f and $\Delta\tilde{z}^c_f/\Delta\tilde{z}^p_f$ both increase with the firm's knowledge stock and knowledge scope, consistent with the **Observation 2**.

4.2.1 Sequential Sectoral Entry

Sequential sectoral entry **(Observation 3)** can be better illustrated by first considering a simplified case in which every firm faces the same fixed cost F (i.e. no idiosyncratic fixed cost risk, $\zeta^i_f = F$). In this case, free entry by newborn firms implies that entry stops when the net value of entry is zero. That is, $\max_i\{u^i\} = F$. The sector that offers the maximum public knowledge value is the first sector that every new firm enters. This condition along with the sectoral entry/exit condition (24) implies that firms enter different sectors *sequentially*: they start developing blueprints in a sector that offers the largest public knowledge value, build up their private knowledge and gradually venture into other sectors using their accumulated knowledge capital.

Figure III explains this pattern of sequential sectoral entry. Suppose sectors are ranked by their value of public knowledge as $u^1 > u^2 > ... > u^K$. The horizontal line of fixed costs (F) intersects with u^1 according to the free entry condition of newborn firms. Every newborn firm enters sector 1 first in this scenario of no idiosyncratic risks ζ_f^i. In order to enter more sectors, the firm then needs to accumulate more private knowledge to fill up the gap between the fixed cost and the value of public knowledge (denoted by $\Delta_i = F - u^i$). As in (24) when the private knowledge value covers the gap Δ_i, the firm enters another sector i.

Figure III: Determination of Firm's Entry into Multiple Sectors



We note that in our more general setup, firms face idiosyncratic shocks to innovation and fixed costs. Hence, not all firms follow the exact same path expanding across the technology space. However, their entries are path-dependent: depending on which sectors they have entered and are actively conducting research in, the inter-sectoral knowledge linkages help to direct the next optimal step. This is one of the features that distinguish our model from others in the literature.

4.2.2 Innovation Rate

In Observation 4, controlling for firm size, firms possessing highly applicable knowledge subsequently tend to innovate faster. A firm innovates by both expanding across sectors and creating new blueprints in the existing sectors, the overall innovation rate (in expectation) can be captured in the model by $g_{f,t+1} = E\frac{\sum_{i\in\mathcal{J}} \Delta\tilde{z}^i_{f,t}}{\sum_{i\in\mathcal{J}} \tilde{z}^j_{f,t}}$, which in equilibrium can be derived as:

$$g_{f,t+1} = \frac{1}{(1-\alpha)\rho} \sum_{j\in\mathcal{S}_{f,t}} s^j_{f,t} \sum_{i\in\mathcal{S}_{f,t}} \frac{\hat{\omega}^{ij}}{v^i}, \tag{42}$$

where $s^j_{f,t} = (\left(1+\kappa(\tau^{ij})\tau^{ij}\right)\tilde{z}^j_{f,t} + \theta\tilde{\bar{z}}^j)/\sum_{j\in\mathcal{J}1}\tilde{z}^j_{f,t}$ and can be broadly interpreted as firm's allocation of patents across sectors (if θ is small).

(42) implies that a firm's innovation rate (i.e. the growth rate of its blue prints) depends on the firm's existing knowledge distribution. Firms that concentrate more in highly applicable knowledge (reflected in a high positive correlation between s^j_f and $\sum_{i\in\mathcal{S}_{f,t}}\frac{\hat{\omega}^{ij}}{v^i}$) are able to develop new products faster, as applicable sectors offer extensive knowledge spillovers to other sectors, leading to higher firm growth. This is in line with the **Observation 4**.

Although it is not the focus of this paper, (42) also explains why firm innovation rate decreases with its patent stock but increases with its patent scope as shown in Table VI. The model predicts that a larger firm which has already accumulated large amount of private knowledge, innovates more slowly because it benefits less from the public knowledge. On the other hand, expansive knowledge scope—reflected by a larger set $\mathcal{S}_{f,t}$—allows the firm to utilize and apply knowledge from more sectors, and hence increases its overall innovation rate.

4.3 Innovation Allocation and Aggregate Growth

What are the aggregate implications of growth generated by the model? As in standard variety expanding models, real output growth is driven by the "variety effects": expansion in varieties is associated with a decrease in goods prices. Therefore, real output growth strictly increases with the innovation rate. Define δ^{ij} as the fraction of j's knowledge (including both private and public) that is eventually absorbed and utilized in innovation in sector i:

$$\delta^{ij} = \tilde{s}^{ij} + \kappa(\tau^{ij})(1-\tilde{s}^{ij}) + \theta\frac{M^i}{M^j}. \tag{43}$$

Summing up all firm's accumulated knowledge in (12), we can derive the (gross) growth rate of the number of varieties in the whole economy in the BGP equilibrium as:

$$g = \ (1-\beta)\left[(1-\alpha)\beta\frac{\sum_{i\in\mathcal{J}}\pi^i + \sum_{i\in\mathcal{J}}\sum_{j\in\mathcal{J}}\hat{\omega}^{ij}}{\sum_{i\in\mathcal{J}}\sum_{j\in\mathcal{J}}\hat{\omega}^{ij}\delta^{ij}} - 1\right]^{-1}. \tag{44}$$

See Appendix B.2 for details of derivation.

This equation provides three insights. First, everything else being the same, strengthening knowledge linkages across sectors enhances growth (because $\hat{\omega}^{ij}$ increases). Second, in the presence of fixed costs, not every firm innovates in every sector. Therefore, the fraction of private knowledge that is used across sectors $\tilde{s}^{ij}+\kappa(\tau^{ij})(1-\tilde{s}^i), \forall i, j$, is strictly less than one (full utilization of knowledge across sectors). Hence, (44) implies that sectoral fixed costs reduce the aggregate innovation rate in the economy by blocking the knowledge circulation across sectors.

Third and more interestingly, when a firm's sectoral selection decision is dictated more by the "luck" factor (i.e. a firm innovates sector i whenever it has a draw of low fixed cost in i) and less by the "fundamental" knowledge linkages, $\hat{\omega}^{ij}$ and δ^{ij} become less correlated. Firms with sufficient background knowledge may not be able to conduct research in many sectors, while firms with insufficient background knowledge may enter many sectors but cannot innovate much. This random sorting of firms in different sectors—manifested in low $\sum_i \sum_j \hat{\omega}^{ij}\delta^{ij}$—reduces aggregate innovation rate. Thus, the uncertainty in the fixed costs of research generates an additional negative "resource misallocation effect" on aggregate growth—a new insight yielded by the model.

Scale-Free Growth Jones (1999) first pointed out that the "scale effects" that exist in many endogenous growth models are not consistent with empirical evidence. By assuming the efficiency of R&D workers to be proportional to the average knowledge stock in that sector, we eliminate the "scale effects" of population on economic growth. As shown by (14), (18), (21), (35) and (36), π^i, $\hat{\omega}^{ij}$, ω^{ij}, v^i, M^i and PY are all proportional to the total population (L) in the economy, while δ^{ij} does not change with L. Therefore, according to (44), aggregate innovation rate is independent of the population size.

5 Simulation

Although much of the equilibrium can be characterized by closed form expressions that provide useful intuitions and implications, matching the model's predictions to our empirical observations quantitatively requires solving the full general equilibrium, which in turn relies on a knowledge of joint distribution of firm's expected gains of sectoral entry $\varphi^i_{f,t}$ and firm size $\tilde{z}^i_{f,t}$ that cannot be derived analytically. Therefore, in this section we simulate the model economy with a large panel of firms (70,000) innovating in various multiple sectors and assess the performance of the model in matching the empirical observations that motivated our work. A counterfactual experiment is conducted in the end to examine the role of sector-specific fixed innovation costs on aggregate growth.

Since it is computationally costly to simulate a large number of sectors, in this section we consider the same 42 SIC sectors as those examined in Section 2.2. NBER firm patenting data and patent citation data over the same set of sectors during 1976-2006 are employed to discipline the parameters.

5.1 Estimation of Parameters

We assume that idiosyncratic shocks to fixed innovation costs in sector i are drawn from a Gamma distribution $H^i(\zeta)$ with mean F^i and variance σ^i_ζ, and shocks to firms' innovation are drawn from a Gamma distribution $G(\varepsilon)$ with mean one and variance σ_ε and $\varepsilon \in (0, \bar{\varepsilon}]$. We set the absorption capacity function to $\kappa(\tau^{ij}) = \left(1 + b\tau^{ij}\right)^{-a}$, where $1 \geq a > 0$, $b > 0$.[26] The set of parameters of the

[26]This functional form ensures that the market application value ω^{ij} is positive according to (21).

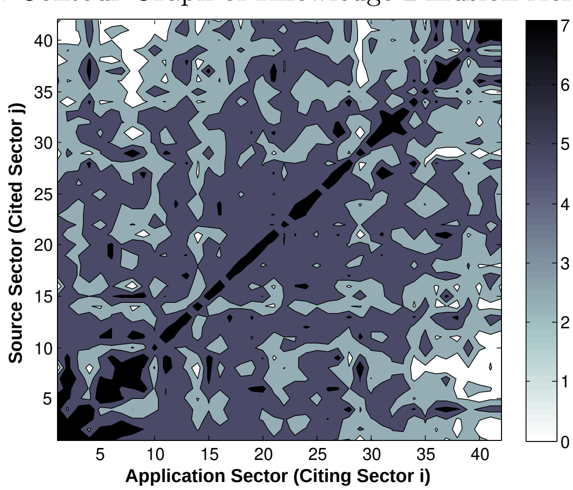
model to be calibrated include elements of the intersectoral knowledge diffusion matrix $[A^{ij}]_{i\times j\in\mathcal{J}\times\mathcal{J}}$ and $\{\beta,\eta,\alpha,\theta,a,b,\sigma^i,s^i,F^i,\sigma^i_\zeta,\sigma_\varepsilon\}$. We explain in turn how to estimate each of these parameters.

Inter-sectoral Knowledge Diffusion Matrix A We proxy the knowledge linkages by the fraction of citations made to sector j by sector i (knowledge flow from j to i). Since sectors with more patents tend to be cited more frequently, we normalize this fraction of citations by the relative importance of sector j, measured by the share of citations received by j in total citations (citationshare^j). Formally,

$$\tilde{A}^{ij} = \frac{\text{no. of citations from } i \text{ to } j/\text{total citations made by } i}{\text{citationshare}^j}. \tag{45}$$

Figure IV shows a contour graph of the knowledge diffusion matrix (in log) for these 42 sectors ranked by their applicability (a lower sector index indicates higher applicability). The darker color signals a larger element in **A**. The contour graph shows that sectors with high knowledge applicability are densely linked to each other, whereas less applicable sectors are loosely connected to only a few other sectors. The darkest area on the diagonal reflects the fact that a large proportion of citations go to patents in the same sector. This is not particularly surprising given that sectors in this case are not highly disaggregated; however, most sectors also allocate a significant number of citations to patents from other sectors, reflecting the importance of cross-sector knowledge spillovers. We normalize the knowledge diffusion matrix by a scale parameter, A_0, such that $A^{ij} = A_0 \times \tilde{A}^{ij}$. A_0 governs the average innovation rate.

Figure IV: Contour Graph of Knowledge Diffusion Across Sectors



Note: The figure represents the knowledge diffusion matrix constructed from the NBER Patent Citation data for 42 SIC sectors. A darker color implies that the sector is cited by another at a higher rate.

Other Parameters Total labor force, L, is normalized to 100, and we choose the following standard values: the growth rate of real output is 2%, the interest rate $r = 5\%$ and $\beta = 0.99$. The average annual (gross) growth rate of patents, g, equals 1.11 as in the data (for the period 1976-2006). (5) then implies that the household's risk aversion parameter $\eta = 3$. The sectoral elasticity of substitution between differentiated goods σ^i is set to match the estimates provided in Broda and Weinstein (2006). The production share, s^i, is estimated using the sectoral output data provided by NBER-CES Manufacturing Industry Database. To save space, the values of sector-specific parameters are relegated to Table A.4 in Appendix B.4.

The equilibrium conditions characterized in previous sections are employed to estimate some of model parameters, $\vartheta \equiv \{A_0, \alpha, \theta, a, b, v^i, F^i, \sigma^i_\zeta\}$, using the Generalized Method of Moments (GMM). The collated complete set of equilibrium conditions are listed in Appendix B.5. In the patent dataset we observe all firms' patent stock and patent applications in 42 sectors in 1997 (assuming the economy is on the BGP in that year): $\{(S^1_f, S^2_f, ...S^{42}_f)\}_{f\in\mathcal{F}}$ and $\{(P^1_f, P^2_f, ...P^{42}_f)\}_{f\in\mathcal{F}}$.[27] Based on these observations, we calculate the relative patent stock across sectors (corresponding to $\{n^i/n^j\}_{ij}$ in the model), the fraction of innovating firms in each sector ($\{M^i/M\}_i$), the share of firms that patent in s number of different sectors ($\{M^s/M\}_s$), the average share of patents applied by new patentees in sector i ($\{\bar{\tilde{z}}^i_0\}_i$), the fraction of new patentees that enter sector i ($\{h^i_0\}_i$) and the fraction of patents in j that are owned by firms innovating in sector i in that year, $\{\tilde{s}^{ij}\}_{ij}$. Then according to (25) τ^{ij} is estimated as $(1 - \tilde{s}^{ij})/\tilde{s}^{ij}$. In addition, we obtain the ratio between the number of firms and total population, M/L, from Axtell (2011).

We set general equilibrium conditions (14), (27), (28), (34), (36), (38), (42) and the equilibrium share of firms that simultaneously innovate in s sectors as our targeted moments. (14), (36) and (38) underpin sectoral knowledge values. (27), (34) and firms co-patenting help to determine parameters governing the distribution of shocks to fixed innovation costs $\{F^i, \sigma^i_\zeta\}_i$. (27) and (28) concerning newborn firms entry pin down the imitation parameter θ. (42) guarantees that every sector grows at the same rate as the model predicted.

More specifically, we substitute the empirical observation of $n^i/n^j, A^{ij}, M^i/M, \bar{\tilde{z}}^i_0, h^i_0, \tau^{ij}, M/L$ into the above equations. Define $\Gamma(\vartheta)$ as the vector of differences between data moments and model-generated moments. Our GMM estimators solve:

$$\hat{\vartheta} = \arg\min_{\vartheta} \Gamma_t(\vartheta) W_\vartheta^{-1} \Gamma_t(\vartheta)', \tag{46}$$

in which W_ϑ is a diagonal weighting matrix with weight being one for all sectoral moments. Once calibrated parameter set $\hat{\vartheta}$ is available, the empirically unobserved licensing fees $\{\omega^{ij}\}_{ij}$ and public knowledge value $\{u^i\}_i$ can be backed out using (18), (21) and (23).

[27] We only keep firms which have filed for at least one patent during the five year period of 1993-1997. Although many firms do not apply for patent every year, the probability of a firm to resume patenting after being inactive for five years is very small.

Our estimated $\alpha = 0.245$, implying a significant input from R&D researchers in the knowledge creation process. The imitation efficiency parameter θ captures how much on average a firm can learn from the public knowledge pool per period, where the size of public knowledge pool is measured by the average firm size. $\theta = 0.002$ implies that every period 0.2% of an average firm's knowledge capital is available to every firm, which is quite substantial for small firms relative to their own knowledge stock.[28] $a = 1.058$ and $b = 4.6 \times 10^3$, implying that the absorption capacity $\kappa(\tau^{ij}) = \left(1 + b\tau^{ij}\right)^{-a} \ll 1$ for large values of τ^{ij}. $A_0 = 0.014$ and parameters related to sector-specific fixed costs of innovation $\{F^i, \sigma^i_\zeta\}_i$ are reported in Table A.4. The estimated average fixed costs are sensible: the three sectors with the highest costs are guided missiles and space vehicles, ship and boat building, ordinance except missiles, and the three sectors with the lowest costs include professional and scientific instruments, electronic components, and general industry machinery and equipment. Lastly, the variance of the innovation shocks $\sigma^2_\varepsilon = 0.2$ so that the variance of the total firm size distribution in the simulation matches that in the data.

5.2 Goodness of Fit at Sectoral Level

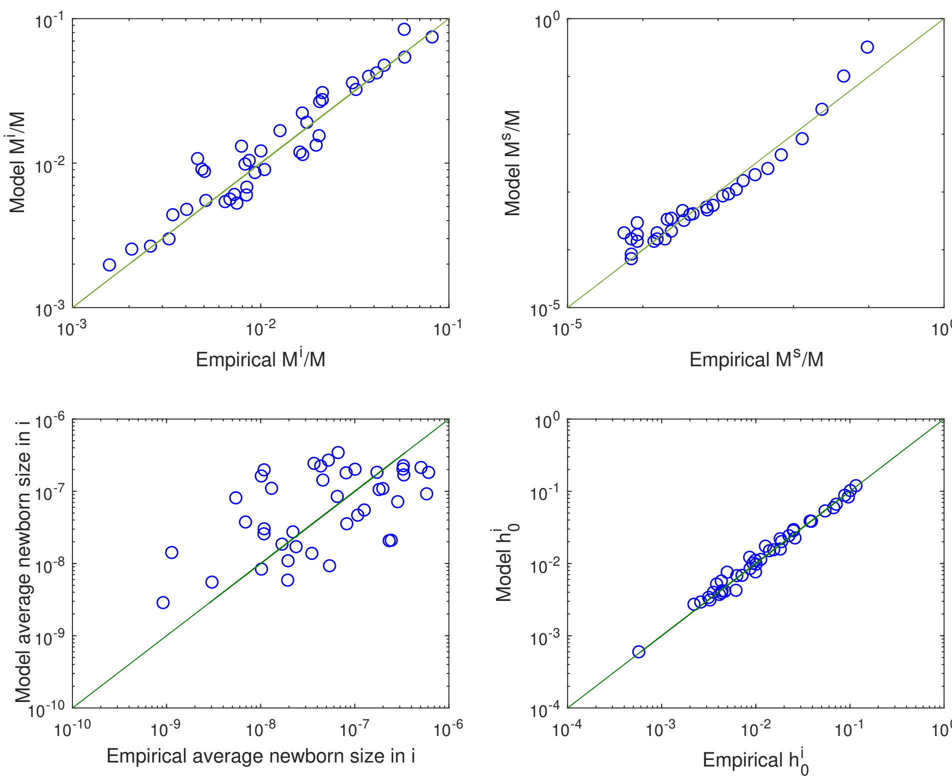
We now show how well our model fits the data by comparing the model's targeted and untargeted moments with their empirical counterparts. All numbers reported are in log-scale. First, we look at the moments that we targeted in our GMM estimation. The upper-left penal of Figure V plots the cross-sector observations of the model-generated share of firms in sector i, M^i/M, and the empirical share of firms in corresponding sectors, together with a 45-degree line signaling a perfect fit. The correlation between the two is as high as 0.96. The upper-right panel plots the share of firms that simultaneously patent in s number of sectors (M_s/M) and its empirical counterpart, the correlation is also high at 0.95. The lower two panel compares the model-generated average newborn size per sector ($\bar{\tilde{z}}^i_0$) and the model-generated entry probability of a potential newborn firm (h^i_0) with their respective empirical counterparts. The correlations are 0.54 and 0.99, respectively.

Next, we examine our model's performance in terms of three moments that were not targeted in GMM. This includes R&D allocation, R^i/R, which is predicted from (39), the external knowledge to own knowledge ratio, τ^{ij}, which has a one-to-one relationship with the share of sector j knowledge that is used sector i, $\tilde{s}^{ij}$, according to (25), and the new knowledge capital created by each firm, $\Delta\tilde{z}^i_{f,t}$, according to (12).

The left panel of Figure VI compares the model-implied share of R&D expenditure across sectors with the empirical sectoral R&D share (correlation=0.61). We also find a significant and positive

[28] For example, the average total patent stock is 32 during 1976-2006, while the smallest firm has only one patent. Therefore acquiring 0.2% of the average firm's knowledge capital by imitation increases its own knowledge capital by about 6.5%, which is sizable for small firms. In addition, when comparing the cross-firm citations per patent with self-citations per patent by calculating the ratio=(cross-firm citation$_{f,t}/\sum_{k\neq f} S_{k,t}$)/(self-citations$_{f,t}/S_{f,t}$), we find that the average ratio is 0.07% in 1997. This implies that on average firms utilize way more private knowledge than public knowledge, consistent with our estimation.

Figure V: Targeted Moments (log scale)



Notes: These figures compare the model generated variables of interest (y-axis) with their empirical counterparts (x-axis). M^i/M is the fraction of firms innovate in sector i, M^s/M is the fraction of firms that innovate in s number of sectors (firms co-patenting pattern), h_0^i stands for the fraction of newborn firms that enter sector i.

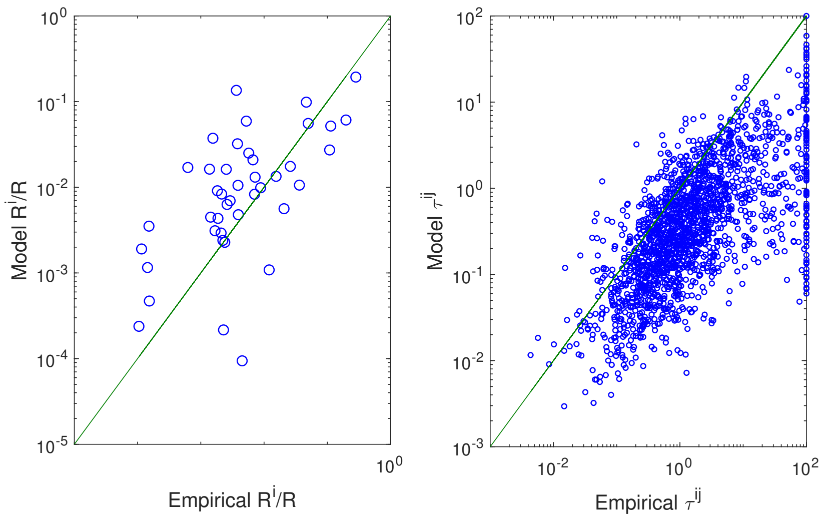
correlation between τ^{ij} predicted by the model and its empirical counterparts (correlation=0.66), presented in the right panel. Lastly, although not reported in the figure, the model-implied new knowledge created and the empirical number of new patent applications are highly correlated at 0.83.

These results indicate that the model performs well in generating cross-sector differences in R&D investment, firms co-patenting, cross-sector knowledge utilization and knowledge creation at firm level.

5.3 Model-Predicted Knowledge Value and Empirical Applicability Measure

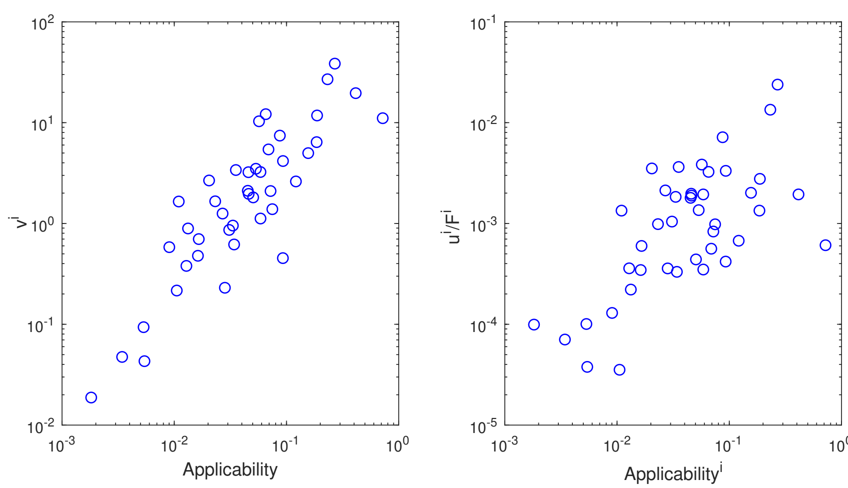
We have commented throughout the model how the cross-sector differences in v^i and u^i/F^i explain the observed sector-level and firm-level facts. While these variables are not directly observable in the data, the model predicts that they are all driven by the fundamental heterogeneity of knowledge linkages across sectors, suggesting a positive relationship between v^i, u^i and the empirical measure of applicability. Figure VII plots the estimated v^i and u^i/F^i against the empirical measure of

Figure VI: Untargeted Moments (log scale)



technology applicability constructed in Section 2. It shows that both v^i and u^i are indeed highly positively correlated with sectoral knowledge applicability (app^i). The correlation between the estimated $\log(v^i)$ and $\log(app^i)$ is 0.89. The correlation between the estimated $\log(u^i/F^i)$ and $\log(app^i)$ is 0.67, suggesting that highly applicable sector does also provide highly valuable public knowledge that attracts small firms with little private knowledge capital to enter.

Figure VII: Model-Generated Knowledge Value and Empirical Measure of Applicability (log scale)

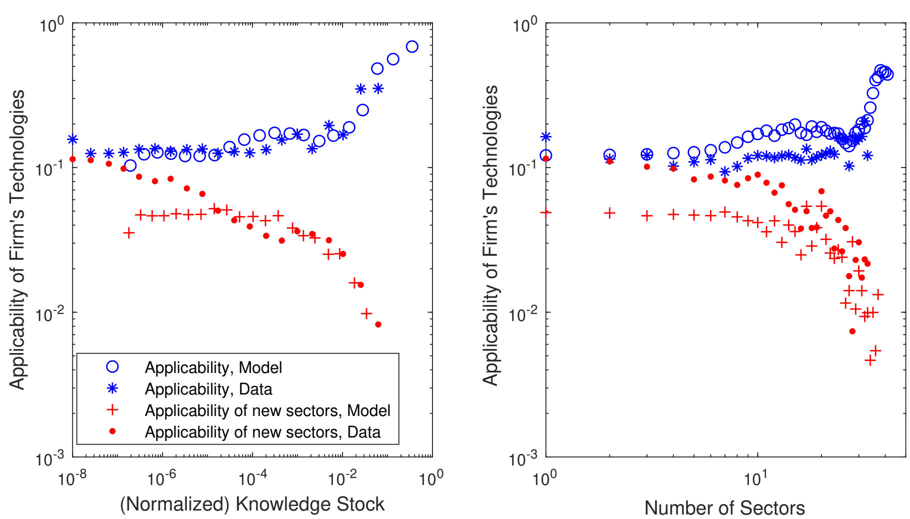


5.4 Simulated Firm-level Observations

The observation that the applicability of a firm's existing technology matters for its innovation activity is one of the key observations that motivates the presence of inter-sectoral knowledge linkages in our model. We have discussed previously in the text how our model might potentially explain these facts. This section presents the simulated observations to show that the model indeed can account for the firm-level observations. To obtaining comparability, we re-generate empirical observations in Section 2 using 42 sectors (instead of 428 Nclasses) and compare these with the model-generated observations. Details on the simulation algorithms are provided in the Computational Appendix C.

Knowledge Stock, Knowledge Scope and Technology Applicability Figure VIII provides a graphical comparison of the model-implied relationship between firm's technology applicability and its knowledge stock/scope with the empirical relationship. Similarly to Figure II in Section 2, all firms are divided into 20 bins according to their knowledge stock (patent stock in the data) in the left panel or according to their numbers of sectors in the right panel. The graphs show that the model is able to replicate closely how a firm's knowledge stock and knowledge cope matter for its technology allocation and entry. In line with the Observation 2 in Figure II, our simulation shows that firm's technology applicability (weighted average applicability of its technologies using applicability calculated for these 42 sectors) increases with its total number of innovations (knowledge stock) and its number of sectors (knowledge scope), while the knowledge applicability of its new sectors is negatively related to both. As the firm accumulates knowledge in many related sectors, it can slowly afford to enter peripheral sectors which have weaker knowledge linkages with its existing sectors. Therefore when taking a snapshot of the innovation outcome across firms, larger firms with more knowledge stock tend to enter sectors closer to the periphery.

Figure VIII: Firm's Technology Applicability, Knowledge Stock and Multi-Technology Innovation (log scale)



Notes: Similarly to Section 2, the large number of simulated firms are divided into 20 bins and observations are based on an average firm in each bin.

Firm Innovation Rate and Initial Technology Applicability Using simulated data from the last 40 periods, we examine the relationship between firm's initial TA_f and its subsequent innovation rate based on the regression specification in (3). Robust standard errors are reported, observations are clustered at the firm level. The result is given as:

$$g_{f,t} = \underset{(0.001)^{***}}{-0.061} \log(S_{f,t-1}) + \underset{(0.000)^{***}}{0.002} \log(N_{f,t-1}) + \underset{(0.001)^{***}}{0.005} TA_{f,t-1} + \mu_f + \eta_t + \upsilon_{f,t}. \tag{47}$$

Replacing the overall innovation rate by the intensive innovation rate and the extensive innovation rate as the dependent variable, we have the following regression results:

$$g^{in}_{f,t} = \underset{(0.001)^{***}}{-0.023} \log(S_{f,t-1}) + \underset{(0.000)^{**}}{0.001} \log(N_{f,t-1}) \underset{(0.001)}{-0.001} TA_{f,t-1} + \mu_f + \eta_t + \upsilon_{f,t}, \tag{48}$$

$$g^{ex}_{f,t} = \underset{(0.001)^{***}}{-0.037} \log(S_{f,t-1}) \underset{(0.000)^{***}}{+0.002} \log(N_{f,t-1}) + \underset{(0.001)^{***}}{0.006} TA_{f,t-1} + \mu_f + \eta_t + \upsilon_{f,t}. \tag{49}$$

Again, the results using simulated data are consistent with empirical observations. It is evident that the model replicates the empirical observation that firms' innovation rates, more so for extensive growth rates, are positively related to the applicability of their initial technological position. A more central positioning in the technology space opens more potential routes for a firm to expand across sectors, thus boosting the overall firm innovation rate.

5.5 Counterfactual Experiments: the Role of Fixed Costs of Innovation

One of the new insights gained from the model, as discussed in Section 4.3, is that the fixed costs of innovation in every sector potentially has negative growth impact through their "knowledge under-utilization effect" and "R&D misallocation effect". In this section, we explore the quantitatively implication of raising the mean and standard deviation of the shocks to fixed R&D costs using counterfactual experiments. The technical procedure of this exercise is given in Appendix C.2. The counterfactual results are presented in Table VII.

In the first experiment, we double the size of the average fixed cost, F^i_i, in all sectors. Since the barrier to conduct R&D is now higher, fewer firms would innovate in each sector (smaller M^i) and those that innovate would simultaneously innovate in fewer sectors (smaller M^s). The average number of sectors per firm drops from the baseline 0.76 to 0.51. Cross-sector knowledge application sightly decreased with the average s^{ij} falling from 0.50 to 0.49. The correlation between cross-sector knowledge licensing share δ^{ij} and knowledge linkages ω^{ij} also decrease, from 0.50 in the baseline to 0.28. As a consequence, aggregate innovation rate, g, falls from 11% to 9.3%. Meanwhile, overall firm size distribution becomes even more homogeneous: The standard deviation of log-scale (normalized) total firm size $\sum_i \tilde{z}^i_{f,t}$ decreases from 2.58 to 0.53. This is because relatively large firms with applicable background knowledge find it harder to enter new sectors and grow further, as the per sector fixed costs of R&D rise. Because the peripheral sectors' F^is are higher than those of the central sectors to begin with, when doubling all F^i, the likelihood to enter drops disproportionally more in the peripheral sectors.

In the second experiment, we quintuple the standard deviation of idiosyncratic shocks to sectoral fixed costs, σ_ς, while keeping the average number of sectors per firm the same. Firms now are distributed more evenly across sectors (less cross-sectoral variation in M^i). Cross-sector knowledge application dramatically decreased, reflected in an significant drop in the average s^{ij} (from 0.50 to 0.08). Meanwhile, the correlation between log-scaled cross-sector knowledge licensing share δ^{ij} and knowledge linkage ω^{ij} falls to 0.11, because the sectoral selection is less based on firm's background of knowledge and the associated applicability and more on luck. Eventually, growth rate is lower, dropping from 11% to 1.6%. Moreover, firm size distribution becomes more homogeneous, with the standard deviation of log-scale firm size lowering from 2.58 to 1.73. This is simply an outcome of sectoral selection becoming more random. Large firms with more related knowledge capital are less likely to keep their position and expand.

Table VII: Counterfactual Experiments

Scenario	$\bar{M}^s$	$\bar{s}^{ij}$	$\bar{\rho}(\log \delta^{ij}, \log \omega^{ij})$	$\bar{\sigma}(\log(\sum_i \tilde{z}^i_{f,t}))$	g
Baseline	0.76	0.50	0.51	2.58	1.11
$2F^i$	0.51	0.49	0.49	0.53	1.09
$5\sigma_\zeta$	0.76	0.08	0.11	1.73	1.02

6 Concluding Remarks and Directions for Future Research

Technological advances are complementary and sequential; interconnections between them are, however, highly heterogeneous. Our goal is to forge a link between observations of firm innovation in multiple technologies and theories of aggregate technological progress. We provide a theoretical framework which builds on micro-level observations and helps to elucidate how innovating firms choose to position themselves in the technology space and allocate their R&D investment. We have attempted to demonstrate that our model can replicate key firm-level facts; as such, the resulting aggregate model is likely to provide a more credible tool for policy analysis. As an example, in two counterfactual experiments, we show that both larger fixed costs and higher degree of randomness in the fixed costs shocks reduce growth by lowering the correlation between cross-sector knowledge utilization and the underlying knowledge linkages.

Our study has important implications for economic growth and R&D policies. First, government policies directed at stimulating innovation in certain technologies need to be based on better understanding of the inter-sectoral knowledge linkages. Heterogenous sectoral knowledge spillovers suggest that industrial or R&D policies that favor highly applicable sectors may boost growth. In a related cross-country study, Cai and Li (2013) find that countries which specialize more in applicable technologies tend to grow faster. Second, policies that lower barriers to diversification help to reinforce the effect of industrial policies, as it can be challenging to shift to more advanced industries given the fixed cost of learning and adapting knowledge in new sectors, and more diversification encourages spillovers between different technologies. Third, competition policies that encourage joint R&D ventures in highly related sectors can benefit growth, because firms are better able to internalize knowledge spillovers.

One direction for future research is to relax the assumption of an efficient and competitive knowledge licensing market, which is assumed in this paper to allow for the basic tractability of the model. Relaxing this assumption would introduce another type of heterogeneity in that innovating firms would bid a firm-specific price for a given knowledge capital depending on the set of sectors that a firm innovates in. This is certainly an intriguing case. If firms price the same knowledge stock differently, a current all-sector firm is likely to bid a higher price as it is able to apply knowledge in greater scope. Therefore, large firms is likely to invest relatively more in R&D and are more active in the knowledge market than small firms. In equilibrium, small firms would find it difficult to survive and expand, but once succeed, they are more likely to defend their status. The firm dynamics would become more persistent and the firm size distribution is likely to be more heterogeneous. A greater share of knowledge stock would then be concentrated in a few superstar firms who are able to apply it in almost all sectors. Unfortunately, this scenario is not feasible computationally in the current setup given the already complicated heterogeneity in multiple dimensions (i.e. at the sector level, at the sector-pair level, at the firm level and at the firm-sector level). Each firm is a multi-dimensional vector as it creates various numbers of blueprints in multiple sectors. Therefore the total number of state vectors is extremely large, implying an enormous size of the transition matrix across different states. If realistic features of inefficient and illiquid knowledge markets were to be incorporated in this model, one needs to make other assumptions to reduce the dimensions of heterogeneity, such as allowing for random and exogenous arrival of new innovation. However such a setup would not be able to explain the endogenous and path-dependent sectoral selection in firm's R&D allocation decisions, which is the focus and novelty of this paper, but could be an interesting venue for future exploration.

References

[1] Acemoglu, D. and D. Cao. 2015. "Innovation by Entrants and Incumbents", *Journal of Economic Theory*, 157:255-294.

[2] Acemoglu, D, Carvalho, V., Ozdaglar and A. Tahbaz-Salehi. 2012. "The Network Origins of Aggregate Fluctuations", *Econometrica*, 80(5):1977-2016.

[3] Aghion P. and P. Howitt. 1992. "A Model of Growth Through Creative Destruction", *Econometrica*, 60(2):323-351.

[4] Akcigit U. and W. Kerr. 2016. "Growth Through Heterogeneous Innovations", *Journal of Political Economy*, 124(1):52-104.

[5] Akcigit U., D. Hanley and N. Serrano-Velarde. 2016. "Back to Basics: Basic Research Spillovers, Innovation Policy and Growth", *CEPR Discussion Paper*, 11707.

[6] Atkeson, A. and A. Burstein. 2010. "Innovation, Firm Dynamics and International Trade", *Journal of Political Economy*, 118(3):433-484.

[7] Axtell, R.. 2001. "Zipf Distribution of U.S. Firm Sizes", *Science*, 293(5536):1818-1820.

[8] Alvarez, F. E., Buera, F. J., and Lucas Jr, R. E. 2014. "Idea flows, economic growth, and trade," NBER working paper, No. 19667.

[9] Balasubramanian, N. and J. Sivadasan. 2011. "What Happens When Firms Patent? New Evidence from U.S. Economic Census Data". *Review of Economics and Statistics*, 93(1):126-146.

[10] Bartelme, D. and Y. Gorodnichenko. 2015. "Linkages and Economic Development". National Bureau of Economic Research (No. w21251).

[11] Benhabib J., A. Bisin and S. Zhu. 2011. "The Distribution of Wealth and Fiscal Policy In Economics With Finitely Lived Agents". *Econometrica*, 79(1):123-157.

[12] Benhabib, J., A. Bisin and S, Zhu. 2015. "The wealth distribution in Bewley economies with capital income risk". *Journal of Economic Theory*, 159: 489-515.

[13] Bernard, A., S. J. Redding and P. K. Schott. 2010. "Multi-Product Firms and Product Switching". *American Economic Review*, 100(1):70-97

[14] Bernard, A., S. J. Redding and P. K. Schott. 2009. "Product and Productivity". *NBER working paper* No.11575.

[15] Bernstein and Nadiri. 1988. "Interindustry R&D Spillovers, Rates of Return, and Production in High-Tech Indintersectoralustries". *American Economic Review*. 78(2):429-439.

[16] Bloom, N., M. Schankerman and J. V. Reenen. 2010. "Identifying Technology Spillovers and Product Market Rivalry". *Econometrica*, 81(4):1347–1393.

[17] Bresnahan, T., and M. Trajtenberg. 1995. "General Purpose Technologies: Engines of Growth?" *Journal of Econometrics*, 65(1):83-108.

[18] Brada, C. and D. E. Weinstein. 2006. "Globalization and The Gains From Variety", *Quarterly Journal of Economics*, 121(2):541-585.

[19] Cai, J. 2010. “Knowledge Spillovers and Firm Size Heterogeneity”. Shanghai University of Finance and Economics, Working Paper.

[20] Cai, J. and N. Li. 2013. “The Composition of Knowledge and Long-Run Growth In a Path-dependent World”, Shanghai University of Finance and Economics, Working Paper.

[21] Cohen, W. M., R. C. Levin and D. C. Mowery. 1987. “Firm Size and R&D Intensity: A Re-Examination.”, *Journal of Industrial Economics*, 35(4): 543-565.

[22] Cohen, W. and D. Levinthal. 1990. “Absorptive Capacity: A New Perspective on Learning and Innovation”, *Administrative Science Quarterly*, 35(1):128-152.

[23] David, P. 1990. “The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox”. *American Economic Review*, 80(2):355-361.

[24] Gabaix, X. 2009. “Power Laws in Economics and Finance”. *Annual Review of Economics*, 1:255-294.

[25] Grossman, G. and E. Helpman. 1991a. “Innovation and Growth in the Global Economy”. Cambridge, MA: MIT Press.

[26] Grossman, G. and E. Helpman. 1991b. “Quality Ladders in the Theory of Growth”. *Review of Economic Studies*, 68:43-61.

[27] Hall, B, A. Jaffe, and M. Trajtenberg. 2001. “The NBER Patent Citations Data File: Lessons, Insights, and Methodological Tools”. NBER Working Paper (No. 8498).

[28] Hall, B, A. Jaffe, and M. Trajtenberg. 2005. “Market Value and Patent Citations”. *Rand Journal of Economics* 36:16-38

[29] Hausmann, R., J. Hwang and D. Rodrik. 2007. “What You Export Matters”. *Journal of Economic Growth*, 12(1):1-25.

[30] Hay, D., R. Rastegar, and A. Roitershtein. 2011. “Multivariate linear recursions with Markov-dependent Coefficients”. *Journal of Multivariate Analysis* 102(3):521-527.

[31] Helpman, E. (ed.). 1998. “General Purpose Technologies and Economic Growth”. Cambridge and London: MIT Press.

[32] Hidalgo, C. A., B. Klinger, A.-L. Barabasi and R. Hausmann. 2007. “The Product Space Conditions the Development of Nations”. *Science*, 317:482-487.

[33] Hirschman, A. O. 1958. “The Strategy of Economic Development”. Yale University Press.

[34] Hopenhayn, H. 1992. “Entry, Exit and Firm Dynamics in Long Run Equilibrium”. *Econometrica*, 60(5):1127-50.

[35] Horvath, M. 1998. “Cyclicality and Sectoral Linkages: Aggregate Fluctuations from Independent Sectoral Shocks”. *Review of Economic Dynamics*, 1:781-808.

[36] Garfield, E. 1972. “Citation Analysis as A Tool in Journal Evaluation”. *Science*, 178:471-479.

[37] Jovanovic B. and P. L. Rousseau. 2005. “General Purpose Technologies”. In Philippe Aghion and Steven Durlauf (Eds.), *Handbook of Economic Growth*, 1:1181-1224.

[38] Jeffe, A. 1986. "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value". *American Economic Review*, 76(5):984-1001.

[39] Jones, C. I. 2011. "Intermediate Goods and Weak Links in the Theory of Economic Development". *American Economic Journal: Macroeconomics*, 3(2):1-28.

[40] Jones, C. I. 1995. "R&D–Based Models of Economic Growth", *Journal of Political Economy*, 103(4):759-784.

[41] Jones, C. I. 1999. "Growth: With or without scale effects?" *Papers and Proceedings of the One Hundred Eleventh Annual Meeting of the American Economic Association*, 89:139-144.

[42] Kesten, H. 1973. "Random Difference Equations and Renewal Theory for Products of Random Matrices". *Acta Mathematica*, 131:207-248.

[43] Kleinberg, R. 1999. "Authoritative Sources in a Hyperlinked Environment", *Journal of Association for Computing Machinery*, 46(5):604-632.

[44] Klenow, P. 1996. "Industry Innovation: Where and why?" *Carnegie-Rochester Conference Series on Public Policy*, 44:125-150.

[45] Klette, T. J. and S. Kortum. 2004. "Innovating Firms and Aggregate Innovation". *Journal of Political Economy*, 112:986-1018.

[46] Landes, D. 1969. "The unbound Prometheus". Cambridge University Press.

[47] Levin, R. C., Cohen, W.M., and Mowery, D.C. 1985. "R&D Appropriability, Opportunity, and Market Structure: New Evidence on Some Schumpeterian Hypotheses", *American Economic Review,* 75:20-24.

[48] Levin, R. C., A. K. Klevorick, R. R. Nelson, S.G. Winter, R. Gilbert and Z. Griliches. 1987. "Appropriating the Returns from Industrial Research and Development". *Brookings Papers on Economic Activity*, 1987(3):783-831.

[49] Leontief, W. 1936. "Quantitative Input and Output Relations in the Economic System of the United States". *Review of Economics and Statistics*, 18(3):105-125.

[50] Lucas, R. E. 1981. "Understanding Business Cycles". In *Studies in Business Cycle Theory.* Cambridge, MA: MIT Press.

[51] Luttmer, E. G. 2007. "Selection, Growth, and the Size Distribution of Firms". *Quarterly Journal of Economics*, 122(3):1103-1144.

[52] Luttmer, E. G. 2012. "Technology Diffusion and Growth". *Journal of Economic Theory*, 147:602-622.

[53] Mirek, M., 2011. "Heavy tail phenomenon and convergence to stable laws for iterated Lipschitz maps.". *Probability Theory and Related Fields*, 151(3-4): 705-734.

[54] Monge-Naranjo A. 2012. "Foreign Firms and the Diffusion of Knowledge". Federal Reserve Bank of St. Louis Working Paper No. 2012-055A.

[55] Nelson, R. and S. Winter. 1977. "In Search of Useful Theory of Innovation". *Research Policy*, 6:36-76.

[56] Nelson, R. 1988. "Modelling the connections in the Cross Section between Technical Progress and R&D Intensity." *Rand Journal of Economics*, 19(3): 478-485.

[57] Ngai, R. and R. Samiengo. 2011. "Accounting for Research and Productivity Growth Across Industries". *Review of Economic Dynamics*, 14(3):475-495.

[58] Peretto, P. F. 1998. "Technological Change and Population Growth". *Journal of Economic Growth* 3.4: 283-311.

[59] Pinski, G., and F. Narin. 1976. "Citation Influence of Journal Aggregates of Scientific Publications: Theory, with Applications to the Literature of Physics". *Information Processing and Management*, 12:297-312.

[60] Romer, P. M. 1986. "Increasing returns and long-run growth". *Journal of political economy* 94.5:1002-1037.

[61] Romer, P. M. 1990. ""Endogenous Technological Change". *Journal of Political Economy*, 98:71-102.

[62] Rosenberg, N. 1982. "Inside the black box: Technology and economics". Cambridge University Press.

[63] Rossi-Hansberg E. and Wright M. 2007. "Establishment Size Dynamics in the Aggregate Economy". *American Economic Review*, 97(5):1639-1666.

[64] Saporta, B. 2005. "Tail of the Stationary Solution of the Stochastic Equation $Y_{n+1} = a_n Y_n + \gamma_n$ with Markovian Coefficients". *Stochastic Processes and their Applications* 115(12): 1954-1978.

[65] Segerstrom, P. S. , T. C. A. Anant, and E. Dinopoloulos. 1990. "A Schumpeterian Model of the Product Life Cycle". *American Economic Reviews*, 80(5):1077-1091.

[66] Stelzer, R. (2008). "Multivariate Markov-switching ARMA Processes with Regularly Varying Noise". *Journal of Multivariate Analysis*, 99(6), 1177-1190.

[67] Trajtenberg, M., Jaffe, A., and R. Henderson. (1997). "University versus Corporate Patents: A Window on the Basicness of Invention". *Economics of Innovation and New Technology*, 5 (1): 19-50.

[68] Wieser, R. 2005. "Research and Development Productivity and Spillovers: Empirical Evidence at the Firm Level". *Journal of Economic Survey*, 19(4): 587-621.

A Empirics Appendix

A.1 Data Sources

Firm Patenting and Patent Citations We use patent applications in the 2006 edition of the NBER Patent Citation Data (see Hall, Jaffe and Trajtenberg, 2001 for details) to characterize firms' innovation activities. We use their citations to trace the direction and intensity of knowledge flows and to construct indices of knowledge linkages among sectors. The data provides detailed information of every patent granted by the United States Patent and Trade Office (USPTO) and their citations from 1976 to 2006. We summarize each firm's patent stock in each disaggregated technological class (intensive margin of innovation) and the number of categories (extensive margin of innovation) for each year.

Each patent corresponds to one of the 428 3-digit United States Patent Classification System (USPCS) technological classes and also one of more than 800 7-digit International Patent Classification (IPC) classes. Figure I presents the intersectoral network corresponding to the patent citation share matrix for 428 technological category. We mostly report the results based on USPCS codes, but we check for robustness using the IPC classes. We also present some evidence based on industrial sector classification, as the model is estimated based on this categorization. To translate the data into the industrial classifications, we use the 2005 edition of the concordance table provided by the USPTO to map USPCS into SIC72 (Standard Industrial Classification in 1972) codes, which constructs 42 industrial sectors.[29] We summarize citations made to patents that belong to the same technological class and use the cross-sector citations to form the intersectoral knowledge spillover network.

Firm R&D Data Information on firm sizes (i.e. sales) and firm's R&D expenditure is from the U.S. Compustat database. Firm-level R&D intensity is defined as R&D expenditure divided by sales. We consider three measures of the industry-level R&D intensity: industry aggregate R&D expenditure divided by industry aggregate sales, median firm R&D intensity and average firm R&D intensity.

A.2 Robustness of Empirical Observations

Time-variant Measure of Applicability In order to capture the fundamental long-run linkages between different technological fields, we construct the sector-specific measure of applicability using 30-year pooled patent citation data. However, knowledge linkages may be formed due to the state of scientific knowledge at a certain point in time and thus may change over time. Especially, the interconnectedness of technologies relevant to firms may also be time-varying. To test if this concern

[29] The patents are classified according to either the intrinsic nature of the invention or the function for which the invention is used or applied. It is inherently difficult to allocate the technological category to economically relevant industries in a differentiation finer than 42 sectors, even with detailed firm level information. First, most of the patents are issued by multi-product firms that are present in multiple SIC-4 industries. Second, in the best scenario, one only has industry information about the origin of the patents but not the industry to which the patent is actually applied.

would affect our observation, we compute a rolling-window based measure of applicability for each sector, $\log(app_t^i)$ using data from $[t-10, t-1]$, and run the same regressions using this time-variant applicability. The results are reported in Column "Time-variant" in the following Tables A.1–A.3, corresponding to the original results in Table IV –VI in Section 2.

Quality-Adjusted Measure of Applicability Not all patents are created equal. The economic impacts of individual patented inventions have demonstrated large heterogeneity (Hall et al. 2007, Harhoff et al. 1999). To handle this concern, we follow the suggestion by Hall et al. (2007) and use forward citations received by a patent as a proxy for its quality to adjust our measure of applicability. Specifically, when calculating aw using (1), we weigh a citation from patent p in sector j by the number of forward citations received by p. That is, $W^{ij} = \sum_{p \in \{\text{patents citing } i\}} C_p^j$, where C_p^j is the number of forward citations received by p and p is a patent in sector j. For example, if patent p received 5 forward citations, then its citation weight count as 5 instead of 1. In this way, the citation received from a patent which itself is well-cited counts more. Column "Quality-adj" in Tables A.1–A.3 reports results using this measure.

Depreciated Patent Stock The R&D literature often assumes that R&D capital stocks depreciate (with a typical annual rate of 15%) (e.g. Hall et al. (2005)). In our paper, patent stock is used as a measure for firm's knowledge stock. Although it is also possible for ideas to depreciate, the paper does not assume physical depreciation of knowledge—as that would imply that some knowledge is exogenously forgotten. However the model does allow for economic depreciation as knowledge can become less valuable (lower market share) when newer knowledge accumulates in the same sector. Nevertheless, to test that if depreciation of knowledge capital would change our empirical results, we assume the same 15% depreciation rate to recalculate firm's patent stock, i.e. $S_{f,t} = 0.85 \times S_{f,t-1} + P_{f,t}$ and reconstruct firms' knowledge portfolio. Column "Depreciated S_f" in Tables A.1–A.3 presents the results.

Overall, all our firm-level observations are robust to these alternative measure of applicability or firms' knowledge portfolio.

Table A.1: Robustness: Firm's Innovation Allocation

Dependent Variable: $TA_{f,t}$	Time-variant		Quality-Adjusted		Depreciated S_f	
	(1)	(2)	(3)	(4)	(5)	(6)
$\log(S_{f,t-1})$	0.137		0.0056		-0.000	
	(0.006)**		(0.007)**		(0.004)	
$\log(N_{f,t-1})$		0.114		0.127		0.039
		(0.008)**		(0.010)**		(0.007)**
Dependent: $TA^{newsec}_{f,t-1}$	(1)	(2)	(3)	(4)	(5)	(6)
$\log(S_{f,t-1})$	-0.252		-0.274		-0.310	
	(0.004)**		(0.005)**		(0.003)**	
$\log(N_{f,t})$		-0.381		-0.421		-0.330
		(0.006)**		(0.006)**		(0.004)**
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Regressions include firm and year fixed effects. Regression coefficients are reported, with robust standard errors adjusted for clustering by firms in brackets. Sample covers every year between 1976 and 2006. ** and * indicate significance at the 1 percent level and 5 percent level respectively.

Table A.2: Robustness: Firm's Sectoral Entry Selections

Dependent Variable: $\log(app^i_{f,t})$	Time-variant		Quality-Adjusted		Depreciated S_f	
	(1)	(2)				
$Newsec^i_{f,t}$	0.044	0.036	0.052	0.053	0.018	0.061
	(0.004)**	(0.004)**	(0.004)**	(0.004)**	(0.004)**	(0.006)**
$\log(S_{f,t-1})$	0.094		0.029		-0.049	
	(0.007)**		(0.009)*		(0.005)**	
$Newsec^i_{f,t} \times \log(S_{f,t-1})$	-0.070		-0.069		-0.108	
	(0.002)**		(0.003)		(0.004)**	
$\log(N_{f,t-1})$		0.102		0.046		0.107
		(0.010)**		(0.012)**		(0.008)**
$Newsec^i_{f,t} \times \log(N_{f,t-1})$		-0.094		-0.100		-0.160
		(0.004)**		(0.004)**		(0.005)**
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Regressions include firm and year fixed effects. Regression coefficients are reported, with robust standard errors adjusted for clustering by firms in brackets. Sample covers every year between 1976 and 2006. ** and * indicate significance at the 1 percent level and 5 percent level respectively.

Table A.3: Robustness: Firm Innovation Rate

Dependent Variable: g	Time-variant		Quality-Adjusted		Depreciated S_f	
	Main	Selection	Main	Selection	Main	Selection
$TA_{f,t-1}$	0.033	0.048	0.054	0.057	0.111	0.047
	(0.002)**	(0.003)**	(0.002)**	(0.003)**	(0.005)**	(0.003)**
$\log(S_{f,t-1})$	-0.345	1.065	-0.339	1.068	-0.353	1.043
	(0.006)**	(0.010)**	(0.006)**	(0.010)**	(0.013)**	(0.006)**
$\log(N_{f,t-1})$	0.131	0.113	0.128	0.109	0.636	0.544
	(0.007)**	(0.013)**	(0.007)**	(0.013)**	(0.012)**	(0.009)**
age		-0.067		-0.067		-0.022
		(0.001)**		(0.001)**		(0.000)**
Year FE	Yes		Yes		Yes	
Firm FE	Yes		Yes		Yes	

Notes: Regressions are based on Heckman two-step procedure and include firm and year fixed effects. Regression coefficients are reported, with robust standard errors adjusted for clustering by firms in brackets. Sample covers every year between 1976 and 2006. ** and * indicate significance at the 1 percent level and 5 percent level respectively.

B Technical Appendix

B.1 Identical Innovation Rates Across Sectors

This section shows that on the BGP, the innovation rates across sectors are the same (i.e. $g = g^i$, $\forall i$), as long as the intersectoral knowledge linkage $A = [A^{ij}]_{i \times j \in J \times J}$ satisfies the following condition: $\exists A^{ij} > 0, \forall i$. That is, every sector is applying knowledge from at least another sector.

Based on the knowledge accumulation equation (11) and (12), the evolution of the number of blueprints in sector i can be derived as:

$$\begin{aligned} n^i_{t+1} &= n^i_t + \int\limits_{f \in \mathcal{F}^i_t} \triangle z^i_{f,t} df \\ &= n^i_t + \sum_{j \in \mathcal{J}} (A^{ij})^{\frac{1}{1-\alpha}} \left(\frac{\alpha \rho^i v^i}{M^i} \right)^{\frac{\alpha}{1-\alpha}} \left[\int\limits_{f \in \mathcal{F}^i_t} ((1 + \kappa^{ij} \tau^{ij}) z^j_{f,t} + \theta \bar{z}^j_t) df \right] \\ &= n^i_t + \left(\frac{\alpha \beta v^i}{g^i M^i} \right)^{\frac{\alpha}{1-\alpha}} \sum_{j \in \mathcal{J}} (A^{ij}) \delta^{ij} n^j_t \end{aligned} \tag{B.1}$$

where $\delta^{ij} = \tilde{s}^{ij} + \kappa^{ij}(1 - \tilde{s}^{ij}) + \theta \frac{M^i}{M^j}$ stands for the fraction of knowledge in sector j (including both private and public) that is ultimately utilized in innovation in sector i. It comprises of three component: $\tilde{s}^{ij} = \int_{f \in \mathcal{F}} H^i(\varphi^i_f) \tilde{z}^j_f df$ is the fraction of private sector-j knowledge that are used by its own inventor in sector i, $\kappa^{ij}(1 - \tilde{s}^{ij})$ is the fraction of sector-j private knowledge that are acquired and absorbed by other firms who innovate in sector i, and the last term in the last bracket ($\theta \frac{M^i}{M^j}$) represents the fraction public knowledge in sector j that is utilized for innovation in sector i. All components are constant in the BGP. Since acquired knowledge is not fully absorbed, $\tilde{s}^{ij} + \kappa^{ij}(1 - \tilde{s}^{ij}) \leq 1$.

Rearranging the terms in (B.1), we have:

$$(g^i - 1)(g^i)^{\frac{\alpha}{1-\alpha}} = \left(\frac{\alpha \beta v^i}{M^i} \right)^{\frac{\alpha}{1-\alpha}} \sum_{j \in \mathcal{J}} \left(A^{ij} \right)^{\frac{1}{1-\alpha}} \left[\tilde{s}^{ij} + \kappa^{ij}(1 - \tilde{s}^{ij}) + \theta \frac{M^i}{M^j} \right] \frac{n^j_t}{n^i_t}. \tag{B.2}$$

Now suppose that sector i had been growing more slowly than other sectors for a lengthy period, its number of goods would be extremely small relative to other sectors. (B.2) implies that the cross-sector knowledge spillovers would increase g^i tremendously through a large ratio of n^j_t / n^i_t until g^i is the same as the innovation rates in other sectors. And vice versa for sectors starting with a faster growth rate than others. Therefore, in the BGP, $g^i = g^j = g$. Since the number of varieties innovated grow at the same rate across sectors, n^j_t / n^i_t is constant. That is,

$$\frac{n^i_t}{n^j_t} = \frac{n^i}{n^j}, \forall t.$$

Thus the distribution of sector sizes is stable and rank-preserving. Intuitively, the number of goods in every sector grows at the same speed, because inter-sector knowledge linkages keep all sectors on the same track.

B.2 Deriving Aggregate Growth Rate and Cross-Sector Research Intensity

Given that all sectors grow at the same rate, we can rewrite (B.2) as

$$g = 1 + \sum_{j\in\mathcal{J}} \frac{\hat{\omega}^{ij}\delta^{ij}}{(1-\alpha)\rho v^i}, \tag{B.3}$$

Based on (14), we can rewrite the equation above as:

$$g = 1 + \frac{(1-\rho)}{(1-\alpha)\rho} \frac{\sum_{j\in\mathcal{J}}\sum_{j\in\mathcal{J}} \hat{\omega}^{ij}\delta^{ij}}{\sum_{j\in\mathcal{J}} \pi^i + \sum_{j\in\mathcal{J}}\sum_{j\in\mathcal{J}} \hat{\omega}^{ij}}. \tag{B.4}$$

Substituting out $\rho = \beta/g$ leads to (44) after rearranging the terms.

The sectoral research intensity is defined as the overall sectoral R&D expenditure divided by sectoral revenue: $RI^i \equiv \frac{1}{s^i PY} \sum_{j=1}^{K} \int_{f\in\mathcal{F}^i\cap\mathcal{F}^j} R_f^{ij} df$. Substitute the optimal R&D expenditure (17) and (B.3) into the equation, we have:

$$\begin{aligned} RI^i &= \frac{\alpha}{1-\alpha}\frac{1}{s^i PY}\sum_{j\in\mathcal{J}} \hat{\omega}^{ij} \frac{\int_{f\in\mathcal{F}^i}[\left(1+\kappa^{ij}\tau^{ij}\right) z_{f,t}^j + \theta^j \bar{z}_t^j]df}{n_t^j} \\ &= \frac{\alpha}{1-\alpha}\frac{1}{s^i PY}\sum_{j\in\mathcal{J}} \hat{\omega}^{ij}\delta^{ij}. \end{aligned}$$

Combining the above equation with (B.3) yields the R&D intensity in sector i:

$$RI^i = \alpha\rho(g-1)\frac{v^i}{s^i PY}. \tag{B.5}$$

Therefore,

$$\frac{RI^i}{RI^k} = \frac{v^i/s^i}{v^k/s^k},$$

B.3 The Evolution of (Normalized) Firm Size

Based on knowledge accumulation (11), knowledge production (12) and optimal R&D investment (17), firm f accumulates its knowledge in sector i according to

$$\begin{aligned} z_{f,t+1}^i &= z_{f,t}^i + \sum_{j\in\mathcal{J}}[A^{ij}\left(\bar{z}_t^i R_{1f,t}^{ij}\right)^{\alpha}\left(\left(1+\kappa^{ij}\tau^{ij}\right) z_{f,t}^j\right)^{1-\alpha}\varepsilon_{f,t}^{ij} + \theta A^{ij}\left(\bar{z}_t^i R_{2f,t}^{ij}\right)^{\alpha}\left(\bar{z}_t^j\right)^{1-\alpha}] \\ &= z_{f,t}^i + I_{f,t}^i \sum_{j\in\mathcal{J}} A^{ij}\left(\bar{z}_t^i \frac{\alpha}{1-\alpha}\frac{\hat{\omega}^{ij}}{n_t^j}\right)^{\alpha}\left(\left(1+\kappa^{ij}\tau^{ij}\right) z_{f,t}^j \varepsilon_{f,t}^{ij} + \theta\bar{z}_t^j\right). \end{aligned}$$

Dividing both sides by n^i_{t+1}, we can write the dynamics of (normalized) firm size ($\tilde{z}^i_{f,t+1} = z^i_{f,t+1}/n^i_{t+1}$) as

$$\begin{aligned}
\tilde{z}^i_{f,t+1} &= \frac{n^i_t}{n^i_{t+1}}\frac{z^i_{f,t}}{n^i_t} + I^i_{ft}\frac{n^i_t}{n^i_{t+1}}\sum_{j\in\mathcal{J}}\frac{\left(1+\kappa^{ij}\tau^{ij}\right)z^j_{f,t}}{n^j_t}\frac{n^j}{n^i}\left[A^{ij}\left(\frac{A^{ij}\alpha\rho v^i}{M^i}\right)^{\frac{\alpha}{1-\alpha}}\varepsilon^{ij}_{f,t}\right] \\
&\quad + I^i_{ft}\frac{n^i_t}{n^i_{t+1}}\sum_{j\in\mathcal{J}}\frac{\theta\bar{z}^j_t}{n^i_t}\left[A^{ij}\left(\frac{A^{ij}\alpha\rho v^i}{M^i}\right)^{\frac{\alpha}{1-\alpha}}\right] \\
&= \frac{1}{g}\tilde{z}^i_{f,t} + I^i_{ft}\sum_{j\in\mathcal{J}}\left(1+\kappa^{ij}\tau^{ij}\right)\tilde{z}^j_{f,t}\frac{1+r}{(1-\alpha)v^i}\left[\frac{1-\alpha}{\alpha}\frac{n^j}{n^i}(A^{ij}\rho\alpha v^i)^{\frac{1}{1-\alpha}}\left(M^i\right)^{\frac{\alpha}{\alpha-1}}\right]\varepsilon^{ij}_{f,t} \\
&\quad + I^i_{ft}\sum_{j\in\mathcal{J}}\frac{\theta}{M^j}\frac{1+r}{(1-\alpha)v^i}\left[\frac{1-\alpha}{\alpha}\frac{n^j}{n^i}(A^{ij}\rho\alpha v^i)^{\frac{1}{1-\alpha}}\left(M^i\right)^{\frac{\alpha}{\alpha-1}}\right] \\
&= \left[\frac{1}{g}\tilde{z}^i_{f,t} + I^i_{ft}\sum_{j\in\mathcal{J}}\left(1+\kappa^{ij}\tau^{ij}\right)\frac{(1+r)\hat{\omega}^{ij}}{(1-\alpha)v^i}\varepsilon^{ij}_{f,t}\tilde{z}^j_{f,t}\right] + I^i_{ft}\sum_{j\in\mathcal{J}}\frac{\theta}{M^j}\frac{(1+r)\hat{\omega}^{ij}}{(1-\alpha)v^i}.
\end{aligned}$$

Define $\xi^{ij} \equiv \frac{(1+r)\hat{\omega}^{ij}}{(1-\alpha)v^i}$, and $\phi^{ij}_{f,t}(\varepsilon^{ij}_{f,t}) \equiv \left[\frac{1_{\{\text{if } i=j\}}}{g} + I^i_{ft}(1+\kappa^{ij}\tau^{ij})\xi^{ij}\varepsilon^{ij}_{f,t}\right]$, $\psi^{ij}_{f,t} \equiv I^i_{ft}\xi^{ij}$, we can rewrite the above equation as in (29):

$$\tilde{\boldsymbol{z}}_{f,t+1} = \boldsymbol{\Phi}_{f,t+1}\tilde{\boldsymbol{z}}_{f,t} + \boldsymbol{\Psi}_{f,t+1}\boldsymbol{b}.$$

where $\tilde{\boldsymbol{z}}_{f,t} \equiv (\tilde{z}^1_{f,t},\ldots,\tilde{z}^K_{f,t})'$, the constant vector $\boldsymbol{b} \equiv (\theta/M^1,\ldots,\theta/M^K)'$.

As both $\boldsymbol{\Phi}_{f,t+1}$ and $\boldsymbol{\Psi}_{f,t+1}$ depend on firm's binary decision of whether to innovate in sector i in period t ($I^i_{f,t}(\tilde{\boldsymbol{z}}_{f,t})$), they themselves are function of $\tilde{\boldsymbol{z}}_{f,t}$. We can thus rewrite (29) as a recursive process:

$$\tilde{\boldsymbol{z}}_{f,t+1} = \boldsymbol{\Gamma}_\varepsilon(\tilde{\boldsymbol{z}}_{f,t}) = \boldsymbol{\Phi}_{f,t+1}(\tilde{\boldsymbol{z}}_{f,t},\boldsymbol{\varepsilon}_{f,t},\boldsymbol{\zeta}_{f,t},)\tilde{\boldsymbol{z}}_{f,t} + \boldsymbol{\Psi}_{f,t+1}(\boldsymbol{\zeta}_{f,t},\tilde{\boldsymbol{z}}_{f,t})\boldsymbol{b}. \tag{B.6}$$

Since the right-hand side of (B.6) is not linear in $\tilde{\boldsymbol{z}}_{f,t}$, we cannot use the results of Kesten (1973). Nonetheless, we can show that it is asymptotically linear, a result which allows us to apply Mirek (2011)'s generalisation of Kesten (1973) under appropriate assumptions.

Assumption 1 *Innovation shocks $\varepsilon^{ij}_{f,t}$ are stochastic processes, independent and identically distributed (i.i.d.) over time, across sector-pairs and across firms. Their cumulative density function $G(\varepsilon)$ is on bounded support $(0,\bar{\varepsilon}]$ and has a mean equal one.*

Assumption 2 *Let $\bar{\boldsymbol{\Phi}}_f$ denote the matrix $\boldsymbol{\Phi}_f$ in which I^i_f is always one, i.e. $\bar{\boldsymbol{\Phi}}_f = [\bar{\varphi}^{ij}_f]_{i\times j\in\mathcal{J}\times\mathcal{J}}$, where $\bar{\varphi}^{ij}_f = \frac{1_{\{if\ i=j\}}}{g} + \left(1+\kappa(\tau^{ij})\tau^{ij}\right)\xi^{ij}\varepsilon^{ij}_f$. There exists $\varkappa \in (0, s_\infty)$ where $s_\infty = \sup\{s \in \mathbb{R}_+ : \mathbb{E}|\bar{\boldsymbol{\Phi}}_f|^s < \infty\}$, such that*

$$\mathbb{E}(|\bar{\boldsymbol{\Phi}}_f|^\varkappa) = 1. \tag{B.7}$$

Moreover,

$$\mathbb{E}(|\bar{\boldsymbol{\Phi}}_f|^\varkappa|\log|\bar{\boldsymbol{\Phi}}_f||) < \infty. \tag{B.8}$$

Here we show that the recursion process in firm dynamics (29) satisfies condition (H1) to H(7) in Assumptions 1.6 and 1.7 of Mirek (2011). (H1) and (H2) require that the recursive process is close to an affine recursion. (H3) to (H7) are standard moment conditions on the heavy tail that also required as assumptions in Kesten (1973) and its later extensions.

Verification of Assumption 1.6 (Shape of the mappings) of Mirek (2011).

Let $\epsilon = (\varepsilon, \zeta)$, and

$$\boldsymbol{\Gamma}_\epsilon(\tilde{\boldsymbol{z}}_{f,t}) = \boldsymbol{\Phi}_{f,t+1}(\tilde{\boldsymbol{z}}_{f,t})\tilde{\boldsymbol{z}}_{f,t} + \boldsymbol{\Psi}_{f,t+1}(\tilde{\boldsymbol{z}}_{f,t})\boldsymbol{b}$$

For every $s > 0$, let

$$\boldsymbol{\Gamma}_{\epsilon,s}(\tilde{\boldsymbol{z}}_{f,t}) = s\boldsymbol{\Gamma}_\epsilon(\frac{1}{s}\tilde{\boldsymbol{z}}_{f,t}).$$

$\boldsymbol{\Gamma}_{\epsilon,s}$ are called dilatations of $\boldsymbol{\Gamma}_\epsilon$. Let

$$\bar{\boldsymbol{\Gamma}}_\epsilon(\tilde{\boldsymbol{z}}_{f,t}) = \lim_{s\to 0}\boldsymbol{\Gamma}_{\epsilon,s}(\tilde{\boldsymbol{z}}_{f,t}).$$

Then we have

$$\bar{\boldsymbol{\Gamma}}_\epsilon(\tilde{\boldsymbol{z}}_{f,t}) = \lim_{s\to 0}\boldsymbol{\Gamma}_{\epsilon,s}(\tilde{\boldsymbol{z}}_{f,t}) = \lim_{s\to 0} s\boldsymbol{\Gamma}_\epsilon(\frac{1}{s}\tilde{\boldsymbol{z}}_{f,t}) = \bar{\boldsymbol{\Phi}}_{f,t+1}\tilde{\boldsymbol{z}}_{f,t},.$$

because $\frac{1}{s}\tilde{\boldsymbol{z}}_{f,t}$ transforms firms into mega firms that have abundant knowledge stock and innovate in all sectors (i.e. $I^i_{f,t}(\tilde{\boldsymbol{z}}_{f,t}) = 1, \forall i$) and $\lim_{s\to 0} s\boldsymbol{\Psi}_{f,t+1} = 0$.

Let

$$N_\varepsilon = |(\bar{\boldsymbol{\Phi}}_f - \frac{1}{g}\boldsymbol{I}_{K\times K})\tilde{\boldsymbol{z}}_{f,t} + \boldsymbol{\Xi}\boldsymbol{b}|, \tag{B.9}$$

where $\boldsymbol{I}_{K\times K}$ is a $K \times K$ identity matrix, $\boldsymbol{\Xi} = [\xi^{ij}]_{K\times K}$. That is N_ε is the distance in terms of $\tilde{\boldsymbol{z}}_{f,t+1}$ between the firms that innovate in all sectors and firms that are completely inactive in all sectors. By design, we have

$$|\Gamma_\epsilon(\tilde{\boldsymbol{z}}_{f,t}) - \bar{\boldsymbol{\Phi}}_{f,t+1}\tilde{\boldsymbol{z}}_{f,t}| \le N_\varepsilon, \tag{B.10}$$

for $\forall\ \tilde{\boldsymbol{z}}_{f,t}, \varepsilon^{ij}_{f,t}, \zeta^i_{f,t}$. Hence, the Assumption 1.6 in Mirek (2011) is satisfied.

Verification of Assumption 1.7 (Moments condition for the heavy tail) of Mirek (2011)

Condition (H3) is satisfied because $\bar{\boldsymbol{\Phi}}_f$ depends on $\{\varepsilon^{ij}_{f,t}\}_{ij}$ only, which is i.i.d. over time, sector-pair and firms; and the support of $\varepsilon^{ij}_{f,t}$ is closed due to Assumption 1. The non-arithmeticity assumption in (H4) is satisfied since the support of $\varepsilon^{ij}_{f,t}$ is an interval of real numbers. Assumption 2 directly guarantees (H5) and (H6). In addition, $|N_\varepsilon|^\varkappa < \infty$ because $\tilde{\boldsymbol{z}}_{f,t}$ is bounded by one and $\varepsilon^{ij}_{f,t}$ is bounded by $\bar{\varepsilon}$. Hence (H7) is satisfied too. Overall, the Assumption of 1.7 in Mirek (2011) is satisfied.

Therefore, according to Theorem 1.8 of Mirek (2011), the (normalized) firm size distribution in this economy converges to a stationary Pareto distribution:

$$1 - F^i(z) = \Pr(\tilde{z}^i_f > z) \sim (\frac{z}{k^i})^{-\mu^i}, \tag{B.11}$$

where the shape parameter is μ^i and the scale parameter is k^i. k^i is governed by the public knowledge capital in this sector (i.e. newborn firm's imitated varieties). Therefore, $k^i = \sum_{j\in\mathcal{J}} \xi^{ij} \frac{\theta}{M^j}$. By definition, $M^i \bar{z}^i = n^i$. Together with the definition of u^i in (23), we can derive an explicit expression of the shape parameter based on

$$\begin{aligned}
1 &= M^i \int_{f\in\mathcal{F}_i} \tilde{z}^i_{f,t} dF^i(z) = \frac{\mu^i}{\mu^i - 1} k^i \\
&= \frac{\mu^i}{\mu^i - 1} \frac{(1+r)M^i}{(1-\alpha)v^i} \sum_{j\in\mathcal{J}} \frac{\theta\hat{\omega}^{ij}}{M^j} \\
&= \frac{\mu^i}{\mu^i - 1} \frac{1+r}{1-\alpha} \frac{u^i M^i}{v^i}.
\end{aligned}$$

Rearranging the terms in the above equation yields

$$\mu^i = \left(1 - \frac{1+r}{1-\alpha}\frac{u^i M^i}{v^i}\right)^{-1}.$$

Therefore, the Pareto shape parameter strictly increases in the ratio between the public knowledge value per firm to private knowledge per firm, $\frac{u^i}{v^i/M^i}$.

B.4 Sector-specific parameter values

Table A.4 lists the estimated values for sector-specific expenditure share, s^i, elasticity of substitution σ^i, average fixed cost F^i and the variance of the fixed cost shocks σ^i_ζ.

Table A.4: Sector-specific Parameter Values

Sector	SIC code	Sector Name	s^i	F^i ($\times 10^{-4}$)	σ^i ($\times 10^{-8}$)	σ^i_ζ
1	20	Food And Kindred Products	13.09%	3.48	1.46	2.02
2	22	Textile Mill Products	1.95%	4.00	1.10	1.21
3	281	Industrial Inorganic Chemistry	0.79%	3.04	0.90	0.93
4	286	Industrial Organic Chemistry	2.70%	3.46	0.29	0.12
5	282	Plastics Materials And Synthetic-resins	1.66%	5.73	1.03	1.50
6	287	Agricultural Chemicals	0.64%	7.62	0.26	0.12
7	284	Soaps, Detergents, Cleaners,Perfumes, Cosmetics And Toiletries	1.38%	1.20	1.51	2.22
8	285	Paints, Varnishes, Lacquers, Enamels, And Allied Products	0.54%	5.00	0.68	0.61
9	289	Miscellaneous Chemical Products	0.74%	6.31	0.81	0.87
10	283	Drugs And Medicines	3.32%	3.03	0.17	0.04
11	13, 29	Petroleum And Natural Gas Extraction	7.33%	6.43	1.62	2.62
12	30	Rubber And Miscellaneous Plastics Products	3.66%	4.10	0.17	0.05
13	32	Stone, Clay, Glass And Concrete Products	2.28%	1.88	0.28	0.09
14	331,332,3399, 3462	Primary Ferrous Products	2.63%	4.23	1.15	1.30
15	333-336, 339(exp 3399),3463	Primary And Secondary Non-Ferrous Metals	0.81%	1.50	1.02	1.08
16	34(exp 3462,3463, 348)	Fabricated Metal Products	5.60%	2.90	0.12	0.02
17	351	Engines And Turbines	0.70%	8.40	0.63	0.71
18	352	Farm And Garden Machinery And Equipment	0.58%	2.42	0.49	0.32
19	353	Construction, Mining And Material Handling Machinery And Equipment	1.22%	2.46	0.32	0.14
20	354	Metal Working Machinery And Equipment	0.94%	1.80	0.52	0.37
21	357	Office Computing And Accounting Machines	1.44%	2.84	0.17	0.07
22	355	Special Industry Machinery, Except Metal Working	0.83%	1.89	0.19	0.06
23	356	General Industrial Machinery And Equipment	1.15%	7.98	0.13	0.04
24	358	Refrigeration And Service Industry Machinery	0.97%	2.25	0.34	0.19
25	359	Miscellaneous Machinery, Except Electrical	0.89%	6.18	0.70	0.69
26	361, 3825	Electrical Transmission And Distribution Equipment	0.64%	2.10	0.41	0.28
27	362	Electrical Industrial Apparatus	0.65%	2.21	0.33	0.16
28	363	Household Appliances	0.54%	1.40	0.81	0.86
29	364	Electrical Lighting And Wiring Equipment	0.63%	1.94	0.85	0.93
30	369	Miscellaneous Electrical Machinery, Equipment And Supplies	0.65%	5.58	0.72	0.82
31	365	Radio And Television Receiving Equipment Except Communication Types	0.29%	5.91	0.70	1.00
32	366-367	Electronic Components And Accessories And Communications Equipment	2.75%	3.81	0.10	0.03
33	371	Motor Vehicles And Other Motor Vehicle Equipment	8.28%	4.90	0.43	0.28
34	376	Guided Missiles And Space Vehicles And Parts	0.59%	18.40	6.01	41.17
35	373	Ship And Boat Building And Repairing	0.43%	30.78	4.12	26.78
36	374	Railroad Equipment	0.20%	11.58	2.02	6.26
37	375	Motorcycles, Bicycles, And Parts	0.08%	2.50	1.58	3.84
38	379(exp 3795)	Miscellaneous Transportation Equipment	0.23%	2.30	1.14	1.84
39	348, 3795	Ordinance Except Missiles	0.24%	2.12	2.21	5.81
40	372	Aircraft And Parts	2.91%	3.95	0.54	0.38
41	38(exp 3825)	Professional And Scientific Instruments	3.65%	3.86	0.07	0.01
42	99	All Other Sic's	19.40%	5.00	0.10	0.01

B.5 The Collated Stationary BGP Equilibrium Conditions

This section describe the collated set of conditions for the stationary BGP equilibrium. The set of parameters in the economy is

$$\Theta = \{A^{ij}, v^i, \sigma^i, s^i, a, b, \theta, \alpha, \beta, \eta, L, F^i, \sigma^i_\zeta, \sigma_\varepsilon, M, \}.$$

The list of equilibrium variables of interest are

$$\left\{[v^i, u^i, M^i, \ \frac{n^i}{n^1}]_K, [\omega^{ij}, \hat{\omega}^{ij}, \tau^{ij}, \tilde{s}^{ij}]_{K\times K}, \rho, g, r, PY\right\}.$$

A stationary BGP equilibrium of this model is described by the following systems of equations:

$$\begin{aligned}
v^j &= \frac{1}{1-\rho}(\frac{s^i PY}{\sigma^i} + \sum_{i\in\mathcal{J}} \omega^{ij}), \\
\omega^{ij} &= \left[\kappa(\tau^{ij}) + \kappa'(\tau^{ij})\tau^{ij}\right]\hat{\omega}^{ij}, \\
\tau^{ij} &= \frac{1-\int H^i(\varphi^i_f)\tilde{z}^j_f df}{\int H^i(\varphi^i_f)\tilde{z}^j_f df}, \\
\tilde{s}^{ij} &= \frac{1}{\tau^{ij}+1}, \\
\hat{\omega}^{ij} &= \frac{n^j}{n^i}\frac{1-\alpha}{\alpha}(A^{ij}\alpha\rho v^i)^{\frac{1}{1-\alpha}}(M^i)^{\frac{\alpha}{\alpha-1}}, \\
u^i &= \sum_{j\in\mathcal{J}} \frac{\theta\hat{\omega}^{ij}}{M^j}, \\
(g-1)(g)^{\frac{\alpha}{1-\alpha}} &= \left(\frac{\alpha\beta v^i}{M^i}\right)^{\frac{\alpha}{1-\alpha}} \sum_{j\in\mathcal{J}} \left(A^{ij}\right)^{\frac{1}{1-\alpha}} \left[\tilde{s}^{ij} + \kappa^{ij}(\tau^{ij})(1-\tilde{s}^{ij}) + \theta\frac{M^i}{M^j}\right]\frac{n^j_t}{n^i_t}, \\
M^i_t &= M\int_{f\in\mathcal{F}} H^i(\varphi^i_{f,t})dG^\varphi(\varphi^i_{f,t}), \\
1 &= \beta(1+r)g^{(\eta-1)\sum_i \frac{s^i}{1-\sigma^i}} \\
\rho &= \frac{1}{1+r}\frac{1}{g}, \\
L &= \frac{\sigma^i - 1}{\sigma^i}PY + \sum_{i\in\mathcal{J}}\left[\alpha\rho(g-1)v^i + M^i F^i\right], \\
PY &= L + r\sum_{i\in\mathcal{J}} v^i.
\end{aligned}$$

where $\varphi^i_{f,t} = \sum_{j\in\mathcal{J}}[1-\kappa(\tau^{ij})-\kappa'(\tau^{ij})\tau^{ij}(\tau^{ij}+1)]\hat{\omega}^{ij}\tilde{z}^j_{f,t}+u^i$ is firm f's additional gain in innovating in sector i, which depends on firm's existing knowledge portfolio, $\tilde{\boldsymbol{z}}_{f,t}$. Unfortunately, we cannot solve the equilibrium analytically using the above equations, as there is no closed form expressions for the mass of innovating firms and $G^\varphi(\varphi^i_f)$. We thus resort to simulation with a large number of firms to explore the implications of the model.

C Computational Appendix

C.1 Simulation Algorithm

This section explains the simulation procedure for replicating the firm-specific observations in Section 5.4. As explained previously, we assume the economy is at steady state in 1997 and estimate the model parameters based on observations of firms which filed for patent at least once during the five year period of 1993-1997.

First, based on the values of g, τ^{ij}, α, θ and v^i estimated in Section 5.1, we construct estimates for u^i and $\hat{\omega}^{ij}$ according to (23) and (18), respectively. We then populate the simulated economy with $N = 70,000$ firms and $K = 42$ sectors. The initial firm sizes in each sector are randomly drawn from a log-normal distribution with mean one and standard deviation 10 in each sector.

The following process is then iterated until the (normalized) firm size distribution stabilizes in each sector, which takes about 100 periods.

1. For each firm in each sector, we calculate the expected gain from innovating, $\varphi^i_{f,t}$, according to (24), given the estimated u^i and $\hat{\omega}^{ij}$ and $\{\tilde{\boldsymbol{z}}_{f,t}\}_f$ from the last period.

2. Calculate the probability of innovation in sector i for each firm $H^i(\varphi^i_{f,t})$, given $\{\tilde{\boldsymbol{z}}_{f,t}\}_f$, the GMM estimated F^i, σ^i_ζ, and the other parameters.

3. Simulate the innovation decision indicator $I^i_{f,t}$ using a binomial random variable generator such that $I^i_{f,t}$ equals one with the estimated probability $H^i(\varphi^i_{f,t})$.

4. For each firm, randomly draw an innovation shock ($\varepsilon^{ij}_{f,t}$) from a bounded Gamma distribution with mean one and the estimated variance σ_ε and upper bound $\bar{\varepsilon} = 2$. With the simulated $I^i_{f,t}$ from the last step, estimated model parameters and the innovation shock, we update the relative firm size $\tilde{z}^i_{f,t+1}$ of the next period according to the equilibrium firm dynamics in (29). Specifically, if $I^i_{f,t} = 1$,

$$\tilde{z}^i_{f,t+1} = \sum_{j=1}^{K} \phi^{ij}_{f,t+1} \tilde{z}^j_{f,t} + \theta \sum_{j=1}^{K} \frac{(1+r)\hat{\omega}^{ij}}{(1-\alpha)v^i M^j},$$

where

$$\phi^{ij}_{f,t+1} = \frac{1_{\{\text{if } i=j\}}}{g} + \left(1 + \kappa(\tau^{ij})\tau^{ij}\right) \frac{(1+r)\hat{\omega}^{ij}}{(1-\alpha)v^i} \varepsilon^{ij}_{f,t}$$

Otherwise, $I^i_{f,t} = 0$, then

$$\tilde{z}^i_{f,t+1} = \frac{\tilde{z}^i_{f,t}}{g}.$$

5. The updated set of $\{\tilde{\boldsymbol{z}}_{f,t+1}\}_f$ then enters the next period as the new initial normalized firm sizes.

6. In simulation if firm f has been inactive for five consecutive periods in all sectors (i.e. $\Pi_{\tau=t}^{t+5}\Pi_{i=1}^{K} I^i_{f,t} = 0$), this firm is treated as if it has exited the economy entirely. In the sixth

period, we treat the firm as a successful newborn firm in at least one sector, with the entry probability to i given by $\frac{H^i(u^i)}{\Pi_i(1-H^i(u^i))}$.[30]

C.2 Counterfactual Simulations Algorithm

In this counterfactual experiment, we change the average and variance of the idiosyncratic fixed costs of conducting R&D, F^i and σ^i_ζ, to examine their impact on the aggregate growth rate, firm size distribution, firm co-patenting and cross-sector knowledge utilization.

The main difference between the simulation process in Appendix C.1 and the counterfactual experiments is that with the counterfactual fixed costs distribution, the economy is operating in a different steady state from what we observe in the data. Therefore, we need to first compute the new steady state value of $g, M^i, n^i, v^i, \tau^{ij}$, and then use the new parameter values to examine the firm dynamics in (29).

The counterfactual simulation iterates through the following steps.

1. Guess an initial value of $g, M^i, n^i, v^i, \tau^{ij}$;

2. Simulate the firm dynamics according to (29) for T periods, which generates a new set of M^i, τ^i;[31]

3. Bring the new M^i, τ^i to equations (14) and (B.2) and run GMM to solve for new value of v^i, n^i, g. We apply the Perron-Frobenius Theorem to solve for the systems of equations in (B.2), where the new steady state $(g-1)g^{\frac{1}{1-\alpha}}$ is the largest eigenvalue of the matrix $\boldsymbol{A_g}$ in (C.1) and $\boldsymbol{n}$ is the corresponding eigenvector. Meanwhile, (14) gives the new steady state value of v^i.

$$(g-1)g^{\frac{1}{1-\alpha}}\boldsymbol{n} = \boldsymbol{A_g n}, \tag{C.1}$$

where

$$A^{ij}_g = (\frac{\alpha\beta v^i}{M^i})^{\frac{\alpha}{1-\alpha}}(A^{ij})^{\frac{1}{1-\alpha}}\delta^{ij}. \tag{C.2}$$

4. When the initial and new values of $g, M^i, n^i, v^i, \tau^{ij}$ do not converge, set the new value of $g, M^i, n^i, v^i, \tau^{ij}$ as initial guess and go back to 1; otherwise exit the loop.

[30] In the patent data, when a firm has been inactive for five periods, its likelihood of resuming patenting is almost zero. In the estimated model, such firms still collect knowledge licensing fees, but are unlikely to resume R&D, because the estimated c.d.f. of the Gamma distribution $H^i(\varphi^i_{f,t})$ is flat and almost zero when $\varphi^i_{f,t}$ is close to zero given our estimated values of F^i and σ^i_ζ. Therefore, treating the persistently inactive firms in the simulation as newborn firms is sensible, as these firms behave almost the same as new entrants in terms of their entry decision and innovation outcome.

[31] The simulation does generate a new set of g, n^i, but due to the granularity with finite number of firms, volatile R&D shocks and fixed cost shocks, g, n^i fluctuate too much over different periods, while M^i, τ^i depends on the distributions of $\tilde{\boldsymbol{z}}_{f,t}$, which are relatively stable. That is why we take M^i, τ^i from the simulation, but not g, n^i.