



University of St.Gallen

Moving to the Innovation Frontier

Edited by Christian Keuschnigg

Moving to the Innovation Frontier

Centre for Economic Policy Research

Centre for Economic Policy Research
33 Great Sutton Street
London EC1V 0DX
UK

Tel: +44 (20) 7183 8801
Fax: +44 (20) 7183 8820
Email: cepr@cepr.org
Web: www.cepr.org

ISBN: 978-1-907142-97-0

© CEPR Press, 2016.

Moving to the Innovation Frontier

Edited by Christian Keuschnigg



CEPR PRESS

Centre for Economic Policy Research (CEPR)

The Centre for Economic Policy Research (CEPR) is a network of over 1000 research economists based mostly in European universities. The Centre's goal is twofold: to promote world-class research, and to get the policy-relevant results into the hands of key decision-makers. CEPR's guiding principle is 'Research excellence with policy relevance'. A registered charity since it was founded in 1983, CEPR is independent of all public and private interest groups. It takes no institutional stand on economic policy matters and its core funding comes from its Institutional Members and sales of publications. Because it draws on such a large network of researchers, its output reflects a broad spectrum of individual viewpoints as well as perspectives drawn from civil society.

CEPR research may include views on policy, but the Trustees of the Centre do not give prior review to its publications. The opinions expressed in this report are those of the authors and not those of CEPR.

Chair of the Board	Sir Charlie Bean
President	Richard Portes
Director	Richard Baldwin
Research Director	Kevin Hjortshøj O'Rourke
Policy Director	Charles Wyplosz

About the Contributors

Philippe Aghion is the Robert C. Waggoner Professor of Economics at Harvard University, and a fellow of the Econometric Society and of the American Academy of Arts and Sciences. His research focuses on the economics of growth. With Peter Howitt, he pioneered the so-called Schumpeterian Growth paradigm which was subsequently used to analyze the design of growth policies and the role of the state in the growth process. Much of this work is summarized in their joint books *Endogenous Growth Theory* (MIT Press, 1998) and *The Economics of Growth* (MIT Press, 2009), in his book with Rachel Griffith on *Competition and Growth* (MIT Press, 2006), and more recently in *Repenser l'Etat* (Seuil, 2011). In 2001, Philippe Aghion received the Yrjo Jahnsson Award of the best European economist under age 45.

Ufuk Akcigit is an Assistant Professor of Economics at the University of Chicago. As a macroeconomist, his research centers on economic growth, productivity, firm dynamics and economics of innovation. Before joining the University of Chicago, Akcigit was an Assistant Professor at the University of Pennsylvania. In addition, he serves as a faculty research fellow in the National Bureau of Economic Research. Akcigit has received several awards for both research and teaching. In 2014 he was named a Kauffman Junior Faculty Fellow in recognition of his outstanding scholarship. Akcigit received his bachelor's degree in economics from Koc University in Istanbul. He then went on to earn his PhD from the Massachusetts Institute of Technology in 2009.

William Kerr is a Professor at Harvard Business School. Bill is the faculty chair of the Launching New Ventures program, and his work focuses on how companies explore new opportunities and generate growth. He has received Harvard's Distinction in Teaching award and the Ewing Marion Kauffman Prize Medal for Distinguished Research in Entrepreneurship. Bill has worked with companies worldwide on the development of new ventures and transformations for profitable growth.

Christian Keuschnigg is Professor of Economics with specialization in public economics at the University of St. Gallen. He is a research fellow of the research CEPR, CESifo and the Oxford University Centre for Business Taxation. From June 2012 to December 2014 was Director of the Institute for Advanced Studies in Vienna. He is coeditor of *Empirical Economics* and of *FinanzArchiv/Public Finance Analysis* and was previously coeditor of *European Economic Review*. His research interests include the theory of taxation; innovation and growth; corporate finance and capital market frictions; and the welfare state, aging and the labour market.

Ramana Nanda is an Associate Professor of Business Administration in the Entrepreneurial Management unit at Harvard Business School. His research focuses on understanding the drivers of financing constraints for startups and on the ways in which the structure of the financial sector impacts innovation and entrepreneurship in the economy. Ramana is a Faculty Research Fellow in the Productivity, Innovation and Entrepreneurship Program at the NBER, and a Faculty Affiliate at the SME Initiative of Innovations for Poverty Action.

Matthew Rhodes-Kropf is an Associate Professor in the Entrepreneurial Management Unit at Harvard Business School, and a faculty research fellow at the National Bureau of Economic Research. Professor Rhodes-Kropf specializes in mergers and acquisitions, venture capital, and corporate governance. His work seeks to understand how capital markets interact with the creation of new firms, their financing, growth, governance, and their ultimate exit through a successful IPO or sale or through failure. A graduate of Duke University, Professor Rhodes-Kropf holds a BA in computer science and economics and an MA and PhD in economics.

Mark Schankerman is Professor in Economics at the London School of Economics, and Research Fellow at the Centre for Economic Policy Research. He has a PhD in Economics from Harvard University, formerly taught at New York University, and was Research Associate at the NBER for ten years. He was Director of Policy Research and Advisor to the Chief Economist at the EBRD from 1995-2003, has worked at the World Bank on Africa, and has extensive consultancy experience in the private and public sectors. He is widely published in the areas of patents, innovation, university technology transfer, R&D and productivity, open source software, and emerging economies. He serves on the Advisory Board of the *Journal of Industrial Economics* and was an editor of several journals. In 2010-11 he served on the Panel of Experts for HMG Hargreaves Review of Intellectual Property and Growth.

Contents

<i>About the Contributors</i>	v
<i>Foreword</i>	ix
Editorial	1
<i>Christian Keuschmigg</i>	
1 Competitiveness and Growth Policy Design	5
<i>Philippe Aghion</i>	
2 R&D Policies and Economic Growth	31
<i>Ufuk Akcigit</i>	
3 Innovation and Business Growth	41
<i>William R. Kerr</i>	
4 Regional Variation in Venture Capital: Causes and Consequences	55
<i>Ramana Nanda and Matthew Rhodes-Kropf</i>	
5 Patent Rights and Cumulative Innovation: Causal Evidence and Policy Implications	73
<i>Mark Schankerman</i>	

Foreword

In January 2015 CEPR held a workshop with the aim of discussing high quality research focussing on R&D, innovation and growth. "Moving to the Innovation Frontier" was hosted by the Institute of Advanced Studies in Vienna and chaired by Professor Christian Keuschnigg of the University of St. Gallen. It brought together researchers from several fields of economics, presenting a mix of theoretical and empirical studies which explored the determinants of innovation and R&D based growth as well as the rationale and effectiveness of public policy interventions.

This new eBook collects essays from the workshop's five keynote speakers on the increasing importance of innovation, competition and structural change and how it feeds through to economic growth. The view that diversification from domestic demand as a main driver of economic growth and the move to a more integrated and globalised world is shared by these authors. Their focus lies in determining the most effective ways to enable and encourage innovation, through which they explore policy design; research and development; business opportunities; venture capitalism; and property rights.

Each chapter presents ideas of how these five aspects can be improved to implement policy that increases economic innovation. This eBook is a comprehensive collection of ideas that clarifies how nations should modernise policy decisions in order to expand their economic potential in a changing global economy.

CEPR is grateful to Professor Christian Keuschnigg for his editorship of this eReport. Our thanks also go to Anil Shamdasani for the excellent and efficient handling of its production. CEPR, which takes no institutional positions on economic policy matters, is delighted to provide a platform for an exchange of views on this topic.

Tessa Ogden
Deputy Director, CEPR
March 2016

Editorial

Christian Keuschnigg

University of St. Gallen, FGN-HSG, and CEPR

Moving to the innovation frontier requires adopting a systemic approach to innovation policy which spans the entire value chain, starting with idea generation and ending with investment and employment after successful market introduction. When approaching the knowledge frontier, an economy's capacity to innovate must shift from imitation and differentiation towards more radical and riskier innovations that aim at entirely new products and services. In this process, tertiary education, basic research and technological infrastructure become more critical factors in activating private innovation and boosting the returns to research and development (R&D).

Successful innovation is a key determinant of a firm's competitive advantage in product markets and is a necessary condition for employment and investment to be profitable. In a firm's life-cycle, innovation-driven growth ultimately creates the need to enter world markets – where competition is worldwide and therefore most intense – for further growth. In the cross-section, exporting firms and multinational companies are thus substantially more productive and larger than firms that are mostly active in domestic markets only. Patent protection allows firms to cash in on successful innovations for a while, but tense competition from potential and actual new competitors forces them to continuously invest in new R&D to keep ahead of their rivals and protect their growth potential. International market integration and competition policy facilitating the entry of new firms are important drivers of growth.

Innovation-based growth is a process of creative destruction; the new replaces the old. Innovation is associated with market entry and exit and therefore selection among younger firms. The creation of new product lines and the closing down of old ones rejuvenates the product cycle in large firms. In all cases, innovation leads to a reallocation of labour and capital. Investment and employment must flow into more profitable activities with higher value-added growth, away from activities with low returns and declining market prospects. Empirical research suggests that roughly half of a country's productivity growth is due to targeted allocation and ongoing reallocation of investment and employment to more valuable uses. When a country moves closer to the knowledge frontier, innovation must become riskier and more radical and factor reallocation must occur on a larger scale. One must conclude that the full potential of innovation can be exploited only if factor reallocation is supported by flexible capital and labour markets. For workers to more readily accept the employment risk in innovative industries, welfare policy should probably become more of the 'flexicurity' type, combining generous unemployment insurance with low job protection and

2 *Moving to the Innovation Frontier*

active labour market policies for retraining and supporting job search. Regarding capital markets, credit financing is exposed to a firm's bankruptcy risk, but it is not endowed with control rights to intervene in case of problems and neither does it participate in a firm's upside potential. Credit financing is suitable for more established firms and relatively safe investments, and is not particularly innovation friendly. When an economy shifts towards riskier but potentially more profitable innovations, the balance of financing should shift from credit to more equity financing, giving a larger role to stock markets, venture capital and private equity.

This report collects five non-technical essays based on the invited lectures at the "Moving to the Innovation Frontier" workshop organised jointly by CEPR, the University of St. Gallen and the Institute of Advanced Studies, sponsored by the Austrian Council, and hosted by IAS on 19-20 January 2015 in Vienna.¹ Philippe Aghion investigates the process of creative destruction from a firm perspective, contrasts the nature of productivity growth in advanced countries and emerging economies, and discusses the design and governance of industrial policy for innovation-based growth. Ufuk Akcigit argues that the design of optimal innovation policy should focus relatively more on the selection of firms, which may be more or less R&D intensive and feature higher or lower growth potential. It should also focus more on the complementary roles of basic and applied research. Basic research is an essential responsibility of government. It also occurs in the private sector and sometimes creates unexpected applications in other sectors. Large firms that are active in many different industries invest relatively more, since they can better exploit the general nature of basic research.

William Kerr investigates innovation in large firms facing a trade-off between improving existing product lines and creating entirely new business opportunities. He develops a theoretical framework capturing these two alternative directions for private R&D. He then explores a case study of how IBM introduced a new innovation strategy to better identify and implement newly emerging business opportunities and was thereby able to manage a turnaround from a declining and loss-making company into a dynamic and most profitable one. Ramana Nanda and Matthew Rhodes-Kropf discuss the role of venture capital in funding high-risk and high-potential innovations in new start-up firms. Venture capital-backed investment is a rather small part of innovation financing, but one which is disproportionately effective in generating employment and growth. The experimental orientation and high risk of start-up investment make it unsuitable for more conservative bank financing; it requires active venture capital support with participation in the upside potential of a firm. The experimental nature of investment typically calls for staged financing. Venture capitalists can save scarce funds by abandoning unpromising projects early on and continue financing in later stages only as more reliable information becomes available.

Ensuring effective innovation incentives is a central element of innovation policy, and property rights in the form of patents is one of the main policy instruments to achieve this. Mark Schankerman, however, discusses empirical

¹ www.fgn.unisg.ch/en/profkeuschnigg/workshops.

evidence that patent protection can also discourage follow-on innovation by downstream firms if bargaining between upstream patent holders and potential downstream licensees breaks down. For the vast majority of patents, the evidence indicates that patents do not impede downstream innovation. However, blocking occurs in complex technology areas where later innovators need many different patents to conduct research (e.g., information technology and electronics), but not in other important sectors like pharmaceuticals and chemicals. Blocking appears to be concentrated in cases where large firms with patents interact with small downstream innovators. The finding that the impact of patent rights on cumulative innovation is localized rather than pervasive calls for more targeted policies rather than a general restriction in patent rights.

1 Competitiveness and Growth Policy Design

Philippe Aghion
Harvard University

1.1 Introduction

After decades during which governments in developed countries would privilege domestic demand as a main driver of economic growth, the advent of globalisation has forced governments to increasingly turn their attention to the competitiveness of the domestic economy, i.e. the extent to which a country can export its production abroad and thereby "exchange goods and services in which it is abundant for goods and services that it lacks" (Altomonte et al., 2012).¹

Meanwhile, trade economists have themselves evolved with regard to how the issue of competitiveness should be approached. Thus, as clearly explained by Bernard et al. (2011), while theories of international trade used to emphasise inter-industry trade and therefore the view that international competition is between countries, with each country playing on the industries where it has a comparative advantage, recent theories emphasise firm-level competition worldwide and intra-industry trade. As well put by Altomonte et al. (2012), "it is not really the country that exchanges [...] goods and services, but rather its firms". According to this view, what makes a country competitive is primarily what makes its individual firms competitive.

And what makes an individual firm competitive on the world market are both its productivity and its size. Here, the seminal theoretical contribution is by Melitz (2003), who develops a model of intra-firm trade with heterogenous firms where only firms that are sufficiently productive can become exporters, as being more productive allows firms to secure a market share that covers the fixed cost of exporting.²

This prediction is confirmed by cross-country firm-level evidence (see, for example, Altomonte et al., 2012) and it has important policy implications for how to enhance competitiveness of the domestic economy. In particular, in a departure from 'vertical' or 'top-down' policies that would emphasise national comparative advantage based on current national factor endowment, the new theories call for more horizontal policies to favour productivity growth and size growth of individual firms in the country.

1 The recent fiscal devaluation in France through the introduction of the Credit d'Impot Competitivite (CICE) gave rise to a heated debate between the advocates of demand-driven policies and those who emphasise the need to increase the country's competitiveness.

2 In practice, credit constraints, labour regulations and other types of market imperfections may prevent more productive firms from growing in size sufficiently to become viable exporters.

This chapter is organised as follows. Section 2 summarises the main arguments in the recent trade literature in favour of a firm-level approach to competitiveness. Section 3 links firm-level competitiveness to productivity. Section 4 discusses potential determinants of firm-level productivity and productivity growth. Section 5 considers potential barriers to the growth in firm size. Section 6 revisits the role for vertical targeting (or sectoral policies). Section 7 draws on our discussion to propose some elements of a new growth strategy for China. Section 8 concludes.

1.2 From industry-level to firm-level competitiveness

Until Melitz's seminal contribution to trade economics (Melitz, 2003), mainstream theories of international trade would commonly rely on the assumption of a representative firm in each domestic economy. These include both the Heckscher-Ohlin model emphasising comparative advantage under perfect competition as the main driver of international trade, and the more recent theories of Krugman (1980) and Helpman (1981) focusing instead on increasing returns and consumers' preference for variety as drivers of trade. However, recent evidence shows both that there is a high degree of heterogeneity across firms in a same domestic industry, and that this heterogeneity is in firm-level productivity, in firm size, in firm-level skills and wages, and in capital intensity. Moreover, as predicted by Melitz (2003), this heterogeneity with regard to productivity in particular is a key determinant of whether and to what extent firms are involved in international trade and of how well they perform as exporters.

First, the extent of intra-industry differences is shown, for example, by Syverson (2004). Within an average US sector, the top 10% of firms in terms of productivity are twice as productive as the bottom 10%. More recently, Hsieh and Klenow (2009) show that intra-industry differences are even bigger in emerging economies: in China and India, the top 10% produce more than five times as much as the bottom 10% in the average industry.

Moreover, the distribution of firm-level performance – whether it is measured by productivity or by firm size – tends to be highly skewed (typically Pareto distributed rather than normally distributed), with many low performing firms and only few high performing firms. This in turn implies that intra-industry firm heterogeneity is much greater than the heterogeneity in average performance across industries across countries. In other words, ignoring firm heterogeneity within industries and looking instead directly at industry-level or country-level averages introduces a significant 'aggregation bias' (Altomonte et al., 2012). And it may lead to inappropriate policy prescriptions if the heterogeneity in firm performance within an industry is mirrored by a heterogeneity in firms' ability to export on the world market.

Now suppose, as in Melitz (2003), that only firms above a given performance level cut-off are able to export. Then, as explained well by Altomonte et al. (2012), any policy that would simply aim at increasing average industry-level productivity without affecting the distribution of firm performance within the industry, and in particular without affecting the number of firms that pass the export threshold, will have no effect on the industry's overall exports, and therefore on its competitiveness. And indeed, as we will see in the next section, there is evidence of a 'happy few' phenomenon, i.e. that only few firms above

a given performance threshold are able to become exporters or more generally 'internationalised'. Hence the importance of looking directly at firm-level differences and of understanding how firm-level characteristics affect firms' ability to export,³ in other words, not just their individual productivity levels but also whether those levels exceed the export threshold.

One remark to conclude this section. We have tried to explain here why more productive firms are more likely to engage in international activities, but there is also the reverse causality from trade openness to firm-level productivity growth. For example, using a new firm-level panel data across twelve European countries over the period 1996-2007, Bloom et al. (2011) show that increased competition from Chinese imports has spurred technical change within firms in those countries (whether technical change is measured by IT diffusion, R&D expenses, TFP growth or improvements in management practices).

1.3 Productivity and other key characteristics of exporting firms

Table 1.1 from Bernard et al. (2011) is based on 2002 data from the US Census of Manufactures. It regresses the various firm characteristics (size, productivity, skills, etc.) on a dummy variable indicating whether the firm is an exporter or not. The results summarised in the table (the first column includes no fixed effect, the second column includes industry fixed effects, and the third column includes industry fixed effects plus log firm employment as an additional control) point to an 'exporter premium' in terms of firm size, productivity, skill and capital intensity.

³ For example, by appealing to the heterogeneity in firm performance within industries, Antras et al. (2010) account for the so-called 'Spanish paradox', i.e. the fact that Spain increased its overall degree of competitiveness (measured by its share of world exports) over the decade 2000-2009 even though average productivity (measured by unit labour costs) deteriorated over that period. What happened is that productivity improved for firms already beyond the export threshold, but it underwent significant deterioration for firms below that threshold. Thus the same number of firms kept exporting and they exported more due to their increased productivity.

Table 1.1 Regression of various firm characteristics on a dummy variable indicating whether the firm is an exporter or not

	(1)	(2)	(3)
Log employment	1.19	0.97	-
Log shipments	1.48	1.08	0.08
Log value added per worker	0.26	0.11	0.10
Log TFP	0.02	0.03	0.05
Log wage	0.17	0.06	0.06
Log capital per worker	0.32	0.12	0.04
Log skill per worker	0.19	0.11	0.19
Additional covariates	None	Industry fixed effects	Industry fixed effects. Log employment

Note: OLS regressions.

Source: Bernard et al. (2011).

More recently, Altomonte et al. (2012) have gone somewhat further by looking at the relationship between firm characteristics and firm-level openness over multiple countries and considering several dimensions of openness. In order to perform a reliable comparative analysis, with the support of the Bruegel think tank the authors carried out a large cross-country, firm-level survey (European Firms in a Global Economy, or EFIGE). To construct the EFIGE dataset, the authors selected seven countries – Germany, France, Italy, Spain, the UK, Austria and Hungary – and for each of these countries they selected a large number of firms to which they sent a survey questionnaire. The overall sample included 3,000 firms in each of the first four countries, more than 2,000 firms in the UK, and around 500 firms in Austria and in Hungary. All of the firms had more than ten employees. Based on the answers to the questionnaire, the authors constructed ‘openness’ indicators reflecting the nature or extent of the firms’ international involvement. Thus, a firm would be called “exporter” if it provided a positive answer when asked if it sold abroad. Similarly, binary indicators were constructed for importing versus non-importing firms and to distinguish between firms that were involved in foreign direct investment (FDI) or outsourcing from firms that were not.

Table 1.2, from Altomonte et al. (2012), provides interesting descriptive statistics on the mapping between various dimensions of firm performance and various indicators of firms’ degree of openness. In particular, we see that larger or more capital-intensive firms tend to be more ‘open’ along the various openness scales. Moreover, the export performance threshold appears to be lower than the FDI threshold. Table 1.3 shows that the same conclusion applies when looking at firm-level productivity (whether measured by TFP, by unit labour costs, or labour productivity): more productive firms tend to be more open, and again the export performance threshold appears to be lower than the FDI threshold.

Table 1.2 Descriptive statistics of firm performance and indicators of firms' degree of openness

	No. of firms	Avg. turnover per firm (in €1,000)	Avg. no. of employees	Avg. capital stock per employee (in €1,000)
Non-active abroad	3,402	4,443.33	31.44	152.16
Active abroad	11,357	19,273.46	139.85	196.4
Of which				
Exporter	9,849	20,494.12	151.42	199.03
Importer of services	3,449	38,659.98	332.12	223.57
Importer of materials	7,298	24,976.44	191.17	200.36
Global exporter	4,016	24,777.71	103.43	222.93
Passive outsourcer	5,799	17,052.42	83.96	204.98
Active outsourcer	590	24,657.11	119.55	225.28
FDI	719	77,637.20	334.13	239.55
Whole sample	14,759	15,589.29	114.52	189.59

Source: EFIGE dataset; Altomonte et al. (2012).

Both the fact that better performing firms tend to be more open and the ranking between the thresholds associated with the different measures of openness appear even more clearly when looking at performance deciles. Altomonte et al. (2012) thus show that around 85% of firms within the top TFP decile in the corresponding industry are exporters, around 45% of firms in the same decile are global importers, fewer than 15% are involved in FDI, and around 5% are involved in outsourcing.

Table 1.4, again from Altomonte et al. (2012), reports the results from the OLS regression of TFP on the various openness dummies. First, we see that the correlations between TFP and the various openness indicators are all positive and significant. Second, being involved in FDI commands a higher TFP premium than being an exporter, which is again consistent with the notion that the fixed cost of FDI involvement is higher than that of exporting.

Table 1.3 Descriptive statistics of firm performance and indicators of firms' degree of openness including firm level productivity

	No of firms	Avg. turnover per firm (in €1,000)	Avg. no. of employees	Avg. capital stock per employee (in €1,000)	Total factor productivity	Unit labour cost (in euros per unit of added value)	Labour productivity (added value per employee in €1,000)
Non-active abroad	1,514	5,298.51	31.67	156.14	0.872	0.77	50.71
Active abroad	5,921	26,104.12	152	200.01	1.024	0.78	57.55
Of which:							
Exporter	5,201	26,104.12	164.41	203.19	1.033	0.77	58.09
Importer of services	1,900	50,004.76	372.81	230.61	1.159	0.84	61.81
Importer of materials	3,939	31,647.82	208.25	203.31	1.058	0.79	58.43
Global exporter	2,211	38,345.27	104.42	224.77	1.094	0.79	62.56
Passive outsourcer	2,965	20,763.66	84.31	208.06	1.06	0.79	59.86
Active outsourcer	306	32,991.62	127.39	224.94	1.066	0.76	56.03
FDI	387	98,554.23	359.7	238.08	1.293	1.05	63.35
Whole sample	7,435	20,303.82	125.6	190.39	0.991	0.78	56.05

Note: Numbers are weighted averages. TFP is the Solow residual of the production function.

Source: FIGE dataset; Altamonte et al. (2012).

Table 1.4 Results from the OLS regression of TFP on the various openness dummies

Dep. Variable: TFP	(1) OLS	(2) OLS	(3) O.Probit	N
Active abroad	0.0906*** (0.0132)	0.0353*** (0.0128)	0.261*** (0.0290)	7,259
Exporter	0.0999*** (0.0136)	0.0399*** (0.0131)	0.272*** (0.0298)	6,563
Importer of services	0.171*** (0.0171)	0.0626*** (0.0171)	0.620*** (0.0531)	3,334
Importer of materials	0.118*** (0.0142)	0.0449*** (0.0138)	0.394*** (0.0332)	5,320
FDI	0.257*** (0.0329)	0.0980*** (0.0357)	0.750*** (0.0750)	1,862
Passive outsourcer	0.122*** (0.0151)	0.0558*** (0.0150)	0.329*** (0.0342)	4,372
Active outsourcer	0.134*** (0.0309)	0.0477 (0.0306)	0.364*** (0.0755)	1,777
Global exporter	0.156*** (0.0168)	0.0699*** (0.0167)	0.425*** (0.0368)	3,652
Country fixed effects	Yes	Yes	Yes	-
Sector fixed effects	Yes	Yes	Yes	-
Firm size	No	Yes	No	-

Source: Altamonte et al. (2012).

A key issue raised by the above tables is of course that of the direction of causality – in particular, whether the above correlations reflect the impact of firm-level performance on firms' ability to become more 'open', or the fact that increased openness raises firms' productivity growth. Melitz (2003) models the performance-to-openness causality, whereas we have developed models that capture the reverse causality from openness to productivity growth by including both a reallocation effect (towards more productive firms) and an 'escaping competition through innovation' effect. Both effects contribute to increasing average productivity in the domestic economy. That both causalities should be at work comes out clearly from the recent empirical literature on trade, reallocation and firm heterogeneity.⁴

1.4 Enhancing productivity

This section looks at the determinants of productivity growth, based on two questions: How can we enhance productivity growth in advanced versus emerging market economies? Is there something to learn from observing the big technological waves and their diffusion patterns across different countries? We first present a simple framework to think about the sources of productivity growth. We then look at the sources of productivity growth in advanced countries, before turning our attention to the sources of productivity growth in emerging market

⁴ In particular, see Sections 2.3 and 2.4 in Bernard et al. (2011).

economies. Finally, we analyse the technological waves and draw a few lessons from a comparison of the differences in their diffusion patterns across countries.

1.4.1 A framework for thinking about the sources of productivity growth

In 1956, Robert Solow developed a model to show that in the absence of technical progress, there can be no long-run growth in GDP per capita (Solow, 1956). On the other hand, historical evidence suggests that productivity growth is an increasingly important component of growth (see, for example, Helpman, 2004). But what are the sources of productivity growth?

A useful framework for thinking about productivity growth and its determinants is the so-called ‘Schumpeterian paradigm’, which revolves around four main ideas.

First idea: productivity growth relies on profit-motivated innovations. These can be process innovations (i.e. to increase the productivity of production factors such as labour or capital), product innovations (introducing new products), or organizational innovations (to make the combination of production factors more efficient). Policies and/or institutions that increase the expected benefits from innovation should induce more innovation, and thus faster productivity growth. In particular, better (intellectual) property rights protection, R&D tax credits, more intense competition, and better performing schools and universities are all policies that foster productivity growth.

Second idea: creative destruction. New innovations tend to make old innovations, old technologies and old skills obsolete. This underlies the importance of reallocation in the growth process.

Third idea: innovations may be either ‘frontier’ innovations, which push the frontier technology forward in a particular sector, or ‘imitative’ or ‘adaptive’ innovations, which allow the firm or sector to catch up with the existing technological frontier. The two forms of innovation require different types of policies and institutions.

Fourth idea: Schumpeterian waves. Technological history is shaped by the big technological waves that correspond to the diffusion of new ‘general purpose technologies’ (the steam engine, electricity, information and communication technologies (ICT), etc.) to the various sectors of the economy.

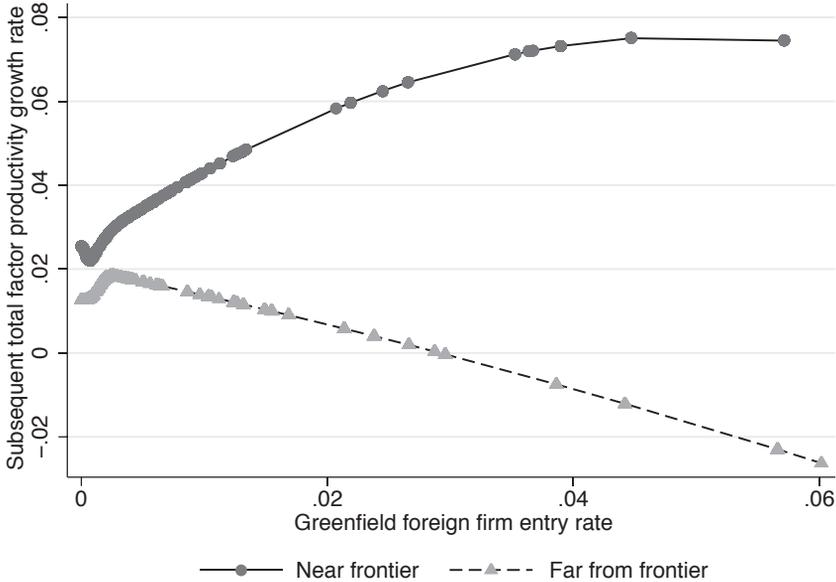
1.4.2 Enhancing productivity growth in advanced countries

To enhance productivity growth in advanced countries, where growth relies more on frontier innovations, it helps to invest more in (autonomous) universities, to maximize the flexibility of product and labour markets, and to develop financial systems that rely heavily on equity financing.

Figure 1.1 (from Aghion et al., 2009a) shows how competition (here measured by the lagged foreign entry rate) affects productivity growth in domestic incumbent firms. The upper curve shows the average TFP growth among domestic firms that are closer than the median to the worldwide technological frontier in their sector. We see that, on average, productivity growth in those firms responds positively to more intense competition. This reflects an ‘escape competition effect’, i.e. such firms innovate more to escape the more intense competition. In contrast, productivity growth in firms that are farther below the worldwide

technological frontier in their sector than the median reacts negatively to more intense competition. This reflects a discouragement effect. The closer a country is to the world-leading productivity level, the higher the fraction of 'above median' firms, and therefore the more productivity-enhancing product market competition is.

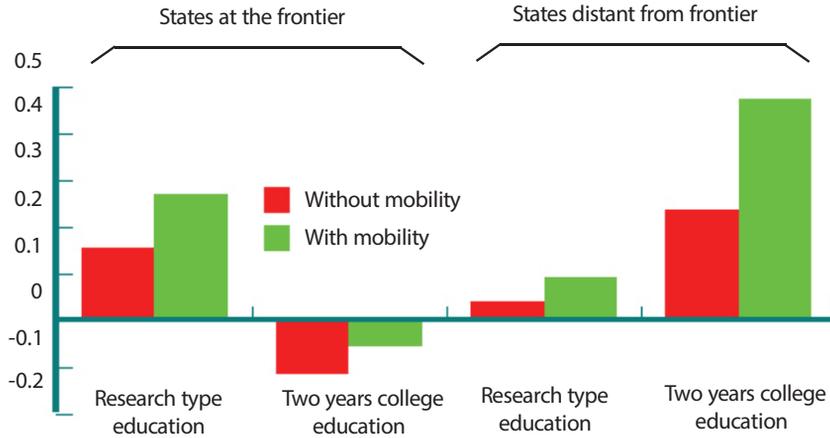
Figure 1.1 Total factor productivity growth and lagged foreign firm entry rate



Source: Aghion et al. (2009a).

Similarly, one can show that more flexible labour markets (which facilitate the process of creative destruction) foster productivity growth more in more advanced countries.

A third lever of productivity growth in advanced countries is graduate education; indeed, frontier innovation requires frontier researchers. Figure 1.2, drawn from Aghion et al. (2009b), shows that research education enhances productivity growth more in US states that are closer to the frontier – i.e. in states with higher GDP per capita (California, Massachusetts, etc.) – whereas a two-year college education is what enhances productivity growth more in less-advanced states (Alabama, Mississippi, etc.). The same is true across countries: higher (and especially graduate) education enhances productivity growth more in countries with higher GDP per capita.

Figure 1.2 Long-term growth effects of spending \$1,000 per person on education in US states

Source: Aghion et al. (2009b).

A fourth lever of productivity growth is the organisation of the financial sector. As shown in Figure 1.3 (drawn from Koch, 2014), a bank-based financial system enhances productivity growth more in less-advanced countries, whereas a more market-based financial system enhances productivity growth more in countries closer to the frontier.

Aghion et al. (2009c) performed cross-country panel regressions of productivity growth on the share of ICT in total value added and found a positive significant coefficient. But interestingly, once they control for product market regulation, the coefficient on ICT becomes non-significant. This in turn suggests that liberalising product markets is key to enhancing productivity growth in developed economies, also because they facilitate the diffusion of the ICT wave throughout the various sectors of the economy. This result is confirmed by Cetto and Lopez (2012); Figure 1.4, taken from their paper, shows that the Eurozone⁵ and Japan suffer from a lag in ICT diffusion compared to the United States.

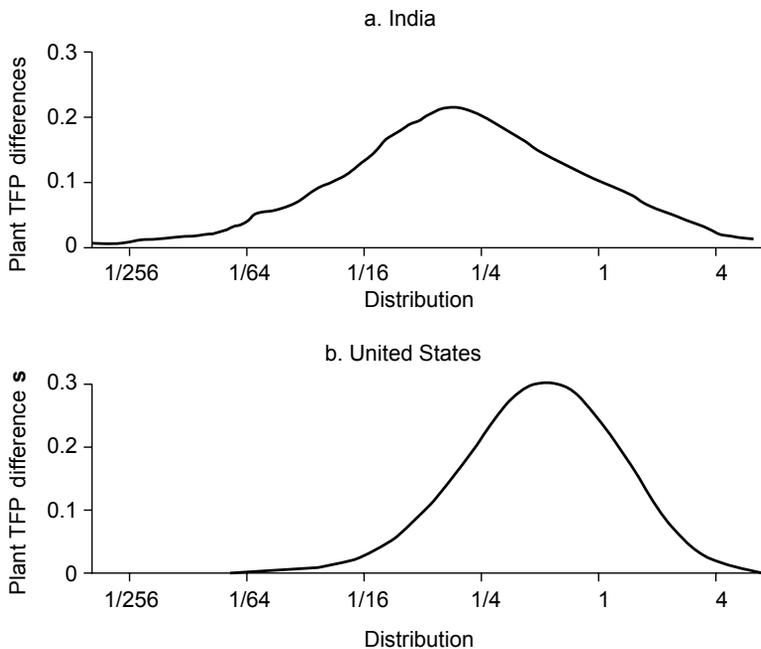
⁵ The Eurozone here is the aggregation of Germany, France, Italy, Spain, the Netherlands, Austria and Finland. These seven countries together represented in 88.5% of the total GDP of the Eurozone in 2012.

an unbalanced country-industry panel dataset covering 15 OECD countries over the period 1987-2007, the authors find that anticompetitive upstream regulations have a significant detrimental effect on productivity growth downstream, and that this effect operates in part (but not entirely) through R&D and ICT investments in downstream industries.

1.4.3 Productivity growth in emerging market economies

Turning now to the sources of productivity growth in emerging market economies, where adaptive innovation and factor accumulation are the main sources of growth, Hsieh and Klenow (2009) have emphasised the importance of input reallocation effects. In particular, if we compare the distribution of firm productivity in India and the United States (see Figure 1.5), we see that the United States has a thinner tail of less productive plants and a fatter tail of more productive plants than India. In other words, it is harder for a more productive firm to grow but also easier for a less productive firm to survive in India than in the United States; the creative destruction process operates more efficiently in the United States.

Figure 1.5 Distribution of plant TFP differences in US vs India



Notes: US mean=1; Higher US TFP due to reallocation – thinner ‘tail’ of less productive plants. Source: Hsieh and Klenow (2009).

This difference can be attributed to various potential factors, in particular, more rigid capital markets and labour/product markets in India, the lower supply of skills in India compared to the United States, the lower quality of infrastructure in India, and the lower quality of institutions to protect property rights and enforce

contracts in India compared to the United States. These factors in turn operate on productivity growth through several potential channels. One particularly interesting channel is that of management practices. Recent work (for a review, see Bloom and Van Reenen, 2010) shows that management practices are far worse in India than in the United States, and that the average management scores across countries are strongly correlated with levels of GDP per capita.

1.5 Technological waves

1.5.1 Two productivity growth waves

Using annual and quarterly data on labour productivity and TFP for 13 advanced countries (the G7 plus Spain, the Netherlands, Finland, Australia, Sweden and Norway) plus the reconstituted Eurozone over the period 1890-2012, Bergeaud et al. (2014) show the existence of two large productivity growth waves during this period.

The first wave, which culminated in 1941, corresponds to the second industrial revolution (in electricity, internal combustion and chemistry) (Gordon, 2000). The second wave, which culminated in 2001, is the ICT wave. The second wave is of a smaller magnitude than the first, and a big question is whether this second wave has now ended in the United States.

1.5.2 Diffusion patterns

Bergeaud et al. (2014, Figure 1.6) show that Japan, the UK and the Eurozone have benefited from both waves, although with delays in both cases. Thus the first wave fully diffused to the current Eurozone, Japan and the UK only after World War II. As for the second productivity wave, so far it has not shown up in the Eurozone or in Japan. Market rigidities contribute to explaining such delays; the lower quality of research and higher education appears to also matter.

5.3 Global breaks

One observes several global breaks in the evolution of productivity growth over the period 1890-2012. Bergeaud et al. (2014) show that there have been types of global break: 1) the global breaks associated with the two World Wars; 2) the global breaks attributable to the two financial crises of 1929 and 2008; and 3) the break corresponding to the global oil supply shock.

Several interesting observations are proposed by Bergeaud et al. (2014) from observing these breaks. First, the global war shocks affected different countries differently; more precisely, they were downward shocks for countries such as France, Germany and Japan where battles were waged, but they correspond to an upward shock for the United States, which was not directly submitted to the confrontations. Second, the rebound from the Great Depression was stronger in the United States and Canada than in other developed countries. Also, most countries exited the Depression through World War II. Third, the impact of the global oil supply shock was generalised, although the United States got in and out of it earlier than the other countries, partly through deregulating its markets.

1.5.4 Country-specific shocks and the role of reforms

Bergeaud et al. (2014) show, in Figure 7 of their paper, a positive break in labour productivity and in TFP growth in Sweden after 1990. This stands in contrast with the case of Japan shown in their Figure 8, where we see no such break but instead decelerating labour productivity and TFP growth since 1980. Our explanation is that Sweden implemented sweeping structural reforms in the early 1990s; in particular, a reform of the public spending system to reduce public deficits, and a tax reform to encourage labour supply and entrepreneurship. No significant reform has taken place in Japan over the past 30 years.

Consider the four countries from Bergeaud et al.'s study that are commonly presented as lead reformers over the past three decades. The reforms initiated in Sweden in the early 1990s saw the rate of TFP growth increase from an average of 0.4% over the period 1976-1992 to an average of 1.9% over the period 1992-2008. Similarly, a 1982 reform in the Netherlands (the Wassenaar Agreement) is associated with an increase from an average TFP growth rate of 0.5% over the period 1977-1983 to an average TFP growth rate of 1.5% over the period 1983-2002. The reforms initiated in the early 1990s in Canada are associated with a break from an average TFP growth rate of 0.3% over the period 1974-1990 to an average rate of 1.1% over the period 1990-2000. Finally, the reforms initiated in the early 1990s in Australia are associated with a break from an average TFP growth rate over the period 1971-1990 of 0.4% to an average growth rate of 1.4% over the period 1990-2002.

These findings are in line with cross-country panel regressions suggesting that structural reforms play a key role in speeding up the diffusion of technological waves.

1.6 Obstacles to firm growth

In Section 3, we reported on recent theoretical and empirical work pointing to the importance of firm size for competitiveness. More precisely, we are after policies that emphasise productivity growth to an extent that should allow firms to eventually reach and pass the thresholds required to become 'open' (i.e. to first become an exporter and then to become involved in FDI and/or outsourcing activities).

There exists an entire literature on firm dynamics and its impact on aggregate productivity growth. On the theory side, the state of the art on the interplay between growth, reallocation and firm dynamics is represented by the recent papers by Klette and Kortum (2004), Acemoglu et al. (2012) and Acigit et al. (2014). These papers build on the Schumpeterian growth paradigm (see Aghion and Howitt, 1992; and Aghion et al., 2013a) to model firms as multi-line producers and innovators. Innovations improve a firm's productivity in producing a particular intermediate input, and they allow an incumbent firm to expand its scope, i.e. the number of product lines it operates in. If the incumbent firm innovates on a new line, the firm drives out the previous producer on that line through Bertrand competition, as it outcompetes the previous producer on that line. Also, a successful innovation by an outsider on a product line currently covered by an incumbent firm eliminates that line from the incumbent firm's

range of products, thereby shrinking the number of product lines covered by that firm.

This framework generates an ergodic steady-state firm size distribution that depends upon the innovation technology, upon government policy towards incumbent firms and/or towards potential entrants, and upon regulatory or credit market characteristics, which will also affect firms' ability to enter and/or grow post-entry.

In particular, this framework can account for various stylised facts about firm dynamics and firm size distribution, including: (i) the firm size distribution is highly skewed; (ii) firm size and firm age are highly correlated (in this framework new firms are one-line firms, and to become large with a sufficient number of lines, a firm needs to have innovated on all these lines and also have survived creative destruction on a sufficient number of lines it used to operate on); (iii) small firms exit more frequently (it takes only one outside innovation to eliminate a one-line firm, whereas it takes several successful outside innovations to eliminate an initially multi-line firm), but the ones that survive tend to grow faster than average (the firm is more likely to be an efficient innovator, and also it can exploit R&D synergies across its multiple lines); (iv) a large fraction of R&D in the United States is done by incumbents; and (v) reallocation of inputs between entrants and incumbents is an important source of productivity growth.

The framework can also explain why factors that inhibit firm size growth in developing countries also inhibit aggregate productivity growth. For example, Akcigit et al. (2014) argue that in developing countries, contractual frictions become more dramatic as firms grow in size – it becomes increasingly hard to avoid hold-up by firm managers as the number of product lines controlled by the firm increases. This in turn inhibits the growth of most efficient firms (i.e. of firms with higher innovation capabilities); such firms have lower incentives to grow as firm owners want to mitigate the hold-up problem with their manager. But this in turn enables less efficient firms to remain active for a longer period before being replaced by more efficient firms.

While contractual incompleteness and lack of trust are obvious obstacles to firm growth, previous studies have also emphasised (a) adjustment costs induced by the R&D and/or advertising of incumbent firms; (b) the administrative costs of creating a new firm; and (c) labour market regulations.

Aghion et al. (2007) present empirical evidence on the effect of financial development on the entry of new firms of different size and on the post-entry growth of successful entrants. They use harmonised firm-level data on entry and post-entry growth by industry, size class and over time for a sample of industrialised, transition and Latin American countries (see Bartelsman et al., 2004). They consider two main indicators of financial development: the ratio of private credit, and stock market capitalisation. They instrument these financial development variables using a detailed set of regulatory indicators that characterise the banking and securities markets. Also, following Rajan and Zingales (1998), to minimise problems of omitted variable bias and other misspecifications, they interact different indicators of financial development with the relative dependence on external financing of the corresponding sector in the United States.

The main results in the study by Aghion et al. (2007) are as follows. First, higher financial development enhances new firm entry in sectors that depend more heavily upon external finance. Second, the entry of the smallest firms benefits

the most from higher financial development, whereas financial development has either no effect or a negative effect on entry by larger firms. Third, financial development enhances post-entry growth of firms in sectors that depend more upon external finance, even when controlling for labour market regulations.⁶

The effect of regulations on firm dynamics and firm size is itself a fascinating topic that has barely been touched upon. An interesting paper by Garicano et al. (2013) analyses the static welfare effects of the 50 employee regulatory threshold in France, and points to a significant source of allocative inefficiency (namely, the inefficient concentration of firm size just below the threshold). Yet, how such a threshold or other types of regulations more generally affect the size distribution of firms and aggregate productivity growth remains an open question.

1.7 Do we still need vertically targeted policies?

1.7.1 The debate

The change of emphasis from industry-level to firm-level competitiveness, the evidence on the relationship between firm-level competitiveness and firm-level productivity, and finally our discussion of the determinants of productivity growth and policies to enhance it – all this points towards giving priority to ‘horizontal targeting’, i.e. policies (competition, labour market liberalisation, patent and R&D policy, etc.) that enhance productivity growth in all sectors, instead of focusing on ‘vertical targeting’, i.e. policies aimed at promoting particular industries in the worldwide competition with similar industries in other countries.

Vertical targeting used to be popular in the aftermath of World War II. For example, the World Bank and other international financial institutions would welcome the import substitution policies in Latin American countries whereby local industries would more fully benefit from domestic demand. Similarly, they would support East Asian countries such as Korea or Japan when they engaged in export promotion, for example through tariffs and non-tariff barriers and partly through maintaining undervalued exchange rates. For at least two or three decades after World War II, these policies – which form part of what is commonly referred to as ‘industrial policy’ – remained fairly uncontroversial, as both groups of countries were growing at fast rates.

However, vertical targeting has come under increasing criticism since the early 1980s among academics and policy advisers in international financial institutions. In particular, it has been criticised for allowing governments to pick winners and losers in a discretionary fashion, and consequently for increasing the scope for capture of governments by local vested interests. Empirical studies by Frankel and Romer (1999) and Wacziarg (2001) pointing to a positive effect of trade liberalisation on growth would, of course, reinforce the case against vertical targeting, as would recent work on competition and growth (see, for example, Aghion et al., 2005; and Aghion et al., 2008).

⁶ Previous work on the subject includes Beck et al. (2004a) who find that financial development is more growth-enhancing for industries that rely more on small firms (this is measured by the share of small firms in the respective US industry). In the same vein, Beck, Demirgüç-Kunt and Maksimovic (2005) use a firm survey to assess firms' perception of financial constraints. They find that small firms tend to be more affected by financial as well as legal and corruption issues than larger firms.

However, three phenomena that have occurred in the recent period invite us to rethink the issue. The first is climate change and the increasing awareness of the fact that without government intervention aimed at encouraging clean production and clean innovation, global warming will intensify and generate all kinds of negative externalities (droughts, deforestations, migrations, conflicts) worldwide. The second is the recent financial crisis, which revealed the extent to which laissez-faire policies has led several countries – especially in southern Europe – to allow the uncontrolled development of non-tradable sectors (in particular, real estate) at the expense of tradable sectors that are more conducive to long-term convergence and innovation. The third is the emergence of China, which has become so prominent on the world economic stage in large part thanks to its constant pursuit of industrial policy. Also, we now see an increasing number of scholars (in the United States in particular) denouncing the danger of laissez-faire policies that lead developed countries to specialise in upstream R&D and in services, while outsourcing all manufacturing tasks to developing countries where unskilled labour costs are lower. They point to the fact that countries such as Germany or Japan have managed to maintain intermediate manufacturing segments of their value chain better through pursuing more active industrial policies, and that this in turn has allowed them to benefit more from outsourcing the other segments.

As mentioned above, the most recurrent counter-argument to industrial interventionism is the ‘picking winners’ argument. True, industrial policy is to some extent always about ‘picking winners’, but as Vince Cable, the former UK Secretary of State for Business, Innovation and Skills, points out, “the ‘winners’ in this sense are the skills we judge we will need for the future, and the sectors they support”. However, we will argue below that the ‘picking winners’ argument loses bite first when the government chooses to pick sectors rather than particular firms, and second when it ‘governs’ its sectoral interventions in such a way that preserves or even enhances competition and Schumpeterian selection within the corresponding sectors. A second criticism of traditional industrial policy is the risk of capture and rent-seeking behaviour that it involves. Here again, setting clear principles for the selection of sectors and for the governance of support to these sectors (competitiveness, exit mechanisms, etc.) should help address this criticism.

More fundamentally, a main theoretical argument supporting growth-enhancing sectoral policies is the existence of knowledge spillovers. For example, firms that choose to innovate in dirty technologies do not internalise the fact that current advances in such technologies tend to make future innovations in dirty technologies more profitable. More generally, when choosing where to produce and innovate, firms do not internalise the positive or negative externalities this might have on other firms and sectors. A reinforcing factor is the existence of credit constraints which may further limit or slow down the reallocation of firms towards new (more growth-enhancing) sectors. Now, one can argue that the existence of market failures on its own is not sufficient to justify sectoral intervention. On the other hand, there are activities – typically, high-tech sectors – that generate knowledge spillovers to the rest of the economy and where assets are highly intangible, which in turn makes it more difficult for firms to borrow from private capital markets to finance their growth. Then there might indeed be a case for subsidising entry and innovation in the corresponding sectors, and for doing so in such a way that guarantees fair competition within the sector.

Note that the sectors that come to mind are always the same four or five sectors: energy, biotech, ICT and transportation.

1.7.2 Rethinking the design and governance of industrial policy

To our knowledge, the first convincing empirical study in support of properly designed industrial policy was by Nunn and Trefler (2010). These authors use micro data on a set of countries to analyse whether – as suggested by the ‘infant industry’ argument – the growth of productivity in a country is positively affected by the extent to which tariff protection is biased in favour of activities and sectors that are ‘skill-intensive’, i.e. that use more highly skilled workers. They find a significant positive correlation between productivity growth and the ‘skill bias’ due to tariff protection. Of course, such a correlation does not necessarily mean there is causality between the skill bias due to protection and productivity growth; the two variables may themselves be the result of a third factor, such as the quality of institutions in the countries considered. However, Nunn and Trefler show that at least 25% of the correlation corresponds to a causal effect. Overall, their analysis suggests that adequately designed (here, skill-intensive) targeting may actually enhance growth, not only in the sector which is being subsidised but also in the country as a whole.

More recently, Aghion et al. (2012a) argue that sectoral policy should not be systematically opposed to competition policy. First, they develop a simple model showing that targeted subsidies can be used to induce several firms to operate in the same sector, and that the more competitive the sector, the more firms will be induced to innovate in order to ‘escape competition’ (see Aghion et al., 2005). Of course, a lot depends upon the design of industrial policy. Such policy should target sectors, not particular firms (or ‘national champions’). This in turn suggests new empirical analyses in which productivity growth, patenting or other measures of innovativeness and entrepreneurship are regressed over some measures of sectoral intervention interacted with the degree of competition in the sector, and also with the extent to which intervention in each sector is not concentrated on one single firm, but rather distributed over a larger number of firms.

Unfortunately, data showing how much state aid each sector receives are not available for EU countries. Thus, to look at the interaction between state subsidies to a sector and the level of product market competition in that sector, Aghion et al. (2012a) use Chinese firm-level panel data. More precisely, they look at all industrial firms from the Chinese National Business Survey, an annual survey of all firms with sales of more than 5 million RMB. The sample period is 1988-2007, and the survey contains information on inputs and outputs, firm-level state subsidies, and so on. Product market competition is measured by 1 minus the Lerner index, which in turn is calculated as the ratio of operating profits minus capital costs over sales. The authors show that TFP, TFP growth and product innovation (defined as the ratio between output value generated by new products to total output value) are all positively correlated with the interaction between state aid to the sector and market competition in the sector. Thus, the more competitive the recipient sector, the more positive the effects of targeted state subsidies to that sector on TFP, TFP growth and product innovation in that sector. In fact, the authors show that for sectors with a low degree of competition the effects are negative, but the effects become positive in sectors

with a sufficiently high degree of competition. Finally, the authors show that the interaction between state aid and product market competition in the sector is more positive when state aid is less concentrated. In fact, if one restricts attention to the second quartile in terms of degree of concentration of state aid (this refers to sectors where state aid is not very concentrated), then state aid has a positive effect on TFP and product innovation in all sectors with an above-median level of product market competition.

1.7.3 Climate

Firms in a *laissez-faire* economy may innovate in ‘the wrong direction’, for example in polluting energy activities just because they have acquired expertise in such activities, not taking into account the environmental and also the knowledge externalities that their choice entails. Aghion et al. (2010) explore a cross-country panel data set of patents in the automotive industry. They distinguish between ‘dirty innovations’ that affect combustion engines and ‘clean’ innovations, such as those relating to electric cars. Then they show that the larger the stock of past ‘dirty’ innovations by a given entrepreneur, the ‘dirtier’ the current innovations by the same entrepreneur. This ‘path dependence’ phenomenon, together with the fact that innovations have been mostly dirty so far, implies that in the absence of government intervention, our economies would generate too many dirty innovations. There is therefore a role for government intervention to ‘redirect technical change’ towards clean innovations.

As argued in Acemoglu et al. (2012), delaying such directed intervention not only leads to further deterioration of the environment. In addition, the dirty innovation machine continues to strengthen its lead, making the dirty technologies more productive and widening the productivity gap between dirty and clean technologies even further. This widened gap in turn requires a longer period for clean technologies to catch up and replace the dirty ones. As this catching-up period is characterised by slower growth, the cost of delaying intervention, in terms of foregone growth, will be higher. In other words, delaying action is costly.

Not surprisingly, the shorter the delay and the higher the discount rate (i.e. the lower the value put on the future), the lower the cost will be. This is because the gains from delaying intervention are realised at the start in the form of higher consumption, while the loss occurs in the future through more environmental degradation and lower future consumption. Moreover, because there are basically two problems to deal with – the environmental one and the innovation one – using two instruments proves to be better than using one. The optimal policy involves using (i) a carbon price to deal with the environmental externality, and, at the same time, (ii) direct subsidies to clean R&D (or a profit tax on dirty technologies) to deal with the knowledge externality. This again calls for vertical targeting.⁷

⁷ Of course, one could always argue that a carbon price on its own could deal with both the environmental and the knowledge externalities at the same time (discouraging the use of dirty technologies also discourages innovation in dirty technologies). However, relying on the carbon price alone leads to excessive reduction in consumption in the short run. And because the two-instrument policy reduces the short-run cost in terms of foregone short-run consumption, it reinforces the case for immediate implementation, even for values of the discount rate under which standard models would suggest delaying implementation.

1.7.4 Summarising

Overall, our discussion in this section suggests that adequately targeted sectoral intervention, for example to more skill-intensive or more competitive sectors, can be growth enhancing. Also, we have argued in favour of not concentrating subsidies across firms in a sector. However, this is just the starting point for what we see as a much broader research programme on how to govern industrial policy so as to make it more competition friendly and more innovation enhancing. In particular, how can industrial policy be designed so as to ensure that projects that turn out to be non-performing will not be refinanced? How should governments update their doctrines and competition policy practices so as to factor in renewed thinking industrial policy design and implementation? The conjunction of the debates on climate change, the recent financial crisis and the new dominance of China in the world market reinforce our conviction that while market competition is certainly a main engine of growth, specialisation cannot be left entirely to the dynamics of *laissez-faire*. Also, one increasingly realises that the specialisation model, whereby the most advanced countries focus on upstream R&D and services and outsource everything else to emerging market economies, may not be sustainable in the long run.

1.8 Implications for policy design in emerging markets: The example of China

China is deservedly admired worldwide for its outstanding growth performance over the past three decades; this forces modesty and humility on people like me who are asked to provide economic advice. Yet, the growth in China so far has largely been a 'catch-up' growth, i.e. growth based on imitating or adapting technologies introduced elsewhere.

This catch-up growth has been spurred by the market reforms under the dual track approach as of the early 1980s, and by the establishment of a system of growth-based yardstick competition between provincial leaders. This in turn has favoured the reallocation of resources and investment from agriculture to industry, and from state-owned enterprises to (credit-constrained) new private enterprises. And technological catch-up has been further enhanced by encouraging foreign direct investment.

While more catch-up or reallocation-based growth can be achieved by improving management practices in existing firms (see our discussion above), by further liberalising labour flows from rural to urban areas, by further developing the financial sector, and by liberalising capital flows (as China already plans to do), there are several reasons to believe that this will not be sustainable in the long run. In particular, (i) the efficiency gains from reallocating resources from agriculture to industry and from absorption of imported technologies will be exhausted once the reallocation is complete; and (ii) wage increases will reduce China's comparative advantage in what it currently exports to the rest of the world.

Then the questions naturally arise: How can China avoid the middle-income trap and make a successful transition from catch-up growth to innovation-led growth? And how can China achieve higher quality growth in this process? The above discussion on firm-level productivity growth as the ultimate source of

competitiveness, as well as on the drivers of productivity growth, suggests five pillars of an innovation-based economy:

1) **Competition and creative destruction.** Frontier innovation is fostered by competition and free entry to a much larger extent than imitation. The reason for this is both that incumbent firms at the technological frontier can escape competition and the threat of entry by innovating, and that most path-breaking innovations are made by new entrants. Checks and balances are necessary to guarantee free entry and full competition, because this helps minimise the scope for collusion between (local) politicians and (large) incumbent firms.

2) **Top research universities, i.e. universities with very high Shanghai rankings.** Recent work on the subject suggests that to achieve such rankings, one needs not only to invest more in the university system, but also to grant universities autonomy on budget management, wage policy, hiring/firing decisions and the design of programmes. This autonomy has to come hand-in-hand with more effective competition between universities as well as between researchers. Thus, as for other sectors of the economy, here also less upward accountability has to be replaced by more downward accountability and competitive pressure.

3) **A dynamic labour market system** which combines (i) flexibility for firms to hire and fire; (ii) a good training system to help workers rebound from one job to another; and (iii) a good social safety net, i.e. with well-developed portable social security and pension rights from job to job, and also with a generous unemployment benefit system (conditional upon the unemployed worker training and then accepting a new job). A 'flexsecurity' system such as this makes creative destruction, and therefore innovation-led growth, work at greater speed.

4) **A financial system that relies more on venture capital, private equity and stock markets;** the reason being that innovative investments are riskier and therefore investors require both to obtain a share of the upside returns and to obtain control rights.

How can one ensure that innovation-led growth will be high-quality growth? My feeling is that the Chinese leadership has been concerned by two negative by-products of growth so far: the deteriorating environment, and the very fast increase in inequality. In other words the challenge is to achieve inclusive and sustainable innovation-led growth! As it turns out, implementing the above pillars helps achieve these objectives, in particular: (i) the combination of competition, education and 'flexsecurity' enhances social mobility; and (ii) the checks and balances (at the local level) which guarantee full competition should also help improve the environment.

A natural question then arises: Which organisational and/or institutional changes (if any) does China need to introduce in order to move toward full-steam innovation-led growth? Obviously, we do not have the answer to this question at hand because we lack knowledge on how the current institutional system is organised and how it works in practice.

Yet, empirical and casual evidence suggests that a smart state can stimulate its innovation-led machinery by: (i) setting up a fiscal system which achieves the triple goal of (a) raising revenue to make innovation-enhancing investments in education, universities infrastructure, (b) being redistributive to avoid excessive inequality and poverty traps, and (c) encouraging innovation by not expropriating innovators; and (ii) setting up adequate institutional mechanisms to strengthen checks and balances on the different levels of government to make sure both that

competition is fully enforced (as argued above) and that state investments aimed at enhancing innovation are properly targeted and monitored.

It would appear somewhat paradoxical to recommend that China move from imitation-led to innovation-led growth by simply imitating the institutional arrangements of existing innovation-led economies. Instead, China must find its own way to reform its state institutions so as to make the above pillars work fully. It must find its own answers to questions such as the following. How can we set up fully effective competition policy instruments and mechanisms starting from the current Chinese institutional context? Which contractual, organisational or institutional changes should we introduce, in particular at the regional/local level, in order for China to reach full steam in implementing sustainable and inclusive innovation-led growth? How can we factor in environmental and social (i.e. inclusiveness) dimensions in addition to GDP growth when evaluating regional or local leaders and organising the yardstick competition among them? How can we improve the tax and welfare system to achieve best standards and practices among innovating countries, and in particular to reconcile the need for redistribution and the need to finance good public infrastructure and services with innovation incentives.

1.9 Conclusion

In this chapter we have taken on board modern trade economics, and in particular the idea that a country's competitiveness boils down to the competitiveness of its individual enterprises. We have then reported on recent empirical work showing that firm-level competitiveness is related to firms' productivity and their ability to grow. We have looked at determinants of firm-level productivity, and also at potential obstacles that may inhibit firm size growth. Finally, we have argued that while enhancing firm-level productivity growth calls first for horizontal policies (product and labour market liberalization, trade liberalisation, more education investment, etc.), there may yet be a case for vertically targeted (sectoral) policies provided they are properly designed and governed.

To conclude our discussion, we would like to touch upon the delicate issue of macroeconomic policy. Recent studies (e.g. Aghion, Hemous, and Kharroubi, 2014; Aghion et al., 2012b) performed at the cross-country/cross-industry level show that more counter-cyclical fiscal and monetary policies enhance growth. Fiscal policy counter-cyclicity refers to countries increasing their public deficits and debt in recessions, and reducing them in upturns. Monetary policy counter-cyclicity refers to central banks letting real short-term interest rates fall in recessions, and having them increase again during upturns. Such policies can help credit-constrained or liquidity-constrained firms to pursue innovative investments (R&D, skills and training, etc.) over the cycle in spite of credit tightening during recessions, and it also helps maintain aggregate consumption, and therefore firms' market size, over the cycle, as argued in the previous section (see Chapter 13 in Aghion and Howitt, 2009). This suggests that an innovation-based economy would benefit from more counter-cyclical macroeconomic policies – with higher deficits and lower real interest rates in recessions, and lower deficits and higher real interest rates in booms – in order to help credit-constrained, innovative firms maintain their R&D and other types of growth-enhancing investments over the business cycle.

References

- Acemoglu, D., P. Aghion, L. Burszty and D. Hemous (2012), "The Environment and Directed Technical Change", *American Economic Review* 102(1), pp. 131-166.
- Aghion, P. and P. Howitt (1992), "A Model of Growth through Creative Destruction", *Econometrica* 60(2), pp. 323-351.
- Aghion, P. and P. Howitt (2009), *The Economics of Growth*, Cambridge, MA: MIT Press.
- Aghion, P., N. Bloom, R. Blundell, R. Griffith and P. Howitt (2005), "Competition and Innovation: An Inverted-U Relationship", *Quarterly Journal of Economics* 120(2), pp. 701-728.
- Aghion, P., T. Fally and S. Scarpetta (2007), "Credit Constraints as a Barrier to the Entry and Post-Entry Growth of Firms", *Economic Policy* 22, pp. 731-779.
- Aghion P., R. Burgess, S. Redding and F. Zilibotti (2008), "The Unequal Effects of Liberalization: Evidence from Dismantling the License Raj in India", *American Economic Review* 94(4), 1397-1412.
- Aghion, P., R. Blundell, R. Griffith, P. Howitt and S. Prantl (2009a), "The Effects of Entry on Incumbent Innovation and Productivity", *Review of Economics and Statistics* 91(1), pp. 20-32.
- Aghion, P., L. Boustan, C. Hoxby and J. Vandenbussche (2009b), "Exploiting States' Mistakes to Identify the Causal Effects of Higher Education on Growth", mimeo, Harvard University, Cambridge, MA.
- Aghion, P., P. Askenazy, R. Bourles, G. Cette and N. Dromel (2009c), "Education, Market Rigidities and Growth", *Economics Letters* 102(1), pp. 62-65.
- Aghion, P., M. Dewatripont, C. Hoxby, A. Mas-Colell and A. Sapir (2010), "The Governance and Performance of Universities: Evidence from Europe and the US", *Economic Policy* 25, pp. 7-59.
- Aghion, P., M. Dewatripont, L. Du, A. Harrison and P. Legros (2012a), "Industrial Policy and Competition", mimeo, Harvard University, Cambridge, MA.
- Aghion, P., E. Farhi and E. Kharroubi (2012b), "Monetary Policy, Liquidity and Growth", mimeo, Harvard University, Cambridge, MA.
- Aghion, P., U. Akcigit and P. Howitt (2013a), "What Do We Learn from Schumpeterian Growth Theory?", mimeo, Harvard University, Cambridge, MA.
- Aghion, P., A. Dechezlepretre, D. Hemous, R. Martin and J. Van Reenen (2013b), "Carbon taxes, Path Dependence and Directed Technical Change: Evidence from the Auto Industry", mimeo, Harvard University.
- Aghion, P., D. Hemous and E. Kharroubi (2014), "Cyclical Fiscal Policy, Credit constraints and Industry Growth", *Journal of Monetary Economics* 62(C), pp. 41-58.
- Akcigit, U., H. Alp and M. Peters (2014), "Lack of Selection and imperfect Managerial Contracts: Firm Dynamics in Developing Countries", mimeo, University of Pennsylvania.
- Altomonte, C., T. Aquilante and G. Ottaviano (2012), "The Triggers of Competitiveness: The EFIGE Cross-Country Report", Bruegel Blueprint No 17, Brussels.

- Antràs, P., R. Segura-Cayuela and D. Rodríguez-Rodríguez (2010), "Firms in International Trade (with an Application to Spain)", SERIEs Invited Lecture at the XXXVth Symposium of the Spanish Economic Society, 16-18 December, Madrid.
- Bartelsman, E., J. Haltiwanger and S. Scarpetta (2004), "Microeconomic Evidence of Creative Destruction in Industrial and Developing Countries", IZA Discussion Papers No. 1374, Bonn.
- Beck, T., A. Demirguc-Kunt, L. Laeven and R. Levine (2004), "Finance, Firm Size and Growth", World Bank Research Working Paper No. 3485, Washington, DC.
- Beck, T., A. Demirguc-Kunt and V. Maksimovic (2005), "Financial and Legal Constraints to Firm Growth: Does Size Matter?", *Journal of Finance* 60, 137-177.
- Bergeaud, A., G. Clette and R. Lecat (2014), "Productivity Trends from 1890 to 2012 in Advanced Countries", Banque de France Working Paper No. 475, Paris.
- Bernard, A., J. B. Jensen, S. Redding and P. Schott (2011), "The Empirics of Firm Heterogeneity and International Trade", NBER Working Paper No. 17627, Cambridge, MA.
- Bloom, D., M. Draca and J. Van Reenen (2011), "Trade Induced Technical Change? The Impact of Chinese Imports on Innovation, IT and Productivity", CEP Discussion Paper No. 1000, London.
- Bloom, D. and J. Van Reenen (2010), "Why Do Management Practices Differ Across Firms and Countries?", *Journal of Economic Perspectives* 24(1), pp. 203-224.
- Clette, G. and J. Lopez (2012), "ICT Demand Behavior: An International Comparison", *Economics of Innovation and New Technology, Taylor and Francis Journals* 21(4), pp. 397-410.
- Clette, G., J. Lopez and J. Mairesse (2013), "Upstream Product Market Regulations, ICT, R&D and Productivity", NBER Working Paper No. 19488, Cambridge, MA.
- Frankel, J. and D. Romer (1999), "Does Trade Cause Growth", *American Economic Review* 89(3), pp. 379-399.
- Garicano, L., Lelarge, C. and J. Van Reenen (2013), "Firm Size Distortion and the Productivity Distribution: Evidence from France", NBER Working Paper No. 18841, Cambridge, MA.
- Gordon, R. (2000), "Interpreting the 'One Big Wave' in U.S. Long-Term Productivity Growth", NBER Working Paper No. 7752, Cambridge, MA.
- Helpman, E. (1981), "International Trade in the Presence of Product Differentiation, Economies of Scale and Monopolistic Competition: A Chamberlin-Heckscher-Ohlin Model", *Journal of International Economics* 11(3), pp. 305-340.
- Helpman, E. (2004), *The Mystery of Economic Growth*, Cambridge, MA: Harvard University Press.
- Hsieh, C.-T. and P. Klenow (2009), "Misallocation and Manufacturing TFP in China and India", *Quarterly Journal of Economics* 124(4), pp. 771-807.
- Klette, T. and S. Kortum. (2004), "Innovating Firms and Aggregate Innovation", *Journal of Political Economy* 112, pp. 986-1018.
- Koch, W. (2014), "Bank-Based Versus Market-Based Finance as Appropriate Institution", Mimeo, Université du Québec à Montréal (UQAM).
- Krugman, P. (1980), "Scale Economies, Product Differentiation and the Pattern of Trade", *American Economic Review* 70(5), pp. 950-959.

- Melitz, M. (2003), "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity", *Econometrica* 71(6), pp. 1695-1725.
- Nunn, N. and D. Trefler (2010), "The Structure of Tariffs and Long-Term Growth", *American Economic Journal: Macroeconomics* 2(4), pp. 158-94.
- Rajan, R. and L. Zingales (1998), "Financial Dependence and Growth", *American Economic Review* 88(3), pp. 559-586.
- Solow, R. (1956), "A Contribution to the Theory of Economic Growth", *Quarterly Journal of Economics* 70(1), pp. 65-94.
- Syverson, C. (2004), "Market Structure and Productivity: A Concrete Example", *Journal of Political Economy* 112(6), pp. 1181-1222.
- Wacziarg, R. (2001), "Measuring the Dynamic Gains from Trade", *The World Bank Economic Review* 15(3), pp. 393-429.

2 R&D Policies and Economic Growth

Ufuk Akcigit

University of Pennsylvania

2.1 Introduction

How should the optimal R&D policy be designed? This question is at the heart of any policy debate which targets technological progress through R&D and innovation. Many governments are providing massive subsidies to foster innovation. As an example, the United States spends more than \$130 billion per year at the federal level to support innovation (NSF + NIH + Army Research Office + R&D tax credit).¹ The proponents of R&D subsidies have argued that R&D has spillovers that are not internalised by the innovating firms. The opponents claim that product market competition already provides sufficient incentives to firms and that any additional subsidy would be wasteful.

In this chapter, summarising the findings from recent research, I argue that there are at least two more dimensions that the design of optimal R&D policy should consider. First, R&D support could distort the selection mechanism among firms and may be welfare reducing. Second, there are different types of research investments – for instance, basic and applied – and the spillovers associated with each type of research could be very different. Identifying these two margins and incorporating them into the current policy debate is an important step forward. Below I describe two recent studies that take important steps in this direction.

2.2 R&D policies and firm selection

The goal of R&D policies is to incentivise firms to undertake greater R&D investment, produce more innovations, increase productivity, and create more jobs. However, these policies do not affect every firm in the economy in the same way. For instance, Criscuolo et al. (2012) have shown that large incumbents are better at obtaining government subsidies. One can therefore argue that R&D subsidies to incumbents might be preventing the entry of new firms, and thus slowing down the replacement of inefficient incumbents by more productive new entrants. The turnover and factor reallocation between incumbents and entrants is an important source of productivity growth. Foster et al. (2000, 2006) have shown empirically that the reallocation of factors across firms accounts for more than 50% of productivity growth in the United States. Given the importance

¹ <http://www.whitehouse.gov/sites/default/files/microsites/ostp/Fy%202015%20R&D.pdf>

of this reallocation margin, it is necessary for R&D policy to take into account the interaction between innovation and factor reallocation. This is our focus in Acemoglu et al. (2013).

A recent literature has emphasised the importance of firm size and age for the firm-level heterogeneity that is observed in the data (Haltiwanger et al., 2013; Akcigit and Kerr, 2015). In Acemoglu et al. (2013), we use data from the US Census Bureau's Longitudinal Business Database and Census of Manufacturers, the National Science Foundation's Survey of Industrial Research and Development, and the NBER Patent Database. Our analysis focuses on innovative firms that were in operation during the period 1987-1997,² and our sample covers over 98% of the industrial R&D conducted in the United States during this period. The empirical heterogeneities are summarised in Figures 2.1 to 2.4.

Figure 2.1 Transition rates

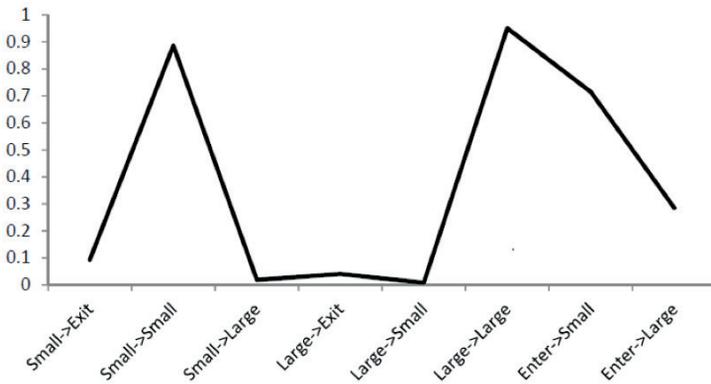
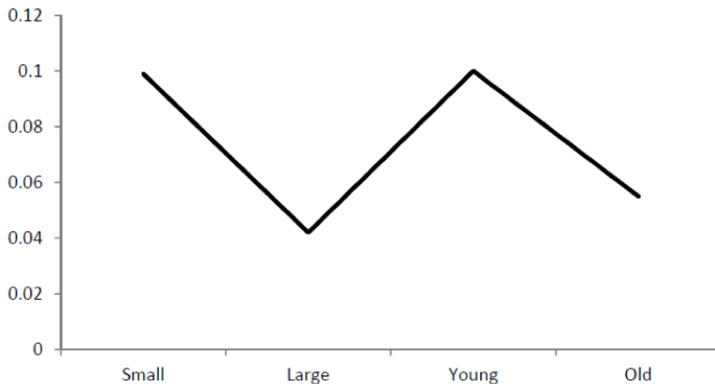
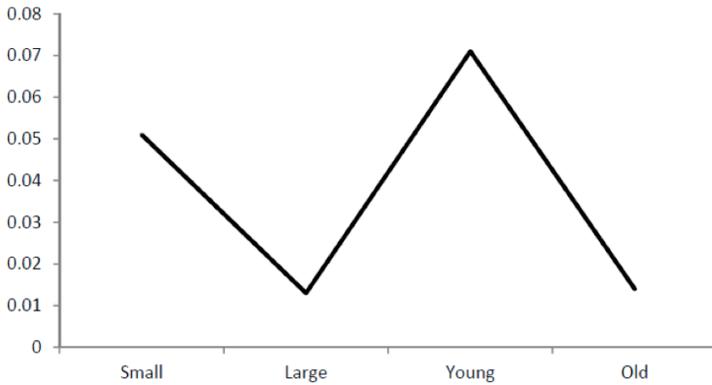
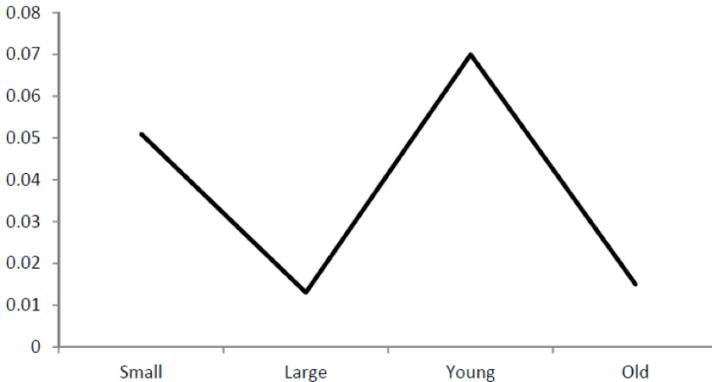


Figure 2.2 R&D intensity



² Non-innovative firms, by definition, do not participate in this process and nor do they compete for these resources; hence, including firms that do not conduct innovation in the sample would create a mismatch between both our focus and our model and the data. Though it would be possible to add another selection margin to the model whereby non-innovative firms choose to transition into innovation, this appears fairly orthogonal to our focus, and we view it as an area for future work.

Figure 2.3 Sales growth**Figure 2.4** Employment growth

Source: Acemoglu et al. (2013).

Figures 1 to 4 show R&D expenditures by shipments, employment growth and exit rates between small, large, young and old firms. A firm is “small” or “large” depending on its size relative to the median employment in the sample by year; a firm is “young” or “old” depending on whether or not it is older than ten years. The figures clearly indicate that in this sample, small and young firms are more R&D intensive and grow faster.³ Thus, industrial policies that discourage the reallocation of resources towards younger firms might indeed be costly in that they slow the movement of R&D resources from less efficient innovators (struggling incumbents) towards more efficient innovators (new firms).

In Acemoglu et al. (2013), we estimate our model by matching empirical moments capturing key features of firm-level R&D behaviour, shipments growth, employment growth and exit, and the variation of these moments with size and

³ Likewise, in Akcigit and Kerr (2015) we regress firm growth on log firm size and find an estimate of -0.04; and innovation intensity (number of innovations relative to the firm size) on log firm size and find an estimate of -0.18.

age (including those that are plotted in Figures 1-4). We then use the estimated model as a lab to run counterfactual experiments and test the impacts of various R&D policy designs on economic growth and welfare. The policies that we consider include a subsidy to new entrants, a subsidy to R&D by incumbents, and a subsidy for the continued operation of incumbents.

Our main results can be summarised as follows. Interestingly, all the policies that we consider have small effects, and some of them even reduce welfare in the economy. When incumbents are subsidised, both the equilibrium growth rate and welfare decrease. This result might suggest that the decentralised equilibrium is already efficient, and any subsidy in this environment makes the economy move away from its efficient level. To the contrary, the decentralised equilibrium is highly inefficient due to the usual intertemporal R&D spillovers and (Schumpeterian) competition effects. However, in this model there is another important margin: firm selection.

In order to understand the role of selection, we first solve for the economy's allocation from the viewpoint of a social planner who internalises all the externalities of R&D spending. In particular, we assume that the social planner can observe firm types. What we find is that the social planner forces low-type firms to exit the economy much more frequently, so that all their production resources are reallocated to the high-type firms. Then we turn to the public policy experiments, in which we assume that the policymaker cannot observe firm types but has access to the usual policy tools such as an R&D subsidy, an entry subsidy and a subsidy/tax to firm operations. What we find is that the optimal policy requires a substantial tax on the operation of incumbents, combined with an R&D subsidy to incumbents. The reason for this result is that taxing operations makes it harder for low-type firms to survive and forces them to exit. This way, the freed-up factors of production are reallocated to high-type firms, which make use of them much more effectively. Our analysis also highlights the importance of the entry subsidy and the incumbent R&D subsidy – these subsidies would not be as effective if the selection margin were ignored.

Overall, our general equilibrium analysis, which incorporates both reallocation and selection effects, highlights the fact that the economy in equilibrium might contain too many low-type firms, and policies that ignore the selection effect might help these low-type firms to survive. Another point that is highlighted is the fact that intertemporal spillovers are sizable and the overall R&D investment is too little. Therefore, a combination of R&D subsidies and taxes on firm operations could be an effective way of providing innovation incentives to firms, while also leveraging the selection margin in the economy.

2.3 Basic versus applied R&D

National funds allocated to basic research have been among the top items in many governments' policy agendas. For instance, in a recent report by the US Congress Joint Economic Committee, it is argued that despite its value to society as a whole, basic research is underfunded by private firms precisely because it is performed with no specific commercial applications in mind. The level of federal funding for basic research is deemed "worrisome" and it is claimed that it must be increased in order to overcome the underinvestment in basic research (JEC,

2010). However, the report also complains about the lack of research studies that actually quantify the extent of this underinvestment and about the lack of data.⁴

For similar reasons, governments introduce programmes to promote collaboration between basic academic researchers and private firms, with the hope that synergies generated from these interactions could lead to breakthrough technological advances. For instance, the United States government has aggressively promoted collaboration between universities and industrial researchers through specific funding programmes. Among many others, the National Science Foundation (NSF) sponsors the Fundamental Research Program for Industry-University Cooperative Research (FRP), the Industry-University Cooperative Research Centers Program (I/UCRC) and Grant Opportunities for Academic Liaison with Industry (GOALI).

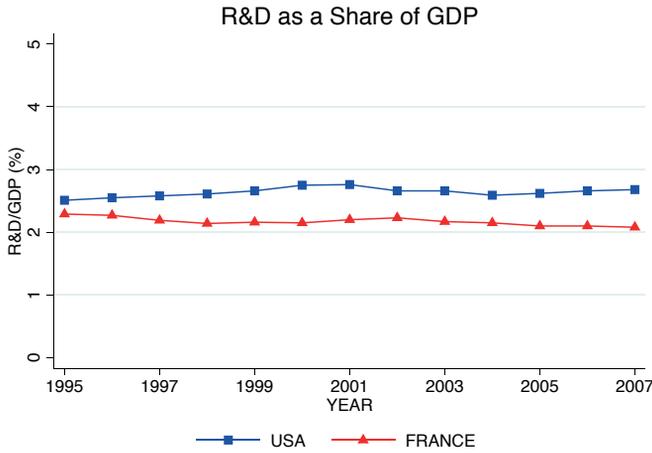
Although the different characteristics of basic and applied research, on the one hand, and academic and corporate research, on the other, have been widely recognised to be of first-order importance by policymakers, these issues have received insufficient attention in the economic literature on productivity and economic growth. In particular, the endogenous growth literature (e.g. Romer, 1990; Aghion and Howitt, 1992) has mainly considered a uniform type of (applied) research and has overlooked basic research investment by private firms.

What are the key roles of basic and applied research for productivity growth? How should R&D policy be geared towards basic versus applied research? What are the incentives of private firms to conduct basic research? How does academic research contribute to innovation and productivity growth? In Akcigit et al. (2014), we attempt to answer these questions. In order to understand the potential inefficiencies involved in different types of research investments and to design appropriate industrial policies to address them, it is necessary to adopt a structural framework that explicitly models the incentives for different types of research investments by private firms. In Akcigit et al. (2014) we take an important step towards developing this theoretical framework, identifying the potential spillovers, and studying their macroeconomic implications for innovation policy.

Our analysis starts with an empirical investigation. Figure 2.5 shows that countries allocate a significant share of their GDP to R&D (around 2-3%). Less well known, however, is the role the composition of this research plays in determining growth, particularly when considering the breakdown between basic and applied research.

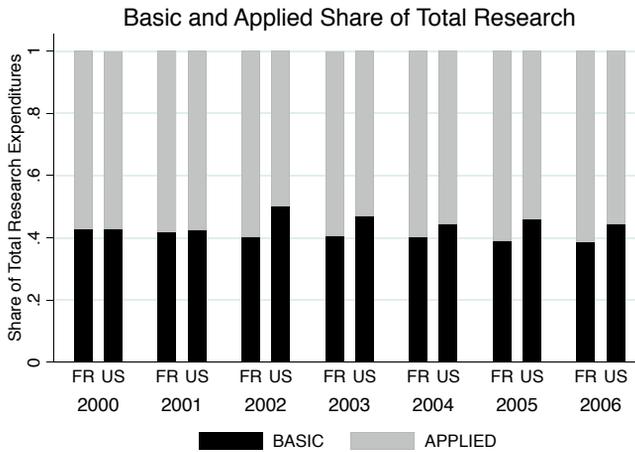
4 http://jec.senate.gov/public/?a=Files.Serve&File_id=29aac456-fce3-4d69-956f-4add06f111c1

Figure 2.5 R&D as a share of GDP



Source: Akcigit et al. (2014).

Figure 2.6 Basic and applied share of total research



Source: Akcigit et al. (2014).

Before we proceed further, it might be helpful to provide the relevant definitions. According to the NSF, basic research investment refers to a “systematic study to gain more comprehensive knowledge or understanding of the subject under study without specific applications in mind”; applied research is defined as a “systematic study to gain knowledge or understanding to meet a specific, recognized need”. Figure 2.6 shows the composition of the overall R&D spending in the United States and in France. The interesting result is that almost half of overall spending goes into basic research.

What kind of spillovers does basic research generate? In our analysis, we follow the influential literature on basic science and consider the possibility that basic research not only generates large spillovers within an industry, but it can also be

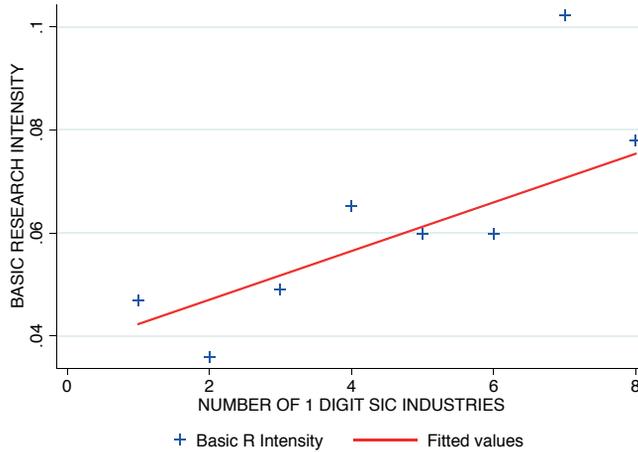
applicable to many different industries. The historical example of Du Pont de Nemours' financing of William Carothers' research serves as a fine showcase of these spillovers. As Nelson (1959) describes it:

Carothers' work in linear super-polymers began as an unrestricted foray into the unknown, with no practical objective in mind. But the research was in a new field in chemistry and Du Pont believed that any new chemical breakthrough would likely be of value to the company. In the course of research Carothers obtained some super-polymers that became viscous solids at high temperatures, and the observation was made that filaments could be made from this material if a rod were dipped in the molten polymer and withdrawn. At this discovery the focus of the project shifted to these filaments and Nylon was the result.

Nylon is now used in many industries including textiles, automobiles and military hardware, three industries in which Du Pont had operations.

Ideally, in order to capture the full return to new scientific knowledge in industries where it could have an application but in which the innovating firm is not present, the innovator would first patent and then license or sell the innovation to other firms in those industries. However, the applications of basic scientific advances are often not immediate and firms are often only able to transform them into patentable applications in their own industries. This is the well-known appropriability problem of basic research that has been discussed in a vast literature. It follows that firms operating in more industries will be able to utilise more facets of a given basic innovation. As Nelson hypothesised it, “[i]t is for this reason that firms which support research toward the basic-science end of the spectrum are firms that have fingers in many pies”. Note that the key concept that is being emphasised here is not the size of the firm per se, but the diversity of its operations. This interesting argument (which we will refer to as "Nelson's hypothesis") will be the central building block of our analysis in this chapter.

We first test Nelson's hypothesis, namely that the main investors in basic research would be those firms that have fingers in many pies. According to this argument, as the range of its products and industries becomes more diversified, a firm's incentive for investing in basic research relative to applied research should increase due to better appropriability of potential knowledge spillovers. To measure multi-industry presence, we count how many distinct SIC codes a firm is present in. Using micro-level data on French firms, Figure 2.7 plots average basic research intensity against the total number of distinct one-digit SIC codes in which the firm is present. The figure also shows a simple linear fitting line.

Figure 2.7 Average basic research intensity against total number of distinct one-digit SIC codes

Source: Akcigit et al. (2014).

Figure 2.7 shows a positive and statistically and economically significant relationship between multi-industry presence and basic research spending. A broader technological base is associated with higher investment in basic research relative to applied research. Our findings are therefore supportive of Nelsons' hypothesis on the link between multi-industry presence and relative research incentives. These correlations are robust to a large variety of potential confounding factors. This result suggests that cross-industry spillovers are sizable, and using the variation in the technology base of firms we can estimate the cross-industry spillovers associated with basic research.

In order to study the policy implications of these spillovers, we build a general equilibrium, multi-industry framework with private firms and a public research sector. Firms conduct both basic and applied research, whereas the public sector focuses exclusively on basic research. In our model, basic research generates fundamental technological innovations and generates spillovers, both within and across industries, that affect subsequent applied innovations.⁵ In line with the 'ivory tower' theory of academic research, basic research by private firms in our model will turn into consumer products faster than that undertaken by public research labs. Applied research, on the other hand, will be done only by private firms and will generate follow-on innovations building on the existing basic knowledge stock.

We then undertake a quantitative investigation of the impacts of various innovation policies on the aggregate economy. We first estimate the model by targeting some of the key moments in the data, especially public and private spending on basic and applied research in France. We use the estimated model to assess the extent of inefficiencies in basic and applied research and to study the implications of several important innovation policies.

⁵ By fundamental innovation, we mean major technological improvements that generate larger than average contributions to the aggregate knowledge stock of society. In addition, these will have long-lasting spillover effects on the size of subsequent innovations within the same field.

Our main results can be summarised as follows. We find that a large fraction of spillovers from basic research across industries is not internalised. As a result, there is a dynamic misallocation of research efforts, which reduces welfare significantly. One striking result is that the decentralised economy and the social planner's economy use the same overall level of resources for research. However, the compositions of the total research efforts are very different. While the social planner allocates more resources to basic research, it allocates less resources to applied research. This implies that the dominant misallocation here is not misallocation between production and research, but among the various types of research activities – in this case, applied and basic research. There is actually overinvestment in applied research in the decentralised economy because of product market competition, whereas there is underinvestment in basic research due to within-industry and cross-industry spillovers that are not internalised.

This raises an important question. To what extent can public policies address this inefficiency? The first policy we analyse is a uniform research subsidy to private firms. In this environment, subsidising overall private research is ineffective since this will over-subsidise applied research, which is already excessive due to product market competition. Therefore, the welfare improvement from such a subsidy is limited unless the policymaker is able to discriminate between types of research projects at the firm level, a difficult task in the real world.

We therefore analyse another policy tool: the level of funding for public research labs. We show that due to the ivory tower nature of public basic research, allocating more money to the academic sector without giving property rights to the researchers (i.e. ownership over their inventions) is not necessarily a good idea. To demonstrate this, we simulate a policy similar to the Bayh-Dole Act enacted in the United States in 1980. We consider alternative scenarios in which public researchers have no property rights, then 50% and 100% property rights. We find a complementarity between the level of property rights and the optimal allocation of resources to academic research. The optimal combination turns out to be granting full property rights to the academic researcher and allocating a larger fraction of GDP to public research. This reduces the welfare gap significantly.

2.4 Conclusions

In this chapter, I have summarised some recent findings from research on optimal innovation policy. The two new elements introduced were firm selection and the distinction between basic and applied research. The former implies that R&D policy could affect firm survival and resource reallocation between more productive and less productive firms, or between incumbent and entrant firms. The latter highlights the fact that different types of research – in this case, basic and applied – could have different spillovers, and R&D policy should take into account its impact on the distinct types of research.

There are still many unexplored directions for future research. One such direction is the labour market consequences of the R&D policies. While the literature typically assumes frictionless labour mobility across firms, it takes time for workers to find new jobs when a firm has to exit. It would be important to study the potential reallocation costs of such policies. Another important issue is the transitional dynamics. The current focus of the literature is typically on

steady-state dynamics. Clearly any new policy is likely to entail a transition path that might generate additional costs for the economy. These are important questions that will we hope will be answered in future research.

References

- Acemoglu, D., U. Akcigit, N. Bloom and W. Kerr (2013), "Innovation, Reallocation and Growth", NBER Working Paper No. 18993, Cambridge, MA.
- Aghion, P. and P. Howitt (1992), "A Model of Growth Through Creative Destruction", *Econometrica* 60(2), pp. 323-351.
- Akcigit, U. and W. Kerr (2015), "Growth through Heterogeneous Innovations", University of Pennsylvania Working Paper, Philadelphia, PA.
- Akcigit, U., D. Hanley and N. Serrano-Velarde (2014), "Back to Basics: Basic Research Spillovers, Innovation Policy and Growth", NBER Working Paper No. 19473, Cambridge, MA.
- Criscuolo, C., R. Martin, H. Overman and J. Van Reenen (2012), "The Causal Effects of Industrial Policy", CEP Working Paper No. 1113, London.
- Foster, L., J. Haltiwanger and C.J. Krizan (2000), "Aggregate Productivity Growth: Lessons from Microeconomic Evidence", in C.R. Hulten, E.R. Dean and M.J. Harper (eds), *New Developments in Productivity Analysis*, Chicago, IL: University of Chicago Press.
- Foster, L., J. Haltiwanger and C.J. Krizan (2006), "Market Selection, Reallocation and Restructuring in the U.S. Retail: Trade Sector in the 1990s", *Review of Economics and Statistics* 88(4), pp. 748-758.
- Haltiwanger, J., R. Jarmin and J. Miranda (2013), "Who Creates Jobs? Small Versus Large Versus Young", *Review of Economics and Statistics* 95(2), pp. 347-361.
- Joint Economic Committee (JEC) (2010), "The 2010 Joint Economic Report", Washington, DC: Government Printing Office.
- Nelson, R.R. (1959), "The Simple Economics of Basic Research", *Journal of Political Economy* 67(3), pp. 297-306.
- Romer, P.M. (1990), "Endogenous Technological Change", *Journal of Political Economy* 98(5-2), pp. S71-S102.

3 Innovation and Business Growth

William R. Kerr

Harvard Business School

3.1 Introduction

Innovation and the pursuit of new business opportunities is essential for growth at the firm level; moreover, it provides the foundation for an economy to achieve new levels of technological prowess, productivity and, ultimately, prosperity. This chapter describes recent work in economics and management scholarship on how firms grow. Given the other contributions in this conference volume, we focus specifically on questions surrounding the types of innovations that large and small firms pursue and how it impacts their relative growth rates. Developing evidence suggests that as firms become larger they have trouble maintaining the external innovations that are most powerful for growth, instead focusing increasingly on internal work and enhancements. Section 3.2 outlines a theoretical model for these dynamics to fix ideas and highlight some key economic considerations. In some cases, the growing internal focus on firm size is optimal, but in most cases it is not and it reflects struggles of larger companies to maintain dynamic capabilities that they otherwise desire. Section 3.3 then provides a case study of IBM – how it observed these limitations within itself in the late 1990s, and then the actions it took to correct the gaps. We use the IBM story to highlight in Section 3.4 several emerging best practices for how firms can best structure themselves to maintain the innovations that are important for their growth.

3.2 Theoretical background

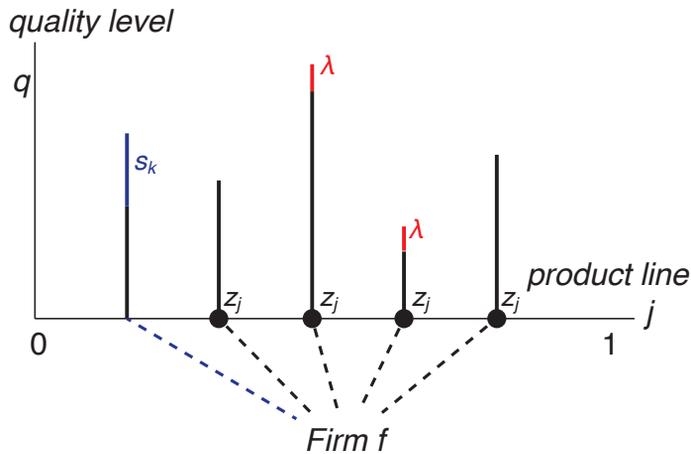
One model of corporate choices towards innovation and the underlying heterogeneity in these processes is depicted by Akcigit and Kerr (2015), who build upon prior micro-macro work such as Klette and Kortum (2004) and Lentz and Mortensen (2008). The model considers why types of corporate venturing and innovation vary along the firm size distribution; this provides an important input into understanding the relationships between innovation and business growth for firms. The Akcigit-Kerr model draws a distinction between two types of innovation that companies may perform: internal innovation and external innovation. Internal innovation, sometimes called ‘exploitation’ innovation by organisational behaviour scholars, concentrates on improvements to a company’s existing product lines, enhancing the capabilities and offerings that the company already has in order to increase profits. External innovation, sometimes referred to as ‘exploration’ innovation, focuses on creating new ideas to add to the

company's product range. In the Akcigit-Kerr model, this exploration process can be understood as improving upon and taking ownership of product lines belonging to other companies.

Figure 3.1 depicts an illustrative firm in this model, which we label f , and how this firm engages in these two types of innovation. Each product line is represented by one of the vertical lines on the chart, and every possible product line in the economy falls somewhere on the 0-to-1 continuum. There is an infinite number of possible product lines, and each firm owns some finite subset of these lines. The 'quality' of the underlying technology for each product line is represented by the height of the line in this figure, following classic 'quality ladder' depictions for studying technological progress (e.g. Aghion and Howitt, 1992). A higher quality product line brings in more profit than a lower quality one.

To the right of the figure are the four product lines originally owned by the firm (the lines shown in black and with z_j at their base). The firm can engage in internal R&D and innovation to improve the quality of the lines they hold. Firms have profit incentives to improve these technologies, but they also face costs for conducting R&D. The rate at which they make these investments is z_j , with some probability of success in each period thus determined (innovation outcomes are stochastic). A firm will invest money up until the costs outweigh the expected benefits. If an innovation is realised, the quality of the product line is incremented by an amount, λ , and the company gains more profits from the improved line. In the figure, two of firm f 's internal R&D attempts have been successful as an example.

Figure 3.1 Internal and external innovation



Source: Akcigit and Kerr (2015).

Companies may also undertake external R&D to 'capture' a product line owned by another company. As with internal R&D, exploration R&D efforts succeed with a probability that depends upon the amount of the investment being made. The firm will spend an amount of money, x , on exploration R&D up to the point that the costs equal the expected gains. If a company's attempt is successful, then it acquires a new product line – chosen at random along the 0-to-1 interval – at

the quality level that said product line has reached prior to this acquisition. The firm then increments the quality of this product line by an amount, s_k . This is shown in the figure by the line on the left with x at the base. The black section represents the quality of the product line at the time that firm f 'captured' it, and its quality is then incremented by s_k as a result of the innovation undertaken by firm f . The magnitude of s_k is determined in the model by a number of factors, including technology waves, how long a product line has existed, and various other inputs that are beyond the scope of this chapter. This form of innovation is also sometimes called 'horizontal innovation' and is closely related to the frequently discussed concept of 'creative destruction'.

Firms are constantly pursuing both forms of innovation, and are thus competing with each other on two fronts: trying to improve the quality of the product lines that they already own, and trying to capture product lines away from other companies. The model also takes into account entrepreneurs or new entrants by modelling individuals who own no product lines but wish to enter the industry by engaging in this creative destruction. The ability to consider both internal and external innovations and to jointly model them in a fully specified, general-equilibrium setting is one of the major theoretical contributions of this model. This is an important step, as it begins to allow economic models to take better account of why differences in the number of small firms versus big firms might matter for the types of innovations undertaken and the economic impact observed.

The key feature of this model is the manner in which the different types of innovation scale up as firm size increases. In particular, the model predicts that *internal* innovation scales with firm size much more than external innovation does. As firms grow larger, the proportion of their R&D budget that they allocate to internal R&D will scale in a linear fashion as more product lines are added. However, external R&D does not scale up with company size as completely. This observation has been made at times in the empirical literature regarding innovation, and is being applied here to theory; the full version of the model also undertakes a more complicated quantification analysis to formally measure these properties.

As an example, consider the extremes of firm size. A new entrant or entrepreneur starts with zero existing product lines, so they cannot, by definition, engage in internal R&D. Their entire budget will therefore be allocated to external innovation. Similarly, a very small firm with only one or two product lines still has a very limited opportunity to spend money on internal innovation, but there is considerable opportunity for external R&D. At the other end of the scale, a firm with 1,000 product lines has a much greater opportunity to spend money on internal efforts, and we see the proportions shift in that direction. It is important to note that this does not reflect the absolute amounts of money spent – a large firm may spend more in aggregate on external R&D than a small firm does, but these exploratory expenditures will account for a smaller proportion of their budget than at a small firm.

Data collected from the US Census Bureau and the NBER Patent Database on firm R&D and patenting behaviour exhibit the scaling that the model predicts. For example, using the 2008 Business R&D and Innovation Survey, there is a -0.16 correlation between firm size and the share of R&D that the firm reports is directed towards business areas and products where it does not have existing revenues. Similar negative correlations are found for questions about the share

of firm R&D being directed to technologies new to markets. Similarly, using the citations that firms make on the patents they file, there is a 0.11 correlation between firm size and the share of backward citations that are made to a firm's own prior work. Firms with larger past patent portfolios are mechanically more likely to self cite, and Akcigit and Kerr (2015) shows that larger firms are more likely to exhibit abnormal rates of self citations compared to Monte Carlo simulations of their expected self-citation rate. Other evidence is also provided in the paper.

By itself, these differences in innovation behaviour across the firm size distribution might not result in important economic outcomes, but the study by Akcigit and Kerr goes further and shows how external innovation is associated with greater employment growth than internal innovation. That is, the average firm growth impact that comes from exploratory work is larger than when firms focus on just enhancing their existing product lines, and also the growth spillovers into the broader economy are larger. The data thus indicate that firm growth rates depend on the kinds of innovation undertaken, and that firms that engage in relatively more internal innovation have slower growth rates than firms that spend proportionately more on external efforts. Thus, we often see larger firms growing at a slower rate than smaller firms or new entrants, and we also find that these smaller, newer firms make disproportionately large contributions to major innovations. This again connects back to allowing for a model that can link firms of different sizes to different types of innovation investments, and ultimately to growth consequences for the firm and the economy as a whole. This is where the academic literature is currently pushing and is starting to make substantial traction.

With this model in mind, the sensible next step is to examine the choices that firms make to see why they engage in the types of innovation that they do. The fact that larger firms devote less resources to external work can have both 'efficient' and 'inefficient' underlying reasons. The model can operate the same in both cases, but the business and policy prescriptions would be different. Why might larger firms engage efficiently in less external R&D? Akcigit and Kerr (2015) describes several reasons, with the most intuitive one being limits on the effective use of manager time. If a skilled CEO does not have the time or resources to add another product line to their workload, it would be a reasonable decision to focus on the existing lines rather than trying to add new ones. In this setting, because new entrants and small firms have fewer product lines, they have competitive advantages in pursuing external-oriented work.

On the other hand, many management scholars have noted inefficient reasons why larger companies conduct less external R&D or are generally less successful at achieving external innovations. Among the issues discussed in the paper are overly bureaucratic organisations and short-term stock market pressures. In each of these settings, the CEO of the large company may in fact want to obtain more product lines and the associated growth, but struggles to do so. This is the scenario in which IBM found itself in the late 1990s, and we will use this case study to describe the setting further. After a successful turnaround following a near bankruptcy at the beginning of the decade, IBM's new CEO was horrified to find that the innovation initiatives that he had set up at the company were failing because IBM's culture and organisation was not conducive to that sort of exploratory R&D. We will use this case study to describe some reasons why large companies can struggle with external innovation and also to identify how

one firm sought to change itself to allow for better innovation outcomes and dynamic growth.

3.3 The IBM Emerging Business Opportunity story¹

Founded in 1911, IBM focused for most of the second half of the twentieth century on creating and selling computer mainframes and minicomputers. In the 1960s and 1970s, it controlled a 70% share of the mainframe industry market, and by the 1980s it was the most profitable company in the world. However, by the end of the decade the company had begun to decline, and by 1991 it was losing money. Between 1991 and 1993, IBM lost approximately \$16 billion, and its market share dropped from 76% to 26%. This happened for several reasons. When smaller, startup companies began to make personal computers more easily available to individual consumers, IBM's leadership believed – based on past success – that they could enter and easily control the PC market. They did not recognise early enough that they needed to continue to innovate their PC platform and its marketing, and they did not realise until too late that the PC would bring about a seismic shift. While IBM's senior executives recognised that they were heading into trouble, they were unable to fix the problem.

Hoping to stop their slide and turn the company around, IBM brought in Lou Gerstner in 1993 to be the new CEO. Gerstner had led American Express and had been CEO of RJR Nabisco, and he was the first outside CEO to lead IBM. When he arrived, he found that IBM had fallen victim to what Donald Sull (1999) termed “active inertia”, in which a company's set of assumptions about its core business become blinders to new ways of thinking that will promote growth.

IBM had spent the past several years focusing on existing products and short-term goals, with little attention paid to customers and their changing needs. In addition to ‘mainframe blindness’, Gerstner found that the processes for managing each of the individual 39 business units had continued to follow unproductive routines that rewarded existing product offerings and short-term results. Processes for starting new ventures were unclear and without a supportive infrastructure. The business units had their own profit and loss statements, but sales, manufacturing and distribution were spread across the company. This organisational structure, and the fact that there was no formal process for acquiring funding or strategy for development, made starting a new venture within the company haphazard at best. Missed opportunities were many. Finally, the culture at IBM fostered relationships that had, in some instances, become shackles that were maintained by a powerful bureaucracy, inflexible hierarchy and interdivisional rivalries.

Gerstner's first move was to stop the steady losses of money and customers, before he could start thinking about how to restart company growth. He decided to keep most of the senior management on board, and strove to create a sense of urgency by requiring them to write memos describing steps to fix their departments and the company as a whole. He also took steps to cut \$7 billion in costs by shutting down underperforming departments and units, and established the “One IBM” philosophy, setting a vision for the company as a

¹ This section and the next draw extensively from Applegate and Kerr (2015). The IBM story is described in detail in Garvin and Levesque (2004) and Applegate et al. (2008), from which this section also pulls material.

global information business, not just a computer company, and allowing them to strategise around new opportunities such as the internet. The company focused its new IBM Global Services business on value partnerships with clients and on eBusiness consulting.

IBM's setbacks to innovation

Gerstner's changes brought an almost immediate improvement, and by 1999 IBM was on a stable financial footing and looking to position itself to be able grow and make its way back to the top of the industry. A large part of the plan to do so involved being able to identify promising new ideas and directions in which to take the company, but IBM was having trouble in this area. IBM researchers were coming up with plenty of promising new ideas, but Gerstner was horrified to learn that, rather than giving them the opportunity to grow, some managers seemed to be obstructing progress or allowing new initiatives to fail. After learning on a Sunday morning that funding had been cut for a promising life sciences initiative due to short-term pressures, Gerstner demanded that action be taken.

Gerstner turned to Bruce Harreld, IBM's Vice President of Corporate Strategy, to investigate why things were going wrong. Harreld and his team discovered that this life science example was part of a very consistent pattern across the company, and that IBM's organisational structure was still fairly hostile to corporate venturing and the creation of new businesses, despite intentions otherwise. IBM's business units were having difficulty integrating new products and ideas that came out of R&D efforts, and managers frequently reduced budgets of growth initiatives or, having failed to commercialise the results of research, even cut the programmes altogether.

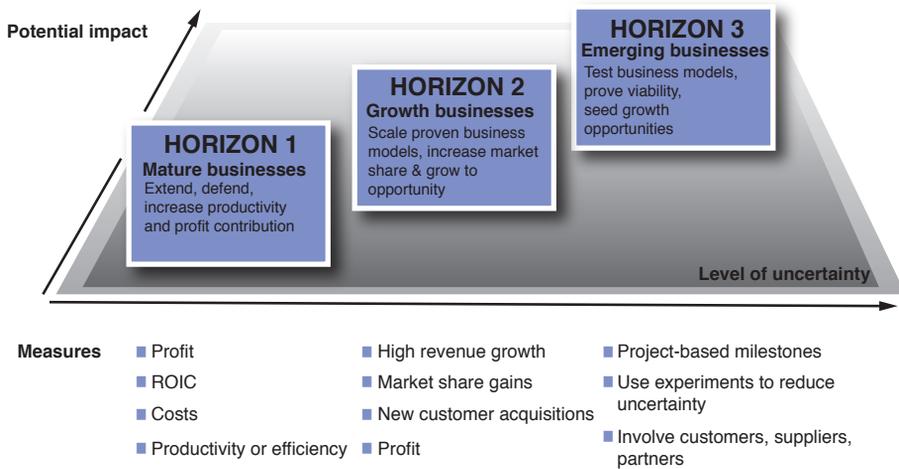
After interviewing individuals within the company who had been involved in several dozen missed opportunities and failed or struggling new venture startups, and documenting their findings in detailed case studies, Harrell and his team identified a number of high-level problems that were leading to the failure of new ventures. First, the company was mainly focused on serving the needs of existing customers, and managers were usually under considerable short-term pressures that restricted the amount of time that they could dedicate to exploring and supporting new ideas. IBM also had no useful approaches to learning about new ideas or identifying strategic needs, and no processes in place for selecting projects or funding them. The company used a complex 'matrix' organisational structure that was focused on existing brands and on geographies and industries for sales and marketing, and new ventures that did not fit well into the rigid matrix were frequently abandoned. IBM also tended to rely on profit-oriented metrics to evaluate projects and business units, which were ill-suited to measuring the progress of early-stage ventures that might not have reached the revenue-generating stage. This meant that R&D efforts were easy targets and often the first to be cut when a unit was having budget issues. And, the new ventures that IBM did undertake tended to be contained in separate 'silos' away from the rest of the company, which meant that it was difficult to effectively integrate new developments into the core business.

The rise of emerging business opportunities

In a very real sense, IBM had become too good at executing, reducing costs and achieving short-term success. While each of these outcomes are desirable, they also placed the company in a position where it struggled to undertake the longer-

term exploratory innovation that would be necessary for the company’s sustained success. In short, IBM found itself in a position similar to that described by the Akcigit-Kerr model, but did not want to be there! Harreld and the rest of the IBM executive team addressed this issue by suggesting the creation of an Emerging Business Opportunity (EBO) initiative. This was based on a framework from the influential book on management from McKinsey Consultants, *Alchemy of Growth* (Baghai et al., 1999), which described a three-horizon model that classified business ventures and innovations according to the length of time until expected impact, return potential and level of uncertainty, as shown in Figure 3.2. The book posits that a company’s sustained growth rests on what the authors call a continuous pipeline of business-building initiatives that is attained by balancing short-term pressures for results with creating the space to conduct long-term and external innovation.

Figure 3.2 Horizons model to classify innovations



Source: Adapted from Baghai et al. (1999).

In the conceptual model in *Alchemy of Growth*, horizons are managed concurrently within an organisation, and each horizon requires its own separate management strategy. Horizon 1 (H1), situated at the lower left (low impact, low uncertainty), covers a company’s core business – the one around which a company has formed its identity, is organised, and has profited. H1 innovations extend or incrementally improve this business (e.g. the development of a new type of bumper by a car manufacturer for an existing line of sales). These efforts connect very closely to the lambda internal innovations we described earlier. While necessary to generate cash and provide resources for growth, H1 businesses – where most companies focus the bulk of their attention – will eventually flatten or become disrupted.

Horizon 2 (H2) encompasses emerging, fast-rising businesses that have the capacity to eventually transform the company and become an H1 business – for example, the development and scaling up of a new type of engine that will be the basis of a new model of car. These innovations and business opportunities often have already exhibited some signs that they will work out well, but much investment remains to be done to prove the opportunities and place them into

position for long-term profitability. Located in the middle of the figure, these ventures are medium term and have the potential for medium amounts of growth or transformation, but also come with an associated level of risk.

Horizon 3 (H3) is where the seeds of totally new ideas and business concepts are created in initiatives such as research projects and pilot programmes (for example, experimentation with rechargeable batteries for the purpose of developing an electronic car in the future). H3 initiatives carry with them a high risk of failure, and are often not completely aligned with a company's existing goals or product lines (and in some cases may even cannibalise current operations if ultimately launched), but they also have the highest growth potential. These are the businesses that can potentially transform a company and provide it with a long-term platform for growth. Although not all of the new H3 ventures will mature to become H1 businesses, nourishing them is necessary for a company's long-term future.

Measurements, expectations and leadership needs differ for each of the horizons. If the three horizons are managed concurrently to ensure healthy and continued growth, they 'cascade' through an organisation. IBM was already well set up to handle Horizon 1 ideas and projects, which returned reliable, short-term gains and could be managed within existing business units; there was never a misalignment of incentives between the managers and these profitable investments. By contrast, Horizon 3 businesses were where IBM was struggling, as they usually required extensive experimentation or research and took a long time to realise their potential, which did not fit well with IBM's current short-term focus. These were the ideas that IBM was most interested in cultivating, and they were the ideas that the EBO initiative targeted. This is very common for larger companies (and very fast growing smaller companies), where the core of the company's operations can limit the ability for other ventures to take root around them. It connects to and reflects the limited scaling built into the Akcigit-Kerr framework.

IBM quickly moved from investigation to action. To begin, Corporate Strategy and managers of individual business units worked together to identify Horizon 3 businesses. They decided upon seven EBOs that met their inclusion criteria, which included: the need for cross-business cooperation and resources; the maturity of the business plan and strategy (e.g. key market and technology risks appeared manageable and expertise was available to build the first offering and take it to market); the forecasted size of the market; and the potential for generating over \$1 billion in three to five years.

Gerstner selected John Thompson, a 34-year veteran of the company, to oversee and coordinate the EBO initiative. He was highly respected at IBM, which gave the programme instant credibility. Thompson and Gerstner began rigorous monthly reviews of each of the seven initial EBOs, focusing on project milestones and developing business plans, rather than on meeting strict financial goals. Corporate Strategy also worked with Finance to identify expenses and revenue for each EBO, and Harreld set a goal of two percentage point annual incremental revenue growth from EBOs. Gerstner also began using the "horizons of growth" terminology in his speeches to the company. This helped to send the message that EBOs were not just a fad, but were something that IBM was taking seriously. By 2002, 18 EBOs had been identified and shepherded through the programme.

One of the first challenges that the EBO group encountered was the question of where to place EBOs organisationally. If innovation was to be the foundation

for success in IBM's future, it could not be delegated to a 'corporate incubator' that received separate funding and was left on its own to build businesses that would later be thrown 'over the wall' to IBM's business unit leaders, who were relentlessly focused on meeting the projections promised to Wall Street. Nor did the IBM team believe that accountability could be delegated to IBM's research labs. While the contributions of the labs were a significant component of the company's innovation culture and brand, executives wanted to ensure that EBOs were integrated into IBM operating businesses that interacted with the marketplace on a daily basis.

After much debate, IBM determined that both the business units and Corporate Strategy should share accountability for EBO efforts. Placement of the EBO teams within the business units facilitated the effective transition to high growth. Simultaneous oversight by Corporate Strategy, however, insured that the EBO initiative would secure significant senior management attention. Corporate Strategy also facilitated initial startup funding and, with business unit leaders, approved additional funding on an ongoing basis.

A second challenge involved managing risk. Horizon 3 ventures, like all new business ideas, came with an inherent uncertainty and a high chance of failure. To help mitigate the risk, IBM began by first thoroughly monitoring customers' use of technology. By understanding how clients were using (or struggling with) current technology, IBM could better predict what future breakthroughs were needed, and hence determine where best to place its research bets. To ensure customer involvement, IBM also introduced its First of a Kind programme that required that IBM researchers identify a customer willing to partner on research projects and provide minimal financing of the project. IBM also managed the uncertainty risk inherent in breakthrough research by borrowing an approach used by oil companies when prospecting for oil (i.e. 'test wells') and staging financial and other resource commitments based on specific timelines and goals for each project (Kerr et al., 2014). Finally, in 2004, IBM launched a venture capital group to help monitor breakthrough innovations outside of the firm and serve as a technology transfer unit facilitating the commercialisation of discoveries and technologies developed in the company's research labs and businesses.

By mid-2002, most of the initial EBO efforts had made considerable progress, and revenues were up sharply. Equally important, there was considerable enthusiasm for the programme. However, processes remained informal, and success depended on Thompson and Harreld's personal interventions and networks. Financial and tracking systems, reporting relationships, review meetings, leader-selection criteria, and incentive mechanisms remained loosely defined. While this had worked for a small number of EBO projects, the informality and intensive hands-on management could not be effectively scaled. EBO leaders differed on when to move out from under the EBO umbrella and into an H2 business. Some were concerned about how they would weather the transition from qualitative measurements such as milestones, while others argued that the tough financial goals expected of an H2 were healthy and necessary for the EBO system to be taken seriously within company.

By now, Sam Palisamo, another IBM veteran, was CEO, and he challenged Harreld and his team to come up with a way to scale and systematise the EBO programme so that it did not require constant hands-on help. Harreld and the Corporate Strategy group assumed formal responsibility for the EBO process. They recognised that different categories of innovation had different

risk profiles and, as a result, different approaches were needed to manage risk during implementation. The type of risk and the approach to implementation were based on: (1) the size of the opportunity; and (2) the timing and level of resources required to exploit the opportunity. These parameters defined different categories of innovation with different risk profiles. Each category of innovation also required a different leadership and organisation model.

By the end of the decade, IBM's adjustments to their EBO scheme had proven their effectiveness. By 2011, the company was making \$19 billion in revenues – 20% of their top line – from businesses that started as EBOs, and IBM seemed to have successfully positioned themselves for further growth and innovation.

Lessons from IBM and corporate innovation for growth

IBM's story provides a useful example of the importance of establishing an innovative structure within large, existing companies. Companies need to have the capabilities to engender dynamic growth. A number of lessons and best practices can be drawn from the successful efforts of others to jump start innovation and new business pursuit (Applegate and Kerr, 2015):

- Innovation is necessary for a company to continue to grow and survive. Eventually, even the most productive core businesses will run out of room to grow and will face loss of market share. Disruptive innovations from other players in the same industry can create even greater pressures to find new ways to grow. The empirical work in Akcigit and Kerr (2015) confirms this point.
- As companies grow larger, it can be harder to innovate. Established patterns and processes at large companies can hinder the ability of those companies to generate new products or businesses, even when it becomes clear that such a change is necessary. Donald Sull (1999) termed this phenomenon “active inertia”. Managers at all levels should be aware of common obstacles that can stifle innovation or new corporate ventures:
 - managers are frequently subject to short-term pressures, leaving them with little time/resources to devote to new ventures;
 - corporate objectives are often misaligned with the goals of the innovative process, and profit-oriented metrics that are a poor fit for early-stage innovations make these efforts easy targets for cuts during budget crises;
 - established structures, bureaucracy and internal politics such as interdivisional rivalries can make it difficult for changes to take root; and
 - companies can be restricted by the expectations of their customers and stockholders, both of whom are less likely to take a long-term view.
- Innovations can be broadly classified, and it is necessary for companies to be able to engage all three horizons simultaneously (what Mike Tushman calls “organizational ambidexterity”). If done correctly, innovations will continually cascade through the company, moving from uncertain H3 ventures to generating H1 ideas for mature ventures.
 - Separate horizons have different needs in terms of management, organisation, evaluation, and so on. Innovation or R&D units

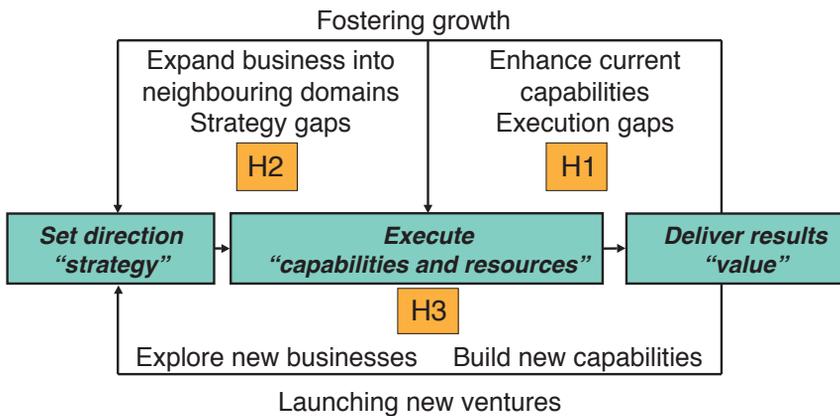
have their own unique requirements in terms of, for example, management and metrics, different from those of established businesses/product lines. Profit levels or revenue growth may be more appropriate for Horizons 1 or 2, while Horizon 3 ventures may be better evaluated using project milestones and less-rigid metrics.

- Breakdowns in the execution of a company's strategy can provide clues to where the company needs to focus on innovation. (As a starting point, since long-term companies should be engaged in all the horizons at once to build truly dynamic organizations.) Figure 3.3 illustrates this framework, where 'gaps' can occur, and how innovations in different horizons can address these gaps.
 - If a company's or unit's strategy is determined to be correct overall, but there is a breakdown between execution and delivering value (an 'execution gap'), then typically this is an opportunity for an Horizon 1 innovation, which mainly enhances current offerings and improves execution.
 - If the breakdown seems to occur between the setting-strategy and execution phases (a 'strategy gap'), then this is more likely addressable by Horizon 2 innovations, which are longer-term and more uncertain than Horizon 1, but still adhere to the same overall strategy. The goal here is building new capabilities to deliver against the strategy.
 - Horizon 3 ventures, by contrast, do not (necessarily) address strategy or execution gaps, but are attempts to expand into new businesses within a corporation or create new capabilities, possibly even creating entirely new strategic elements.
- One of the first and most important decisions encountered by IBM and other companies seeking to innovate is where to locate the new initiatives within the company. There is no 'one size fits all' solution. If there is a risk of cannibalisation of time or resources by core businesses, it may be beneficial to keep a new venture separate from the rest of the company – but this risks a situation in which an innovation is not well-aligned with the company's goals and is difficult to integrate and move to a H2 business. On the other hand, while integrating new ventures into existing business units from the very start can afford them better access to funding and resources, it can also position them under managers who don't have the time or know-how to properly nurture them.
- Senior management must create a sense of 'urgency' around the changes and new initiatives, and it is responsible for ensuring that there is buy-in at all levels of the company and that innovation and new ventures are taken very seriously. It is also important to staff new ventures with some of the firm's best talent – although it is tempting to reserve the most capable workers for existing businesses with guaranteed returns, innovation efforts cannot succeed without skilled and dedicated workers.
- Although hands-on involvement from the CEO and senior management can be helpful in the early days of innovation initiatives, this may not be sustainable for the largest companies like IBM, and it is necessary

at some point to formalise the process of shepherding early-stage ideas through the stages up to Horizon 1. This includes financial and tracking systems, leader selection, processes for meetings and reviews, and incentive mechanisms. For other organisations, the CEO may retain more direct control over the moving parts.

- Companies should involve outside parties in the ideation and innovative process to minimise risk. In particular, firms can use customers and other outsiders as ‘early discovery systems’ by monitoring customers’ use of existing products to provide clues to their needs and generate likely ideas for new ventures.
- Perhaps most importantly, failure must be an option. Just as venture capital firms rely on their ability to terminate investments in projects that are not working out, large firms like IBM must be able to halt work on ideas that are not panning out and reallocate their resources elsewhere. This can be difficult for large firms – the relative availability of funding may lead to allowing struggling ventures to flounder for much too long and managers are likely incentivised to avoid or mask failures. Proper continuation choices are essential, and some of the best companies use outsiders to obtain objective opinions about which projects to push forward or to terminate.

Figure 3.3 Strategy execution framework for business growth



Source: Taken from Applegate and Kerr (2015).

In summary, the development of new businesses and innovations helps drive the growth of firms of the economy as a whole. Recent academic work is pushing the boundaries to understand better how firms differ in this regard, and we have collected empirical and case evidence of the challenges that large companies face in maintaining the pursuit of exploratory powers. In some cases, the shift towards an internal focus is warranted; in other cases, such as that depicted by the IBM story, it is inefficient and may ironically be an outcome of attributes that makes the organisation otherwise successful. In managerial research, we are discerning a set of best practices for how to keep organisations more dynamic. These ideas need to be customised to each company and situation, and not all apply to every firm, but corporations should be learning from others as they discern how to

best foster new business opportunities in their companies to provide growth for tomorrow.

References

- Aghion, P and P. Howitt (1992), "A Model of Growth Through Creative Destruction", *Econometrica* 60(2), pp. 323-351.
- Akcigit, U. and W. Kerr (2015), "Growth through Heterogeneous Innovations", Harvard Business School Working Paper, Cambridge, MA.
- Applegate, L. and W. Kerr (2015), "Leading Breakthrough Innovation in Established Companies", Working Paper.
- Applegate, L., R. Austin, and E. Collins (2008), "IBM's Decade of Transformation: Turnaround to Growth", Harvard Business School Case No. 9-805-130, Harvard Business School Publishing, Boston, MA.
- Baghai, M., S. Coley and D. White (1999), *The Alchemy of Growth*, New York: Basic Books.
- Garvin, D. and L. Levesque (2004), "Emerging Business Opportunities at IBM (A)", Harvard Business School Case No. 9-304-075, Harvard Business School Publishing, Boston, MA.
- Kerr, W., R. Nanda and M. Rhodes-Kropf (2014), "Entrepreneurship as Experimentation", *Journal of Economic Perspectives* 28(3), pp. 25-48.
- Klette, T.J. and S. Kortum (2004), "Innovating Firms and Aggregate Innovation", *Journal of Political Economy* 112(5), pp. 986-1018.
- Lentz, R. and D. Mortensen (2008), "An Empirical Model of Growth Through Product Innovation", *Econometrica* 76(6), pp. 1317-1373.
- Sull, D. (1999), "Why Good Companies Go Bad", *Harvard Business Review*, July-August, Reprint #99410.

4 Regional Variation in Venture Capital: Causes and Consequences

Ramana Nanda and Matthew Rhodes-Kropf¹
Harvard Business School

4.1 Introduction

Entrepreneurship is a central element of the Schumpeterian process of creative destruction (Schumpeter, 1942). Startups have been associated with the birth of important new industries such as semiconductors and computers, the internet and biotechnology, and there is increasing evidence of the important role that startup firms play in driving aggregate productivity growth in the economy (Aghion and Howitt, 1992; King and Levine, 1993; Foster et al., 2008).

The availability of finance, and in particular venture capital (VC), seems to be an important part of this phenomenon, despite its extremely small size. Kerr et al. (2014b) highlight that there are less than 500 active VC firms investing in startup ventures across the United States in a given year, and Ewens and Rhodes-Kropf (2015) find that approximately 2,000 individuals accounted for 84% of all dollars invested in venture capital in the United States between 1987 and 2012. In fact, only about 1,000 of the 600,000 new firms that are founded each year receive initial venture capital financing, but VC backed firms constitute over 50% of the initial public offerings (IPOs) on US stock markets (Kaplan and Lerner, 2010) and about 10% of private sector employment (IHS, 2011),² highlighting the disproportionate impact that this industry has on the economy.

Several papers have documented the role that VC plays in the economy. For example, Kortum and Lerner (2000) find that increased VC availability leads to increased levels of patenting. Samila and Sorenson (2011) find that an expanded supply of VC raises employment and aggregate income within different regions in the United States. This work also suggests that in most regions even within the United States, an increase in VC of a dollar would lead to an increase of more than a dollar in local employment. Kerr et al. (2014b) use census data in the United States to compare startups that received VC with those that did not. Looking at firms founded in the period, 1986-1997, they find that by 2007, 75% of the VC-backed firms had shut down, compared to 66% of the non-VC-backed firms. The surviving VC-backed firms had grown to the point where their total employment was equal to 364% of the total employment of the original firms at the time of VC investment (including those that eventually failed). On the other hand, the

1 The ideas contained in this chapter arise from a number of other pieces we have written, including a similar brief written for the NBER's Innovation Policy and Economy Working Group entitled "Financing Entrepreneurial Experimentation", Kerr et al. (2014b) and Nanda (2015). All errors are our own.

2 Based on data provided by the National Venture Capital Association.

larger number of non-VC-backed firms still employed only 67% of the original sample. Puri and Zarutskie (2012) also find that venture-backed firms grow larger and employ more people. Chemmanur et al. (2011) report that venture backing improves the efficiency of firms. Several other papers have documented the role that VC plays in driving innovation through venture capitalists' roles in monitoring and governing startup ventures (Hellmann and Puri, 2000, 2002; Sorensen, 2007; Chemmanur et al., 2011; Puri and Zarutskie, 2012; Bernstein et al., 2014). This suggests that the availability of VC may be a central factor that determines the degree to which radical new ideas are commercialised in a given region or at a given point in time.

A notable feature of venture capital is the uneven nature of VC investment across regions and time. For example, VC investment per capita is a lot larger in the United States than in Europe, and within the United States, Silicon Valley, New York and Boston account for the lion's share of VC investment. In addition, VC investment has been documented to occur in cycles, where certain industries receive a disproportionate share of investment relative to others across time (Gompers and Lerner, 2004; Kaplan and Schoar, 2005; Gompers et al., 2008).

This chapter develops a framework for understanding the uneven distribution of venture capital across industries, regions and time periods. We highlight how the extreme uncertainty facing startup ventures at their earliest stages leads venture capitalists to engage in a process of experimentation across multiple rounds of funding, abandoning investments where intermediate information is negative and investing more in startups where intermediate information is positive. While these real options are a central element of the investment process, we also point out that financiers, rather than markets, dictate investment and continuation decisions as they choose which experiments to attempt, how to interpret the results, and whether to continue with or abandon the investment. These financiers' actions are impacted by a myriad of incentive, agency and coordination problems that shape their ability to effectively experiment. We document two important costs to experimentation: constraints to exercising abandonment options when intermediate information is poor, and shocks to the supply of capital that impact the ability to raise capital even when intermediate information is positive. We show how these can vary across regions and time, thereby not only impacting the distribution of venture capital across regions but also, in doing so, impacting the rate and trajectory of startup innovation.

The financial benefit of running an experiment stems from an ability to abandon the investment if intermediate information is poor, or to replace the founder with a new CEO (e.g. Hellmann and Puri, 2002; Kaplan et al., 2009; Ewens and Marx, 2014). The first constraint we consider is that it is often difficult or costly to shut down a firm. One cost from quickly shutting down a firm, for example, is the disutility felt by the entrepreneurs who suddenly lose their jobs. This and other costs create a trade-off between the rapid abandonment of projects, which encourages investors, and tolerance of failure, which encourages entrepreneurs (Manso, 2011). In Nanda and Rhodes-Kropf (2015), we note that this trade-off is even more troublesome when it cannot be solved optimally for each project, and is instead set by a law, culture or level of bureaucracy that will apply to all projects. Countries with laws designed to make it difficult to fire employees and shut down firms may encourage innovation, but financiers in these countries will be unwilling to back very experimental projects. This may help explain the remarkable dearth of VC backing of innovation in some

European countries (Nanda and Rhodes-Kropf, 2015). More generally, this work develops how formal and informal institutions in an economy play an important role in the level of innovation through their role in promoting the amount of experimentation that investors undertake.

The next constraint on the use of abandonment options is that those experiments that turn out well will need to be funded in a future, unknown capital market. The financing available for startups engaged in innovation is notoriously volatile (Gompers and Lerner, 2004; Kaplan and Schoar, 2005; Gompers et al., 2008), leading entrepreneurs and VC investors to worry about the availability of capital even if initial experiments go well. In Nanda and Rhodes-Kropf (2014), we model investors' responses to this financing risk. We show that times or places with high financing risk (low capital availability) are also the times/places that high expected value – but safe – projects will be run. This fits the intuition that good, solid firms are funded when capital is not freely available. The results also suggest, however, that investors are more willing to experiment in boom times or in places with a great deal of capital.

In Nanda and Rhodes-Kropf (2013), we examine early stage investments between 1984 and 2004 and follow them to 2010 to allow time for exits. We find that increased VC availability leads to increased rates of failure among venture-backed firms, but also that those that succeed are more successful and more innovative.³

This suggests that increases in capital caused investors to back not just riskier firms, but more innovative firms. Money not only chased deals (Gompers and Lerner, 2000), but also changed the deals that were funded to more innovative projects.

The results from Nanda and Rhodes-Kropf (2013, 2014) suggest that the most innovative startups may need hot financial markets to facilitate their initial diffusion. Investors cannot fund experiments in areas with low capital availability because there is no future funding. This creates a 'chicken and egg' problem in that available capital in an area cannot be deployed if there is not enough other capital in the same area. Therefore, policies may have larger effects if they are able to encourage a concentration of investors that breaks the bad equilibrium. Alternatively, policies that help the local successful experiments reach the resources they need may allow much more local experimentation.

Overall, our framework can be used to help guide policy by helping to provide an understanding of where the costs of experimentation can be reduced.

4.2 The importance of experimentation

High-impact entrepreneurship requires, almost by definition, going against the grain. Rajan (2012) argues that an entrepreneur "must be willing to strike out, largely on the basis of intuition, on courses of action in direct opposition to the established settled patterns". A consequence of this environment is extreme

³ In the paper, we show that firms financed at more active times have higher valuations when they go public, controlling for the level of the stock market and the year they go public. Thus, the finding compares firms funded in hot times to those funded in cold times that go public at the same time. Furthermore, those funded in hot markets filed for more patents and their patents were more highly cited.

uncertainty over whether a particular technology, product or business model will be successful.

In this context of extreme uncertainty, experimentation plays a powerful role in increasing the chances that the most promising ideas succeed. One form of experimentation entails a variety of different entrepreneurs commercialising what they believe to be the superior product or technology, and where the ensuing competition leads to the 'survival of the fittest'. As Stern (2006) argued, "a favorable environment for entrepreneurship and a high level of economic experimentation go hand in hand". For example, Klepper (1996) has documented a consistent pattern in which a multitude of new startups emerge at the birth of an industry, followed by a shakeout once the dominant technology has been found. Indeed, Rosenberg (1994) has argued that one of the defining features of capitalism is the freedom it provides entrepreneurs to pursue novel approaches to value creation in the pursuit of economic gain. The promise of large rewards drives entrepreneurs to experiment with new ideas, helping to create a dynamic and growing economy. An institutional environment that facilitates this form of experimentation is thus central to maintaining a vibrant entrepreneurial ecosystem. This not only requires an environment in which it is easy to start new ventures, but also one in which it is easy to shut ventures down (given the high failure rates of startup ventures).

This first form of experimentation depicts experimentation at the level of the economy. A second form of experimentation is one in which investors learn about the potential of individual startups over time, by investing in stages instead of providing the full amount upfront. The ability to invest in stages, with the possibility to abandon the investment along the way,⁴ is particularly valuable for high-potential ventures where it is extremely hard, even for professional investors, to know the true potential of a startup without providing money and to learn about the startup's viability over time. A good example of the difficulty in determining how well a new venture will do comes from Kerr et al. (2014b), who study internal data from a single large and successful US VC firm. They look at ratings the partners at this firm gave each deal at the time of the first investment and study how this score relates to the ultimate outcome of the same startups. They find that the correlation between these initial scores and the ultimate performance of the startups was 10%, showing how even successful professional investors have a hard time distinguishing among the most promising startups at the earliest stages of investment. Using similar data from an angel investment group, Kerr et al. (2014a) find the correlation between the interest levels assigned to funded deals and their ultimate success to be less than 10%. More generally, the fact that the majority of VC investments fail – nearly 60% of this VC firm's investments returned less than the money invested – is itself indicative of the difficulty in predicting which firms will be successful and which will fail.

VC firms therefore invest in stages, and learn about the viability of startups through a sequence of investments over time. Since each stage of financing is typically tied to achieving milestones that create information about the future prospects of the venture, each round of funding can be seen as an experiment that generates information about the venture's probability of success and its value conditional on that success. Experiments that generate positive information therefore increase the value of the company and allow the entrepreneur to seek

⁴ See, for example, Gompers (1995), Cornelli and Yosha (2003), Bergemann and Hege (2005), Fluck et al. (2007) and Bergemann et al. (2008).

the next round of funding without giving up as much equity. On the other hand, experiments that generate negative intermediate information allow the investor to abandon the investment without have committed the full amount upfront. Therefore, this process of experimentation – whereby investors learn about the viability of a radical new idea through an initial investment, interpret intermediate results, and decide whether to continue with or abandon their investment – is a key aspect of entrepreneurial finance. It is this second aspect of experimentation that is the focus of this chapter, although we also highlight important interactions and policy implications that stem from the first notion of experimentation.

4.3 A simple model of multi-stage financing

In this section, we set up a simple model of multi-stage financing that we use to demonstrate the key benefits and costs associated with experimentation across rounds of funding. We use this to highlight how costs from such experimentation that can arise in certain regions or points in time can have important implications for the degree to which investors are willing to finance startups commercialising the most radical innovations. In doing so, we hope to demonstrate that costs of and constraints to experimentation can play a first-order role in impacting the supply of VC, and hence play a central role in driving the rate and trajectory of innovation – independent of the availability of novel ideas and talent to commercialise such ventures.

Consider the following investment. A startup requires $\$X$ to commercialise its technology that may or may not work. The probability it will be successful and worth V is p , while the probability it will be worth nothing is $(1 - p)$. The expected value of the project is $pV - X$. Thus, this project will not be financed if $X > pV$.

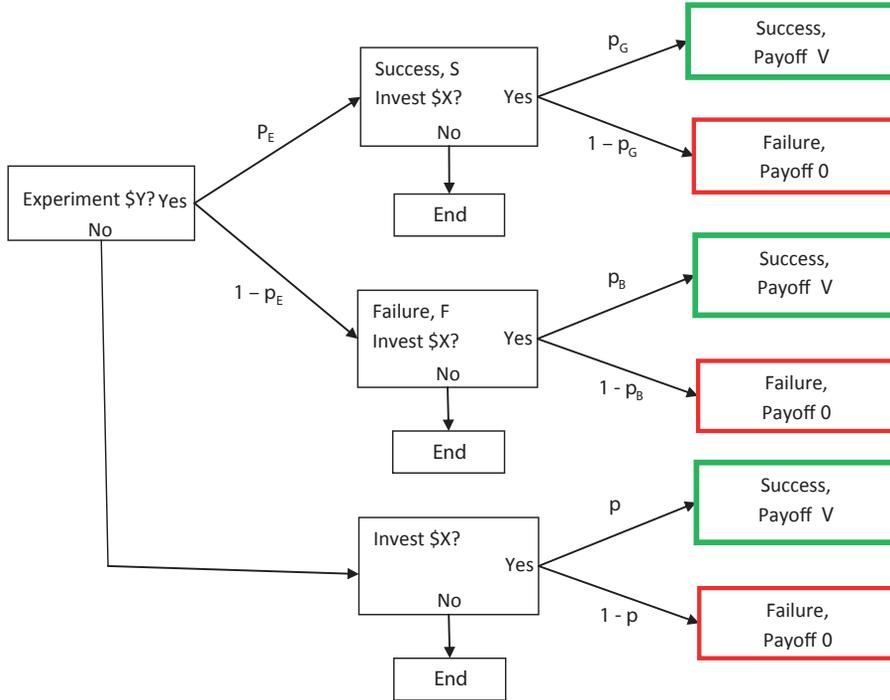
Then imagine that the entrepreneur can conduct an experiment before fully funding the startup. The likelihood that the experiment generates positive intermediate information is p_E , while the likelihood of the intermediate information being negative is $(1 - p_E)$. If the results from the experiment look promising (the “Good” outcome), the chance of ultimate success is p_G , while if the the results from the experiment are not promising (the “Bad” outcome), the chance of success is p_B . The experiment costs $\$Y$ to run. To be equivalent to the project when no experiment is run, $p_G * p_E + p_B * (1 - p_E) = p$, i.e. the unconditional probability of success is the same whether or not the experiment is run. Thus, the experiment reveals information about the quality of the project.

To make this example concrete, consider a project that requires $\$11$ million (X) to be commercialised and that has a 99% probability of being worth $\$0$, and a 1% (p) probability of being worth $\$1$ billion (V). This project will not be pursued as its expected value is negative ($-\$1$ M), i.e. $\$11M > 0.01 * \$1B$. But what if the entrepreneurs could conduct an experiment that will reveal that the project either has a 10% (p_G) chance of working or a 0% chance of working (p_B)? Furthermore, assume this experiment will reveal the more promising news with a 10% probability. Thus, the ex ante probability of success is the same whether or not the entrepreneurs conduct the experiment, i.e. $0.10 * 0.10 + 0 * (1 - 0.1) = .01 = p$. The decision tree of the investor is shown in Figure 4.1.

The question facing the investor is whether it is worthwhile to finance the initial experiment. Intuition might suggest that since running the experiment

increases the amount the investor has to pay from $\$X$ to $\$X + \Y , the experiment is not worth pursuing. However, the value in the experiment arises because it may prevent the investor from spending $\$X$ at all.

Figure 4.1 The investor's decision tree



The experiment can thus be thought of as an investment that pays off $p_G * V - X$ ($\$89M$) with probability p_E , and pays off $\text{Max}[p_B * V - X, 0]$ ($\$0$) with probability $(1 - p_E)$. Note that if the results of the experiment are not promising, the investor will only invest $\$X$ if the project has an expected value greater than zero – the max function accounts for this decision. In our example, $p_E = 10\%$, and therefore the expected value of the experiment is $10\% * \$89 \text{ million} = \8.9 million . Thus, as long as the experiment costs less than $\$8.9 \text{ million}$, it should be run.

Even though the original investment of $\$11 \text{ million}$ (X) was not a good idea, an investment of up to $\$8.9 \text{ million}$, followed by an investment of $\$11 \text{ million}$ if the experiment is successful, is a good idea – it represents positive expected value. Spending an additional $\$8.9 \text{ million}$ to learn about the viability of the project is more valuable than simply spending $\$11 \text{ million}$ directly. This is the benefit of experimentation.

We emphasise that the value of experimentation is not driven by the specific numbers chosen in this example. Rather, the experiment is valuable any time when:

$$p * V - X < p_E * (p_G * V - X) + (1 - p_E) \text{Max}[p_B * V - X, 0] - Y \tag{1}$$

i.e. when the expected value without the experiment is less than the expected value with the experiment. When is this true? This cannot hold, for example, for any project that has a positive expected value even after the experiment fails. In this case, $Max[p_B * V - X, 0] = p_B * V - X$. Since $p_G * p_E + p_B * (1 - p_E) = p$, we see that $p_E * (p_G * V - X) + (1 - p_E)(p_B * V - X) = p * V - X$ and running the experiment is really just a waste of resources. This is because it changes no decision, as the investor invests $\$X$ no matter what the experiment reveals. However, if $p_B * V - X < 0$ then the investor would like to avoid investing when the true probability of success is p_B . The investor would therefore be willing to pay to learn whether the probability is p_G or p_B . How much the investor is willing to pay depends on how much the investor learns from the experiment.

In an extreme case, an experiment might demonstrate nothing, i.e. $V * p_G = V * p_B$. That is, the probability of earning V is the same no matter the experiment's outcome. Alternatively, the experiment might provide a great deal of information. In this case, $V * p_G$ would be much larger than $V * p_B$. We can think, therefore, of $V * p_G - V * p_B$ as the amount or quality of the information revealed by the experiment. $V * p_G - V * p_B$ is larger if the experiment revealed more about what might happen in the future.^{5, 6}

Overall, we see that experimentation is very valuable in situations when an investment of relatively few dollars can reveal information that results either in a valuable project going forward or a mistaken investment being prevented. We next demonstrate two important constraints to experimentation and document how institutional features that govern experimentation can play a role in leading these costs to be systematically different across regions. This naturally sets up potential roles for policy.

Costs of exercising abandonment options

As was seen above, the benefit of running an experiment from the investor's perspective stems from an ability to abandon the investment if intermediate information is poor (or to replace the founder with a new CEO). This form of failure can be frustrating to entrepreneurs, who often tend to feel that a breakthrough requires only a little more funding and patience. Thus, entrepreneurs often look for investors willing to allow them a second go if the intermediate information is negative, or even look for investors who are willing to fund the project more fully up front. In an extreme case, entrepreneurs may not be willing to take an investment from investors who have a reputation for exercising their abandonment options.

To incorporate this idea into our simple model, we will assume that the effort decision by the entrepreneurs is all or nothing, i.e. they either start the new venture or they do not. They also face a cost of u_f if the project is shut down after the experiment. This can be thought of as the disutility they experience when they fail. In this case, even for firms where experimentation may be valuable ($p_B * V - X < 0$), disutility for failure may hamper experimentation.

5 Note that we can think of p_G and p_B as posterior probabilities with a prior of p . Thus, one special case is martingale beliefs with prior expected probability p and updating that follows Bayes' rule. In this case, projects with weaker priors would have more valuable experiments.

6 Note also that the experiment is no more or less important if the project is riskier. A riskier project might be one with a larger V and smaller probabilities of success, p_G and p_B , but the information revealed by the experiment, $V * p_G - V * p_B$, could be the same. Thus, the value of the experiment and the risk of the project are related but are not the same.

The total value of experimentation, including both the financial payoff and the costs borne by the failed entrepreneur, is:

$$p_E^* (p_G^* V - X) - Y - (1 - p_E)^* u_F \quad (2)$$

Including a cost of early failure reduces the value of experimentation by $(1 - p_E)^* u_F$. Note that this will also affect the financier even though they do not directly pay the failure costs. This is because the financier and entrepreneur must negotiate over any surplus generated by the project. The loss from early failure lowers the entrepreneur's expected payoff. If the total expected value of the project does not generate enough to cover the costs borne by both the entrepreneur and investor, then the entrepreneur and investor will not be able to find a deal that will induce them to both participate.

If the costs of early failure are too high, then the entrepreneur will not participate in the project if it is funded via experimentation. For example, one can imagine that an aspiring entrepreneur who could receive a \$100,000 investment but then may be forced to shut down in six months due to a lack of further funding may be less willing to quit his day job than if funded with millions of dollars, even if the quality of the project is the same in either case. This is the intuition of failure tolerance – an investor may have to agree to fund the project significantly in order to induce the entrepreneur to start the project. In Manso (2011), for example, principals decide how to reward agents in an interim period as well as when the final output is revealed. Manso (2011) demonstrates how the optimal payments may involve leniency in the case of bad interim outcomes. This reduces incentives for effort, but simultaneously induces the agent to experiment. Hellmann and Thiele (2011) also suggest that low-powered incentives may induce low effort in standard tasks but may encourage experimentation. This is a very intuitive result, and a number of empirical papers consider the impact on innovation of policies that create a failure tolerance.⁷

Interestingly, however, many innovations are commercialised by new ventures that are backed by VC investors, who tend to be remarkably intolerant of early failure (Hall and Woodward, 2010). It is standard for venture capitalists to negotiate control rights that allow the investors to fire management and/or abandon the project (Gompers and Lerner, 2004; Sahlman, 1990; Hellmann, 1998). Even among venture backed firms that are 'successful', Hellmann and Puri (2002) and Kaplan et al. (2009) show that many end up with CEOs who were not the founders.

In Nanda and Rhodes-Kropf (2015), we explain this apparent contradiction by arguing that the principals who are financing innovation (Venture capitalists, corporations, and even governments) cannot set an optimal failure tolerance policy on a project-by-project basis. Bureaucratic constraints, laws, policies, or possibly a desire to maintain a consistent reputation lead investors to fix an 'innovation policy' upfront.⁸

7 See, for example, Burkart et al. (1997), Myers (2000), Acharya and Subramanian (2009), Ferreira et al. (2011), Aghion et al. (2009) and Tian and Wang (2014).

8 For example, the manifesto of the VC firm the Founders Fund (investors in Facebook) reads "companies can be mismanaged, not just by their founders, but by venture capitalists who kick out or overly control founders in an attempt to impose 'adult supervision.' Venture capitalists boot roughly half of company founders from the CEO position within three years of investment. Founders Fund has never removed a single founder..." (emphasis added); see <http://www.foundersfund.com/the-future>.

They may do so by committing not to shut down projects quickly. Alternatively, a company culture or level of bureaucracy will apply to all projects. Or, for example, a government looking to stimulate innovation may pass laws making it harder to fire employees. These levels of ‘failure tolerance’ will apply to all employees, regardless of the project. Put differently, a principal often has an innovation policy that is set *ex ante* – one that is a blanket policy that covers all projects in the principal’s portfolio.

This pre-set policy, culture or bureaucracy may then affect what projects the principal chooses to pursue. Intuition can again be gained from our simple model by assuming that a failure-tolerant investor commits to funding the project regardless of the outcome of the experiment. Thus, the expected value of the project if run by a committed investor is $p^*V - X$ (because the experiment is not run). With the alternative uncommitted strategy, the expected value of the project is as in equation (2). Thus, a project will be done by an uncommitted investor if:

$$p^*V - X < p_E^*(p_G^*V - X) - Y - (1 - p_E)^*u_F \quad (3)$$

In this case, the value of the project is large enough with an uncommitted investor that enough value can be shared with the entrepreneur to make up for their potential disutility from failure.

When will this be the case? In those companies where the experiment reveals a large amount of information. As we saw above, when the value of the experiment is high, then $p^*V - X < p_E^*(p_G^*V - X) - Y$. Since entrepreneurial disutility lowers the value of the experiment, the information from the experiment has to be even more valuable to be financed. Thus, it is the uncommitted, failure-intolerant investors that will select the most experimental projects. Meanwhile, those organisations that are more tolerant of failure will only be willing to back the less experimental projects, because with safer projects they will not need to extract value by terminating if bad information occurs.

Combining this with the idea of tolerance of failure in Manso (2011), we should expect that large, bureaucratic corporations may encourage innovation, but will be unwilling to back very experimental projects, as it would imply a negative expected value to do so without shutting them down after early bad news. Venture capitalists, on the other hand, will choose to fund radical experiments, but many entrepreneurs may be unwilling to leave safe jobs to pursue these projects since they have a significant chance of early failure. Interestingly, corporate venture capitalists are thought to be more tolerant of failure than regular venture capitalists, and Chemmanur et al. (2012) report that this encourages greater innovation. In Nanda and Rhodes-Kropf (2015), we suggest that this might explain why corporate VC earns lower returns than typical VC.⁹

In the same vein, countries with laws designed to make it difficult to fire employees and shut down firms may encourage innovation, but financiers in these countries will be unwilling to back very experimental projects – again, those that would have a negative expected value if they could not be shut down after

⁹ Corporate venture capitalists do not seem to have had adequate financial performance, but Dushnitsky and Lenox (2006) have shown that corporations benefit in non-pecuniary ways (see the theory by Fulghieri and Sevilir, 2009).

early experiments. This may help explain the remarkable dearth of innovation in some European countries (Saint-Paul, 1997; Bozkaya and Kerr, 2014).

The standard culprit for the lack of entrepreneurship in Europe is that there is thought to be a higher stigma attached to failure (Landier, 2002). We can see the intuition for this from equation (2): if u_F is larger, then the value of experimentation is lower. Thus, there will be a tendency towards more certain, or less experimental, projects. However, although the stigma of failure can explain a reduction in entrepreneurship, it has more challenges explaining the virtual absence of radical new economy companies emerging from many countries. Surely some entrepreneurs are willing to take the risk? In fact, what entrepreneurs complain about in many countries is that they cannot get their idea funded. Even Skype, a huge venture-backed success that was started by European entrepreneurs Niklas Zennström and Janus Friis, received its early funding from US venture capitalists Bessemer Venture Partners and Draper Fisher Jurvetson.

A stigma of failure cannot explain this phenomenon by itself. In an environment with a high stigma of failure, capital will be even cheaper as it fights to attract entrepreneurs (Landier, 2002). But European entrepreneurs complain that they cannot find capital to fund their novel ideas even if they are willing to take the risk and potentially suffer the consequences of failure. In Nanda and Rhodes-Kropf (2015), we build on Landier (2002) to show that the problem is two-sided: venture capitalists look for less experimental projects to help them form a reputation for being tolerant of failure, because most entrepreneurs want a more failure-tolerant backer. But doing so potentially results in an equilibrium with no investor willing to fund radical experiments, even if they have positive expected value and the entrepreneur is willing to take the risk. Martin Varsavsky, one of Europe's leading technology entrepreneurs, noted in an interview with *Fortune* magazine that "Europeans must accept that success in the tech startup world comes through trial and error. European [investors] prefer great plans that don't fail".¹⁰

More generally, this work implies that formal and informal institutions in an economy can play an important role in the level of innovation through their role in promoting the amount of experimentation that investors undertake. First, certain financial intermediaries are, by design, limited in the amount of experimentation they can engage in. Banks, for example, do not share proportionately in the benefits when a startup does extremely well, but do suffer the losses when the startup fails. Banks cannot, therefore, fund an experiment with a high chance of failure, even if it is a positive expected-value experiment. Indeed, Black and Gilson (1998) argue that bank-oriented economies are less likely to encourage startups engaged in innovation. In a similar vein, regulations surrounding the amount of money that can be committed by pension funds to asset classes such as VC can have important implications for the amount of capital available to support the financing of experimentation (Kortum and Lerner, 2000).

Second, policies that are aimed at motivating experimentation by entrepreneurs can limit the degree to which investors are willing to finance this experimentation. For example, lenient bankruptcy laws may encourage entrepreneurs to take on bolder experiments, but at the same time make investors less willing to fund the experimentation, since their return if things go badly is reduced (Guler, 2007a,b;

¹⁰ <http://tech.fortune.cnn.com/2012/08/14/europe-vc/>

Cerqueiro et al., 2013; Nanda and Rhodes-Kropf, 2015;). On a similar note, employment protection laws might encourage employees in large companies to engage in more experimentation, but can limit the attractiveness for VC investors who need to hire and fire employees to effectively engage in experimentation (Bozkaya and Kerr, 2014).

Finally, societal norms can have important interactions with the formal institutional environment and with the organisational strategies of investors. Cultures in which there is a high stigma attached to failure are ones in which entrepreneurs are less likely to want financing from investors with a reputation for shutting down projects. This can lead investors to pick more failure-tolerant strategies and, in doing so, only finance the less-experimental startups in the economy. Thus, programmes aimed at celebrating the entrepreneur and venture investors, even if unsuccessful, may have important effects.

4.4 Shocks to the availability of capital

Having discussed the costs associated with exercising abandonment options when intermediate information is bad, we turn next to constraints associated with experimentation even when intermediate information is positive. This is because the financing available for startups engaged in innovation is notoriously volatile (Gompers and Lerner, 2004; Kaplan and Schoar, 2005; Gompers et al., 2008), leading entrepreneurs and VC investors to worry about the availability of capital even if initial experiments go well. Venture capitalists refer to this concern as ‘financing risk’ – the risk that the survival of an otherwise healthy startup might be threatened by a negative shock to the supply of capital in its sector when it is looking for the next round of funding.¹¹

This worry seems rational given the ebbs and flows of capital that have occurred within various venture sectors at different times and in distinct places.

In Nanda and Rhodes-Kropf (2014), we model investors' responses to financing risk and explain why investors' responses have a larger effect on the most novel technologies in the economy. Investors can respond to financing risk by providing firms more upfront funding, thus making startups less vulnerable to the future state of the capital markets. This response can effectively eliminate financing risk, but it also comes at a cost – providing firms greater upfront funding reduces investors' ability to abandon their investment if intermediate information on its prospects is poor. In fact, the value of the lost real option can be high enough that it makes the investment unviable. This trade-off between wanting to protect firms from financing risk and wanting to preserve the option to abandon the investment is most salient for firms engaged in radical innovations. Thus, the startups most susceptible to financing risk are those commercialising radical innovations – these are the ventures that are most likely to be funded when financing risk is low, and are most likely to be constrained when financing risk is high. Their work thus provides an intuitive mechanism linking hot and cold financial markets to innovation in the real economy.

In Nanda and Rhodes-Kropf (2014), we show how investors with small pools of capital, who depend more on other investors' willingness to fund the startup in its next round of funding, are more exposed to financing risk. Regions with

¹¹ Large firms who finance with debt face a similar risk, referred to as ‘rollover risk’, when trying to issue new bonds to replace maturing bonds (Acharya et al. 2011; He and Xiong, 2012a,b).

a small number of investors and investors with small funds are therefore more likely to be subjected to financing risk. As shown by Kortum and Lerner (2000), the Prudent Man Rule in the United States, which allowed pension funds and other large institutional investors to make substantial commitments to private equity, seems to have been pivotal in generating a large pool of capital to fund innovation. A big distinction between the United States and Germany, for example, is the number of active ‘large’ VC firms (i.e. with more than \$300 million under management). The size of the fund can have a direct bearing on the degree to which venture capitalists push for bold commercialisation strategies but, as seen above, can be reinforced by the presence of financing risk, which is much more salient for smaller VC investors.

This insight from Nanda and Rhodes-Kropf (2014) can be seen in the context of our model by assuming that there is a probability $(1 - \theta)$ that the firm cannot find $\$X$ when it is ready for the next round of funding. Since $\theta < 1$, including financing risk in the model shows how it reduces the value of experimentation:

$$p_E^* \theta^* (p_G^* V - X) - Y < p_E^* (p_G^* V - X) - Y \quad (4)$$

The introduction of θ implies that some experimental projects will no longer be undertaken. These are firms that were not viable without an experiment (that is, the most novel investments), but are now also not profitable even with an experiment, because of the presence of financing risk. Other startups are likely to be financed with all-or-nothing bets. These latter startups are firms for which:

$$p^* V - X > p_E^* \theta^* (p_G^* V - X) - Y \quad (5)$$

i.e. rising financing risk (smaller θ) causes the expected value with the experiment to be less than without the experiment. These are startups that were not particularly novel, so that the value of the lost abandonment option is not as high. They are better off being protected against financing risk and being funded all in one go.

These results show that times or places with high financing risk (times with low capital availability) are the times/places when high expected value – but safe – projects will be run. This fits the intuition that good, solid firms are funded when capital is not freely available. The results also suggest, however, that investors are more willing to experiment in boom times or places with a great deal of capital. Thus, these become times when, or places where, more novel, experimental startups are financed. Startups funded in boom times/places should be more likely to fail (when investors exercise their abandonment options), but are also likely to have bigger successes. This is what locations with limited capital available for new ideas miss out on – the great success that comes from the few remarkable outcomes (Nanda and Rhodes-Kropf, 2013).

This way of thinking about the funding of innovation suggests that there can be a ‘good’ equilibrium that increases innovation in places like Silicon Valley and in booming time periods, and alternative equilibria that are bad for innovation in other places and times. If we believe that this is an important part of the phenomena, then policy designed to increase innovation should be aimed at trying to break the ‘bad’ equilibria and switch it to the ‘good’ equilibria. This is easier said than done, as there are many arguments for why one equilibrium

or another might currently exist. However, two notions are helpful guides to thinking about policy in this context.

First, concentrated policies are likely to have a larger effect. That is, something that encourages investment in a particular area or sector is more likely to have an impact than a broad-based initiative. In this framework, a policy will have a large impact if it increases the perception that several investors are interested in backing a certain sector, thereby lowering potential shocks to the availability of capital. Thus, a broad policy designed to have a small effect on many companies seems fundamentally less likely to engineer a regime switch. The analogue to policies aimed at encouraging innovation are those aimed at popping technology ‘bubbles’ or preventing investors from losing money in risky ventures. What may look to policymakers as unsound investments in areas with a great deal of failed companies may be vital to experimentation and innovation. In fact, the results from Nanda and Rhodes-Kropf (2014) suggest that the most innovative startups may even need hot financial markets to facilitate their initial diffusion.

The second notion that stems for our model's intuition is that local areas could potentially break the ‘bad’ equilibrium and encourage local innovation by, counterintuitively, creating a mechanism to help the best local companies leave to go to innovation hubs. This should encourage local entrepreneurs and small investors to fund and start companies locally, because they would know if the companies work locally they could be moved to areas where they could get the funding and other resources needed to scale the idea. Once a vibrant startup community has formed locally, investors would naturally arise trying to fund the best before they moved away. Thus, this idea breaks the ‘chicken and egg’ problem.

4.5 Conclusion

A large body of literature in entrepreneurial finance has shown how financing frictions arising from asymmetric information between entrepreneurs and VC investors can lead to credit constraints for high-potential ventures. This chapter complements prior research by focusing on another possible source of financing frictions: the fundamental uncertainty facing startups in their earliest stages, when neither the entrepreneur nor the investor knows about the true potential of the venture without investing in learning about its viability. In this context of extreme uncertainty, multi-stage financing allows investors to learn about a venture's potential over time, without committing the full amount upfront. These real options can be particularly valuable in the context of entrepreneurship, because most new ventures fail completely and only a few go on to become extremely successful. We have shown how constraints to staged financing reduce the value of these real options, and thus influence the degree to which investors can effectively experiment. We show how this has important consequences for the degree to which radical new technologies are commercialised across regions, with important consequences for policies looking to stimulate high potential entrepreneurship.

Formal regulations and informal cultural institutions that make it harder to abandon investments when intermediate information is bad can lead investors to only finance startups when the value of abandonment options is low. These are startups with safer, less novel innovations, with the implication that regions

or firms where it is harder to engage in experimentation are likely to see fewer startups engaged in innovation. In addition, potential shocks to the availability of capital can reduce the value of staged financing. This risk is more salient in regions with a small number of investors, or investors with smaller funds. Again, these constraints to experimentation impact the most novel startups in the economy. Overall, these insights also suggest caution in trying to prevent failure of startup ventures. Failure is a natural part of the experimental process and, in fact, extreme failure and extreme success may be two sides of the same coin.

References

- Acharya, V. and K. Subramanian (2009), "Bankruptcy Codes and Innovation", *Review of Financial Studies* 22(12), pp. 4949-4988.
- Acharya, V., D. Gale and T. Yorulmazer (2011), "Rollover risk and market freezes", *Journal of Finance* 66(4), pp. 1177-1209.
- Aghion, P. and P. Howitt (1992), "A Model of Growth through Creative Destruction", *Econometrica* 60(2), pp. 323-351.
- Aghion, P., J. Van Reenen and L. Zingales (2009), "Innovation and Institutional Ownership", Harvard University Working Paper, Cambridge, MA.
- Bergemann, D. and U. Hege (2005), "The Financing of Innovation: Learning and Stopping", *RAND Journal of Economics* 36(4), pp. 719-752.
- Bergemann, D., U. Hege and L. Peng (2008), "Venture Capital and Sequential Investments", Working Paper.
- Bernstein, S., X. Giroud and R. Townsend (2014), "The impact of venture capital monitoring: Evidence from a natural experiment", *Journal of Finance*, forthcoming
- Black, B.S. and R.J. Gilson (1998), "Venture capital and the structure of capital markets: banks versus stock markets", *Journal of Financial Economics* 47(3), pp. 243-277.
- Bozkaya, A. and W.R. Kerr (2014), "Labor Regulations and European Venture Capital", *Journal of Economics and Management Strategy* 23(4), pp. 776-810.
- Burkart, M., D. Gromb and F. Panunzi (1997), "Large Shareholders, Monitoring and the Value of the Firm", *Quarterly Journal of Economics* 112(3), pp. 693-728.
- Cerqueiro, G., D. Hegde, M.F. Penas and R. Seamans (2013), "Debtor Rights, Credit Supply and Innovation", SSRN Working Paper No. 2246982, Rochester, NY.
- Chemmanur, T., K. Krishnan and D. Nandy (2011), "How does venture capital financing improve efficiency in private firms? A look beneath the surface", *Review of Financial Studies* 24(12), pp. 4037-4090.
- Chemmanur, T.J., E. Loutskina and X. Tian (2012), "Corporate Venture Capital, Value Creation and Innovation", Working Paper, Boston College.
- Cornelli, F. and O. Yosha (2003), "Stage Financing and the Role of Convertible Securities", *Review of Economic Studies* 70(1), pp. 1-32.
- Dushnitsky, G. and M.J. Lenox (2006), "When does corporate venture capital investment create firm value?", *Journal of Business Venturing* 21, pp. 753-772.
- Ewens, M. and M. Rhodes-Kropf (2015), "Is a VC Partnership Greater than the Sum of its Partners?", NBER Working Paper No. 19120, Cambridge, MA (forthcoming in *Journal of Finance*).

- Ewens, M. and M. Marx (2014), "Executive Replacement in Venture Capital-Backed Startups", MIT Working Paper, Cambridge, MA.
- Ferreira, D., G. Manso and A. Silva (2011), "Incentives to Innovate and The Decision to Go Public or Private", LSE Working Paper, London.
- Fluck, Z., K. Garrison and S. Myers (2007), "Venture Capital Contracting: Staged Financing and Later-Stage Syndication of Venture Capital Investments", MIT Working Paper, Cambridge, MA.
- Foster, L., J. Haltiwanger and C. Syverson (2008), "Reallocation, Firm Turnover and Efficiency: Selection on Productivity or Profitability", *American Economic Review* 98(1), pp. 394-425.
- Fulghieri, P. and M. Sevilir (2009), "Organization and Financing of Innovation and the Choice between Corporate and Independent Venture Capital", *Journal of Financial and Quantitative Analysis* 44(6), pp. 1291-1321.
- Gompers, P. (1995), "Optimal Investment, Monitoring and the Staging of Venture Capital", *Journal of Finance* 50(5), pp. 1461-1489.
- Gompers, P. and J. Lerner (2000), "Money chasing deals? The impact of fund inflows on private equity valuations", *Journal of Financial Economics* 55(1), pp. 281-325.
- Gompers, P. and J. Lerner (2004), *The Venture Capital Cycle*, Cambridge, MA and London: MIT Press.
- Gompers, P., A. Kovner, J. Lerner and D. Scharfstein (2008), "Venture capital investment cycles: The impact of public markets", *Journal of Financial Economics* 87(1), pp. 1-23.
- Guler, I. (2007a), "An Empirical Examination of Management of Real Options in the U.S. Venture Capital Industry", in J.J. Reuer and T.W. Tong (eds), *Real Options Theory (Advances in Strategic Management, Volume 24)*, Emerald Group Publishing Limited, pp.485-506.
- Guler, I. (2007b), "Throwing Good Money After Bad? A Multi-Level Study of Sequential Decision Making in the Venture Capital Industry", *Administrative Science Quarterly* 52(2), pp. 248-285.
- Hall, R.E. and S.E. Woodward (2010), "The Burden of the Nondiversifiable Risk of Entrepreneurship", *American Economic Review* 100(3), pp. 1163 -1194.
- Hellmann, T. (1998), "The Allocation of Control Rights in Venture Capital Contracts", *RAND Journal of Economics* 29(1), pp. 57-76.
- Hellmann, T. and M. Puri (2000), "The Interaction between Product Market and Financing Strategy: The Role of Venture Capital", *Review of Financial Studies* 13(4), 959-984.
- Hellmann, T. and M. Puri (2002), "Venture Capital and the Professionalization of Start-Up Firms: Empirical Evidence", *Journal of Finance* 57(1), 169-197.
- Hellmann, T. and V. Thiele (2011), "Incentives and Innovation: A Multitasking Approach", *American Economic Journal: Microeconomics* 3(1), pp. 78-128.
- He, Z. and W. Xiong (2012a), "Dynamic debt runs", *Review of Financial Studies* 25(6), pp. 1799-1843.
- He, Z. and W. Xiong (2012b), "Rollover Risk and Credit Risk", *Journal of Finance* 67(2), pp. 391-429.
- IHS Global Insight (2011), *Venture Impact* (6th edition), Waltham, MA.
- Kaplan, S. and A. Schoar (2005), "Private Equity Performance: Returns, Persistence and Capital Flows", *Journal of Finance* 60(4), pp. 1791-1823.

- Kaplan, S., B. Sensoy and P. Stromberg (2009), "Should Investors Bet on the Jockey or the Horse? Evidence from the Evolution of Firms from Early Business Plans to Public Companies", *Journal of Finance* 64(1), pp. 75-115.
- Kaplan, S.N. and J. Lerner (2010), "It Aint Broke: The Past, Present and Future of Venture Capital", *Journal of Applied Corporate Finance* 22(2), pp. 36-47.
- Kerr, W., J. Lerner and A. Schoar (2014a), "The consequences of entrepreneurial finance: Evidence from angel financings", *Review of Financial Studies* 27(1), pp. 20-55.
- Kerr, W., R. Nanda and M. Rhodes-Kropf (2014b), "Entrepreneurship as Experimentation", *Journal of Economic Perspectives* 28(3), pp. 25-48.
- King, R. and R. Levine (1993), "Finance and Growth - Schumpeter Might Be Right", *Quarterly Journal of Economics* 108(3), pp. 717-737.
- Klepper, S. (1996), "Entry, Exit, Growth and Innovation Over the Product life Cycle", *American Economic Review* 86(3), pp. 562-583.
- Kortum, S. and J. Lerner (2000), "Assessing the impact of venture capital on innovation", *Rand Journal of Economics* 31(4), pp. 674-92.
- Landier, A. (2002), "Entrepreneurship and the Stigma of Failure", MIT Working Paper, Cambridge, MA.
- Manso, G. (2011), "Motivating Innovation", *Journal of Finance* 66(5), pp. 1823-1860.
- Myers, S. (2000), "Outside Equity", *Journal of Finance* 55, 1005-1037.
- Nanda, Ramana. 2015. "Experimentation and Real Options in Entrepreneurial Finance." Harvard Business School Note 815-056, 1-35
- Nanda, R. and M. Rhodes-Kropf (2013), "Investment Cycles and Startup Innovation", *Journal of Financial Economics* 110(2), 403-418.
- Nanda, R. and M. Rhodes-Kropf (2014), "Financing Risk and Innovation", Harvard University Working Paper, Cambridge, MA.
- Nanda, R. and M. Rhodes-Kropf (2015), "Innovation Policies", Harvard University Working Paper, Cambridge, MA.
- Puri, M. and R. Zarutskie (2012), "On the lifecycle dynamics of venture-capital-and non-venture-capital-financed firms", *Journal of Finance* 67(6), pp. 2247-2294.
- Rajan, R. (2012), "The corporation in finance", *Journal of Finance* 67(4), pp. 1173-1217.
- Rosenberg, N. (1994), "Economic Experiments", in N. Rosenborg, *Inside the Black Box: technology and experiments*, Cambridge, UK: Cambridge University Press.
- Sahlman, W.A. (1990), "The Structure and Governance of Venture-Capital Organizations", *Journal of Financial Economics* 27(2), pp. 473-521.
- Saint-Paul, G. (1997), "Is labour rigidity harming Europe's competitiveness? The effect of job protection on the pattern of trade and welfare", *European Economic Review* 41(3), 499-506.
- Samila, S. and O. Sorenson (2011), "Venture capital, entrepreneurship and economic growth", *Review of Economics and Statistics* 93(1), pp. 338-349.
- Schumpeter, J. (1942), *Capitalism, Socialism and Democracy*, New York: Harper.
- Sorensen, M. (2007), "How smart is smart money? An empirical two-sided matching model of venture capital", *Journal of Finance* 62(6), 2725-2762.
- Stern, S. (2006), "Economic Experiments: The role of Entrepreneurship in Economic Prosperity", *Melbourne Review: A Journal of Business and Public Policy* 2(2), pp. 53-56.

Tian, X. and T.Y. Wang (2014), "Tolerance for Failure and Corporate Innovation",
Review of Financial Studies 27(1), pp. 211-255.

5 Patent Rights and Cumulative Innovation: Causal Evidence and Policy Implications

Mark Schankerman

London School of Economics

5.1 Introduction

Cumulative research is a dominant feature of modern innovation. New genetically modified crops, computers, memory chips, medical instruments and many other modern innovations are typically improvements on prior generations of related technologies. Of course, cumulative innovation is not new. Economic historians have emphasised the role of path dependence in the development of technology, documenting how past successes and failures serve as ‘focusing devices’ that guide the direction of later technological inquiry (Rosenberg, 1976). However, the increasing importance of basic science in shaping the direction of technological development has intensified this process.

Cumulative innovation is underpinned by knowledge spillovers, as later innovators build on earlier research. This process lies at the heart of the recent macroeconomic literature on innovation and growth – so-called ‘endogenous growth’ models (Grossman and Helpman, 1991; Aghion and Howitt, 1992; Acemoglu and Akcigit, 2012). At the same time, there is a large body of evidence showing that R&D creates positive ‘knowledge spillovers’ that increase productivity growth and subsequent innovation (e.g. Bloom et al., 2013). This consensus on the centrality of knowledge spillovers to innovation, and innovation to productivity growth, is the primary justification for government policies to support R&D.

There has been an intensifying academic and public policy debate over the role of patents in stimulating innovation and growth. The debate has been driven by several factors. The first is the recognition that modern economies are increasingly based on intangible knowledge assets, and that this is no longer limited to particular sectors. As a consequence, an effective growth strategy requires policies and institutions that promote the generation and diffusion of innovation. The patent system is one of the main instruments governments use to increase research and development incentives, while at the same time promoting follow-on innovation. However, there is growing concern among academic scholars and policymakers that patent rights are themselves becoming an impediment to innovation, rather than the incentive they were originally intended to be. The increasing proliferation of patents and the fragmentation of ownership rights among firms are believed to raise transaction costs, constrain

the freedom of action to conduct research and development, and expose firms to ex post holdup through patent litigation (Heller and Eisenberg, 1998). In the extreme case where bargaining failure in patent licensing occurs, follow-on innovation can be blocked entirely.

These issues are thought to be particularly acute in ‘complex technology’ industries where innovation is highly cumulative and requires the input of a large number of patented components held by diverse firms – leading examples are information technology, software and biotechnology. On top of that, critics claim that (large) firms strategically accumulate patents to use them to resolve disputes through cross-licensing, and this puts small firms, without such patent ‘chits’ to trade, at a disadvantage in enforcing their patent rights. These dangers have been prominently voiced in public debates on patent policy in the United States (National Research Council, 2004, Federal Trade Commission, 2011) and recent decisions by the Supreme Court (e.g. *eBay Inc. v. MercExchange*, 547 U.S. 338, 2006). Similar concerns have also been raised in European policy discussions on the implementation of a unitary European Patent (European Commission, 2011).

In order to design evidence-based government policies that effectively address this potential problem, it is first important to quantify the extent to which patent rights *do* in fact impede follow-on innovation, and to identify whether their impact is pervasive or instead is localised in particular types of technology fields and transacting firms. Broad reforms of the patent system may be required if this blocking effect is widespread and has a substantial blocking effect on follow-on innovation by firms across different technology areas. On the other hand, more targeted policies may be preferable if patents appear to block innovation only in very specific environments.

To date, most of the economic research on the impact of patent rights on cumulative innovation has been primarily theoretical. The main conclusion from these studies is that anything can happen – patent rights may impede, have no effect on, or even facilitate subsequent technological development. It depends critically on assumptions about the bargaining environment and contracting efficiency between different generations of innovators. In an early contribution, Kitch (1977) argues that patents enable an upstream inventor to coordinate investment in follow-on innovation more efficiently and to mitigate the dissipation of profit from downstream competition that can lead to underinvestment. By allowing the upstream innovator to serve as the gatekeeper to coordinate downstream investments, patent rights can facilitate cumulative innovation. In contrast, Green and Scotchmer (1995) show that upstream patent rights will not impede follow-on innovation that increases total value (joint profit) as long as bargaining between the parties is efficient, i.e. if there are no transaction costs and perfect information. While these assumptions are not likely to hold perfectly in most environments, this work is important because it focuses our attention on bargaining failure as the source of any blocking effect patent rights might create. The question, then, is in what kind of environments is bargaining failure more likely?

Finally, a number of papers have shown how patent rights can block innovation when bargaining failure occurs. This can arise from two main sources. First, asymmetric information about the value of the initial or follow-on innovation can lead to the parties failing to agree on a license even though there is joint profit that could be shared (Bessen and Maskin, 2009). Second, bargaining failure can

occur when downstream innovators need to license multiple (complementary) upstream patents that are held by distinct patent holders. Not only does this increase transaction costs but, since bargaining is typically done bilaterally rather than coordinated across the different licensors, this creates the ‘complements (or royalty stacking) problem’ – value maximisation requires coordinated resolution, which is ignored by independent claimants (Shapiro, 2001; Galasso and Schankerman, 2010).

This diversity of theoretical models highlights the need for empirical research. It is important not only to establish whether patent rights block subsequent innovation, but also to identify how this effect depends on the characteristics of the bargaining environment and the transacting parties. Who exactly is blocking whom, and in what settings? Understanding these issues is essential in order to design appropriate policy remedies.

In order to provide a solid foundation for formulating policy in this (and other) areas, we need credible evidence of the *causal relationship* (not just correlations) between patents and later innovation. Given the importance of the issue, there is surprisingly little econometric evidence on this link. In two influential papers, Murray and Stern (2007) and Williams (2013) provide the first causal evidence that patent rights block later research in the biomedical field. Murray and Stern exploit patent-paper pairs to study how citations of scientific papers are affected when a patent is granted on the associated invention. They show that citations of scientific publications fall (by about 15%) when a patent is granted on the innovation associated with that publication. Williams studies the impact of contract-based intellectual property (not patents) on specific genes on subsequent human genome research and a measure of medical diagnostic tests developed on the basis of the specific genes. Interestingly, both papers find roughly similar magnitudes – property rights appear to cause roughly a 15-30% reduction in follow-on research. These important studies focus on very specific (albeit significant) innovations in human genome and biomedical research, and it is hard to know whether their conclusions generalise to other industries.

In this chapter, we report on recent evidence of how patent rights affect the process of cumulative innovation, based on Galasso and Schankerman (2015). This research adopts a novel identification strategy to estimate the causal effect of patents on cumulative innovation. We use the decisions to invalidate patents by the U.S. Court of Appeals for the Federal Circuit, which has exclusive jurisdiction in appellate cases involving patents. Because patents constitute prior art, later applicants are still required to cite patents when relevant even if they have been invalidated and thus put into the public domain. This allows us to trace how the loss of the patent right affects the rate of subsequent citations to that patent, relative to those patents that are upheld by the Court.

The main concern is that unobserved factors might be affecting both the decision to invalidate a patent and the follow-on innovation, leading us to conclude wrongly that the loss of the patent causes the later change in innovation (this is called the ‘endogeneity’ problem). We are able to avoid this potential problem by exploiting the fortunate institutional fact that Federal Circuit judges are assigned to patent cases through a computer programme that randomly generates three-judge panels, with decisions governed by majority rule. This random allocation of judges allows us to pin down the causal relationship between the loss of the patent right and later innovation by other firms.

There are three main empirical findings. First, the loss of patent rights causes about a 50% increase in subsequent citations of the focal patent, on average, and this finding stands up to a wide variety of tests for robustness. Second, this average impact is misleading because there is a huge amount of variation in the effect of patent invalidation on later innovation. For most patents, there is no statistically significant effect; the positive (unblocking) effect of invalidation on citations is concentrated on a small subset of patents which have unobservable characteristics that are associated with a lower probability of invalidity (i.e. stronger patents).

There is also large variation across broad technology fields in the impact of patent invalidation, and the effect is concentrated in fields that are characterised by two features: complex technology and high fragmentation of patent ownership. This finding is consistent with predictions of the theoretical models that emphasise bargaining failure in licensing as the source of blockage. Patent invalidation has a significant impact on cumulative innovation only in the fields of computers and communications, electronics and medical instruments (including biotechnology). We find no effect for drugs, chemicals or mechanical technologies. Importantly, we also are able to confirm these results using measures of later innovation that do not rely on patent citations. In two technology fields – pharmaceuticals and medical instruments – we use data on new product developments (available because of government registration requirements) and in both fields our findings are the same as with citations – patents have no blocking effect in drugs, but do in medical instruments.

Lastly, we show that the effect of patent rights on later innovation depends critically on the characteristics of the transacting parties. The impact is entirely driven by the invalidation of patents owned by large firms, which increases the number of small innovators subsequently citing the focal patent. We find no statistically significant effect of patent rights on later citations when the invalidated patents are owned by small or medium-sized firms. This result suggests that bargaining failure between upstream and downstream innovators is not widespread, but is concentrated in cases involving large patentees and small downstream innovators.

Taken together, our findings indicate that patent rights block cumulative innovation only in *very specific* environments, and this suggests that government policies should be targeted at facilitating more efficient licensing in those environments. Since innovation is the key to sustained productivity growth, policies that improve the market for licensing will make an important contribution to promoting economic growth over the long term.

5.2 Strategy for identifying the causal effect of patent rights on innovation

There are two main challenges in studying the impact of patent rights on cumulative innovation. The first is that we need to identify comparable technologies with and without patent protection. The second is that follow-on innovation is difficult to measure.

In our analysis, we exploit patent invalidation decisions by the U.S. Court of Appeals for the Federal Circuit, established in 1982. We use comprehensive data on 1357 Federal Circuit decisions from 1983 to 2008, and record whether each patent was invalidated. About 40% of the decisions in our sample lead to a loss of patent protection for the technology. We use the number of citations by subsequent patents of the ‘focal’ patent as a measure of cumulative innovation. Patent applicants are required to disclose known prior art that might affect the patentability of any claim and any wilful violation of this duty can render the patent unenforceable. Importantly for our purposes, the expiration or invalidation of a patent has no impact on its prior art status, so the requirement to cite it remains in place. Citations have been widely used in the economics of innovation literature as a proxy for follow-on research (Griliches, 1992), and are the only practical measure of cumulative innovation for studies such as ours that cover a wide range of technology fields. We also show that our results are robust to non-patent measures of cumulative innovation that we are able to construct for two technology fields: pharmaceuticals and medical instruments.

To estimate the effect of patent rights on follow-on innovation, we compare the number of citations received by patents that are invalidated to those that are upheld by the Federal Circuit Court, in a five-year window following the decision. A fundamental challenge with this approach is that invalidated patents may differ from those that are upheld in a variety of dimensions that may affect patent citations. For example, patents covering technologies with greater commercial potential are both more likely to be an attractive target for follow-on innovation and to induce the patentee to invest heavily in the case to avoid invalidation. It is crucial to address this ‘endogeneity’ issue in order to estimate the true causal impact of patent protection on cumulative innovation. We show that failure to do this leads to misleading and incorrect findings.

As mentioned earlier, our empirical strategy exploits the fact that judges are assigned to patent cases through a computer programme that randomly generates three-judge panels, with decisions governed by majority rule. We show that judges on the Federal Circuit Court exhibited very different ‘propensities to invalidate’ in their tenure at the Court – some voted for invalidation much more often than others (varying from about 25% to 75%). The random allocation of judges to cases, together with this variation in their propensity to invalidate patents, essentially means that invalidation of patents is a randomised outcome and thus can be used to identify the true causal impact of removing patent protection (econometrically, we implement this approach using instrumental variables). In conducting this exercise, we control for a number of patent characteristics such as the age of the patent, the technology field, the number of patent claims, and the number of citations received before the Federal Circuit decision. This approach allows us to identify the *causal impact of removing patent rights on later innovation*.

5.3 What does the evidence show?

5.3.1 The 'average effect' of patents

The baseline finding, using our instrumental variable identification strategy, is that the removal of patent protection on a patent leads to about a 50% increase in subsequent citations to that patent, on average. This evidence shows that, at least on average, patents block cumulative innovation, and we emphasise that this is evidence of a causal relationship. It is critically important to use an appropriate identification strategy to pin down causal effects here, especially if one wants to make policy recommendations on the basis of the evidence. If we instead use a simple (OLS) regression that fails to account for the fact that the patent invalidation decision is endogenous, the results indicate that there is no effect on subsequent citations. But this is a false result, since formal statistical tests confirm that patent invalidation is in fact endogenous (i.e. influenced by unobserved factors that also affect subsequent citations). This highlights the importance of using an appropriate identification strategy, and the dangers of drawing policy conclusions from evidence that is not causal.

As additional checks on this key finding, we examine other possible explanations. First, we show that the jump in later citations following the invalidation of a patent is not simply due to a 'publicity effect' from the court's decision – where subsequent innovators become more aware of the patent and thus cite it. The impact begins only after about two years following the court decision, which is consistent with the onset on follow-on innovation rather than simply being a media effect from press coverage associated with the court decision. Moreover, when we introduce a measure of the actual press publicity around the case, the results are the same – on average, patents block later innovation. Second, we examine whether part of the jump in later citations that we observe might reflect greater use of the invention covered by the invalidated patent by later innovators, because it is now cheaper to use when no longer protected by the patent. There is some evidence of this kind of 'substitution', but it can only account for a small part (about 15%) of the overall blocking effect we find.

While the average blocking effect of patents is large, we also find that the impact of patent invalidation on subsequent innovation is highly heterogeneous. This means that the average effect is misleading, and should not form the basis for policy prescriptions. There is a lot of variation across patents – there is essentially *no significant blocking effect for most patents, but a strong effect for a minority of patents*. From a policy perspective, it is very important to understand when patents block, and when they do not so that appropriate, targeted policy remedies can be designed. In our research, we show that the blocking effect depends critically on key features of the technology area and the contracting environment, as we summarise in the next section.

5.3.2 Unbundling the impact: When do patents block?

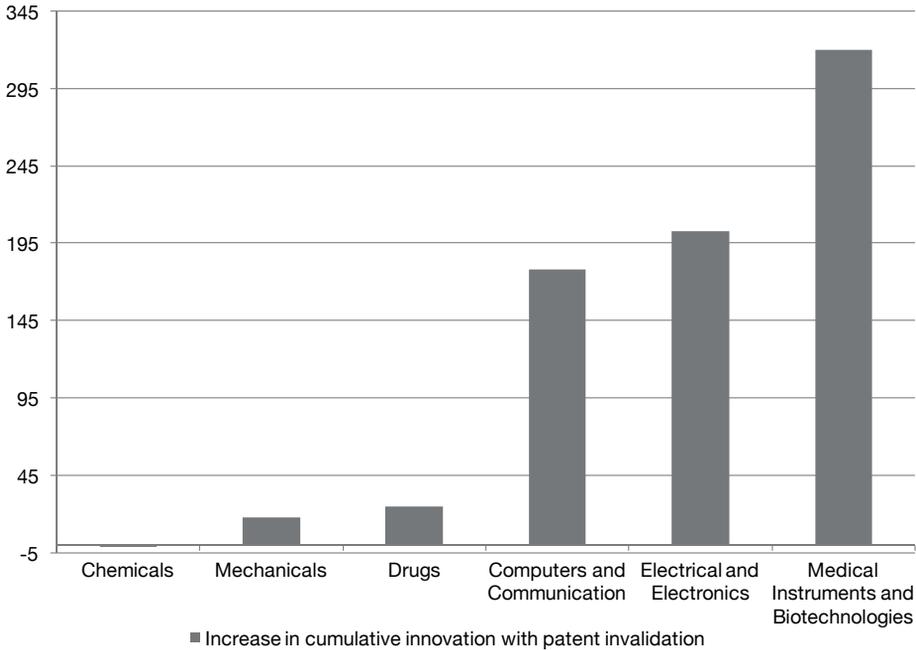
In which technology fields does blocking occur?

Previous empirical studies emphasise two features of the innovation environment that affect bargaining between upstream and downstream firms, and thus the incentives to invest in follow-on innovation. The first is the *fragmentation of patent ownership* in the technology field (Zeidonis, 2004). When patent ownership is fragmented rather than concentrated in a few hands, downstream innovators need to engage in multiple negotiations, which exacerbates the risks of bargaining failure and thus make it more likely that patents end up blocking later innovation. The second feature is the *'complexity' of the technology field*. In complex fields, new products – such as mobile telephones or medical instruments – embody numerous patentable elements, as contrasted with 'discrete' technology areas where products build only on few patents, such as pharmaceuticals or chemicals. When products typically incorporate many patented inputs, and they are held by different owners, licensees need to engage in multiple negotiations and the risk of bargaining failure is higher. Thus we expect the impact of patent rights on cumulative innovation to be more pronounced in complex technology fields.

To test these ideas, we construct two variables. The first is a measure of how concentrated patenting is in the technology field of the litigated patent – we use the share of patenting accounted for by the four largest patent owners in that technology subcategory during the five years preceding the Federal Circuit Court decision. The second is a control variable that identifies which technology fields are complex and which are not, building on earlier survey research by Levin et. al. (1987) and Cohen et al. (2000). Complex technology fields include electronics, computers and communication, medical instruments and biotechnology. Non-complex fields include pharmaceuticals, chemicals and mechanical technologies.

The evidence strongly confirms these hypotheses. We find that the 'blocking effect' of patents is much stronger when patent ownership is fragmented (i.e. where concentration is low) and in complex technology fields. The results indicate that the effect of invalidation is more than twice as large in complex technology areas as compared to the non-complex technology fields, and the blocking effect is much weaker when concentration of patent ownership is greater. Increasing the level of concentration by one standard deviation reduces the blocking effect of patents by about 32% in complex technology fields.

We can use these econometric estimates of the effect of concentration and complexity to compute the implied effect of patent invalidation on citations for each of the technology fields, based on the observed values of concentration and complexity that correspond to each field. The results are summarised in Figure 5.1, and they are striking. Patent rights have no statistically significant effect on cumulative innovation in the pharmaceuticals, chemicals and mechanical technology fields. By contrast, the effect is large and statistically significant in the fields which are complex and where patent ownership is more fragmented: patent invalidation raises citations by 320% in medical instruments/biotechnology, 203% in electronics and 178% in computers and communications.

Figure 5.1 Impact of patent invalidation on follow-on innovation

We want to emphasise that these key findings continue to hold when we use alternative measures of cumulative innovation that do not rely on patent citations. We are able to construct more direct measures of follow-on innovation for two of our technology fields – pharmaceuticals and medical instruments – thanks to government regulation that requires registration of new product developments. These two fields encompass both a ‘complex’ technology area (medical instruments) where we found a strong blocking effect, and a non-complex technology field (drugs) in which we found no blocking effect using the citations measure.

We begin with the medical instruments technology field. The Food and Drug Administration (FDA) in the United States has primary authority to regulate medical devices sold in the country. These products are subject to a regulatory process that requires detailed product information and evidence of safety from clinical trials. The FDA releases data on approvals requested for medical instruments. To use these FDA approval requests as a measure of follow-on innovation, we link them to the medical instrument patents in our sample using two different approaches (for details, see Galasso and Schankerman, 2015). Using these FDA approval requests of new medical devices as the measure of follow-on innovation in our empirical model, we find again that patent invalidation increases cumulative innovation by about the same magnitude as when we use patent citations to measure follow-on innovation. This analysis confirms our conclusion that patent invalidation has a significant impact on cumulative innovation in the complex technology field of medical instruments.

We were also able to do something similar for the pharmaceuticals technology field, again made feasible by exploiting FDA data on approvals of subsequent clinical trials. We construct a measure of follow-on innovation by identifying the subsequent clinical drug trials that are related to the active ingredient of the litigated drug patent. We are then able to match Federal Circuit drug patents with clinical trials by several different methods (details in Galasso and Schankerman, 2015). Using this clinical trials measure of cumulative innovation in our empirical model, we find that the loss of patents through invalidation has no statistically significant effect on cumulative innovation in the non-complex field of pharmaceuticals.

Overall, this analysis with product-based measures of innovation confirms our earlier conclusions from regressions based on patent citation data.

Who is blocking whom?

We showed that the blocking effect of patents on later innovation depends on how concentrated patent rights are, i.e. on the structure of technology markets. However, the influence can also run in the other direction. Patent rights can shape the industrial structure of innovation by impeding the entry of new innovators or the expansion of existing firms, and this potential blocking effect may be stronger for certain kinds of patentees or downstream innovators. We also examine this issue and show that the blocking effect of patents depends critically on the size of the patentee and the downstream innovators.

To understand better where bargaining (licensing) failures occur, we examine whether the blocking effect is stronger for certain kinds of patentees or downstream innovators. We split patentees and citing innovators in three size groups, based on the size of their patent portfolio: 'small' (fewer than 5 patents), 'medium' (6-101 patents), and 'large' (more than 102 patents, which is the 75th percentile of the distribution). This means that we can study the effects of patent invalidation on later citations for six different pairings of patentees and later innovators in terms of their size: small-small, small-medium, small-large, medium-small, medium-medium, medium-large, large-small, large-medium and large-large. The results are very striking. We find that the loss of patent rights has a statistically significant effect only for the large-small pair, that is, patents appear to block only when the patent is owned by a large firm and their impact is only on later citations by small firms.

This finding indicates that patent rights held by large firms appear to impede the 'democratisation' of innovation among small innovating firms. This is of public policy concern, especially because of the increased focus on entrepreneurial, high technology firms. However, it is equally important that we find that patents do not have any significant blocking effect among other types of patent holders and potential licensees. The blocking problem appears to be highly localised, both in terms of the types of technology fields, as described earlier, and the types of contracting parties.

These findings show that fragmentation of patent ownership and complexity of technology fields, and the types of contracting parties – in particular, their size – are key empirical determinants of the relationship between patent rights and cumulative innovation. Of course, other factors can also affect the impact of patent rights on subsequent innovation. One is product market competition. Aghion et al. (2013) provide evidence that strong patent protection stimulates

innovation only when product market competition is fierce. A second factor is the degree to which ‘tacit cooperation’ can be used by firms to mitigate potential bargaining failures and litigation that might otherwise arise from dispersed ownership of patent rights (Lanjouw and Schankerman, 2001, 2004). Understanding where and how these differences operate is a valuable direction for future theoretical and empirical research.

5.4 Policy implications and challenges

Governments use the patent system as an important policy instrument to provide incentives for innovation, and thereby to promote long-run productivity and economic growth. In recent years, however, many scholars and other commentators in the public debate over patent reform have argued that patents are getting in the way of innovation and have recommended scaling back patent rights in various ways. The core concern is that patents are increasingly making it harder for firms to license inputs required for their research, exposing them to hold-up through patent litigation, and generally raising the cost of doing R&D.

If this is true, we should see evidence that patent rights are blocking follow-on innovation. A few recent, high-quality studies have provided credible, causal evidence that patents block cumulative innovation in very specific biomedical subfields. Our research, using a completely different identification strategy to pin down causal effects, demonstrates that, while there is some blocking effect of patents, it is localised and not pervasive. We find that patents block only in very specific technology areas (including biomedical) and only between specific types of contracting parties (large patentees and small later innovators). In other technology fields, and between other contracting parties, there is no evidence that patents block follow-on innovation.

The fact that the impact of patent rights on cumulative innovation is localised, rather than pervasive, suggests that remedial government policies should be targeted. In particular, a ‘broad-based’ scaling back of patent rights is unlikely to be the appropriate policy. As we argued, blocking occurs when patent owners and potential licensees fail to exploit profitable opportunities for follow-on research. This could be because they are unaware of these opportunities, or because bargaining between the parties breaks down for some reason. In the first case, an appropriate policy response is to promote private institutions, or if necessary to set up public ones, that disseminate information to potential licensees – some form of information repository that can be easily and affordably accessed. If the source of the problem is bargaining failure – in particular, as we have shown, between large patent owners and small follow-on innovators – the appropriate response is to design policies and institutions that facilitate more efficient bargaining (as with arbitration and other dispute resolution mechanisms, for example). One interesting example of such institutions are the biological resource centres in the United States studied by Furman and Stern (2011), which reduce the transactional costs of accessing knowledge inputs for biomedical research.

The key focus in patent reform should be on finding ways to reduce transaction costs and bargaining failure in licensing. In this way, governments can promote the process of cumulative innovation (and the long-run productivity growth it creates) without diluting the innovation incentives that patent rights provide.

Finally, while we have focused on the link between patents and cumulative innovation, in formulating public policy toward patent rights it is also important to bear in mind that patents can encourage innovation through a variety of other channels. Perhaps the most important of these is their role in facilitating access to the capital markets for high-technology entrepreneurial firms, both as a source of investment capital and as a means of exit for successful startups. For such firms, whose primary assets are their innovations, patents help secure their rights in these assets and thus allow them to signal their potential more effectively to venture capital and the stock market. There is growing evidence of the importance of this function of patents (Conti et al., 2013). And beyond the capital markets, patents enhance knowledge and technology diffusion across firms (and countries) by allowing innovators to capture part of the benefits from such transfers, most notably through international trade and foreign direct investment (Branstetter et al., 2006; Delgado et al., 2013). These other socially valuable functions of the patent system must also be considered in any evaluation and policy proposals for patent reform.

References

- Acemoglu, D. and U. Akcigit (2012), "Intellectual Property Rights Policy, Competition and Innovation", *Journal of the European Economic Association* 10(1), pp. 1-42.
- Aghion, P. and P. Howitt (1992), "A Model of Growth through Creative Destruction", *Econometrica* 60(2), pp. 323-351.
- Aghion, P., P. Howitt and S Prantl (2013), "Patent Rights, Product Market Reforms, and Innovation", NBER Working Paper No. 18854, Cambridge, MA.
- Bessen, J and E. Maskin (2009), "Sequential innovation, patents, and imitation", *RAND Journal of Economics* 40(4), pp. 611-635.
- Bloom, N., M. Schankerman and J. Van Reenen (2013), "Identifying Technology Spillovers and Product Market Rivalry", *Econometrica* 81(4), pp. 1347-1393.
- Branstetter, L., R. Fishman and C.F. Foley (2006), "Do Stronger Intellectual Property Rights Increase International Technology Transfer? Empirical Evidence from U.S. Firm-level Panel Data", *Quarterly Journal of Economics* 12(1), pp. 321-349.
- Cohen, W., R. Nelson and J. Walsh (2000), "Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)", NBER Working Paper No. 7552, Cambridge, MA.
- Conti, A., J. Thursby and M. Thursby (2013), "Patents as Signals for Startup Financing", *Journal of Industrial Economics* 61(3), pp. 592-621.
- Delgado, M., M. Kyle and A.M. McGahan (2013), "Intellectual Property Protection and the Geography of Trade", *Journal of Industrial Economics* 61(3), pp. 733-762.
- European Commission (2011), "Implementing enhanced cooperation in the area of the creation of unitary patent protection", Commission Staff Working Paper, Brussels.
- Federal Trade Commission (2011), *The Evolving IP Marketplace: Aligning Patent Notice and Remedies with Competition*, Washington, DC: Government Printing Office.

- Furman, J. and S. Stern (2011), "Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research", *American Economic Review* 101(5), pp. 1933-1963.
- Galasso, A. and M. Schankerman (2010), "Patent Thickets, Courts and the Market for Innovation", *RAND Journal of Economics* 41(3), pp. 472-503.
- Galasso, A. and M. Schankerman (2015), "Patents and cumulative innovation: causal evidence from the courts", *Quarterly Journal of Economics* 130(1), pp. 317-369.
- Green, J. and S. Scotchmer (1995), "On the Division of Profit in Sequential Innovation", *RAND Journal of Economics* 26(1), pp. 20-33.
- Griliches, Z. (1992), "The Search for R&D Spillovers", *Scandinavian Journal of Economics* 94, pp. 29-47.
- Grossman, G. and E. Helpman (1991), *Innovation and Growth in the Global Economy*, Cambridge, MA: MIT Press.
- Heller, M. and R. Eisenberg (1998), "Can Patents Deter Innovation? The Anticommons in Biomedical Research", *Science* 280, 698-701.
- Kitch, E. (1977), "The Nature and Function of the Patent System", *Journal of Law and Economics* 20(2), pp. 265-290.
- Lanjouw, J. and M. Schankerman (2001), "Characteristics of Patent Litigation: A Window on Competition", *RAND Journal of Economics* 32(1), pp. 129-151.
- Lanjouw, J. and M. Schankerman (2004), "Protecting Intellectual Property Rights: Are Small Firms Handicapped?" *Journal of Law and Economics* 47(1), pp. 45-74.
- Levin, R., A. Klevorick, R. Nelson and S. Winter (1987), "Appropriating the returns from industrial R&D", *Brookings Papers on Economic Activity* 3, 783-820.
- Murray, F. and S. Stern (2007), "Do formal intellectual property rights hinder the free flow of scientific knowledge? An empirical test of the anti-commons hypothesis", *Journal of Economic Behavior and Organization* 63(4), pp. 648-687.
- National Research Council (2004), *A Patent System for the 21st Century*, Report by the Committee on Intellectual Property Rights in the Knowledge-Based Economy, Board of Science, Technology and Economic Policy, Washington, DC: National Academies Press.
- Shapiro, C. (2001), "Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard Setting", in A. Jaffe, J. Lerner and S. Stern (eds), *Innovation Policy and the Economy*, Vol. 1, Cambridge, MA: MIT Press.
- Williams, H. (2013) "Intellectual Property Rights and Innovation: Evidence from the Human Genome", *Journal of Political Economy* 121(1), pp. 1-27.
- Ziedonis, R. (2004), "Don't Fence Me In: Fragmented Markets for Technology and the Patent Acquisition Strategies of Firms", *Management Science* 50(6), pp. 804-820.

When approaching the knowledge frontier, an economy's capacity to innovate must shift from imitation and differentiation towards more radical and more risky innovations that aim at entirely new products and services. Tertiary education, basic research and technological infrastructure become more critical factors in activating private innovation and generating continued growth.

Patent protection allows firms to cash in on successful innovations for a while, but tense competition from potential and actual new competitors forces them to continuously invest in new R&D. In a firm's lifecycle, innovation-driven growth creates the need to enter world markets for further growth. In the cross-section, exporting firms and multinational companies are thus substantially more productive and larger than other firms with domestic sales only.

Innovation-based growth is a process of creative destruction, reflecting market entry and exit of young firms, and the creation of new product lines and closing down of old ones by large firms. Labour and capital must flow to new uses. About half of a country's productivity growth is due to a targeted allocation and ongoing reallocation of investment and employment to more valuable uses. When a country moves closer to the knowledge frontier, innovations become more risky and factor reallocation must occur on a larger scale. Flexible capital and labour markets can support innovation by facilitating factor reallocation. Welfare policy should combine unemployment insurance with low job protection and active labour market policies for retraining and supporting job search. Financing should shift from credit to relatively more equity financing, giving a larger role to stock markets, venture capital and private equity.

These and other ideas are explored in this report in five essays by Philippe Aghion, Ufuk Akcigit, Ramana Nanda and Matthew Rhodes-Kropf, William Kerr, and Mark Schankerman, based on the invited lectures at the CEPR conference "Moving to the Innovation Frontier" held on 19-20 January 2015 in Vienna.



CEPR

33 Great Sutton Street, London EC1V 0DX

Tel: +44 (0)20 7183 8801

Email: cepr@cepr.org www.cepr.org