Corporate Innovation, R&D Personnel and External Knowledge Utilization
Corporate Innovation, R&D Personnel and External Knowledge Utilization

By Wenjing Wang

A PhD thesis submitted to
School of Business and Social Sciences, Aarhus University,
in partial fulfilment of the requirements of
the PhD degree in
Economics and Business

August 2015
# Contents

**Introduction**

**Chapter 1.** Firm-level innovation activity, employee turnover and HRM practices – evidence from Chinese firms  
  
**Chapter 2.** More outsourced R&D, fewer internal specialists?  

**Chapter 3.** The effect of knowledge from market research on product innovation: Evidence from longitudinal data
Acknowledgements

Six years ago, when I chose the master program at former Aarhus School of Business, I had zero intention to further pursue a PhD. At that time, I felt fearful towards the PhD, which is popularly interpreted as the abbreviation of “Permanent Head Damage” in China. Today, as my PhD study approaches its finishing line, I feel blessed and grateful, not only because my head is still intact (thanks to the Danish habit of wearing a helmet), but also because my fear retreats and my view broadens - with the help from my remarkable supervisors, colleagues, friends and other scholars whom I have encountered along the way.

First and foremost I want to thank my supervisor Tor Eriksson. He has taught me, through not only words but also his own behaviors, about how a successful economist contributes to academic society. His questions always trigger deeper thinking; his way of interpreting results is mind-opening; and he generously shares these valuable ideas to all. Whenever I get caught by negative thoughts, he helps me see the bright side - his office is full of sunshine, where I can get guidance, encouragement and energy whenever I need it. He is always ready to provide instant and effective help, while being very patient with my slow working style. He gives me trust; his powerful recommendation helps me find the next excellent employer. He reads all my drafts very carefully, and provides insightful and constructive suggestions on all dimensions - from research question formation, to paper structure, to grammar details. Only recently have I begun to understand the first things he
told me - so I expect that he must have taught me more than I realize at this point; as time goes by, I will realize more things to be appreciated.

Professor Qin from Renmin University is my second supervisor. During my visits to Beijing, he provided the data for our first paper, gave me insightful comments, and introduced me to his other PhD and EMBA students, through whom I had a glimpse of some thriving business leaders in China.

Frederic Warzynski and Valerie Smeets both have major roles in my academic trajectory. Actually Frederic is the first professor who I met in Aarhus, when I attended his lecture Industrial Economics. He speaks really clearly and his slides are well-structured. He explains complex things in a simple way. His lecture has built my confidence in studying a program in English language in Denmark. Valerie’s lecture Strategy of Economics is also excellent. Her well-selected reading materials actually got me interested in the research on personnel economics and firm strategy. I appreciate the important information they provide to us students – particularly at lunch table, workshops and seminars.

My PhD committee members - Frederic Warzynski, Keld Laursen and Ulrich Kaiser, have given many helpful comments and suggestions for this thesis. Their questions are to the point, and their suggestions are precise and implementable. The new references mentioned in committee report are highly relevant and inspiring, which I appreciate very much. After having revised according to their suggestions, the chapters are in much better shape.
Special thanks go to Anders Frederiksen, who gives me a great future work opportunity. It is such a pleasure to work with him; in light of the coming work prosperity, the final phase of the thesis writing process has been increasingly enjoyable. He generously supported my thesis revision process and trips to Druid15 conference and Ce2 workshop, so that I was able to talk to top researchers in the area and receive high-quality feedbacks, encouragements and inspirations from them.

Sincere appreciation goes to my dear colleagues. Susanne Christensen has been giving me all kinds of effective help during my PhD study. She is very considerate and proactive; many severe headaches (such as that of my residence permit) just disappear with her help. I appreciate Bo Rasmussen’s kindness, trust, and encouragement to us PhD students. He works hard on providing us with the best possible opportunities and encourages us to make the best use of them. Thomas Stephansen has helped with the proof reading; with his professional help, the errors in grammar and word choice etc. have been reduced significantly. The administrative team in our department has been very supportive and provided me the best possible working environment. Special thanks go to Lene Bavnbek Enevoldsen and Christel Brajkovic Mortensen, who have given me generous help in addition to doing an excellent job.

I would like thank former and current PhD students and other colleagues at our department, in particular my officemate, office neighbors and other colleagues who I have often met at the lunch table or coffee machine. I appreciate the great conversations, cheerful atmosphere, useful information, and the fun times I have had together with them.
Thanks to my friends. They make my expat life colorful. I appreciate our get-togethers, when we share meals, play card games or go for excursions, etc. My good friends, in particular my gym mates, have been helping me develop a healthy life-style and a positive outlook to my future career and life.

Finally, I would thank my family. Chinese culture generally prioritizes work to family. Although I am trying to follow the Danish value of work-life balance, my family prioritizes my work automatically. Language is not enough to describe my appreciation to their love, care and all that they have done for me. For now, I practice working hard and effectively, so that we will have more quality time together.

I appreciate this journey of my PhD study. At the end of this voyage, the view becomes broader and brighter, while the shadow of permanent head damage fades away.
Introduction

This thesis comprises three self-contained studies related to firm’s R&D and innovation activities. They endeavor to uncover some regular patterns about how a firm’s innovation interacts with its own management and marketing practices, internal R&D personnel, and external R&D partnerships. These studies share a common empirical approach in the sense that the analyses are based on econometric models estimated from rich datasets, but each study applies a different set of estimation methods catered for the specific research context.

Compared to the conventional approaches such as case studies, the empirical approach based on econometric models can uncover some regular pattern among a large number of observations and quantify the effects derived from theoretical or case analyses. This systematic empirical approach can add to our knowledge of firm R&D and innovation, an area with many interesting theories to be tested or quantified in a systematic way. Recent development in econometrics provides more effective methods for isolating the effect of interest from unobserved influences, and these methods are especially beneficial for studies on firm decisions and outcomes. Alongside a development of advanced econometric methods, statistic agencies have now accumulated the data of firm innovation practices and outcomes from many dimensions in a consistent way. In a nutshell, the resources are ready to be tapped into, and it is a good time to apply advanced econometric methods to the high-quality data and to uncover systematic evidence for the antecedents and outcomes related to
the firm R&D and innovation activities, which are regarded as major drivers of economic development.

The thesis exploits the above-mentioned developments in methodology and data and empirically examines how R&D and innovation activities interact with several other factors of the firm. The following provides an overview of each chapter.

Chapter 1 examines employee turnover, HRM practices and innovation activities in Chinese high-tech firms. We estimate hurdle negative binomial models for count data. In addition to exploiting the discrete nature of outcome variables, these models also allow each explanatory variable to have distinct impacts on the probability of innovating and on the intensity of innovation. Innovation activities are measured from two dimensions: one is innovation effort, which is measured by the number of ongoing innovation projects; the other is innovation performance, which is measured by the number of new products. The results show that higher R&D employee turnover is associated with a higher probability of having innovation, but a lower intensity of innovation activities in innovating firms. Innovating firms are more likely to have adopted high performance HRM practices, and the impact of employee turnover varies with the number of HRM practices implemented by the firm.

Chapter 2 is motivated by the question: “Does the firm employ fewer R&D specialists as it outsources more R&D?” Both theoretical analysis and previous evidence suggest that the employment implications of R&D outsourcing activities vary across two distinct dimensions: outsourcing depth and outsourcing breadth, which reflect how deeply and
widely the firm leverages on external knowledge. By estimating Correlated Random Effects (CRE) Tobit, CRE Selection and CRE Fractional Response Models on a panel dataset of Danish firms, the study finds that both the absolute number and the employment intensity of R&D specialists decrease when R&D outsourcing deepens but increase when it broadens.

Chapter 3 is motivated by the question: “To what extent can producers improve product innovation performance by utilizing the knowledge acquired from users through different market research practices?” Inspired by March (1991), this study categorizes producers’ market research practices (MR) into two types - exploitive MR and explorative MR. Then it examines how product innovation responds to the use of knowledge from MR in early stages of new product development. By applying program evaluation methods to longitudinal data from over 3000 firms in Denmark, this study finds that both the knowledge from exploitive MR and the knowledge from explorative MR improve product innovation performance. It also reveals that the knowledge from explorative MR has slightly larger effects than the knowledge from exploitative MR. In addition, higher effectiveness of innovation investment is found among the firms which utilize the knowledge from more types of MR into the early stage of new product development. Overall, knowledge from exploitive MR, knowledge from explorative MR, and innovation investment improve product innovation performance in a coordinated way.
Introduktion

Denne afhandling består af tre selvstændige studier relateret til virksomheders forskning og udvikling (FoU) og innovationsaktiviteter.

Formålet er at vise nogle regelmæssige mønstre af, hvordan en virksomheds innovation interagerer med dets egen ledelses og marketings praktikker, med det interne FoU personale og eksterne samarbejdspartnere.

De tre studier har en fælles empirisk tilgang forstået på den måde, at alle tre analyser er baseret på økonometriske modeller som er estimeret ud fra store datasæt. Dog anvender hvert enkelt studie en differentieret økonometrisk model som tager højde for den konkrete forskningskontext.

Sammenlignet med de traditionelle tilgange, såsom casestudier, kan den empiriske tilgang baseret på økonometriske modeller afdække nogle regulære mønstre blandt et stort antal observationer og kvantificere effekterne, som kan uddrages fra teoretiske analyser eller casestudier.

Den systematiske empiriske tilgang kan tilføje os viden om FoU og innovation på virksomhedsniveau, hvor der er fundet rigeligt med interessante mønstre via casestudier eller teoretiske analyser, som blot venter på at blive generaliseret eller kvantificeret.

Nyere udvikling indenfor økonometri giver metoder til at isolere effekten af uobserveret heterogenitet; disse avancerede metoder er meget nyttige i særlig for studier af virksomheders
beslutninger, hvor man normalt møder store udfordringer i forhold til uobserveret heterogenitet og endogenitet.

Sideløbende med udviklingen indenfor avanceret økonometrisk metode, hvad enten tilsigtet eller utilspigt, har statistiske organisationer konsekvent akkumuleret rigelige mængder data vedrørende virksomhedens innovationspraktik samt resultater i mange dimensioner.

Kort sagt er ressourcerne klar til at blive udnyttet, og nu er et godt tidpunkt at anvende avancerede økonometriske metoder til data af høj kvalitet og dermed finde systematisk bevis for fortilfælde og resultater relateret til virksomhedens FoU og innovationsaktiviteter, som vurderes som værende en væsentlig drivkraft indenfor økonomisk udvikling.

Denne afhandling benytter sig af ovennævnte udviklinger indenfor metode og data og undersøger empirisk hvordan en virksomheds FoU og innovationsaktiviteter spiller sammen med adskillige andre faktorer. I det følgende gives en oversigt over hvert kapitel.

Kapitel 1 undersøger personaleomsætning, human resource-management praktik og innovationsaktiviteter i kinesiske high-tech virksomheder. Vi estimerer forhindringerne ved negative binomiale modeller for count data. Ud over at undersøge resultatvariablenens diskrete natur tillader disse modeller også hver forklarende variable at have distinkte virkninger på sandsynligheden for innovation og på innovationsintensiteten.

Innovationsaktiviteterne måles ud fra to dimensioner: den ene er innovationsindsatsen, som måles ud fra det antal af igangværende innovationsprojekter, den anden er innovationsresultater, som måles ud fra antallet af nye produkter.

Kapitel 2 er motiveret af spørgsmålet: rummer virksomheden færre interne FoU specialister i takt med at den udliciterer mere forskning og udvikling? Både teoretisk analyse og tidligere evidens tyder på at beskæftigelseskonsekvenserne i forbindelse med virksomhedens FoU udliciteringsaktiviteter varierer på tværs af distinkte dimensioner: udliciteringens dybde og udliciteringens bredde, hvilket afspejler henholdsvis hvor dybt og hvor bredt virksomheden udnytter ekstern viden.

Ved at estimere Correlated Random Effects (CRE), Tobit, CRE Selection og CRE Fractional Response modeller på paneldata af danske virksomheder, finder man frem til, at både the absolutte antal samt intensiteten af interne FoU-specialister falder når FoU udliciteringen fordybes, men stiger når den udvides.

Kapitel 3 er motiveret af spørgsmålet: i hvilket omfang kan producenterne forbedre produktinnovationen ved at udnytte viden erhvervet fra brugerne gennem forskellige former for praktik?

Under henvisning til March (1991), kategoriserer dette studie producenterernes praktik ved indhentning af brugerviden i to typer: udnyttende læring og udforskende læring og dernæst undersøges deres indvirkning på produktinnovationen.
Ved at anvende program evalueringsmetoder til longitudinalt data fra 3,150 virksomheder i Danmark, fastslår denne afhandling at både udnyttende og udforskende læring af brugerviden øger producenternes produktinnovation, samt at udforskende læring har en smule større indvirkning på produktinnovation.

Desuden finder man blandt de virksomheder, der integrerer mere brugerviden på et tidligt stadie af produktudviklingen, et højere niveau af effektivitet fra innovationsinvestering.

Samlet set supplerer udnyttende og udforskende læring af brugerviden og innovationsinvestering hinanden og de bidrager begge positivt til produktinnovation på en koordineret måde.
Chapter 1

Firm-Level Innovation Activity, Employee Turnover and HRM Practices – Evidence from Chinese Firms
Firm-Level Innovation Activity, Employee Turnover and HRM Practices – Evidence from Chinese Firms*

Tor ERIKSSON, Zhihua QIN, Wenjing WANG

ABSTRACT
This paper examines the relationship between employee turnover, HRM practices and innovation in Chinese firms in five high technology sectors. We estimate hurdle negative binomial models for count data on survey data allowing for analyses of the extensive as well as intensive margins of firms’ innovation activities. Innovation is measured both by the number of ongoing projects and new commercialized products. The results show that higher R&D employee turnover is associated with a higher probability of being innovative, but decreases the intensity of innovation activities in innovating firms. Innovating firms are more likely to have adopted high performance HRM practices, and the impact of employee turnover varies with the number of HRM practices implemented by the firm.

Keywords: Innovation, HRM practices, employee turnover
JEL classification: L22, M50, O31

* This paper has been published in China Economic Review (2014), 30: 583–597.
1. Introduction

Employee turnover can be an important mechanism for innovation activities in firms. Persistent differences in turnover between two otherwise identical organizations will evolve very different tenure distributions, with implications for stability and organizational culture which in turn may have considerably different implications for innovation. The level of turnover can be a result of the human resources management (HRM) practices chosen by the firm, but the HRM practices can also have a direct effect on innovation activities of the firm. This paper examines the relationship between employee turnover, HRM practices and innovation activity in Chinese high technology sector firms.

In the current stage of China’s economic development, innovation is considered as one of the key factors for continued increase in total factor productivity and hence sustaining high growth; see e.g., World Bank (2011). Very little systematic evidence of the drivers of innovation activities based on firm-level data exists for China. Empirical results from other (mostly advanced industrialized) countries, which are also rather scarce, do not necessarily generalize to the Chinese context, as labor markets in China are still relatively underdeveloped and protection of intellectual property rights remains weak. Moreover, Chinese firms also differ from Western firms with respect to corporate culture and a more important role for business groups and other networks.

For our empirical analysis we make use of data from a survey carried out by researchers at Renmin University (Beijing) in 2011. The sample consists of firms in China from five (high technology) industries: energy, electronic information, biotechnology, equipment manufacturing and environmental protection. In addition to standard controls in the analysis of innovation activities, the data set includes information about the firm’s HRM practices as well as measures of employee turnover for different categories, including technical personnel. The dependent variables in our analysis are the number of ongoing R&D projects and new commercialized products during 2010.

\[1\]

In fact, we are only aware of one article (written in English) by Wei, Liu and Herndon (2011) on this topic.
The econometric analysis is performed using a hurdle negative binomial model for count data. An advantage of this model is that it allows for analyses of both extensive and intensive margins.

The empirical analysis shows that a higher turnover rate of R&D personnel is associated with a higher likelihood that a firm is innovating but a lower level of R&D effort and innovation performance in innovating firms. Particularly important HRM practices for enhancing innovation are the use of job description manuals and training programs. Notably, employee turnover has larger impact on innovation performance for firms using more high performance HRM practices. Among the other drivers of innovation, external network cooperation attaches an especially large and positive marginal effect. This is perhaps not so surprising in view of the importance of networks and business groups in the Chinese corporate system.

The remainder of the paper unfolds as follows. Next, a brief review of the previous studies of the relationship between HRM practices, employee turnover and innovation is given. The third section describes the data and the econometric method used. The results are presented and discussed in sections four and five, respectively. Section six briefly concludes.

2. Previous Research

Since the mid-nineties a fairly large literature has built up dealing with HRM and firm performance. Performance is typically measured by productivity (surveyed in Bloom and van Reenen, 2011), while there is rather little (beyond case studies) on HRM and innovation.² Instead, the large innovation literature has mainly been concerned with firm size, product market competition, knowledge spillovers and R&D collaboration.

It is somewhat surprising that there is relatively little amount of work on HRM practices and innovation in view of the fact that the interest in new work practices emphasizing delegation of authority, empowerment of employees, information sharing and employee involvement, originated

² Notable exceptions are Michie and Sheehan (2003), Laursen and Foss (2003), Jimenez-Jimenez and Sanz-Valle (2005) and Zoghi, Mohr and Meyer (2010). See also the recent survey by Foss and Laursen (2012).
from the focus on the Japanese firms’ organization of workplaces in which horizontal information flows play a key role. The interest in the Japanese work organization and job design was to a high extent due to the fact that they were largely considered as the main determinants of the high level of innovation and quality improvement that characterized Japanese firms; see e.g., Applebaum and Batt (1994).

The first two papers to look at the relationship between HRM practices and innovation were Michie and Sheehan (1999), (2003) in which the authors examined British firms’ use of so-called high- and low-road HRM practices and how these were related to firms’ R&D expenditures (the 1999 study) and process and product innovations during a three year period (the 2003 paper), respectively. They find that extensive use of modern (that is high-road) practices is positively correlated with investments in R&D and with process (but not product) innovations. Laursen and Foss’s (2003) study investigates bundles of work practices and the degree of novelty in product innovation in Danish firms and finds a positive relationship. Jimenez-Jimenez and Sanz-Valle’s (2005) analysis of a relatively small sample of Spanish firms find that participative practices and promotion plans significantly increases the firm’s innovation orientation.

In a more recent study, Zoghi, Mohr and Meyer (2010) use Canadian longitudinal data to study how workplace organization is correlated with the adoption of process and product innovations. They find that decentralized decision-making, information sharing programs and (individual) incentive pay are associated with more innovations. Another recent study by Zhou, Dekker and Kleinknecht (2011) makes use of four waves of survey data from the Netherlands and finds that functional flexibility (measured by the rate at which people change their function or department within the firm) has a positive effect on the percentage of sales due to new products. Thus, this, as well as other studies, finds some evidence suggesting that internal labor mobility (functional flexibility, job rotation) is important for innovation activities.

3 Notably, they also find, but do not discuss, that firms with a high vacancy rate (which is likely to be a sign of high employee turnover) are also more likely to innovate.
Although there are a number of studies suggesting that especially the new, high involvement/performance work practices are implemented more frequently in innovative firms, the evidence is not very strong. Summarizing and concluding from the earlier empirical studies is difficult because these have not only made use of many different measures of innovation as the dependent variable but also included quite different measures of HRM practices. Moreover, it should be pointed out that the mechanisms behind the relationship are not well understood. A key candidate is that HRM practices promote learning processes of individuals as well as organizations (Cohen and Levinthal (1989), Shankar and Ghosh (2013)); for a systematic study of this mechanism for a developing country, see Santiago and Alcorta (2012).

Could we expect the relationship between HRM practices and innovation to be different in the Chinese case? A central element in the modern work practices is delegation of decision rights to employees. This may not, however, function well in a Chinese context where keeping distance to superiors and showing respect to elders is deeply rooted in the culture. Participative decision making also presupposes a high level of trust between employees at different levels in the hierarchy, which is often said not to be present in Chinese workplaces; see Wang, Yeung and Zhang, (2011) for empirical evidence. Another cultural difference that may weaken the effect of introducing modern HRM practices is that, in appraisals of performance, the employee’s attitude and behavior is traditionally considered more important than the results of her performance.4

There is to the best of our knowledge, only one earlier study, Wei et al. (2011), of the relationship between HRM practices and innovation in firms operating in China. The data used in that study was collected by a survey questionnaire sent to both CEOs and HR managers in firms in various industries (manufacturing accounts for only 24 of the respondents). Strategic HRM is measured using Huselid’s (1995) eight-item instrument and the firm’s product innovativeness (relative to industry average) is self-reported (that is, is assessed by the respondents). The results show a

---

4 Nevertheless, a number of studies have documented a positive relationship between the firm’s use of strategic HRM practices and its performance (typically measured by subjective ratings of the overall performance of the firm); see e.g., Björkman and Fan (2002), Chow, Huang and Liu (2008), Ngo, Lau and Foley (2008), and Wei, Liu, Zhang and Chiu (2008).
positive relationship between the strategic HRM measure and product innovation. The correlation is stronger for firms with flatter structure and developmental culture.

One particular aspect of firms’ internal labor markets that has attracted some attention recently (Møen, 2005; Kaiser et al., 2008; Müller and Peters, 2010) is the role of worker flows and employee turnover for firms’ innovation activities. As knowledge and competencies are embodied in people it is important to consider how these are transferred between firms. Two broad hypotheses have been put forward. The idea behind the first hypothesis is that with low employee turnover the result is likely to be result in too little experimentation and innovation. This is especially the case if the relevant employees are hired after graduation from college or university (or some vocational education) and therefore possess little professional experience from other firms or industries. As this brings few ideas from other companies, the firm itself becomes less capable of exploring new environments and adapting to changing technologies. Instead the focus will be on existing product performance – the improvements are typically small – and on efficiency within existing technology and product variety. As long as there is not sufficient flow of “new blood”, economic incentives, employee empowerment and involvement, cross-functional teams and adoption of new information technologies can do little to radically change the innovation activities within the firm. All in all, this implies that a too low turnover of personnel is associated with a low level of innovation activity.

The idea that employee turnover above a certain threshold is good for firm performance is related to the risky hires hypothesis put forward by Lazear (1995), according to which employees are thought of as real options and the firm’s choice is between a candidate with relatively predictable performance and one more risky. As long as firing costs are low, it may pay off for the employer to hire the risky candidate because s/he has an option value. Potential benefits are likely to be largest for positions where small differences in talent can have large impacts on performance outcomes.

---

5Two recent papers by Balsvik (2011) and Parrotta and Pozzoli (2012) examine the impact of between firm employee flows on firm productivity.
(such as R&D and leadership positions). Hiring risky workers is likely to result in higher employee turnover, but may nevertheless give better performance in terms of innovation and creativity.

The alternative hypothesis states that in order to promote innovation a firm needs to employ and retain appropriate staff. As innovative employees need to undergo relatively much training, a lower rate of employee turnover is also typically associated with lower training costs. A related argument is that innovative firms should provide employment security as a means to get the employees involved in their firm as this is important for innovation. Another argument goes back to Jovanovic’s (1979) job-worker matching paper according which long job tenures reflect good matches between the job (or employer) and the employee.

All in all, whether the relationship between worker turnover and innovation is positive or negative is an empirical matter. However, the evidence is so far rather scant. An early paper by Ettlie (1985) examines (by means of cross-tabulations) the role of new personnel and net manpower flows on process and product innovation in a small sample of food processing firms. He finds that “new blood” is good for major process innovations, whereas the opposite holds for product and minor process innovations. Kaiser et al. (2008) carry out a considerably more elaborated analysis in which they distinguish between R&D employees leaving and joining the firms. The innovation outcome is number of patents applied for. Their results indicate that the rate of R&D employees leaving the firm has, not surprisingly, a negative impact on its patent activity, while an inflow of R&D employees affects patents positively. The net mobility effect is found to be positive. A recent paper by Müller and Peters (2010) makes use of the firm level churning rate of R&D employees as a measure of workforce turnover and their empirical analysis allows for non-linearities in the turnover-innovation relationship. They find that an increase in the churning rate, up to a certain threshold, is associated with a higher likelihood that the firm has innovated during the previous three years. Plausibly, the threshold is lower for process innovations than for product innovations.

---

6 The direction of causality could in both cases go the other way.

7 The churning rate is the employee turnover rate at which employment is unchanged, that is, it is a measure of the extent of replacement hires during a given time period.
There are two reasons for why the role of employee turnover for innovation is particularly interesting in a Chinese context. One is that after the removal of the lifelong employment ("iron rice bowl") system, employment security seems to have lost some of its importance. Hence, average turnover rates are reported to be quite high (annual rates of 20-40 percent have been mentioned); see Schmidt (2011). Another reason is that the new Labor Contracts Law which came into force in 2008 aims at providing more employment security by requiring formal employment contracts and introduces costs for employers in connection with employee displacements. Thus, evidence shedding light on whether and to which extent employee mobility enhances or decreases firms’ innovation behavior is called for. It should be noted that a traditional feature of Chinese internal labor markets has been a relatively slow involvement of new employees in firms and organizations. Decision rights typically lie with small informal groups, and so, to become a member takes time. Thus, in firms where these features still are present, a positive relationship between employee turnover and innovation could be weaker, and even negative.

3. Data Description and Econometric Method

3.1. Data and Variables

The data used in this study comes from a survey collected by researchers at Department of Organization and Human Resources at Renmin University (Beijing) in 2011. The survey targeted firms in five industries: energy, electronic information, biotechnology, equipment manufacturing and environmental protection, all of which are considered as high tech industries in China (World Bank, 2011). For each firm, the data set contains in addition to basic firm information such as total number of employees, establishment year, industry, etc., information about its overall performance outcomes, inputs and outputs of innovation activities, the number of total employees as well as of technical employees who voluntarily left the firm in each year during the period 2008-2010, detailed information about the innovation environment, organizational strategy and the use of HRM practices.

The sample comprises 582 companies. The five industries account for about 20% each. 45.7% of them are state-owned enterprises, private companies make up 26.6% and the rest have a mixed ownership structure. 89.4% are financed domestically, 2.2% are financed by foreign capital while
8.4% are joint ventures. The firm size ranges from 17 employees to 300,000 employees, though 95% of observations have fewer than 55,000 employees; the median firm has 2,534 employees.

Next we turn to discuss the key variables used in our empirical analysis. The dependent variables are the firm’s R&D efforts, which is measured by the number of ongoing R&D projects during year 2010, and innovation performance, which is measured by the number of new commercial products in 2010. The number of R&D projects is a direct measure of R&D effort and also highly relevant to outcomes of both product and process innovation; the number of new commercial products is a typical measure for innovation performance. They contain richer information than the binary indicators which are used in many earlier innovation studies. Together, these two variables provide a more detailed picture of the firm’s innovation activity, not only of the innovation outcome, but also the intermediate role of R&D efforts, which are in the black box transforming the labor and management input into innovation outcomes.

On average, a sample firm has 56 ongoing R&D projects in 2010. The distribution is heavily skewed: 14.4 per cent of the firms have no R&D project at all; among firms with at least one R&D project, the median number is 30, which is only half of the average. 95% of the firms have 280 or fewer ongoing R&D projects, while the most R&D active firms have over 1,000 projects. The distribution of the number of new commercial products in 2010 is also skewed: 167 firms in our sample do not have commercialized any new product in 2010 and the median innovating firm has 15 new commercial products, while the average firm has 33.4. About 95% of the firms have 100 or less new commercial products.

INSERT TABLE 1 ABOUT HERE

---

8 A weakness of this measure is that what constitutes a new product is a subjective assessment and that new products can also be developed by buying external resources rather than through internal innovation activities.
One explanatory variable of interest is the voluntary turnover rate of technical employees. It is a proxy for the turnover of R&D employees, which is not available in our data set. This is measured by the number of technical employees who left the firm voluntarily\(^9\) divided by the total number of technical employees. In general the voluntary turnover rate of technical employees is much lower than the turnover rate of all employees, which is not what is found in more advanced economies, such as the U.S., where the opposite pattern is strong (Shankar and Ghosh, 2013). Thus, 21.5 per cent of the firms did not experience any turnover of technical employees in 2010 and 95% of the firms had less than 6.5 per cent of their technical staff leaving voluntarily. On the other hand, the highest turnover levels observed in the sample exceed sixty per cent.

The HRM practices used by the firm make up the other category of variables of special interest in the current study. First, in order to measure high performance practices, we include four binary variables indicating whether the firm uses: job rotation schemes, job description manuals, operation standard manuals, and training programs. In order to capture the degree of the usage of these high performance HRM practices, we construct a variable counting the number of these practices adopted by the firm. Moreover, we use a binary variable to track a special HRM practice that is often highlighted in Chinese context – whether there is a formal channel for employees’ suggestions. Two of the HRM practices are measured as continuous variables: the proportion of base salary of an R&D employee’s annual income, and the firm’s training expenditures in 2010.

Table 1 contains some descriptive statistics. For two key variables, we also report the subsample statistics for top and bottom quintiles of their distributions. The technical employee turnover rate in the top (bottom) quintile with respect to the number of R&D projects is clearly below (above) average. The turnover rate in the top quintile with respect to new commercial products is almost twice as high as in the bottom quintile. The average number of the HRM practices examined in our study is 3.9. This number varies only little by quintiles of the number of R&D projects or new

\(^9\) This concept presents no ambiguities in the context of Chinese firms as this is part of the regular reporting of their HR offices.
commercial products. The operation standards and job description manuals are adopted in almost all the high-tech firms in our sample.

There seems to be a negative relationship between R&D efforts and innovation outcomes and employee turnover, whereas there appears to be no systematic relation regarding the number of HRM practices adopted by the firms. Of course, a concern is that the measures of HRM practices or the technical employee turnover rate are proxying some other traits of the firm (like size or industry). To examine this we include observable controls in the more formal regression analysis presented below.

First of these firm characteristics we include is firm size, which is measured by two indicators - the total number of employees and firm output (sales) in 2010. A sizable literature has shown that large firms are more likely to innovate.\(^\text{10}\) Another factor which has been extensively studied in the innovation literature is product market structure. We do not have access to measures of the degree of competition faced by the sample firms, but expect the dummy variables for industry affiliation to pick up the variation in competitive pressure across (but not within) industries. A third factor which potentially needs to be accounted for, especially in the Chinese context, is the ownership structure of the company. We use two dummy variables, one for state owned enterprise (SOEs) and another for foreign owned firms. This is motivated by the fact that SOEs, for historical reasons, may differ from private firms in several respects, especially regarding management style, which may influence both innovation activity and employee turnover. Foreign firms may choose to limit their R&D activities in their Chinese subsidiaries due to the relatively weak protection of intellectual property rights (Yang and Jiang, 2007).

Firm-level network expansion is measured by the number of new cooperating partners in 2010. Networks have been found to be related to innovation outcomes in earlier research (e.g., Rittera and Gemünden, 2004; Whittington et al., 2009; Huggins et al., 2012). It is also likely that employee

\(^{10}\) Employee turnover has also been shown to be positively correlated with firm size, which is another reason for controlling for it.
mobility could be facilitated by inter-firm cooperation. Hence it is important to control this factor when examining the relation between the employee turnover rate and innovation activity.

As we are treating R&D as a production process, it is also necessary to control for R&D inputs. Here we enter the total number of technical employees and total investments in R&D and innovation activities in order to control for the scale factor which influences the technical employee turnover rate, R&D and innovation. Furthermore, we control for profitability, which is measured by the annual profits divided by the value of assets. Previous studies suggest that higher profits are associated with both more innovation activity and better management practices, and so it is important to control for this factor to mitigate omitted variables bias.

Finally, we include the turnover of non-technical employees as a proxy variable for unobserved time-invariant effects, such as management quality. Following Blundell et al. (2002), we enter the average number of R&D projects/new commercial products in previous years as a proxy for unobserved fixed effects such as the firm’s knowledge capital.

### 3.2. Econometric Model

This paper utilizes a logit hurdle negative binomial model\(^\text{11}\), which assumes that the probability of observing no R&D projects or new commercial products in firm \(i\) during year 2010 is:

\[
P(Y_i = 0 | Z_i) = \frac{1}{1 + \exp(Z_i' Y)}
\]

and the probability of observing a positive number of R&D projects or new commercial products is:

\[
P(Y_i = y_i | X_i) = \frac{[1 - P(Y_i = 0 | X_i)] g(Y_i = y_i | X_i)}{1 - g(Y_i = 0 | X_i)}
\]

---

\(^{11}\) An alternative model for dependent variables that spread over a large range of positive values and cluster at zero is the Tobit model. This assumes variables are continuous and that the process generating zero outcomes is the same as the process generating the positive value outcomes. We have estimated our model using a Tobit specification and obtain largely similar results.
where $Y_i$ denotes the positive number of R&D projects or the number of new products by firm $i$ during 2010, $Z_i$ and $X_i$ denote firm characteristics, and $g(\cdot)$ is a probability function following a negative binomial II model.

(1) is a logit model, which describes the binary process determining whether the firm carries out R&D projects (or has new commercial products) or not; equation (2) is a zero truncated negative binomial model, in which $g(\cdot)$ describes the count process determining the intensity of R&D activity or new product commercialization. For further details of the model, see Appendix 1.

### 3.2.1. Choice of Model

The reason for choosing this particular econometric model is as follows. First, as our dependent variable is of count data type, it can (and does) attach multiple values and has no upper bound. This implies we have to make a choice between two major categories of count data models: basic models including the Poisson model and the negative binomial model, and furthermore, between one and two parts models, that is, zero inflated models and hurdle models, respectively.

A key feature of the data set used in this paper is that the proportion of zero outcomes is large relative to the other count values. It is therefore important to check whether the processes generating zero outcomes and strictly positive outcomes differ. Since the factors influencing firm’s decision regarding whether to innovate or not can be quite different from those influencing the decision concerning the intensity of innovation, it makes intuitively sense to relax the constraint that the zeroes are generated in the same way as positives. Furthermore, even if the same factors influence both the decision to innovate and the innovation intensity, their influences may exhibit different patterns.

The next choice is between hurdle count models and zero inflated models. The difference between them is that in the zero inflated models the zero outcomes can arise in two ways: as a consequence of strategic decisions (with probability $w$) and incidentally (with probability $1-w$), while the hurdle count model assumes that zeroes are exclusively the outcomes of strategic decisions (for details, see Lambert, 1992; Winkelmann, 2008). In the current case, it seems very unlikely that a firm that has adopted an innovation strategy would not have any ongoing R&D projects/new commercial products during the year. Consequently, we think the hurdle count model is a more natural specification to adopt in this context.
There is a further choice between a hurdle Poisson model and a hurdle negative binomial model. The former model nests in the latter: the Poisson model assumes that the variance equals the conditional expectation, while the hurdle negative binomial allows the variance to grow faster than the expectation (over-dispersion). The likelihood ratio test of over-dispersion factor alpha confirms that in our case over-dispersion is indeed present in the second part of model (see Table 2, below).

3.2.2. Marginal Effects

As the coefficient estimates from the hurdle negative binomial models are difficult to interpret, we have computed marginal effects implied by the estimated models. For binary part of the model, the marginal effects describe the influences on the probability of being an innovative firm; for the count part of the model, the marginal effects describe the influences on the expected number of R&D projects/new commercial products among innovating firms. These computations and their relations to the model estimations are discussed in more detail in Appendix 1.

Subsequently, for each part of the model, two types of marginal effects will be reported: (i) average marginal effects and (ii) conditional marginal effects. The average marginal effects are calculated for all observations by first obtaining individual marginal effects by inserting the true values of the regressors into the marginal effect formula for each observation after which we compute the average of the individual marginal effects. The conditional marginal effects for certain groups of firms are calculated by inserting given values (e.g., median group values) into the marginal effects formula.

4. The Roles of Employee Turnover Rate and HRM practices

In the following presentation of the estimation results we will first report estimates from a specification with HRM practices measured as a count variable (Table 2) and next consider estimates of dummy variables for the individual HRM practices (Table 3). For the purpose of comparison, each table presents estimates of the same regressors on two dependent variables: the number of R&D projects and new commercialized products. This is followed by a discussion of other factors and the interaction of technical employee turnover and the HRM practices.
4.1 The Role of Technical Employee Turnover

From Table 2 it can be seen that the turnover rate of technical employees\textsuperscript{12} is negatively and significantly correlated with both the number of R&D projects and the number of new commercialized products. For the median firm, a one percentage point increase in the technical employee turnover rate is associated with about 0.7 fewer R&D projects and 0.8 fewer new commercial products. Thus, firms with a higher technical employee turnover rate exert less R&D effort and have less innovation output.

\textbf{INSERT TABLE 2 ABOUT HERE}

On the other hand, the turnover rate of technical employees is positively and significantly associated with the probability of having at least one new commercialized product; on average, a one percentage point higher turnover rate is accompanied by about a one per cent increase in the probability of having a new commercial product. For the probability of having a R&D project, its impact is also positive, albeit imprecisely estimated.

In sum, while a higher technical employee turnover increases the probability of having at least one new commercial product on a yearly basis, it is negatively associated with the intensity of both R&D effort and innovation output. In other words, it takes a certain level of R&D labor mobility to bring in new ideas that initiate R&D projects, but once firms cross the threshold and engage in innovation, their R&D efforts and innovation performance are discouraged by a higher technical employee turnover.

\textbf{INSERT TABLE 3 ABOUT HERE}

\textsuperscript{12} To reduce the problem of possible simultaneity bias, the employee turnover rate is entered with a one-year lag.
4.2. The Role of HRM Practices

The number of high performance HRM practices adopted by the firm is significantly and positively correlated with both the number of R&D projects and the number of new commercial products. For a median firm in our sample, one additional HRM practice is associated with three more R&D projects and 2.5 additional new commercial products per year; see Table 2.

As for the effects of the individual high performance HRM practices, shown in Table 3, many of the individual practices attach insignificant estimates. To some extent this may be due to collinearity among some of the more widely implement practices. Still, a few things are worth noting. First, introducing training programs in firms that are already innovating gives rise to a higher R&D and innovation intensity. Second, firms using job rotation schemes are more likely to have produced at least one new commercial product. This makes sense as job rotation facilitates communication and exchange of ideas which in turn activates innovation. Third, job description manuals exist in nearly all firms with R&D projects and consequently have no impact on the probability component. However, they are strongly and positively associated with the number of new commercial products. Finally, we may note that a higher share of performance pay (that is, a lower base salary share) increases the probability of having at least one new commercial new product in a given year. So, while higher incentive pay does not lead to more ongoing R&D projects, it seems to give rise to more commercialized products.
4.3. The Impact of Other Factors

As for the control variables, a few results are also worth noting. In line with the earlier literature, the size of the firm is found to be positively related to its innovation activities. A firm with a higher output level is more likely to produce a new commercial product every year and an increase in the firm’s employees increases both the number of R&D projects and commercialized new commercial products. Firms with more external cooperation partners have significantly more R&D projects. For the median firm it means that a one per cent increase in the number of cooperation partners gives rise to two additional R&D projects and three additional new commercial products. More profitable firms have a higher number of R&D projects. Of course, here it should be noticed that causality can go both ways.

The proxy for unobserved heterogeneity, the pre-2010 sample average of the dependent variable, turned out highly significant in all models. The estimated effect is not very large, however. The turnover rate of non-technical employees is negatively associated with probability of having at least one new commercial product.

4.4. The Interaction between Employee Turnover and HRM Practices

It seems plausible to assume that a firm’s employee turnover rate is related to its HRM policies. Thus, we first experimented with excluding either the HRM practices variables or the technical

---

13 As was noticed in Table 1, there is considerable variation in the key variables and hence a concern is that the estimates may be influenced by outliers. Therefore, we have estimated the models on shaved samples excluding the top 5 per cent observations on the number of R&D projects, new commercial products and firm size (number of employees). For the binary part the estimates are very robust whereas there are some changes and in the magnitudes (not sign and significance) for the count part estimates. Overall, the estimates for the saved samples do not differ much from those of the full sample. We also estimated the models separately for SOEs and non-SOEs. As the SOEs make up 46 per cent of the sample, this means the estimation samples are substantially smaller. In the R&D project models the coefficient estimates and conditional marginal effects for the technical employee turnover rate and the number of HRM practices are slightly larger in magnitude (and associated with higher significance levels) for the SOEs than for the full sample. For the models of new commercial products, employee turnover has a larger impact for SOEs and the number of HRM practices attaches a positive and significant estimate for non-SOEs.
employee turnover rate in the models in Tables 2 and 3. The estimates, which are not shown to save space, turned out to be quite insensitive to the inclusion of the exclusion of the abovementioned variables. This suggests that although the turnover rate of technical employees and HRM practices both influence the number of R&D projects and new commercial projects, their influences seem to operate via different channels. Simple (unreported) Tobit model regressions also show that the HRM practices are only weakly correlated with differences in the technical employee turnover rate.

However, a closer inspection reveals that the technical employee turnover rate may have different influences depending on firms’ use of different sets of HRM practices. This is shown in Table 4, which contains the conditional marginal effects. The firms are divided into subgroups according to the number of HRM practices adopted, and for each subgroup, the marginal effects are calculated by setting other values equal to their subsample median values.

For the number of R&D projects we find nonlinear relations for the number of HRM practices, the technical employee turnover rate and the number of new cooperating partners. One additional HRM practices has a larger marginal effect for firms with a few practices (three or less), but as the firm has implemented more than three practices, the marginal effect of adopting an additional practice decreases. The same pattern is observed for the number new cooperating partners. The negative effect of the turnover rate of technical employees is also first increasing in the number of HRM practices but is lower when the firm has more than three practices. For the other dependent variable, the number of new commercial products, the magnitudes of the marginal effects are increasing in the number of HRM practices implemented by the firm. Thus, for instance, is the negative effect of the turnover rate of technical employees almost three times larger for a firm having only one of the HRM practices examined than for a firm that has adopted five of the practices.

5. Discussion

We have found that the technical employee turnover rate is higher in innovating firms than in non-innovating firms and that for firms that are carrying out R&D projects and are producing new commercial products further increases in the turnover rate discourage the firm’s R&D efforts and lower its innovation performance. These findings are in accordance with the inverted U-shaped
relations between employee turnover and innovation which have also been documented by some previous studies; see e.g., Müller and Peters (2010). The theoretical reasons for a positive relationship – new ideas brought into the firm by new employees – are supported by our study, but we also find that as the marginal benefits of employee turnover decreases while its costs increases, the impact of higher turnover eventually turns negative.

Thus, one of the key findings of the empirical analysis is that technical employee turnover in Chinese high-tech firms has a negative influence on both R&D effort and innovation performance. While, as discussed in Section 2, this is not necessarily inconsistent with theory, several distinguishing features of the Chinese innovation environment may also contribute to the differences observed with respect to studies for more developed economies.

First, despite significant improvements, the protection of intellectual property in China is still relatively poor. For a given level of employee mobility, Chinese firms face a higher risk of being copied by competitors and losing benefits of innovation due to the transmission of information via leavers. Second, due to the large population, the average number of job candidates for one position in China is very large, which results in higher recruitment cost and higher risk of mismatch. Both factors reduce the net benefit brought by newcomers to innovation. So, again, firms facing high labor mobility have weaker incentives to innovate more because the high labor mobility creates leakages of the gains from innovation.

Third, the Chinese corporate culture is often considered to be more conservative than Western culture. It is generally believed that the best strategy for new joiners is “shut the mouth and open the eyes and ears”, which limits the spillover of new ideas from new joiners. Since the positive effect of labor mobility on innovation (which is mainly brought by new employees joining the firm) is smaller, it is less likely that high labor mobility contributes to more R&D efforts activities or higher innovation performance.

Moreover, in Chinese firms it is usually relatively small informal groups which actually make the decisions. A newcomer’s ideas are not be valued unless she is involved in one of these small groups or a member thereof speaks for her. As it takes time for small groups to accept newcomers, Chinese firms are slower in observing a newcomer’s innovative ideas and in reaping the benefits thereof. Although Chinese firms are claimed to have spent large amounts of money on poaching high level technical employees, the overall impact on firm level innovation seems to be negative. Lastly, the general level of trust in China is lower (Wang et al., 2011), which means that the newcomers are
less trusted and as a consequence their ideas are less valued. This is reinforced by the fact that as lack of trust is mutual, new employees do not commit themselves to innovate either.

Overall, there are a number of reasons for why the higher employee turnover among technical employees is less likely to facilitate R&D activity in China. This list of characteristics specific to the internal labor markets of Chinese firms is of course mainly speculative. It should be noted, however, that Aoshima (2008) also finds a negative effect of the mobility of engineers on Japanese companies’ innovation performance. As China and Japan have many elements of a conservative corporate culture in common, this may explain why the labor mobility seems to influence innovation activity in China and Japan differently than in Western countries.

6. Concluding Remarks

In this paper we have examined the empirical relationships between high-tech firms’ technical employee turnover rates, HRM practices, R&D efforts and innovation performance using Chinese firm-level survey data from five high-tech industries. Unlike earlier studies, the analysis distinguishes between decisions whether or not to innovate and decisions of how much to innovate and therefore estimates logit hurdle negative binomial models.

A notable feature of the high-tech studied is that the turnover rate of technical employees is lower than that of the firm’s overall workforce. We find that technical employee turnover is higher in innovating firms but that higher turnover in firms that are already innovating has negative effects on R&D effort as well as on innovation performance. As for high performance HRM practices, we find that they contribute both higher R&D effort and innovation performance. Moreover, the negative relationship between employee turnover and R&D effort and innovation performance is stronger in firms which have adopted more of the high performance HRM practices, which indicates that these practices increase the value of the employees to the firm for its innovation activities.
References


APPENDIX 1: The Logit Hurdle Negative Binomial Model

Mullahy (1986) first offered solutions for how to deal with the situation when the zero outcomes of the data generating process differ from the positive ones within a hurdle model framework. Cameron and Trivedi (1986, 1998) further developed the model. Generally, hurdle models contain two parts: a binary probability model which determines whether the outcome is zero or not, and a truncated model which describes the positive outcomes.

This paper utilizes a logit hurdle negative binomial model. The first part of the model captures the probability of being non-innovative firm, which can be expressed as:

$$P(Y_i = 0|Z_i) = \frac{1}{1+\exp(Z_i'\gamma)} \quad (A-1)$$

The second part of the model captures the process generating positive outcomes, which follows negative binomial model. The probability of observing $y_i$ R&D projects can be expressed as:

$$P(Y_i = y_i|X_i) = \frac{[1-P(Y_i=0|X_i)]g(y_i=y_i|X_i)}{1-g(Y_i=0|X_i)} \quad (A-2)$$

Where $g(.)$ is a probability function following the negative binomial II model.

The negative binomial model is obtained by generalizing the Poisson model by introducing an individual unobserved effect $\epsilon_i$ into the conditional mean:

$$\mu_i = E_{NB}(Y_i|X_i, \epsilon_i) = \exp(\beta_0 + \beta_1X_{i1} + \beta_2X_{i2} + \cdots + \beta_pX_{ip} + \epsilon_i) \quad (A-3)$$

The distribution of $Y_i$ conditional on $X_i$ and $\epsilon_i$ follows the Poisson form (Zaninotto and Falaschetti, 2011):

$$g(Y_i = y_i|X_i, \epsilon_i) = \frac{e^{-\mu_i\exp(\epsilon_i)}y_i^{y_i}}{y_i!} \quad (A-4)$$

The unconditional distribution $g(Y_i = y_i|X_i)$ is the expected value over $\epsilon_i$ of $g(Y_i = y_i|X_i, \epsilon_i)$:

$$g(Y_i = y_i|X_i) = \int_0^\infty e^{-\mu_i\exp(\epsilon_i)}\frac{y_i^{y_i}}{y_i!}h(\exp(\epsilon_i)) d(\exp(\epsilon_i)) \quad (A-5)$$

The choice of density $h(.)$ for $\exp(\epsilon_i)$ defines the unconditional distribution. In the negative binomial II model, $\exp(\epsilon_i)$ is assumed to be Gamma distributed with $E[\exp(\epsilon_i)]=1$: 
\[ h(\exp(\epsilon_i)) = \frac{\theta^\theta}{\Gamma(\theta)} e^{-\theta \exp(\epsilon_i)} [\exp(\epsilon_i)]^{\theta-1} \]  

where \( \Gamma(\cdot) \) is the Gamma function, such that \( \Gamma(s) = \int_0^\infty z^{s-1} e^{-z} \, dz \) for \( r > 0 \) (Winkelmann, 2008). Let

\[ \lambda_i = \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip}) \]  

(A-7)

Then, \( E(\mu_i) = E[\lambda_i \exp(\epsilon_i)] = E[\lambda_i] \cdot E[\exp(\epsilon_i)] = E[\lambda_i] = \lambda_i \), and the unconditional distribution (A-5) can be expressed as (Greene, 2012):

\[ g(Y_i = y_i | X_i) = \frac{\Gamma(\theta + y_i)}{\Gamma(\theta) \Gamma(y_i + 1)} \left( \frac{\theta}{\theta + \lambda_i} \right)^\theta \left( \frac{\lambda_i}{\theta + \lambda_i} \right)^{y_i} \]  

(A-8)

and

\[ g(Y_i = 0 | X_i) = \left( \frac{\theta}{\theta + \lambda_i} \right)^\theta = (1 + \theta^{-1} \lambda_i)^{-\theta} \]  

(A-9)

Hence, the latent heterogeneity \( \epsilon_i \) induces over-dispersion:

\[ Var[Y_i | X_i] = \lambda_i \left[ 1 + \frac{1}{\theta} \lambda_i \right] = \lambda_i [1 + \kappa \lambda_i], \text{where } \kappa = Var[h_i], \]  

while preserving the conditional mean:

\[ E_{NB}(Y_i | X_i) = \lambda_i = \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip}) \]

Insert functions (A-1), (A-8) and (A-9) into (A-2) and obtain the probability functions of the count part model:

\[ P(Y_i = y_i | X_i, Y_i > 0) = \frac{\exp(z_i' \gamma) \Gamma(\theta + y_i)}{[1 + \exp(z_i' \gamma)] \Gamma(y_i + 1) \Gamma(\theta)(1 - (1 + \theta^{-1} \lambda_i)^{-\theta})} \left( \frac{\theta}{\theta + \lambda_i} \right)^\theta \left( \frac{\lambda_i}{\theta + \lambda_i} \right)^{y_i} \] for \( y_i = 1, 2, 3 \ldots \)  

(A-10)

Insert (A-7) into (A-10), then parameters \( \beta, \gamma \), and \( \theta \) can be estimated using Maximum Likelihood.

Firm \( i \)'s expected number of R&D projects, conditional on \( Y_i > 0 \) and \( X_i \) is:

\[ E(Y_i | Y_i > 0, X_i) = \frac{E_{NB}(Y_i | X_i)}{1 - g(Y_i = 0 | X_i)} = \frac{\exp(x_i' \beta)}{1 - [1 + \theta^{-1} \exp(x_i' \beta)]^{-\theta}} \]  

(A-11)

The expected number of R&D projects conditional on \( X_i \) is:

\[ E(Y_i | X_i) = [1 - P(Y_i = 0 | X_i)] E(Y_i | Y_i > 0, X_i) = \frac{\exp(z_i' \gamma + x_i' \beta)}{[1 + \exp(z_i' \gamma + x_i' \beta)] \{1 - [1 + \theta^{-1} \exp(x_i' \beta)]\}^{-\theta}} \]  

(A-12)
For a continuous variable $x_{ij}$ that appears only in $\mathbf{X}_i$ (count part), its conditional marginal effect of the count part model is obtained by differentiating A-11 with respect to $x_{ij}$:

$$\frac{\partial E(Y_i | Y_i > 0 | X_i)}{\partial x_{ij}}$$  \hspace{1cm} (A-13)

which depends on all the values of regressors $\mathbf{X}_i$.

Since $\frac{E(Y_i | X_i, x_{ij}+\delta)}{E(Y_i | X_i, x_{ij})} = e^{\beta_j \delta}$, the interpretation of $\beta_j$ is that: for a change of $\delta$ in $x_{ij}$, the expected number of R&D projects increases by a factor of $\exp(\beta_j \delta)$, or by $100 \times \exp(\beta_j \delta) \%$.

For the continuous variable $z_{ik}$ that appears only in $\mathbf{Z}_i$ (the binary part), its marginal effect is found by differentiating A-1 with respect to $z_{ij}$:

$$\frac{\partial P(Y_i = 0 | z_{ik})}{\partial z_{ik}} = -\frac{\gamma_k \exp(z_{ik}' \gamma)}{[1 + \exp(z_{ik}' \gamma)]^2}$$  \hspace{1cm} (A-14)

In the logit model, the reported marginal effect is calculated as:

$$\frac{\partial P(Y_i > 0 | z_{ik})}{\partial z_{ik}} = \frac{\gamma_k \exp(z_{ik}' \gamma)}{[1 + \exp(z_{ik}' \gamma)]^2}$$  \hspace{1cm} (A-15)

which also depends on all the values of regressors in $\mathbf{Z}_i$.

Since the coefficient of the binary equation $\gamma_k = \partial \log \left[ \frac{P(Y_i > 0 | \mathbf{Z}_i)}{P(Y_i = 0 | \mathbf{Z}_i)} \right] / \partial z_{ik}$ [from (A-1)], it can also be interpreted directly as marginal change in the log value of relative probability of being an innovative firm to being a non-innovative firm with respect to the change in $z_{ik}$.

For continuous variables that appear in both $\mathbf{X}_i$ and $\mathbf{Z}_i$ such that $x_{ij} = z_{ik}$ for some $j, k$, the overall marginal effect is obtained through differentiation of (A-12) with respect to $x_{ij}$ ($z_{ik}$):

$$\frac{\partial [P(Y_i > 0 | \mathbf{Z}_i)]}{\partial z_{ik}} \text{E}(Y_i | Y_i > 0, X_i) + \frac{\partial E(Y_i | Y_i > 0 | X_i)}{\partial x_{ij}} [1 - P(Y_i = 0 | \mathbf{Z}_i)]$$  \hspace{1cm} (A-16)

which depends on all the values of regressors in $\mathbf{X}_i$ and $\mathbf{Z}_i$.

For a discrete variable, its partial effect is the difference in predicted values as the variable changes from 0 to 1 while all other variables are held constant at specified values.

For the binary part, the partial effect of a discrete variable $z_{ik}$ is:

$$P(Y_i > 0 | z_{ik} = 1, x_{1,2..,k-1,k+1..m} - P(Y_i > 0 | z_{ik} = 0, x_{1,2..,j-1,j+1..n}, z_{1,2..,k-1,k+1..m})$$
\[ P(Y_i = 0| z_{ik} = 0, z_{i,1,2..,k-1,k+1..m}) - P(Y_i = 0| z_{ik} = 1, x_{1,2..,j-1,j+1..n}, z_{1,2..,k-1,k+1..m}) \] (A-17)

which can be calculated using equation (A-1).

For the count part, the partial effect of a discrete variable \( x_{ij} \) is:

\[ E(Y_i | Y_i > 0, x_j = 1, x_{1,2..,j-1,j+1..n}) - E(Y_i | Y_i > 0, x_j = 0, x_{1,2..,j-1,j+1..n}) \] (A-18)

which can be calculated using equation (A-11).
<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of R&amp;D Projects</td>
<td>55.98</td>
<td>126.25</td>
</tr>
<tr>
<td>Number of New Commercial Products</td>
<td>573</td>
<td>33.43</td>
</tr>
<tr>
<td>Voluntary Turnover Rate of Technical Employees</td>
<td>1.27</td>
<td>3.81</td>
</tr>
<tr>
<td>Number of R&amp;D Projects in Subsamples</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zero</td>
<td>0.15</td>
<td>0.34</td>
</tr>
<tr>
<td>Top 20% Firms</td>
<td>0.74</td>
<td>1.11</td>
</tr>
<tr>
<td>Bottom 20% Firms</td>
<td>2.21</td>
<td>7.56</td>
</tr>
<tr>
<td>Number of New Commercial Products in Subsamples</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zero</td>
<td>0.34</td>
<td>1.49</td>
</tr>
<tr>
<td>Top 20% Firms</td>
<td>1.49</td>
<td>2.13</td>
</tr>
<tr>
<td>Bottom 20% Firms</td>
<td>2.70</td>
<td>9.61</td>
</tr>
<tr>
<td>Firms with 5 HRM Practices</td>
<td>0.59</td>
<td>1.43</td>
</tr>
<tr>
<td>Firms with 2 or fewer HRM Practice</td>
<td>2.52</td>
<td>4.48</td>
</tr>
<tr>
<td>Number of HRM Practices in Subsamples</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zero</td>
<td>4.38</td>
<td>0.83</td>
</tr>
<tr>
<td>Top 20% Firms</td>
<td>4.11</td>
<td>0.76</td>
</tr>
<tr>
<td>Bottom 20% Firms</td>
<td>3.95</td>
<td>0.84</td>
</tr>
<tr>
<td>Number of New Commercial Products in Subsamples</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zero</td>
<td>4.35</td>
<td>0.79</td>
</tr>
<tr>
<td>Top 20% Firms</td>
<td>3.92</td>
<td>0.76</td>
</tr>
<tr>
<td>Bottom 20% Firms</td>
<td>3.75</td>
<td>0.90</td>
</tr>
<tr>
<td>Training Programmes</td>
<td>0.71</td>
<td>0.45</td>
</tr>
<tr>
<td>Operation Standards Manuals</td>
<td>0.98</td>
<td>0.14</td>
</tr>
<tr>
<td>Job Description Manuals</td>
<td>0.90</td>
<td>0.30</td>
</tr>
<tr>
<td>Job Rotation Schemes</td>
<td>0.53</td>
<td>0.50</td>
</tr>
<tr>
<td>Base Salary Share of Total Pay</td>
<td>0.68</td>
<td>0.16</td>
</tr>
<tr>
<td>Training Expenses (10,000 Yuan)</td>
<td>8.80</td>
<td>32.70</td>
</tr>
<tr>
<td>New Cooperation Partners</td>
<td>40.91</td>
<td>130.34</td>
</tr>
<tr>
<td>State Owned Enterprise</td>
<td>0.46</td>
<td>0.50</td>
</tr>
</tbody>
</table>
Table 2. Estimates of Hurdle Negative Binomial Model: Controlling for Unobserved Heterogeneity with Pre-Sample Average and Turnover Rate of Non-Technical Employees+

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>R&amp;D Projects (1) Coefficient</th>
<th>R&amp;D Projects (2) CME^1</th>
<th>New Commercial Products (3) Coefficient</th>
<th>New Commercial Products (4) CME^1</th>
<th>R&amp;D Projects (5) Coefficient</th>
<th>R&amp;D Projects (6) AME^2</th>
<th>New Commercial Products (7) Coefficient</th>
<th>New Commercial Products (8) AME^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technical Employee Turnover Rate</td>
<td>-0.029*</td>
<td>-0.680*</td>
<td>-0.036**</td>
<td>-0.793*</td>
<td>3.014</td>
<td>0.017</td>
<td>0.812**</td>
<td>0.007**</td>
</tr>
<tr>
<td>Number of HRM Practices</td>
<td>0.137**</td>
<td>3.249**</td>
<td>0.113*</td>
<td>2.506*</td>
<td>-1.030</td>
<td>-0.006</td>
<td>1.229</td>
<td>0.011</td>
</tr>
<tr>
<td>Formal Channel for Employee Suggestions</td>
<td>-0.231*</td>
<td>-5.479</td>
<td>-0.037</td>
<td>-0.819</td>
<td>4.402</td>
<td>0.025</td>
<td>-1.753</td>
<td>-0.015</td>
</tr>
<tr>
<td>Base Salary Share of Total Pay</td>
<td>-0.304</td>
<td>-7.189</td>
<td>0.178</td>
<td>3.950</td>
<td>4.296</td>
<td>0.025</td>
<td>-12.476*</td>
<td>-0.107*</td>
</tr>
<tr>
<td>Training Expenses</td>
<td>-0.002</td>
<td>-0.051</td>
<td>-0.004***</td>
<td>-0.091**</td>
<td>0.010</td>
<td>0.000</td>
<td>0.014</td>
<td>0.000</td>
</tr>
<tr>
<td>Log Number of New Cooperating Partners</td>
<td>0.110**</td>
<td>2.595**</td>
<td>0.156***</td>
<td>3.461***</td>
<td>-0.437</td>
<td>-0.003</td>
<td>3.070**</td>
<td>0.026**</td>
</tr>
<tr>
<td>Pre-Sample Average of Dependent Variable</td>
<td>0.006***</td>
<td>0.143***</td>
<td>0.009***</td>
<td>0.205***</td>
<td>4.882***</td>
<td>0.028***</td>
<td>12.185***</td>
<td>0.104***</td>
</tr>
<tr>
<td>Non-Tech Employee Turnover Rate</td>
<td>0.026</td>
<td>0.609</td>
<td>0.037</td>
<td>0.822</td>
<td>0.055</td>
<td>0.000</td>
<td>-6.475**</td>
<td>-0.055**</td>
</tr>
<tr>
<td>Log Investments in Innovation</td>
<td>0.057</td>
<td>1.346</td>
<td>-0.036</td>
<td>-0.798</td>
<td>0.058</td>
<td>0.000</td>
<td>-1.626*</td>
<td>-0.014**</td>
</tr>
<tr>
<td>Profits to Asset Ratio</td>
<td>0.670*</td>
<td>15.869*</td>
<td>-0.108</td>
<td>-2.407</td>
<td>-16.260</td>
<td>-0.094</td>
<td>5.155</td>
<td>0.044</td>
</tr>
<tr>
<td>Log Number of Technical Employees</td>
<td>-0.091</td>
<td>-2.164</td>
<td>-0.102*</td>
<td>-2.280</td>
<td>2.050</td>
<td>0.012</td>
<td>5.316**</td>
<td>0.045**</td>
</tr>
<tr>
<td>Log Number of Employees</td>
<td>0.251***</td>
<td>5.947***</td>
<td>0.245***</td>
<td>5.450***</td>
<td>-0.477</td>
<td>-0.003</td>
<td>-5.309**</td>
<td>-0.045**</td>
</tr>
<tr>
<td>Log Output</td>
<td>-0.015</td>
<td>-0.343</td>
<td>0.040</td>
<td>0.889</td>
<td>-1.064</td>
<td>-0.006</td>
<td>1.462*</td>
<td>0.013*</td>
</tr>
<tr>
<td>State Owned Enterprise</td>
<td>-0.076</td>
<td>-1.792</td>
<td>0.047</td>
<td>1.043</td>
<td>1.970</td>
<td>0.011</td>
<td>-1.153</td>
<td>-0.010</td>
</tr>
<tr>
<td>Foreign Owned Enterprise</td>
<td>-0.287</td>
<td>-6.798</td>
<td>0.231</td>
<td>5.139</td>
<td>---</td>
<td>---</td>
<td>-6.245</td>
<td>0.007**</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>481</td>
<td>399</td>
<td>565</td>
<td>564</td>
<td>565</td>
<td>564</td>
<td>565</td>
<td>564</td>
</tr>
<tr>
<td>LR (\chi^2)</td>
<td>545.76</td>
<td>488.70</td>
<td>451.35</td>
<td>648.66</td>
<td>451.35</td>
<td>648.66</td>
<td>451.35</td>
<td>648.66</td>
</tr>
<tr>
<td>Prob &gt; c(\chi^2)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Notes: 1. Conditional Marginal Effect, conditional on Median Values; 2. Average Marginal Effect
***: Significant at 1%; **: significant at 5%; *: significant at 10%.

+: The model included industry dummies which are not shown. They are jointly significantly different from zero in both intensity equations and in probability equation for new commercial products. The alpha is a test statistic for testing between the negative binomial model and the Poisson model.
Table 3. Estimates of Hurdle Negative Binomial Models with Individual HRM Practices*

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Intensity Equation</th>
<th>Probability Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R&amp;D Projects</td>
<td>New Commercial Products</td>
</tr>
<tr>
<td></td>
<td>Coeff.</td>
<td>CME(^1)</td>
</tr>
<tr>
<td>Technical Employee Turnover Rate</td>
<td>-0.028*</td>
<td>-0.735</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.461)</td>
</tr>
<tr>
<td>Training programs</td>
<td>0.232*</td>
<td>6.107*</td>
</tr>
<tr>
<td></td>
<td>(0.121)</td>
<td>(3.365)</td>
</tr>
<tr>
<td>Job Rotation Schemes</td>
<td>0.149</td>
<td>3.913</td>
</tr>
<tr>
<td></td>
<td>(0.131)</td>
<td>(3.530)</td>
</tr>
<tr>
<td>Cross Functional Teams</td>
<td>0.076</td>
<td>2.009</td>
</tr>
<tr>
<td></td>
<td>(0.120)</td>
<td>(3.313)</td>
</tr>
<tr>
<td>Formal Channel for Employee Suggestions</td>
<td>-0.234</td>
<td>-6.169</td>
</tr>
<tr>
<td></td>
<td>(0.145)</td>
<td>(4.040)</td>
</tr>
<tr>
<td>Job Description Manuals</td>
<td>0.154</td>
<td>4.066</td>
</tr>
<tr>
<td></td>
<td>(0.161)</td>
<td>(4.317)</td>
</tr>
<tr>
<td>Base Salary Share of Total Pay</td>
<td>-0.353</td>
<td>-9.285</td>
</tr>
<tr>
<td></td>
<td>(0.360)</td>
<td>(9.320)</td>
</tr>
<tr>
<td>Training Expenses</td>
<td>-0.002</td>
<td>-0.056</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.042)</td>
</tr>
<tr>
<td>Log (Number of New Cooperating Partners)</td>
<td>0.105*</td>
<td>2.760*</td>
</tr>
<tr>
<td></td>
<td>(0.058)</td>
<td>(1.516)</td>
</tr>
<tr>
<td>Pre-Sample Average of Dependent Variable</td>
<td>0.006***</td>
<td>0.160***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.024)</td>
</tr>
<tr>
<td>Non-Tech Employee Turnover Rate</td>
<td>0.022</td>
<td>0.572</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.797)</td>
</tr>
<tr>
<td>Log (Investment in Innovation)</td>
<td>0.054</td>
<td>1.428</td>
</tr>
<tr>
<td></td>
<td>(0.056)</td>
<td>(1.487)</td>
</tr>
<tr>
<td>Profits to Assets Ratio</td>
<td>0.656*</td>
<td>17.261</td>
</tr>
<tr>
<td></td>
<td>(0.398)</td>
<td>(10.637)</td>
</tr>
<tr>
<td>Log Number of Technical Employees</td>
<td>-0.090</td>
<td>-2.359</td>
</tr>
<tr>
<td></td>
<td>(0.072)</td>
<td>(1.896)</td>
</tr>
<tr>
<td>Log Number of Employees</td>
<td>0.245***</td>
<td>6.457***</td>
</tr>
<tr>
<td></td>
<td>(0.073)</td>
<td>(2.050)</td>
</tr>
</tbody>
</table>
The model included industry dummies which are not shown. They are jointly significantly different from zero in both intensity equations and in the probability equation for commercialized new products. The alpha is a statistic testing between the negative binomial model and the Poisson model.

<table>
<thead>
<tr>
<th>Model</th>
<th>Log Output (SE)</th>
<th>State Owned Enterprise (SE)</th>
<th>Foreign Owned Enterprise (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log Output</td>
<td>-0.021 (0.061)</td>
<td>-0.079 (0.091)</td>
<td>-0.276 (0.254)</td>
</tr>
<tr>
<td></td>
<td>-0.547 (1.610)</td>
<td>-2.073 (2.557)</td>
<td>-7.262 (6.868)</td>
</tr>
<tr>
<td></td>
<td>0.008 (0.060)</td>
<td>0.050 (0.090)</td>
<td>0.237 (0.256)</td>
</tr>
<tr>
<td></td>
<td>0.183 (1.335)</td>
<td>1.129 (1.953)</td>
<td>5.310 (5.736)</td>
</tr>
<tr>
<td></td>
<td>-1.042 (1.296)</td>
<td>1.009 (2.201)</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>-0.006 (0.007)</td>
<td>0.006 (0.012)</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>2.246** (1.103)</td>
<td>-1.873 (1.580)</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td>0.017** (0.007)</td>
<td>-0.014</td>
<td>---</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-2225.56</td>
<td>-1590.91</td>
<td>-10.91</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>481</td>
<td>399</td>
<td>565</td>
</tr>
<tr>
<td>LR (χ²)</td>
<td>547.53</td>
<td>496.53</td>
<td>453.24</td>
</tr>
<tr>
<td>Prob &gt; (χ²)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Pseudo R²</td>
<td>0.110</td>
<td>0.135</td>
<td>0.954</td>
</tr>
<tr>
<td>alpha(χ²)</td>
<td>0.704***</td>
<td>0.538***</td>
<td>---</td>
</tr>
</tbody>
</table>

Notes: 1. Conditional Marginal Effect, conditional on Median Values; 2. Average Marginal Effect
***: Significant at 1%; **: significant at 5%; *: significant at 10%.
Table 4. Marginal Effects Across Firms with Different Number of High Performance HRM Practices

<table>
<thead>
<tr>
<th># HRM Practices</th>
<th>Number of R&amp;D Projects</th>
<th></th>
<th>Number of New Commercial Products</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Turnover Rate of Technical Employees</td>
<td>-0.488* (0.295)</td>
<td>-0.574* (0.347)</td>
<td>-1.101* (0.661)</td>
<td>-0.691* (0.418)</td>
</tr>
<tr>
<td>Number of HRM Practices</td>
<td>2.334*** (0.755)</td>
<td>2.74*** (1.058)</td>
<td>5.262** (2.342)</td>
<td>3.303** (1.410)</td>
</tr>
<tr>
<td>Log (Number of New Cooperating Partners)</td>
<td>1.865* (0.977)</td>
<td>2.194* (1.127)</td>
<td>4.204* (2.227)</td>
<td>2.639** (1.348)</td>
</tr>
</tbody>
</table>

***: Significant at 1%; **: significant at 5%; *: significant at 10%.
Chapter 2

More outsourced R&D, fewer internal specialists?
More outsourced R&D, fewer internal specialists?

Wenjing Wang
Aarhus University
Sino-Danish Center for Education and Research

ABSTRACT
Does the firm employ fewer R&D specialists as it outsources more R&D? Although this question has many important implications, it has not been systematically examined by previous studies. This paper posits that the employment implication of firm’s R&D outsourcing activity varies across two distinct dimensions: outsourcing depth and outsourcing breadth, which reflect, respectively, how deeply and widely the firm leverages on external knowledge. Through estimating Correlated Random Effects (CRE) Tobit, CRE Selection, and CRE Fractional Response Models on a panel dataset of Danish firms, this paper finds that both the absolute number and the employment intensity of R&D specialists decrease as R&D outsourcing deepens but increase as it broadens.

Keywords: Correlated Random Effect models, employment of R&D specialists, R&D outsourcing breadth and depth, R&D strategy

JEL Code: J21, M51, O32

Acknowledgements: Sincere appreciation goes to Tor Eriksson and Frederic Warzynski for providing data, encouragement and valuable guidance; I thank Zhihua Qin, Keld Laursen, Ulrich Kaiser, Alex Pedrosa and Joonmo Ahn for inspiring comments and suggestions. This paper has also received helpful inputs from participants at Druid Academy 2014, IAB-AU workshop 2014, the Annual Conference of Academy of Innovation and Entrepreneurship 2014. Financial support from Sino-Danish Center for Education and Research is gratefully acknowledged.
1. INTRODUCTION

Aggregate-level studies suggest that outsourcing to (or trade with) low-cost foreign countries may cause employment polarization and skill upgrading in local labor market (see e.g., Acemoglu and Autor, 2011; Autor et al., 2013). At firm-level, only a few studies on the association between outsourcing and personnel upgrading exist (see e.g., Biscourp and Kramarz, 2007; Eriksson, 2010). But it is still unknown whether R&D outsourcing crowds out or increases internal employment of R&D specialists - a key indicator for the quality of R&D personnel. Tracking firm’s employment of R&D specialists contributes better understanding of “what has happened to the firm’s core competency in the wave of open R&D and innovation”, which is raised by Christensen (2008) as one of the most important questions in open R&D and innovation research. The employment of R&D specialists is also related to the strategic dimension of firm’s openness decisions, the investigation of which has been called for in the recent open innovation literature (Chesbrough, 2006; Dahlander and Gann, 2010; Laursen and Salter, 2014). Nevertheless, existing literature on this topic is sparse.

Related empirical studies have examined (1) the effects of R&D outsourcing on innovation performance (e.g., Grimpe and Kaiser, 2010), (2) whether internal R&D and external R&D are complements or substitutes (e.g., Cassiman and Veugelers, 2006; Hagedoorn and Wang, 2012), and (3) the contextual factors determining the effect of external knowledge, such as organizational structure (Arora, et al. 2014), firm’s previous experience and absorptive capacity (e.g., Cohen and Levinthal, 1990; Ceccagnoli et al., 2014), appropriability (Laursen
and Salter, 2014) and regions (Laursen et al., 2012). However, little attention has been paid to the implication of R&D outsourcing on R&D personnel. An insightful study suggests high technical skills are complementary with R&D collaboration (Leiponen, 2005). Nevertheless, R&D outsourcing is not included in her analysis.

To our knowledge, only one closely related empirical study based on a relatively large dataset exists. The study finds that firm’s internal R&D employment intensity decreases when it decides to start, to increase, or to stop R&D outsourcing (Teirlinck et al., 2010). However, they focus on overall R&D employment rather than the composition of employment within the R&D function - especially the employment of R&D specialists. Thus it does not inform us about the quality aspect of R&D personnel. As for the employment of R&D specialists, existing evidence is based on case studies. For example, one pioneering study finds that, in several firms that adopt open innovation, the role of senior scientists is undermined, whereas the role of engineers and the business innovation team has been highlighted (Petroni et al., 2012). One aim of this paper is to examine whether the pattern observed by Petroni et al. (2012) is common in a broader range of firms.

This paper helps fill the gap by examining the question “How does firm’s employment of R&D specialists changes with R&D outsourcing?” using longitudinal data and methods. The number of R&D specialists is measured by the number of full-time-equivalent (FTE) researchers and other specialists who work on R&D within a firm. Inspired by Laursen and Salter (2006), R&D outsourcing is measured on two dimensions – breadth and depth. R&D outsourcing depth is measured by the expenditure on purchased R&D divided by total R&D
expenditure. R&D outsourcing breadth mainly refers to the diversity of R&D outsourcing partners, which is measured by the number of types of external agents from which the firm purchases R&D.

While the depth of R&D outsourcing reflects the degree of reliance on external sources (which is usually the focus of previous literature), the breadth of R&D outsourcing captures the variety of external R&D sources utilized by the firm. While the depth mirrors specialization and exploitation, the breadth reflects diversification and exploration. This two-dimensional measurement helps describe how deeply and widely the firm leverages on external sources for R&D, and it also allows observation of how employment of R&D specialists may change differently on these two distinct dimensions of R&D outsourcing.

Related theories such as the diseconomies of scope, the resource-based view, the knowledge based view, and especially the absorptive capacity suggest that both complementary and substitution relationships may link firm’s employment of R&D specialists to R&D outsourcing. The overall direction of the association is likely to depend on why and how the firm outsources R&D, which can be inferred by the changes in outsourcing breadth and depth. Further analysis suggests that firms employ more and a higher proportion of R&D specialists when R&D outsourcing broadens. The opposite is true when R&D outsourcing deepens.

Next, the above predictions are examined empirically, using panel data from Statistic Denmark’s annual survey on R&D and innovation activities in over 2200 firms during the period 2007-2010. Three types of econometric models are used - Correlated Random Effects
(CRE) Tobit models, CRE Selection models, and CRE Fractional Response Models. The former two types of models explain the absolute number of R&D specialists, while the third type explains the share of specialists within the R&D function. The empirical results support the above-mentioned theoretical predictions.

The remainder of this paper unfolds as follows. Section 2 presents the related previous studies and theoretical analysis, based on which two hypotheses are proposed. Section 3 describes the empirical strategy for testing the hypotheses. Section 4 presents the estimation results. Section 5 discusses the implications and then concludes.

2. THEORY AND HYPOTHESES

2.1. Firm’s employment of R&D specialists and R&D outsourcing

According to knowledge based view, the primary task of the firm is to establish the coordination necessary for integrating knowledge resident in individuals (Kogut and Zander, 1992; Grant, 1996). This indicates that tasks of the firm change with the knowledge to be integrated. When innovating through assembling new components made by suppliers, firms have to perform additional tasks of experimentation with new system architectures and coordination the work of suppliers. These additional tasks require the firm to broaden in-house capabilities while keeping a cognitive overlap with suppliers (Brusoni et al, 2001). Similarly, when purchasing R&D from a new type of external partner, additional internal tasks are required for coordinating and integrating the newly purchased R&D. At the same time, some old tasks of integrating many small pieces of knowledge residing in internal R&D
personnel are replaced by the new tasks of integrating a few bigger chunks of knowledge residing in purchased R&D. Therefore, as the firm outsources R&D work, the tasks and corresponding internally required knowledge and skills have changed. Because each individual is endowed with specific knowledge and skills suitable for specific tasks, the change in internal tasks may lead to change in employment (Acemoglu and Autor, 2011). Consequently, the internal amount and composition of R&D personnel, including R&D specialists, is subject to change with R&D outsourcing. In addition, the degree of the change in internal R&D employment is probably associated with the distance and innovativeness of the newly outsourced R&D.

Then, how could employment of R&D specialists be associated with R&D outsourcing?

On one hand, more outsourced R&D may take additional R&D specialists to integrate it into a firm’s own knowledge base. In this sense, outsourced R&D and internal R&D specialists are complementary and thus positively associated with each other. According to resource based view (RBV) and knowledge based view (KBV), R&D outsourcing may (1) provide client firms with access to resources that are not available internally (Grimpe and Kaiser, 2010; Lavie, 2006; Pernrose, 1959; Weigelt, 2009) and (2) broaden the existing knowledge base (Berchicci, 2013; Kogut and Zander, 1992). In order to recognize, assimilate, and apply new information to commercial ends, firms need absorptive capacity (AC) (Cohen and Levinthal, 1990). Because absorbing new external information requires a certain depth of internal knowledge (Cohen and Levinthal, 1990), which is endowed by R&D specialists, one major way of strengthening AC is to employ more R&D specialists. This is in line with previous research
suggesting that transferring knowledge is more effective if accompanied by relocating employees (Argote and Ingram, 2000). Accordingly, as AC complements with external knowledge, the employment of individual R&D specialist should complement and evolve in the same way with external knowledge acquired by outsourcing R&D. In sum, the combination of related theories of AC, KBV, and RBV indicates that firms need to strengthen internal capacity and especially the employment of R&D specialists to complement extended R&D outsourcing.

On the other hand, the additional outsourced R&D may substitute the R&D work previously performed by internal R&D specialists. Thus the employment of these may decrease with R&D outsourcing. This relates to the concern that outsourcing may hollow corporations and their capabilities (Bettis et al., 1992; Weigelt, 2009). From this perspective, outsourced R&D and internal R&D specialists are negatively associated with each other. The substitutive relationship may arise from similar roots of the substitution relationship between internal and external R&D. For example, external knowledge may reduce resources and incentives for internal research (Hitt et al., 1990; Arora et al., 2014). Especially, in the presence of diseconomies of scope (Hess and Rothaermel, 2011), firms can reduce cost by replacing internal non-core R&D with outsourced R&D, the practice of which may reduce the demand of internally employed R&D personnel - including specialists. In addition, the availability of cheaper external knowledge may make it more profitable for a firm to switch from knowledge production towards knowledge application. Unlike knowledge production that needs deep insight endowed with R&D specialists, knowledge application may require
shallower but more practical knowledge. This may lead to a decreased demand for R&D specialist relative to R&D support staff. This kind of evolvement within R&D personnel has been suggested by several case studies (e.g. Petroni et al., 2012).

2.2. Two distinct dimensions of R&D outsourcing - breadth and depth

Inspired by Laursen and Salter (2006), which has developed the concepts of breadth and depth as two components of the openness of individual firms’ external search strategies, this paper also uses two dimensions to describe the way through which the firm outsources R&D – the breadth and depth. The breadth measures the variety of R&D outsourcing sources; the depth measures the quantity of outsourced R&D relative to internal R&D. Together, they reflect how broadly and deeply a firm leverages on outsourced R&D.

Changes in R&D outsourcing along these two distinct dimensions usually originate from different purposes, require different antecedents and complementary adjustments, and have different consequences. To some extent, deepening and broadening of R&D outsourcing is parallel to the twin concepts of exploitation and exploration in March (1991). Deepening of R&D outsourcing is like exploitation, which is usually motivated by efficiency and current performance; while broadening R&D outsourcing is more like exploration, which is motivated by new knowledge, experimentation and future return. While the former emphasizes exploiting existing external R&D sources, the later emphasizes exploring new alternatives. Therefore, compared with deepening R&D outsourcing, broadening R&D outsourcing is likely to bring in more new knowledge, which requires more internal efforts to
be integrated into firm’s original knowledge base. Due to the new alternatives and experimentation, broadening R&D outsourcing is also associated with more uncertainty.

2.3. Diverging implications on employment of R&D specialists

Because of the different antecedents and characteristics discussed above, deepening and broadening of R&D outsourcing may have diverging implications on employment of R&D specialists in both absolute and relative terms. According to Brusoni et al. (2001), assembler firms need more additional in-house capabilities when the potential imbalances in component performance and unpredictability of product-level interdependencies are higher. Applying this insight to the context of R&D outsourcing, it indicates that broadening R&D outsourcing requires more supporting in-house R&D capacities than deepening R&D outsourcing. This is because broadening R&D outsourcing means involving more distant external agents in terms of location and industry, who are not necessarily to coordinate with each other. As a result, the imbalance of performance is more likely to occur and the interdependencies are more difficult to predict. Therefore, it takes more internal capabilities to manage and coordinate the broadening of R&D outsourcing than the deepening of R&D outsourcing. The following will discuss in how employment of R&D specialists changes differently with R&D outsourcing breadth and depth.

2.3.1. R&D outsourcing breadth and employment of R&D specialists

R&D outsourcing broadens when the firm purchases R&D from a new type of external contractor. Because purchasing from a new distant source is always associated with higher
transaction costs and higher uncertainty than purchasing from existing sources, a premise of broadening R&D outsourcing decision is that the firm expects higher return from this distant outsourcing than from existing outsourcing. In order to realize the higher external return from leveraging on more distant external knowledge, firms have to build corresponding stronger internal AC. Because AC mainly depends on individuals (gate keepers) who work at the interface of the firm and the external environment and “breadth of knowledge cannot be superficial to be effective” (Cohen and Levinthal, 1990, pp 135), firms need more specialists with deeper knowledge to work at the extended interface as the new “gate keepers” and leverage the broader knowledge more effectively. Relative to specialists, other supporting employees at R&D department are less relevant to AC. So their employment may not increase as much as that of the R&D specialists in this scenario. Hence, along the breadth dimension, the previously discussed complementary association between R&D specialists and R&D outsourcing is likely to dominate - especially when the outsourcing depth stays the same.

An alternative argument based on personnel economics also reaches to the same prediction. According to personnel economics theory, for high risk and high return tasks, the best strategy is to hire employees with high ability. In the context of R&D work, which is carried out by R&D specialists, technicians and other supporting staffs, high-ability employees corresponds to R&D specialists. As discussed before, the work of integrating the knowledge from broadening R&D sourcing contains more uncertainty and higher expected return.
Therefore, to do the additional work of integrating the knowledge from the new R&D outsourcing partners, the best strategy is to hire more R&D specialists.

The above discussions all suggest the following hypothesis:

**Hypothesis 1**: As R&D outsourcing broadens, the firm employs more and a higher proportion of R&D specialists.

### 2.3.2. R&D outsourcing depth and employment of R&D specialists

A deepening of R&D outsourcing implies that a higher proportion of R&D is purchased from external sources. Under budget constraint of total R&D investment, deepening (without broadening) R&D outsourcing implies reducing internal R&D investment while increasing outsourcing R&D to existing external channels. Because existing outsourcing channels are endowed with less new knowledge compared to new channels, outsourcing more R&D to existing channels (i.e. deepening R&D outsourcing) is more likely to be motivated by reducing cost or narrowing down internal R&D activities (i.e. focusing on specific areas) than by acquiring new knowledge. In this scenario, firms typically replace R&D work previously performed internally with the cheaper alternative offered by existing R&D partners. Hence, a deepening of R&D outsourcing usually indicates substituting internal R&D and corresponding employment of R&D specialists. Meanwhile, because deepening R&D outsourcing is mainly through exploiting existing channels vertically, it does not require an extension of the interface with the external environment horizontally. This limited extension of the interface implies limited additional demand for absorptive capacity and R&D specialists. Thus the
complementary effect discussed in the previous section is rather small. Hence, as R&D outsourcing deepens, the substation effects are likely to be higher than the complementary effects. Consequently, internal R&D tends to decline.

For similar reasons, the employment of supporting staff within R&D functions reduces with deepening R&D outsourcing, but it is less sensitive:

On one hand, increasing purchase from existing external R&D partners may still need complementary internal R&D employees to manage the additional transaction. In other words, the complementary effects still exist - despite being smaller than the negative substitution effects. In addition, because deepening R&D outsourcing involves little extra knowledge integration at which R&D specialists are better, supporting R&D staff is more cost-effective than specialists in this scenario. Hence the positive complementary effects of deepening R&D outsourcing are likely to be larger for supporting R&D staff than for specialists.

On the other hand, the negative substitution effects are likely to be no smaller for R&D specialists than for supporting R&D employees, because the newly outsourced R&D work is likely to constitute the same or more work from R&D specialists, compared with the R&D work kept internally. This is because the aim of cost minimization requires the firm to outsource the more expensive part of R&D work until the average cost of internal R&D equals external R&D. Because capital cost is relatively unchanged in the short run, the higher cost of R&D must be due to higher labor cost. As R&D specialists require higher salary, it is
likely that more expensive R&D work constitutes a higher portion of the work from R&D specialists. Hence, the negative substitution effects are likely to be larger for R&D specialists. Consequently, the overall negative employment effects are larger for R&D specialists than for supporting staff, which means a decrease in the proportion of R&D specialists employed within a firm’s R&D function.

The above analyses suggest the following hypothesis:

**Hypothesis 2**: As R&D outsourcing deepens, the firm employs fewer and a lower proportion of R&D specialists.

### 3. EMPIRICAL ANALYSIS

#### 3.1. Data

The dataset is constructed by merging survey data on firms’ R&D and Innovation (FoU) activities with basic firm characteristics (FIRE). FoU survey is conducted annually by Statistics Denmark since 1990s. Considering the availability and consistency of the variables of interest, this paper uses only the surveys conducted during 2007-2010. Each year’s survey contains around 4000 firms. However, only a proportion of these have R&D related activity. FIRE data provides basic information of the firm, such as location, industry, total number of employees, profits, etc. Only firms that appear in both datasets are used. For the purpose of this analysis, the sample (the population of interest) is further restricted to firms with positive R&D expenditure. Because a firm may not participate in the survey or have positive R&D
expenditures every year, the panel data are unbalanced. In total, the dataset contains 3973 observations from 2285 different firms.

3.2. Variables

3.2.1. Dependent Variables

The FoU survey categorizes R&D personnel into three types: R&D specialists, technicians and supporting staffs. The number of R&D specialists is measured by the number of full-time-equivalent (FTE) researchers and other specialists who work on R&D within a firm, which is a direct measure of the employment opportunity for R&D specialists.

The share of R&D specialists is measured by the number of R&D specialists divided by the number of all the employees within the firm’s R&D function, which include R&D specialists, technicians and supporting staffs.

While the absolute number of R&D specialists embodies the capacity of internal R&D, the share of R&D specialists emphasizes the quality aspect of R&D personnel. Together, these two dependent variables reflect R&D competency and absorptive capacity from the facet of labor input.

3.2.2. Main Explanatory Variables

R&D outsourcing activity is measured by two dimensions - depth and breadth.

R&D outsourcing depth is measured by the expenditure on purchased R&D divided by total R&D expenditure.
R&D outsourcing breadth mainly refers to the diversity of R&D outsourcing partners, which is measured by the number of types of external agents from which the firm purchases R&D. In the survey, there is information about expenditures on purchased R&D from each of the following eight types of mutually exclusive external agents: firms from the same business category in Denmark, other firms in Denmark, firms from the same business category abroad, approved technological service institutes (ATS) in Denmark, universities and colleges in Denmark, other public research institutions in Denmark, other firms abroad, and other public research institutions abroad. Based on this information, eight binary variables, each indicating whether a firm has purchased R&D from a certain category, are generated, and then a variable counting the total types of external R&D purchasing partners is constructed to measure the breadth of R&D outsourcing.

3.2.3. Control Variables

Several factors that may influence both R&D specialists’ employment and R&D outsourcing breadth/depth are controlled for:

Total R&D expenditure. Recent research shows that firms with higher R&D expenditure are more likely to engage in open innovation (e.g. Mina A. et al., 2014). At the same time, breadth and depth of R&D outsourcing may also relate to open innovation. Thus, R&D expenditure may correlate with R&D outsourcing breadth/depth through open innovation strategy. Meanwhile, R&D expenditure captures the scale of R&D, which directly links to R&D specialists’ employment. Thus both R&D outsourcing breadth/depth and R&D
specialists’ employment may relate to total R&D expenditure, which must be controlled for to avoid omission bias.

For similar reasons, it is also necessary to control for:

Firm size. This is captured by the value of asset and the number of full-time-equivalent employees, which may relate to both R&D employment and outsourcing. Log values are used in the models.

R&D department. This is a binary variable indicating whether a firm has an R&D department or not. It reflects the degree of importance a firm attaches to R&D activities, which may relate to both the internal employment of R&D specialists and R&D outsourcing.

Profit per employee. This is a proxy capturing a group of unobservable factors that may influence both the capability and efficiency of hiring R&D specialists and R&D outsourcing depth and breadth.

Industry. Previous research has pointed out that persistent differences across industries, especially in terms of technological opportunities and social institutions, result in differences in collective invention (Powell and Ginnella, 2010). As collective inventions may relate to internal employment of R&D specialists and R&D outsourcing activities, they may link back to the industry differences, which are thus necessary to control. To balance between precise industry classification and consumption of degree of freedom, the first digit of (NACE) industry classification is used, which classifies the firms into seven different industries.
Location. Differences in social institutions and labor supply may influence firm’s choice on R&D employment and outsourcing. These differences are controlled by a location indicator. In total, the sample covers eight different locations.

### 3.3. Descriptive Statistics

Table I provides a brief descriptive statistics of the major variables. Among the firms investing in R&D, 75% hire R&D specialists. Compared with firms without R&D specialists, firms with R&D specialists are on average larger (in terms of asset and employment scale), enjoy more profit per employee, invest more in R&D, and outsource a smaller proportion of it but to a broader range of external agents. In the sample, the average outsourcing depth is 15%, while the outsourcing breadth is 0.74. On average, a firm employs fifteen R&D specialists, who account for 55% of R&D employment.

Table II provides a more detailed picture of R&D outsourcing depth and breadth, summarized across quintiles. Among the firms that outsource R&D, the average outsourcing depth in the medium quintile is 23%, and the average outsourcing breadth is two. Separately examining each quintile ladder of outsourcing breadth or depth, we can see that the average depth or breadth roughly doubles when moving to the next quintile. Examination on the quintiles between outsourcing breadth and depth reveals that outsourcing breadth evolves
along an inverted-U shape curve as outsourcing deepens: firms with a medium level of outsourcing depth have two types of outsourcing partners on average. This level is around 30% broader than that of the firms with outsourcing depth at the 1st quintile (1.52) or the 5th quintile (1.32).

3.4. Econometric Models

To identify the impact on employment of R&D specialists, a Correlated Random Effect (CRE) Tobit model is estimated. The estimates are compared with estimates from three sample selection models, in which the second stage estimation uses fixed effects (FE), CRE, and pooled OLS specification respectively. Then CRE Fractional Response Model is used to analyze the impacts on the share of R&D specialists.

The CRE specification allows for correlation between unobserved heterogeneity and independent variables, and it is a more reliable estimation than Random Effect (RE) models (e.g. RE Tobit model) which are often found in previous studies. For studies on R&D and innovation activity, where the unobserved heterogeneity and independent variables are very likely to be correlated, the advantage of the CRE framework becomes more significant.

The comparison between CRE Tobit estimations and CRE sample selection estimations is also relatively new to the literature. CRE sample selection models place even fewer restrictions
and serve reality better than the popular models in existing literature, such as RE sample selection models or the hurdle models. Besides the advantage of CRE device, the CRE sample selection models have additional advantages: they allow not only (1) the process of deciding whether or not to hire R&D specialists to differ from the process deciding how many R&D specialists to hire, but also (2) the correlation between these two processes.

In addition, the CRE Fractional Response Model makes use of the fractional nature of the dependent variable - the share of the R&D specialists, so that the estimators and predictions fit the real situation better. This improvement is comparable to the advance from the linear model to the Probit or the Logit model for binary response variables.

In sum, these recently developed models are good candidates for empirical studies on R&D and innovation activity, thus enhancing the validity of this study.

3.4.1. CRE Tobit Model

The Tobit model allowing for unobserved heterogeneity assumes an underlying equation determining the employment of R&D specialists as (3-1):

\[
y_{it}^{*} = x_{it}\beta + c_i + u_{it}
\]

\[
y_{it} = \begin{cases} y_{it}^{*}, & \text{if } and y_{it}^{*} > 0 \\ 0, & \text{if } and y_{it}^{*} \leq 0 \end{cases}
\]

, where \(y_{it}\) and \(y_{it}^{*}\) are latent and observed outcome variable respectively, \(x_{it}\) is a vector of explanatory variables, \(c_i\) is firm specific unobserved heterogeneity, and \(u_{it}\) is an idiosyncratic error.
The CRE approach, which dates back to Mundlak (1978), allows correlation between \( c_i \) and \( x_i \), thus loosening the assumption for the traditional Random Effect (RE) method and making RE a special case of CRE. Following Wooldridge (2010), the conditional distribution of heterogeneity is modeled as:

\[
c_i | x_i \sim Normal (\psi + \bar{x}_i \xi, \sigma_a^2).
\]

(3-3)

Given (3-3), equation (3-1) and (3-2) can be summarized as:

\[
y_{it} = \max(0, \psi + x_{it} \beta + \bar{x}_t \xi + a_i + u_{it})
\]

(3-4)

, which assumes \( a_i | x_i \sim Normal (0, \sigma_a^2) \) and \( u_{it} | x_i \sim Normal (0, \sigma_u^2) \) - so that (3-4) can be estimated by joint maximum likelihood estimation (conditional on \( x_i \)).

3.4.2. CRE Selection Models

Although CRE Tobit estimates are more reliable than traditional RE Tobit estimates, the CRE Tobit model may still be rather restrictive, because it assumes that the explanatory variables and the signs of marginal effects are the same between the two processes deciding whether or not to hire R&D specialists (participation decision) and how many R&D specialists to hire (intensity decision). To distinguish between the two processes, a group of previous literature makes use of hurdle models. Still, hurdle models are a special case of a more general group of models – selection models. To check whether the participation decision process differs from the intensity decision process, this paper makes use of CRE selection models, following Semykina and Wooldridge (2010). Generally, selection models also use equation (3-1) to describe the intensity decision, which captures the expectation of dependent variable,
conditioning on positive outcomes. In addition, it introduces a selection equation (3-5) to relax condition equation (3-2):

\[ s_{it} = 1 \left[ s_{it}^* > 0 \right] = \max \left[ x_{it2} \delta_t + c_{i2} + u_{it2} > 0 \right]. \]  

(3-5)

Then equation (3-2) becomes:

\[ y_{it} = \begin{cases} y_{it}^*, & \text{if and } s_{it} = 1 \\ 0, & \text{if and } s_{it} = 0 \end{cases} \]  

(3-6)

where \( s_{it} \) and \( s_{it}^* \) are observed and latent selection indicators respectively; \( x_{it2} \) is a vector of variables explaining participation; and \( c_{i2} \) is firm specific unobserved heterogeneity. Both \( x_{it2} \) and \( c_{i2} \) in equation (3-5) can be different from \( x_{it} \) and \( c_i \) in equation (3-1). In this way, the participation decision, which is captured by (3-5), is allowed to differ from the intensity decision, which is captured by (3-1). In this study, a binary variable indicating whether the firm has R&D department is included in \( x_{it2} \) (but not in \( x_{it} \)) as exclusion restriction to increase the reliability of the selection model. Equations (3-1), (3-5), and (3-6) form the basic framework for the selection models.

Following Semykina and Wooldridge (2010), the selection models are estimated using a two-step procedure incorporated with the CRE device from Mundlak (1978). The first step is to estimate the selection equation (3-5): \( c_{i2} \) is assumed to relate to \( x_{i2} \) only through its time averages \( \bar{x}_{i2} \), so that \( a_{i2} \) is independent of \( x_{i2} \):

\[ c_{i2} = \bar{x}_{i2} \pi + a_{i2} \]  

(3-7)

\[ a_{i2} | x_{i2} \sim \text{Normal} \left( 0, \sigma_{a2}^2 \right). \]  

(3-8)
Then equation (3-5) becomes:

\[ s_{it} = 1 \left[ s_{it}^* > 0 \right] = 1 \left[ x_{it2} \delta_t + \bar{x}_{i2} \pi_t + v_{it2} > 0 \right] \]  

(3-9)

where \( v_{i2} = a_{i2} + u_{it2} \) and \( v_{i2} | x_i \sim \text{Normal} \left( 0, 1 + \sigma_{a2}^2 \right) \). Then equation (3-9) can be estimated with a Probit model for each time period, and an inverse Mills ratio for each observation \( \hat{\lambda}_{it} \) can be obtained.

One advantage of the selection model is that it allows the correlation between the participation equation and the intensity equation through error terms and unobserved factors. The correlation between error terms is assumed to be linear:

\[ E(u_{it} | x_i, c_i, v_{it2}) = E(u_{it} | v_{it2}) = \rho_t v_{it2}, \quad t = 1, \ldots, T. \]  

(3-10)

Further assuming that

\[ E(c_i | x_i, v_{it2}) = \bar{x}_i \xi + \psi_t v_{it2} + a_{i1}. \]  

(3-11)

Taking expectation of (3-1) conditional on \( x_i, v_{it2} \) and replacing \( E(u_{it} | x_i, v_{it2}) \) and \( E(c_i | x_i, v_{it2}) \) with (3-10) and (3-11) gives:

\[ E(y_{it} | x_i, v_{it2}) = x_{it} \beta + \bar{x}_i \xi + \gamma_t v_{it2} + a_{i1}, \]  

(3-12)

where \( \gamma_t = \rho_t + \psi_t \).

Conditioning on \( s_{it} = 1 \), (3-12) becomes:

\[ E(y_{it} | x_i, s_{it} = 1) = x_{it} \beta + \bar{x}_i \xi + \gamma_t \lambda_{it} (x_{it2} \delta_t^a + \bar{x}_{i2} \pi_t^a) + a_{i1} \]

where \( \lambda_{it} (\cdot) \) is the inverse Mills ratio obtained by previous estimation of equation (3-9).

Thus, the equation for \( s_{it} = 1 \) is:
\[ y_{it} = x_{it} \beta + \bar{x}_i \xi + \gamma_t \lambda_{it} (x_{it2} \delta_t^a + \bar{z}_i \pi_t^a) + a_{i1} + \epsilon_{it1} \]  \hspace{1cm} (3-13)

where \( \delta_t^a = \delta_t / \sqrt{1 + \sigma_{a2}^2} \), \( \pi_t^a = \pi_t / \sqrt{1 + \sigma_{a2}^2} \).

The second step estimates the final equation (3-13) by using either FE, CRE, or pooled OLS after substituting \( \lambda_{it} \) by \( \hat{\lambda}_{it} \). The consistency of the estimators depends on different assumptions. The major differences among these three models are: for \( t = 1, 2, ..., T \), OLS requires \( E[x_{it}'(a_{i1} + \epsilon_{it1})] = 0 \), which practically means that \( E[x_{it}'\epsilon_{it1}] = 0 \) and \( E[x_{it}'a_{i1}] = 0 \); both CRE and FE require \( E[\epsilon_{it1}|x_i, a_{i1}] = 0 \); in addition, CRE requires that \( x_{it} \) and \( a_{i1} \) are not correlated: \( E[a_{i1}|x_i] = E(a_{i1}) = 0 \), while FE permits that (Semykina and Wooldridge, 2010).

Following Semykina and Wooldridge (2010), the standard error is obtained through bootstrap procedure.

### 3.4.3. CRE Fractional Response Model

The CRE Fractional Response Model is used to analyze the impacts of outsourcing on the percentage of specialists among all the R&D employees.

Following Papke and Wooldridge (2008), the fractional response \( y_{it} \) can be modeled with the following function:

\[ E(y_{it} | x_{it}, c_i) = G(x_{it} \beta + c_i), t = 1, 2, ..., T \]  \hspace{1cm} (3-14)
where \( G(\cdot) \) can have any function form as long as \( G(\cdot) \in (0,1) \) for \( y \) in \([0,1]\); \( x_{it} \) is a vector of explanatory variables, \( c_i \) is firm specific unobserved heterogeneity, and \( u_{it} \) is an idiosyncratic error term.

A CRE approach with a Chamberlain-Mundlak device allows for the correlation between \( c_i \) and \( x_i \), by further assuming that \( c_i = \psi + \bar{x}_i \xi + a_i \), where \( a_i | x_i \sim \text{Normal} (0, \sigma_a^2) \). Then (3-14) can be written as:

\[
E(y_{it} | x_{it}, a_i) = G(x_{it} \beta + \bar{x}_i \xi + a_i), t = 1, 2, ..., T. \tag{3-15}
\]

This paper assumes that \( G(\cdot) \) takes logit form and estimates (3-15) by a traditional Random Effect method. The average marginal effects are the same as in the Logit model, except that these are partial effects on a mean response (Wooldridge 2010). The standard errors are estimated via bootstrap procedure.

4. RESULTS

Table III reports the estimates from the Correlated Random Effect (CRE) Tobit model and CRE Selection models. Following Wooldridge (2010), CRE Selection models are estimated by a two-stage procedure. The first stage estimates the selection function (3-9) by the CRE Probit model, while the second stage estimates the intensity function (3-13) by fixed effects (FE), CRE, and pooled OLS. As discussed in section 3.4.2, these three second-stage methods require different assumptions on the association between independent variables, unobservable time-invariant, and time-variant individual characteristics. Furthermore, the choice among them faces a trade-off between efficiency and consistency: FE estimators tend
to be the most consistent but the least efficient; pooled OLS estimators can be the least consistent but the most efficient; CRE estimators locates somewhere in between. As it is unknown which assumption best serves the reality, the second-stage estimates from all the three methods are reported.

Although the econometric models are different, they all reveal a similar picture: the estimates are roughly stable across the four models, especially in terms of the signs and significant levels. The magnitudes of marginal effects exhibit some differences, but when they are interpreted together with the units of variables, the differences are actually small.

The breadth of outsourcing, which is measured by the number of types of R&D outsourcing partners, has a positive effect on R&D specialists’ employment. The estimates are significant in three models, and the magnitude of marginal effect goes from 2.81 to 3.35, depending on estimation method. On average, increasing one type of R&D outsourcing partner leads to about three more full-time-equivalent hires of R&D specialists, holding the other factors constant.

On the contrary, the depth of outsourcing, which is measured by the share of purchased R&D, has a negative effect on R&D specialists’ employment. Again, the estimates are significant in three models; the magnitude of marginal effect goes from -0.18 to -0.35,
meaning that a 5 percentage increase in the share of purchased R&D leads to one to two fewer full-time-equivalent hires of R&D specialists, depending on the estimation method.

In the selection models, the estimators associated with the selection indicator lambda are only significant at a 10 percentage level at the most, regardless of the second-stage estimation method. This indicates that only slight difference exists in the impacts of the examined factors on the two processes determining “hiring R&D specialists or not” and “how many R&D specialists to hire”. This also explains why the estimates are roughly stable across the CRE Tobit and CRE Selection models. In other words, the comparable estimates from the two different types of models, not only further confirm the robustness of the positive/negative association between R&D outsourcing breadth/depth and R&D specialists’ employment, but also indicate that the decisions “hire or not” and “how many to hire” are influenced by the examined factors in similar ways.

The analysis above shows how the absolute number of R&D specialists evolves with R&D outsourcing. Because labor input directly links to production scale, the absolute number of employment actually reflects the firm’s internal R&D ability (absorptive capacity) through the facet of quantity.

Another facet reflecting the firm’s internal R&D ability is employment quality. An important question under this strand is whether firms adjust their employment composition within the R&D function when resorting to external R&D resources. The answers can be inferred by looking at how the share of specialists within the R&D function changes with R&D outsourcing.
Table IV reports the estimates from the CRE fractional response Probit model for the share of specialists among total R&D employment.

INSERT TABLE IV ABOUT HERE

Generally speaking, the share of R&D specialists relates to the two dimensions of R&D outsourcing in parallel to the way the absolute number of R&D specialists relates to R&D outsourcing.

On one hand, a significant and positive association is found between the share of R&D specialists within R&D function and the breadth of R&D outsourcing. On average, establishing one extra type of R&D outsourcing partner is associated with an around 2.5 percent increase in the share of R&D specialists. Combining this with previous finding, it is confirmed that establishing more types of R&D outsourcing partner is associated with an upgrading of R&D personnel. Not only the absolute number but also the share of specialists within the R&D function increase, which corresponds to an upgrade in both capacity and capability of internal R&D, realizing a higher absorptive capacity.

On the other hand, a significant and negative association is found between the share of R&D specialists within the R&D function and the depth of R&D outsourcing. On average, one percent increase in the share of purchased R&D links to a reduction in the share of R&D specialists by around 0.16 percent. So in general, an increasing reliance on purchased R&D
undermines the role of internal R&D specialists – in terms of both absolute employment and relative intensity within R&D function, which embodies deterioration in R&D capacity and capability, implying a lower absorptive capacity.

As for control variables, several results are worth noticing. First, the scale indicators associate positively with the absolute number of R&D specialists but negatively with the share of R&D specialists: as firms acquire more assets or invest more in R&D, they employ more R&D specialists – which is not surprising. Yet, at the same time, they also tend to hire even more supporting staff (non-specialists) into R&D functions - which is somehow interesting. In addition, consistent with previous literature which finds that several aspects of innovation differ across industries, this paper also finds industrial differences in internal employment of R&D specialists, in terms of both absolute numbers and the shares within R&D function.

One major concern about the validity of the findings is that unobserved firm-level heterogeneity may weaken the association or bias our estimates. A possible source of such unobserved heterogeneity is firm quality, which may influence both employment of R&D specialists and R&D outsourcing decisions. High-quality firms that are able to orchestrate the outsourcing of R&D across many different partners may also find it profitable to employ more R&D specialists; conversely, low-quality firms that can only manage a limited range of R&D partners may find it optimal to hire fewer R&D specialists. Fortunately, this kind of heterogeneity such as firm quality is likely to persist across time (time-invariant), thus it is taken good care of by the CRE device or FE framework. CRE methods deal with the time-
invariant unobserved heterogeneity using mean values of time-variant independent variables as additional controls. In this way, the rest unobserved heterogeneity left in error term is unlikely to correlate with independent variables (see section 3). By design, FE framework allows for correlation between unobserved heterogeneity and independent variables, no matter the panel is balanced or not (Wooldridge, 2010).

5. DISCUSSION AND CONCLUSION

This study aims to raise the awareness of how firm’s internal employment of R&D specialists changes with its R&D outsourcing activities.

Theoretical analysis based on AC, KBV, and RBV indicates that firms tend to employ fewer and a lower proportion of R&D specialists when R&D outsourcing deepens, while the opposite is true when R&D outsourcing broadens.

These hypotheses are supported by systematic evidence based on firm-level longitudinal data and econometric analyses using Correlated Random Effect (CRE) Tobit, CRE Selection, and CRE Fractional Response Models. These methods increase the reliability of the results compared with those from the Random Effect (RE) Tobit or Hurdle models usually used in previous studies. This paper also provides the comparison across CRE Tobit estimates and CRE Selection estimates from different second-stage models, which may yield some new insights about the suitability of these alternative models in the context of empirical studies on firm’s R&D and innovation.
The estimations of econometric models quantify the changes in R&D specialist’s employment due to the changes in R&D outsourcing breadth and depth. However, these estimations alone, though statistically significant, do not claim any direction of causality. As suggested by Gelman and Imbens (2013), the focus of interpretation of econometric estimations should be on “effects of causes” rather than “causes of effects”. The causality has already been established through theoretical discussion in section 2, before estimating econometric models in section 3. Although this study focuses only on the effects from one direction, i.e., how employment of R&D specialists changes with R&D outsourcing, it does not exclude the possibility of reverse causality. Actually, as discussed in section 2, internal R&D specialists and R&D outsourcing can complement or substitute each other, depending on why and how the firm outsources R&D. Complementary and substitution effects can go both ways. But the reverse relationship, though also very interesting, is beyond the scope of this paper.

The findings of this paper only reveal the tip of the iceberg. For future research, it would be interesting to examine how employment of R&D specialists and R&D outsourcing mutually and dynamically affect each other. In addition, it is worth to further explore the role of firm strategy in determining the interactions between employment of R&D specialists and R&D outsourcing. Another interesting question is whether R&D employment has a mediation/moderation role in determining how R&D outsourcing influences innovation performance.
References


Strategic Management Journal, 18, 509–33.


<table>
<thead>
<tr>
<th>Variables</th>
<th>Whole sample</th>
<th>Employing R&amp;D specialists</th>
<th>Without R&amp;D specialists</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N = 3973</td>
<td>N = 2987</td>
<td>N = 986</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
</tr>
<tr>
<td>R&amp;D Outsourcing depth</td>
<td>0.17</td>
<td>0.30</td>
<td>0.09</td>
</tr>
<tr>
<td>(share of purchased R&amp;D)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R&amp;D outsourcing breadth</td>
<td>0.74</td>
<td>1.11</td>
<td>0.75</td>
</tr>
<tr>
<td>(types of R&amp;D outsourcing partners)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of R&amp;D specialists</td>
<td>15.16</td>
<td>90.16</td>
<td>20.16</td>
</tr>
<tr>
<td>Share of R&amp;D specialists within R&amp;D function</td>
<td>0.55</td>
<td>0.35</td>
<td>0.67</td>
</tr>
<tr>
<td>Total R&amp;D expenditure (1000.000 DKK)</td>
<td>34.22</td>
<td>275.32</td>
<td>44.45</td>
</tr>
<tr>
<td>Average profit per employee (1000.000 DKK)</td>
<td>0.0070</td>
<td>1.22</td>
<td>0.0126</td>
</tr>
<tr>
<td>Number of full-time-equivalent employees</td>
<td>229.97</td>
<td>865.28</td>
<td>265.71</td>
</tr>
<tr>
<td>Asset value (1000.000 DKK)</td>
<td>784.13</td>
<td>5873.20</td>
<td>941790.2</td>
</tr>
</tbody>
</table>
Table II. Breadth and depth of R&D outsourcing across quintiles (subsample with positive R&D outsourcing, N=1745)

<table>
<thead>
<tr>
<th>Variable quintiles</th>
<th>R&amp;D outsourcing depth</th>
<th>R&amp;D outsourcing breadth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(share of purchased R&amp;D)</td>
<td>(R&amp;D outsourcing partner types)</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
</tr>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt; Quintile</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt; Quintile</td>
<td>0.11</td>
<td>0.02</td>
</tr>
<tr>
<td>3&lt;sup&gt;rd&lt;/sup&gt; Quintile</td>
<td>0.23</td>
<td>0.05</td>
</tr>
<tr>
<td>4&lt;sup&gt;th&lt;/sup&gt; Quintile</td>
<td>0.51</td>
<td>0.13</td>
</tr>
<tr>
<td>5&lt;sup&gt;th&lt;/sup&gt; Quintile</td>
<td>0.99</td>
<td>0.03</td>
</tr>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt; Quintile</td>
<td>0.41</td>
<td>0.38</td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt; Quintile</td>
<td>0.33</td>
<td>0.31</td>
</tr>
<tr>
<td>3&lt;sup&gt;rd&lt;/sup&gt; Quintile</td>
<td>0.31</td>
<td>0.27</td>
</tr>
</tbody>
</table>
### Table III. Estimates from CRE Tobit and CRE Selection models for number of R&D specialists

<table>
<thead>
<tr>
<th>Variables of Interest</th>
<th>CRE Tobit model</th>
<th>CRE Selection model: second stage estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff.</td>
<td>AMEs</td>
</tr>
<tr>
<td><strong>Dependent Variable: Number of R&amp;D Specialists</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breadth of outsourcing (Types of R&amp;D outsourcing partner)</td>
<td>5.910***</td>
<td>2.812***</td>
</tr>
<tr>
<td>(Share of purchased R&amp;D, %)</td>
<td>(0.858)</td>
<td>(0.410)</td>
</tr>
<tr>
<td>Depth of outsourcing</td>
<td>-0.465***</td>
<td>-0.221 ***</td>
</tr>
<tr>
<td>(Share of purchased R&amp;D, %)</td>
<td>(0.054)</td>
<td>(0.026)</td>
</tr>
<tr>
<td><strong>Control Variables</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total R&amp;D expenditure (1000.000 DKK)</td>
<td>0.200***</td>
<td>0.095***</td>
</tr>
<tr>
<td>(Share of purchased R&amp;D, %)</td>
<td>(0.007)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Annual profit per employee (1000.000 DKK)</td>
<td>1.138</td>
<td>0.541</td>
</tr>
<tr>
<td>(Share of purchased R&amp;D, %)</td>
<td>(0.939)</td>
<td>(0.447)</td>
</tr>
<tr>
<td>Log(Number of employee)</td>
<td>5.216**</td>
<td>2.482*</td>
</tr>
<tr>
<td>(Share of purchased R&amp;D, %)</td>
<td>(2.889)</td>
<td>(1.375)</td>
</tr>
<tr>
<td>Log(asset, 1000.000 DKK)</td>
<td>-0.542</td>
<td>-0.258</td>
</tr>
<tr>
<td>(Share of purchased R&amp;D, %)</td>
<td>(2.231)</td>
<td>(1.062)</td>
</tr>
<tr>
<td>R&amp;D Department (Binary)</td>
<td>0.074</td>
<td>0.035</td>
</tr>
<tr>
<td>(Share of purchased R&amp;D, %)</td>
<td>(2.389)</td>
<td>(1.137)</td>
</tr>
<tr>
<td>6 Industry dummies</td>
<td>Chi$^2(6)=34.40***</td>
<td>--</td>
</tr>
<tr>
<td>7 Location dummies</td>
<td>Chi$^2(7)=11.80</td>
<td>--</td>
</tr>
<tr>
<td>3 Year dummies</td>
<td>Chi$^2(3)=3.80</td>
<td>--</td>
</tr>
<tr>
<td>Lambda</td>
<td>--</td>
<td>17.921 (14.020)</td>
</tr>
<tr>
<td>Lambda*3 Year dummies</td>
<td>--</td>
<td>17.921 (14.020)</td>
</tr>
<tr>
<td>Wald Chi2</td>
<td>3885 .89</td>
<td>Prob. &gt; Chi2</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-16014.335</td>
<td>--</td>
</tr>
<tr>
<td>Rho</td>
<td>0.649***</td>
<td>(0.012)</td>
</tr>
<tr>
<td>Observations</td>
<td>3973</td>
<td></td>
</tr>
</tbody>
</table>

***: Significant at 1%; **: Significant at 5%; *: Significant at 10%. Bootstrap standard errors for CRE Selection models (500 repetitions).
Table IV. Estimates from CRE Fractional Response Probit model for share of specialists

<table>
<thead>
<tr>
<th>Variables of Interest</th>
<th>Coefficients</th>
<th>Marginal Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breadth of outsourcing (Types of R&amp;D outsourcing partner)</td>
<td>0.154**(0.067)</td>
<td>0.025*** (0.011)</td>
</tr>
<tr>
<td>Depth of outsourcing (Share of purchased R&amp;D)</td>
<td>-0.990**(0.422)</td>
<td>-0.160** (0.063)</td>
</tr>
</tbody>
</table>

Control Variables

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficients</th>
<th>Marginal Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>R&amp;D Department</td>
<td>-0.157 (0.124)</td>
<td>-0.025 (0.020)</td>
</tr>
<tr>
<td>Total R&amp;D Expenditure (1000.000 DKK)</td>
<td>0.018 (0.012)</td>
<td>0.003 (0.002)</td>
</tr>
<tr>
<td>Annual Profit per Employee (1000.000 DKK)</td>
<td>0.010 (0.090)</td>
<td>0.002 (0.015)</td>
</tr>
<tr>
<td>Log(Number of Employee)</td>
<td>-0.012 (0.153)</td>
<td>-0.002 (0.025)</td>
</tr>
<tr>
<td>Log(asset, 1000.000 DKK)</td>
<td>-0.342** (0.148)</td>
<td>-0.055** (0.024)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dummies</th>
<th>Chi² (degrees of freedom)</th>
<th>Prob. &gt; Chi²</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 Industry Dummies</td>
<td>Chi² (6) = 12.65**</td>
<td>Prob. &gt; Chi² = 0.049</td>
</tr>
<tr>
<td>7 Location Dummies</td>
<td>Chi² (7) = 29.03***</td>
<td>Prob. &gt; Chi² = 0.0000</td>
</tr>
<tr>
<td>3 Year Dummies</td>
<td>Chi² (3) = 30.09***</td>
<td>Prob. &gt; Chi² = 0.0001</td>
</tr>
</tbody>
</table>

Observations: 3634

Wald Chi²: 157.47

Prob. > Chi²: 0.000

Log likelihood: -1478.9955

Rho: 0.539*** (0.047)

***: Significant at 1%; **: Significant at 5%; *: Significant at 10%. Based on bootstrap standard errors (400 replications).
Chapter 3

The effect of knowledge from market research on product innovation - Evidence from longitudinal data
The effect of knowledge from market research on product innovation - Evidence from longitudinal data

Wenjing Wang
School of Business and Social Sciences, Aarhus University
Sino-Danish Center for Education and Research

ABSTRACT

To what extent can producers improve product innovation performance by utilizing the knowledge acquired through different market research practices? Inspired by March (1991), this study first categorizes market research practices into two types - exploitative and exploratory market research, and then examines their implications for product innovation. By analyzing longitudinal data from more than 3000 firms in Denmark, this study finds that the knowledge from exploitative market research and the knowledge from exploratory market research can both improve product innovation performance, and that the knowledge from exploratory market research has slightly larger effects than the knowledge from exploitative market research. In addition, innovation investments are more effective in improving innovation performance for the firms which utilize more types of knowledge acquired through market research.

Keywords: user knowledge, exploratory and exploitative market research, product innovation
1. INTRODUCTION

Despite the fact that users have been increasingly recognized as important sources of innovation (e.g., von Hippel, 1976, 1978, 1986, 1988; Bogers, Afuah and Bastian, 2010), few studies have systematically estimated the impacts of user’s inputs on producer’s innovation performance (Chatterji and Fabrizio, 2014). In particular, regarding to the question whether the knowledge from market research (MR) practices improves product innovation performance, there is no previous evidence based on longitudinal data and analyses (as to our knowledge; see section 2).

Using firm-level longitudinal data, this paper examines how innovation performance changes with the integration of user knowledge acquired from different market research (MR) practices into early stage of innovation development.

Following Laursen and Salter (2006), product innovation performance is measured by the fraction of revenue attributed to new product sales. Inspired by March (1991), producer’s MR practices are first categorized into two groups – exploitative MR and exploratory MR. Exploitative MR refers to the practices which acquire knowledge from current users through established communication channels (e.g., regular contact). Exploratory MR refers to the practices which explore potential market (e.g., anthropological studies and in-depth interviews). In the regression, exploitative MR and exploratory MR are both indicated by the binary variable, each of which equals to 1 if the knowledge acquired from corresponding MR practices is integrated into the early stage of innovation development.

The dataset is constructed by merging Statistics Denmark’s survey data on firm R&D and innovation activities with firm accounting data. The final dataset contains observations from more than 3000 firms during 2007-10. Three estimation methods are applied here: fixed effect (FE), correlated random effect (CRE) Tobit (Wooldridge, 2010), and nearest neighbor matching (NNM) (Abadie and Imbens, 2006, 2011; see section 3). These methods are new in the context of corporate practice evaluations, but their estimators turn out to be consistent and robust.

The results indicate that knowledge from exploratory MR and knowledge from exploitative MR both have positive impacts on product innovation, and the effects of knowledge from exploratory MR are slightly larger than that from exploitative MR. In addition, innovation
investment is most effective for product innovation performance among the firms that incorporate the knowledge from both types of MR into innovation development. Overall, exploratory MR, exploitative MR, and innovation investment complement each other and contribute to product innovation in a coordinated way (see section 4). The paper ends with a discussion on the limitation of this study and ideas for future research (see section 5).

2. THEORY AND HYPOTHESES

As there is always the choice of not doing MR, the existence of MR itself should indicate enough benefits that at least justify the costs - it has been widely recognized that MR contributes to successful overall sales (from old and new products). Nevertheless, few have noticed that the knowledge from MR may also influence the structure within overall sales, particularly the proportion of new product sales, which reflects product innovation performance. As for the question whether the knowledge from MR impels product innovation, existing studies have not yet reached consensus. Just as pains and gains of tapping into external knowledge co-exist for innovation performance (Kaiser and Grimpe, 2010; Martinez-Noya et al., 2013), two competing forces of using knowledge from MR may drive the overall effects on innovation to one direction or another. This section will discuss the benefits, costs, contingencies and overall effects of knowledge from MR on product innovation, as well as how the effects may differ between two types of knowledge acquired from various MR practices.

2.1. Knowledge from MR and product innovation outcomes

In general, knowledge about users has been recognized as crucial to innovation. While some previous studies highlight the benefits of user knowledge (e.g. Hippel, 1998), some are more concerned with its potential caveats (Tzeng, 2009), others identify several contingencies that determine the overall effects of user knowledge on product innovation (e.g., Foss, Laursen, and Pedersen, 2011). Before addressing the overall effects, we will first discuss its benefits, costs and contingencies of integrating knowledge from MR for product innovation performance.

2.1.1. Benefits of user knowledge from MR for product innovation

The major benefit of user knowledge for innovation is that, producers can recombine their prior knowledge with distinctive ideas from users, the process of which gives rise to innovations
(Chatterji and Fabrizio, 2014; Cohen and Levinthal, 1990). In addition, MR gives producers a better understanding of user’s needs. This market insight enables producers to select and prioritize the innovative ideas that users value most, which in turn translates into more sales of new products and higher return to the innovation investment. In this sense, knowledge about users should be able to contribute significantly to producer’s innovation performance.

Some evidence suggests various positive effects of integrating knowledge about users on innovation. For example, collaborating with physicians can enhance innovation of medical device firms (Chatterji and Fabrizio, 2014). Coordination between marketing/sales and R&D (Ernst, Hoyer, and Rübsaamen, 2010) and customer relationship management (Ernst, Hoyer, and Rübsaamen, 2010) can be beneficial for new product development. More generally, opening to external sources can increase innovation performance (Laursen and Salter, 2006).

2.1.2. Costs of user knowledge from MR for product innovation

However, contradicting to the above-mentioned rationale and evidence in favor of “listening to customers” for innovation, a few studies emphasize potential caveats of doing so. First, a firm’s initiatives of involving users may lead to missing opportunities for radical innovation. Even initiated by in-house efforts, a radical innovation may have to first “escape” from current users and other conservative stakeholders, because there is a general tendency of avoiding disruption in existing procedures and markets (Christensen, 1997), presumably due to inertia or habits. Second, when focusing and researching on users, firms tend to overlook the “non-consumption” group and external incremental innovation designed for serving it. As the external innovation improves gradually and finally goes beyond serving the “non-consumption”, the original focus on easily identifiable users is not only harmful to relative innovation performance, but also threatening the survival of firm. Third, because users rely on existing products as reference point for expressing preferences, their suggestions may be too conservative (Brockhoff, 2003). Last but not least, MR distracts resources which can otherwise be allocated to innovation and R&D activities. Consequently, the user’s extensive participation can undermine the creativity of in-house innovators and the process of integrating users may result in “mass mediocrity” (Greer and Lei, 2012). Tzeng (2009) even suggests firms can perform better by not listening to their users.
2.1.3. Contingencies

Some evidence suggests that the link between user integration and innovation is indirect or depending on a third group of factors (contingencies). Indeed, these contingencies may decide whether the benefits can outweigh the costs, pulling the overall effects in one direction or another. For example, Foss, Laursen, and Pedersen (2011) find that the effects of customer integration are completely mediated by organizational practices. Mahr, Lievens and Blazevic (2014) summarize that the effect of user co-creation is “contingent on the richness and reach of the communication channels enabling cocreation”. Fang, Lee and Yang (2015) suggest that the outcomes of co-development are contingent on a firm’s position in the value chain and the factors that facilitate effective cooperation. More generally, market turbulence (Lichtenthaler, 2009), market competitiveness (Kim and Atuahene-Gima, 2010), appropriability (Levinthal and March, 1993), etc. may influence the overall effects of various kinds of organizational learning.

In sum, prior studies have addressed the question of “whether listening to users is beneficial for product innovation” from different angles. Instead of a simple yes or no, the answer may depend on whom to listen to, how to capture and communicate user’s knowledge, as well as where and when to apply the knowledge. Nevertheless, even for a much narrower question inquiring “whether the knowledge acquired from MR impels product innovation”, existing studies have not yet reached consensus. In order to address this question, it is necessary to control for contingencies and other firm characteristics, and then compare innovation outcomes between observations that differ only in the decision on integrating knowledge acquired from MR. In this context, systematical analysis based on longitudinal data should be able to advance our understanding.

2.2. User knowledge from two types of MR practices

Another gap in previous studies is that they treat user knowledge acquired from MR as unitary. However, MR may include various practices designed for acquiring user knowledge of different types, which may have heterogeneous implications for innovation. Nevertheless, although it is ideal to examine the knowledge from each MR practice separately, this is not practical because some MR practices are adopted as a bundle so that it is difficult isolate the effect of one practice from another. To balance between oversimplification and redundancy, we can re-categorize
these various MR practices into a few distinguishable groups, and then examine the effect of each type of MR practices.

In this context, the twin concepts of “exploitation” and “exploration” provide a helpful framework for this re-classification. According to March (1991), the essence of exploitation is the refinement and extension of existing competencies, technologies, and paradigms, while the essence of exploration is experimentation with new alternatives. In a systematical review of research on exploitative and exploratory learning, Gupta, Smith and Shalley (2006) suggest differentiating these two terms by “focusing on the type or amount of learning rather than on the presence or absence of learning.” (pp: 694). Following the above interpretation, this study classifies user knowledge into two types: one is acquired from exploitative MR, and the other is acquired from exploratory MR.

Exploitative MR refers to acquiring information from current users through established communication channels. Consistent with the general concept of exploitation, exploitative MR features with refinement, extension and implementation. Thus, the traditional marketing methods, which mainly focus on learning from current users about their explicit needs through standardized procedures, belong to this category. Examples include using surveys to gather users’ opinions, or documenting users’ feedback through regular contact. These MR methods are mainly used for exploiting user’s real-world experience and identifying current problems of existing products.

Exploratory MR mainly refers to acquiring information about the potential market through the practices beyond routine communication with current users. Consistent with the general term of exploration, the exploratory MR features with risk taking, experimentation and flexibility. Therefore, the advanced marketing methods, which are designed for learning tacit knowledge and latent needs from both current and potential users, belong to this category (Leonard and Rayport, 1997; Narver, Slater and MacLachlan, 2004). One example is the “lead user method”, which means producers involve lead users in the product development process, in order to collect information about their needs and solutions (Lilien, Morrison, Searls, Sonnack and von Hippel, 2002). Another example is participatory design (also called design anthropology), in which developers may work and live with the users being studied (Buur and Matthews, 2008;
Pals, Steen, Langley and Kort, 2008). Compared with the knowledge acquired from exploitative MR, the knowledge acquired from exploratory MR (1) emphasizes more on future products and potential users (2) involves higher risks and (3) could be more relevant to radical innovation.

This classification allows for the examination of the knowledge from MR practices from a comfortable distance, which is neither too near to manage the numerous detailed practices, nor too far to tell the distinct internal patterns.

Interestingly, marketing literature propagates a parallel pair of concepts: responsive market orientation and proactive market orientation. They are defined by Narver et al. (2004) with the purpose of extending the measure of market orientation introduced by Narver and Slater (1990) and Kohli and Jaworski (1990). The responsive market orientation addresses producer’s attempts to understand customers’ expressed needs, while the proactive market orientation addresses producer’s attempts to understand customers’ latent needs, which include opportunities for customer value of which the customer is unaware (Narver et al., 2004). Both types of market orientation focus on understanding users; the major difference between them is that the proactive orientation includes a second round of thinking, i.e. the producer’s attempts to put herself into the user’s shoes and infer user’s latent and future needs, which are beyond the reach of responsive market orientation (Blocker, Flint, Myers, and Slater, 2011).

Both concepts of responsive/proactive market orientation and exploitative/exploratory MR can distinguish producer’s efforts of understanding user’s current and future needs. However, these two pairs of concepts take different angles to reflect producer-user interaction: while the concepts of market orientation departure from the perception of user’s needs, the concepts of exploitative/explorative MR center around producer’s behaviors. Compared with the former pair of concepts, the later pair is objective and it pins down to actual practices in a more concrete way. For example, while responsive/proactive market orientation concepts mainly focus on current and potential needs of existing users, exploitative/explorative MR concepts allow equal emphasis on potential users. This is because the exploitation/exploration angle allows the producer to view existing users from a longer distance (rather than be pushed into current user’s shoes). In this way, producer’s view becomes broad enough to aware alternative opportunities, e.g., potential users. In addition, exploitation/exploration concepts are more frequently
associated with objective measures of producer’s actual behaviors, while market orientation concepts mainly associate with subjective measures of producer’s perceptions. Therefore, this study takes a different angle but still complements previous market orientation literature such as Blocker et al. (2011) and Kim, Im and Slater (2013).

2.3. The effect of using knowledge from exploitative and exploratory MR

The above discussion indicates that the knowledge acquired from exploitative and exploratory MR may differ in costs, benefits and overall effects on product innovation.

From cost side, the knowledge acquired from exploratory MR can be more expensive than that from exploitative MR. This is because exploratory MR aims to extract tacit and sticky knowledge. Since tacit knowledge is impossible to codify (Collins, 2001), and sticky knowledge is costly to transfer (Von Hippel, 1994), extracting these kinds of knowledge requires a higher degree of user involvement and interaction, which means higher cost. Therefore, exploratory MR may distract more resources from R&D and innovation activities. Moreover, as exploratory activities feature experimentation with new alternatives, they are generally more risky. Since most firms are risk-adverse, the additional unpredictability means additional cost on performance.

From the benefit side, the knowledge acquired from exploratory MR can be more beneficial to product innovation than exploitative MR. First, exploratory MR brings more distant knowledge, which allows for more innovative associations and ideas. A major concern for traditional MR is that it may impede or only incrementally improve innovation, because of its narrow focus on current and explicit customer needs (Berghman, MatthysSENS, and Vandenbempt, 2006; Pedrosa, 2012). Exploratory MR alleviates this concern, because it reaches out to potential users and advances into unique interpretation of users’ latent needs. In this sense, exploratory MR further expends the range and diversity of knowledge, which prepares it for incubation of more innovative ideas.

In sum, prior discussion does not yield clear-cut predictions about the overall implications of integrating the knowledge from exploitative/exploratory marketing research for product innovation. In order to know whether their benefits for outweigh the costs in general, it is necessary to resort to data and examine the following hypotheses:
Hypothesis 1: Knowledge from exploratory MR impels product innovation.

Hypothesis 2: Knowledge from exploitative MR impels product innovation.

Prior comparison between knowledge from exploitative/exploratory MR infers that knowledge from exploratory MR has higher return, higher cost, and are more risky, compared with knowledge from exploitative MR. According to economics theory, a rational firm allocates investments between two alternatives of MR until their marginal net (risk-adjusted) returns become equal. Because the risk information is unobserved and uncontrolled in the model, the estimated returns also include the compensation of additional risk. Therefore, other firm characteristics being equal (which include investment on obtaining and utilizing knowledge from MR, in-house innovation investment, firm size, etc.), knowledge acquired from exploratory MR is expected to bring slightly more benefits, so that the premium can compensate for the additional risk. Although the data is not enough to test the underlying (risk-compensating) mechanism which explains the additional return from exploratory MR, this empirical analysis should be able to reveal that the return of exploratory MR is higher than the return of exploitative MR. This corresponds to the following hypothesis:

Hypothesis 3: Other conditions being equal, knowledge from exploratory MR contributes more to product innovation than knowledge from exploitative MR.

In addition to direct effects, integrating knowledge from MR may also affect product innovation indirectly. As discussed in section 2.1, market insight help producers to know which new features or products are valued most by users, and then prioritize and select the innovation investment accordingly. In this way, innovation investment can be translated into more new product sales. This could be tested through the following hypothesis:

Hypothesis 4: Innovation investment is more effective in the firm that utilizes more knowledge acquired from MR.

3. EMPIRICAL STRATEGY

The effects of integrating user knowledge acquired from exploratory/exploitative MR activities are estimated by several econometric models using longitudinal data.
3.1. Data

The dataset is constructed by merging Statistics Denmark’s survey data on firms’ R&D and Innovation activities (FoU) with basic information data on firms (FIRE). Compared with the harmonized Community Innovation Survey (CIS), FoU survey is a more detailed and frequent version for firms in Denmark. FoU data is available from 1990s, but only since 2007 have the variables and measurements become highly consistent across years. FoU survey data contains information of firm’s R&D innovation activities, such as their financial sources, investments, expenditures, number of R&D employees, revenue share of new products etc. FIRE data provides basic information of the firm, such as location, industry, total number of employees, assets, etc. Only firms that appear in both datasets are used. Considering the availability and consistency of the variables of interest, this paper mainly uses the data from 2007 to 2010. Data from 2001-2006 is partially exploited for obtaining the pre-sample mean of outcome variable\(^1\). Each year’s survey between 2007 and 2010 contains around 4,000 firms. However, only a proportion of these are engaged in innovation activities. Including pre-sample mean and lagged values exclude a number of observations. For the purpose of this analysis, the sample (and the population of interest) is further restricted to firms with positive innovation expenditures. Because firms may not participate in the survey or have positive innovation expenditures every year, the panel data is unbalanced. The final dataset contains observations from 3,150 firms.

3.2. Measures

3.2.1. Product Innovation

Following Laursen and Salter (2006), this paper examines corporate innovation levels through the share of new product sales in total revenue. This measure reflects the performance of new products relatives to old ones. Its increase indicates higher returns from new products and

---

\(^1\) Inspired by Blundell et al. (2002), this pre-sample mean of the outcome variable is used as a proxy for time-invariant unobserved heterogeneity.
continuous improvement of overall product performance. This measure also reflects the product updating rate, a higher level of which indicates a more vigorous product profile. It actually balances between the two most popular measures for innovation, i.e. the number of patents/innovations (innovation antecedent) and market return (innovation outcome). This niche position allows shedding light on the underlying mechanism of how innovation antecedents are translated into outcomes.

The share of new product revenue \( y \) is transformed into \( y^* = \ln(1 + y) \), and \( y^* \) is used in the regressions. This log transformation of \( y \) is based on several considerations. Generally speaking, when the dependent variable is in log form, assumptions for the classical linear model are more likely to be satisfied. Moreover, log transformation also reduces the extrema in the data, so that it curtails the effects of outliers (Wooldridge, 2010). In our data, the original distribution of \( y \) is quite skewed and there is a long tail of extreme values; when logs are applied, the distribution behaves better. However, directly transforming \( y \) into \( \ln(y) \) will exclude observations with \( y = 0 \), leading to a waste of information. An alternative solution is to transform \( y \) into \( \ln(1 + y) \). The neat feature of this transformation is that, it exploits the advantages of log transformation without dropping observations with \( y = 0 \). In fact, when \( y = 0 \), \( y^* = 0 \) too. More generally, when \( y \) is small, \( y^* = \ln(1 + y) \approx y \). As the sample average of \( y \) is only 0.11 in our data (see table 1), the marginal effects for \( y^* \) should be very similar with those for \( y \).

### 3.2.2. The exploitative and exploratory MR

The primary independent variables of interest are two binary indicators: each takes on a value equal to 1 if the firm integrates the knowledge about users into corresponding MR practices into

---

\(^1\) More precisely, the magnitude of marginal effect for \( y \) is slightly larger than that estimated for \( y^* \). This is because given \( y^* = \ln(1 + y) \) and \( 0 < y < 1 \), we have \( \left| \frac{dy}{dx} \right| = \left| \frac{(1+y)dy^*}{dx} \right| > \left| \frac{dy^*}{dx} \right| \), where \( dy/dx \) is marginal effect for \( y \) and \( dy^*/dx \) is marginal effect for \( y^* \). For discrete variables, similar logic applies. In this sense, the estimates using \( y^* \) are relatively conservative.
early stages of new product development (i.e. concept development or innovation implementation), and 0 otherwise.

More specifically, the binary indicator for using knowledge acquired from exploitative MR takes on a value equal to 1 if the firm integrates the knowledge acquired from regular marketing methods, such as contact with users for example through daily dialogue and other routines. The binary indicator for using knowledge acquired from exploratory MR takes on a value equal to 1 if the firm uses the knowledge acquired from at least one of the following advanced MR methods: observation or interview (e.g., anthropological studies or in-depth interviews), involvement of ordinary users (e.g., prototype tests or internet communities with ongoing feedback in the innovation process), and inclusion of advanced users (e.g., lead user methods).

The key distinction between exploitative and exploratory MR is that, the exploitative MR attaches more weight to learning from existing users about explicit needs, while the exploratory MR aims to hedge future and implicit needs of both existing and potential users.

3.3.3. Controls

Major control variables include the firm’s R&D expenditures, innovation expenditures (unit: DKK million), and the diversity of R&D partners (measured by the number of R&D partner types among the following categories: Danish firms from the same industry, foreign firms from the same industry, other firms in Denmark, technology service institution in Denmark, higher education institute in Denmark, other public research institute in Denmark, other types of organizations in Denmark, other type of foreign organizations and other foreign research institute). In addition, firm basic characteristics are controlled for, including salary expenditures, industry, location indicators, and the firm size (which is measured by asset value and the number of employees except those who work on R&D or innovation). When applicable, a one-year lag value of new product’s share in total revenue ($y_{t-1}$) and the average of $y$ over the pre-sample periods 2001-2006 are also controlled for to capture unobserved heterogeneity.

Table 1 presents the descriptive statistics. The sample contains 5344 observations: 28.9% of them (1543 observations) do not use knowledge from exploitative or exploratory MR during product innovation; 26.6% of them (1423 observations) use both knowledge from exploitative and knowledge from exploratory MR during product innovation. On average, sales of new
products account for 11% of total revenue. Among the observations which adopt knowledge from exploitative and exploratory MR, sales of new products account for 21% of total revenue on average. Compared to knowledge from exploitative MR, knowledge from exploratory MR is less commonly used: 3704 observations adopt the knowledge from exploitative MR, while 1520 observations adopt the knowledge from exploratory MR. It is relatively rare that a firm adopts knowledge from exploratory MR without adopting knowledge from exploitative MR: only 1.8% of the observations show this combination. In contrast, 42.7% of the observations use knowledge from exploitative MR without using knowledge from exploratory MR.

3.3. Estimation Methods
The main objective of the empirical analysis is to estimate the effects of incorporating user knowledge acquired from MR into early stage of product innovation. The effects of the knowledge from exploitative MR and that from exploratory MR are distinguished and estimated separately. Three estimation methods are applied here: fixed effect (FE), correlated random effect (CRE) Tobit, and nearest neighbor matching. Fixed effect model imposes fewest assumptions for consistency; CRE Tobit model requires additional (but not unrealistic) assumptions in exchange for more efficient estimators; matching methods exploit the binary nature of treatment variable. Matching methods use a different strategy for causal inference, and they become increasingly popular in the program evaluation literature aiming to estimate “causal effects”. The FE estimates are the baseline; CRE Tobit and matching estimates are for comparison and robustness check.

3.3.1. Fixed Effect Analysis
Under the assumption of ignorability conditional on unobserved heterogeneity, FE analysis can lead to consistent estimators of treatment effects (Wooldridge, 2010). This assumption is more likely to be satisfied when the potentially influencing past values of explanatory variables are controlled together with contemporary ones, so that the rest of the unobserved heterogeneity
does not correlate with “treatments”, i.e., binary variables of interest. The FE estimators in this paper are based on the following model:

\[ y_{it} = \alpha + \beta m_{it} + \gamma w_{it} + x_{it}' \delta + \lambda m_{it-1} + \mu w_{it-1} + x_{it-1}' \xi + \theta d_t + a_i + u_{it} \] (1)

Where i is firm indicator and t is year indicator, \( y_{it} \) is the share of revenue from new product sales, \( m_{it} \) is binary variable indicating whether the firm incorporates knowledge acquired through exploitative MR practices (e.g. regular contact with customers); \( w_{it} \) is a binary variable indicating whether the firm incorporates knowledge acquired through exploratory MR practices (e.g. advanced marketing research methods), \( x_{it} \) is a vector of control variables including innovation expenditure, asset, etc.; \( a_i \) denotes firm specific unobserved heterogeneity; \( d_t \) denotes a vector of year dummy; \( u_{it} \) denotes idiosyncratic error; \( w_{it-1}, m_{it-1}, x_{it-1} \) are one-year lag values of \( w_{it}, m_{it}, x_{it} \) respectively.

Our primary interests are \( \beta \) and \( \gamma \), which reflect how innovation performance responses to the integration of user knowledge acquired from exploitative MR (\( m_{it} \)) or exploratory MR (\( w_{it} \)). Following Wooldridge (2010), under the key assumption that the distribution of \( m_{it} \) and \( w_{it} \) is independent of \( a_i \) and \( u_{it} \) (conditional on other covariates), the fixed effect estimators of \( \beta \) and \( \gamma \) in equation (1) consistently reflect the “treatment effects” of \( m_{it} \) and \( w_{it} \).

3.3.2. CRE Tobit Model

A Tobit model applies to the situation where outcome variable has strictly positive probability of being zero. The Tobit model has a high potential to fit our data well, since there are a significant number of observations with zero new product sales. The CRE approach, which dates back to Mundlak (1978), allows for correlations between unobserved heterogeneity and explanatory variables, so that CRE approach requires less restrictive assumptions than traditional random effects (RE) approach. By assuming that the unobserved time-invariant heterogeneity can be fully captured by average values of each time-variant dependent variable, CRE approach allows for correlations between unobserved time-invariant heterogeneity and independent variables to some extent. Following the method that combines the Tobit model with the CRE device (Wooldridge, 2010), this paper includes average values of each time-variant dependent variable as additional controls in random effect Tobit regression. The CRE Tobit model estimated in this paper is:
\[ y_{it} = \max (0, \psi + \alpha m_{it} + \beta w_{it} + x_{it} \gamma + \delta \bar{m}_i + \rho \bar{w}_i + \bar{x}_i \theta + \lambda \bar{y}_i + d_i \tau + a_i + u_{it}) \]  

(2)

where \( w_{it}, m_{it}, x_{it}, d_t, a_t, u_{it} \) have the same meanings as in (1), and \( \bar{y}_i \) is the pre-sample average of \( y_i \); \( \bar{m}_i, \bar{w}_i, \bar{x}_i \) denote the average of \( m_{it}, w_{it}, x_{it} \) across years 2008-2010 respectively.

Following Wooldridge (2010), under the following assumptions

\[ a_i \mid (m_t, w_t, x_t, \bar{y}_i, d_t) \sim \text{Normal} (0, \sigma^2_a) \]

\[ u_{it} \mid (m_t, w_t, x_t, \bar{y}_i, d_t) \sim \text{Normal} (0, \sigma^2_u) \]

model (2) can be estimated by joint maximum likelihood estimation. The standard errors are obtained by panel bootstrap procedure, in which cross section units are re-sampled. According to Wooldridge (2010), in the setting of CRE Tobit model, usual standard errors have to be adjusted, while panel bootstrap standard errors are valid.

The mean function for share of new product sales (average structural function, ASF, defined by Blundell and Powell, 2004) can be estimated as:

\[ \bar{ASF}_t = N^{-1} \sum_{i=1}^n f (\hat{\psi} + \hat{\alpha} m_t + \hat{\beta} w_t + x_t \hat{\gamma} + \hat{\delta} \bar{m}_i + \hat{\rho} \bar{w}_i + \bar{x}_i \hat{\theta} + \hat{\lambda} \bar{y}_i + \hat{d}_t \tau, \hat{\sigma}^2_a + \hat{\sigma}^2_u) \]

(3)

where \( f (z, \sigma^2) = \Phi \left( \frac{z}{\sigma} \right) z + \sigma \varphi (z/\sigma) \) is the mean function of the Tobit model.

The average marginal effect (AME) is estimated by taking partial derivative (or calculating the change) of function (2) with respect to the corresponding element of interest among \((m_t, w_t, x_t)\).

### 3.3.3. Nearest-Neighbor Matching under Dynamic Ignorability Assumption

In the setting of panel data, the matching method based on dynamic ignorability assumption (Lechner, 2005) is a very powerful alternative to estimate treatment effects. First, the assumption of this method is more realistic (less restrictive), which allows the treatment to depend on past observed outcomes, treatments and covariates (i.e., dynamic ignorability assumption); while most other methods rarely allow past outcomes to appear as independent variables in panel data setting. Under the dynamic ignorability assumption, matching estimators are consistent (Lechner, 2005). Second, this method makes good use of both cross-sectional and longitudinal dimensions of the available data. It applies matching methods for each cross-sectional dimension, while
allowing error terms from different time periods to be clustered by firm identification number. Third, matching methods complement FE or CRE Tobit methods by approaching causal inference from another angle; in general, matching methods have been proven to be especially helpful for causal inference in situations where experiments are impractical (e.g., in the context of this study).

For each cross-section, the estimation is based on nearest-neighbor matching (NNM) method derived by Abadie and Imbens (2006, 2011). NNM method estimates treatment effects by averaging the differences between actual outcome of each observation and its counterfactual outcome, which belongs to the “nearest-neighbor” observation with the other treatment state. This paper uses Mahalanobis distance to decide the nearest neighbor, which completely depends on their covariates.

The matching analysis in this paper is based on the following model:

\[ y_{it} = \alpha_t + \sigma_t y_{it-1} + \beta_t m_{it} + \gamma_t w_{it} + x_{it} \delta_t + \lambda_t m_{it-1} + \mu_t w_{it-1} + x_{it-1} \xi_t + \alpha_t + u_{it} \]  

(4)

where \( y_{it-1} \) is the one year lagged value of \( y_{it} \), and all the other notations are the same with (1).

The estimation procedure is: first estimate the treatment effects in equation (4) using NNM method (with Mahalanobis distance metric) for each time period, and then use formula (5) to obtain the overall average treatment effects:

\[ \beta, \gamma_{ate/att} = T^{-1} \sum_{t=1}^{T} \beta, \gamma_{ate/att,t} \]  

(5)

where \( \beta, \gamma_{ate} \) are the estimates of average causal treatment effects; \( \beta, \gamma_{att} \) are the estimates of average causal treatment effects on treated; \( \beta, \gamma_{ate/att,t} \) are the corresponding NNM estimators for year \( t \). The standard errors are calculated by panel bootstrap method, which is valid in the estimation of dynamic model using panel data, according to Wooldridge (2010).

### 3.3.4. Heterogeneous Marginal Effects and Complementarities

Formula (3) indicates that marginal effects of CRE Tobit model depend on all covariates. In other words, marginal effects can be heterogeneous across subgroups with different characteristics. These heterogeneous marginal effects could further reveal the potential interaction among covariates. The later part of this paper also reports marginal effects for
different subgroups. These effects are estimated by first obtaining the change in the predicted outcome due to the change in variable of interest for each observation within a certain group, and then averaging over the changes across all the observations in the group. Again, panel bootstrap methods are applied for the inferences.

The complementarities among exploitative MR, exploratory MR and innovation investment are inferred through comparing the marginal effects for different subgroups of observations. The following explains why complementarity can be inferred in this way.

According to the definition of complementarity, two variables $x$ and $z$ are complementary in determining $f(x, z)$ if the marginal effect of $x$ increases as $z$ increases. Therefore, two continuous variables $x$ and $z$ are complementary if the second order partial derivative of function $f(x, z)$ is positive: $\frac{\partial^2 f(x, z)}{\partial x \partial z} > 0$.

In the context of this study where the mean function $f(x, w)$ is as in (3), $w$ is binary ($w = 1$ if using knowledge from exploratory MR) and $x$ (innovation investment) is continuous, the condition for complementarity between $w$ and $x$ corresponds to:

$$\frac{\Delta [df(x, w)]}{\Delta w} = \frac{df(x, 1)}{dx} - \frac{df(x, 0)}{dx} > 0 \quad (6)$$

since $\Delta w = 1$. Note that $\frac{df(x, 1)}{dx}$ is the marginal effect of $x$ when $w = 1$ and $\frac{df(x, 0)}{dx}$ is the marginal effect of $x$ when $w = 0$.

Similarly, if both variables of interest are binary, such as $w, m$ in the mean function $f(w, m)$, the condition for complementarity between $w$ and $m$ corresponds to:

$$\frac{\Delta [df(w, m)]}{\Delta w} = \frac{df(1, m)}{\Delta m} - \frac{df(0, m)}{\Delta m} = f(1, 1) - f(1, 0) - [f(0, 1) - f(0, 0)] > 0 \quad (7)$$

Again, $f(1, 1) - f(1, 0)$ is the partial effect of $m$ when $w = 1$ and $f(0, 1) - f(0, 0)$ is the partial effect of $m$ when $w = 0$. Similar framework for identifying complementarity can be found in Leiponen (2005).

In sum, if the marginal effect of $x$ (or $m$) is larger for $w = 1$ than for $w = 0$, then $x$ (or $m$) and $w$ are complementary.
4. RESULTS

Table 2 presents fixed effect estimates and CRE Tobit estimates. Theoretically, fixed effect estimators are better in terms of consistency, while CRE Tobit estimation is more efficient and able to reveal additional information on the effects of time-invariant characteristics. Comparing the estimates from the two methods, both the magnitude and significance levels of variables of interest are actually highly consistent. On average, incorporating knowledge from exploratory MR (advanced methods) improves product innovation performance by approximately 4 percentage points. Considering the new product sales only contribute around 11 percentage points of total revenue on average (see table 1), this marginal effect actually corresponds to approximately 36 percentage increase relative to sample average. Similarly, incorporating knowledge from exploitative MR (regular contact with users) improves product innovation performance by approximately 3 percentage points. This is equivalent to approximately 27 percentage increase in relative term. Estimates in both models are significant at 5 percentage or 1 percentage level.

As for control variables, innovation expenditures significantly contribute to product innovation performance. Each percentage point increase of innovation expenditure can increase product innovation performance by 0.7-0.8 percentage points. Similar pattern is found for diversity of R&D partners: on average, introducing one additional type of R&D partner improves product innovation by 2.6 percentage points, according to fixed effects estimator and 1.1 percentage points according to the CRE Tobit estimator. Increasing firm size (as measured by assets), however, has a lagged negative impact on product innovation performance.

........................................................................................................

INSERT TABLE 2 ABOUT HERE
........................................................................................................

Table 3 shows estimates from the nearest neighbor matching (NMM) method under dynamic ignorability assumption. The average treatment effect on treated refers to the effect for the firm who has already integrated knowledge from MR into innovation development. The average treatment effect refers to the effect for a firm in general, regardless whether it has already
integrated knowledge from MR or not. In addition to average treatment effects across three years, those for each year are also reported. The effects are relatively stable across years. The NMM estimates are generally larger, compared to the FE and CRE Tobit estimates. Both effects are almost doubled here: incorporating knowledge from exploratory MR leads to approximately 8 percentage points increase in innovation performance, and incorporating knowledge from exploitative MR improves innovation performance by around 6 percentage points on average.

Because the consistency of FE estimators requires fewest assumptions, FE model is usually considered as the baseline for comparison between alternative estimating methods. In our analysis, as CRE Tobit estimators are generally consistent with FE estimators, we can infer that the averages of covariates and pre-sample mean of outcome variable used in CRE Tobit model are effective in taking care of time-invariant unobserved heterogeneity. Compared with FE estimators, NNM estimators have larger magnitudes. The major reason can be that matching estimators are likely to pick up the positive effects of unobserved time-invariant factors which are also correlated with the use of knowledge from exploitative/exploratory MR. The comparison between the CRE Tobit model and the NNM model further indicates that, the mean values of observables ($\bar{m}_i$, $\bar{w}_i$, $\bar{x}_i$ and $\bar{py}$ in equation (2) for the CRE Tobit model) are better in capturing the unobserved heterogeneity than the previous-year outcome ($y_{it-1}$ of in equation (4) for the NNM model). These mean values of observables may correspond to the contingency factors mentioned by previous theory (see section 2). In this sense, this difference (between matching and FE estimators) also implies the importance of contingency factors in determining how innovation performance responses to user knowledge, which has been pointed out in previous studies (e.g., Foss, Laursen, and Pedersen, 2011).

The above results confirm the hypotheses 1 and 2 that both types of user knowledge impel product innovation. Despite the potential negative effects, integrating knowledge from exploitative/exploratory MR is still beneficial enough to have positive overall effects on innovation performance.
Quantifying the effects reveals a slightly larger impact of exploratory MR compared to exploitative MR, which supports hypothesis 3; the difference is only about 1 percentage point according to FE or CRE Tobit estimations. As discussed in section 2, this difference may reflect the premium that compensates for additional risk of obtaining and using knowledge from exploratory MR practices.

In order to examine whether innovation investment is more effective for firms using a broader range of MR methods (hypothesis 4), it is necessary to compare the effects of innovation investment across the subgroups that utilize knowledge from different combinations of MR practices.

Table 4 summarizes the overall marginal effects for four types of subgroups: observations not using knowledge from MR in early stage of product innovation, observations using only the knowledge acquired from exploitative MR (e.g., regular customer contact), observations using only the knowledge acquired from exploratory MR (e.g., advanced marketing methods), and observations using both the knowledge from exploitative MR and the knowledge from exploratory MR. Not surprisingly, the general pattern for almost every significant estimator is that, the effect is larger for observations that use more types of knowledge acquired from MR practices.

The effectiveness of innovation investment can be inferred by the marginal effects of innovation expenditure on innovation performance. Table 4 shows that a one percentage increase in innovation expenditures leads to a 0.011 percentage point increase in product innovation performance for the firms using knowledge from both exploratory and exploitative MR, but only to a 0.003 percentage point increase in product innovation performance for the firms not using knowledge from MR. In other words, the effectiveness of innovation investment in firms using both knowledge from exploratory and exploitative MR is almost 3.67 times higher than firms not using knowledge from MR. By applying the estimates in Table 4 to formula (6), we can get
the complementary effect between innovation investment and exploratory MR, which is 0.003 for firms not using any knowledge from MR and 0.004 for firms only using knowledge from exploitative MR. The complementary effect between innovation investment and exploitative MR is 0.004 for firms not using any knowledge from MR and 0.005 for firms only using knowledge from explorative MR. In any of the above or similar situations, the complementarity between innovation investment and using knowledge from MR is positive. Therefore, hypothesis 4 is supported.

Notably, the return of integrating additional knowledge from MR also follows the above pattern: the firm using knowledge from both exploratory and exploitative MR will experience a decrease of product innovation performance by 5.6 percentage points if it stops using the knowledge from exploratory MR. However, the firm not using any knowledge from exploitative or exploratory MR will improve innovation performance by only 1.8 percentage points if it begins to integrate knowledge from exploratory MR. Similar pattern applies to the effects of exploitative MR. Applying the estimates in Table 4 to formula (7), we can get the complementary effects between using knowledge from exploitative MR and using knowledge from explorative MR, which is a) 0.031 - 0.015 = 0.016 for firms not using any knowledge from MR and b) 0.047 - 0.029 = 0.018 for firms using only exploitative MR. Similar positive complementary effects can be found for other combinations. Therefore, using knowledge from exploitative MR and using knowledge from explorative MR are complementary in improving product innovation performance.

Similarly, an increase of R&D partnership diversity may also bring larger improvement in innovation performance for the firm using knowledge from more types of MR.

These observations indicate that knowledge from exploratory and exploitative MR, innovation investment and the diversity of R&D partners improve product innovation performance in a coordinated and complementary way.

One may challenge the validity of the above results by raising the issue of endogeneity. One major source of endogeneity is unobserved heterogeneity. But it should not cause too much concern in this study. The time-invariant unobserved heterogeneity is eliminated to the minimum (if not all) by using longitudinal data and methods along with a rich set of control variables including pre-sample mean. The time-variant unobserved heterogeneity is reduced to
the lowest possible level by using the NNM method which allows for powerful control variables such as lagged outcome. Another source of endogeneity is reverse causality. It is quite weak in the context of this study, because a firm’s decisions regarding the use of the knowledge acquired from exploratory/exploitative MR practices are not likely to be affected by innovation performance in the same time period. Innovation performance is usually a “side product”, which is not the key factor which firm monitors frequently or adjusts decisions according to. Even if a firm does decide to integrate knowledge from MR due to the change of innovation performance, it usually takes significant amount of time to do so: the firm has to get feedback about innovation performance, analyze it, and decide whether to integrate knowledge from MR to product innovation. Therefore contemporary influence from innovation performance on integrating knowledge from MR is limited.

5. DISCUSSION AND CONCLUSION

This study provides systematic evidence that both knowledge from exploitative market research (MR) and knowledge from exploratory MR improve product innovation performance. Integrating knowledge about users into innovation development rejuvenates a firm’s product profile in general, no matter whether the knowledge is acquired from MR on leading users, ordinary users, current or potential users. The effect of integrating the knowledge from exploratory MR is slightly larger than that from exploitative MR. Innovation investment exhibits higher returns among the firms which utilize the knowledge from both exploratory and exploitative MR during product innovation. This implies that, through increasing the effectiveness of innovation investment, the knowledge from exploitative/exploratory MR also enhances innovation performance indirectly. In addition, the more types of knowledge from MR the firm has already been using in innovation development, the larger are the impacts of integrating additional knowledge from MR. In sum, knowledge from exploitative MR, knowledge from exploratory MR, and innovation investment improve product innovation performance in a complementary way.

These findings contribute to the scarce empirical literature that systematically assesses the implications of the knowledge from MR for corporate product innovation. In general, the findings support the broader notion that firms should be open to diverse external sources during
innovation (Laursen and Salter, 2006; Vanhaverbeke, Chesbrough, and West, 2008). In particular, this study is closely related to previous studies which examine the effects of learning knowledge about users either as a whole (e.g., Foss et al, 2011) or through a specific practice such as the lead user method (e.g. Lilien et al, 2002; Chatterji and Fabrizio, 2014). In addition, this study focuses on how innovation performance changes with integrating knowledge acquired from MR into early stages of new product development. So it complements with previous studies which explain why customer relationship management influences new product performance (Ernst et al., 2011) and why sales, marketing and R&D cooperation has a positive effect on new product performance (e.g. Ernst et al., 2010).

For management practitioners, this study suggests an effective and concrete step leading to better innovation performance: to integrate both the knowledge acquired from exploitative MR and the knowledge from exploratory MR into early stages of innovation and new product development (such as concept design). It usually pays off to listen to users and address their needs when conceptualizing the new product. This practice may also make innovation investment more effective. To some extent, market research is also a kind of R&D, which contributes to innovation. Therefore, policy makers who would like to accelerate product innovation may consider encouraging the learning, sharing and using of the knowledge from the market in both public and private R&D and innovation development.

This study has several limitations. First, the key explanatory variables, which indicate the use of knowledge from exploitative/exploratory MR, are binary rather than continuous. Due to this limitation, the estimators are not informative about how innovation performance responds to the degree of integrating knowledge from MR. Second, the high dimensional problem does not allow for estimating the significant levels of complementary effects. Although the differences in conditional marginal effects among subgroups of observations suggest the presence of complementary relations between knowledge from exploratory/exploitative MR and effectiveness of innovation investment, it would be more informative to measure the significant levels of the complementarities by including a group of dummy indicators into the regressions, just as Leiponen (2005). However, these issues have to be left for the future, when the data becomes rich enough to estimate models of high dimension. Third, the limitation of time window does not allow us to examine the long-run effects of using knowledge from
explorative/exploitative MR. When more years of data become available, it will become possible to examine how the returns of incorporating user’s opinion distribute along a longer time line. In addition, it would be interesting to examine the sequential and dynamic dimensions of integrating user knowledge, e.g., to examine whether the sequence of integrating knowledge from different MR practices matters. Finally, this study does not distinguish radical innovation from incremental innovation. Because knowledge acquired from explorative/exploitative MR may affect radical and incremental innovation in different ways, it is possible to get more interesting findings by examining the effects of MR on each type of innovation separately.

ACKNOWLEDGEMENTS

Sincere appreciation goes to Tor Eriksson, Frederic Warzynski, Keld Laursen, Ulrich Kaiser, John van Reeneen, Ina Drejer, Hsing-fen Lee, Gregor Pfeifer for their very insightful comments and constructive advices. I also thank participants at the Druid15 conference, the Ce2 workshop and the WIEM conference for their inspiring questions and discussions.

Financial support from Sino-Danish Center for Education and Research, Department of Economics and Business Economics, Department of Business Development and Technology, and Arctic Research Center at Aarhus University is gratefully acknowledged.
### Table 1. Descriptive Statistics

<table>
<thead>
<tr>
<th>Variables</th>
<th>Whole sample</th>
<th>Ignore user knowledge from MR</th>
<th>Integrate knowledge from exploitative MR</th>
<th>Integrate knowledge from exploratory MR</th>
<th>Integrate knowledge from both MR types</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
</tr>
<tr>
<td>Ln (1 + new products sales/total revenue)</td>
<td>0.11</td>
<td>0.22</td>
<td>0.04</td>
<td>0.14</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>0.11</td>
<td>0.21</td>
<td>0.11</td>
<td>0.21</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>0.11</td>
<td>0.21</td>
<td>0.21</td>
<td>0.26</td>
<td></td>
</tr>
<tr>
<td>Exploratory market research</td>
<td>0.28</td>
<td>0.45</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Exploitative market research</td>
<td>0.69</td>
<td>0.46</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
</tr>
<tr>
<td>R&amp;D expenditure (1000,000 DKK)</td>
<td>18154</td>
<td>218598</td>
<td>18523</td>
<td>326058</td>
<td>5204794</td>
</tr>
<tr>
<td></td>
<td>18523</td>
<td>326058</td>
<td>32550</td>
<td></td>
<td>86397</td>
</tr>
<tr>
<td></td>
<td>86397</td>
<td>712381</td>
<td>33859</td>
<td>165735</td>
<td></td>
</tr>
<tr>
<td>Ln(Innovation expenditure, 1000,000 DKK)</td>
<td>2.72</td>
<td>3.42</td>
<td>1.10</td>
<td>2.43</td>
<td>2.63</td>
</tr>
<tr>
<td></td>
<td>2.63</td>
<td>3.29</td>
<td>4.01</td>
<td>3.65</td>
<td>4.54</td>
</tr>
<tr>
<td></td>
<td>4.01</td>
<td>3.65</td>
<td>4.54</td>
<td>3.62</td>
<td></td>
</tr>
<tr>
<td>Diversity of R&amp;D partners</td>
<td>0.32</td>
<td>0.87</td>
<td>0.14</td>
<td>0.60</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>0.22</td>
<td>0.70</td>
<td>0.51</td>
<td>1.02</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>0.51</td>
<td>1.02</td>
<td>0.66</td>
<td>1.21</td>
<td></td>
</tr>
<tr>
<td>Ln(number of employees, other type)</td>
<td>3.98</td>
<td>1.72</td>
<td>3.91</td>
<td>1.66</td>
<td>3.83</td>
</tr>
<tr>
<td></td>
<td>3.83</td>
<td>1.63</td>
<td>4.57</td>
<td>2.02</td>
<td>4.24</td>
</tr>
<tr>
<td></td>
<td>4.57</td>
<td>2.02</td>
<td>4.24</td>
<td>1.85</td>
<td></td>
</tr>
<tr>
<td>Ln(asset, 1000,000 DKK)</td>
<td>10.97</td>
<td>2.11</td>
<td>10.79</td>
<td>1.97</td>
<td>1.08</td>
</tr>
<tr>
<td></td>
<td>1.08</td>
<td>2.01</td>
<td>11.78</td>
<td>2.26</td>
<td>11.44</td>
</tr>
<tr>
<td></td>
<td>11.44</td>
<td>2.26</td>
<td>11.44</td>
<td>2.30</td>
<td></td>
</tr>
<tr>
<td>Number of Observations</td>
<td>5344</td>
<td>1543</td>
<td>2281</td>
<td></td>
<td>97</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1423</td>
</tr>
</tbody>
</table>
Table 2. Impacts of the knowledge from exploratory/exploitative market research on product innovation performance: Estimates from fixed effect and CRE Tobit models

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>Fixed Effects Model</th>
<th>CRE Tobit Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sale of new products (percentage of total revenue)</td>
<td>Coeff. (S.E.)</td>
<td>Coeff. (S.E.)</td>
</tr>
<tr>
<td>Knowledge about users acquired from:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exploratory MR (advanced marketing methods)</td>
<td>0.038*** (0.012)</td>
<td>0.096*** (0.027)</td>
</tr>
<tr>
<td>Exploitative MR (regular customer contact)</td>
<td>0.034*** (0.011)</td>
<td>0.079** (0.037)</td>
</tr>
<tr>
<td>Exploratory MR (lag)</td>
<td>-0.015 (0.011)</td>
<td>-0.033 (0.024)</td>
</tr>
<tr>
<td>Exploitative MR (lag)</td>
<td>0.014 (0.011)</td>
<td>0.052 (0.033)</td>
</tr>
<tr>
<td>R&amp;D expenditure (1000.000 DKK)</td>
<td>-9.95E-08 (2.02E-07)</td>
<td>-9.4E-09 (2.60E-07)</td>
</tr>
<tr>
<td>R&amp;D expenditure (lag) (1000.000 DKK)</td>
<td>3.09E-07* (1.85E-07)</td>
<td>3.18E-07 (2.99E-07)</td>
</tr>
<tr>
<td>Ln(Innovation expenditure, 1000.000 DKK)</td>
<td>0.008*** (0.002)</td>
<td>0.017*** (0.003)</td>
</tr>
<tr>
<td>Ln(Innovation expenditure, 1000.000 DKK) (lag)</td>
<td>0.003** (0.001)</td>
<td>0.005* (0.003)</td>
</tr>
<tr>
<td>Salary level</td>
<td>-1.23E-07 (1.08E-07)</td>
<td>-2.12E-07 (2.23E-07)</td>
</tr>
<tr>
<td>Diversity of R&amp;D partners</td>
<td>0.026*** (0.008)</td>
<td>0.027** (0.012)</td>
</tr>
<tr>
<td>Ln(number of employees, other type)</td>
<td>0.009 (0.018)</td>
<td>0.039 (0.054)</td>
</tr>
<tr>
<td>Ln(asset, 1000.000 DKK)</td>
<td>-2.81E-04 (1.26E-02)</td>
<td>-0.032 (0.048)</td>
</tr>
<tr>
<td>Salary level (lag)</td>
<td>2.49E-07*** (8.17E-08)</td>
<td>2.71E-07* (1.65E-07)</td>
</tr>
<tr>
<td>Diversity of R&amp;D partners (lag)</td>
<td>-0.002 (0.007)</td>
<td>-0.001 (0.010)</td>
</tr>
<tr>
<td>Ln(number of employees, other type) (lag)</td>
<td>0.003 (0.017)</td>
<td>0.056 (0.053)</td>
</tr>
<tr>
<td>Ln(asset, 1000.000 DKK) (lag)</td>
<td>-0.020** (0.010)</td>
<td>-0.095** (0.049)</td>
</tr>
<tr>
<td>Pre-sample mean of % revenue from new products: 2001-2006</td>
<td>---</td>
<td>0.002*** (4.46E-04)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.170 (0.183)</td>
<td>-0.741 (0.136)</td>
</tr>
<tr>
<td>F Statistic/ Wald chi2</td>
<td>F(18,3149)=9.14</td>
<td></td>
</tr>
<tr>
<td>Log pseudo likelihood</td>
<td>---</td>
<td>-866.8783</td>
</tr>
<tr>
<td>Number of observations</td>
<td>5344</td>
<td>left-censored: 1219; uncensored: 961</td>
</tr>
<tr>
<td>Number of firms</td>
<td>3150</td>
<td>1060</td>
</tr>
</tbody>
</table>

***: Significant at 1%; **: Significant at 5%; *: Significant at 10%. Panel bootstrap standard errors for CRE model (400 repetitions).
Table 3. Impacts of the knowledge from exploratory/exploitative market research on product innovation performance: Estimates from nearest neighbor matching method under dynamic ignorability assumption

<table>
<thead>
<tr>
<th>Dependent Variable:</th>
<th>Treatment - Using the Knowledge Acquired from:</th>
<th>Number of Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sale of new products (percentage of total revenue)</td>
<td>Exploratory Market Research</td>
<td>Exploitative Market Research</td>
</tr>
<tr>
<td>ATT in each Year</td>
<td>Observed Coef.</td>
<td>S.E.</td>
</tr>
<tr>
<td>2008</td>
<td>0.0814***</td>
<td>(0.0208)#</td>
</tr>
<tr>
<td>2009</td>
<td>0.0734***</td>
<td>(0.0189)#</td>
</tr>
<tr>
<td>2010</td>
<td>0.0813***</td>
<td>(0.0156)#</td>
</tr>
<tr>
<td>Weighted ATT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td>0.0786***</td>
<td>(0.0110)+</td>
</tr>
<tr>
<td>ATE in each Year</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td>0.1093***</td>
<td>(0.0335)#</td>
</tr>
<tr>
<td>2010</td>
<td>0.0519***</td>
<td>(0.0189)#</td>
</tr>
<tr>
<td>Weighted ATE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td>0.0979***</td>
<td>(0.0376)#</td>
</tr>
<tr>
<td>2010</td>
<td>0.0849***</td>
<td>(0.0158)+</td>
</tr>
</tbody>
</table>

***: Significant at 1%;
**: Significant at 5%;
*: Significant at 10%.
#: AI Robust standard errors
+: Panel bootstrap standard errors: 400 repetitions, seed 1000.
Distance metric: Mahalanobis
Table 4. Implication of knowledge from exploratory/exploitative market research on new product innovation for subsamples: Heterogeneous marginal effects based on CRE Tobit estimates

| Knowledge acquired from: | A.M.E. on E(y*|y>0) | For observations that are using the knowledge acquired from: |
|--------------------------|----------------------|----------------------------------------------------------|
|                          | Neither methods | Exploitative MR | Exploratory MR | Both methods |
| Share of new product sales (in total revenue) | | | | |
| Exploratory market research | 0.018*** (0.006) | 0.044*** (0.012) | 0.0305*** (0.008) | 0.056*** (0.014) |
| Exploitative market research | 0.015** (0.007) | 0.029** (0.012) | 0.0310** (0.014) | 0.047** (0.019) |
| Exploratory market research (lag) | -0.005 (0.003) | -0.013 (0.009) | -0.012 (0.008) | -0.021 (0.014) |
| Exploitative market research (lag) | 0.009 (0.006) | 0.020 (0.013) | 0.019 (0.012) | 0.032 (0.020) |
| R&D expenditure (1000.000 DKK) | -1.51E-09 (4.04E-08) | -3.81E-09 (1.02E-07) | -3.37E-09 (9.02E-08) | -5.94E-09 (1.59E-07) |
| R&D expenditure (lag) (1000.000 DKK) | 5.12E-08 (4.85E-08) | 1.29E-07 (1.20E-07) | 1.14E-07 (1.03E-07) | 2.01E-07 (1.87E-07) |
| Ln(Innovation expenditure, 1000.000 DKK) | 0.003*** (0.001) | 0.007*** (0.001) | 0.006*** (0.001) | 0.011*** (0.002) |
| Ln(Innovation expenditure, 1000.000 DKK) (lag) | 0.001 (0.001) | 0.002 (0.001) | 0.002 (0.001) | 0.003 (0.002) |
| Salary level | -3.42E-08 (3.38E-08) | -8.63E-08 (8.53E-08) | -7.63E-08 (7.69E-08) | -1.34E-07 (1.33E-07) |
| Diversity of R&D partners | 0.004*** (0.002) | 0.011*** (0.005) | 0.010*** (0.005) | 0.017*** (0.008) |
| Ln(number of employees, other type) | 0.006 (0.009) | 0.016 (0.022) | 0.014 (0.020) | 0.024 (0.035) |
| Ln(asset, 1000.000 DKK) | -0.005 (0.008) | -0.013 (0.020) | -0.012 (0.017) | -0.020 (0.030) |
| Salary level (lag) | 4.36E-08 (2.83E-08) | 1.10E-07 (7.10E-08) | 9.75E-08 (6.41E-08) | 1.710E-07 (1.11E-07) |
| Diversity of R&D partners (lag) | -1.85E-04 (0.002) | -4.67E-04 (0.004) | -4.129E-04 (0.004) | -0.001 (0.006) |
| Ln(number of employees, other type) (lag) | 0.009 (0.009) | 0.023 (0.021) | 0.020 (0.019) | 0.035 (0.033) |
| Ln(asset, 1000.000 DKK) (lag) | -0.015 (0.008) | -0.039 (0.019) | -0.034 (0.017) | -0.060 (0.029) |
| Pre-sample mean of y (% revenue from new products: 2001-2006) | 2.95E-04 (8.38E-05) | 7.46E-04 (1.85E-04) | 6.60E-04 (1.78E-04) | 1.16E-03 (2.88E-04) |

***: Significant at 1%; **: Significant at 5%; *: Significant at 10%. Panel bootstrap standard errors for CRE models (400 repetitions).
References


PhD Theses since 1 July 2011

2011-4  Anders Bredahl Kock: Forecasting and Oracle Efficient Econometrics
2011-5  Christian Bach: The Game of Risk
2011-6  Stefan Holst Bache: Quantile Regression: Three Econometric Studies
2011:12 Bisheng Du: Essays on Advance Demand Information, Prioritization and Real Options in Inventory Management
2011:13 Christian Gormsen Schmidt: Exploring the Barriers to Globalization
2011:16 Dewi Fitriasari: Analyses of Social and Environmental Reporting as a Practice of Accountability to Stakeholders
2011:22 Sanne Hiller: Essays on International Trade and Migration: Firm Behavior, Networks and Barriers to Trade
2012-1 Johannes Tang Kristensen: From Determinants of Low Birthweight to Factor-Based Macroeconomic Forecasting
2012-2 Karina Hjortshøj Kjeldsen: Routing and Scheduling in Liner Shipping
2012-3 Soheil Abginehchi: Essays on Inventory Control in Presence of Multiple Sourcing
2012-4 Zhenjiang Qin: Essays on Heterogeneous Beliefs, Public Information, and Asset Pricing
2012-5 Lasse Frisgaard Gunnersen: Income Redistribution Policies
2012-6 Miriam Wüst: Essays on early investments in child health
2012-7 Yukai Yang: Modelling Nonlinear Vector Economic Time Series
2012-9 Henrik Nørholm: Structured Retail Products and Return Predictability
2012-10 Signe Frederiksen: Empirical Essays on Placements in Outside Home Care
2012-11 Mateusz P. Dziubinski: Essays on Financial Econometrics and Derivatives Pricing
2012-12 Jens Riis Andersen: Option Games under Incomplete Information
2012-14 Laurent Callot: Large Panels and High-dimensional VAR
2012-15 Christian Rix-Nielsen: Strategic Investment
2013-1 Kenneth Lykke Sørensen: Essays on Wage Determination
2013-2 Tue Rauff Lind Christensen: Network Design Problems with Piecewise Linear Cost Functions
2013-4 Rune Bysted: Essays on Innovative Work Behavior
2013-5 Mikkel Nørlem Hermansen: Longer Human Lifespan and the Retirement Decision
2013-7 Mark Strøm Kristoffersen: Essays on Economic Policies over the Business Cycle
2013-8 Philipp Meinen: Essays on Firms in International Trade
2013-9 Cédric Gorinas: Essays on Marginalization and Integration of Immigrants and Young Criminals – A Labour Economics Perspective
2013-12 Paola Andrea Barrientos Quiroga: Essays on Development Economics
2013-13 Peter Bodnar: Essays on Warehouse Operations
2013-14 Rune Vammen Lesner: Essays on Determinants of Inequality
2013-15 Peter Arendorf Bache: Firms and International Trade
2013-16 Anders Laugesen: On Complementarities, Heterogeneous Firms, and International Trade
Anders Bruun Jonassen: Regression Discontinuity Analyses of the Disincentive Effects of Increasing Social Assistance

David Sloth Pedersen: A Journey into the Dark Arts of Quantitative Finance

Martin Schultz-Nielsen: Optimal Corporate Investments and Capital Structure

Lukas Bach: Routing and Scheduling Problems - Optimization using Exact and Heuristic Methods

Tanja Groth: Regulatory impacts in relation to a renewable fuel CHP technology: A financial and socioeconomic analysis

Niels Strange Hansen: Forecasting Based on Unobserved Variables

Ritwik Banerjee: Economics of Misbehavior

Christina Annette Gravert: Giving and Taking – Essays in Experimental Economics

Astrid Hunghøj: Papers in purchasing and supply management: A capability-based perspective

Nima Nonejad: Essays in Applied Bayesian Particle and Markov Chain Monte Carlo Techniques in Time Series Econometrics

Tine L. Mundbjerg Eriksen: Essays on Bullying: an Economist’s Perspective

Sashka Dimova: Essays on Job Search Assistance

Rasmus Tangsgaard Varneskov: Econometric Analysis of Volatility in Financial Additive Noise Models

Anne Floor Brix: Estimation of Continuous Time Models Driven by Lévy Processes

Kasper Vinther Olesen: Realizing Conditional Distributions and Coherence Across Financial Asset Classes

Manuel Sebastian Lukas: Estimation and Model Specification for Econometric Forecasting

Sofie Theilade Nyland Brodersen: Essays on Job Search Assistance and Labor Market Outcomes

Jesper Nydam Wulff: Empirical Research in Foreign Market Entry Mode
2015-6  Sanni Nørgaard Breining: The Sibling Relationship Dynamics and Spillovers
2015-7  Marie Herly: Empirical Studies of Earnings Quality
2015-8  Stine Ludvig Bech: The Relationship between Caseworkers and Unemployed Workers
2015-9  Kaleb Girma Abreha: Empirical Essays on Heterogeneous Firms and International Trade
2015-10 Jeanne Andersen: Modelling and Optimisation of Renewable Energy Systems
2015-11 Rasmus Landersø: Essays in the Economics of Crime
2015-12 Juan Carlos Parra-Alvarez: Solution Methods and Inference in Continuous-Time Dynamic Equilibrium Economies (with Applications in Asset Pricing and Income Fluctuation Models)
2015-13 Sakshi Girdhar: The Internationalization of Big Accounting Firms and the Implications on their Practices and Structures: An Institutional Analysis
2015-14 Wenjing Wang: Corporate Innovation, R&D Personnel and External Knowledge Utilization