



RESEARCH PAPER 2/2017

The role of spillovers in research and development expenditure in Australian industries

Sasan Bakhtiari¹

Office of the Chief Economist
Department of Industry, Innovation and Science

Robert Breunig

Crawford School of Public Policy
Australian National University

April 2017

Abstract

Using administrative data from firms in Australia that conduct research and development (R&D), we examine how R&D activity of other firms and public institutions affect a firm's own R&D expenditure. We distinguish between the impact of peers, suppliers and clients. We examine whether geographical proximity and industrial clustering affect R&D spillovers. Overall, we detect positive effects on R&D expenditure from spillovers from peers and clients to firms that are nearby; within 25 or 50 km. R&D expenditure by academia, unlike by government bodies, has a positive influence on a firm's own R&D expenditure within state boundaries. We fail to find any significant role for industrial clusters in augmenting spillover effects.

JEL Codes: D22, O31, O33, R12

Keywords: Research and Development, Innovation, R&D Spillovers, Supply Chains, Industrial Clustering



For further information on this research paper please contact:

Name: Sasan Bakhtiari

Section: Industry and Firm Analysis

Department of Industry, Innovation and Science

GPO Box 9839

Canberra ACT 2601

Phone : +61 2 9397 1639

Email: Sasan.Bakhtiari@industry.gov.au

Disclaimer

The views expressed in this report are those of the author(s) and do not necessarily reflect those of the Australian Government or the Department of Industry, Innovation and Science or the Australian government. Use of any results from this paper should clearly attribute the work to the authors and not to the department or the government.

© Commonwealth of Australia 2017.

This work is copyright. Apart from use under Copyright Act 1968, no part may be reproduced or altered by any process without prior written permission from the Australian Government. Requests and inquiries concerning reproduction and rights should be addressed to chiefeconomist@industry.gov.au. For more information on Office of the Chief Economist research papers please access the Department's website at: www.industry.gov.au/OCE



Creative Commons Licence

With the exception of the Coat of Arms, this publication is licensed under a Creative Commons Attribution 3.0 Australia Licence.

Creative Commons Attribution 3.0 Australia Licence is a standard form license agreement that allows you to copy, distribute, transmit and adapt this publication provided that you attribute the work. A summary of the licence terms is available from <http://creativecommons.org/licenses/by/3.0/au/deed.en>. The full licence terms are available from <http://creativecommons.org/licenses/by/3.0/au/legalcode>.

The Commonwealth's preference is that you attribute this publication (and any material sourced from it) using the following wording:

Source: Licensed from the Commonwealth of Australia under a Creative Commons Attribution 3.0 Australia Licence. The Commonwealth of Australia does not necessarily endorse the content of this publication.

Key points

- The empirical evidence presented in this paper points to R&D spillovers between Australian firms.
- R&D spillovers are the strongest between geographically proximate firms.
- The results also point to a positive influence from academia on firms' R&D.
- R&D expenditures are higher in industrial clusters.
- However, there is no evidence of any extra spillovers within industrial clusters.

1. Introduction

Research and development (R&D) play a central role in long-run productivity and economic growth. Theory suggests that R&D spillovers (where the R&D activity of a firm affects the well-being of consumers or the profitability of other firms) also play an important role in economic growth and that the benefits of R&D extend well beyond the firm that makes the R&D investment. Thus, the social returns to R&D as a whole may be greater than the sum of the private returns to firms who make R&D investment decisions.

The existence of R&D spillovers may also increase the incentive for firms to invest in R&D if other firms' R&D is complementary — that is if it makes a firm's own R&D more productive. This could happen if spillovers from other firms make a firm's R&D more likely to succeed or if knowledge from other firms' R&D combines with an individual firm's R&D to increase the returns to a firm's own R&D expenditure.

However, the existence of R&D spillovers also has the potential to disincentivise firms from investing in R&D. The partially public nature of knowledge and competition in markets reduces the firm's ability to appropriate rents from their innovative activities. This may lead firms to reduce their R&D expenditure. A priori, it is not known whether the positive or negative effects of spillovers on firm-level R&D expenditure will dominate.

Recognising the latter possibility, governments around the world offer a range of incentives such as patents and licenses (that grant a temporary monopoly to the inventor) or R&D grants and subsidies. These measures compensate for the lack of incentive for firms to invest in R&D.

In this paper we focus on the R&D expenditure decisions of individual firms and how they are affected by the R&D activity of other firms. We find that overall, the negative disincentives dominate. The presence of spillovers results in firms making less R&D investment than they otherwise would.

However, both distance and relationship matter. For peers and clients, we find a positive role for proximity. After trying a few discrete radii, we find that spillovers from peers within 25 km, and for clients within 50 km, result in higher R&D expenditure. For suppliers, we find that spillover effects are always negative but less strongly so at greater distances. This gives us insight into which types of spillovers might be most important for which types of relationships, as we discuss below.

We also test for the role of industrial clustering in spillovers. We find that R&D activity is on average higher in industrial clusters but that clustering does not amplify the effect of spillovers on R&D expenditure. For public sources of R&D expenditure, we find that higher education expenditure has a positive influence on firm-level R&D expenditure. Direct government spending on research seems to crowd out private R&D expenditure.

In the next section we elaborate on the conceptual background of our approach. We provide a short literature review in Section 3 and a discussion of our data in Section 4. Our model and methodology are described in Sections 5 and 6. Our results are presented in Section 7 with the geographical refinements

presented in 7.2 and the results on industrial clustering discussed in 7.3. We conclude in Section 8.

2. Background

In the presence of spillovers, the decisions of individual firms about how much R&D activity to undertake may be influenced by the R&D decisions of other firms. However, the relationship is ambiguous, as noted above.

We use a simple model where a firm's future R&D expenditure depends on its current sales, R&D expenditures, and the number of the firm's own R&D staff. As in Bloom et al. (2013), we assume that R&D staff within a firm are the vehicle by which spillovers are translated into more or less R&D activity. We treat R&D staff in other firms as both the source of knowledge about other firms' R&D activities and as a proxy for their R&D activity. Firms in our data are geo-coded, and we use the geographic coordinates of firms to test whether R&D spillovers are affected by distance.

We classify firms as peers (those within the same industry), suppliers, or clients to partially account for technological similarities and the type of spillover (see below). We also control for the impact of R&D activity by academia and government institutions on the R&D expenditure of private firms.

We differ from most prior empirical research which has focused on the effect of one firm's R&D expenditure on another firm's productivity or level of output. Our approach is closest to that of Jaffe (1988), who looks at the productivity of R&D expenditure in the presence of spillovers.

There are four reasons why we think our focus on the effect of spillovers on R&D expenditure is interesting: First, follow-up R&D is a natural output of newly acquired knowledge which may complement a firm's own knowledge. Such combining of new knowledge with firm investment may help to increase the productivity of R&D activity. Changes to output or productivity, on the other hand, are an indirect and uncertain consequence of innovation, new knowledge, and spillovers. Innovation may or may not lead to increases in productivity when one considers that innovation encompasses process, organisational and product innovations. Product innovation, in particular, does not necessarily enhance productivity. On the contrary, there are cases where it will be detrimental to the firm's productivity (Schoar, 2002). Our approach is to look directly at the effect of spillovers on firm-specific R&D expenditure. It encompasses the effect of R&D on all types of innovation.

Second, and related to the above, is that our approach reflects that even in the presence of free-flowing knowledge, firms have to make an effort to effectively assimilate the knowledge and make it useful for the firm. The presence of R&D staff and expenditure on R&D makes this possible.

Third, we can separate out spillover effects by peers (competitors), suppliers and clients. The conventional definition of spillover suffers from an identifiability problem when the type of relationship (horizontal versus vertical) is taken into account: R&D by a supplier that results in a cheaper input reflects as a surge in the productivity of the downstream firm. This productivity effect in general cannot be decoupled from the actual spillover effect and estimates of the size

of spillovers will be biased upwards (Kenta et al., 2015). Since we are looking directly at R&D expenditure of firms, we do not suffer from this problem.

Fourth, most government programmes in this domain have as an explicit goal to increase firm-level R&D expenditure. A better understanding of the relationship between spillovers and firm-level R&D expenditure is thus crucial to designing and evaluating such policies.

Spillovers can happen in a variety of ways. Three distinct types of spillovers have been recognised in the literature. Jaffe (1998) identifies knowledge spillovers, market spillovers, and network spillovers (see also Jones & Williams (1998)). All three types of spillovers create a gap between the private rate of return to R&D and the social rate of return which includes benefits to both other firms and consumers.

The work of Jaffe et al. (1993) and others has tried to sort out the differences between knowledge, market, and network spillovers by using the location of firms in technology or product market space. The focus has usually been on firm productivity or output or innovative activity such as patenting.

For all three types of spillovers — knowledge, market or network — the effect on a firm's R&D expenditure may be either positive or negative. Knowledge spillovers can be either complements or substitutes for a firm's own R&D activity. Depending upon the nature of the knowledge and the spillover, they could affect the R&D activity of any type of firm — competitors, clients, or suppliers. Market spillovers will primarily affect peers and downstream firms. Market competition may keep prices from rising to fully reflect the value of the innovation, thus consumers and other firms that purchase from a firm will benefit from these spillovers. Market spillovers may be either complementary to other firms' R&D activities or make them less valuable. Theory suggests that spillovers between firms in a supplier-client relationship could have a network effect that boosts the value of an innovation when it is concurrent with similar innovations in the supply chain (Jaffe, 1998). This may make R&D more productive within the network. For all three types of spillovers, there are pecuniary externalities that impose a downward pressure on prices due to the firm's imperfect monopoly over the new innovation, which could act as a disincentive to engage in R&D activity.

Given that each mechanism affects R&D in a firm in multiple and potentially opposing way, the final direction of R&D spillovers on R&D expenditure is ambiguous. The sign of the overall R&D spillover effect will depend upon which externality dominates within different groups of firms (peers, clients and suppliers) and may be determined by their geographic positioning. We allow for both.

Finally, industrial clusters are potential loci of spillovers owing to the co-location of similar and often competing firms. Such clusters may feature stronger inter-personal professional relationships, and might facilitate talent poaching and collaboration. We test whether such clusters affect firm-level R&D expenditure in Australia.

One drawback of our approach (and the limitations of our data) is that we cannot disentangle the effect of different types of spillovers. Nonetheless, we think it is informative to estimate the total overall effect of spillovers on R&D

expenditure and to see how this varies by relationship (peer, client, supplier), geographical proximity, and industrial clustering.

Next, we briefly review the literature before turning to the details of our data and results.

3. Literature

In the search for R&D spillovers, Griliches (1992) concludes that:

“R&D spillovers are present, their magnitude may be quite large, and social rates of return remain significantly above private rates.”

Griliches (1979) emphasizes the distance aspect in R&D spillover. Outside stock of useful knowledge is traditionally represented as the aggregate sum of R&D activity of firms in an economy or industry. The sum can be formed by weighting all firms uniformly. More coherently, every firm can be weighted by a *distance* measure to the firm that is the potential recipient to account for diffusion chances. Distance in physical location, technological fields and product space have been considered and shown to matter (Jaffe, 1988; Bloom et al., 2013; Chatterji et al., 2013). Audretsch & Feldman (1996) and Ellison et al. (2010) go one step further and suggest that one reason firms cluster geographically in the first place is to benefit from increased knowledge spillovers.

Kenta et al. (2015) use detailed firm-level data to identify vertical relations among Japanese companies and find positive spillovers emanating from both suppliers and client firms. They are also able to show that spillovers are localized within similar firms but do not detect any geographic aspect to spillovers from vertical chains. However, their study is limited by the fact that they can identify vertical relations only within a one-year cohort.

The positive impact of spillovers, especially from nearby firms, has been found with data from many countries including the US (Lychagin et al., 2016), France (Mairesse & Mulkay, 1991), Europe (Bottazzi & Peri, 2003), and Japan (Kenta et al., 2015). There is a dearth of comparable evidence in Australia mainly due to the lack of appropriate firm-level data. The only relevant work that the authors are aware of is Elnasri & Fox (2014); they use aggregate time series data and find no impact on national-level productivity of direct government spending on innovation. They also fail to find any impact of indirect spending such as tax incentives on innovation.

Most of the existing evidence considers spillovers as productivity boosting without specifying a direct channel through which that happens (see, for instance, Lychagin et al., 2016; Bloom et al., 2013; Belderbos et al., 2013). Alternatively, other papers look at the impact of R&D spillovers on patenting (see Jaffe, 1986; Jaffe et al., 1993).

We now turn to a detailed description of our data followed by our econometric model.

4. Data

Our main source of firm-level data is the Australian Department of Industry's R&D Tax Concession programme (see Appendix A). These data have been collected since the Australian government enacted a R&D Tax Concession programme in 1985 with the goal of promoting R&D expenditure and innovation among Australian businesses, especially small businesses, by offering them a tax rebate payable on expenditures for research and development projects. Businesses qualify if total annual expenditure is \$20,000 or above. As part of this programme, the government gathers data from all recipient firms in each year, and the resulting panel is the basis of our study. By the Department of Industry's estimate, the data cover more than 87 per cent of R&D active firms in Australia.

Note that our data omit firms who do not engage in R&D activity. The absence of these firms does not detract from our ability to answer our research question. Our focus is on the R&D activity of other firms spilling over to a firm that is doing R&D, and this dataset provides a comprehensive source from which to answer our research question.

We use data for the period 2001 to 2011, a stable period for the programme with no major regulatory changes.¹ Starting in July 2011, the programme was replaced by the R&D Tax Incentive and programme settings changed substantially.²

The data contain information on annual turnover, annual R&D expenditure, number of employees and number of R&D staff. The reported number of employees is a head-count, whereas the number of R&D staff is reported in full-time equivalents. The data also contain the address of the firm and the country of headquarters, which we use to determine whether the firm is a subsidiary of a foreign multinational. For holding and foreign companies the address is specifically of the subsidiary that is registering the R&D expenditure and not the parent. The Department of Industry, Innovation and Science has provided us with addresses matched to the geo-coordinates of each firm.

The number of firms registered in the programme has grown over time, and as of 2011, there were 9,281 firms across Australia registered in the programme (Table 4.1). The panel of firms in the data is not balanced and there is a constant flow of entries into and exits from the programme in every year. Entrants out-number exits each year, causing the programme to expand in scale. 1,568 firms appear in all ten years of the data, whereas 2,140 appear in at least nine years. These numbers illustrate a large degree of continuity in the R&D activities of registered firms; however, there is a size bias and firms in the

¹ Note that years in this text are financial years, not calendar years. In Australia, the financial year begins on 1st of July and ends on 30th June of the following year. Throughout the paper, we refer to financial years based on when they end. Therefore, 2007 in this context means July 2006 to June 2007.

² See Appendix A and <https://www.ato.gov.au/Business/Research-and-development-tax-incentive/>

balanced panel are larger in size, on average, than the sample of firms in the unbalanced panel. For our analysis, we make use of the unbalanced panel.

Table 4.1 shows that about a quarter of exiting firms in every year re-appear in some later year. This pattern is suggestive of firms concluding an innovation project which causes R&D expenditure to fall below the \$20,000 programme threshold. The firm drops out of the programme until the commencement of a new project followed by re-registration in the programme. This pattern is typical of smaller firms; they tend to lack a dedicated research division and a constant flow of R&D expenditures.

Table 4.1: Annual count of firms, entries and exits.

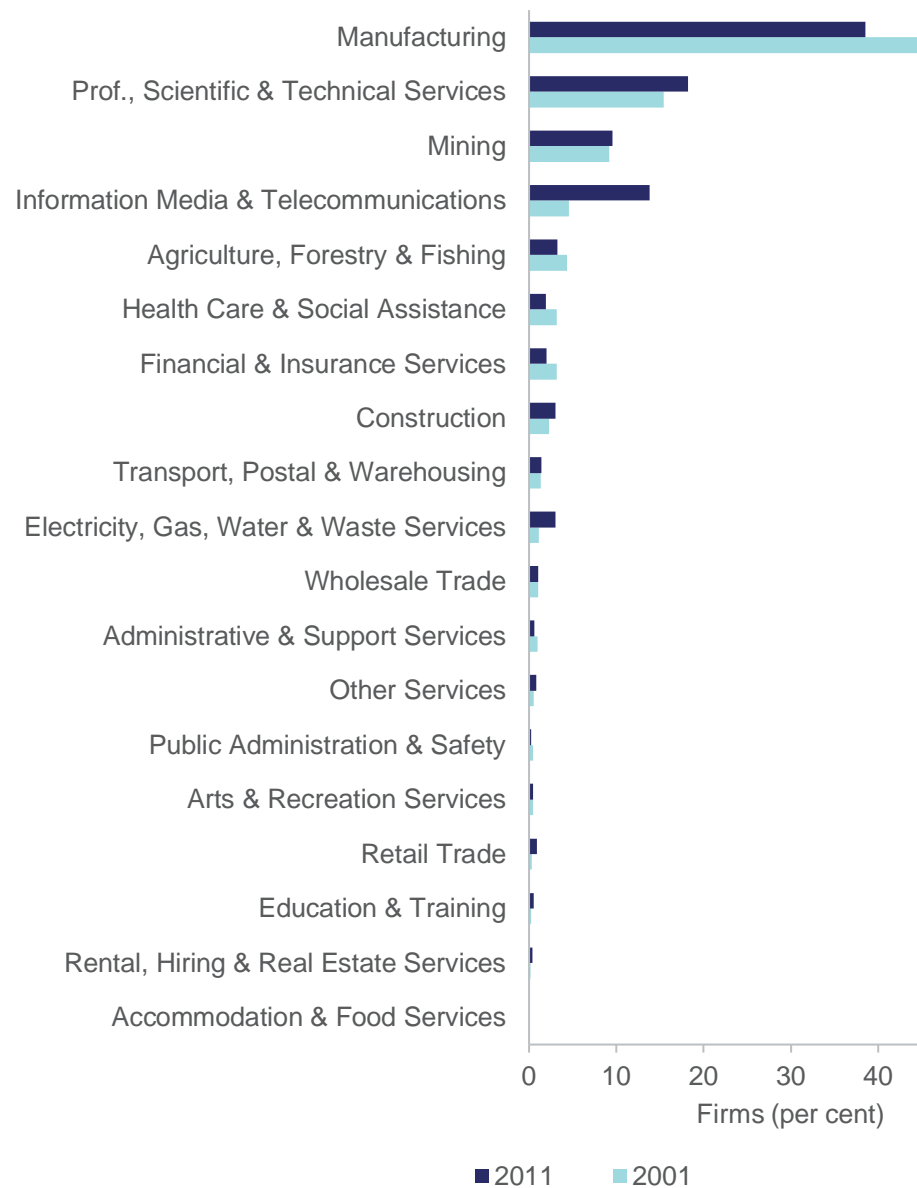
Year	Number of				
	Firms	Total Exits	Total Entries	Temporary Exits	Re-entries
2001	3,732	699	-	196	-
2002	4,755	1,050	1,722	261	-
2003	5,097	975	1,392	252	75
2004	5,646	1,160	1,524	278	129
2005	5,997	1,123	1,511	277	192
2006	6,421	1,183	1,547	282	236
2007	6,967	1,277	1,729	228	230
2008	7,911	1,537	2,221	225	326
2009	8,582	1,769	2,208	184	302
2010	8,767	1,755	1,954	-	302
2011	9,281		2,269	-	391
Total	73,156	12,528	18,077	2,183	2,183

Notes: Exit is a firm that is not observed the following year. Entry is a firm that was not observed the preceding year.

Source: Dept. of Industry, Innovation and Science (2016)

Firms in the data come from industries ranging from manufacturing to mining, construction and services. This diversity is illustrated in Figure 4.1, where we show the distribution of firms by main industry of operation in 2001 and 2011.

Figure 4.1: The distribution of firms in the R&D Tax Concession programme by industry in 2001 and 2011.



Source: Dept. of Industry, Innovation and Science (2016)

Manufacturing has the largest presence in the data; more than half of all registered firms in the year 2001 are manufacturing. The next largest groups are Professional, Scientific and Technical Services, Mining, and Information Media and Telecommunication. The latter group represents publishing, radio and television broadcasting but also software, internet, data processing, and telecommunication firms. Not surprisingly, the data reveal that about 78 per cent of registered firms in this sector specialize in digital media and telecommunication. Firms from all other industries together make up about 10 per cent of the cohort. However, these firms operate in sectors as diverse as Rental, Hiring, and Real Estate Services, Retail Trade and Construction.

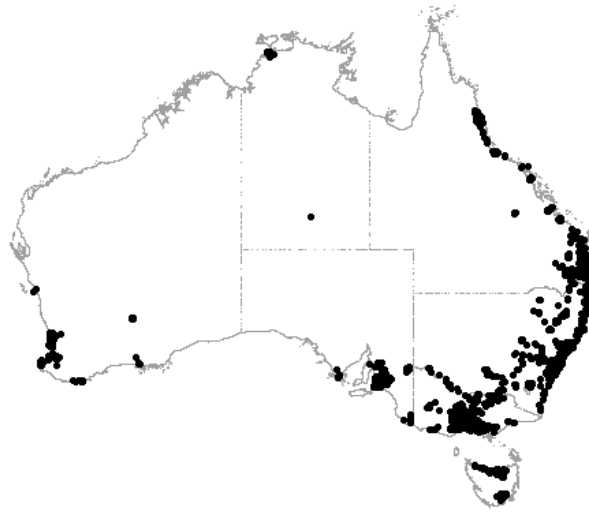
Comparing the distributions from 2001 and 2011 reveals some shifts in the industry composition. The top participants — Manufacturing, Scientific Services, Mining and Information Media — are still the largest participants; nonetheless, the proportion of manufacturing firms in the programme has dropped substantially during the decade, while scientific services and digital media have increased their shares. This trend is driven more by a shift in the relative prevalence of R&D activity in these industries than by changes in the number of firms in these industries, which in fact has increased.

There are also new industry groups in 2011. Accommodation and Food Services, for example, had no presence in 2001 but is represented by a small number of firms in 2011. The pattern as a whole indicates an increased diversity in the range of R&D expenditure undertaken in Australia and the shift, taking place in most advanced industrialized countries, away from manufacturing and towards services.

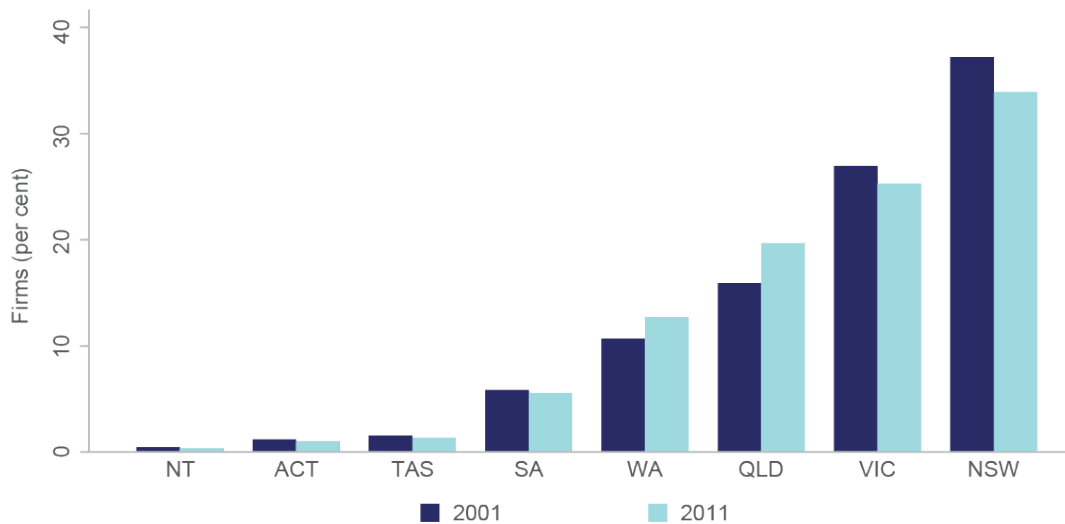
Figure 4.2(a) presents the geographical distribution of firms in the data. As is the case with the distribution of population in Australia, the economic and innovative activity is concentrated along the coastal areas and especially in the vicinity of metropolitan areas (Figure 4.2(a)). Of all the states and territories in Australia, the most populated are New South Wales (NSW), Victoria (VIC), and Queensland (QLD), and these three states account for more than 75 per cent of firms in our data (Figure 4.2 (b)). The shares of New South Wales and Victoria have declined slightly over time with a shift towards Queensland and Western Australia (WA). This trend matches a decline in manufacturing — NSW and VIC are major manufacturing hubs — in tandem with a boost in resource industries, primarily based in Queensland and Western Australia.

Figure 4.2: The geographic distribution of firms in the R&D Tax Concession programme (a) in Australia and (b) by state.

(a)



(b)



Notes: Due to the confidentiality of data, all firms with fewer than five neighbours within $\pm 1^\circ$ of coordinates are suppressed in panel (a) to protect individual identities.

Source: Dept. of Industry, Innovation and Science (2016)

4.1 Foreign Multinationals

We believe that the process by which R&D expenditure of subsidiaries of foreign multinationals is determined is substantially different from that of domestically-owned firms. Since we lack the means to properly measure international spillovers, we focus on explaining the R&D expenditure of domestic firms.

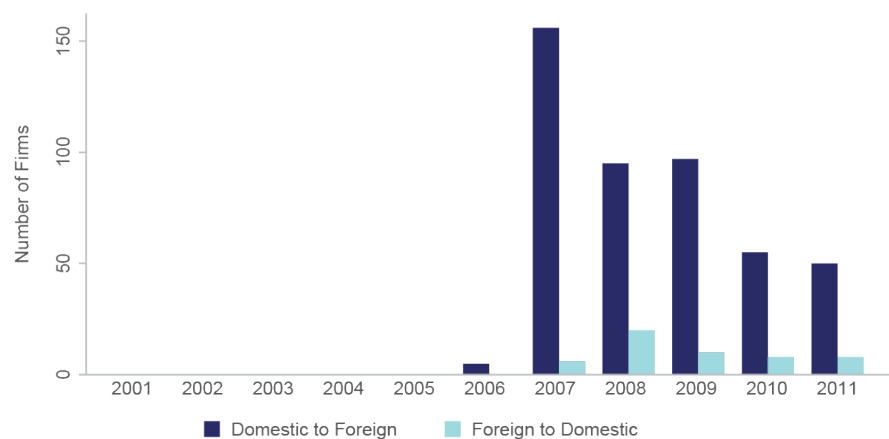
Despite the focus on domestic firms, it seems reasonable that the R&D activity of foreign-owned subsidiaries based in Australia could influence domestic firms. Therefore, we keep these observations in the data but we use them only as explanatory variables for the R&D expenditure of domestic firms. There is evidence that domestic firms can benefit from the presence of foreign

multinationals while the reverse is not necessarily true (Blomström & Kokko, 1998).

Complicating implementation is the non-trivial number of firms in our sample who change status from domestic to multinational or vice versa in the course of the eleven years (Figure 4.3). The number of firms changing status to foreign multinational peaks in 2007 — when restrictions were lifted on foreign multinationals' claiming the tax concession for their incremental R&D — but continues during the years that follow.³

Part of this activity is associated with the acquisition of Australian firms by foreign multinationals. The other part could be foreign subsidiaries that misrepresented themselves as Australian entities prior to 2007 in order to benefit from those aspects of the R&D tax concession programme unavailable to foreign multinationals. We are unable to distinguish between the two cases using the information available to us. For the purposes of our regression modelling, we classify firms as domestic if they declare an Australian headquarters in all years of the data in which they appear.

Figure 4.3: The number of firms that change headquarters from Australian domestic to foreign and vice versa by year.



Source: Dept. of Industry, Innovation and Science (2016)

To see if this affects our results, we repeat the estimation exercise and include all firms, domestic or foreign, as possible recipients of spillovers. We report the results in Appendix B. We do not detect any substantial change in our findings as a result of this addition.

³ In appendix A we explain the R&D Tax Concession program in detail. Foreign firms could benefit from two of the three elements of the program before 2007 and were allowed to access all aspects of the program after 2007. Somewhat surprisingly, we find no firms that change headquarters in either direction prior to 2006. We double-checked this in the data and with the Department of Industry and this appears to be correct.

5. Spillovers

The modelling and measurement of R&D spillovers is a key element of our study. We model inward spillovers to firm j from firm j' as depending upon the random encounters between the R&D staff of the two firms. This follows the spirit of Lychagin et al. (2016) and Bloom et al. (2013), although our implementation is somewhat different.

Our approach is to model the spillover possibilities for firm j averaged across all other firms (N in total) in the economy as

$$\begin{aligned}\overline{SPILL}_{jt} &= \frac{1}{N_{t-1}} \sum_{j' \neq j} \log(RDSTAFF_{jt} \times RDSTAFF_{j't}) \\ &= \log(RDSTAFF_{jt}) + \frac{1}{N_{t-1}} \sum_{j' \neq j} \log(RDSTAFF_{j't}) \\ &= \log(RDSTAFF_{jt}) + S_{jt},\end{aligned}\tag{1}$$

In this representation, R&D staff in firm j are the conduit by which spillovers affect the future R&D expenditure of firm j . At the same time, R&D staff are one proxy for R&D activity in the firm. Equation (1) incorporates the idea that spillovers are augmented by the interaction of R&D staff in firm j with R&D staff in other firms, j' . If other firms have more R&D staff (and hence more R&D activity), there is more to spill over to firm j . Likewise, if firm j has more R&D staff, the ability of firm j to benefit from the knowledge and networks of other firms increases. After taking logs, the second term, S , proxies for the stock of knowledge available from an average external firm and held by its R&D staff. Equation (1) is the natural log of the geometric mean of R&D staff in all other firms in the economy weighted by the number of R&D staff in firm j .

Firms with no R&D staff cannot benefit from spillovers in our model. Either they have no R&D expenditure, in which case they are excluded from our data. Or, they are outsourcing their research and are excluded from analysis.

For estimation, we expand on this simple model in three ways. First, we separate firms j' by status with respect to firm j into peers, suppliers and clients. We consider peers as those firms that are in the same 3-digit industry as firm j . This classification allows for different spillover effects from different types of relationships with potentially different types of technologies.

Second, we enhance our model by looking at spillovers from public sources of knowledge and separately consider R&D expenditure from federal government, state government and academia. Third, we introduce a set of parameters that allow for different effects of different sources of spillover. Initially, we do not include any role for geography but we introduce this in Section 7.2.

Spillovers from other firms, S_{jt} can come from three sources:

$$S_{jt} \text{ Peers} = \frac{1}{N_t^p - 1} \sum_{j' \neq j} w_{j',j}^p \log(RDSTAFF_{j't})\tag{2}$$

$$S_{jt} \text{ Suppliers} = \frac{1}{N_t^s - 1} \sum_{j' \neq j} w_{j',j}^s \log(RDSTAFF_{j't})\tag{3}$$

$$S_{jt} \text{ Clients} = \frac{1}{N_t^c - 1} \sum_{j' \neq j} w_{j,j'}^c \log(RDSTAFF_{j't}) \quad (4)$$

As discussed above, the relative weight of negative and positive effects of spillovers on R&D expenditure may differ by the source of the spillover. This can happen, for example, if spillovers from different sources differ in their proximity or adaptability to the technology used in firm j . Hence, we model them separately. Unlike Kenta et al. (2015), we do not know who is buying from whom. Also, unlike Bloom et al. (2013), we do not have enough information to construct a measure of technological proximity. Instead, we follow Brown & Conrad (1967) to use weights based on input-output tables reported for 3-digit industry groups.

For peers (horizontal spillover), $w_{j,j'}^p = 1$ if firms j and j' are in the same 3-digit industry, otherwise $w_{j,j'}^p = 0$. Firms included in this count mostly compete in the same market as that of firm j . R&D expenditure may allow rapid uptake of innovations from firms with similar markets. However, market externalities and competition also generate disincentives for R&D investment for these firms.

For clients and suppliers, we set weights equal to zero if firms j and j' are in the same 3-digit industry. For firms in different industries, firm j' could be in either a supply- or client-type vertical relationship with firm j . We use the Australian Bureau of Statistics (ABS) Input-Output tables for year 2007 and compute the share of output from industry of firm j' that is directed towards the industry of firm j . The Input-Output shares change very little over our data window, so we use 2007 data. We make the simplifying assumption that the proportion of researchers in firm j' that are relevant to firm j 's line of research follow the same share as outputs. $w_{j,j'}^s$ is the share of output by the industry of firm j' that is consumed by the industry of firm j . $w_{j,j'}^c$ reverses the direction and looks at the share of output from the industry of firm j that is consumed by the industry of firm j' . Following Javorcik (2004), we exclude firms that are in the same 3-digit industry from clients and suppliers, despite some intra-industry sourcing. This exclusion eliminates double counting and helps to reduce the correlation between our proxies.

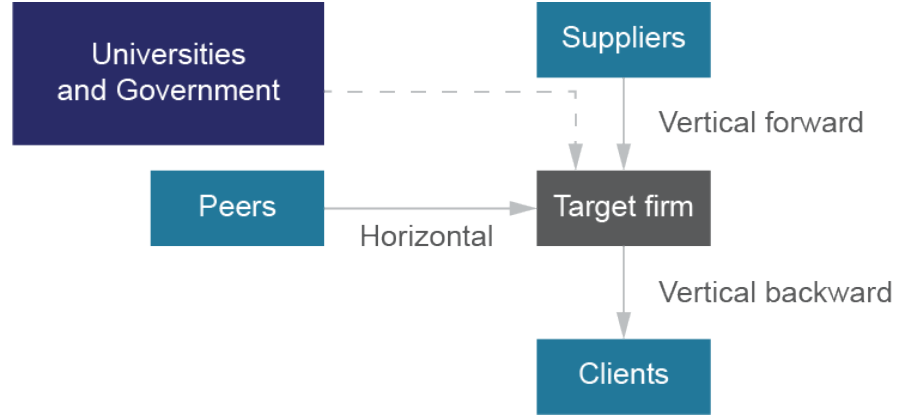
We model inward spillovers from all firms to firms based in Australia. Thus, the index j spans only those firms headquartered in Australia during the full range of years, whereas index j' covers both Australian firms and foreign subsidiaries as sources of spillovers.

Lastly, we enhance our model by accounting for public sources of knowledge spillover. The ABS publishes the amount of R&D expenditures by state and commonwealth governments (ABS Cat.No.8109) and by institutions of higher education (ABS Cat.No.8111) within each Australian state and territory. Using the information, we supplement our data with the following, measured at the state level:

- *FEDERAL*: R&D expenditure by commonwealth government.
- *STATE*: R&D expenditure by local state governments.
- *ACADEMIA*: R&D expenditures by institutes of higher education in Australia.

These data are collected biennially, and we fill the gap years by linear interpolation. Not to be confused with government R&D subsidies and other forms of assistance to private firms, the reported expenditures by the commonwealth and state governments on R&D are *strictly intramural* (ABS Cat.No.8109, Explanatory Notes). Figure 5.1 illustrates the full picture of proxies we use in the modelling of spillovers.

Figure 5.1: Modelled channels of spillovers



Source: Dept. of Industry, Innovation and Science (2016)

In the econometric specification in the next section, we bring all of these terms together and add parameters, α_p which allow for different effects of different sources of R&D activity. Put formally:

$$\begin{aligned} \overline{SPILL}_{jt} = \Sigma_{jt} \alpha = & \alpha_1 \log(RDSTAFF_{jt}) + \alpha_2 S_{jt}(Peers) + \\ & \alpha_3 S_{jt}(Supplier) + \alpha_4 S_{jt}(Client) + \alpha_5 \log(FEDERAL_t) + \\ & \alpha_6 \log(STATE_t) + \alpha_7 \log(ACADEMIA_t) \end{aligned} \quad (5)$$

We will refer to these spillovers and the associated parameters as Σ_{jt} and α .

6. Econometric Methodology

The goal of this study is to investigate the factors that impact the size of research and development expenditure (R&D) conducted by a firm with a focus on the role of spillovers from external sources. To begin, define

$$r = \log(R\&D), \quad s = \log(SALES),$$

where $R\&D$ and $SALES$ are the annual R&D expenditure and sales turnover, in nominal terms. Σ is the set of proxies defined in the previous section that measure the strength of inward spillover from external sources. We model the effect of these factors on the size of R&D expenditure in firm j at time $t+1$ as:

$$r_{j,t+1} = \sum_{jt} \alpha + b_1 r_{jt} + b_2 s_{jt} + T(r_{jt}, s_{jt}) + c_j + \tau_t + \varepsilon_{jt}. \quad (6)$$

c_j captures firm-specific and time-invariant characteristics and τ_t are a set of time dummy variables. In our model, we allow for a one year lag between current encounters and future R&D expenditures to account for the time it takes for knowledge to diffuse between firms. The function $T(\cdot, \cdot)$ captures the size of the subsidies that are provided through the R&D tax concession. Our administrative data do not include information on the exact amount of the subsidy — this data is held by the Australian Taxation Office, not by the Department of Industry. Subsidies depend upon a firm's reported accounting profits and losses as well as firm behaviour, and we proxy this using a non-linear (translog) function of firm sales and R&D expenditure:

$$T(r_{jt}, s_{jt}) = b_3 r_{jt} + b_4 s_{jt} + b_5 r_{jt}^2 + b_6 r_{jt} s_{jt} + b_7 s_{jt}^2. \quad (7)$$

Note that when we plug equation (7) into equation (6), b_1 and b_3 and also b_2 and b_4 cannot be separately identified. As our main objective is estimating vector α , this does not present any problem for our estimation strategy. Putting these two equations together produces the following dynamic panel representation

$$r_{j,t+1} = \rho r_{jt} + \sum_{jt} \alpha + \mathbf{X}_{jt} \beta + c_j + \tau_t + \varepsilon_{jt}. \quad (8)$$

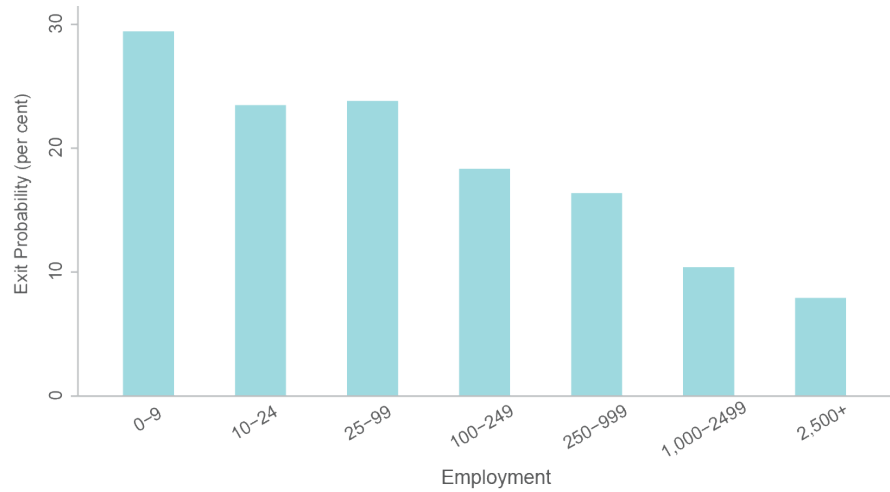
Note that $\rho = b_1 + b_3$, β contains the remaining b_k and \mathbf{X} contains s_{jt} , r_{jt}^2 , $r_{jt} s_{jt}$, s_{jt}^2 , and also the log of R&D staff.

Two econometric issues arise. The first is that there is entry and exit which can cause selection bias. The second is the inconsistency of estimates caused by the propagation of the fixed effects in a dynamic panel model (Hsiao, 2003, Page 72). We will adopt the approach of Semykina & Wooldridge (2013) to deal with both problems simultaneously.

According to Table 4.1, firms leave the sample in each year, either temporarily to reappear later or 'permanently' within our window of observation. One eligibility criteria for continuing in the programme, and therefore in the data, is to maintain at least \$20,000 in R&D expenditure. One reason for exit from the programme is that the firm's expected future R&D expenditures fall below this threshold and the firm thus decides not to sign up for the programme.

Unsurprisingly, there is a systematic relationship between size and exit. Small firms appear to carry out R&D in spells while larger firms often have a continuous and dedicated R&D programme. This systematic relationship is illustrated in Figure 6.1, where we predict the average probability of exit using a Probit projection on a set of firm size categories and year, industry and state dummy variables. We will use this systematic relationship to help address the selection problem, as described below.

Figure 6.1: Probability of exit by employment size categories.



Notes: Sample is all firm-year observations

Source: Dept. of Industry, Innovation and Science (2016)

We adopt the approach of Semykina & Wooldridge (2013) who use the recursive nature of equation 8 to compute r_{t+1} as a function of the initial condition and the history of the remaining explanatory variables. Owing to the nonlinear nature of the estimation, firm fixed effects are modelled as in Mundlak (1978) using a linear combination of the initial condition and the firm-level time averages of the continuous variables:

$$c_j = \eta + (\bar{\Sigma}_j, \bar{\mathbf{X}}_j)\xi + \gamma r_{j0} \quad (9)$$

$\bar{\Sigma}_j$ only includes the average of the peers, clients and suppliers as public R&D does not vary across firms. These latter averages will be subsumed in the constant, η . $\bar{\mathbf{X}}_j$ contains firm-level averages of s_{jt} , r_{jt}^2 , $r_{jt}s_{jt}$, s_{jt}^2 , and log of R&D staff over time. We drop all firms that appear only once in the panel (5,640 observations).

As in Semykina & Wooldridge (2013), we conduct the estimation in two steps. In step one, the probability that a firm continues into $t+1$ is projected onto a constant, r_{j0} , Σ_{jt} , \mathbf{X}_{jt} and the log of firm employment using Probit estimation. The log of employment is used in this step to avoid collinearity in the second step. It plays the role of an exclusion restriction in the sample selection model. In step two, nonlinear least squares is used to estimate:

$$r_{j,t+1} = \rho^t r_{j0} + \left(\sum_{k=0}^t \rho^k (\Sigma_{j,t-k}, j, t-k) \right) \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \left(\frac{1-\rho^t}{1-\rho} \right) c_j + \varphi_t \lambda_{jt} + \tau_t + e_{jt} \quad (10)$$

where λ_s are the inverse Mill's ratios from the first step.⁴

⁴ The e_{jt} will be a recursive sum of previous error terms and the key econometric assumption is orthogonality with respect to current period regressors.

7. Estimation Results

7.1 Baseline Estimations

Table 7.1 reports descriptive statistics for the covariates in our model. Sales, employment and R&D expenditures and staff relate to firm scale. The distribution of these variables is skewed with the median showing a large presence of small businesses while the average values are driven by a few very large firms. This pattern is typical of Australian businesses.

Table 7.1: Table of descriptive statistics for the set of covariates used in the modelling.

<i>Variable</i>	<i>Mean</i>	<i>Std.Dev.</i>	<i>1st Decile</i>	<i>Median</i>	<i>9th Decile</i>	<i>#obs</i>
Idiosyncratic						
<i>SALE</i>	82.2	1062.3	0.00	0.92	43.8	58,338
<i>EMP</i>	207.5	4038.0	1.00	10.0	150.0	58,338
<i>R&D</i>	1.36	8.61	0.06	0.31	1.80	58,338
<i>RDSTAFF</i>	7.63	409.8	0.00	2.00	9.90	57,977
Public knowledge						
<i>ACADEMIA</i>	1203.3	623.3	450.1	1062.3	2210.1	58,337
<i>FEDERAL</i>	319.56	148.29	120.57	313.2	541.0	58,337
<i>STATE</i>	227.7	75.8	113.6	256.9	308.7	58,337
Private knowledge						
<i>S</i>	1.10	0.07	1.09	1.11	1.14	51,307
<i>S Peers</i>	1.10	0.25	0.83	1.07	1.37	51,281
<i>S Suppliers</i>	1.08	0.09	1.01	1.09	1.18	51,307
<i>S Clients</i>	1.10	0.12	1.00	1.10	1.19	50,777

Source: Dept. of Industry, Innovation and Science (2016)

A few firms in our sample are not hiring any R&D staff despite spending non-zero sums on R&D. We believe that these firms are most likely paying for outsourced R&D activity rather than undertaking research internally. They are dropped in estimation.

The variables on public knowledge from universities and government vary across states and years but not across firms. Academia spends much more on R&D than the combined government sectors.

The bottom part of the table lists sample values of the spillover variables which we described in equations (1) through (4). The spillover proxies show greater dispersion when we split by peers, clients and suppliers.

In Table 7.2 we report the correlation coefficients between the key variables in our modelling specification. The firm size variables are positively correlated with one another. Spillovers do not show a strong correlation with any of the firm size measures. In particular, there is almost no correlation with number of employees. This reinforces the case for using this latter variable as an exclusion restriction in our selection model.

Table 7.2: Table of correlations between the set of covariates used in the modelling.

Variable	SALES	EMP	R&D	RDSTAFF	S	S Peers	S Suppliers	S Clients
EMP	0.350							
R&D	0.417	0.194						
RDSTAFF	0.021	0.011	0.042					
S	0.017	-0.003	0.003	0.004				
S Peers	0.093	0.034	0.091	0.013	0.268			
S Suppliers	0.064	0.025	0.027	0.006	0.733	0.348		
S Clients	0.022	-0.004	0.037	0.002	0.530	0.261	0.522	

Source: Dept. of Industry, Innovation and Science (2016)

There are positive correlations between 0.26 and 0.52 for the peer, client, and supplier proxies. In other words, engagement of larger R&D staff by an industry in general coincides with the engagement of more R&D staff by the industries that are positioned as suppliers and clients to that industry. This hints at some degree of synergy between the R&D activities of vertically linked industries and a networking effect.

We start by first estimating the average effect of each spillover channel by estimating (10) using the following set of proxies:

$$\Sigma_{jt} = \{S_{jt} Peers, S_{jt} Suppliers, S_{jt} Clients\} \cup \Sigma_{Public} \quad (11)$$

where

$$\Sigma_{Public} = \{\log(FEDERAL_{jt}), \log(STATE_{jt}), \log(ACADEMIA_{jt})\}$$

In addition to reporting the estimates from our main econometric approach, we report results from OLS and truncated regression estimates of the same specification to investigate whether selection and fixed effect biases are sizable enough to justify the use of our main methodology. In comparison to OLS, truncated regression accounts for the fact that firms with less than \$20,000 of future R&D expenditure are absent from the data. Truncated regression makes strong parametric assumptions and fails to account for the dynamic panel aspect of the problem. The full set of results is reported in columns (1) to (3) of Table 7.3.

Table 7.3: Estimated coefficients in the main model.

	OLS	Truncated	SW	SW
Variables	(1)	(2)	(3)	(4)
$\log(RDSTAFF)$	-0.184*** (0.006)	-0.190*** (0.009)	0.036*** (0.009)	0.077*** (0.009)
Peers	-0.001 (0.026)	0.007 (0.033)	-0.088*** (0.020)	
Suppliers	0.142 (0.089)	0.145 (0.104)	-0.035*** (0.010)	
Clients	0.130** (0.059)	0.147* (0.083)	-0.015*** (0.005)	
All				-0.245*** (0.034)
$\log(ACADEMIA)$	0.038*** (0.014)	0.034** (0.014)	0.101*** (0.029)	0.239*** (0.028)
$\log(FEDERAL)$	-0.053*** (0.010)	-0.054*** (0.011)	-0.112*** (0.035)	-0.179*** (0.023)
$\log(STATE)$	-0.015* (0.008)	-0.012 (0.008)	-0.002 (0.004)	-0.107*** (0.028)
r	0.821*** (0.005)	0.819*** (0.008)	0.580*** (0.009)	0.518*** (0.010)
s	-0.002 (0.002)	-0.002 (0.004)	0.012*** (0.004)	0.028*** (0.005)
r^2	0.115*** (0.002)	0.112*** (0.004)	0.017*** (0.004)	0.017*** (0.003)
rxs	-0.016*** (0.001)	-0.016*** (0.001)	0.019*** (0.004)	0.023*** (0.004)
s^2	0.005*** (0.001)	0.005*** (0.001)	-0.006*** (0.001)	-0.010*** (0.002)
Log Likelihood	-34,207.1	-33,286.0	-44,579.7	-45,302.8
N	38,166	38,166	38,166	38,566
σ^2	0.593	0.349	0.779	0.784

Notes: Dependent is the log of R&D expenditure at time $t+1$. SW refers to Semykina & Wooldridge (2013) method. Numbers in parentheses are robust standard errors. ***, **, and * indicate significance at 1 per cent, 5 per cent and 10 per cent, respectively. Dummy variables for year and industry are included but not reported.

Source: Dept. of Industry, Innovation and Science (2016)

Using the approach of Semykina & Wooldridge (2013) (SW) results in important differences relative to OLS and truncated regression estimates, which are similar to each other. In the OLS estimates, the coefficient for R&D staff is negative and spillover effects are insignificant or slightly positive. In the SW approach, the number of R&D staff has a positive effect on expenditure and we find that spillover effects are negative and significant. These are economically important differences.

We don't find much difference in the effect of public knowledge spillovers using SW as opposed to OLS or truncated regression. The latter models produce a higher estimate for the auto-regressive coefficient on past R&D expenditure. This is consistent with selection bias in these models. If larger R&D expenditure indicates appetite for future R&D, then the positive correlation that will ensue between R&D and the likelihood of staying in the programme will bias the OLS autoregressive coefficient upwards. With this in mind, we focus on the results from column (3) as our preferred model.

Overall, the coefficients on sales and R&D expenditures and their interactions suggest a stimulating role for the R&D tax concession. In our model, these variables are proxying for the amount of concession that firms receive. The estimate is confounded by correlation across time in firm R&D expenditure so we cannot quantify the effect of the tax concession. The only negative coefficient in this group is the one for s^2 . There appears to be a diminishing impact on R&D expenditure as sales increase.

Turning attention to the spillover terms, we observe a positive and significant role for the log of R&D staff. R&D staff and expenditures are complementary, but imperfect measures of actual research effort. The results support the notion that larger research activity facilitates the absorption and conversion of external knowledge for the firm's specific use as suggested by Cohen & Levinthal (1989).

As mentioned earlier, overall R&D spillovers can be negative or positive depending on which form of externality dominates. In our estimates, all of the spillover effects are negative and statistically significant, indicating that the disincentives dominate. In column (4), we estimate an overall effect by combining all spillovers into one term (see equation (1)). Unsurprisingly given column (3), the overall effect of spillovers in the last column is negative.

Academic expenditure appears to stimulate private R&D; the estimated coefficient is positive and statistically significant. This result echoes Jaffe (1989), who finds a positive influence of academic research on the private sector in the U.S., especially in high-technology fields. The impact from R&D expenditures by state and federal governments on private R&D is, however, negative in our estimates. It appears that private firms are dissuaded from carrying out R&D where government agencies are actively undertaking research projects. ABS reports (Cat.No.8109 and Cat.No.8111) indicate that more than 60 per cent of R&D expenditure in academia is for basic research. On the other hand, state and federal governments spend about 70 per cent of their total R&D expenditures on applied research. These estimates could be read to imply that basic research generates more knowledge spillovers than applied research or it could be that the way the research is being

communicated to the business community differs between academia and government.

7.2 Geography of Spillovers

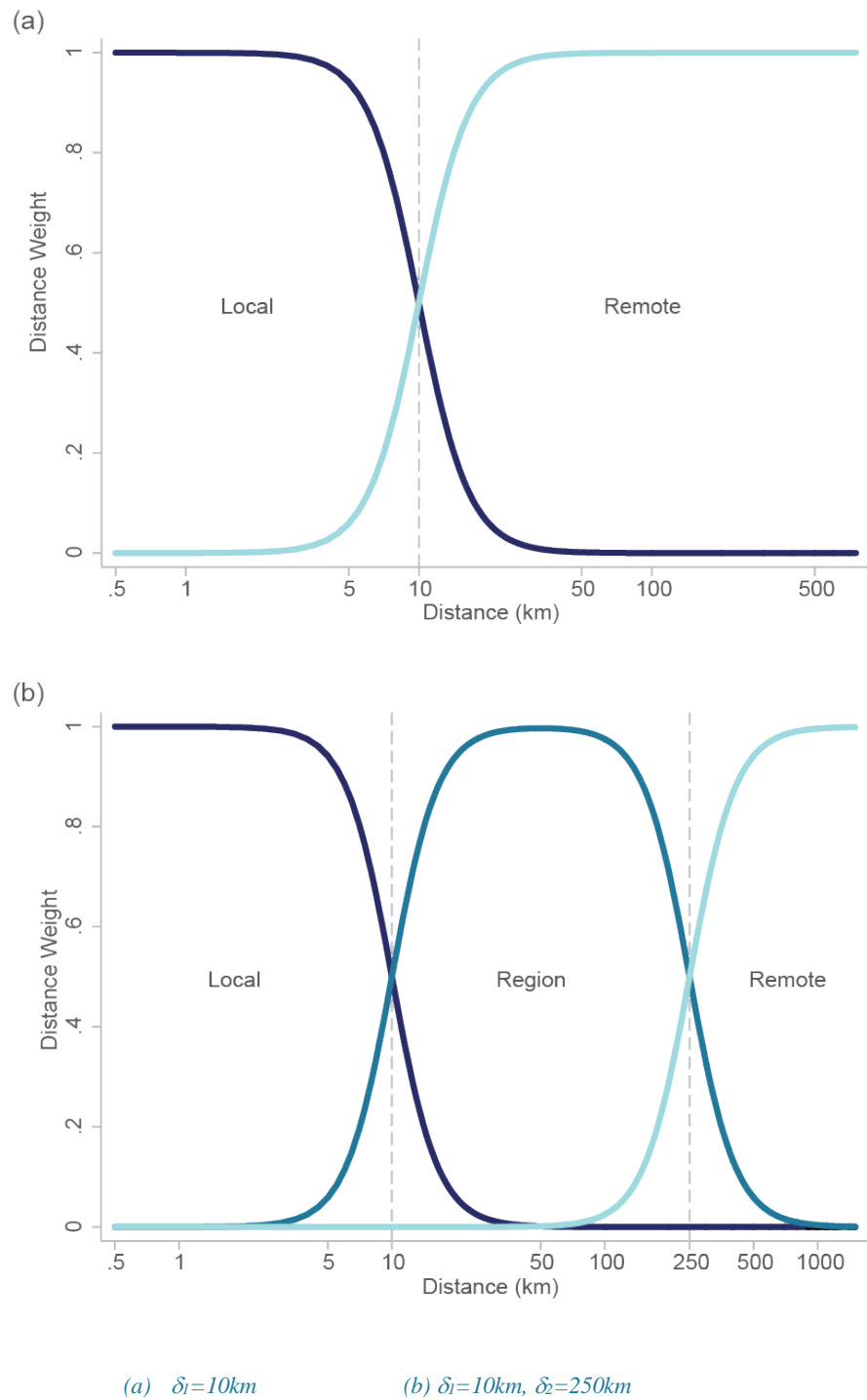
There is mounting evidence that the inter-firm leakage of knowledge is strongly correlated with the physical proximity of firms and happens in geographical clusters; see Chatterji et al. (2013) for a comprehensive survey of the related literature. Some recent work approaches the issue by formulating geography as an exponentially decaying function of distance followed by parametric or semi-parametric approximations which make estimation possible (Lychagin et al., 2016; Belderbos et al., 2013). Our econometric methodology, highly nonlinear and already extremely computationally intensive, renders this approach infeasible. We thus adopt a cruder approach by splitting distance into two (or three) discrete classes to separate out impact from local versus regional or remote firms.

We create weights, $\mu_{j,j'}$ based upon the physical proximity of firms j and j' using

$$\mu_{j,j'}^{LOC} = \frac{1}{1 + \left(\frac{d_{j,j'}}{\delta_1}\right)^\kappa}, \quad \kappa \geq 1 \quad (12)$$

d is the distance between firms j and j' and δ_1 is the radius of proximity such that firms at distance δ_1 are weighted by 50 per cent. κ controls the sharpness of the transition from local to remote classification as distance increases. With $\kappa=4$, which we use throughout, firms at distance $2\delta_1$ or farther are weighted by less than 6 per cent (Figure 7.1(a)). Similar weighting functions have been extensively used in fuzzy logic to make between class transitions — in this case, from local to remote firms — a smooth one (e.g., Zimmermann, 1996). The smoothness feature is especially helpful in rendering small adjustments to the distance parameter inconsequential. This is preferable to a sharp zero/one boundary where a small change in the radius can result in a large impact on results, for instance, when a cluster of firms lies close to the border.

Figure 7.1: Illustration of geo-weights for local, regional and remote firms.



Notes: In all cases $\kappa=4$.

Source: Dept. of Industry, Innovation and Science (2016)

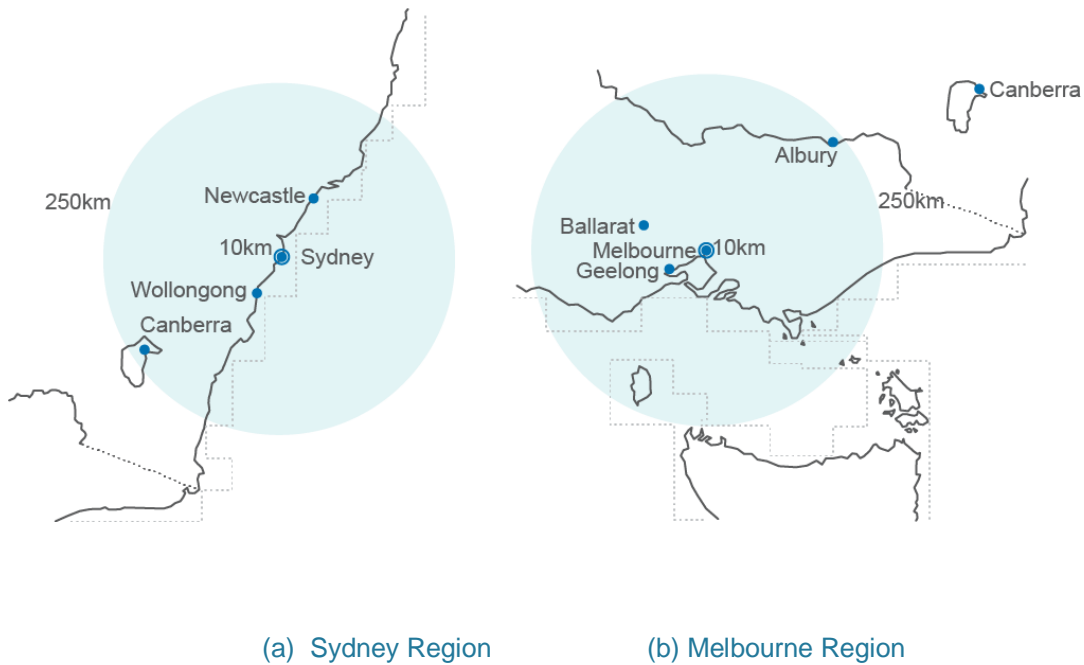
The classification can be extended to an arbitrary number of geographic distinctions but at the cost of increasing complexity. We mainly use two classes (local and remote) but also look at three classes (local, regional and remote)

for a finer classification of geography. For the case of three classes, we define the weights for remote (μ^{REM}) and region (μ^{REG}), as⁵:

$$\begin{aligned}\mu_{j,j'}^{REM} &= \frac{1}{1 + \left(\frac{\delta_2}{d_{j,j'}}\right)^\kappa}, \\ \mu_{j,j'}^{REG} &= 1 - \mu_{j,j'}^{LOC} - \mu_{j,j'}^{REM}\end{aligned}\quad (13)$$

In the same spirit as before, firms located at distances δ_1 and δ_2 are equally members of two classifications (Figure 7.1(b)). In Figure 7.2 we illustrate these distinctions for a firm located in the centre of two metropolitan areas Sydney and Melbourne using $\delta_1=10km$ and $\delta_2=250km$. In the case of Sydney, for instance, the definition of region covers industrial areas in the neighbouring cities of Newcastle and Wollongong and areas as far away as Canberra.

Figure 7.2: An exhibition of the locality (10km radius) and region (250km radius) for a firm located at the centre of Sydney and Melbourne metropolitans.



Source: Dept. of Industry, Innovation and Science (2016)

Firms in Australia can be as far as 4,000 kilometers from one another; not accounting for the curvature of the earth's crust could lead to significant computational errors. Hence, instead of the Euclidean distance, we compute the great arc distance between every two pairs of firms using the available longitude and latitude coordinates in the Haversine formula. In particular, if the coordinates of two firms are (x_i, y_i) and (x_j, y_j) , then

⁵ Note that where we only use two classifications, the remote weight is $1 - \mu_{j,j'}^{LOC}$; see Figure 6(a).

$$d_{j,j'} = 2R \arcsin \left(\sqrt{\sin^2 \left(\frac{y_j - y_{j'}}{2} \right) + \cos(y_j) \cos(y_{j'}) \sin^2 \left(\frac{x_j - x_{j'}}{2} \right)} \right) \quad (14)$$

in which R is the average radius of the earth and is set equal to 6,371.009km.

Incorporating the geographic weights into the proxies introduced earlier generates an extra set of distance-based measures of spillovers originating from local firms. For example, in the case of spillovers that originate from peers, one can define

$$S_{jt} \text{ Local Peers} = \frac{1}{N_t^p - 1} \sum^{j' \neq j} \mu_{j,j'}^{LOC} w_{j',j}^p \log(RDSTAFF_{j't}) \quad (15)$$

We construct the local proxies for suppliers and clients with the same convention but additionally using the appropriate vertical weights ($w_{j,j}^S$ and $w_{j,j}^C$ as previously defined in equations (3) and (4)). An extra set of proxies would characterize spillovers generated by all firms in the regional radius or

$$S_{jt} \text{ Regional Peers} = \frac{1}{N_t^p - 1} \sum^{j' \neq j} (\mu_{j,j'}^{LOC} + \mu_{j,j'}^{REG}) w_{j',j}^p \log(RDSTAFF_{j't}) \quad (16)$$

This measure accumulates all firms within the radius of δ_2 (within both the local and regional areas).⁶

To study whether geographic distances play any part in spillovers, we extend our specifications by using the following set of proxies:

$$\Sigma_{jt} = \{S_{jt} \text{ Peers}, S_{jt} \text{ Suppliers}, S_{jt} \text{ Clients}\} \cup \{S_{jt} \text{ Local Peers}, S_{jt} \text{ Local Suppliers}, S_{jt} \text{ Local Clients}\} \cup \Sigma_{Public} \quad (17)$$

In this specification, the first set of proxies measure the average effect of each source, as is the case in (11). The second set of *local* proxies measure the premium for being proximate. Chatterji et al. (2013) provide evidence for a spillover radius ranging from one kilometer to hundreds of kilometers. We use an initial radius of 10km for proximity to account for a fair amount of exposure between the R&D staff. We then follow up by using radii of 25km and 50km to see if the results are affected. Sydney and Melbourne, Australia's two largest cities, are almost fully covered by the 50 km radius.

⁶ Note that the weight $\mu_{j,j'}^{LOC} + \mu_{j,j'}^{REG}$ is equal to $1 - \mu_{j,j'}^{REM}$, which is the weight of being 'not remote'.

To get a more refined view, we estimate (17) using three geographical splits

$$\Sigma_{jt} = \{ \begin{array}{l} S_{jt} \text{ Peers}, S_{jt} \text{ Suppliers}, S_{jt} \text{ Clients} \\ \cup \{S_{jt} \text{ Local Peers}, S_{jt} \text{ Local Suppliers}, S_{jt} \text{ Local Clients}\} \\ \cup \{S_{jt} \text{ Regional Peers}, S_{jt} \text{ Regional Suppliers}, S_{jt} \text{ Regional Clients}\} \cup \Sigma_{Public} \end{array} \} \quad (18)$$

In this specification, the regional proxies account for the premium from spillovers generated within the radius δ_2 , while local proxies account for the premium on top of the regional one for being co-located within the radius δ_1 .

Information on the spillover terms which include input–output weights and the locality weights is reported in Table 5. As we move from the smaller radius, $\delta_1=10\text{km}$, to the larger one, $\delta_1=50\text{km}$, the size of average spillover terms increase — that is more firms are defined as contributing to ‘local’ spillovers as the radius expands.

Table 7.4: Average value for S (see equations (2) through (4)) for two values of radius, δ_1 .

Variable	Mean	Std.Dev.	1st Decile	Median	9th Decile	#obs
Idiosyncratic						
Panel A	$\delta_1=10\text{km}$					
S Local Peers	0.11	0.14	0.00	0.07	0.28	51,281
S Local Suppliers	0.08	0.08	0.00	0.07	0.19	51,307
S Local Clients	0.09	0.09	0.00	0.07	0.21	50,777
Panel B	$\delta_1=50\text{km}$					
S Local Peers	0.22	0.19	0.02	0.18	0.46	51,281
S Local Suppliers	0.19	0.13	0.03	0.19	0.36	51,307
S Local Client	0.20	0.13	0.03	0.20	0.35	50,777

Source: Dept. of Industry, Innovation and Science (2016)

We report the estimated coefficients for the distance-based proxies in Table 7.5 where the first three columns correspond to the use of proximity radii of 10km, 25km, and 50km, respectively. The last column presents the results when we use three geographic classifications.

We consistently find that geography matters. As before, the overall average effect estimated for spillovers among peers is negative. However, we estimate a positive and statistically significant premium for geographic proximity amongst peers. The premium is strong enough that the total spillover effect is also positive within 10 km (the sum of the two coefficients for peers in column two is statistically greater than zero.) When we expand out to 25 or 50 km, we still find a positive influence of proximity on R&D spillovers. The total effect is positive within 25 km and turns negative before 50 km.

There is evidence from the literature (see Jaffe et al. (1993) for instance) that knowledge spillovers are localized and the strength of these spillovers diminishes with distance. The pattern we observe is consistent with this evidence: if knowledge and network spillovers are strongest within localized areas then they dominate and create a positive total spillover effect. As distance increases, negative externalities arising from market effects or other pecuniary factors dominate and the total spillover effect becomes negative.

For clients, we find a similar result. Specifically, the average effect of spillovers from clients is negative but within the local region there are both positive premiums and positive overall effects. Even up to 50 km, the total R&D spillover effect is still positive. This pattern, again, points to positive spillovers within localized areas.

Table 7.5: Estimated coefficients in the distance-based specification.

	(1)	(2)	(3)	(4)
Parameters				
κ	4	4	4	4
$\bar{\delta}_1$	10km	25km	50km	10km
$\bar{\delta}_2$				250km
Spillovers				
Peers	-0.028*** (0.006)	-0.159*** (0.016)	-0.119*** (0.013)	-0.082*** (0.011)
Local Peers	0.071*** (0.027)	0.287* (0.040)	0.072*** (0.010)	0.250*** (0.066)
Regional Peers				-0.109*** (0.024)
Suppliers	0.235*** (0.038)	0.229*** (0.030)	0.294*** (0.035)	0.231*** (0.027)
Local Suppliers	-1.394*** (0.240)	-1.350*** (0.129)	-0.775*** (0.101)	-1.125*** (0.171)
Regional Suppliers				0.106*** (0.026)
Clients	-0.361*** (0.036)	-0.205*** (0.030)	-0.300*** (0.035)	-0.263*** (0.027)
Local Clients	1.063*** (0.181)	0.930*** (0.102)	0.529*** (0.071)	0.657*** (0.131)
Regional Clients				-0.097*** (0.025)
Log Likelihood	-44,415.5	-44,478.8	-44,493.5	-44,378.8
N	38,166	38,166	38,166	38,166
σ^2	0.775	0.777	0.777	0.775

Notes: Dependent variable is the log of R&D expenditures at $t+1$. The estimation method in all columns is that of Semykina & Wooldridge (2013). Numbers in parentheses are robust standard errors. *** indicates significance at 1 per cent. Dummy variables for year and industry and additional controls are included but not reported.

Source: Dept. of Industry, Innovation and Science (2016)

Spillovers from suppliers also show a significant geographic-dependent aspect but in an unexpected way: there appears to be a negative effect of proximity for suppliers. The addition of a third geographical classification (column (4)) does not change the overall picture. This may be a result of different spillover effects dominating at the local level. We also explore whether these results

could be driven by Australia's unique geographical distribution of firms or the result of industry composition in our data.

First, we re-estimate our model using only those firms that are in the densely populated part of coastal Australia. Specifically, we restrict our sample to firms that are located within a 50km radius of the city centres of Sydney, Melbourne, and Brisbane. We still define our spillover variables in the same way to include all firms, even those outside these cities. We want to check if our results from Table 7.5 might be driven by firms located in remote or sparsely populated areas of Australia. These firms may have no local peers, clients or suppliers and could be sourcing knowledge in a different way. We re-estimate and report the results in Table 7.6. We do not observe any qualitative change in the findings. The main difference in these results is that we now find a positive spillover effect for local peers for all distances considered — 10, 25 and 50 km. The other results are unchanged, in particular, the positive effect of spillovers from local clients and the negative effect for local suppliers.

Table 7.6: Estimated coefficients when sample is restricted to the metropolitan areas of Sydney, Melbourne, and Brisbane.

	(1)	(2)	(3)	(4)
Parameters				
κ	4	4	4	4
$\bar{\delta}_1$	10km	25km	50km	10km
$\bar{\delta}_2$				250km
Spillovers				
Peers	-0.239*** (0.011)	-0.158*** (0.016)	-0.212*** (0.019)	-0.167*** (0.007)
Local Peers	0.742*** (0.057)	0.492*** (0.034)	0.351*** (0.023)	0.545*** (0.048)
Regional Peers				-0.103*** (0.008)
Suppliers	0.458*** (0.020)	0.333*** (0.011)	0.265*** (0.021)	0.449*** (0.014)
Local Suppliers	-2.281*** (0.118)	-1.134*** (0.075)	-0.940*** (0.048)	-1.011*** (0.087)
Regional Suppliers				-0.600*** (0.025)
Clients	-0.775*** (0.021)	-0.704*** (0.015)	-0.678*** (0.012)	-0.883*** (0.018)
Local Clients	1.492*** (0.113)	1.182*** (0.056)	1.264*** (0.060)	0.521*** (0.056)
Regional Clients				0.988*** (0.043)
Log Likelihood	-28,086.5	-28,123.5	-28,103.5	-28,026.6
N	24,600	24,600	24,600	24,600
σ^2	0.759	0.760	0.759	0.757

Notes: Dependent variable is log of R&D expenditures at $t+1$. The estimation method is Semykina & Wooldridge (2013). Numbers in parentheses are robust standard errors. *** indicates significance at 1 per cent. Dummy variables for year and industry and additional controls are included but not reported.

Source: Dept. of Industry, Innovation and Science (2016)

For some industries, particularly those that primarily interact online, distance may be irrelevant. Two such industries in our dataset are Media and Telecommunication and Professional, Scientific and Technical Services, which also have sufficient presence in the data to make an impact. We re-estimate our model dropping these industries from our sample. We re-define all of the

spillover proxies using this reduced sample. The results are listed in Table 7.7 columns (1) to (3) and do not show any qualitative change.

Table 7.7: Estimated coefficients with restrictions on subsamples of industries.

	(1)	(2)	(3)	(4)
Industry Subset	Excludes Media & Telecom	Excludes Prof. Services	Excludes Media & Telecom, Prof. Services	Includes Manufacturing only
Parameters				
κ	4	4	4	4
δ_1	10km	10km	10km	10km
Spillovers				
Peers	-0.180*** (0.018)	-0.180*** (0.028)	-0.120*** (0.020)	-0.052*** (0.009)
Local Peers	0.286*** (0.041)	0.308*** (0.072)	0.249*** (0.085)	0.125*** (0.023)
Suppliers	0.585*** (0.058)	0.214*** (0.031)	0.331*** (0.054)	0.205*** (0.027)
Local Suppliers	-1.374*** (0.200)	-1.041*** (0.162)	-0.902*** (0.195)	-3.389*** (0.464)
Clients	-0.260*** (0.025)	-0.022*** (0.005)	-0.021*** (0.005)	-0.017*** (0.003)
Local Clients	0.868*** (0.119)	0.067*** (0.026)	0.452*** (0.130)	1.926*** (0.291)
Log Likelihood	-39,689.9	-36,368.8	-32,198.8	-19,150.6
N	33,679	30,583	26,504	17,656
σ^2	0.787	0.795	0.816	0.717

Notes: Dependent variable is log of R&D expenditures at $t+1$. Estimation method is Semykina & Wooldridge (2013). Numbers in parentheses are robust standard errors. *** indicates significance at 1 per cent. Dummy variables for year and industry and additional controls are included but not reported.

Source: Dept. of Industry, Innovation and Science (2016)

Finally, we estimate the model restricting our sample to manufacturing firms to check whether the pattern is being driven by the service industry. In this case, we only keep manufacturing firms and recalculate all of our proxy variables using only manufacturing firms. These results appear in the last column of Table 7.7 and again point to the same pattern found above.

Considering all this evidence, proximity appears to have a very different effect on R&D spillovers for suppliers compared to peers and clients. The disincentives to R&D investment appear stronger for suppliers who are close by than for those that are far away. It seems unlikely that distance would be an

advantage in generating spillovers. One possibility is that this is the result of different types of interactions between firms and suppliers. Another possibility is that this is driven by a difference in the mix and nature of the externalities generated by spillovers for suppliers as opposed to clients and peers. If we think that knowledge and network spillovers are likely to be stronger with proximity, then these results would indicate that those types of spillovers from supplier firms are creating dis-incentives for R&D expenditure for firms. Unfortunately, our data does not allow us to examine different types of spillovers, but this raises an interesting question for future research.

Kenta et al. (2015) find positive local effects for peers, but they find insignificant geographic effects for clients and suppliers combined together. Our results highlight that one might want to treat clients and suppliers differently.

7.3 Spillover and Clustering

As in many other industrial countries, some industries such as car manufacturing or pharmaceuticals are geographically clustered in Australia. Industrial clustering becomes significant when similar firms co-locate to benefit from increased knowledge spillovers in order to accelerate their innovation rate (Ellison et al., 2010). Clustering also adds an extra dimension to knowledge spillover by raising the likelihood of collaboration, information sharing and talent poaching. To explore this possibility in our data, we construct the M-index of clustering introduced by Marcon & Puech (2010). We apply a log transformation to control for extreme values, so that

$$CLUSTER_j = \log\left(\frac{N_j^p/N_j}{N^p/N}\right) \quad (19)$$

N_j^p is the number of firms in the same 3-digit ANZSIC division as firm j in a radius of 10km around this firm, N_j is the total number of firms in the data in the same 10km radius. N^p is the total number of firms in the same 3-digit ANZSIC, and N is the total number of distinct firms in the data.⁷ None of the firms in our data change location, therefore, this index is time-invariant for each firm. The index measures how the local distribution of similar firms differs from a uniform distribution. In the case of clustering, the value of $CLUSTER$ will be positive and increases as firms are more clustered. Conversely, if a firm does not have many similar firms as neighbours, the value of $CLUSTER$ will be negative. Table 7.8 shows the range of values the index takes in our data.

Table 7.8: Descriptive statistics for the clustering index.

	<i>Mean</i>	<i>Std.dev.</i>	<i>Min.</i>	<i>Qrtl.1</i>	<i>Median</i>	<i>Qrtl.3</i>	<i>Max.</i>
<i>CLUSTER</i>	0.279	0.678	-2.992	-0.110	0.293	0.548	5.933
						N	19,300

Source: Dept. of Industry, Innovation and Science (2016)

⁷ Recall that our data only include R&D active firms.

To investigate whether industrial clustering intensifies spillovers, we now use the following set of proxies:

$$\begin{aligned}
\Sigma_{jt} = & \{S_{jt} \text{ Peers}, S_{jt} \text{ Suppliers}, S_{jt} \text{ Client}\} \\
& \cup \{S_{jt} \text{ Local Peers}, S_{jt} \text{ Local Suppliers}, S_{jt} \text{ Local Clients}\} \\
& \cup \{CLUSTER_j, CLUSTER_j \times S_{jt} \text{ Peers}, CLUSTER_j \\
& \times S_{jt} \text{ Local Peers}\} \cup \Sigma_{Public}.
\end{aligned} \tag{20}$$

The interaction terms in this specification test whether there is a premium for peers to be co-located within a 10km radius or not. The clustering hypothesis is by its nature a hypothesis about peer relationships so we only interact the clustering variable with peers. The estimated coefficients are reported in Table 7.9. The first column of results in the table only looks at the effect of clustering. Column (2) estimates the full model; column three is copied from Table 6 for ease of comparison.

Table 7.9: Estimated coefficients with clustering interactions.

<i>Variables</i>	(1)	(2)	(3)
Parameters			
κ	4	4	4
$\bar{\delta}_1$	10km	10km	10km
Variables			
$CLUSTER_t$	0.048*** (0.007)	0.056*** (0.008)	
Peers		-0.004*** (0.001)	
Local Peers		-0.003*** (0.000)	
Peers		-0.134*** (0.012)	-0.028*** (0.006)
Local Peers		0.236*** (0.039)	0.071*** (0.027)
Suppliers		0.414*** (0.043)	0.235*** (0.038)
Local Suppliers		-1.809*** (0.174)	-1.394*** (0.240)
Clients		-0.426*** (0.043)	-0.361*** (0.036)
Local Clients		1.187*** (0.123)	1.063*** (0.181)
Log Likelihood	-45,607.5	-43,826.1	-44,415.5
N	38,861	37,697	38,166
σ^2	0.783	0.774	0.775

Notes: Dependent variable is log of R&D expenditures at $t+1$. Estimation method is Semykina & Wooldridge (2013). Numbers in parentheses are robust standard errors. *** indicates significance at 1 per cent. Dummy variables for year and industry and additional controls are included but not reported.

Source: Dept. of Industry, Innovation and Science (2016)

Overall, the clustering index by itself shows a positive and significant coefficient, that is, within-cluster firms invest more on R&D. The estimated coefficients for the interaction terms are negative, but small. In other words, we do not detect any additional spillover within clusters of similar firms as was the initial hypothesis. The overall impacts of average and local R&D spillover effects remain unchanged after the addition of clustering terms.

8. Conclusion and discussion

The existing literature is abundant with evidence for the role of spillovers on innovation and productivity and emphasizes the localized nature of those spillover effects. In this paper, we focus on one aspect of spillovers — the role of spillovers in affecting firm-level R&D expenditure — and find similar evidence. We go further by allowing for different R&D spillover effects from peers, clients and suppliers and also for spillovers from government and academia.

Our first notable finding is that, as is the case in other countries, geography plays an important role for R&D spillovers in Australia. For clients and peers, we find a positive effect of spillovers on firm-level R&D expenditure if firms are located near each other. We find this effect when we use radii of 25 or 50 km. We find positive spillover effects for suppliers which then become negative with proximity. Our interpretation is that negative disincentives dominate overall but they are mitigated by geography. This is especially true for clients and peers but not for suppliers. This would be the case if the mix of spillovers and their effects are different for clients, peers and suppliers. We are unable to explore this further as our data does not allow us to separately identify different types of spillovers. However, it highlights the fact that it is important to consider peers, clients and suppliers as having separate and distinct impacts on R&D spillovers

Our second important finding is that public expenditure has significant but opposite effects depending upon the source. Academic R&D expenditure has positive spillover effects on firm-level R&D expenditure whereas government expenditure at both local and federal levels has negative effects. Government R&D seems to crowd out firm-level R&D. This could be due to academic R&D being more focused on basic research than government R&D or could be the result of some difference in how the R&D expenditure actually spills over into firms. We can not distinguish between these two possibilities. These results suggest that shifting expenditure from government R&D to academic R&D might bring positive effects for firm-level R&D expenditure. It also suggests that a shift by government agencies towards more basic research could be rewarding.

Finally, we find a positive effect of clustering on R&D expenditure. This effect could be a consequence of more intensive competition or alternatively a result of building collaborations within clusters. We tested for the latter effect by interacting clustering with peers and local peers but we find no significant extra effect on local spillovers from higher levels of clustering. The competition effect seems to be the dominating factor.

A few caveats are also worth mentioning. As is often the case with firm-level data, we had difficulty identifying vertical relationships. We used input-output tables to create industry averages which we then used to weight peer, client and supplier effects. More detailed firm-level data that would allow for identification of these relationships would be better. Given the computational demands of our non-linear model, we use a rather crude classification of geography but managed to mitigate this problem by making our classification a fuzzy one. Finally, due to lack of information, we had to treat all R&D employees as identical. It is likely that there is considerable heterogeneity among R&D employees in their tenures and skills. More disaggregated data on

the characteristics of these employees would be desirable. We look forward to seeing whether our results hold once richer firm-level data become available in Australia.

Appendix A R&D Tax Concession Programme

As of 2001, the R&D Tax Concession offered (Australian Taxation Office, 2009):

- a 125 per cent tax concession that applies to all firms,⁸
- a R&D tax offset to small businesses in case they incurred tax losses,
- a 175 per cent tax concession that applies to the incremental part of R&D expenditures. The incremental part is defined as any R&D expenditure above the average expenditure over the preceding three years.

The first two parts of the programme were available for all firms operating in Australia irrespective of ownership. Originally, the 175 per cent concession only applied to domestic firms. In 2007, the government extended the 175 per cent tax concession for incremental R&D to the subsidiaries of multinational firms operating in Australia.

Firms need to meet three main eligibility criteria to participate in the programme. First, the business or the subsidiary must be registered in Australia. Second, at least 90 percent of R&D expenditure must be incurred on Australian soil. Third, the annual R&D expenditure has to be at least \$20,000 (all monetary values are in AUD) with the exception of when R&D is conducted jointly with an approved research institute. The government's guideline for what constitutes research and development and defines R&D may be found at (Australian Taxation Office, 2009):

1. Systemic, investigative and experimental activities that involve innovation or high levels of technical risk and are carried on for the purpose of
 - 1.1 acquiring new knowledge (whether or not that knowledge will have a specific practical application); or
 - 1.2 creating new or improved material, products, devices, processes or services; or
2. Other activities that are carried on for a purpose directly related to the carrying on of activities of the kind referred to in paragraph (1).

Only those costs conforming to the guideline above are considered eligible expenses. Firms that satisfy all the eligibility conditions have to register no later than 10 months after the end of the fiscal year in which the expenses were incurred in order to claim the benefits. For more details on the programme visit.

⁸ In the US taxation system, the equivalent of a tax concession is an income credit; a deduction from income that results in a reduction in taxes. In turn, an Australian tax offset is equivalent to a tax credit in the US system.

Appendix B Impact of foreign multinationals

Our main focus has been the channels of knowledge spillover within Australia for domestic firms and we thus dropped all firms that were foreign-owned or subsidiaries of foreign multinational companies. As a robustness check, we estimate the model on all firms including those that are foreign-owned. The estimation results for the full sample of domestic plus multinational firms is reported in Table B1.

Although some of the coefficient values are slightly different, the signs and significances of the key variables are all unchanged from the main analysis. Our exclusion or inclusion of foreign-owned firms does not alter the results.

Table B1: Estimated coefficients in the main model when including both domestic and foreign firms in the sample.

Variables	(1)	(2)
$\log(RDSTAFF_t)$	0.037*** (0.004)	0.019*** (0.002)
Peers	-0.118*** (0.015)	
Suppliers	-0.238*** (0.017)	
Clients	0.146*** (0.017)	
All		-0.192*** (0.012)
$\log(ACADEMIA_t)$	0.260*** (0.024)	0.284*** (0.018)
$\log(FEDERAL_t)$	-0.206*** (0.024)	-0.208*** (0.018)
$\log(STATE_t)$	-0.098*** (0.023)	-0.127*** (0.020)
Log Likelihood	-54,648.4	-55,563.1
N	44,465	44,976
σ^2	0.827	0.833

Notes: Dependent variable is the log of R&D expenditure at time $t+1$. Estimation method is that of Semykina & Wooldridge (2013). Numbers in parentheses are robust standard errors. *** indicates significance at 1 per cent. Dummy variables for year and industry and additional controls are included but not reported.

Source: Dept. of Industry, Innovation and Science (2016)

References

- Audretsch, David B., and Maryann P. Feldman (1996) "R&D Spillovers and the Geography of Innovation and Production," *American Economic Review*, 86(3), 630–640.
- Australian Taxation Office (2009) "Tax Concession for Research and Development," (<https://www.ato.gov.au/Business/Research-and-development-tax-concession/>).
- Belderbos, Rene, Kenta Ikeuchi, Kyoji Fukao, Young Gak Kim, and Hyeog Ug Kwon (2013) "Plant Productivity Dynamics and Private and Public R&D Spillovers: Technological, Geographic and Relational Proximity," *Center for Economic Studies*, Working Paper 2013-5, Hitotsubashi University.
- Blomström, Magnus and Ari Kokko (1998) "Multinational Corporations and Spillovers," *Journal of Economic Surveys*, 12(3), 247-277.
- Bloom, Nicholas, Mark Schankerman, and John Van Reenen (2013) "Identifying Technology Spillovers and Product Market Rivalry," *Econometrica*, 81(4), 1347–1393.
- Bottazzi, Laura, and Giovanni Peri (2003) "Innovation and spillovers in regions: Evidence from European patent data," *European Economic Review*, 47(4), 687–710.
- Brown, Murray, and Alfred H. Conrad (1967) "The influence of Research on CES Production Relations," in *The Theory and Empirical Analysis of Production*, Edited by M.Brown, NBER, 341–394.
- Chatterji, Aaron, Edward L. Glaeser, William R. Kerr (2013) "Clusters of Entrepreneurship and Innovation," *NBER Working Paper*, No.19013.
- Cohen, Wesley M., and Daniel A. Levinthal (1989) "Innovation and Learning: The Two Faces of R&D," *Economic Journal*, 99(397), 569–596.
- Ellison, Glenn, Edward L. Glaeser, and William Kerr (2010) "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns" *American Economic Review*, 100(3), 1195–1213.
- Elnasri, Amani, and Kevin J. Fox (2014) "The Contribution of Research and Innovation to Productivity and Economic Growth," *Univ. of New South Wales*, Working Paper 2014-08.
- Griliches, Zvi (1979) "Issues in Assessing the Contributions of Research and Development to Productivity Growth," *Bell Journal of Economics*, 10(1), 92–116.
- Griliches, Zvi (1992) "The Search for R&D Spillovers," *Scandinavian Journal of Economics*, 94(Suppl.), S29–S47.
- Hsiao, Cheng (2003) *Analysis of Panel Data*, 2nd Edition, Cambridge University Press, Cambridge, UK.

- Jaffe, Adam B. (1986) "Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits and Market Value," *American Economic Review*, 76(5), 984–1001.
- Jaffe, Adam B. (1988) "Demand and Supply Influences in R&D Intensity and Productivity Growth," *Review of Economics and Statistics*, 70(3), 431–437.
- Jaffe, Adam B. (1989) "Real Effects of Academic Research," *American Economic Review*, 79(5), 957–970.
- Jaffe, Adam B., Manuel Trajtenberg, and R. Henderson (1993) "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics*, 108, 577–598.
- Jaffe, Adam B. (1998) "The importance of Spillovers in the policy Mission of the Advanced Technology Program," *Journal of Technology Transfer*, 23(2), 11–19.
- Javorcik, Beata S. (2004) "Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers through Backward Linkages," *American Economic Review*, 94(3), 605–627.
- Jones, Charles I., and John C. Williams (1998) "Measuring the Social Return to R&D," *Quarterly Journal of Economics*, 113(4), 1119–1135.
- Kenta, Ikeuchi, Rene Belderbose, Fukao Kyoji, Young Gak Kim, and Kwon Hyeog Ug (2015) "Buyers, Suppliers, and R&D Spillovers," *RIETI Discussion Paper*, No. 15-E-047.
- Lychagin, Sergey, Joris Pinkse, Margaret E. Slade, and John Van Reenen (2016) "Spillovers in Space: Does Geography Matter?" *Journal of Industrial Economics*, 64(2), 295–335.
- Mairesse, Jacques, and Benoit Mulkey (2008) "An Exploration of Local R&D Spillovers in France," *NBER Working Paper*, No.14552.
- Marcon, Eric, and Florence Puech (2010) "Measures of the Geographic Concentration of Industries: Improving Distance-based Methods," *Journal of Economic Geography*, 10, 745–762.
- Mundlak, Yair (1978) "On the Pooling of Time Series and Cross Section Data," *Econometrica*, 46(1), 69–85.
- Schoar, Antoinette (2002) "Effects of Corporate Diversification on Productivity," *Journal of Finance*, 57(6), 2379–2403.
- Semykina, Anastasia, and Jeffrey M. Wooldridge (2013) "Estimation of Dynamic Panel Data Models with Sample Selection," *Journal of Applied Econometrics*, 28(1), 47–61.
- Zimmermann, Hans-Jürgen (1996) *Fuzzy Set Theory and Its Applications*, 3rd Edition, Kluwer Academic Publishers, Boston, USA.