Centre for Transformative Innovation

Global Prosperity Through Knowledge and Innovation

SWINBURNE UNIVERSITY OF TECHNOLOGY

INNOVATION = IDEAS + CHANGE

Measuring R&D spillovers from Australia industry: Uses and limitations of using the Extended Analytic Business Longitudinal Database (EABLD)

Paper for the Department of Industry and Science

Alfons Palangkaraya and Elizabeth Webster

Centre for Transformative Innovation
Swinburne University of Technology

Introduction

Access to firm-level datasets from within the Australia Bureau of Statistics (ABS) is a watershed moment for empirical research into Australian firm performance. Hitherto, empirical work in this area has been limited to small and possibly biased datasets comprising firm-level information.

To produce robust results and identify causality, datasets should include both cross-sectional and time dimensions. Cross-sectional data is data on units (eg individual firms; industries; countries or people) for one period of time. Time series data is data on a unit (firm, individual, industry, country etc) for a series of time periods. Panel data is both time series and cross-sectional and thus contains data on multiple units over multiple time periods.

Small datasets, which are typically contain a single cross-section, rarely produce results that have the degree of robustness to which policy makers are now accustomed. In the areas of health, education, social security and labour market, panel data analysis is the norm. The paucity of firm panel data analysis makes it hard to convince broader audiences of the effects – or lack thereof – of industry programs and policies. Since 2013, the ABS has been allowing the research community to interrogate their unpublished firm-level datasets, albeit under strict confidentiality protocols.

There are two main ABS datasets. The first is based on the ABS Business Characteristics Survey (BCS), an annual mail-out/mail-back survey designed to collect detailed information on firm characteristics. Each year the survey contains a consistent set of core questions on business structure and operations, finance, markets and competition, innovation and many others to allow comparison over time (ABS, 2014). The respondents comprise a sample of Small and Medium Enterprises (SMEs) and a census of large firms. The observation unit is Type of Activity Unit (TAU), [1] which is equivalent to a "firm" having either single or multiple Australian Business Number (ABN). When the data is linked to ABS held taxation records from Australian Taxation Office (ATO), there are about 9000 units from 2005-06 to 2011-12. This dataset is called the Super Main Unit Record File (SMURF).

The second dataset is constructed based on ABS held ATO Business Activity Statements and Business Income Tax (BAS-BIT) administrative records. The first year is 2001-02. Although the ATO data is collected on an ABN level, with certain assumptions it can be concorded to a TAU level data and linked to TAU level ABS datasets such as SMURF and the ABS R&D survey. In addition, using the same concordance, non-ABS data such as intellectual property applications can be linked as well the BAS-BIT dataset. As of May 2014, this dataset contains about 9 million records with reported sales revenue value. There are about 800,000 to 850,000 TAUs with both revenue and wage data per year. The ABS (2014) calls this dataset that links BAS-BIT, ABSBR, BCS and Economic Activity Survey (EAS) the Extended Analytic Business Longitudinal Dataset (EABLD).

Both datasets are firm-level panel data which means they include both cross-sectional and time series dimensions. The time dimension of panel data allows the researcher to more carefully identify factors that precede others in time; and the cross-sectional aspect allows the researcher to identify

[1] A Type of Activity Unit is a producing unit comprising one or more business entities, sub-entities or branches of a business entity that can report production and employment activities.

factors that are associated with one unit and not another. Together the data gives a stronger and more sensitive test for causal inference. Panel data analysis is now the standard for micro-economic research.

In this paper, we review the uses and limitations of using the EABLD for policy analysis – with reference where applicable to the SMURF. We then undertake some preliminary investigations to test how knowledge spillovers from neighbourhood firms will affect a given firm's productivity.

Uses and limitations of using the SMURF and EABLD

A description of the dataset

The SMURF comprises a rolling 5-year window sample of about 7000 Australian SMEs and a census of about 2000 large firms who are respondents to annual Business Characteristics Survey. To limit respondent burden, each SME remains in the survey for 5 waves before being replaced by another SME. The greatest value of the SMURF is its rich array of variables from various topics covered by the BCS Survey: Business structure and operations, Finance, Markets and Competition, Innovation, Barriers to Business performance, Skills, and Use of IT (ABS, 2014). For example, SMURF contains basic economic data (revenue, employment, wages, investment, export and capital) and qualitative responses (i.e. yes/no) on main industry, innovation, market, business focus, and government assistance *inter alia*. Unfortunately, the SMURF does not include management or human resource variables. As a major current theory argues that the explanation for the wide variance in productivity differences can be traced to differences in management and human resource practices, this is a critical omission. The way to test this empirically is to model productivity from a firm level dataset that includes both production and managerial variables.

The EABLD is a proof of concept for ABS Firm level Strategy currently being developed to address increasing demand from policy makers and the research community (ABS, 2014). It is a linked dataset comprising ABS held BAS-BIT tax data, ABS Business Register, Business Characteristics Survey and Economic Activity Survey data. It contains about 40 million records since 2001-02. However, not all these 40 million contain records for the basic economic fields. There are about 1.3 million ABNs per year with some economic data, but only 250,000 to 300,000 ABNs per year with complete wage, revenue and assets records. The main missing variable is assets – which are missing for 32% of firms which have complete records for revenue and wages. We estimate that the number of trading businesses is about 800.000 to 850,000, and that firms with missing asset values but positive wage and revenue data are going concerns in the economic sense.

We believe it is desirable to test our models on both the sample of 250,000 to 300,000 firms with reported assets and the larger population of 800,000 to 850,000 firms. Although the former is more complete, it is biased towards large firms and may therefore return biased parameter estimates (depending on our research question). The latter is a relatively unbiased sample and given our econometric estimates generally compare a firm's current state with its past state, the lack of asset data may not compromise the estimates.

The EABLD is a more limited dataset than the SMURF in terms of variables – mainly limited to revenue, wages, costs, exports, 4-digit industry classification – but it does covers all tax paying ABNs in Australia that can be linked to the datasets mentioned above. The EABLD does not contain employment numbers and these need to be estimated from the wage bill data and average weekly earnings.[2] Notwithstanding, the power of this dataset is tremendous. It means we can extract very granular findings on different sectors of the economy, subject to confidentiality. Furthermore, it is possible for us to send the ABS ABN-linked data and then analyse our own data on their dataset.

Use of linked administrative data

Administrative data are data collected without any direct survey of the original sources of information such as individuals, households, firms, etc. (Chetty, 2012). For example, data collected by the public sector when administering school records, income tax filing and the social welfare payment systems are parts of public administrative data. Similarly, "administrative" data such as supermarket inventory and cashier scanner data collected by the private sector are also parts of the administrative data.

There are many benefits from the use of administrative data in empirical research and the potential for these benefits have increased rapidly as more data become available electronically and computer technology required for processing and analysing them progresses. Chetty (2012), for example, has argued that administrative data provide higher quality information without the common problems faced by survey data such as missing observations or sample attrition. In addition, the large sample size (covering the universal or near-universal population) and long longitudinal aspects of administrative data are particularly attractive for development of more robust and rigorous analytical approach. For example, administrative data allows researchers to "conduct sharper tests of existing models and tests of theories that had previously been difficult to assess" (Einav and Levin, 2014).

These benefits and their impacts on research quality are evidenced by the rising adoption of linked administrative data in studies published in top economic journals (Chetty, 2012; Einav and Levin 2014). Looking at micro-data based articles published in *American Economic Review*, *Journal of Political Economy*, *Quarterly Journal of Economics* and *Econometrica*, Chetty (2012) found a significant decrease in the proportion of articles which used only pre-existing survey data and significant increase in the proportion of those which use administrative data over the period 1980-2010. For example, in *Quarterly Journal of Economics*, the proportion of published micro-data studies using pre-existing survey dropped from around 90 per cent in 1980 to around 10 per cent in 2010. In contrast, the proportion of studies using administrative data jumped from less than 15 per cent to around 75 per cent in the same period.

Similarly, Einav and Levin (2014) reviewed a number of important studies to illustrate how the use of public sector administrative records can produce high impact research outcomes across fields. One of such studies is Piketty and Saez (2014) which relies on administrative tax record data to study the long-run income inequality in the US and Europe. Without the administrative data the study would not have been possible because available survey data "cannot measure top percentile incomes accurately because of the small sample size and top coding". In this type of study, it is crucial to include the top percentile population given their "very large role in the evolution of inequality".

[2] The ABS (2015) is developing a Linked Longitudinal Employer-Employee Database based on BAS-BIT and personal income tax information which can provide estimates for employment numbers.

Furthermore, survey data “have a much shorter time span—typically a few decades—than tax data that often cover a century or more”.

In addition to their potential contribution for academic research, administrative data are potentially highly valuable for conducting policy evaluation with clear causal inference. The reason is the universalness or near-universalness of administrative data which allow for data linking across different records or existing survey data in order to track policy outcomes on target population. Einav and Levin (2014) cited a recent study by Akerman et al.’s (2013) on the effects of broadband Internet access by exploiting regional variation in Norway in terms of the timing of broadband rollout and its effects on wages and firm productivity using linked individual and firm tax record data.

Administrative data availability

Administrative data in the form of linked public record data have only been available for researchers in few countries, notably Scandinavian countries such as Denmark and Finland. Researchers from other countries would almost always need to collaborate with the researchers based in which administrative data are made available.

Unfortunately, as stated in Productivity Commission’s annual report (2013, Chapter 1, page 1), “[u]nlike many other countries, Australia makes relatively little use of its public data resources even though the initial costs of making data available would be low relative to the future flow of benefits”. This is despite the abundant of evidence on the value of administrative data for research. Australia clearly lags other countries such as United Kingdom, Denmark, Sweden, Finland, Germany, New Zealand and the Netherlands in providing researchers and policy makers with access to administrative data to conduct research and policy evaluation. Not only the type of linked Australian administrative data accessible to researchers is very limited compared to these countries, the method to access the data is also very costly both financially and, most significantly, in terms of required time for conducting the data analysis.

In the current setup, data analysis using the ABS data is subject to strict confidentiality rules that have significant cost implications. External analysts cannot access the data – they can only submit command file to the ABS who will then run the file.[3] Output from the statistical analysis is then scrutinised by ABS officials who check it does not breach confidentiality before releasing it to the external analysts. It cannot reveal information about any firm. The ABS also scrutinises any report or paper that is released by the external analysts. This process can be time consuming and is costly for the ABS to provide. Accordingly, the external analyst needs to pay the ABS for the data service and allow considerable time for the completion of their work. In contrast, in Denmark, researchers can analyse de-identified micro data directly by remote access to Statistics Denmark’s research server (Statistics Denmark, 2014). In UK, a country which has only recently made linked administrative more readily available for research, there are four research centres in England, Norther Ireland, Scotland and Wales under UK’s Administrative Data Research Network[4] where researchers can come

[3] Subject to approval from the Australian Statistician, secondment to the ABS is a more direct and possibly more cost effective option available to employees from other government agency (ABS, 2014).
[4] This is a “UK-wide partnership between universities, government departments and agencies, national statistics authorities, the third sector, funders and researchers” which accredit researchers and manage

and analyse linked administrative data once they obtain approval. In New Zealand, researchers approved to use the Integrated Data Infrastructure which links administrative data from the tax office, customs service, education ministry and many others can analyse the data directly in one of three data labs located in Wellington, Auckland and Christchurch.[5]

Even with direct access to the data, considerable work cleaning and linking records still needs to be done within the ABS. This task is not trivial. Not only are the datasets large, but different variables are collected at different entity levels (ABN, TAU, enterprise group) and the relationship between them is complex and every changing. It would be most efficient for a central group, such as the ABS, to undertake this upstream level of linking, for all users. ABS officers have a level of technical expertise that is difficult and expensive to replicate.

Why use ABS datasets?

It is not viable for an agency outside the ABS to generate its own firm-level dataset for 2 reasons:

1. Collecting data from businesses is extremely expensive for both the collector and the business. On efficiency grounds, there is only room for one business data collector in the country. Because of this, the ABS collaborates with other primary data collectors to minimise respondent burden.

2. Response rates to business surveys are notoriously low and falling. As of 2014, the expected rate would be about 10 per cent. By contrast the ABS achieves about a 95 per cent response rate because business are required by law to answer the survey. Low response rates are a problem because they may introduce an unknown bias into the data.

The ideal data system must be capable of integrating data from an array of sources—private and public, business and household and an array of formats—cross-sectional and longitudinal, survey and administrative, national and sub-national. Given the history of enterprise data collections in Australia, the most efficient options are to work with the ABS firm-level data sets. Furthermore, the cost and burden on companies of collecting enterprise data mean that it is most efficient for one party to be responsible for collecting the main or master enterprise dataset. Currently, this is done by the ABS in collaboration with the ATO but other government agencies also collect important firm-level administrative data such as intellectual property registrations and data on participation in grant, information, networking or training programs. Rather than requiring the government agencies to collect information already collected by the ABS or ATO, it is more efficient for the agency to be able to link their data into the ABS micro data collection.

Advantages of unit level datasets over aggregated datasets

Currently, the ABS does offer very accessible *industry*-level datasets. However, detailed firm-level data (also called micro data) is normally preferred because:

- aggregated data can combine effects;

approval panels for the use of UK de-identified administrative data (see http://adrn.ac.uk/about for more details).

[5] http://www.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure.aspx for further information.

- many questions, such as firm-level economic decisions, cannot be considered without micro data;
- dynamics, meaning the effect of an activity in one year on subsequent years performance, are difficult to model using aggregate data;
- firm-level data can decompose effects into gross and net;
- aggregate data is too blunt to capture the effects of specific small-scale policies; and
- longitudinal firm-level data can account for self-selection and unobserved characteristics.

Most economic evaluations depend on *observational* data[6] – either from national statistical offices or the program administering unit. However, much statistical theory assumes the data is drawn from *experimental* data. Extra care is needed when dealing with observational data. If there are no factors that determine both selection into treatment and the outcome of that treatment (called a 'confounding' factor), and the number and spread of observations is large, then the analyst can simply compare the outcomes of two groups to get a measure of the treatment outcomes. However, it is rare that the analyst can be sure that there are no confounding factors in observational data. By this we mean, that when a government runs a program, the most (or least) motivated and able firms chose to enter the program with the consequence that it is hard to distinguish these motivation and ability factors from the pure effect of the program. Controlling for confounding factors requires the use of multiple techniques that demand large datasets.

There are five main econometric techniques to control for confounding influences: *multivariate regression* analysis, which depends on confounding factors being measured and included in the data set; *instrumental variable analysis* which tries to control for unmeasured confounding factors but relies on the presence of suitable 'instruments'*; panel data analysis* which controls for unmeasured confounding factors if they are time invariant; *matching analysis* which involves constructing a synthetic control group but only eliminates the effect of measured confounding factors; and *difference-in-difference estimators* which can wash out both macroeconomic influences and time invariant firm-specific unmeasured confounding factors. Which technique is most suitable depends on the properties of the data set.

Quasi-experimental data collected through a 'random assignment' program gives the strongest and most objective results if the number of observations (i.e. firms) is large. The program design rules can eliminate confounding factors (or selection into the treatment) but does require the program administrators to work closely with the analyst. This type of data is still comparatively rare in economic analysis due to political considerations. A random assignment program requires the program administrator (the government) to deny program benefits to some firms on a 'throw of the dice' basis. This could attract unwanted public comment and few governments in the world have

[6] Observational data is information about activities that occurs naturally in the environment. For firms, this might be sales per week, employment levels, profits per year. Experimental data is information that is artificially generated in a laboratory setting. For a scientist this might be an experiment using different chemical strengths of fertiliser on the growth rate of plants, holding all other factors such as sunlight and water constant. There are few experimental datasets in economics. The most common are data generated by paying people to answer questions in a game that simulates a real world example. It is not clear how well these games translate into real world behaviour.

taken this path.[7] NESTA,[8] in the UK, is however getting a consortium of economic evaluators together to promote this form of industry program.

Uses and limitations of SMURF and EABLD

Table 1 summarises the differences between SMUR and EABLD and the implications on the use of the datasets. For example, the larger sample of EABLD allows for more granular analysis of industry and firm size variation. The higher firm coverage also makes EABLD more useful for program evaluation due to more complete information on treatment and control groups associated with the program.

[7] The World Bank has funded some random assignment management programs in India, see Bloom et al (2013).

[8] National Endowment for Science, Technology and the Arts.

Table 1: Comparison of SMURF and EABLD datasets

	SMURF	EABLD
Sample size	Sample but response rate over 95%. Because the ABS does not know in every case why a business does not respond it is not able to estimate a precise rate.[9]	Population, all ABNs returning a BAS or BIT statement concorded to the Type of Activity Unit (TAU) level.[10] Over the period 2001-02 to 2011-12 there are 1,175,169 observations on 357,059 unique entities with wage, asset and turnover data.
	Too small to allow granular analysis by some industries or firm sizes. This means the analysis of specific industries will be limited (causal analysis especially).	Allows granular analysis by industry or firm size. This is the best dataset in Australia for analysing specific market based industries – subject to the variables on offer.
Program evaluation	Unlikely to be able to control for small government programs.	Able to test for effects of small government programs.
Population estimates	Cannot weight to get population estimates	Do not need to weight but still missing many values.
Coverage	Does give systematic coverage to industries E, K, O, P, Q, S.	Financial information may not be meaningful for industries without sales
Selection	Do not know when firm 'disappears' why it is missing. Evidence is that the least profitable firms prematurely exit the dataset.	Firms will systematically disappear if they do not return BAS or BIT (which most likely means no sales or income).
Link to other information	Not viable to link to external ABN information because it's a sample,	Reasonable scope for linking to external ABN information because it's a population.
Counterfactual	Possible if treatment group not too granular.	Possible for quite granular treatment groups
Variable scope	Rich set of quantitative and qualitative variables. Includes all the BAS-BIT variables	More limited set of accounting and economic variables
	Includes several employment variables as integers.	Does not include employment. This needs to be estimated from total wages data.
Dynamics	Limited to 5 years. First observation 2005-06.	All years since 2001-02.
Causality	Can use dynamic feature to infer causal relationships. As firm-level data, it is more robust that industry level data.	Can use preceding (earlier in time) information to infer causality. As firm-level data, it is more robust that industry level data.
Inter-firm externalities	Limited accuracy because dataset is a sample	High accuracy because dataset is a population
Time to analyse	Relatively quick to run programs	Slow to run, needs a 64bit PC to open the whole dataset.
Controlling for firm level unobservable factors	Possible given panel nature	Possible given panel nature

[9] The business may have ceased, moved, merged or just not responded.
[10] A Type of Activity Unit is a producing unit comprising one or more business entities, sub-entities or branches of a business entity that can report production and employment activities.

Why do spillovers matter for public policy?

It is the presence of externalities – the unrequited flow of benefits to householders – that defines whether or not public money should be spent on what is a seemingly private business activity. In the case of knowledge-based activities, externalities are produced when the knowledge generating activities of one business enhances the knowledge and capabilities of unrelated firms, and subsequently leads, via competition, to markets providing products that are better, cheaper or both.

For productivity growth, the creation of knowledge for one's own use is significant, but the exploitation of it by third parties is critical. We know from deduction that the generation and use of knowledge is the only source of productivity growth over the long run. Physical matter is fixed – humans cannot create more – and there is a limit to how fast a person can toil. The only difference between us and our Neanderthal forebears, therefore, is our ideas. And ideas are special. They do not wear out, and unless proven false, an idea will produce a perpetual flow of benefits. It would be difficult to imagine a situation where all these everlasting benefits are fully appropriated by the creator and developer. Accordingly, we expect that spillover benefits, that is, the value captured by consumers, are substantial. Substantial, not only because knowledge is hard to contain, but substantial because the created value lasts forever.

Given this, it is surprising that there is scant evidence in Australia on whether and how new ideas lead to productivity growth. And if they do not, why not? According to Shanks and Zheng (2006), the absence of consistent and robust evidence of the impact of R&D on productivity might be attributed to the presence of noise in the data which has concealed the true relationship. Although R&D is not the only, or the main, source of ideas in the economy, it is the most tractable and quantifiable given the current state of our datasets and is therefore a common starting point for evidence.

What is knowledge?

Döring and Schnellenbach (2006) define knowledge as comprising all cognitions and abilities that individuals use to solve problems, make decisions and understand incoming information. A common but insightful way to describe knowledge is to assess its attributes along the codified-tacit spectrum. At the codified polar we have knowledge which can be translated into text, symbols, algorithms and formulae. As such, it can be disembodied from both the creator and user. At the other end, is tacit knowledge that is only revealed though experience, either because it cannot be fully articulated (as, for example, the knowledge of how to ride a bike), or because its complete appreciation depends on how the receiver and transmitter decode the information. New and innovative knowledge is typically more tacit than codified but most forms of knowledge contain both codified and tacit elements. Diffusion of knowledge takes time and is often incomplete and the more tacit it is, the greater is the difficulty in its transmission.

The trail of knowledge from discovery to consumption can be convoluted, changing and circuitous, and we should not expect to uncover stable quantitative relations over time or across space. According to Döring and Schnellenbach (2006), this pathway is likely to differ between regional and industry groupings according to their institutions and social norms. Some groupings convey knowledge enthusiastically whereas others may simply ignore it or lack supportive institutional frameworks. The networks holding these groupings together are characterized, to a greater or less extent, by routines to share knowledge internally and receive and handle incoming knowledge spillovers.

Conventionally, there are two types of knowledge spillover:

The first are Marshallian spillovers (following Marshall 1890; 1912). These are knowledge flows within industries such as between researchers, entrepreneurs and businesses. Examples of grouping where Marshallian spillovers flourish include the German chemical industries in the 19th century, the Italian footwear industries, the Japanese car manufacturing districts and the semiconductor industries of Silicon Valley. In these examples, knowledge flows between individuals working to solve similar or related problems. The close technological space between individuals and firms aids the transmission of knowledge.

The second type is inter-industry or Jacobian spillovers. These spillovers create economies of scope. A typical spillover may be the application of a well-known method from one industry or technology to an apparently unrelated setting.

We expect that networks and institutions that promote and support how knowledge is conveyed to third-parties will influence the size of spillovers between firms and industries. Empirical work in this area is limited, however, Meagher and Rogers (2004) have undertaken simulations on how the development of knowledge within industries might be influenced by the network structure of relations among firms. These simulations can formally reveal trade-offs between, for example, the quantity of information processed and the time taken to process the information; or the asymmetries in the flow of knowledge traffic between firms can affect aggregate productivity growth.

Model used to estimate the effect of R&D spillovers

To investigate if, and how, R&D activity affects productivity we first need to estimate the productivity of each firm, while making sure there is no reverse causality (feedback from productivity to a firm's decision to invest into R&D). We follow the existing international literature by specifying that the net output of each firm i in year t (Y_{it}) can be represented by a common across-firm Cobb-Douglas production function of the form[11]:

$$Y_{it} \equiv J_{it} K_{it}^{\alpha_k} L_{it}^{\alpha_l} \tag{1}$$

where J_{it} denotes the Solow or production residual, K_{it} denotes the tangible capital stock and L_{it} denotes the size of employment. J_{it} has also been called the intangible capital stock or total factor productivity. We do not need a coefficient or exponent for J_{it} because it is not defined in natural units such as dollars or people. Using the corresponding lower case letters to denote the logarithmic values of the inputs and output above, equation (1) can be rewritten as:

$$y_{it} \equiv j_{it} + \alpha_k k_{it} + \alpha_l l_{it} \tag{2}$$

[11] According to Hall, Mairesse and Mohnen (2010) estimates of R&D elasticities are not highly sensitive to whether output is define as net or gross of material inputs.

We assume that the log of the current production residual (j_{it}) is determined by the firm's measured ability (A_{it}) such that:

$$j_{it+n} = \beta A_{it} + \theta_i + u_{it} \tag{3}$$

where θ_i and u_{it} denote unobserved time-invariant firm-specific and random effects, respectively. We would expect that θ_i includes slow-changing managerial and worker skills. Equation (3) highlights a further complication in the estimation process which is knowing the appropriate time interval (n) between the investment into knowledge (A) and its ensuing effect on intangible capital stock (J).[12] These time lags could vary by the type of change, the magnitude of the change, the industry of the firm or the technology introduced. In the immediate investment phase of an innovation, the effect of A on the stock of intangible capital could well be negative (Holmes, Levine, and Schmitz 2008; Arrow 1962).[13] Therefore, when we calculate the year-by-year effects, we may be averaging the effects over different phases (i.e. a negative, neutral and positive phase) of the life cycle of different innovation. This means that n can be of variable length and it may be more accurate to estimate J as the average over several years such that:

$$j_{i\overline{t+n}} = \beta A_{it} + \theta_i + u_{it} \tag{3a}$$

Substituting Equation (3a) into Equation (2) yields our augmented Cobb-Douglas function:

$$y_{i\overline{t+n}} = \beta A_{it} + \alpha_k k_{i\overline{t+n}} + \alpha_l l_{i\overline{t+n}} + \theta_i + u_{it} \tag{4}$$

The problem with directly estimating Equation (4) is that analysts rarely have reliable measures of the level of A but datasets often have measure of the change in A if defined as the change in knowledge from own R&D (R^o) and from externally captured R&D (R^e). Hence, we may write

$$\Delta y_{i\overline{t+n}} = \gamma^o R^o_{it} + \gamma^e {R^e}_{it} + \alpha_k \Delta k_{i\overline{t+n}} + \alpha_l \Delta l_{i\overline{t+n}} + \Delta u_{it} \tag{5}$$

where we assume the change in A is reflected by own and spillover R&D ($\beta \Delta A_{it} = \gamma^o R^o_{it} + \gamma^e {R^e}_{it}$). (6)

Hall, Mairesse and Mohnen (2010), for example, refer to this method as 'long differencing' (as opposed to the more common year-on-year differencing) if the change is measured over several years.

The variable for 'external R&D' should be constructed in a way that mimics the viscous way in which knowledge travels. An obvious way is to define all externally available R&D as the total level of R&D spending by other firms in the same industry as the subject firm. One of the problems with measuring spillovers in this way is that we may be measuring a third factor that is correlated with R&D and common to all firms in the same industry (such as government tax or regulation change; change in international trade agreements; changes in industrial relations), not spillovers. This third factor is called a confounding factor and means we have poorly 'identified' R&D spillovers. A way around this can be to limit the measured R&D spillovers to firms that are within the same geographic area (assuming the third factor is not geographically bounded). This can be justified if it is thought

[12] Although we would expect the median lag to vary between industries and technologies, Hall, Mairesse and Mohnen (2010) report seven studies that estimate the median to be between 2-3 years.
[13] There are fixed costs associated with installation, fine-tuning new technology, and retraining workers.

that personal contact is important for the transfer of knowledge. An alternative to the 'same industry' is the 'same technology', although it is more difficult to obtain this information at the firm level. An alternative to the same geographical area is the use of inter-industry trade or labour mobility linkages. What is appropriate depends on what are the main channels for knowledge movement and the quality of the transfer, and of course the availability of reliable data.

A significant estimation challenge is obtaining accurate price indices. Typically, official indices from national statistical offices account poorly for changes in product quality. Motor cars and PC are the exceptions.[14] The result is that when a price rises to reflect an improvement in quality, it is assumed to constitute inflation. The CPI and other industry price deflators thus overstate the extent of inflation and understate production levels. Klette (1999) has suggested that we normalise each variable in the production function with respect to its industry average in each year as a substitute for industry-specific price deflators. The alternative is to use time dummy variables. In the latter, the R&D coefficients are biased only to the extent these time dummies do not accurately reflect industry inflation and this inaccuracy is correlated with R&D spending.

Of greater concern is omitted variable bias in equation (6). If the omitted factors are important explanators and are correlated with R&D then the estimation will over or understate the 'true' R&D coefficient. For example, if R&D intensive firms also employ more skilled workers, but the skills of the workforce are omitted from the model then part of the apparent effect of R&D may be due to the more skilled workforce. Crépon and Mairesse (1993) found that accounting for the correlation between worker skill and R&D spending in firms reduced the R&D elasticity by half. However, this effect was only present in the cross-sectional estimates but not panel estimates as differences in worker skills does not vary much over time.

International evidence on the effects of R&D

The main empirical results from recent reviews of this R&D spillover literature are as follows.

- Hall, Mairesse and Mohnen (2010) reviewed 150 international studies and calculated a median R&D elasticity of 0.08. This estimate was higher for cross-sectional studies and lower for panel analysis (because measurement errors have a much greater effect on differenced variables, and we are more likely to have a correlation between R&D and the error term in cross-sectional estimates). Of the firm-level panel estimates, the median elasticity is about 0.07.

- In the same review, Hall, Mairesse and Mohnen (2010) reported that although the estimate of the spillover R&D elasticity was often significant, the size varied considerably depending on the weighting matrix used and whether the estimates are cross-sectional or panel. Part of this variability of the spillover estimate therefore arises because many spillovers are inadvertent benefits received by the firm rather than a strategic decision. They argued that it is more logical to expect consistency in the own-firm returns to R&D because the decision to invest is centred on a rate of return calculation.

[14] The Australian Bureau of Statistics calculate hedonic price indices for computers and uses the Delphi method for cars.

- There is clear and consistent evidence that the strength of spillovers is related to physical distance (Döring and Schnellenbach 2006)

- The Hall, Mairesse and Mohnen (2010) review also found:

  - Higher rates of return for process than product R&D (but according to this may reflect the poor price deflators for improvements in product quality and that new products can have lower sales/higher costs in the short run due to the fixed costs of marketing and re-tooling).

  - Lower rates of private return to publicly funded R&D compared to privately funded R&D. This may arise because the public sector targets R&D investments that are expected to have externalities rather than private benefits.

  - Higher returns to basic R&D (compared with applied research and development). This may arise because basic R&D is very long term and therefore more risky (and demanding of a higher expected rate of return).

- Some studies find spillovers to be negative (see Kafouros and Buckley 2008). It is most probable that this 'finding' arises from imprecise price deflators. To illustrate this effect, assume that the firm undertaking R&D produces a superior product but sells at the same price as competitors who are still producing the earlier model. Sales from the second firm will fall as consumers switch to the superior product, but because the price deflator is the same for all firms in the industry, it will appear as if the non-R&D firms' productivity has fallen even though their ratio of inputs to outputs (i.e. productivity) has not changed.

- The published modelling of R&D spillovers in Australia has only been done at the industry level (Shanks and Zheng 2006; de Rassenfosse and Jensen 2013). Intra-industry spillovers cannot be estimated using industry level data. de Rassenfosse and Jensen (2013) found that a 10 per cent increase in R&D expenditure in all but the focal industry leads to a 2.76 per cent increase in productivity in the focal industry. Shanks and Zheng (2006) examined the effect of public sector and foreign R&D on Australian industry level productivity but reported that within the market sector, there was great uncertainty as to the magnitude of the effect of public and foreign R&D on Australian productivity, as the estimates are very sensitive to model specification.

- In a review of the agriculture economics literature, Alston (2004) concludes that both intra-national and international public agricultural R&D spillovers were responsible for more than half of total measured agricultural productivity growth. However, these impacts can be sensitive to the specifics of the approach taken.

- The Döring and Schnellenbach (2006) review found that:

  - Knowledge spillovers appear to be most relevant in 'young' industries and firms where new knowledge can be assumed to be of special importance. It is possible that young firms have a greater willingness to use new incoming knowledge compared with mature firms where a large fraction of activities is already following established routines that are costly to change.

- Young and small firms often do not have the capacities to maintain large-scale R&D departments themselves and, therefore, rely on external sources of knowledge to a larger extent than more mature firms with extensive own R&D activities.

- Knowledge is absorbed more easily in areas with higher productivity levels and a larger stock of knowledge. This supports the theory that knowledge is acquired via a cumulative process during which new incoming knowledge can only be used if necessary complementary knowledge already exists.

Preliminary estimation results

Given the time taken to run each model (over a week) it has not been possible to test equation (5) and instead Tables 2 to 6 present preliminary results from testing equation (5a):

$$y_{it+1} = \gamma^{o} R^{o}_{it} + \gamma^{e} {R^{e}}_{it} + \alpha_k k_{it+1} + \alpha_l l_{it+1} + u_{it+1} \quad (5a)$$

Equation (5a) does not difference the output, assets and labour variables and uses a rather simplistic and inflexible time lag between R&D and output of one year. As such the results should not be taken literally. Rather the results below indicate the type of modelling that can be done.

In Table 2 we test using the full dataset for the effects of own R&D and spillover R&D on firm output. Spillover R&D is defined in one of three ways: the R&D done by other firms in the same 2-digit industry and same state; the same 3-digit industry and state; and the same 4-digit industry and state. The estimated coefficient for own R&D is positive and statistically significant but small compared with estimations overseas. The estimated coefficient for spillover R&D is positive and statistically significant for the 2 and 3 digit variables and also small compared with estimations overseas.

These estimations are for all industries (except agriculture, fisheries and forestry). Table 3 presents results from estimating this model for each major industry separately. The coefficients for own R&D are positive and statistically significant for B (mining); C (manufacturing); I (transport, postal and warehousing); J (information media and telecommunications); K (financial and insurance services); M (professional, scientific and technical services) and Q (health care and social assistance). We only tested for 2-digit industry based spillovers and these were only found to be positive in E (construction); G (retail trade); I (transport, postal and warehousing) and M (professional, scientific and technical services). Note however, that these results should not be taken at face value as the model is still at an early stage of development.

Tables 5, 6 and 7 disaggregate these estimates by firm size; first using the 2-digit definition of an R&D spillover and then using the 3-digit and 4-digit definitions. The 2 and 3 digit estimations are consistent: they find that micro firms (under 5 employees) and medium sized firms (20 to 199 employees) exhibit signs of receiving R&D spillovers from other firms in their industry and state. Small firms (6 to 19 employees) and large firms (over 200 employees) did not show any effects.

The proceeding models have defined spillovers as intra-industry knowledge flows. In Table 7 we model inter-industry R&D knowledge flows – that is, how R&D performed in industry X affects productivity in all other industries. The first column tests for the effects of R&D coming out of all industries separately and the subsequent columns test each R&D industry on its own (if there is a high degree of correlation between R&D spending across industries then the first equation could produce unreliable results). These results show that the source of most R&D spillovers is C (manufacturing); H (accommodation and food services) and L (rental, hiring and real estate services). Again, we re-iterate the warning about the provisional nature of these findings.

Taken together these estimations illustrate what can be done with the data. The model has considerable room for improvement which will hopefully make the results more economically sensible and more consistent with overseas studies. In particular, we need to either capitalise R&D or difference output, labour and assets. In addition, we should try the Olley-Pakes estimation to correct for firm exit and possible correction between the error term and labour and capital.

Table 2: Estimation of R&D impact intra-industry R&D spillovers on firm output

VARIABLES	All industries (ex A)	All industries (ex A)	All industries (ex A)
Log (value of tangible assets) (t)	0.0069264***	0.0069244***	0.0069232***
	(0.000)	(0.000)	(0.000)
Log (employment) (t)	0.5021389***	0.5021593***	0.5021447***
	(0.001)	(0.000)	(0.001)
Log (value of own R&D) (t-1)	0.0060273***	0.0060155***	0.0060000***
	(0.001)	(0.001)	(0.001)
Log (value of 2-digit industry state R&D) (t-1)	0.0017328***		
	(0.000)		
Log (value of 3-digit industry state R&D) (t-1)		0.0008397***	
		(0.000)	
Log (value of 4-digit industry state R&D) (t-1)			0.0002529
			(0.000)
Year dummies	Yes	Yes	Yes
Observations	2,971,343	2,971,343	2,971,343
R-squared	0.1805	0.1805	0.180
Number of TAUs	845,703	845,703	845,703

Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Extended Analytic Business Longitudinal Database

Table 3: Estimation of intra-industry spillovers by 1-digit industry on firm output

	1-digit industry							
VARIABLES	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**
Log (value of tangible assets)	-0.0005764	0.0066555***	0.0062685***	0.0086735***	0.0064296***	0.0049799***	0.0072663***	0.0072504***
	(0.003)	(0.000)	(0.002)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Log (employment)	0.5226745***	0.5532750***	0.4991057***	0.4885248***	0.4906930***	0.4866691***	0.5922836***	0.4660392***
	(0.016)	(0.003)	(0.013)	(0.002)	(0.004)	(0.003)	(0.003)	(0.003)
Log (value of own R&D)	0.0207275*	0.0043153*	-0.0222038*	0.0069310	0.0041898	0.0045437	0.0027849	0.0228757***
	(0.011)	(0.002)	(0.013)	(0.007)	(0.006)	(0.010)	(0.023)	(0.008)
Log (value of 2-digit industry state R&D)	-0.0044848	0.0025364	0.0032893	0.0062525***	-0.0035441*	0.0029400***	0.0000954	0.0011962*
	(0.012)	(0.002)	(0.003)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,931	235,194	11,553	522,648	160,769	322,136	227,189	175,056
R-squared	0.205	0.197	0.174	0.191	0.115	0.127	0.220	0.186
Number of TAUs	1,607	61,721	3,352	155,389	44,606	92,977	71,971	49,914

Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Extended Analytic Business Longitudinal Database

	1-digit industry						
VARIABLES	**J**	**K**	**L**	**M**	**Q**	**R**	**S**
Log (value of tangible assets) (t)	0.0076238***	0.0069311***	0.0064554***	0.0071022***	0.0052133***	0.0078115***	0.0063383***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
Log (employment) (t)	0.4807451***	0.4140106***	0.4773163***	0.5357421***	0.4647341***	0.4926519***	0.5070561***
	(0.007)	(0.003)	(0.003)	(0.002)	(0.003)	(0.007)	(0.003)
Log (value of own R&D) (t-1)	0.0200883**	0.0142371**	-0.0064163	0.0046494*	0.0328335***	0.0061507	-0.0048527
	(0.009)	(0.007)	(0.013)	(0.003)	(0.012)	(0.020)	(0.010)
Log (value of 2-digit industry state R&D) (t-1)	0.0040388	-0.0036733*	-0.0011620	0.0058988**	-0.0003887	-0.0005475	-0.0019384
	(0.003)	(0.002)	(0.001)	(0.003)	(0.001)	(0.002)	(0.001)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	29,137	132,658	150,836	465,465	239,386	44,951	248,434
R-squared	0.177	0.144	0.155	0.251	0.164	0.146	0.192
Number of TAUs	8,307	40,098	43,815	129,828	60,873	12,876	70,689

Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Extended Analytic Business Longitudinal Database

Table 4: Estimation of R&D impact 2-digit intra-industry R&D spillovers on firm output, by firm employee size

VARIABLES	all sizes	Under 5 employees	5-19 employees	20-199 employees	Over 200 employees
Log (value of tangible assets) (t)	0.0069264***	0.0070344***	0.0028531***	0.0034466***	0.0025452**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Log (employment) (t)	0.5021389***	0.4403259***	0.8098858***	0.8024972***	0.6921485***
	(0.001)	(0.001)	(0.003)	(0.006)	(0.016)
Log (value of own R&D) (t-1)	0.0060273***	-0.0032042	0.0031237	-0.0012066	0.0008298
	(0.001)	(0.004)	(0.002)	(0.002)	(0.003)
Log (value of 2-digit industry state R&D) (t-1)	0.0017328***	0.0025016***	0.0007669	0.0025965***	0.0008295
	(0.000)	(0.000)	(0.000)	(0.001)	(0.003)
Year dummies					
Observations	2,971,343	2,175,974	752,053	191,722	15,810
R-squared	0.180	0.149	0.131	0.152	0.183
Number of TAUs	845,703	694,868	236,460	57,430	4,689

Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Extended Analytic Business Longitudinal Database

Table 5: Estimation of R&D impact 3-digit intra-industry R&D spillovers on firm output, by firm employee size

VARIABLES	all sizes	Under 5 employees	5-19 employees	20-199 employees	Over 200 employees
Log (value of tangible assets) (t)	0.0069244***	0.0368717***	0.0028507***	0.0034558***	0.0025543**
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)
Log (employment) (t)	0.5021593***	0.4519065***	0.8098596***	0.8023326***	0.6921689***
	(0.001)	(0.001)	(0.003)	(0.006)	(0.016)
Log (value of own R&D) (t-1)	0.0060155***	0.0030361	0.0031084	-0.0012413	0.0008533
	(0.001)	(0.005)	(0.002)	(0.002)	(0.003)
Log (value of 3-digit industry state R&D) (t-1)	0.0008397***	0.0006876***	-0.0000898	0.0016133**	0.0028709
	(0.000)	(0.000)	(0.000)	(0.001)	(0.003)
Year dummies	Yes	Yes	Yes	Yes	Yes
Observations	2,971,343	924,452	752,053	191,722	15,810
R-squared	0.180	0.156	0.131	0.152	0.183
Number of TAUs	845,703	294,231	236,460	57,430	4,689

Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Extended Analytic Business Longitudinal Database

Table 6: Estimation of R&D impact 4-digit intra-industry R&D spillovers on firm output, by firm employee size

VARIABLES	all sizes	Under 5 employees	5-19 employees	20-199 employees	Over 200 employees
Log (value of tangible assets) (t)	0.0069232***	0.0368635***	0.0028510***	0.0034563***	0.0025357**
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)
Log (employment) (t)	0.5021447***	0.4518723***	0.8098625***	0.8022413***	0.6922002***
	(0.001)	(0.001)	(0.003)	(0.006)	(0.016)
Log (value of own R&D) (t-1)	0.0060000***	0.0030386	0.0031101	-0.0012847	0.0008150
	(0.001)	(0.005)	(0.002)	(0.002)	(0.003)
Log (value of 4-digit industry state R&D) (t-1)	0.0002529	0.0000387	0.0000132	0.0025061***	-0.0003805
	(0.000)	(0.000)	(0.000)	(0.001)	(0.002)
Year dummies	Yes	Yes	Yes	Yes	Yes
Observations	2,971,343	924,452	752,053	191,722	15,810
R-squared	0.180	0.156	0.131	0.152	0.183
Number of TAUs	845,703	294,231	236,460	57,430	4,689

Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Extended Analytic Business Longitudinal Database

Table 7: Estimation of inter-industry spillovers by 1-digit industry on firm output

VARIABLES	All ind (ex A)	All ind (ex A)	All ind (ex A)	All ind (ex A)	All ind (ex A)	All ind (ex A)	All ind (ex A)	All ind (ex A)	All ind (ex A)
Log (value of tangible assets) (t)	0.0069114***	0.0069280***	0.0069234***	0.0069230***	0.0069229***	0.0069224***	0.0069248***	0.0069225***	0.0069194***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log (employment) (t)	0.5019618***	0.5021002***	0.5021202***	0.5021375***	0.5021365***	0.5021451***	0.5021303***	0.5021454***	0.5019917***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Log (value of own R&D) (t-1)	0.0060258***	0.0060289***	0.0059933***	0.0060469***	0.0060012***	0.0060036***	0.0060035***	0.0060009***	0.0059822***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Log (value of industry A state R&D) (t-1)	0.0023258***								
	(0.001)								
Log (value of industry B state R&D) (t-1)	-0.0034623***		-0.0022921***						
	(0.001)		(0.001)						
Log (value of industry C state R&D) (t-1)	0.0039346**			0.0050282***					
	(0.002)			(0.001)					
Log (value of industry D state R&D) (t-1)	-0.0034406***				-0.0001667				
	(0.001)				(0.001)				
Log (value of industry E state R&D) (t-1)	-0.0026046***					-0.0025040***			
	(0.001)					(0.001)			
Log (value of industry F state R&D) (t-1)	0.0003376						-0.0014640**		
	(0.001)						(0.001)		
Log (value of industry G state R&D) (t-1)	0.0022469***							0.0007812	
	(0.001)							(0.001)	
Log (value of industry H state R&D) (t-1)	0.0025554***								0.0025485***
	(0.000)								(0.000)
Year dummies									
Observations	2,971,343	2,971,343	2,971,343	2,971,343	2,971,343	2,971,343	2,971,343	2,971,343	2,971,343
R-squared	0.181	0.181	0.180	0.180	0.180	0.180	0.180	0.180	0.181
Number of TAUs	845,703	845,703	845,703	845,703	845,703	845,703	845,703	845,703	845,703

Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Extended Analytic Business Longitudinal Database

VARIABLES	All ind (ex A)	All ind (ex A)	All ind (ex A)	All ind (ex A)	All ind (ex A)	All ind (ex A)	All ind (ex A)	All ind (ex A)	All ind (ex A)
Log (value of tangible assets) (t)	0.0069280***	0.0069232***	0.0069224***	0.0069215***	0.0069257***	0.0069225***	0.0069224***	0.0069222***	0.0069252***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Log (employment) (t)	0.5021002***	0.5021367***	0.5021260***	0.5021153***	0.5021359***	0.5021455***	0.5021378***	0.5021434***	0.5021313***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Log (value of own R&D) (t-1)	0.0060289***	0.0060009***	0.0059980***	0.0060019***	0.0060118***	0.0060042***	0.0060024***	0.0060008***	0.0060112***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Log (value of industry I state R&D) (t-1)	0.0003869	0.0001052							
	(0.000)	(0.000)							
Log (value of industry J state R&D) (t-1)	-0.0020235**		-0.0017273***						
	(0.001)		(0.001)						
Log (value of industry K state R&D) (t-1)	-0.0016506**			-0.0023578***					
	(0.001)			(0.001)					
Log (value of industry L state R&D) (t-1)	0.0033680***				0.0026109***				
	(0.000)				(0.000)				
Log (value of industry M state R&D) (t-1)	-0.0030623**					-0.0050235***			
	(0.001)					(0.001)			
Log (value of industry Q state R&D) (t-1)	-0.0002754						0.0004854		
	(0.001)						(0.001)		
Log (value of industry R state R&D) (t-1)	0.0001012							0.0002161	
	(0.000)							(0.000)	
Log (value of industry S state R&D) (t-1)	-0.0028398***								-0.0021391***
	(0.001)								(0.001)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,971,343	2,971,343	2,971,343	2,971,343	2,971,343	2,971,343	2,971,343	2,971,343	2,971,343
R-squared	0.181	0.180	0.180	0.180	0.181	0.180	0.180	0.180	0.180
Number of TAUs	845,703	845,703	845,703	845,703	845,703	845,703	845,703	845,703	845,703

Notes: Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1
Source: Extended Analytic Business Longitudinal Database

Mechanisms to increase spillovers

It would be premature to say what the ultimate effects of R&D spending have been in Australia from the models estimated above. However, we should be cognisant of the possibility that the magnitude of effects from own R&D and spillover R&D may be small in Australia as a whole compared with overseas studies (as reviewed above); or in some local industries compared with others. We may want to think now about the implications this may have for policy.

A low elasticity (and rate of return) on R&D spillovers compared with international estimates can mean one of two things. Either the nature of R&D in Australia is very tightly held within the originating business or has very little applicability to other Australian businesses in the same industry. Or, the institutions and networks that link businesses with potentially overlapping R&D and technology interests are poor by international standards.

Thin markets and the absence of deep market structures could lead to a lack of applicability between businesses operating in what appears to be the same industry, but in reality may be quite a different type of technology.

The second reason – sparse and poor networking and collaborating institutions – is consistent with international rankings of the Global Innovation Index and the World Economic Forum. Our poor performance in this area is also an opportunity for us. We do not have to reinvent the wheel here, as there are good examples overseas of mechanisms for maximising to external benefits from R&D. These include, but are not limited to:

- The public funding of bipartisan intermediary bodies which organise seminar, workshops and events to bring related industry people together. Depending on the technology and the history of regions, these might be universities, peak academies, public research institutes, industry organisations or even large private businesses. These relate to strengthening the ties between research organisations and industry, making education and training more responsive and improving the financial sector's knowledge of and confidence in manufacturing

- A culture of trust, reciprocity between businesses with common interests. Overseas research has shown that the frequency of interaction between R&D employees from different firms has a positive impact on the frequency of innovations in these firms. It is important that people expect the relationship to be reciprocal regarding the quality and quantity of knowledge that was to be exchanged. A lack of reciprocity results in people refusing to act as a source of knowledge spillover (Schrader 1991).

- Personal social networks and informal, face-to-face communication between individuals who are already mutually acquainted allows the identification of entrepreneurial opportunities (Sorenson 2005).

- Countries such as the USA, Germany and the lowlands of Belgium, Netherlands and Denmark have been operating schemes that underpin the risk of innovation for business and create networks of trust and collaboration between the research sector and businesses for many decades. The rise and dominance of Silicon Valley would not have happened without the US Departments of Defence and Energy DARPA[15] and ARPA-E[16] projects. The SBIR[17] programs in the US are legendary for co-investing in the successful commercialisation of frontier technologies. The Fraunhofer Institutes make Germany a leader in applied research as well as encouraging a flexible, autonomous and entrepreneurial approach to the society's research priorities. Recently, the UK has embarked on similar schemes. Extension services, whereby originators of research and new ideas, physically show their ideas to firms have been found to be effective. It is well know that face-to-face contact is needed to convey tacit knowledge. This type of knowledge is difficult to convey on paper or on-line and can only be learned through showing how.

- At issue for the industries survival is the need for Australia to 'Think global, act local'. We can't achieve the full value from our work and our talents unless we are integrated into the global supply chain. But neither can we achieve frontier advances without a close knit community of researchers, designers, financiers, entrepreneurs and manufacturers who use trust and confidence to make innovation successful. These tight relationships are not happening, or not happening in a comprehensive and supportive way.

- Australia punches above its weight in terms of research quantity and quality but its translation into value is poor and often limited to high profile cases such as CSIRO's wifi or Fraser's Gardasil. There is low hanging fruit here from improving the efficacy of collaboration between industry and public research. We need innovation hubs to develop solutions for global value chain and forge new models of working together; and stronger incentives to ensure our manufacturing capabilities and the diffusion of knowledge are developed in step with our research capabilities.

- There is a poor alignment of incentives between the research and industry sectors. The financial incentives in our universities and research institutes are too tightly focussed on the old 'publish or perish' indicators.

References

ABS (2014) 'Australian Bureau of Statistics submission to the Productivity Commission's Inquiry into Business Set-up, Transfer and Closure" (available online at http://www.pc.gov.au/__data/assets/word_doc/0008/187244/sub024-business.docx; checked on 23-06-2015).

ABS (2015) 'Use of a Prototype Linked Employer-Employee Database to Describe Characteristics of Productive Firms', Research Paper, Cat. No. 1351.0.55.055.

15 Defense Advanced Research Projects Agency.
16 Advanced Research Projects Agency-Energy.
17 Small Business Innovation Research program

Akerman, I. Gaarder, M. Mogstad (2013), The skill complementarity of broadband Internet, Institute for the Study of Labor (IZA) Discussion Paper No. 7762 (downloaded from http://ftp.iza.org/dp7762.pdf).

Alston, J. (2004) 'Spillovers', Australian Journal of Agricultural and Resource Economics, 46, 315–346.

Chetty, R. (2012) 'Time Trends in the Use of Administrative Data for Empirical Research', presentation slides (downloaded from:

http://obs.rc.fas.harvard.edu/chetty/admin_data_trends.pdf on 18-06-2015).

de Rassenfosse, G. and Jensen, P. (2013) Measuring Industry-Level R&D Spillovers in Australia, A Report for Department of Industry, Innovation, Science, Research and Tertiary Education, Melbourne Institute of Applied Economic and Social Research, University of Melbourne, mimeo.

Döring, T. and Schnellenbach, J. (2006) 'What do we know about geographical knowledge spillovers and regional growth?: A survey of the literature', Regional Studies, 40:03, 375-395.

Dowrick, S. (2002) Investing in the Knowledge Economy: implications for Australian economic growth, Paper prepared for the 2002 Economic and Social Outlook Conference, mimeo, ANU, Canberra.

Einav, L. and J. Levin (2014) 'Economics in the age of big data' Science 346 (6210), DOI: 10.1126/science.1243089 (downloaded from:

http://web.stanford.edu/~leinav/pubs/Science2014.pdf on 18-06-2015).

FSI Final Report (2014), (downloaded from http://fsi.gov.au/publications/final-report/ on 18-06-2015).

Griffiths, W. and Webster, E. (2010) 'The Determinants of Research and Development and Intellectual Property Usage among Australian Companies, 1989 to 2002', Technovation, 30, 471–481.

Hall, B., Mairesse, J. & Mohnen, P. (2010), 'Measuring the returns to R&D', in B. Hall & N. Rosenberg, eds, Handbook of the Economics of Innovation, North-Holland, Amsterdam.

Kafouros, M. and Buckley, P. (2008) 'Under what conditions do firms benefit from the research efforts of other organizations?' Research Policy, 37, 225–239

Meagher, K, and Rogers, M, (2004) 'Network density and R&D spillovers' Journal of Economic Behavior & Organization, 53, 237–260

Piketty, T. and E. Saez (2014), 'Inequality in the long run' Science, 344 (6186), 838–843. oi: 10.1126/science.1251936.

Productivity Commission (2013), 'Chapter 1: Using administrative data to achieve better policy outcomes' Annual Report 2012–13, Commonwealth of Australia, Canberra (Available online: http://www.pc.gov.au/about/governance/annual-reports/2012-13/02-annual-report-2012-13-chapter1.pdf).

Shanks, S. and Zheng, S. (2006) Econometric Modelling of R&D and Australia's Productivity, Productivity Commission, Staff Working Paper, Canberra, April 2006.

Statistics Denmark (2014) The Danish System for Access to Micro Data (Available online: http://www.dst.dk/ext/645846915/0/forskning/Access-to-micro-data-at-Statistics-Denmark_2014--pdf).