NPL REPORT IEA 16

EMPLOYMENT GROWTH AND R&D SPENDING AMONG COMPANIES THAT ENGAGE WITH THE NMS LABORATORIES

DR SOLOMON OLAKOJO, MIKE KING

MARCH 2023

Employment growth and R&D spending among companies that engage with the NMS laboratories

Solomon Olakojo, Mike King
Analysis and Evaluation, Partnerships Directorate

ABSTRACT

The impact of employment growth on firms' R&D spending is underpinned in the accelerator principle, which plays a key role in the theory of investment. Existing evidence showed that firms which engaged with NMS experienced employment growth. However, the impact of the observed growth on their R&D spending is yet to be ascertained. This study, therefore, is centred on the short-run relationship between variations in the level of R&D spending among companies engaging with the NMS laboratories and the changing level of employment within these companies. A theoretical model which predicts that an increased rate of employment growth among companies should lead them to increase the level of investment — some portion of which will be investment in intangible assets — was used. Administrative data from NPL (e.g., invoicing records) and financial data from the FAME database were combined to create a company-level panel dataset for the companies that engaged with the NMS between 2009 and 2015. Estimations are based on combining a series of econometric techniques that account for missing R&D data for some of the observations. The headline results confirmed that employment growth leads to higher levels of R&D spending among the growing companies. The elasticity of R&D with respect to employment is about 0.368, meaning that a 10% increase in a company's workforce will generate a 3.68% increase in its level of R&D spending. The study concluded that employment growth explains about 13% of the variation in private R&D spending.

ISSN 2633-4194

https://doi.org/10.47120/npl.IEA16

National Physical Laboratory
Hampton Road, Teddington, Middlesex, TW11 0LW

Approved on behalf of NPLML by
David Skelton, Strategic Programme Leader.

CONTENTS

EXECUTIVE SUMMARY....1
1. INTRODUCTION TO R&D LEVERAGE ANALYSIS....5
2. LITERATURE REVIEW....6
2.1 FIRM GROWTH AND R&D EXPENDITURE....7
2.2 INNOVATIVE FIRMS AND GROWTH....9
2.3 APPLICATION TO R&D LEVERAGE ANALYSIS....9
3. THEORETICAL FRAMEWORK....11
3.1 THE GROWING IMPORTANCE OF INTANGIBLE ASSETS....12
3.2 MODELLING THE EFFECT OF GROWTH ON INVESTMENT....13
4. DATASET, SPECIFICATIONS, AND ESTIMATION TECHNIQUES....14
4.1 DATASET....15
4.2 SPECIFICATION OF THE MODEL....16
4.3 HECKMAN SELECTION MODEL....18
4.4 ESTIMATION USING THE METHOD OF CORRELATED RANDOM EFFECTS....19
5. SUMMARY STATISTICS....19
6. REGRESSION TABLES AND SPECIFICATION TESTS....21
6.1 ESTIMATION OF THE SELECTION MODEL....21
6.2 ELASTICITY ESTIMATIONS....21
6.3 FIRST DIFFERENCE ESTIMATIONS....25
6.4 LONG RUN ESTIMATIONS....26
7. INFERENCES....26
7.1 APPLICATION 1 – SOURCE OF EMPLOYMENT GROWTH AND CONTRIBUTION OF EMPLOYMENT GROWTH TO R&D SPENDING....27
7.2 APPLICATION 2 - EFFECT OF PUBLIC SUPPORT ON R&D SPENDING....28
8. ROBUSTNESS TESTS AND CAVEATS....29
9. CONCLUSION....29
REFERENCES....31
ANNEX A - FULL DERIVATIONS OF THEORETICAL FRAMEWORK....34
MAIN ASSUMPTIONS AND MODEL SET UP....34
FINDING THE EQUILIBRIUM CAPITAL INTENSITY....34
IMPLICATIONS OF FIXING SOME PARAMETERS....35
IMPLICATIONS FOR INVESTMENT....35
ANNEX B: SORTING OUT ISSUES WITH SIC CODES....36
ANNEX C: INTERPOLATION OF MISSING EMPLOYMENT AND REAL TURNOVER VALUES....37
ANNEX D: AUXILIARY TABLES....39

ANNEX E: LONG-RUN ANALYSIS....42

ANNEX G: ASSESSING THE IMPACT OF CONTINOUS INTERVENTION ON EMPLOYMENT GROWTH WITH DOSE RESPONSE MODEL 44

FRAMEWORK OF DOSE RESPONSE MODEL 44

IMPACT OF CONTINUOUS TREATMENT ON EMPLOYMENT GROWTH OF SUPPORTED FIRMS 45

EXECUTIVE SUMMARY

Compared to other OECD countries of a comparable size, the UK invests relatively little in productive capital, and spends only 1.7% of its GDP on R&D. This is a concern given the well-established link between investment, labour productivity, and the earning power of citizens. Hence, the UK government has a target of increasing the UK's R&D expenditure to 2.4% of GDP by 2027; and increasing the private sector's investment in R&D is critical to achieving this target.

The progress toward achieving the 2.4% target is currently somewhat opaque due to recent developments in the methodology used for the business research and development (BERD) survey. The previous approach, used in BERD statistics, was reportedly characterised by under-coverage of small businesses. However, the most recent estimates have not yet been incorporated into the national accounts. There is optimism that some progress has been made, but until the revised estimates have been incorporated into national accounts, a definitive estimate of R&D as a percentage of GDP is yet to be ascertained.

Two important analytical questions for the economists evaluating the NMS have been prompted by government's attempts to meet the 2.4% challenge:

1. Given that successive governments have had mixed success with attempts to incentivise business R&D spending using tax subsidies, through what other mechanisms might government most effectively influence the level of business expenditure on R&D? For example, one might want to consider programmes that increase the supply of human capital and encourage employment growth among inherently R&D intensive companies.
2. How can economists best quantify the net-additional effect of innovation support programmes on businesses' R&D spending? Moreover, given the near impossibility of applying the preferred econometric methods (dynamic panel data models) to estimating the leverage rates of smaller innovation programmes (less than £100 million), what other approaches might one want to consider?

To achieve the goal of raising the UK's R&D intensity, successive governments have tended to focus on policies aimed at reducing the user-cost of R&D, which remains central to a growing portfolio of innovation policies and strategies. That is, policy makers have mainly attempted to raise business spending on R&D through the introduction of subsidies that lower the cost R&D projects. This was done principally through a series of modifications to R&D tax credits (e.g., through the introduction of RDEC).

The traditional focus on offering financial incentives seems reasonable given that basic economic theory tells us that this should encourage firms to fund R&D projects that would have been unprofitable without such subsidies.[1] Because of strenuous efforts of multiple governments over decades, there has been some discernible effect on R&D as a percentage of GDP.[2] However, the current scale of subsidies are such that there is a danger of

1 Reducing the user-cost of capital, by reducing corporation tax in a targeted way, encourages firms to increase the capital intensity (increase the level of capital per employee) of their production process or business operations.

2 Originally, R&D tax credits were split between a generous SME scheme, and less generous scheme for large companies. This meant that, before 2012, large companies were excluded from accessing the payable credits available through the SME scheme. Consequently, large companies only benefited from R&D tax credits if they already paid substantial corporation tax in the UK, which isn't the case for all such corporations. So, to further incentivise R&D spending in the UK, RDEC was introduced in 2012 so that large, multi-national corporations could benefit from an above-the-line payable credit based on their volume of eligible R&D expenditure.

inadvertently crowding out the private spending on R&D.[3] (This concern is principally based a highly cited study by two OECD economists that says the tipping point occurs when the subsidy rate reaches about 25%.) Given that we may be reaching the limits of what tax subsidies can achieve, what other mechanism might policy makers want to try? The implication of our study is that policy makers may want to consider measures to increase the supply of human capital, and boost employment growth among R&D intensive sectors of the economy.

Alongside coming up with ideas for potential policy interventions, economists have also tried to develop ways to quantify the effect of public support on the R&D spending among the supported businesses. Econometric approaches are fully appropriate for assessing the impact of large programmes such as RDEC (accounting for over two billion pounds of public spending each year). Unfortunately, missing data combined with the skewed nature of R&D spending, makes it extremely difficult for smaller innovation programmes to use the same econometric methods. Hence, when it comes to estimating a leverage rate, smaller innovation programmes often end up relying on beneficiary surveys and self-reported data about co-funding. Even granting that the claims about project-level co-funding are accurate[4], it remains unclear how much of the reported R&D spending is truly net-additional.[5] If econometric approaches do not work well for smaller programmes and there are questions over the reliability of self-reported leverage rates, then by what other means might we quantify the effect of public support on private R&D spending? The implication of this study is that a reasonable leverage estimate could be made by combining an estimate of the elasticity of R&D (with respect to changes in employment) with what is known about the programme's effect on employment from existing econometric studies.

The issues raised above encouraged the authors of this study to develop a novel proposal for how to quantify the effect of public support (delivered through smaller programmes) on R&D spending among the supported businesses. Firstly, this study confirms earlier studies that found a strong connection between investment in R&D and employment growth at the firm-level. That is, this study estimates the elasticity of changes in R&D spending with respect to changes in employment. Secondly, this study explores the implication of this result to the development of future policies intended to raise R&D spending. Moreover, the results can be used to quantify the effect of a programme on R&D spending using what is already known about the effect of such a programme on employment growth.

This study has two main implications for policy makers: 1) perhaps, more attention should be focused on skills and employment growth among inherently innovative companies to raise R&D expenditure; and 2) the study provides a means by which economists may quantify the effect that smaller innovation programmes have on businesses' R&D spending.

- Firstly, this study implies that greater R&D spending is a natural corollary of employment growth among R&D intensive companies, particularly with the increased employment of STEM graduates. Note that this happens even if public support has no direct effect on the amount of R&D spending per employee, implying that increased R&D spending is an indirect effect of employment growth. Therefore, it might be possible to raise R&D spending through policies whose immediate effect is to

3 (Guellec & Van Pottelsberghe, 2003) show an increased effectiveness of R&D subsidies up to a threshold of 13%, after which further increases in generosity start to decrease the leverage rate; beyond the 25% threshold, further increases have a negative effect on the level of private funding committed to R&D.

4 There is a concern about the 'recycling' or double counting of public funding when businesses make use of multiple programmes. Also, how do you best classify the R&D spending of for-profit laboratories who receive most of their income from government?

5 How much of the co-funding is deadweight and so would have happened anyway? And, to what extent did the existence of support displace R&D resources from one project to the subsidised project?

incentivise employment growth whilst also increasing the supply of human capital through subsidised education and training.
- Secondly, in many cases, programmes that support innovation among business have robust econometric evidence for effects on employment growth. The production of such evidence is enabled by the availability of accurate and comprehensive employment data from the Business Structure Database (BSD), as well as low variation in employment leading to low standard errors in regression analysis. Hence, a potential application of this result (knowing the elasticity of R&D with respect to employment) is that it gives us a way of estimating the effect of existing programmes on private R&D spending. Essentially, the effect of a programme on R&D spending could be found simply by multiplying the elasticity of R&D with respect to employment, φ, by the programme's effect on employment growth.

The implication spelt out in the first bullet point also has a bearing on the design of future programmes. The implication being that skills-based programmes that encourage employment growth among innovative businesses would also lead to positive effects on business R&D spending in the future.

It is useful to provide some explanation of why receiving support from NPL, or other Public Sector Research Establishments (PSREs), leads to an increase in economic opportunity and employment growth among the supported firms. Knowledge increases the effectiveness with which firms bring together the factors of production (capital and labour) to produce goods or services. Hence, a simple explanation that fits well with the conventional production function framework is that support from PSREs provides access to specialist expertise that most firms cannot afford to maintain in-house. This unaffordability is even more apparent when this expertise isn't needed all the time, despite being intermittently critical to projects and problem solving. (To a large extent, the degree of specialisation of scientists within NPL is what distinguishes it from more general research-based organisations.) By engaging with the PSREs, firms receive an incremental rise in the quality and scope of the knowledge that's available to them; and this increase in knowledge occurs because the PSRE's own knowledge increases as a consequence of the publicly funded R&D being conducted by its scientists.[6] This increase in effectiveness leads to an increase in labour productivity (output per employee), which encourages a supported firm to expand to the point where its marginal revenue product of labour (MRPL) once again equals the wage rate in the labour market. Improved productivity of supported firms draws more labour to them because of their ability to pay higher wages and the need for more resources as these firms approach full capacity (i.e., maximum output that can be produced with existing capital and labour resources). This would lead to higher R&D investment for innovating firms since R&D spending is split between personnel remuneration (scientists and other professional employees) and spending on R&D equipment and materials. This explanation is rooted in accelerator principle — a macroeconomic idea which plays a key role in the theory of investment.

The study detailed in this report estimates the short-run elasticity of R&D spending with respect to employment among businesses that engage with NPL, having previously established (through Belmana's econometric analysis) that 5.5% of the job-years among NPL's regularly supported businesses are attributable to support from the NMS labs (NPL, LGC, and NEL). The data for the analysis in this report was obtained from the FAME database and combined with NPL's administrative data (e.g., invoicing data) covering the period between 2009 and

6 Part of the public funding received by a PSRE each year is used to update and extend its capabilities. The remainder of this funding is used to maintain existing capabilities and infrastructure. Moreover, the expertise of a PSRE's scientists will tend to increase each year as they encounter new challenges and thereby acquire new knowledge.

2015.[7] The yearly sample size, which varies across years, averages 4,248 active businesses with some history of engaging with NPL.

The firms' R&D investment data is heavily censored (not observed for some firms) suggesting that some firms are much more likely to report R&D investment compared to others. Therefore, to avoid a possible selectivity bias, this analysis used a Heckman selection model to endogenized selection into the sample used for our panel-data analysis of firms' R&D spending. The study found that, among the NPL supported firms, the elasticity of R&D spending with respect to employment was about 0.368. That is, a 10% increase in employment generates a 3.68% increase in R&D spending. Other control variables, such as fixed assets and turnover are also positively associated with firms' R&D investments. Overall, employment growth explains about 13% of the variation in private R&D spending.

Many previous studies have focussed on estimating the extent to which public support stimulates private R&D spending. Such studies often include employment as a control variable, but it isn't the focus of analyses. A series of studies by Coad and colleagues (such as Coad & Rekha, 2007; Coad and Rao, 2010; Coad eta al, 2014; Coad and Grassano, 2016) are a notable exception and are discussed in the literature review. Compared to most other studies, our elasticity estimates are a little higher, which might be explained by the special nature of the firms in our study (e.g., unusually knowledge intensive, with a tendency to employ highly skilled staff.) It is theorised that focusing on firms that have experienced growth in employment due to NPL's support might account for this difference.

This study takes a micro econometric approach to analysing how growth in employment leads to effects on R&D spending. One issue with this approach is that the bigger, and more knowledge-intensive companies, are more likely to report R&D investment compared to the smaller ones. (NPL's regularly supported firms are clearly in the former category [Belmana, 2019].) It is acknowledged that one of the supposed weaknesses of this approach is the difficulty in generalising R&D leverage numbers beyond the domain of the firms in the dataset. Hence, the objective of this study is perhaps more modest than some macro studies, accepting at the outset that the study is focused on a self-selecting set of businesses supported by PSREs, such as the NMS laboratories.

7 The two datasets were matched one-to-one using CRNs to identify the same firm in both datasets.

1. INTRODUCTION TO R&D LEVERAGE ANALYSIS

R&D leverage is a measure to evaluate the impact of public R&D spending on private innovation activities. Two broad channels of impact are well documented in the literature: direct and indirect leverage. The direct leverage relates to public R&D spending impact on firms which received support, while indirect leverage is a spill-over impact of public R&D investment on firms which do not receive support but nonetheless increased their R&D undertakings. The latter could be through response to competitors' or clients' R&D investment behaviour (Oxford Economics, 2020). However, previous studies support two conflicting arguments. One claims that public R&D support simulates private R&D activity by reducing the cost and risk of private R&D undertakings, which is referred to as "crowding in". The other argument focuses on the "crowding out" of private investment, where public R&D competes for the same resources as needed by private firms to have profitable R&D undertakings. (For example, in the short term, there is a limited number of research scientists, and those employed in a public research institution are not available to work in a corporate R&D laboratories). This would, inadvertently, make some private R&D activity unprofitable or unnecessary. But "crowd outs" may also happen when supported firms engage in free-riding behaviours and use public grants for R&D projects that they would otherwise have funded on a purely private basis. Evidence in the literature relating to "crowding in" or "crowding out" are mixed. Therefore, the degree to which public R&D spending affects private R&D undertakings is significantly contextual and remains a testable hypothesis.

Public R&D support which simulates private R&D activity by reducing the cost and risk of private R&D is expected to make a firm more competitive and grow. This growth would require more investment in inputs which, for innovating firms, is roughly split between wages and salaries of scientists and spending on R&D materials and equipment (Coad and Rao, 2010). The focus of this study is based on this channel. However, there are other explanations for R&D leverage. One is found in the theory of firms' production where technological progress is labour-augmenting and dynamics of knowledge production is endogenously determined by the proportion of labour and capital in which a firm devotes to its production. Hence, a firm which devotes large proportion of its labour to R&D collaboration with publicly funded research institutions is expected to have extra R&D investment (Mulligan et al, 2022).

Approaches to leverage analysis are of two types: one type is found in macro analysis and the other in micro analysis. The macro approach seeks to find how private R&D investment in an economy is influenced by the aggregate level of public R&D support. (Examples include Oxford Economics, 2020; Sussex et al, 2016; Economic Insight, 2015; and Montmartin and Herrera, 2014.) In contrast, the micro approach investigates whether specific instances of R&D support encourage individual firms to undertake more privately funded R&D than would otherwise have been the case. Key analytical methods that have been widely used to achieve this are Propensity Score Matching (PSM) and Difference-in-Differences (DiD). (Examples include: Czarnitzki and Delanote, 2015; Hottenrott and Lopes-Bento, 2012; Aschhoff, 2009.)

When using the micro approach to analyse R&D leverage, PSM is one of the commonly used methodologies. PSM matches firms with similar observable characteristics (as a way of reducing bias in the estimated treatment effects) and groups firms into the 'treated' (those that received support) and the 'controls' (those without the support). One of the limitations of this approach, however, is that some important characteristics (e.g., managerial ability) are unobservable due to inability of existing data to capture them. But, if these characteristics are important, then DiD can be used to remove the influence of unobservable fixed effects, allowing one to compare the average change over time in the outcome variable for the treated group and the (untreated) matched control group. The key assumption is that the outcome between the treated and control groups would follow the same growth pattern in the absence of the support - though this can be difficult to verify.

If the outcome variable is heavily censored (not observed for some firms), then it is much more difficult to use PSM or DiD; as was seen in the case of the R&D data used as the outcome variable in our analysis. This is due to the characteristics of NPL's regularly supported firms (Belmana, 2019), which tends to be large, knowledge-intensive firms, who in turn are more likely to report R&D investment compared to smaller firms. Hence, it is difficult for the PSM and DiD techniques to adequately explain R&D outcomes of the sampled firms, due to the weighting towards larger and knowledge-intensive firms, leading to selectivity bias which would likely confound attempts to estimate the effect of public support on R&D spending.

The R&D leverage rates analyses in previous studies relate to two different time horizons: the short-run and the long-run impact. A businesses' R&D activities may respond immediately to changes in public R&D support, referred to as the short-run effect, but it may take some time for them to settle into a new equilibrium corresponding to the long-run effect. That is, the short-run analysis focuses on the relationship between public R&D support in one period and private R&D activity within the same period, while the long-run involves the relationship between public R&D and private R&D spending in many subsequent periods. Various degrees of time lag differentiate the short-run from the long-run leverage rates analyses.[8] Oxford Economics (2020) shows that the short-run R&D leverage begins within the same year that the public investment occurs, while the long-run effect takes 15 years to fully materialise and can be five times the size of the short-run effect.

This study is motivated by the observed employment growth of the firms which engaged with NMS laboratories, and the objective is to estimate the degree to which private R&D spending leverages on employment growth. This is related to the study of Coad and Rao (2010). Our study focuses on the short-term impact because the period covered (2009-15) is less suitable for a long-run analysis. Nonetheless, it is expected that the ratio of short-run effects to long-run effects would broadly follow the relationship found by Oxford Economics in their macro study.

2. LITERATURE REVIEW

There is a potential bidirectional causation between firms' R&D investment and employment. The causation could run from firms' R&D investment to firm's employment. This direction has received considerable attention in the empirical literature. For instance, Ciarli et al (2018) Goel and Nelson (2022) and Sokhanvar and Çiftçioğlu (2022) have established this.

Similar explanations which significantly remained a less trodden path is when the causation runs in the opposite direction. That is, a situation in which employment growth explains firms' R&D investment. There are two explanations for this. Firstly, knowledge production which ensures production efficiency requires a great deal of R&D investment. The production of knowledge is endogenously determined among other things by the proportion of labour and capital a firm devotes to its production. Hence, increase in the share of labour and capital devoted to knowledge production would imply increase in R&D spending and vice versa. This explanation is relevant to collaborative R&D. For instance, Mulligan et al (2022) found that R&D collaboration with publicly funded research centres drives extra investment in firms' overall R&D and applied research.

8 Consider a regression in which private R&D is the dependent variable and the only regressors are public R&D and the first lag of private R&D (that is, last year's private R&D). The short-run marginal effect of increasing public R&D is just the coefficient of public R&D (say θ) in this regression, while the long-run effect is θ divided by $(1 - \beta)$, where β is the coefficient of a one-year lag of private R&D spending in the dynamic model of R&D spending. (Typically, one finds that: $0 < \beta < 1$.)

The second explanation is underpinned in accelerator principle enshrined in firms' investment theory. Accelerator effect is a standard macroeconomic concept which defines a positive effect of market economic growth on private investment; and has played a key role in the theory of investment (Knox, 1970). It is expected that an exogenous increase in output would get firms close to operating at full capacity, which would require more resources (including labour) to be employed. This naturally leads to an increase in investment to help the firms meet the rising aggregate demand for their output. For an innovating firm, R&D investment is an important component of these investments. In addition, it is possible that not all the firms experience expanded capacity within the accelerator effect framework. This suggests that firms which have experienced improved capacity due to increased investment would be able to afford a premium wage.

When there is near full employment, labour will be attracted to firms which pay higher wages. Since labour is paid the value of its marginal productivity, it implies that firms which have experienced improved capacity, due to increased investment, must have also experienced productivity growth. The wage premium is expected to further lead to increase in R&D investment since wages and salaries of additional scientists (including other professional staff) and spending on research materials and equipment are key components of firms' R&D expenditure.

Many of the studies that focussed on estimating the extent to which public support stimulates private R&D spending include employment as a control variable, but the role of employment isn't the focus of the analysis. Studies, such as Oxford Economics (2020), a macro study, found an insignificant 0.4 elasticity of private R&D expenditure with respect to employment. A micro study by HMRC assessing the effectiveness of R&D tax credits found elasticities of R&D spending with respect to employment of about 0.3 (Fowkes, Sousa, and Duncan; 2015). In other studies, such as, Yang et al. (2012), elasticities of 0.1 and 0.4 were found for high-tech and non-high-tech firms in Taiwan, respectively. Whilst Czarnitzki and Licht (2006) found elasticities of 0.9 and 0.1 for firms in Eastern Germany and Western Germany, respectively.

The studies mentioned above focus on the direct connection between public and private R&D spending. And, in these studies, employment features merely as a control variable. This contrasts with a series of studies by Coad and colleagues (e.g., Coad and Rao, 2010), that are notable for focussing primarily on the relationship between employment growth and R&D spending.

The approach taken in our study follows the path taken by Coad and colleagues. That is, estimating how increased private R&D spending follows from growth in the workforce of firms that have accessed public support. As this path is probably less well trodden than that taken in recent leverage studies, this section provides further arguments and references to show that our approach can be anchored to academic research, as well as to government evidence reports.

2.1 Firm Growth and R&D Expenditure

Dynamics of knowledge is very essential to the degree to a firm would be competitive. This knowledge could be created through collaborative effort (Mulligan et al, 2022), learning-by-doing and experimenting. Meanwhile, activities which generate knowledge depends on firms' R&D investment.

Scientific and technological knowledge could be divided into three categories: basic science (blue skies research), applied research (generic technologies), and experimental development (proprietary technologies) (King and Renedo, 2020). However, Tassey (2004) argued that the efficiency with which generic knowledge is converted into proprietary knowledge depends on access to research tools, techniques, and standards. Hence, additional category of knowledge referred to as ‘infra-technologies (which is a public goods) determine the productivity of private R&D. The fourth category of technological knowledge is one of the core mandates of NMS laboratories and has been found to improve the productivity of firms which engaged with NMS (Belmana, 2019).

Coad and Rao (2010) investigated the co-evolution of sales growth, employment growth, and the growth of R&D expenditure. Their study was based on a panel of US manufacturing firms in the Compustat database; and the empirical analysis was based on applying the panel version of a vector autoregression (VAR) model to a firm-level longitudinal dataset. The VAR framework makes it possible to disentangle the lead-lag relationships of multiple variables in a dynamic model. (In an autoregressive model for R&D, changes in the level of R&D in the previous period contribute to changes in R&D this period. The vector element comes from modelling the simultaneity of two or more variables using a system of simultaneous equations - changes in x cause changes in y, but changes in y can also cause changes in x.)

The main finding of Coad and Rao (2010) was that firms increase their R&D expenditure following growth in sales and growth of employment. However, the magnitude of this elasticity is less than proportional in the short run, suggesting that firms want to keep the level of R&D activity at a sustainable level due to adjustment costs. The coefficient estimates imply that if the growth rate of employment increases by 1 percentage point, then, all other things being equal, we can expect R&D expenditure to rise by around 0.22 percentage points in the following year.

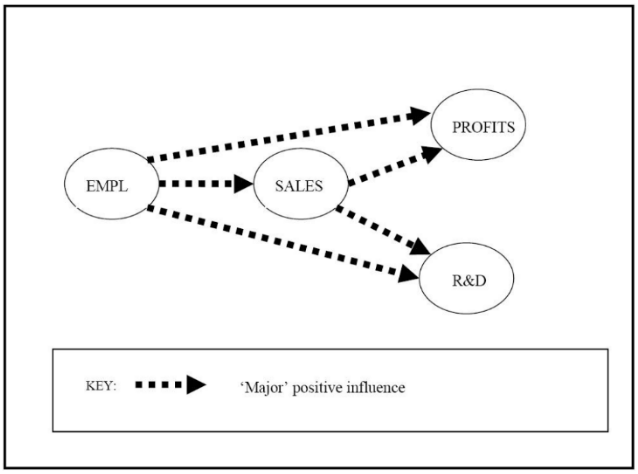
In the short-term employment growth has the largest effect on R&D spending, but sales growth may be more important in the long term. This may be because hiring additional staff tends to bring in people with new ideas. Additional staff may also provide the necessary bandwidth for some employees to be freed from having to focus on the ‘day-job’.

Some other important findings of Coad and Rao (2010) are as follows:

- In the long-term, firms may aim for a roughly constant ratio of R&D spending to employment (or to sales).
- The size of effect differs depending on whether firms are growing or shrinking, suggesting that firms are less willing to reduce R&D following a negative shock to growth than they are to increase it after a positive shock. (The idea is that R&D is a ‘sticky’ investment that cannot just be turned off or paused.)
- Surprisingly, this study found that profit growth had no detectable effect on firms’ R&D spending, suggesting that increased profits are extracted as dividends rather than being used to fund R&D projects. This goes against the Schumpeterian hypothesis that a firm’s internal finance has an important influence on the scale of its innovation activity.

The analysis in Coad and Rao (2010) suggests that past growth of employment and sales are followed by growth of R&D expenditure. Hence, the figure below gives a stylised depiction of the time profile of the processes behind a firm’s growth and its subsequent effect on R&D investment. (‘EMPL’ = employment growth, ‘SALES’ = sales growth, ‘R&D’ = growth of R&D expenditure, ‘PROFITS’ = growth of operating income.)

The main policy implication of Coad and Rao (2010) is that a government initiative to stimulate private R&D should be focussed on trying to remove obstacles to firm growth (because growing firms invest in R&D) rather than, say, reducing corporation tax in the hope that more profitable firms will reinvest the extra profits into R&D. For example, it might be possible to get firms to invest in R&D by encouraging them to grow in overseas markets, should domestic markets offer limited scope for growth.



2.2 Innovative Firms and Growth

There has been interest in this line of research among government economists and policy makers. For example, many of the insights contained in Coad and Rao (2010) can be found in Chapter 3 of an evidence report for the Department of Business, Innovation and Skills (BIS) that was published in 2014. The key findings of Chapter 3 were as follows:

- Past growth does not materially affect current growth. In short, growth does not carry over from one period to another. A firm that is growing this year is no more (or less) likely than any other firm to be growing in the following year.
- Firms begin the innovation process by employing more staff and particularly science graduates with degrees in STEM subjects. This then supports (and rationalises) higher R&D spending, as the returns to R&D spending are higher in firms with the talent to manage it. This then feeds through to the introduction of new-to-market products, which go on to account for an increasingly large share of the firm's total sales.
- The analysis did not find evidence of any feedback loops that perpetuate the process. After going through this dynamic process of growth and innovation once, firms do not often tend to renew this cycle.

2.3 Application to R&D Leverage Analysis

Small programmes that support innovation among businesses often have reasonably robust econometric evidence for their effect on employment growth among the supported businesses. Such studies are made possible by accurate and comprehensive employment data (e.g., from the Business Structure Database) and relatively low year-on-year variation in employment among the supported companies. Hence, a natural application of work done by Coad and colleagues is that it gives us a way of estimating the effect of small innovation programmes on the R&D spending among the supported businesses.

Public support for innovation not only leads the supported businesses to grow their workforce (as found by Belmana's analysis), but it may also have stimulated R&D activity among these businesses through its effect on employment growth. If the R&D intensity of the supported businesses can be found from official data (e.g., from the BERD survey), then the effect of a programme on firms' R&D spending could be found simply by multiplying the elasticity of R&D with respect to employment by the programme's effect on employment growth.

It is clear from the discussion above that our proposed approach to R&D leverage analysis is based on combining the following two claims:

- Direct public support for private sector innovation (e.g., firms collaborating with a PSRE) leads to an increase in economic opportunity, which then results in employment growth among the supported firms, a few years later.[9]
- Employment growth initiates an increase in R&D spending, leading to the introduction of new-to-market products, which finally results in revenue growth.

The first claim is justified by reference to numerous econometric studies showing that public support for innovation leads to employment growth among the supported firms. In the case of NPL, two econometric studies – one by Frontier Economic and another by Belmana Consulting – have found robust evidence of employment effects.

The second claim leads us to estimate the elasticity of R&D spending with respect to employment, which requires some discussion because a reasonable objection is that maybe the causation is running the other way. That is, one might hypothesise that an increase in R&D spending leads to an increase in employment as the firm expands to sell its new products. And, if this were true, our regression analysis would still find a positive correlation between R&D and employment, but the interpretation of what was happening would be very different.

Although this potential objection deserves careful consideration, a mixture of theory and empirics suggest that our preferred interpretation is the correct one. Firstly, there are two basic theoretical reasons why this objection can be discounted:

- It is unlikely that the benefits of R&D could so rapidly be translated into growth, given the time it takes to develop and introduce a new-to-market product. That is, economic studies have consistently found that it takes a two or three years before even successful innovation projects start to generate benefits that show-up in a business's performance.
- Creating value through innovation does not entail having the ability to capture much of the value that's being created (hence, the need for public support). That is, much of the value created by innovation does not go the innovators but rather to the imitators who manage to supply generic versions of once novel products at a more competitive price (e.g., those firms playing a 'fast second' strategy and who do not need to recoup the large R&D costs of the original innovators). Hence, much of the benefits of innovation show-up at the aggregate level in form of spill overs that benefit competitors, as well as benefiting the wider society and consumers.

Secondly, these theoretical arguments are well supported by a series of empirical studies, mainly by Coad and colleagues. For example, Coad et al (2014) explored whether there was any connection between Highly Innovative Firms (HIF) and High Growth Firms (HGF) in an analysis conducted for BIS in 2014. Coad et al found that there is minimal overlap between high growth firms and highly innovative firms. Moreover, at the individual firm-level, episodes of innovation do not appear to be followed by episodes of employment growth, with the innovation process terminating at revenue growth. Coad et al suggest that the frequent inability of innovators to internalise much of the economic value they've created explains way the there is so little overlap between highly innovative firms and high growth firms. (Of course, the inability to innovators to fully internalise the value created does not negate the importance of the spill-over benefit to the wider society). Therefore, it is reasonable to suppose that although R&D spending might lead to revenue growth in the future, it is a much weaker effect than the contemporaneous effect of employment growth on R&D spending.

Lastly, there is one further potential objection to consider: If both aggregate employment and the quantity of human capital (skilled workers) are fixed in the short term, then there cannot really be employment growth at the level of the whole economy. If so, it might seem as if well

9 Engagement with PSREs provides access to specialist expertise, toping up the existing human capital employed by an already innovative firm. The expertise provided by the PSRE increases the effectiveness of the firm's existing innovation activities, increasing the likelihood of successful innovation that increase labour productivity, and, in turn, encourages the firm to increases employment.

intended subsidies are just displacing economic activity from unsupported firms to the supported firms. Moreover, at first sight, it may appear that shifting workers from one firm to another, though the use of targeted public subsidies, just changes where the R&D occurs: The concern is that a decrease in the R&D being performed by the shrinking firms my offset the increase in R&D taking place at the growing firms. There may be a little truth in this objection, however, it is important to recognise that some industries are naturally much more R&D intensive than others (e.g., biotech versus real-estate). Indeed, previous government reports (e.g., Tera Allas Review in 2014) found that the structure of the UK economy goes a long way to explaining it is relatively low R&D intensity. (That is, the proportion of the UK economy in R&D intensive industries is lower than that of leading OECD countries.) Therefore, to shift its R&D intensity the UK would need to grow sectors that are inherently R&D intensive, which implies that one way to achieve the 2.4% target is to use public support to encourage resources (labour and capital) to shift towards the more R&D intensive sectors.

3. THEORETICAL FRAMEWORK

A public subsidy targeted at innovative firms (support from NMS labs) helps to encourage a shift in resources (e.g., labour and capital) towards more productive firms, with inherently higher levels of capital intensity than typical firms in the economy. If resources move into the supported firms, causing them to expand, then they will have to increase their level of investment to maintain their high level of capital intensity. Hence, these firms will increase investment (including in R&D) as their employment grows.

This section shows how the Solow model for the equilibrium capital intensity can be applied at the level of the firm. The Solow model is enlisted to help us explain why an increased rate of employment growth among the supported firms will cause them to increase the level of investment they make in both tangible assets (e.g., capital equipment) and intangible assets (e.g., knowledge assets).

Non-Rival but Excludible Knowledge

Knowledge increases the effectiveness with which firms can bring together the factors of production (capital and labour) to produce goods or services. But unlike the conventional factors of production, a piece of knowledge is non-rival in the sense that its use by one person does not preclude its use by another person, and so there is the potential for positive externalities or spill-overs. However, this stock of knowledge is really composed of two parts: The first part is non-excludible in the sense that it is available to everyone for free (e.g., codified knowledge in a technical standard). The second part is excludible in the sense that the person who possess that knowledge can decide whether and how to share it with others (e.g., tacit knowledge or know-how that is built-up though long experience).

Public Sector Research Establishments (PSREs), such as, NPL and the other NMS labs, are an important source of both excludible and non-excludible knowledge: On the one hand, NPL's scientists help to codify knowledge in the form of technical standards for products and services; and this knowledge becomes freely available to those following the standard. On the other hand, NPL's scientists possess considerable know-how (tacit knowledge) that cannot easily be written down and transmitted to someone else in the form of a document. This know-how is built-up through experience and forms the basis of the proprietary services that NPL sells to businesses, as well as organisations in the public sector, such as, hospitals. The organisations paying for such services gain access to the excludible portion of NPL's knowledge.

The revenue from the sale of services contributes to some extent towards the financial sustainability of a PSRE, such as, NPL. However, because much of the knowledge that NPL's scientists generate is made freely available (e.g., in technical standards), some sustained public funding is required to keep NPL in business, and this funding is used int two ways: (1)

to update and extend NPL's capabilities by conducting R&D; and (2) to maintain existing capabilities and infrastructures (e.g., performing key comparisons).

3.1 The Growing Importance of Intangible Assets

The improvements in technology, that are generated through innovation, are fundamental to improvements in productivity and a source of economic growth. Long term productivity improvements are strongly correlated with innovation and R&D activities. Moreover, investments in intangible capital, which is a good proxy for innovation related expenditure, accounted for 33% of the UK's labour productivity growth during the period 2000-13.[10]

Though plant and machinery (being the most obvious type of tangible capital) are still important, investment in intangibles assets (such as software, design, R&D and organisational capabilities) has increased in importance with total intangible investment surpassing tangible investment in 2018.[11] Advances in digital technologies hold considerable potential to lift the trajectory of the UK's productivity improvements and economic growth, and to create new and better jobs to replace older, less productive ones. Therefore, the evidence suggests that a growing component of investment is accounted for by spending on knowledge assets, including R&D.

Effect of Public Support on Employment Growth

This study analysed the employment elasticity of R&D spending among the NPL supported businesses, having previously established (through Belmana's econometric analysis) that 5.5% of the job-years (created between 2010 and 2015) among these regularly supported firms were attributable to the support they receive from the NMS labs (NPL, LGC, and NEL). This study, using dose response model adjusting for the generalized propensity score, also established this finding. The results are presented in annex G.

It is useful to provide some explanation of why receiving support from NPL, or other Public Sector Research Establishments (PSREs), leads to an increase in economic opportunity and employment growth among the supported firms. A simple explanation that fits well with the conventional production function framework is that support from PSREs provides access to specialist expertise that most firms cannot afford to maintain in-house; even more so, when this expertise isn't needed all the time, despite being intermittently critical to projects and problem solving. (To a large extent, the degree of specialisation among scientists within NPL is what really sets it apart from more general research-based organisations.)

By using the services and support provided by PSREs, firms receive an increment in the quality and scope of the available knowledge, 'A'. That is, by paying for services or engaging in collaborative R&D, the firms gain an increment in their knowledge, ΔA: If 'AL' denotes the number of effective workers, with A being knowledge and L being labour, then paying for access to the specialist expertise being housed and developed at NPL, means that: $AL \rightarrow (A + \Delta A).L$. This increases in effectiveness leads to an increase in labour productivity (output per employee); which lowers the unit cost and subsequently drives down the price in the product market. Providing the product market is reasonably competitive, this increase in labour productivity encourages the firms to expand to the point where their marginal revenue product of labour (MRPL) once again equals the wage rate in the labour market.[12] Hence, firms using

10 Corrado, C., Haskel, J., Jona-Lasinio, C. Iommi, M. 2016. "Intangible investment in the EU and US before and since the Great Recession and its contribution to productivity growth". Access here.

11 Office for National Statistics (ONS) 2018. "Investments in intangible assets in the UK: 2018". Access here.

12 It is possible a monopolistic firm would gain from reducing its employment if the labour productivity of its workers were to rise.

NPL's expertise expand their workforce as NPL's own knowledge increases due to the publicly funded R&D being performed by its specialist scientists.

Tacit knowledge[13] is particularly excludable (which is why it is basically rented by employing expert scientists), and often only large companies with supernormal profits can pay for it on a long-term basis.[14] Hence, public subsidies channelled through schemes, such as, Analysis for Innovators, are required to increase the accessibility of the knowledge housed at NPL and other PSREs.

3.2 Modelling the Effect of Growth on Investment

The model assumes that a firm's output is characterised by a Cobb–Douglas production function: $Y = (AL)^{\alpha}K^{\beta}$, with the firm operating efficiently so that there are constant returns to scale: $\alpha + \beta = 1$. Where: Y is the firm's output; A is labour augmenting technology (excludable and non-excludable knowledge); L is the size of the firm's workforce, and AL is the firm's effective labour (or skilled labour); and K is the firm's capital stock. The growth in firm's workforce is determined by factors outside the model, but the capital stock evolves according to $\dot{K} = s.Y - \delta.K$, where 's' is the rate at which income from sales is retained and reinvested in the firm. Let $I = s.Y$ denote the firm's level of investment.

The focus of the model is on showing how employment growth leads to greater investment. The full derivation of the model is presented in Annex A, but the basis for the argument and the intuition behind the model are as follows:

1. From the Solow model[15], there exists a steady state capital intensity: $\frac{K}{L} = \tilde{k} = constant$. In other words, there exists a long-run equilibrium in which the rate of change in the firm's capital intensity is zero: $\dot{k} = 0$.
2. For a successful firms, in a growing economy, there is likely to be steady growth in the firm's workforce: $\frac{\dot{L}}{L} = g_L$ or equivalently $L = L_0.\exp(g_L t)$, where t denotes time.[16] Note that this growth is determined by factors outside the model, such as, changes in the size of the available labour force and the stock of knowledge.
3. The capital stock evolves according to the equation $\dot{K} = I - \delta.K$, where δ is the depreciation rate; and from this expression we derived two important results:
   (a) The capital stock and the firm's workforce must grow at the same rate for the capital intensity to remain constant; otherwise, the capital intensity decreases as the workforce grows. That is, in the long-run equilibrium, we must have: $\frac{\dot{K}}{K} = \frac{\dot{L}}{L} = g_L$. (Meaning that the growth rate of capital is also g_L.) Moreover, we can find a convenient expression for investment, I, by rearranging our equation for the evolution of the capital stock and then making a substitution for $\dot{K}$ using $\dot{K} =$

13 A piece of tacit knowledge sits somewhere between a skill (e.g., human capital) possessed by someone who can only work on one project at a time and piece of codified knowledge, that isn't embodied in anyone individual, and so that can be used by multiple people at the same time.

14 Please see the beginning of Romer's classic endogenous growth paper for a discussion of the fundamental difficulty of incorporating partially excludable (non-rival) knowledge into any competitive economic model with production that's characterised by constant returns to scale and competitive factor markets. Romer's solution was to assume a kind of monopolistic competition among manufactures of producer durables (capital equipment); which ensures a source of profits to pay for access to the excludible knowledge.

15 The rate of change in capital intensity initially increases as k itself increases, but it then reaches a maximum after which it begins to decrease and would eventually become negative if the capital intensity were to exceed a certain threshold, $\tilde{k}$.

16 It could be that without public support that g = 0. However, this would be a strong assumption and so we allow for the possibility that g > 0.

$g_L K$. By this means, we arrive at the following expression: $I = \dot{K} + \delta . K = g_L . K + \delta . K$.

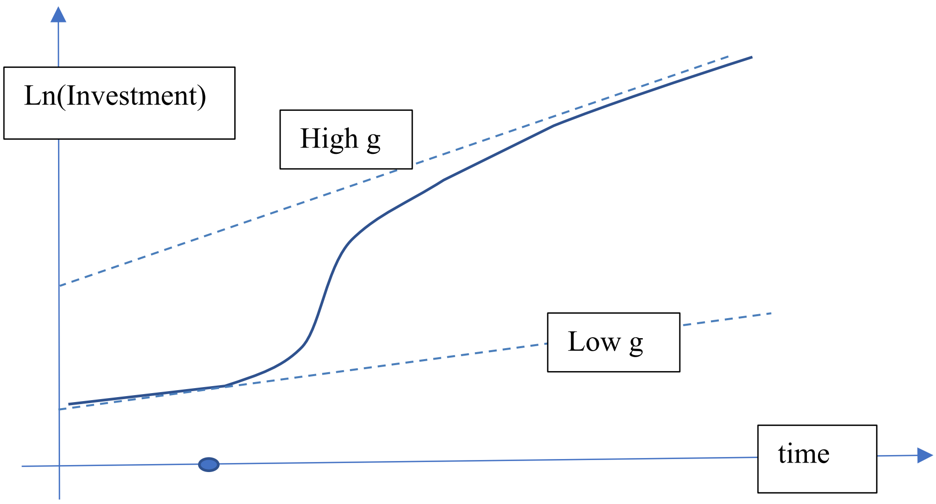
(b) If we have $I = (g_L + \delta). K$, and capital intensity, $\tilde{k}$, is constant, then by using the relation $K = \tilde{k}L$ our equation for I can be re expressed as: $I = (g_L + \delta). \tilde{k}L$, where $L = L_0 . \exp(g_L t)$ is an increasing function of both the growth rate, g_L, and of time, t.

Taking logs of our expression for L linearises its relationship with time. Therefore, the rate of growth in the workforce, g, positively influences both the slope and the intercept of the trend line:

$$ln(I) = ln(g_L + \delta) + \ln{(\tilde{k})} + \ln{(L_0)} + g_L t$$

The figure below shows that, in equilibrium, investment is an increasing function of g, where g is growth in the workforce. The effect of an increase in g on investment can be illustrated by a sketch graph with log investment, $ln(I)$, on the vertical axis and time, t, on the horizontal axis.

If there was a sudden shift in the value of g such that $g \rightarrow g + \Delta g$, then the evolution of system would change from following a long-term path that tracks along the bottom line ('Low g') to following a long-term path that tracks along the top line ('High g'). If innovation support from a PSRE (access to additional expertise) causes a sustained increase the growth rate of the firms' workforce, then our model implies that there is also an increase in investment. Moreover, along with other productive investments, one can expect an increase in the level of R&D spending. (This model seems broadly consistent with the empirical work of Coad and colleagues.)



The thick dot on horizontal time axis represents the instant when the firm starts to receive support from the PSRE (or the PSRE's own knowledge increases), and so the firm gains access to additional knowledge after this point in time. The model sketched above suggests that the firm then undergoes a sudden change in the growth rate of its workforce; so that it subsequently transitions to the 'High g' path. (In time periods to the left of the dot the firm is unsupported, whereas to the right of the dot the firm is supported. This implies that in time periods to the left of the dot the growth rate is just g, but the right of the dot the growth rate is $g + \Delta g$.)

4. DATASET, SPECIFICATIONS, AND ESTIMATION TECHNIQUES

To a large extent, the scope of the analysis was guided by the coverage of sample of supported firms used in Belmana's econometric analysis: The aim being to demonstrate that if NPL's supported firms increase their employment, then the NMS programme (which is NPL's principal source of funding) indirectly increases R&D investment among these supported firms.

This section presents the dataset, the variables it contains, and the various transformations of these variables to create the regressors. In addition, this section also provides a discussion of the econometric model and the estimation techniques used in our analysis.

4.1 Dataset

All the firms in the dataset had some engagement with NPL over the period 2009-15, but it is important to explain what is meant here by 'engaging' with NPL. Among the firms in the dataset, use of NPL ranges from attending free events and using online training courses, at one end of the spectrum, to paying for measurement services and working on collaborative R&D projects, at the other end of the spectrum. Clearly, the latter forms of usage are rather more substantive than the former. Hence, the following terminology is helpful when discussing the varying degrees with which NPL and the firms interact:

- '*Engagement*': Use of free services, such as, online training or attending events and seminars. Also, includes downloading materials available on NPL's website, such as, measurement good practice guides (de facto standards), validated software, or certified reference data.
- '*Support*': Paying for measurement services (e.g., calibrations or reference materials) or working on collaborative R&D projects with NPL's scientists.
- '*Treatment*': Receiving regular support during the six-year period (supported in 85% of the years under consideration).

Hence, it is reasonable to say that all the firms in the dataset made use of NPL, whilst recognising that only a minority of the firms received substantive support. It is also important to keep in mind that our study is not a classic treatments effects analysis. Rather, we need a sample of representative firms to investigate the relationship between employment growth and R&D spending among the relevant population of firms.

The data used for the analysis was obtained from the FAME database, as well as from NPL's administrative records (e.g., invoices) covering the period between 2009 and 2015. As already discussed, the firms' usage of NPL was defined broadly to include downloading good practice guides, attending events, use of paid-for services, and R&D collaborations. The two datasets were merged using one-to-one matching, with CRNs being used to uniquely identify a single firm in both datasets. To create a panel dataset, the data was reshaped from 'wide' to 'long'. The sample size was around 4,260 for active UK-based businesses engaging with NPL during the period 2009-15.

(We chose this period so that the coverage in our study is the same as that of Belmana's analysis. Our intention is that this will make it easier to justify combining the estimated elasticity from our study with the headline result for attributable job-years from Belmana's analysis.)

Key variables used the analysis include the firms' (real) turnover, fixed assets, total assets, number of employees, liquidity ratio (defined as current assets over current liabilities), and (real) R&D spending. These variables, except for the number of employees, were measured in '000 GBP at constant prices, with 2010 being used as the base year. Key variables were log transformed and then trimmed (using Tukey's rule 17) to avoid the danger of 'far out' data points influencing the estimations.

---

17 Tukey's fence's is a nonparametric outlier detection method using the interquartile range (IQR). (This is defined as $IQR = Q_3 - Q_1$, where Q_3 is the 3rd quartile and Q_1 is the 1st quartile.) Tukey's rule involves creating

Missing data on some of the key variables created challenges for our analysis, and so it is helpful to discuss what might have determined the availability or non-availability of such data:

- During the period 2009-15, smaller firms were permitted to submit abridged accounts, and they did not have to declare either employment or turnover. The consequent truncation observed in employment and turnover data were dealt with by interpolating the missing values using the Heckman technique. The specifications chosen for the employment and turnover equations follows those of Huang and Verma (2018) and Yazdanfar and Ohman (2015).
- The R&D data used in our study comes from FAME, which is owned and updated by staff at Bureau van Dijk (BVD) who assemble the data using publicly available sources (companies' accounts and annual reports). Because of the reliance on information in company accounts, the availability of R&D data for a firm in our dataset is likely to depend mainly on whether it has applied for R&D tax credits. (That is, the growing use of the RDEC scheme over this period probably had a big influence on whether firms chose to record and report their R&D spending.) The problem is that systematic patterns in the availability of data violates the random sampling assumption needed for OLS. This issue will be discussed at length in one of the following sections.

The Standard Industrial Classification of Economics Activities (SIC) codes of the sampled firms are included as control variables in our analysis. However, using these codes required fixing some of the issues with them. The detailed explanation of how we made the SIC codes usable for this analysis is presented the annex B

Lastly, two of our control variables require some discussion: Firstly, the number of years since a company's incorporation was thought to be an important control variable in our model of a firm's R&D spending. (The idea being that as a market becomes more mature there is less need or opportunity for radical product innovation.) Hence, age (a fixed variable) was created from the 'birth year' of the firm up to the end of 2008.[18] Secondly, to help us model the availability of R&D data, a dummy was generated for a firm being listed on the stock market. Any delisted firms were categorised as unlisted, and the dummy takes the value of '0' for an unlisted firm, whilst for a listed firm it takes the value of '1'.

4.2 Specification of the Model

The model used in our analysis is very similar to that featured in the first part of Griffith et al (2006). That is, the model used in their study was also static (the regressors do not include lagged R&D spending), and so looked at the contemporaneous effects of a range of variables on R&D intensity. However, because they focus on R&D per employee (pre-normalising by size), a firm's employment only features as a regressor in their selection equation, whereas, in our study, employment is a regressor in both the R&D equation and the selection equation.

In our study, the R&D spending of firm *i* in year $t \in \{0,1,2,\dots,6\}$ is given by the following regression equation:

---

a 'fence' boundary at a distance of k x IQR beyond the 1st and 3rd quartiles. Any data beyond these fences were classed as outliers. The interval for acceptable observations is $[Q_1 - k(Q_3 - Q_1), Q_3 + k(Q_3 - Q_1)]$ for some nonnegative constant k. John Tukey proposed this test, where k = 1.5 indicates an "outlier", and k = 3 indicates data that is "far out". Being nonparametric Tukey Fences are robust methods in detecting outliers. In our study we dropped a handful of 'far out' data points from the sample. That is, we used Tukey's fences with k = 3.

18 Because age is fixed, age dummies were also created for the groups of firms that are less than two years, between 2 and 5 years, 5 and 10 years, 10 and 20 years, and over 20 years.

$$ln(R_{it}) = \alpha + \boldsymbol{X}_{it}'\boldsymbol{\beta} + \varphi.ln(L_{it}) + \delta_t + f_i + \varepsilon_{it} \quad (1)$$

Where R_{it} is the firm's (real) R&D spending each year; L_{it} is the number of staff employed by the firm each year; and $\boldsymbol{X}_{it}$ is the vector of control variables. These control variables include a firm's age, its (real) turnover, its fixed assets, and its liquidity ratio.[19] Let α, $\boldsymbol{\beta}$, and φ denote parameters (coefficients) to be estimated; and note that $\varphi = (\partial R/\partial L)/(R/L)$ is the elasticity of R&D with respect to employment, and so is really the main focus of this report.

The model includes a full set of time dummies, and δ_t is the coefficient corresponding to year *t*. We chose 2009 (t = 0) as the base year, so that $\delta_0 = 0$. These year dummies are included to account for any macroeconomic influences (e.g., the pending referendum in 2016) or policy changes (e.g., the introduction of RDEC in 2013) that might simultaneously influence the R&D spending of all firms.

The final two terms in the regression equation, $u_{it} = f_i + \varepsilon_{it}$, constitutes a composite error term, u_{it}, made up of an unobservable fixed effect, f_i, representing the firm's inventive ability or ambition, and an idiosyncratic error term, ε_{it}, representing other random influences on R&D spending that are supposed to be unconnected with the regressors in the model.

The fundamental assumption is that having accounted for observable factors, as represented by the regressors, and the (unobservable) fixed effects, the remaining influence are essentially random noise. More formally, it is assumed that ε_{it} are independent (across both years and firms[20]) and normally distributed with mean zero, $E[\varepsilon_{it}] = 0$, and variance, σ^2. This can be summarised as: $\varepsilon_{it} \sim N(0, \sigma^2)$. Moreover, ε_{it} is assumed to be uncorrelated with any of the regressors in the model; and, specifically, there is assumed to be no correlation between these random influences and a firm's employment: $E[\varepsilon_{it}L_{it}] = 0$. Lastly, the unobservable nature of f_i creates potential challenges for the estimation of the model, particularly, given that the composite error term is likely to be correlated with some of the regressors[21]; and considerations of how to deal with this problem determined our choice of estimation method.

To simplify the exposition of the model, it is helpful to introduce a notation in which lower-case letters denote the natural log of their upper-case counterparts. Let $r_{it} \equiv ln(R_{it})$ and $l_{it} \equiv ln(L_{it})$, so that the model can be written as:

$$r_{it} = \alpha + \boldsymbol{X}_{it}'\boldsymbol{\beta} + \varphi.l_{it} + \delta_t + f_i + \varepsilon_{it} \quad (2)$$

Unfortunately, applying OLS to the above equation would give a biased estimate, because R&D is only observed for a portion of the firms in the population. (Due to the random sampling condition, the Ordinary Least Squares (OLS) estimator may well be biased if the data are not randomly drawn from the population.) This selection effect, resulting from the partial observability of R&D, means that applying OLS to equation (2) would yield a biased result.[22]

---

19 Liquidity ratio can affect the R&D spending of smaller firms according to Kenneth and Adam (1993).

20 The need to remove common macro influences that would otherwise enter the error terms was the main reason for wanting to include the time dummies in the model.

21 For example, managerial ability may influence the level of investment in R&D, as well as its size and future growth prospects.

22 The direction of the bias depends on the correlation between ε_{it} and ν_{it}, where ν_{it} is the error term in the selection equation. When the correlation is significantly positive, the OLS estimate of the coefficients is biased upward, and the values of the coefficients are biased downward when the correlation is negative. That is, OLS estimator converges to a value higher or lower than the true value, respectively.

4.3 Heckman Selection Model

An issue encountered in many innovation studies is that R&D data is only available for a small fraction of the population. According to Griffith et al (2006), the proportions of firms reporting that they are engaged in R&D and have process and product innovations are greater in France (0.35) and Germany (0.40) than in Spain (0.21) or the UK (0.27).To deal with the problem of missing R&D data Griffith et al (2006) used the Heckman approach, adding the inverse Mills ratio as an extra regressor to control for selection effects created by the truncation of the dependent variable.

To justify using the Heckman approach, Griffith et al (2006) argued that all firms really have some R&D, or at least innovation activity, even if a firm does not record it or report it in official statistics. That is, Griffith et al argue that even in firms without any recorded R&D spending, at least a few employees will occasionally spend time trying to improve the firm's products or processes. Hence, the missing R&D data should not be taken as tantamount to being 'zeros'.

Following Griffith et al (2006), our solution to the missing data problem is to explicitly model the observability of a firm's R&D spending, r. (As above, r_{it} denotes the log of a firm's R&D spending each year.) Suppose that r is not always observed and that its observability depends on a vector of firm characteristics, **Z**. More formally, suppose that r is only observed if $\boldsymbol{Z}'_{it}\boldsymbol{\gamma} + \nu_{it} > 0$, where $\boldsymbol{\gamma}$ is a vector of coefficients, and ν_{it} is the unobserved determinants of whether a firm decides to report its R&D. The contents of ν_{it} might include policies, strategies, or cultures that influence whether the staff responsible for preparing annual reports decide to broadcast the scale of the firm's R&D activities. The vector of variables, **Z**, explaining a firm's R&D spending, includes a dummy variable for being listed on the stock exchange. The idea being that the listed firms may want to make potential investors aware of the firm's investments in R&D. The idea behind our inclusion of this dummy in the selection model, is that, using annual reports to provide the market with evidence of doing some R&D, is probably welcomed by investors as a sign that the firm is investing for the future.

It is further assumed that $\nu_{it} \sim N(0,1)$, and $correl(\nu_{it}, \varepsilon_{it}) = \rho$, where ρ is the correlation between the unobserved determinants of whether to record R&D spending and the unobserved determinants of how much R&D to perform. (Investors may be less than keen to see the firm investing large amounts of money in R&D activities, with an uncertain return, when that money could have been paid out as dividends. If this were true, then we might expect to see a negative correlation between ε_{it} and ν_{it}.) From this, it follows that the probability of observing r is given by $\Pr(r_{it} \text{ is observed}|\boldsymbol{Z_{it}}) = \Phi(\boldsymbol{Z}'_{it}\boldsymbol{\gamma})$, where $\Phi(.)$ is a cumulative normal distribution; and this selection model can be operationalised by estimating a Probit regression.

Consider what happens when the expectation of equation (2) is taken over the whole population of firms, recalling that $E[\varepsilon] = 0$ when the expectation is take over the full population. However, the partial observability of r must lead to the following identity:

$$E\left[r \,\middle|\, r \text{ is observable}, \mathbf{X}\right] = E[r|\boldsymbol{X}] + E[\varepsilon|\, r \text{ is observable}, \mathbf{X}]$$

In the 1970s James Heckman showed that to account for extra term created by the partial observability of the outcome variable (which is R&D, in our case), the inverse Mills ratio[23], $m_{it} = \frac{\phi(\gamma' \boldsymbol{Z}_{it})}{\Phi(\gamma' \boldsymbol{Z}_{it})}$, should be added to the model as an extra regressor:

$$r_{it} = \alpha + \boldsymbol{X}'_{it}\boldsymbol{\beta} + \varphi.l_{it} + \delta_t + \lambda.m_{it} + f_i + \varepsilon_{it} \quad \text{given that } r_{it} \text{ is observable.} \quad (3)$$

23 $\Phi(.)$ is a cumulative normal distribution; and $\phi(.)$ is the PDF for the normal distribution.

Where λ is the coefficient of the inverse Mills ratio. To find the inverse Mills ratio we need to estimate $\Pr(y_{it} \text{ is observed}|\boldsymbol{Z_{it}}) = \Phi(\boldsymbol{Z_{it}}'\boldsymbol{\gamma})$. The vector of coefficients, $\boldsymbol{\gamma}$, is found by estimating the corresponding Probit model, and an estimate of the inverse Mills ratio can then be found using the fitted values from this regression.

4.4 Estimation Using the Method of Correlated Random Effects

A common issue encountered by studies focussed on R&D is that, although, we begin with around 26,000 observations (covering about four thousand firms over seven years), we only have R&D data for around 3,000 observations (covering about eight hundred firms over a maximum of seven years). To estimate the coefficients with any precision requires us to make maximum use of both the within and between variation in the useable part of the panel.

Our first choice of estimation method might have been Random Effects (RE), but this method is known to be vulnerable biases caused by unobserved heterogeneity. So, in practice, the method of Correlated Random Effects (CRE) was used in our analysis, because of the need to account for the presence of (unobservable) firm-level fixed effects, f_i. The concern was that, because some regressors are correlated with these fixed effects, the composite error term, $u_{it} = f_i + \varepsilon_{it}$, would also become correlated with these regressors. And such a correlation would violate the exogeneity assumption needed for standards regression methods (e.g., OLS) to yield consistent estimates of the coefficients.

The solution is to include firm-level time averages of the (potentially) endogenous regressors as extra regressors in the equation, whose inclusion helps to remove, or at least to soften, the correlation between the composite error term and the endogenous regressors. The time average of x_{it} for firm *i* is given by $\bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$. And the part of x_{it} that is correlated with f_i is necessarily constant, and so it must be contained in $\bar{x}_i$, making it a good proxy for the portion of f_i that's associated with x_{it}. If the coefficient of $\bar{x}_i$ is found to be significant after estimation, then this is a sign that x_{it} would indeed have been correlated with the composite error term had $\bar{x}_i$ not been included as a regressor.

If the analysis is robust, then it should be possible to get the same, or at least similar estimates, using other methods that also account for the presence of fixed effects. Hence, other estimations are used to explore the robustness of the results, including both fixed effects (FE) and first differences (FD). (Note that if the fixed effects are truly fixed, then first differencing should sweep them out of the model. And, if this happens, then the model in first differences could be estimated by OLS.) The presence of firm-level heterogeneity can be confirmed with an F-test or by looking the estimate of the rho parameter following estimations performed by fixed effects or random effects, respectively; while choosing between fixed and random effect model, was guided by a Hausman test.

5. SUMMARY STATISTICS

Table 1 reveals certain characteristics of the kind of firms that engage with NPL. One of these attributes is that, during the period 2009-15, some firms remained very innovation-active, despite the 'Great Recession' occurring at the start of the period.[24] Also, note that R&D data (for any of the years) was only available for 19.9% of the 4,120 firms in the sample, and this missing data problem necessarily complicates the analysis in ways that will be discussed in the next section.

24 For instance, despite a decline (after 2013) in the count of firms providing R&D data, the arithmetic mean of the R&D spending among those firms providing R&D data rose by 37.6%. Comparing this to the availability of employment data, within the same period, revealed an increase in the count of firms with employment data, but this does not seem to lead to an increase in the scale of the reported workforce.

Table 1. Summary Statistics for the Raw Data								
Variable		2009	2010	2011	2012	2013	2014	2015
Emp	Arithmetic mean	1633.3	1629.3	1612.3	1581.8	1563.2	1572.6	1433.5
	Median	154	147	155	152	151.5	148.5	116
	Geometric mean	187.3	182.0	188.5	184.5	178.4	168.2	125.8
	Count	1,771	1,085	1,815	1,859	1,906	1,958	2,213
R&D	Arithmetic	22777.9	12227.2	15022.8	17237.8	14028.9	13513.0	18756.9
	Median	1088.2	982.8	1008.6	1070.8	1096.5	888.0	952.4
	Geometric mean	1426.5	940.8	1094.5	1174.0	1131.1	941.3	1045.6
	Count	200	517	568	601	592	462	265

Note: emp (firm's workforce), R&D (real R&D investment). Real R&D is measured in thousand GBP.

The median and geometric mean are more representative of the R&D spending of a typical firm than the arithmetic mean, because they are less influenced by large companies with unusually big R&D investments. However, the geometric mean has additional advantages which makes it more representative of the data and preferred. These include its inclusiveness as its calculation is based on all terms in the sequence with relatively more weight given to small observations, and it is suitable in the case of non-zeros or negative observations which characterised our raw data. Focussing on the geometric means, shows that a typical firm invests around £1 million pounds yearly, and employs about 180 workers. The summary statistics of the raw data for other variables are in the annex (Table D1). Lastly, note that the R&D spending, along with and all the other financial variables, have been deflated, using 2010 as the base year. Hence, the values in the tables are real as opposed to nominal, and so already account for inflation.

The variables used in the estimations were first put into logarithmic form, and then trimmed (based on Tuckey's fences) to remove extreme values which might unduly affect the estimations. Untransformed summary statistics for the raw data are presented in the annex (Table D2). Relating to (real) R&D spending, two firms with extreme values were dropped from the sample (Table 2 and 3). The liquidity ratio suffered most from extreme values; with twelve firms being identified as outliers and subsequently dropped from the dataset. Following the trimming, the variables became a little more symmetrical in their distributions (except for the log of firms' age which was not trimmed), with skewness within the range -0.5 to 0.5.

For all the variables in the dataset there is significant heterogeneity between the firms in the sample. For instance, the decomposition of the standard deviation of R&D spending into the 'within' and 'between' components, indicates that if we were to draw two firms randomly from the data, the difference in their R&D spending is expected to be higher than the difference for the same firm in two randomly selected years. A similar pattern is noticed with the other variables. However, age does not vary with time, which is why its 'within' standard deviation is zero.

Table 2. Summary Statistics (After Trimming for Extreme Values)											
Variable	Obs	Number of firms	Mean	Median	Min	Max	Std. Dev.				
							Overall	Between	Within	Skewness	Variance
logrrd	3,203	821	6.99	6.93	-2.07	14.04	2.28	2.34	0.57	-0.01	5.20
logempx	25,506	3,951	3.80	3.61	-5.31	12.53	2.19	2.20	0.42	0.36	4.78
logemp	13,326	2,367	5.14	4.98	0.00	12.53	1.77	1.88	0.35	0.46	3.12

logrtox	25,622	3,970	8.52	8.46	-3.50	19.28	2.74	2.79	0.50	0.13	7.51
logrto	13,921	2,363	10.09	9.95	-1.37	19.28	2.28	2.48	0.43	-0.07	5.18
logage	26,194	3,742	2.73	2.83	-0.01	4.91	1.00	1.00	0.00	-0.53	1.00
loglr	25,869	4,032	0.40	0.35	-3.08	3.90	0.91	0.84	0.46	0.04	0.83
logrfa	25,526	3,953	6.31	6.33	-6.99	18.68	3.64	3.69	0.72	0.05	13.28

Note: logrrd (log of real R&D investment by firms), logempx (log of interpolated employees), logemp (log of employees), logrtox (log of interpolated real turnover), logrtox (log of real turnover), logage (log of age of firms as at 2009), loglr (log of liquidity ratio), logrfa (log of real fixed assets).

It is important to note that 'logempx' and 'logrtox' in tables 2 and 3 are interpolated values of the logs of firms' employment and turnover. Interpolation was necessary because both variables are truncated due to missing values in the dataset. The interpolated values were predicted using a Heckman model. This procedure added around 1,600 firms that would otherwise have been missing from the dataset. About 40.1% and 43.8% of log of employment and real turnover values are interpolated, respectively. It is acknowledged that interpolation of these variables may create some measurement errors. The implication of this is that the real effect might be more significant than the estimated. This is because measurement error generates type II error, a false negative result. Hence, our estimates could be regarded as lower bound.

Table 3. Summary Statistics (Before Trimming for Extreme Values)

Variable	Obs	No. of firms	Mean	Median	Min	Max	Std. Dev.				
							Overall	Between	Within	Skewness	Variance
logrrd	3,205	823	6.98	6.92	-6.21	14.04	2.30	2.41	0.57	-0.10	5.29
logempx	25,507	3,952	3.80	3.61	-6.31	12.53	2.19	2.21	0.42	0.36	4.78
logrtox	25,623	3,971	8.52	8.46	-5.51	19.28	2.74	2.80	0.50	0.12	7.51
logage	26,194	3,742	2.73	2.83	-0.01	4.91	1.00	1.00	0.00	-0.53	1.00
loglr	26,217	4,044	0.36	0.35	-12.72	4.59	1.08	0.98	0.55	-1.41	1.17
logrfa	25,526	3,953	6.31	6.32	-6.99	18.68	3.64	3.69	0.72	0.05	13.28

Note: logrrd (log of real R&D investment by firms), logempx (log of interpolated employees), logemp (log of employees), logrtox (log of interpolated real turnover), logrtox (log of real turnover), logage (log of age of firms as at 2009), loglr (log of liquidity ratio), logrfa (log of real fixed assets).

6. REGRESSION TABLES AND SPECIFICATION TESTS

This section presents a summary of the results obtained from our estimations.

6.1 Estimation of the Selection Model

To correct for selectivity bias, a Heckman model (presented in Annex D) was estimated to generate the Mills ratio used in subsequent estimations. All the regressors used in the R&D model were included in the selection model, but in addition we also include a dummy for being listed on the stock market, as well as squares of turnover and employment along with their interaction. The additional regressors are highly significant, which helps with identification of the coefficients in the Heckman model. Without the extra regressors, such identification would depend purely on functional form.

6.2 Elasticity Estimations

Table 4 presents the estimated coefficient values coming from three estimation methods: correlated random effects (CRE1, CRE2, CRE3); fixed effects (FE); and random effects (RE). The FE estimation necessitates excluding the time-constant terms that appear in the RE and the three CRE models. The correlated random effect models have many of the characteristics of a fixed effect model but makes it possible to retain all the observed (and unobserved) time-constant terms in the estimation.

The values of the corresponding coefficients generated by each estimation method (RE, CRE1, FE) are similar to one another. As both the dependent variable and the regressors are in logarithmic form the coefficients correspond to elasticities. The size of these coefficients suggests that a change in any of the regressors has a less than proportional effect on R&D spending. (That is, none of the coefficient values are greater than unity.)

Given the significance of Wald Chi-square and F-statistics, the explanatory power of the estimated models is more than adequate. The size of the rho parameter implies that about 88% of the variation in the composite error term is explained by the firm-level fixed effects.

The inverse mills ratio is consistently significant and negative across all three models, indicating that estimates of the R&D equation by OLS would otherwise have been biased downwards. The negative sign of the coefficient for the inverse Mills ratio implies a negative correlation between the idiosyncratic error term in the R&D model, ε_{it}, and the error term in the selection equation, ν_{it}. This means that negative selection has occurred. That is, without the correction the estimates would have been downward biased, a lower bound estimates of R&D.

4. Elasticity Estimates					
	Random Effect	Correlated Random Effect			Fixed Effect
		CRE1	CRE2	CRE3	
R&D	Coef.	Coef.	Coef.	Coef.	Coef.
Employment	0.450*** (0.056)	0.422*** (0.072)	0.372*** (0.059)	0.368*** (0.059)	0.434*** (0.075)
Employment_mean	-	- 0.149 (0.121)	-	-	-
Age	-0.084 (0.068)	-0.142** (0.070)	-0.147*** (0.069)	- 0.143*** (0.069)	-
Turnover	0.143*** (0.035)	0.053 (0.040)	0.062* (0.039)	0.067* (0.039)	0.054 (0.040)
Turnover_mean	-	0.362*** (0.088)	0.302*** (0.074)	0.262*** (0.056)	-
Liquidity ratio	-0.031 (0.031)	-0.057* (0.034)	-0.058* (0.034)	-0.061* (0.033)	-0.056* (0.034)
Liquidity ratio_mean	-	0.178** (0.086)	0.179** (0.086)	0.191** (0.085)	
Capital	0.114*** (0.029)	0.096*** (0.033)	0.100*** (0.032)	0.088*** (0.029)	0.100** (0.034)
Capital _mean	-	-0.020 (0.062)	-0.048 (0.059)	-	-
Year Dummies	yes	yes	yes	yes	yes
SIC sector Dummies	Yes	Yes	Yes	Yes	Omitted

Mills	-0.623*** (0.187)	-0.547*** (0.189)	-0.576*** (0.188)	- 0.572*** (0.187)	-0.469* (0.253)
_cons	2.435*** (0.877)	0.562 (0.967)	0.911 (0.925)	0.921 (0.914)	3.755*** (0.909)
sigma_u	1.582	1.582	1.582	1.582	1.822
sigma_e	0.641	0.642	0.642	0.641	0.642
Rho	0.859	0.859	0.859	0.859	0.889
Wald Chi-square/ F-statistics	Chi2 = 971.17***	Chi2 = 1005.89****	Chi2 = 1004.23***	Chi2 = 1003.59***	F-Statistic = 14.29***
F-test (all u_i=0)	-	-	-		27.34***
R-Square (overall)	0.49	0.50	0.49	0.49	0.38
Number of firms	795	795	795	795	795
Obs	3,132	3,132	3,132	3,132	3,132

Note: ***,**,* implies significant at 1%,5% and 10% level. Standard errors are in the parentheses. The explanatory variables, except employment and employment_mean) and are in real terms. Employment and real turnover are interpolated values. Explanatory variables with underscore mean are time-averages. Capital is a proxy for real fixed assets and liquidity implies liquidity ratio.

The coefficient of primary interest is that of ‘Employment’ (log of employment) and its estimated value varies depending on the estimation technique: 0.450 for RE; 0.422 for CRE1; and 0.434 for FE. These are all very similar values, and their confidence intervals all overlap. Since this coefficient is an elasticity, it tells us how a firm’s R&D spending responds to changes in the size of its workforce. For example, based on an elasticity of 0.45, a 10% increase in employment would lead to a 4.50% increase in R&D spending.

All the models control for changes in a firm’s turnover. The estimated values for the coefficient of ‘Turnover’ (log of real turnover) suggest that changes in turnover also have a positive (inelastic) effect on R&D spending. However, the effect of changes in turnover seems to be much weaker than the effect of changes in employment. Also, the variation among the various models’ estimates of the coefficient of ‘Turnover’ is larger than that for the coefficient of ‘employment’. The random effect model yielded a significantly higher coefficient value than that found by the other two estimation methods: 0.143 for RE, compared to 0.053 for CRE1, and 0.054 for FE.

A firm’s age does not feature in the FE model because it is effectively a time-constant variable. (That is, age must always go up by one each year, and we include a full set of year dummies in the models.) However, age does appear as a regressor in the RE and CRE1 models, and these estimations found strong evidence that R&D activity typically reduces with a firm’s age. This may be because older firms are operating in more mature markets and so have less scope for product innovation than firms operating in new markets. Alternatively, it may be that older firms can become more hierarchical or bureaucratic, so that maybe their inventive capacities are subject to a kind of atrophy.

The liquidity ratio appears to weakly affect R&D spending in the CRE1 and FE models, but it is insignificant in the RE model. The negative sign of the coefficient for ‘Liquidity’ suggests that firms who are holding cash tend to invest less in R&D. Lastly, there is consistent evidence of a strong, positive relationship between fixed assets and investment in R&D. This may indicate a connection between firms investing in productive capital (e.g., plant and machinery) and investments in process innovation.

We controlled for calendar years in our estimations, but the effect of time (macroeconomic influences) appears to have improved R&D spending significantly between 2010 (the base year) and 2015. A joint test of all the time dummies for CRE1 yielded a Chi-squared statistic that was significant at the 5% level (chi2(6) =12.44, p-value=0.05).

Following the FE estimation, an F-test for the joint significance of all firm-level fixed effects found strong evidence of unobserved heterogeneity (F = 27.34, p-value = 0.000). Directly comparable versions of FE and RE were estimated by removing all time-constant terms from the model. If the fixed effects are uncorrelated with the regressors, then both RE and FE are consistent, and so should give similar estimates of the coefficients. A Hausman test examines (random) deviations between the estimated coefficients from both methods under the null hypothesis (H0) that both methods give consistent results. (The alternative hypothesis is that the fixed effects are correlated with the regressors, so that only FE gives consistent coefficient estimates.) However, in our case, the scale of the differences between the coefficients are so large, relative to the estimated standard errors, that the Hausman test rejects the null hypothesis, and so picks the fixed effect model as the preferred model (Table D3). This suggests that the firm-level fixed effects are correlated with at least some of the regressors.

To some extent, the Hausman test (Table D3) can be used to diagnose the source of the problem by examining the difference between the FE and RE estimates of a given coefficient and comparing this difference with its standard deviation. Examining Table D3 shows that FE and RE give similar estimates for the coefficients of ‘Employment’ and ‘Capital’; whereas there is a significant difference between the FE and RE estimate of the coefficients for ‘Turnover’ and ‘Liquidity’.

The correlated random effect (CRE) models effectively demean a regressor by including, as an extra variable, the average value of the regressor for each firm. (The average is taken over the yearly values of the regressor for a given firm. That is, the time-average of x_{it} for firm *i* is given by $\bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$.) If the variable representing the time-average of a regressor, $\bar{x}_i$, has a statistically significant coefficient, then this indicates that the regressor itself, x_{it}, is correlated with the firm-level fixed effects. Hence, it seems that neither ‘Employment’ nor ‘Capital’ are correlated with the fixed effects, and this intuition was followed up with more formal testing.

The important sectors contributing significantly to R&D investments in CRE3 are R&D firms; information and communication; legal, accounting and market research activities firms; random firms (whose activities cut across different SIC sectors); technical services firms; and high-tech manufacturing firms. The R&D elasticities for these firms range between 2.7% and 1.3%. However, water supply, sewage, waste management, and remediation activities R&D elasticity is -2.2%.

A Mundlak test involves performing tests for the joint significance for groups of these time-averages. Joint significance test involving all four of the time-averages (‘Employment_mean’, ‘Turnover_mean’, ‘Liquidity_mean’, ‘Capital_mean’) yielded a very significant chi2 statistic (chi2(6) = 27.40, p-value = 0.0000), which confirms the result of the Hausman test. However, a joint significance test involving ‘Employment_mean’ and ‘Capital_mean’ yielded an insignificant chi2 statistic (chi2(2) = 2.22, p-value = 0.329). In contrast, a joint significance test involving ‘Turnover_mean’ and ‘Liquidity_mean’ yielded a very significant chi2 statistic (chi2(2) = 21.84, p-value = 0.0000). Also, the joint test for year dummies (chi2(6) =12.44, p-value=0.05) and SIC sector codes (chi2(17) = 156.69, p-value=0.0000) are very significant.

Taken together, the tests above provide strong evidence that ‘Employment’ (log of employment) and ‘Capital’ (log of fixed assets) are not correlated with the fixed effects; but that

'Turnover' (log of turnover) and 'Liquidity' (log of the liquidity ratio) are strongly correlated with the fixed effects. Therefore, this Mundlak approach leads us to identify CRE3 as the preferred estimation, because it provides the most precise estimate of the coefficients (smallest standard errors) whilst avoiding the bias that comes from regressors being correlated with the fixed effects. Hence, our preferred estimate for the elasticity of R&D spending with respect to employment is employment is 0.368 ± 0.059 (central estimate ± standard error). Based on the value of 0.368 from CRE3, a 10% increase in employment would lead to a 3.68% increase in R&D spending.

Separate versions of CRE3 for large firms (non-SMEs) and SME firms show that our estimates are biased toward large firms (Table D4). The result for large firms shows that a 10% increase in employment would lead to a 3.10% increase in R&D spending while interpreting the estimate of SME firms could be misleading because the Wald chi2 could not be computed, but not necessarily because there is something wrong with our model.

6.3 First Difference Estimations

We assessed how changes in the explanatory variables may affect changes in R&D investment of the sampled firms (Table 5). The models are adequate, but the precision of the estimates was limited by the small number of observations in the effective dataset – only 649 firms have R&D data in consecutive years.

To estimate the model in First-Differences, pooled OLS was preferred, as the FE and RE estimates show no significant heterogeneity across firms. (Presumably, the first differencing swept out the firm-level fixed effects, a result that broadly confirms our theoretical model.) Specifically, the rho parameter from the random effect estimate was zero; whilst in the case of the FE estimate, an F-test for the joint significance of the fixed effects was not significant.

The inverse Mills ratio is not at all significant in any of the regressions, which is not surprising given that the likelihood ratio test of independent equations was insignificant in the Heckman model (annex D). This implies that using OLS to estimate the model First-Differences, yields unbiased estimates, and suggests that a 10% upward change in firms' workforce shifts private R&D spending upward by about 4.68%.

Again, we did not find much evidence of a significant background change in the level of R&D spending among the sampled firms over the period 2009-15, given the insignificance of the year dummies.

Table 5. First Difference Estimates

	Pooled OLS			Random Effect			Fixed Effect	
$\Delta R\&D$	Coef.	Std. Err.		Coef.	Std. Err.		Coef.	Std. Err.
$\Delta Employment$	0.481***	0.088		0.481***	0.090		0.400***	0.114
$\Delta Turnover$	0.081*	0.045		0.088*	0.045		0.107**	0.054
$\Delta Capital$	0.081**	0.039		0.081**	0.039		0.058	0.049
$\Delta Liquidity$	-0.014	0.039		-0.014	0.039		-0.031	0.047
mills	-0.048	0.039		-0.048	0.039		-0.225	0.337

Year dummies	Yes	-		Yes			Yes	
SIC sector dummies	Yes	-		Yes			Yes	
_cons	0.080	0.135		-0.015	0.139		0.262	0.596
sigma_u	-	-		0.000			0.495	
sigma_e	-	-		0.798			0.798	
rho	-	-		0.000			0.288	
Wald Chi-square/ F-statistics	2.66***	-		71.94***	-		3.25***	
F-test (all u_i=0)	-	-		-	-		0.69	
R-Square (overall)	0.03			0.03			0.02	
Number of firms	656			656			656	
Obs	2,199			2,199			2,199	

Note: ***,**,* implies significant at 1%,5% and 10% level. Standard errors are in the parentheses.

6.4 Long run Estimations

We attempted to estimate the long-run employment elasticity of R&D investment with different estimators including Polled Least Squares, fixed effects, random effects and difference and system General Method of Moment (GMM) dynamic panel estimator. GMM might have been preferable given its ability to create exogenous instruments for potentially endogenous explanatory variables. However, GMM estimators yielded unreliable knife-edge[25] results. That is, a small change in estimation method ("two-step" versus full maximum likelihood estimates) resulted in marked changes in coefficient estimates.

There may be two reasons for the instability of the GMM estimates. Firstly, the small size of our sample nullifies the large sample property of GMM. The lack of adequate instruments, other than the use of lags as instrumental variables, which are unlikely to satisfy the exclusion restriction, may be another reason. This instability is well illustrated by the results presented in Table E1. It is important to note that the long-run estimate is not central to our analysis which is focused on the short run estimates given in Table 4 and 5.

7. INFERENCES

The following conclusions can be reached based on this evidence:

- Firstly, in static form, a firm's R&D spending responds positively to changes in the size of its workforce. Based on an elasticity of 0.368, a 10% increase in employment would lead to a 3.68% increase in R&D spending.
- Secondly, with regards to turnover, larger firms, with more revenue, tend to engage in more R&D than the smaller firms with limited revenue.

25 The GMM estimates in Table E1 are quite different from the estimations coming from the estimated OLS, RE and FE in Table E1. Moreover, we would expect the GMM estimate to lie between the OLS and FE estimates (Bond, 2002). Hence, we considered the GMM estimates unstable.

- Finally, older firms seem to show some decline in R&D investment. It may be that difficulties in replacing older technologies with new ones explains the apparent decline in R&D spending among the older firms.

This study has two main applications: Firstly, it is possible to trace employment growth to treatment and estimate the contribution of employment growth to R&D spending of the sampled firms. Secondly, if public support for innovation causes firms to increase their employment, then based on this study, this employment growth will lead to additional R&D spending. These applications will now be discussed in turn.

7.1 Application 1 – Source of Employment Growth and Contribution of Employment Growth to R&D Spending

It is important to try and explain where employment growth comes from and the fraction of R&D coming from employment growth. That is, 'if a firm's employment hadn't engaged with NPL and grown, what would have been the R&D investment?' Methods employed to address these questions are dose-response model estimated through generalized linear model approach (in annex G), and the use of R-squared. The latter measures the proportion of the variation in R&D that is predictable from employment and the other regressors.

In a typical simple linear regression model, the R-square is simply the square of correlation coefficient between the dependent variable and the explanatory variable. In a multiple regression analysis, splitting the R-squared among the regressors is not straightforward, because the R-squared is determined by pairwise correlations of the regressors with each other, as well as, with the dependent variable. However, using this approach makes it possible for correlated predictors to have negative values, that is negative contribution to the explained variance. With some modification involving normalising the standardized[26] coefficient correlation product by the ratio of R-squared to produce a normed standardized coefficient correlation product, however, Dominance Analysis (DA) can produce estimates of relative importance of predictors that sum to R-squared, and are all nonnegative (Grömping, 2007).

Hence, this study uses DA which decomposes variance in real R&D investments attributable to employment, real turnover, real fixed assets, liquidity ratio, and firms' age. That is, it determines the relative importance of these independent variables in the estimation of real R&D investment based on contribution of each to an overall model fit statistic. The summary result is presented in table 7.

Table 7. Dominance Analysis

	Dominance Stat.	Standardized Domin. Stat.	Ranking
Log (employment)	0.134	0.3526	1
Log (age)	0.005	0.0130	5
Log (real turnover)	0.127	0.3331	2
Log (liquidity ratio)	0.004	0.0114	4
Log (real fixed assets)	0.110	0.2900	3
Observations	3123		
Overall Fit Statistic	0.380		

26 The product of the standardized regression coefficient and the zero-order correlation between the predictor and the dependent variable.

The R-squared from the Table 7 is 0.380. This is less than the R-squared of the CRE3 because other firms fixed factors, year sector dummies are excluded. Out of the 0.38 overall fit, employment and turnover account for 0.134 and 0.127 respectively, while other variables account for the remaining 0.118. In ranking, firms' workforce remains the main predictor of the private R&D investments. The complete dominance designation (table F1) shows that employment dominates real turnover and all other variables' impact on R&D but real turnover dominates all other variables except employment effect. Liquidity ratio only dominates age while fixed assets dominate age and liquidity ratio.

Variables used in the CRE3 is replicated to conduct the DA. The result (Table F2) shows that workforce remains the most important predictor explaining variations in private R&D investment of the sampled firms. Time effect is the least contributor.

7.2 Application 2 - Effect of Public Support on R&D Spending

A formula for the effect of public support on R&D spending can be derived by combining the following pieces of information:

- Firstly, the analysis in this report proves that a firm's R&D spending responds positively to changes in the size of its workforce. Based on the estimated elasticity of 0.368, a 10% increase in employment would lead to a 3.68% increase in R&D spending.
- Secondly, our evidence for a connection between employment growth and R&D spending follows on from having previously established, through Belmana's econometric analysis, that 5.5% of all the job-years among the regularly supported firms (during the period 2010-15) were attributable to receiving support from the NMS labs (NPL, LGC, and NEL).[27] Employment growth of firms which are regularly supported in this study also corroborates this.
- Finally, Belmana matched the sample of regularly supported firms to the Business Structure Database (BSD), as well as other ONS datasets, such as, that for the Business Expenditure on Research and Development (BERD). It was found that 49% of the companies receiving regular support from the NMS labs can be found in the sample-frame used for the BERD survey.[28] The R&D spending among the 175 regularly supported firms - used in Belmana's analysis of the effect of support on job-years - was found to average £2 million per year.

Therefore, the support delivered by the NMS labs not only leads to employment growth among the supported firms (as found in Belmana's analysis), but it also simulates R&D activity among these businesses through the connection we have established between employment growth and increased R&D spending.

A short-term leverage ratio for the NMS programme can be computed as follows:

£2 million (average annual R&D spending among the 175 regularly supported firms used in Belmana's analysis of the job-years attributable to support from the NMS labs) x 0.368 (the estimated elasticity from our own study, as detailed in this report) x 0.055 (the proportion of job-years attributable to regular support from the NMS labs, as found in Belmana's econometric analysis) x 360 (the number of firms that were regularly supported by the NMS labs during the

27 In 2017 Belmana Consulting conducted an econometric analysis focussed on finding the effect of public support for innovation job-years among the supported firms. Of the 360 firms that were regularly supported by the NMS labs between 2010 and 2015, it was possible to find very strong matched controls for 175 of these firms, with propensity scores between the 25 and 75 percentiles of the distribution. Econometric analysis (using DiD and PSM) found that 5.5% of the job-years in this sample of 175 regularly supported firms was attributable to support from the NMS labs. (Where these job-years occurred between 2010 and 2015.)

28 This survey is referred to as BERD – *Business Expenditure on Research and Development*.

period 2010-15) gives us a final estimate of £14.57 million per year in additional business R&D spending.

Based on its record of RDEC submissions, NPL spends around £50 million on publicly funded R&D each year; and NPL accounts for at least 90% of all funding delivered through the NMS programme. Therefore, dividing £14.57 million of private spending by £50 million of public spending gives us a short-term leverage ratio of 0.29. This implies that through the of the effect of the NMS labs on employment growth among supported businesses, £100 of public funding leads to around £29 of additional R&D expenditure in the short run.

However, Oxford Economics (2020) shows that whilst short-run R&D leverage begins within the same year that the public investment occurs, the long-run effect takes 15 years to fully materialise, and long-term effects can be five times the size of the short-run effects. Our study focuses on the short-term impact because the period covered (2009-15) is not suitable for a long-run analysis. Nonetheless, it's expected that the ratio of short-run effects to long-run effects would broadly follow that found by Oxford Economics in their macro study. Therefore, by combining our short-run estimate of 0.29 with the findings of Oxford Economics, we estimate a long-run leverage ratio of around 1.50.

8. ROBUSTNESS TESTS AND CAVEATS

There three main caveats:

- Ideally, we would have estimated the coefficients in the model using just the treated firms (defined as the firms who were supported by NPL in 85% of the years under consideration). However, this gave a sample that was just too small to estimate a stable elasticity.
- The period is too limited for the direct computation of long-run R&D leverage rates, based on dynamic panel data models. Moreover, attempts to include lags of R&D investment as additional explanatory variables (necessary to calculate long run leverage ratio) did not produce stable results.
- We found some hints of a possible quadratic (concave) relationship between R&D, r, and employment: Based on a quadratic specification, we might write the regression model as:

$$r = \alpha + \theta_1 . l + \theta_2 . (l)^2 + \cdots + \varepsilon$$

If this were the case, then the elasticity of R&D with respect to employment would be given by:

$$\frac{\partial r_{it}}{\partial l_{it}} = \theta_1 . l_{it} + 2\theta_2 . l_{it}$$

That is, the elasticity might vary with a firm's size; and, if so, we could estimate the elasticity for different size-classes. Note that if $\theta_2 < 0$, then the elasticity decreases as the size of a firm increases. (The R&D spending of larger firms would be less responsive to changes in employment than that of smaller firms.) Unfortunately, the sample was not large enough to confidently determine whether such a quadratic exists, but it might be something to explore in further studies with larger datasets.

9. CONCLUSION

This study is centred on the short-run relationship between variations in the level of R&D spending among companies engaging with the NMS laboratories and the changing level of employment within these companies. It uses a theoretical framework, underpinned in accelerator principle, which predicts that an increased rate of employment growth among companies should lead them to increase the level of tangible and intangible investments. Estimations are based on combining the Heckman technique with fixed and random effect models using administrative data from NPL and financial data from the FAME database to create a firm-level panel dataset for the companies that engaged with the NMS between 2009 and 2015.

The headline results confirmed that employment growth leads to higher levels of R&D spending among the growing companies. The elasticity of R&D with respect to employment is about 0.368. The estimated elasticity is broadly like that found by other economies studied such as Fowkes et al (2015), Yang et al. (2012) and Czarnitzki and Licht (2006). This study also found results consistent with the work of Coad and colleagues - that growth in employment among the supported firms leads to an increase in R&D spending.

In the future, we plan to use the parameters of the estimated models to develop metrics for tracking and comparing R&D activities of the supported firms. This can be used in varieties of scenarios which could be based on the level of support a firm receives and sector a firm belongs.

REFERENCES

Aschhoff, B. (2009). The Effect of Subsidies on R&D Investment and Success – Do Subsidy History and Size Matter? ZEW Discussion Paper No. 09-032. Available at: https://ftp.zew.de/pub/zew-docs/dp/dp09032.pdf

Bia, M & A. Mattei (2008). A Stata package for the estimation of the dose–response function through adjustment for the generalized propensity score. *Stata Journal* 8: 354–373.

Bond, S.R. (2002). Dynamic panel data models: a guide to micro data methods and practice. *Portuguese Economic Journal* **1,** 141–162. https://doi.org/10.1007/s10258-002-0009-9.

Yang, Chih-Hai, Chia-Hui Huang,Tony Chieh-Tse Hou. (2012). Tax incentives and R&D activity: Firm-level evidence from Taiwan. *Research Policy* 41: 1578-1588.

Coad, A., & Rao, R. (2010). Firm growth and R&D expenditure, *Economics of Innovation and New Technology*, Taylor & Francis Journals, vol. 19(2), pages 127-145.

Coad, A., Cowling, M., Nightingale, P., Pellegrino, G., Savona, M. and Siepel, J. (2014), Innovative Firms and Growth. UK Innovation Survey: UK.

Coad, A., and Grassano, N. (2016). Disentangling the processes of firm growth and R&D investment. No. JRC103175. Joint Research Centre (Seville site).

Coad, A., & Rekha, R. (2007). Firm Growth and R&D Expenditure. Max Planck Institute of Economics, Evolutionary Economics Group, Papers on Economics and Evolution. 19. 10.1080/10438590802472531.

Corrado, C., Haskel, J., Jona-Lasinio, C. Iommi, M. (2016). Intangible investment in the EU and US before and since the Great Recession and its contribution to productivity growth. Access here.

Czarnitzki, D. and G. Licht. (2006). Additionality of public R&D grants in a transition economy: the case of Eastern Germany. Available at: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.467.8013&rep=rep1&type=pdf

Economic Insight (2015). What is the relationship between public and private investment in R&D?" April. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/438763/bis-15-340-relationship-between-public-and-private-investment-in-R-D.pdf

Fowkes, R. K., Sousa, J., and Duncan, N. (2015), Evaluation of Research and Development Tax Credit. HMRC Working Paper 17. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/413629/HMRC_WorkingPaper_17_R_D_Evaluation_Final.pdf

Goel, R.K., Nelson, M.A. (2022). Employment effects of R&D and process innovation: evidence from small and medium-sized firms in emerging markets. *Eurasian Bus Rev* **12,** 97–123. https://doi.org/10.1007/s40821-022-00203-6

Guardabascio, B & M. Ventura (2014). Estimating the dose–response function through a generalized linear model approach, *Stata Journal* 14(1): 141–158

Guellec, D. & Van Pottelsberghe, B., 2003. The impact of public R&D expenditure on business

R&D. *Economics of Innovation and New Technology*, pp. 225-243.

Griffith, Rachel & Huergo, Elena & Mairesse, Jacques & Peters, Bettina. (2006). Innovation and Productivity Across Four European Countries. *Oxford Review of Economic Policy*. 22. 483-498. 10.1093/oxrep/grj028.

Grömping, U. 2007. Estimators of relative importance in linear regression based on variance decomposition. The American Statistician 61(3): 139-147. DOI: 10.1198/000313007X223739

Hottenrott, Hanna; Lopes-Bento, Cindy (2012). (International) R&D collaboration and SMEs: The effectiveness of targeted public R&D support schemes. ZEW Discussion Paper No. 12-086. Available at: https://ftp.zew.de/pub/zew-docs/dp/dp12086.pdf

Huang., X And Verma, A. (2018). Industry- and firm-level determinants of employment relations in China: a two-level analysis, *The International Journal of Human Resource Management*, 29(2): https://doi.org/10.1080/09585192.2016.1155165

Kenneth Y. H., and Adam B.J (1993), Effect of Liquidity on Firms' R&D Spending, *Economics of Innovation and New Technology*, 2(4)

King, M and Renedo, E (2020). Achieving the 2.4% GDP target: the role of measurement in increasing investment in R&D and innovation, NPL Report IEA 3. https://www.npl.co.uk/about-us/who-we-work-with/government/npl-evidence-and-analysis/2-4-gdp-target-iea3.aspx

Kluve, J., H. Schneider, A. Uhlendorff & Z. Zhao (2007). Evaluating Continuous Training Programs Using the Generalized Propensity Score, IZA DP No. 3255.

Knox, A. D. (1970). "The Acceleration Principle and the Theory of Investment: A Survey". In Shapiro, Edward (ed.). Macroeconomics: Selected Readings. New York: Harcourt, Brace & World. pp. 49–74. *ISBN 0-15-551207-2*.

Montmartin, B. and M. Herrera. 2014. Internal and external effects of R&D subsidides and fiscal incentives: Empirical evidence using spatial dynamic panel models. GREDEG Working Paper No. 2014-09. Available at: http://www.gredeg.cnrs.fr/working-papers/GREDEG-WP-2014-09.pdf

Montmartin, B. and M. Herrera. (2015) Internal and external effects of R&D subsidides and fiscal incentives: Empirical evidence using spatial dynamic panel models. Research Policy, Vol. 44/5 1065-1079

Mulligan, K., Lenihan, H., Doran, J., and Roper, S. (2022), Harnessing the science base: Results from a national programme using publicly-funded research centres to reshape firms' R&D, *Research Policy*, 51 (4). https://doi.org/10.1016/j.respol.2021.104468.

Sokhanvar, A., and Çiftçioğlu, S. (2022). The Impact of R&D on Skill-specific Employment Rates in the UK and France. European Review, 30(2), 151-171. doi:10.1017/S1062798720000010

Ciarli, T., Marzucchi, A., Salgado, E., and Savona, M. (2018). The Effect of R&D Growth on Employment and Self-Employment in Local Labour Markets, SPRU Working Paper Series 2018-08, SPRU - Science Policy Research Unit, University of Sussex Business School.

Oxford Economics (2020). The relationship between public and private R&D funding. BEIS Research Paper Number 2020/010. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/897470/relationship-between-public-private-r-and-d-funding.pdf

Office for National Statistics (ONS) 2018. “Investments in intangible assets in the UK: 2018”. Access here.

Romer, Paul M, 1990. "**Endogenous Technological Change**," Journal of Political Economy, University of Chicago Press, vol. 98(5), pages 71-102, October.

Sussex, J., Feng, Y., Mestre-Ferrandiz, J. *et al. (2016).* Quantifying the economic impact of government and charity funding of medical research on private research and development funding in the United Kingdom. *BMC Med* **14,** 32. https://doi.org/10.1186/s12916-016-0564-z

Tassey, G. (2004). Underinvestment in public good technologies. The Journal of Technology Transfer, 3(1-2), pp. 89-113.

Tera Allas (2014) *Insights from international benchmarking of the UK science and innovation system*, BIS ANALYSIS PAPER NUMBER 03, JANUARY 2014

Yazdanfar, D. and Öhman, P. (2015), Firm-level determinants of job creation by SMEs: Swedish empirical evidence", *Journal of Small Business and Enterprise Development* 22 (4): 666-679. https://doi.org/10.1108/JSBED-06-2013-0084

ANNEX A - FULL DERIVATIONS OF THEORETICAL FRAMEWORK

Main Assumptions and Model Set Up

- The rental rate of capital, r, and wage rate, w, may change over time as the supply and demand of labour and capital evolve. However, they are determined by the outcome of market process beyond a firm's control.
- Bowley's Law (the law of constant labour share) was one of Kaldor's facts that macroeconomics seeks to explain. We will suppose that the parameters in the Cob Douglas production function are fixed parameters: $\alpha \approx 0.75, \beta \approx 0.25$.
- Assume that the depreciation rate, δ, is a constant parameter. In contrast, a firm's growth in employment, g, and its reinvestment rate, s, are variables that can change over time and may differ across the population of firms.
- Support from the NMS labs causes firms to grow faster than they would have done otherwise. Suppose that the impact on a firm's growth is Δg.

The production function is $Y = (AL)^{\alpha} K^{\beta}$, with constant returns to scale: $\alpha + \beta = 1$. Output per effective worker is denoted by $y = \frac{Y}{AL}$; and capital intensity is denoted by $k = \frac{K}{AL}$. Combining these definitions of y and k with constant returns to scale gives us $y = k^{\beta}$.

The assumptions needed for the dynamic aspects of the model are as follows:

- There is exogenous growth in the workforce: $\frac{\dot{L}}{L} = g_L$, where g_L is the growth rate. That is, $L = L_0.\exp(g_L t)$, where L_0 is the size of the initial size of the workforce.
- Exogenous growth in knowledge stock: $\frac{\dot{A}}{A} = g_a$
- The capital stock evolves according to $\dot{K} = s.Y - \delta.K$, where s is the rate at which income is retained and reinvested in the firm. Let $I = s.Y$ denote the firm's level of investment.

The profit function is $\Pi = Y - rK - wL$, where r is the rental rate of capital (interest rate) and w is the wage rate.

Finding the Equilibrium Capital Intensity

The first order conditions for profit maximisation gives us:

$$\frac{\partial \Pi}{\partial K} = \beta.\left(\frac{Y}{K}\right) - r = 0 \Rightarrow \frac{Y}{K} = \frac{r}{\beta}$$

$$\frac{\partial \Pi}{\partial L} = \alpha.\left(\frac{Y}{L}\right) - w = 0 \Rightarrow \frac{Y}{L} = \frac{w}{\alpha}$$

The capital stock evolves according to $\dot{K} = s.Y - \delta.K$, where s is the rate at which income is retained and reinvested. Dividing through by K gives us:

$$\frac{\dot{K}}{K} = s.\left(\frac{Y}{K}\right) - \delta$$

Recall that, $k = \frac{K}{AL}$. Linearising and differentiating with respect to time gives:

$$\frac{\dot{k}}{k} = \frac{\dot{K}}{K} - \frac{\dot{A}}{A} - \frac{\dot{L}}{L} = s.\left(\frac{Y}{K}\right) - (\delta + g_a + g_L)$$

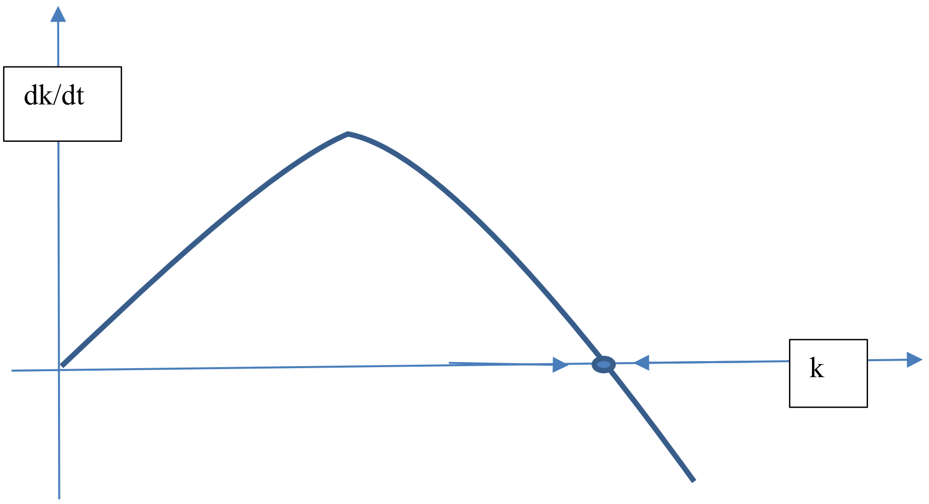
If we let $g = g_a + g_L$, then the previous expression simplified to:

$$\frac{\dot{k}}{k} = s.\left(\frac{yAL}{K}\right) - (\delta + g)$$

Multiplying through by $k = \frac{K}{AL}$ gives $\dot{k} = s.y - (\delta + a + g)k = sk^{\beta} - (\delta + g)k$. Hence, the evolution of capital intensity is governed by: $\dot{k} = s.k^{\beta} - (\delta + g).k$. The RHS of this expression is just a function of k.

- The first derivative of the RHS with respect to k is $s\beta/k^{\alpha} - \delta - g$
- The second derivative of the RHS with respect to k is $-s\alpha\beta/k^{1+\alpha} < 0$

It can be seen that $\dot{k}$ initially increases as k increases, but it then reaches a maximum at $k = (s\beta/(\delta + g))^{1/\alpha}$, after which k reaches its maximum and begins it decreases.



There exists a unique equilibrium for capital intensity:

$$\dot{k} = 0 \Rightarrow s.y = (g + \delta).k \Rightarrow s.(y/k) = \delta + g$$

And, because $y/k = Y/K = r/\beta$, we finally arrive at: $\beta = sr/(g + \delta)$. The equilibrium value of k is defined by: $s. = (\delta + g).k^{\alpha}$. And, from this, we get:

$$k^* = \left(\frac{s}{\delta + g}\right)^{1/\alpha}$$

Implications of Fixing some Parameters

Making β and δ constant parameters have some interesting implications: if g increases, then s must also increase. (The expression $\beta = sr/(\delta + g)$ must hold whilst the fixed parameters remain the same.) The relationship between s and g can be written as: $s = \beta.(\delta + g)/r$. Substituting for s in the earlier expression for k* shows that the equilibrium capital intensity is given by:

$$k^* = \left(\frac{\beta}{r}\right)^{1/\alpha}$$

Notice that this formula for k* is independent of g, and so capital intensity does not depend on the firm's growth rate.

Implications for Investment

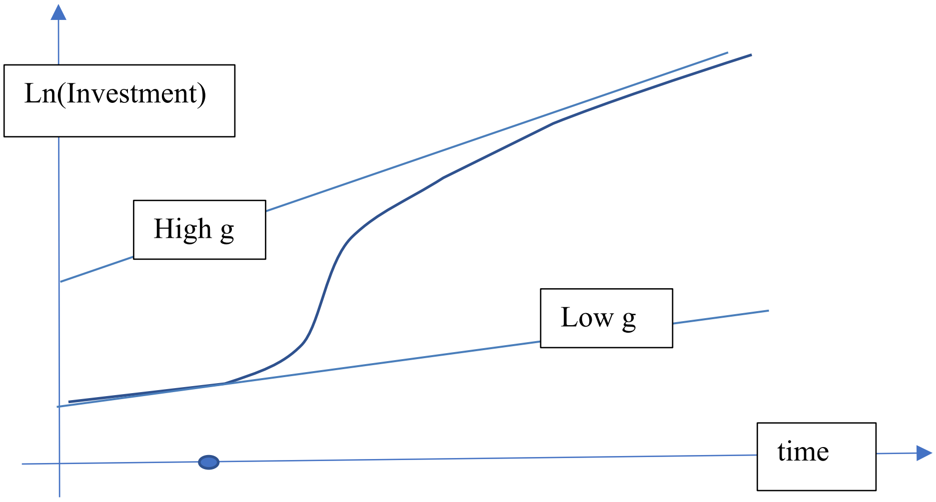
If $k = k^*$, then $\dot{k} = 0$, which implies that: $s.y - (\delta + g).k = 0$. Hence, in equilibrium, the investment *per employee* is given by $(\delta + g).k^*$. It follows that equilibrium investment is given by $I^* = L.(\delta + g).k^*$. By substituting for k* and L, this can be re expressed as:

$$I^* = L_0 \exp(gt).(\delta + g).\left(\frac{\beta}{r}\right)^{1/\alpha}$$

Taking logs of our expression for equilibrium investment gives:

$$\ln(I^*) = \ln(L_0) + \text{nt} + \ln(\delta + g) + \frac{1}{\alpha}\ln(\beta) - \frac{1}{\alpha}\ln(r)$$

The equilibrium level of investment is clearly an increasing function of g. The effect of an increase in g on investment can be illustrated using a graph with log investment as the vertical and time as the horizontal. The parameter n changes both the slope and the intercept of the line.



If there was a shift in the value of g such that $g \to g + \Delta g$, then the evolution of system would change from following bottom line to following the top line. Since support from the NMS cause a sustained increase the size of the firms' workforce, this implies an increase in the rate at which income is reinvested. Along with other productive investments, there is an increase in the level of R&D spending.

ANNEX B: SORTING OUT ISSUES WITH SIC CODES

We included condensed SIC codes of the sampled firms as additional control variables in the analysis. To make the SIC codes usable, we first examine the nature of the SIC codes and found out that some are less than the ideal five digits number. We sorted this out by replacing the incomplete SIC codes with appropriate SIC codes. Subsequently, we created SIC sections following the standard classifications. However, we realized that sector "M" is broadly Professional, scientific, and technical activities and has a lot of activities grouped together. Hence, we regrouped these activities into 4 sub-categories including legal, accounting and market research activities, technical services, R&D firms, and professional, scientific, and other technical services firms.

Further, we realized that NPL had no dealings in terms of collaborations and paid-for-services with some of the SIC sections. These SIC sections include "I", "P", "T" and "U". These are firms in the business of hotel and accommodation, education, activities of households as employers; undifferentiated goods and services producing activities of households for own use, and activities of extraterritorial organizations and bodies, respectively. Also, engagement of NPL with firms in the SIC sectors "K", "L", "Q", "R", and "S" is thin. These are firms in the businesses of finance and banking, real estate and accommodation, hospital, medical, dental and care practices, performing arts and recreation, trade unions, political, and professional associations, respectively. Although the SIC section "K", "L", "Q", "R", and "S" less of conventional businesses but we noticed that few of them are. Hence, we created a random sector/firm for few treated and supported conventional businesses/firms in "K", "L", "Q", "R", and "S". Hence, SIC sections "I", "P", "T", "U and few unconventional firms in SIC section "K", "L", "Q", "R", and "S" are excluded from our sample.

SIC section C (manufacturing activities) was further split into level of manufacturing technology to include low medium-low, medium-high, and high technology manufacturing firms while the SIC section "O" was merged with high-tech manufacturing. This is because SIC sector "O" include firms such as BAE Systems Electronics Limited, Raytheon Systems Limited, GE Aviation Systems Limited General Dynamics United Kingdom Limited BAE Systems Integrated

System Technologies Limited and Datong Limited. There are more of high-tech manufacturing than low tech manufacturing in the sample.

Overall, we included 18 SIC sectors after excluding those that are less relevant to our analysis. The list is indicated in the Table B1.

Table B1. SIC Sector Classifications of Sampled Firms

Sector used as used in the estimations	SIC sector	Description	Freq.	Percent
_Isector_1	B	Mining and Quarrying	50	1.2
_Isector_2	C1	Low-tech Manufacturing	314	7.4
_Isector_3	C2	Medium-Low Tech Manufacturing	489	11.5
_Isector_4	C3	Medium-high tech manufacturing	605	14.2
_Isector_5	C4	High tec manufacturing	507	11.9
_Isector_6	D	Electricity, gas, steam and air conditioning supply.	39	0.9
_Isector_7	E	Water supply, sewerage, waste management and remediation activities.	27	0.6
_Isector_8	F	Construction	140	3.3
_Isector_9	G	Wholesale and retail trade; repair of motor vehicles and motorcycles.	321	7.6
_Isector_10	H	Transportation and storage.	42	1.0
_Isector_11	J	Information and communication	280	6.6
_Isector_12	M_1	Legal, accounting and market research activities.	242	5.7
_Isector_13	M_2	Tech services firms	332	7.8
_Isector_14	M_3	R&D firms	244	5.7
_Isector_15	M_4	Professional, scientific and other technical services firms	236	5.6
_Isector_16	N	Administrative and support service activities	151	3.6
_Isector_17	N.E.C	Not Elsewhere Classified	85	2.0
_Isector_18	RF2*	Random firms	144	3.4
Total			**4,248**	**100**

Note: Random firms (These activities cut across different SIC codes as described above)
N.E.C includes Agriculture, Forestry and Fishing and other activities with missing SIC codes

ANNEX C: INTERPOLATION OF MISSING EMPLOYMENT AND REAL TURNOVER VALUES

The process of interpolating the missing employment and turnover data involved the use of two Heckman regressions. The specifications followed Huang and Verma (2018) and Yazdanfar and Ohman (2015). The estimated Heckman models (Table B1 and B2) passed specification tests (Table B3 and B4). The estimations are also adequate given the significant Wald Chi-2 tests.

In 2016, the UK's accounting rules stipulated that the gross assets of 'small entities' should not exceed a threshold of £5.1m, a rise from £3.26m in earlier years. In terms of 2010 prices, this new threshold was £4.6m. We used this information to create separate dummies for 10% below and 10% above the firms' gross assets threshold. The regressor '*threshold_b*' in Table B1 and B2 is a dummy variable for firms whose gross assets are below this threshold. Also, '*log_ci_n*' is the natural log of capital intensity, where capital intensity is defined at the ratio of fixed assets to total assets.

The estimated Heckman models used in interpolating log of employment and real turnover, respectively is stated as follows:

heckman logemp (c.logrta c.loglr)##(i.threshold_b) c.logci_n c.year, select(c.logci_n c.loglr c.logrta i.year i.threshold_b)
linktest, select(c.logci_n c.loglr c.logrta i.year i.threshold_b)

heckman logrto (c.logrta c.logempx)##(i.threshold_b) c.logci_n loglr c.year, select(c.logci_n c.logrta i.year i.threshold_b logempx loglr)
linktest, select(i.year c.logci_n c.logrta i.threshold_b logempx loglr)

At 5% level, the estimated models passed link specification tests as shown in equations above and in Table C1 and C2.

Table C1. Linktest (log of employment estimate)

logemp	Coef.	Std. Err.	z	P>z
_hat	0.951***	0.030	31.760	0.000
_hatsq	0.004*	0.003	1.680	0.093
_cons	0.134	0.090	1.490	0.136
Wald Chi-2	20146.33***			
Selected Obs	13,078			

Note: *,*** implies significant at 1% and 10% level, respectively.

Table C2. Linktest (log of real turnover estimate)

logrto	Coef.	Std. Err.	z	P>z
_hat	1.002***	0.023	43.96	0.000
_hatsq	-6.8E-05	0.001	-0.07	0.945
_cons	-0.009	0.133	-0.06	0.949
Wald Chi-2	53338.64***			
Selected Obs	13,417			

Note: *** implies significant at 1% level.

To fill-in the missing data, the logs of employment and turnover estimated using the commands 'predict yhat1, xb' and 'predict yhat2, xb', respectively, following each Heckman regression. αSubsequently, to get estimates in levels, we obtained the exponential values of log of employment and turnover by generating *m1* = EXP (*yhat1*) and *m2* = EXP (*yhat2*), respectively.

To deal with issues caused by discreet changes in employment and turnover data, especially among the smaller firms, a negative binomial regression was the preferred approach. Hence, we estimated negative binomial regressions of employment and real turnover using the exponentials of the fitted values from the Heckman regressions. The (predicted) natural logs of employment and turnover entered these binomial regressions as exposure variables (Table B5 and B6). Because employment and turnover are roughly log normal, the exponentiated values of the constant terms (α coefficients) in Table B5 and B6 were used to scaleup 'm1' and 'm2', respectively. We subsequently replaced 'yhat1' and 'yhat2' with natural logarithm of 'm1' and 'm2' respectively.

Finally, we generated fully populated variables for log of employment and turnover by using their observed values wherever possible, and their interpolated values ('yhat1' and 'yhat2') if employment or turnover was missing. This procedure added around 1,600 firms to the dataset whose employment or turnover would otherwise have been missing.

Table C3. Negative Binomial Estimate for Employment

```
Negative binomial regression                    Number of obs     =     12,940
                                                LR chi2(0)        =       0.00
Dispersion     = mean                           Prob > chi2       =          .
Log likelihood = -83915.312                     Pseudo R2         =     0.0000

------------------------------------------------------------------------------
         emp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .4690783   .0070861    66.20   0.000     .4551899    .4829668
      ln(m1) |          1  (exposure)
-------------+----------------------------------------------------------------
    /lnalpha |  -.4506549   .0115668                     -.4733254   -.4279843
-------------+----------------------------------------------------------------
       alpha |   .6372107   .0073705                      .6229273    .6518216
------------------------------------------------------------------------------
LR test of alpha=0: chibar2(01) = 2.0e+07              Prob >= chibar2 = 0.000
```

Table C4. Negative Binomial Estimate for Real Turnover

```
Negative binomial regression                    Number of obs     =     13,417
                                                LR chi2(0)        =       0.00
Dispersion     = mean                           Prob > chi2       =          .
Log likelihood = -150337.38                     Pseudo R2         =     0.0000

------------------------------------------------------------------------------
         rto |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .2276181   .0056175    40.52   0.000      .216608    .2386281
      ln(m2) |          1  (exposure)
-------------+----------------------------------------------------------------
    /lnalpha |  -.8608091   .0114904                     -.8833298   -.8382884
-------------+----------------------------------------------------------------
       alpha |   .4228198   .0048584                       .413404    .4324501
------------------------------------------------------------------------------
LR test of alpha=0: chibar2(01) = 3.6e+09              Prob >= chibar2 = 0.000
```

ANNEX D: AUXILIARY TABLES

Table D1 Summary Average Annual Statistics of Raw Data

Variable	Mean	2009	2010	2011	2012	2013	2014	2015
rfa	Arithmetic	261336.2	253830.3	268945.5	238598.8	241107.5	245946.6	274322.1
	Median	568.1	567.5	560.3	544.6	547.7	556.1	583.5
	Geometric	612.2	581.7	575.9	539.2	537.5	519.0	518.6
	Count	3,422	3,522	3,593	3,685	3,724	3,774	3,807
rto	Arithmetic	503418.8	517970.8	559670.3	526871.2	512200.8	493760.4	443443.5
	Median	19747.8	19823.3	21957.3	21541.4	21137.5	21188.2	20551.6
	Geometric	21998.4	22821.4	25489.9	25316.4	24912.1	24323.0	23374.3

	Count	1,879	1,951	1,963	2,003	2,024	2,052	2,050
lr	Arithmetic	2.30	2.34	2.25	2.39	2.58	2.68	2.83
	Median	1.3	1.3	1.3	1.4	1.5	1.5	1.5
	Geometric	1.37	1.36	1.37	1.41	1.46	1.51	1.51
	Count	3,504	3,605	3,693	3,768	3,835	3,879	3,934
age	Arithmetic	19.4	20.4	21.4	22.4	23.4	24.4	25.4
	Median	14.0	15.0	16.0	17.0	18.0	19.0	20.0
	Geometric	14.9	15.4	15.9	16.4	17.2	18.0	18.4
	Count	4,237	3,742	3,829	3,919	3,982	4,042	4,120

Note: fra (real fixed assets), rta (real turnover), lr (liquidity ratio), age (age of firms)

Note: fra (real fixed assets), rta (real turnover), lr (liquidity ratio), age (age of firms)

Table D2. Summary Statistics of the Raw data

Variable	Obs	Mean	Std. Dev.	Min	Max	Skewness
rrd	3,205	15378.6	78204.4	0.00	1250075.0	10.69
emp	13,327	1570.6	9464.8	1.00	277684.0	14.64
rto	13,922	507785.6	5020083.0	0.25	237000000.0	33.27
age	29,659	22.4	22.2	-8.01	141.0	1.78
lr	26,218	2.5	4.7	0.00	98.2	9.15
rfa	25,527	254799.7	2979656.0	0.00	129000000.0	26.32

The specification of the estimated R&D hackman selection model used to correct selectivity bias in the subsequent panel model elasticity estimations is indicated below:

xi:heckman logrrd logempx logage logrtox loglr logrfa i.year i.sector, select (i.listed_f c.logempx#c.logrtox c.logempx##c.logempx logage c.logrtox##c.logrtox loglr c.logrfa##c.logrfa i.year i.sector) mills(mills)

Specification of hackman model used to correct selectivity bias in the first difference R&D panel model estimations is specified as follows:

xi:heckman d.logrrd d.logempx d.logage d.logrtox d.loglr d.logrfa i.sector i.year if year>2009, select (i.listed_f c.logempx#c.logrtox c.logempx##c.logempx logage c.logrtox##c.logrtox loglr c.logrfa##c.logrfa i.sector i.year) mills(mills2) robust

Table D3. Hausman Test

```
              ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |      fe           re          Difference          S.E.
-------------+----------------------------------------------------------------
    logempx  |   .4338262     .4504147       -.0165885        .0507785
    logrtox  |   .0536069     .1425216       -.0889147        .0192928
      loglr  |  -.0556912    -.0309915       -.0246997        .0128192
     logrfa  |   .0995518     .1140715       -.0145197        .0184417
 _Iyear_2010 |  -.2058543    -.2916467        .0857924        .0838765
 _Iyear_2011 |  -.2380205    -.3298154        .0917949        .0912202
 _Iyear_2012 |  -.1875038     -.287314        .0998101        .0952537
 _Iyear_2013 |  -.1684444    -.2682949        .0998506        .0934674
 _Iyear_2014 |  -.0913056    -.1850448        .0937393        .0692326
 _Iyear_2015 |  -.0908449    -.1425297        .0516848        .0202379
      mills  |  -.4688895    -.6228242        .1539348        .1694303
------------------------------------------------------------------------------
                          b = consistent under Ho and Ha; obtained from xtreg
           B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(11) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                           =  -134.63    chi2<0 ==> model fitted on these
                                         data fails to meet the asymptotic
                                         assumptions of the Hausman test;
                                         see suest for a generalized test
```

Table D4: Estimations by firms sizes.

	Large (non-SME firms)			Small (SME firms)		
logrrd	Coef.	Std. Err.	P>z	Coef.	Std. Err.	P>z
logempx	0.310	0.081	0	0.424	0.131	0.001
logage	-0.052	0.088	0.557	-0.256	0.116	0.028
logrtox	0.115	0.065	0.075	0.043	0.052	0.406
logrtox_mean	0.302	0.087	0.001	0.202	0.099	0.041
loglr	-0.036	0.052	0.489	-0.077	0.048	0.106
loglr_mean	0.201	0.138	0.144	0.252	0.115	0.029
logrfa	0.061	0.043	0.158	0.103	0.048	0.033
Year Dummies	Yes			Yes		
Sector Dummies	Yes			Yes		
mills	-0.600	0.289	0.038	-0.455	0.422	0.28
_cons	0.082	1.338	0.951	2.616	2.541	0.303
sigma_u	1.55			1.615072		
sigma_e	0.617			0.67352		
rho	0.863			0.851856		
Wald Chi-square/ F-statistics	386.8***			Not computed		
R-Square (overall)	0.425			0.28		
Number of firms	412			383		
Obs	1,766			1,366		

Note: *** indicates significance at 1% level.

ANNEX E: LONG-RUN ANALYSIS

Table E1. Long-run elasticity estimate of R&D investment				
	Pooled Least Square	Fixed Effects	Random Effect	Difference and system GMM
R&D				
$R\&D_{t-1}$	0.958*** (0.012)	0.063*** (0.025)	0.818** (0.013)	0.273*** (0.102)
Employment	0.016 (0.026)	0.379*** (0.082)	0.095*** (0.027)	0.783** (0.273)
Capital	0.014 (0.016)	0.103** (0.039)	0.044** (0.017)	-0.098 (0.144)
Liquidity	0.015 (0.026)	-0.071* (0.042)	-0.010 (0.026)	-0.157 (0.132)
Age	0.009 (0.022)	-		-0.043 (0.083)
Constant	-0.289 (0.223)	3.439*** (0.531)	0.010 (0.213)	1.606 (1.041)
Sector dummies	Yes	-		Yes
Year dummies	Yes	Yes		Yes
Rho	-	0.89		-
F test that all u_i=0:	-	3.14**		-
R-squared	0.88	0.45	0.89	-
Wald chi2(28)/ F(28, 2168)	488.16***	7.60***	9521.99**	23875.53***
Arellano-Bond test for AR(1) in first differences: (z-stat)	-	-	-	-3.43***
Arellano-Bond test for AR(2) in first differences: (z-stat)	-	-	-	1.20
Sargan test of overid. restrictions: chi2(32)	-	-	-	67.39***
Hansen test of overid. restrictions: chi2(32)	-	-	-	12.65
Number of obs	2,204	2,204	2,204	2204
Number of groups	656	656	656	656

Note: The variables have been log transformed and expressed in real terms. Employment and real turnover are interpolated values. Standard errors are in the parentheses

The estimated GMM model is specified as:

xi:xtabond2 logrrd l.logrrd logempx logrfa loglr logage i.sector i.year, gmm(l.logrrd logempx logrfa loglr logage, collapse equation(diff)) ivstyle(i.sector i.year logage,equation(level)) ivstyle(l.logempx l.logrfa l.loglr,equation(diff)) twostep robust

The above equation contain mainly vectors of strictly exogenous covariates and predetermined covariates (which includes the lag of log of real R&D) potential endogenous covariates, all of which may be potentially correlated with past errors.

First-differencing of the gmm equation removes the unobserved individual-level effects thus eliminating a potential source of omitted variable bias in estimation. However, differencing variables that are predetermined but not strictly exogenous makes them endogenous.

The ivstyle option specifies a set of variables to serve as standard instruments, these are both at level and difference for time-invariant instruments and a year lag of strictly exogenous covariates and predetermined covariates of potential endogenous covariates. The twostep specifies calculation of two-step estimator.

Table F1. Conditional dominance statistics, complete and strongest dominance designations

```
General dominance statistics: Linear regression
Number of obs             =                  3132
Overall Fit Statistic     =                0.3802
```

logrrd	Dominance Stat.	Standardized Domin. Stat.	Ranking
logempx	0.1340	0.3526	1
logage	0.0049	0.0130	4
logrtox	0.1266	0.3331	2
loglr	0.0043	0.0114	5
logrfa	0.1103	0.2900	3

Conditional dominance statistics

	#indepvars: 1	#indepvars: 2	#indepvars: 3	#indepvars: 4	#indepvars: 5
logempx	0.3586	0.1925	0.0820	0.0237	0.0135
logage	0.0093	0.0044	0.0030	0.0038	0.0043
logrtox	0.3489	0.1840	0.0746	0.0174	0.0081
loglr	0.0081	0.0038	0.0027	0.0033	0.0037
logrfa	0.3225	0.1624	0.0580	0.0060	0.0023

Complete dominance designation

	dominated?: logempx	dominated?: logage	dominated?: logrtox	dominated?: loglr	dominated?: logrfa
dominates?:logempx	0	1	1	1	1
dominates?:logage	-1	0	-1	0	0
dominates?:logrtox	-1	1	0	1	1
dominates?:loglr	-1	0	-1	0	0
dominates?:logrfa	-1	0	-1	0	0

```
Strongest dominance designations

logempx completely dominates logage
logrtox completely dominates logage
logempx completely dominates logrtox
logempx completely dominates loglr
logrtox completely dominates loglr
logempx completely dominates logrfa
logrtox completely dominates logrfa
logage conditionally dominates loglr
logrfa generally dominates logage
logrfa generally dominates loglr
```

Table F2. Replication of CRE model variables in Dominance Analysis

```
General dominance statistics: Linear regression
Number of obs           =                 3132
Overall Fit Statistic   =               0.4008
```

logrrd	Dominance Stat.	Standardized Domin. Stat.	Ranking
logempx	0.0906	0.2259	1
logage	0.0056	0.0139	6
logrtox	0.0814	0.2032	3
logrtox_~n	0.0827	0.2063	2
loglr	0.0020	0.0050	8
loglr_mean	0.0029	0.0072	7
logrfa	0.0726	0.1811	4
mills	0.0631	0.1573	5

Conditional dominance statistics

ANNEX G: ASSESSING THE IMPACT OF CONTINOUS INTERVENTION ON EMPLOYMENT GROWTH WITH DOSE RESPONSE MODEL

Framework of Dose Response Model

Employment growth is treated to be exogenous in the R&D leverage analysis. One questions which may require explanation is how effective are NPL programs in determining employment growth with continuous treatment exposure? Dose response model through a generalized propensity score offers provides some answers. Belmana's analysis made effort at this approach, but we replicated it with our data to ensure that we have narratives which fit together.

Dose response model is a further development from binary treatment which assumes a pair of potential outcomes $Q_i(0), Q_i(1) \perp \mathrm{T}|\mathrm{X} \quad \forall\, T \in (0,1)$. This expression means that the potential outcomes for a firm ($Q_i(0)$ and $Q_i(1)$) become independent of that firm's treatment status after conditioning on the covariates. This eliminates any bias in the estimated treatment effect associated with differences in covariates (X) across the population of firms. The covariates are the firm's specific independent variables that can influence the outcome of a given statistical trial, but which is not of direct interest.

The observed outcome Q_i may be written in terms of the potential outcomes and the observed exposure, T_i as:

$$(1 - T_i)Q_i(0) + T_iQ_i(1) \;\; \forall\, T_i \in (0,1) \qquad (1)$$

Like the previous expression, equation (1) implies that potential outcomes are independent of treatment status after conditioning on the observable covariates, which may be causing changes in outcomes between treated and non-treated firms. This condition disentangles the treatment effect from the influence of other firm-specific factors which may affect outcomes.

Unlike the binary treatment, dose response model assumes a continuous treatment. That is, given a random sample of firms (1,2…, N), there exists a set of potential outcomes (Q) corresponding to different treatment levels, $T_i \in (t_{1,}t_{2,\dots,}, t_d)$ such that:

$$Q(t) \perp \mathrm{T}|\mathrm{X} \qquad \forall\, t \in (t_1, t_d) \qquad (1)$$

Where: $Q(t)_i$ = $Q_i\,(T_i\,)$ potential outcomes corresponding to the level of treatment received which is assumed to be a continuous, $X(t)_i$ is vector of covariates, and T_i is treatment received with $T_i \in (t_1, t_d)$. In the case, the potential outcomes take different forms expressed as:

$$\{Q_i(t_1), Q_i(t_2), .., Q_i(t_d)\} \qquad (3)$$

Equation (3) means that the potential outcomes vary over time depending on whether a firm is treated or not at a particular time. This helps to assess how outcomes changes with treatment level. That is, the observed outcome, Q_i, can be written in terms of the potential outcomes and the observed treatment, T_i , as:

$$Q_i = \sum_{j=1}^{d} t_j\,(T_i)Q_i(t_j) \qquad (4)$$

Where $t_j(T_i)$ is the random set of treatments (Ʈ) that a firm received, with $t_j(T_i) = 1$ if treatment is received and zero otherwise. This is important in generating potential outcomes associated with treatments.

If treatment is continuous, the potential outcomes are $\{Q_i(t), t \in$ Ʈ$\}$ rather than $(1 - T_i)Q_i(0) + T_iQ_i(1)\ \forall\, T_i \in (0,1)$ in the case of binary treatment. Hence, dose response model traces the path of changes in $\mathrm{Q_i(t)}$ as t changes.

Further, minimizing the bias effect of differences in covariates requires co-variate balancing. The dose-response model generates generalized propensity score (GPS) which helps to minimize this bias. According to Guardabascio and Ventura (2014), the *GPS, r* (*T*, *X)*, defines *the* conditional density of the treatment (T) given the covariates (X). Its balancing property helps to assess its adequacy in terms of creating artificial matched control group. That is, within strata with the same covariate value (X) of *r* (*t*, *X*), treatment does not depend on the value of X. This condition makes T independent of any covariate in the trial.

There are two variants of dose response model: the one which is estimated with maximum likelihood estimator and the other which is estimated with generalized linear model. The latter is more recent, flexible, and suitable when intervention variable is not necessarily normally distributed (Guardabascio & Ventura, 2014). Studies (such as Kluve, et al, 2007; Bia& Matti, 2008; Guardabascio & Ventura, 2014;) have estimated dose–response function adjusting for the generalized propensity score to examine the impact of continuous treatment on economic outcomes of interest.

Impact of Continuous Treatment on Employment Growth of Supported Firms

The results from different estimations established that firm's growth influences private R&D spending of users of NPL services. What is lacking however, is the explanation of the source of growth. To unravel this, we estimated dose response model adjusting for the generalized propensity score (gps). Firm's workforce growth was used as a measure of firm's growth and as an outcome variable while incidences of NPL support was used as an intervention variable. Employment growth was normalized using minimum-maximum method. This method maps a value, x_i of X to x_i^I in the range [*new_min*$_X$, *new_max*$_X$] by computing:

$$x_i^I = \frac{x_i - X_{min}}{X_{max} - X_{min}} \qquad (5)$$

This technique gets all the data in the new range (0, 1) while preserving the relationships among the original data values. It also helps to achieve smaller standard deviations which suppresses the impact of outliers.

In this context of this study, this approach helps in the implementation of our dose response model. That is, it enables us to align actual employment growth (which could take negative values; table G1) with the average potential employment outcome for each level of treatment which is stored to a positive 10-dimensional vector ranging between 0 and 1. Our approach follows Hirano and Imbens (2004).

Table G1. Distribution of employment growth

Demp

	Percentiles	Smallest		
1%	-1.386294	-1.640529		
5%	-.7657351	-1.625699		
10%	-.4476862	-1.622683	Obs	1,591
25%	-.1616416	-1.591894	Sum of wgt.	1,591
50%	.0673203		Mean	.0755901
		Largest	Std. dev.	.4753573
75%	.3307414	1.605174		
90%	.6220512	1.609438	Variance	.2259646
95%	.8675008	1.673976	Skewness	-.1914731
99%	1.299283	1.686399	Kurtosis	4.572827

Some of the control variables (such as change in total assets, turnover, and liquidity ratio) on which the gps was generated are trimmed for extreme far out values using Tukey fences method.

It is important to note that using value of the NPL purchased services as an intervention variable did not yield reliable results while invoices from NPL sales proves difficult to model than the number of incidences. This is not surprising given that number of incidences is less affected by the size of the business. However, the incidences of support on employment growth impact violated normality assumption. Hence, the dose response function was estimated through a generalized linear model which addresses this problem.

The estimated dose-response (glmdose/doseresponse2) model allows six possible distribution functions: binomial, gamma, inverse gaussian, negative binomial, normal and Poisson coupled with admissible links. It also tests the balancing property of the general propensity score. The Figure below presents part of the dose response results and it is based on gamma distribution- a continuous probability distribution. Estimations with other discrete probability distribution functions (paying attention to the family-link) such as binomial, inverse gaussian and negative binomial produced less reliable results. This is not surprising given that employment growth (the outcome variable) could take any value in a continuous time.

The average of the number of times services are purchased during the period 2010-2015 is 2.5 (table G2). Belmana (2019) obtained the same number. The regularly supported business (firms having support in at least about 83% of the time between 2010 and 2015) represents only about 16.4% of all sampled firms. In other words, majority of the businesses have only between one (40.25% of the firms) and two (20.84% of the firms) incidents of using measurement services (table G3). These firms are referred to as sometimes supported businesses.

Table G2. Treatment Variable (incidences of NPL support)

x

	Percentiles	Smallest		
1%	1	1		
5%	1	1		
10%	1	1	Obs	1,113
25%	1	1	Sum of Wgt.	1,113
50%	2		Mean	2.513028
		Largest	Std. Dev.	1.685758
75%	4	6		
90%	6	6	Variance	2.841782
95%	6	6	Skewness	.8693656
99%	6	6	Kurtosis	2.471442

Table G3. Statistics of Treated firms

x	Freq.	Percent	Cum.
1	448	40.25	40.25
2	232	20.84	61.10
3	142	12.76	73.85
4	108	9.70	83.56
5	71	6.38	89.94
6	112	10.06	100.00
Total	1,113	100.00	

The dose response function in the left-hand panel of figure G1 presents the normalized employment growth whose distribution is presented in Table G4. However, this cannot be interpreted as employment growth. The employment growth is presented in Table G5.

Table G4. Normalized employment growth

	Percentiles	Smallest		
1%	.0764172	9.65e-08		
5%	.2629434	.0044577		
10%	.3585418	.0053641	Obs	1,591
25%	.4445204	.0146187	Sum of wgt.	1,591
50%	.5133412		Mean	.5158269
		Largest	Std. dev.	.1428818
75%	.5925197	.9755855		
90%	.6800809	.9768673	Variance	.0204152
95%	.7538576	.9962661	Skewness	-.1914731
99%	.8836416	1	Kurtosis	4.572827

Inferences drawn from left-hand panel of figure G1 are in three parts (Table G5):

- The employment growth of the top 10% of supported firms was around 5.9%. These are firms which use NPL very regularly.
- The employment growth of average users of NPL support services was around 1%. These firms could be categorised as mid-way to being regularly treated.
- The last category are those firms which rarely use NPL services, the bottom 10% of the treatment level. These firms did not experience positive employment growth.

The results show that employment growth exhibits diminishing returns as treatment level increases. This implies that low innovating firms characterised with large deficit in R&D activities grew faster with R&D intervention relative to high innovating firms characterised with high level of R&D activities.

The result of the treatment effect function (on the left-hand panel of figure G) shows that intervention gap is somewhat inversely related to the treatment level. This suggests that firms with low intervention gap (the regularly supported) have higher level of treatment than those with higher intervention gap.

Figure G1. Dose Response Model

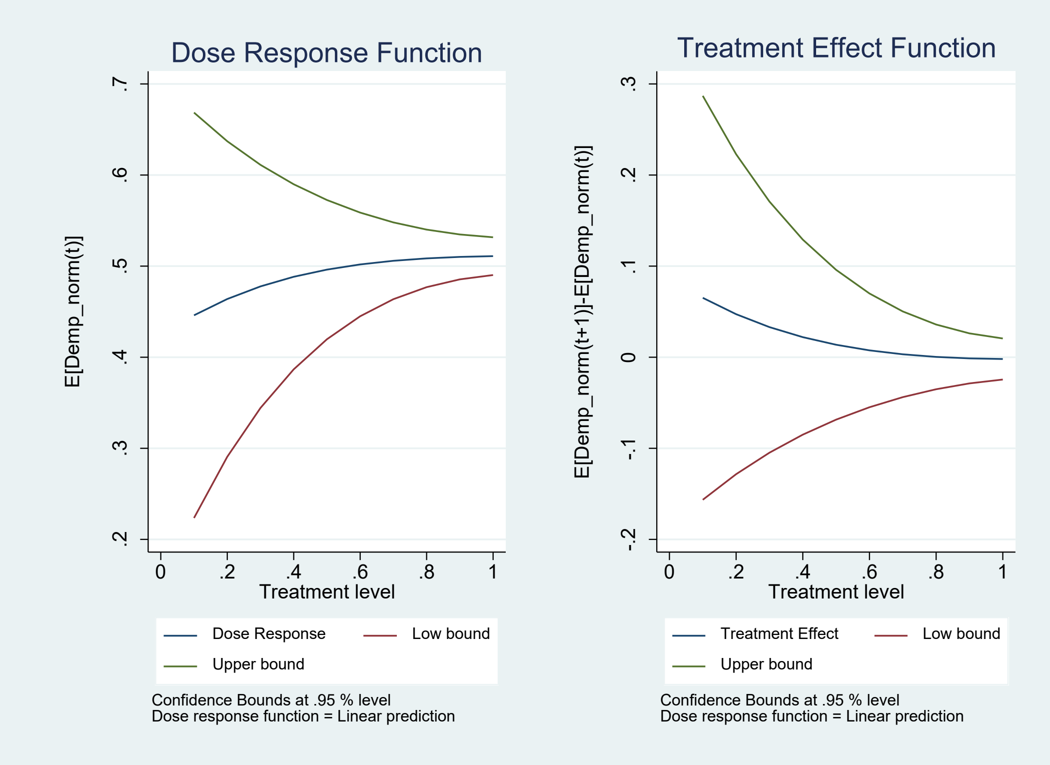


Table G5. Employment Growth Generated from Dose Response Model.

treatment_level	transformed employment growth (dose_response_g)	employment growth
0.1	0.446	-0.157
0.2	0.464	-0.097
0.3	0.478	-0.051
0.4	0.488	-0.016
0.5	0.496	0.010
0.6	0.502	0.029
0.7	0.506	0.042
0.8	0.509	0.051
0.9	0.510	0.057
1	0.511	0.059

Note: We retrieved employment growth using $x_i = x_i^I(X_{max} - X_{min}) + X_{min}$. Where x_i is employment growth to be determined, x_i^I is the transformed employment growth estimated by the dose response model (in Table G5), X_{max} is the maximum value of employment growth (picked from distribution of employment growth in Table G1), and X_{min} is the minimum value of employment growth (in Table G1).

In terms of caveats, the estimated dose-response model accessed the impact of NPL's continuous R&D supports on employment growth of the supported firms. However, this impact may vary between high and low skilled workers. The current data did not allow us to assess this in detail. Besides, the potential spill-overs effect of R&D spending is beyond the scope of this study. Hence, future studies may focus on how public R&D supports for innovation affect employment growth of high and low skilled workers in businesses, and possible spill-over impact of private R&D spending.

Table G6. Quadratic Regression of Outcome

```
The outcome variable ''Demp_norm'' is a continuous variable

The regression model is: Y = T + T^2 + GPS +  GPS^2 + T*GPS
```

Source	SS	df	MS
Model	.222376121	5	.044475224
Residual	7.71078314	454	.016984104
Total	7.93315926	459	.017283571

Number of obs	=	460
F(5, 454)	=	2.62
Prob > F	=	0.0238
R-squared	=	0.0280
Adj R-squared	=	0.0173
Root MSE	=	.13032

Demp_norm	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
x	.3173019	.1371068	2.31	0.021	.0478592	.5867447
x_sq	-.02957	.0118054	-2.50	0.013	-.0527701	-.0063699
pscore_g	4.170728	2.459217	1.70	0.091	-.6621328	9.00359
pscore_g_sq	-5.119762	3.769325	-1.36	0.175	-12.52725	2.287727
x_pscore_g	-.7475777	.4365076	-1.71	0.087	-1.605404	.1102483
_cons	-.3161319	.398001	-0.79	0.427	-1.098285	.4660209

Table G7. Estimated Generalized Propensity Score

Estimated generalized propensity score

	Percentiles	Smallest		
1%	.028787	.0177239		
5%	.0521443	.0223464		
10%	.056982	.0243191	Obs	472
25%	.0915573	.0246079	Sum of Wgt.	472
50%	.182607		Mean	.1856959
		Largest	Std. Dev.	.102622
75%	.3003088	.365084		
90%	.318459	.3660072	Variance	.0105313
95%	.3293158	.3663055	Skewness	.1706122
99%	.3650378	.3678795	Kurtosis	1.585173