Department for Business Innovation & Skills

RESEARCH REPORT

BIS RESEARCH REPORT NUMBER 177

Estimating the impact of publicly funded training on industry and firm-level outcomes

MAY 2016

Authors:
Mr Pietro Patrignani, London Economics
Dr Gavan Conlon, London Economics
Ms Maike Halterbeck, London Economics

We would like to thank a range of people at the Department for Business, Innovation and Skills who provided expert knowledge, guidance and support to London Economics throughout the project, often provided over and above their day-to-day activities. Specifically, we would like to thank Mr Matt Bursnall, Mr Peter Blythe, Mrs Karen Woolgar and Mr Steve Dempsey. All have been generous with their time, as well as thought provoking in the contributions. All are a tribute to the concept of public service. Despite this assistance, all errors and omissions remain the responsibility of London Economics.

Contents

Index of Tables and Figures

Executive Summary

London Economics were commissioned by the Department for Business Innovation and Skills to undertake an assessment of the impact of training on industry-level and firm-level productivity and wages.

In the **first part** of the report, using a range of information on entities engaged in publicly funded training, we undertook a matching exercise of a range of data sources including

- the **Individualised Learner Record (ILR)** (containing information on all publicly funded training in England), and the **Employer Data Service (EDS)** data (containing some basic information on those businesses with employees engaged in publicly funded training);
- the **Inter-Departmental Business Register (IDBR)**, which is the official record of all businesses incorporated in the United Kingdom, and contains information on a range of firm level characteristics (including financial information);
- the **Annual Business Survey (ABS)**, which is sampled from the IDBR and contains further information on a range of measures of interest in the econometric analysis [such as productivity]; and
- the **National Employer Skills Survey**, which is also sampled from the IDBR and contains information on the extent of any form of training undertaken by firms

Data matching exercise

Of around **500,000** unique business entities contained in the EDS that received publically funded training between 2009 and 2014, 70% were successfully matched to the IDBR using the sequential application of 20 different matching rules. By auditing a sample of 60 matches from each 'rule' we estimate that **99.7%** were 'good' matches, with only **0.3%** being mistakenly matched.

However, there are some significant limitations to the matched data set. Specifically, although the matched entities appear generally representative of all entities engaged in publicly funded training in terms of region of incorporation and industry, there are significant differences in the matching rates for some firm characteristics

- Only **26%** of "sole traders" could be matched to the IDBR, compared to between **61%** and **86%** for the other categories of legal status.
- Only **31%** of firms with 1 employee were matched to the IDBR, compared to between **69%** and **85%** for the other categories of firm size.

As such, caution needs to be exercised when undertaking or interpreting any analysis at the firm level. Because of the nature of the analysis, the accuracy of the matching process relies on the **precision** of the underlying data collected through EDS (which is gathered from a range of 3rd party commercial sources as well as the companies engaged in training themselves). For example, consistent collection of top-tier information, such as Company Registration Number (CRN), Parent CRN, *full* company name, *full* trading name *full* postcode will improve the ability to match EDS to IDBR data going forward.

Estimating the impact of training on industry and firm level outcomes

Two strands of econometric analysis were undertaken – the first to assess the impact of training on productivity and the wage bill controlling for both publicly funded training (derived from the matched ILR-EDS-IDBR data) and overall training intensity (derived from the ESS) at **industry-level**. The second strand of analysis replicated the industry-level approach at **firm-level** (although with substantial gaps in the available data).

At **industry-level**, the analysis suggests that raising overall training intensity by **1 percentage point** is associated with an increase on productivity of about **0.74%** and around **0.36%** on the wage. However, it was not possible to identify any unambiguous impact of publicly funded training leading for a formally recognised qualification on either the productivity or wage-bill measure. Although these estimates in relation to any form of training are in line with previous analyses using training intensity at the industry level (e.g. Dearden *et al* (2005)), unlike previous research, **the result is never statistically significant in the specification controlling for the endogeneity of training decisions**[1] (i.e. whether training leads to higher productivity or whether more productive firms are likely to engage more in training).

At **firm-level**, there are currently large evidence gaps in the data. Although estimated coefficients point towards a positive relationship between publicly funded training and firm level outcomes, the lack of firm-level data for key variables (i.e. capital stock and privately funded training); inconsistencies in relation to employment measures (i.e. individual enterprises having different measures of employment depending on the data source); as well as the length of the time series currently available (4 years), make the firm-level results **very speculative** at this stage. Furthermore, any robust analysis using panel data techniques can only be undertaken on firms repeatedly available over time in the Annual Business Survey (ruling out small and micro firms). As a result of these data limitations, the analysis does not indicate the absence of any impact of publicly funded training on productivity, but rather than we are unable to draw any strong conclusions due to deficiencies in the underpinning data.

Scope for future research and analysis

The evidence presented in the study should be seen as a stepping stone for future research and analysis, in particular we recommend that:

- Identifying the firms undertaking ILR training in the IDBR becomes a regular matching exercise to be repeated when a new ILR edition becomes available (with potential further refinement of the matching strategy);
- The Department interacts with the ONS and other data depository to ensure that the data limitations currently affecting the firm-level analysis are overcome;
- The analysis of the impact of publicly funded training on productivity at the firm level (and possibly also at the industry level) is repeated in 36-48 months (potentially) when the data limitations are resolved and a longer time series become available;

[1] The industry-level results presented in this report were estimated using *Random Effects*, *Fixed Effects* and *System GMM*. Results from the first two models are statistically significant, while results from GMM are not. In the Dearden *et al*, (2005) analysis, results are statistically significant across all different models used.

Introduction

London Economics were commissioned by the Department for Business Innovation and Skills to undertake an assessment of publicly funded training on industry-level and firm-level outcomes (such as productivity and salary bill).

In the first part of the report, we present an in-depth assessment of the various data sources used as part of this analysis (i.e. the **Individualised Learner Record (ILR)**, **Employer Data Service (EDS)**, the **Inter-Departmental Business Register (IDBR)**, the **Annual Business Survey (ABS)**, and the **National Employer Skills Survey**). Furthermore, we provide a detailed assessment of the data matching process undertaken and an assessment of the quality of the match undertaken.

In the second part of the report, we provide a comprehensive description of the most important variables relating to measures of training generated from the different data sources and the measures of firm-level and industry-level outcomes (labour productivity and salary bill); a detailed presentation of the econometric models under consideration; as well as the findings from the econometric models.

In the final section of the report, we present our overarching conclusions, as well as recommendations in relation to the feasibility of undertaking alternative strands of research; the issues that might limit future options, as well as the timing of undertaking any repeat analyses.

Main data and matching approach

The Individualised Learner Record (ILR)

The **Individualised Learner Record (ILR)** contains detailed information on course and personal characteristics for publicly funded Further Education (FE) courses and is supplied by learning providers throughout the Further Education system. The ILR is organised by academic year (1st August – 31st July) and the data specification may vary to some extent from year to year. ILR data is collected from providers that are in receipt of funding from the Skills Funding Agency (SFA), the Education Funding Agency (EFA) and co-financed European Social Funds (ESF). For training undertaken through the employer, an employer identifier (**A44**) is attached to the dataset in order to identify the organisation engaging in publicly funded training. The current study focuses **on publicly funded training undertaken through the employer only** (including Apprenticeships) and uses ILR data from the academic years 2010/11 to 2013/14.

Employer Data Service (EDS) – "Blue Sheep" data

The **Employer Data Service (EDS)** (or 'Blue Sheep') data is a composite database containing information on the characteristics of UK firms (or sites within firms, hence 'business entities')[2]. The constituent information is collected from a range of different sources (e.g. Thomson Directories, Companies House, Dun and Bradstreet etc). However, in the case of evidence gaps, the EDS data may also contain information and data that is directly sourced from the EDS Help Desk (which could be potentially provided by the particular entities engaged in training activities). The EDS contains information on entity characteristics including:

- Company and trading name,
- Entity's postcode,
- Number of employees,
- Turnover at site and group level,
- Sector of activity as defined by SIC code,
- Year of foundation,
- Company Registration Number (CRN; where available), and
- A range of other entity-level characteristics.

The version of the EDS data extract received by London Economics contained information on around **18.6 million entities** (not all of them will identify live entities), with Company Registration Number available for around **13 million entities**. Employment data for entities is reported at the **site** and **group** level in the EDS; however, neither may necessarily correspond to the classification of **local unit** or **enterprise unit** strata used in the IDBR (see subsequent section for a description of the IDBR).

[2] In general, to distinguish EDS entities from the various IBDR units (enterprises, local units etc.), we refer to **"entities"** throughout the document **to identify EDS 'firms',** and refer to **"enterprises"** (or other relevant aggregation) when discussing **IDBR units**. When comparing information for matched entities across data sources, IDBR data may refer to EDS entities (labelled "IDBR – entity level"), which involves duplication on the IDBR side, as multiple EDS entities may correspond to the same enterprise or to the enterprise unit (labelled "IDBR – enterprise level"), with no duplication on the IDBR side.

The ILR data on training can be linked to the EDS using the **A44** variable (corresponding to the EDS identifier) to identify companies undertaking publicly funded training in the period 2010/11-2013/14: in total there were **533,033 distinct entities** (as identified by the EDS) **engaging in publicly funded training** in the period considered.

The Inter-Departmental Business Register (IDBR)

The IDBR is the comprehensive list of UK businesses that is used by Government for statistical purposes and provides the main sampling frame for surveys of businesses carried out by the Office for National Statistics and other Government departments. It is also a key data source for analyses of business activity. The IDBR covers over **2.5 million live enterprises** in all sectors of the UK economy, other than some very small businesses (those without employees, and with turnover below the relevant tax threshold) and some non-profit making organisations. The information used in the IDBR is obtained from the following main sources:

- HMRC VAT - Traders registered for VAT purposes with HMRC;
- HMRC PAYE - Employers operating a PAYE scheme, registered with the HMRC;
- Companies House - Incorporated businesses registered at Companies House;
- Dun and Bradstreet for Enterprise Group information;
- The Business Register and Employment Survey (BRES) and other ONS surveys.

Information contained in the IDBR covers:

- Company name and trading name;
- Address including postcode;
- Unit (e.g. enterprise) 'birth' date;
- Unit (e.g. enterprise) 'death' date;
- Standard Industrial Classification (UK SIC 2007 and UK SIC 2003);
- Employment and employees (updated from administrative sources (PAYE and VAT records) and ONS Surveys (Business Register Employment Survey));
- Turnover (updated via administrative sources (HMRC VAT and PAYE records) and ONS Business Surveys (Annual Business Survey));
- Legal status (company, sole proprietor, partnership, public corporation/nationalised body, Local Authority or non-profit body);
- Enterprise group links;
- Country of ownership;
- Company Registration Number.

The IDBR is organised in different datasets, and information is updated each quarter. The **enterprise unit** is at the centre of the IDBR classification, and all other units (i.e. local unit) can be linked to the enterprise. Using this database structure (see Figure 1), it is possible to distinguish between:

- Statistical units (Enterprise, Local Units and Enterprise Group);

- Administrative units (VAT and PAYE units also containing the Company Registration Number);
- Observation unit (Reporting Unit);

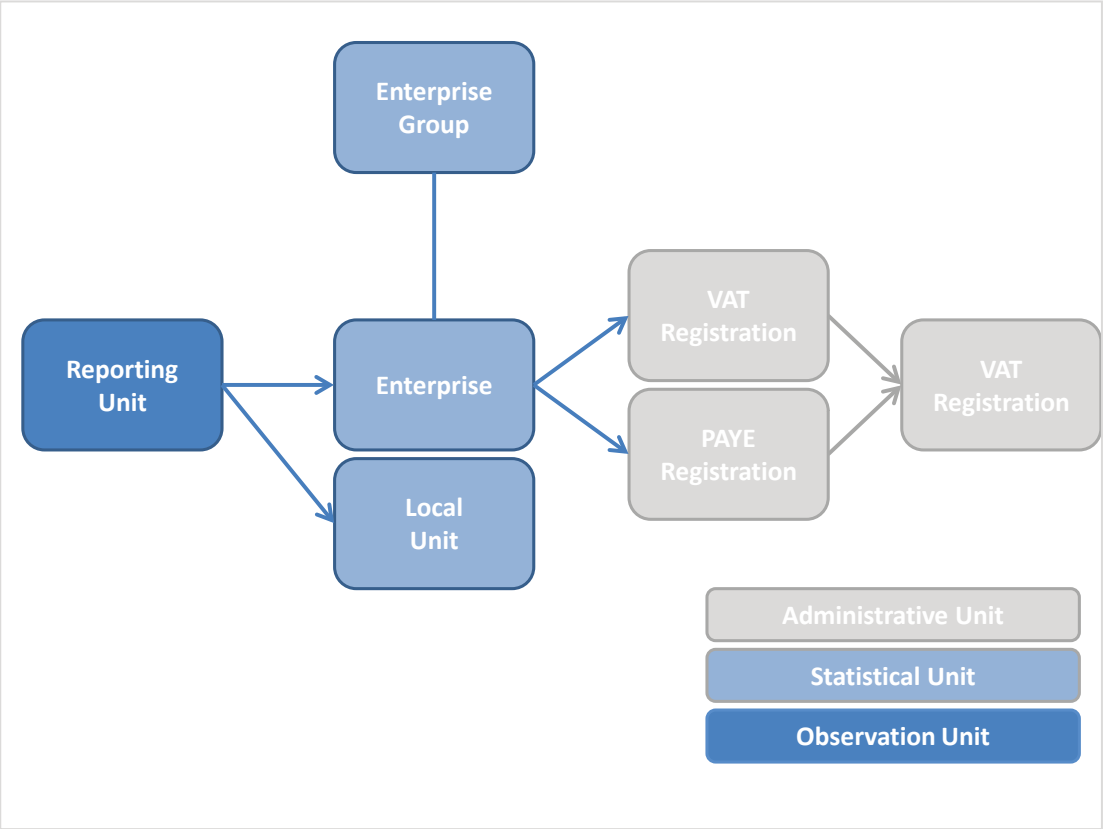


Figure 1: IDBR structure
Source: *http://www.ons.gov.uk/ons/about-ons/products-and-services/idbr/index.html*

In practice, the IDBR is organised in more than 40 files, reporting information on the different units, with a link between each unit level and the enterprise. Detailed information on the geographical location, industry and a variety of other variables is also reported. **For the matching exercise, we received and made use of all quarterly IDBR extracts between December 2007 and September 2014 inclusive**.

One of the key pieces of information contained within the IDBR are the **Company Registration Numbers** (CRN), which are available in the IDBR for VAT and PAYE units and can be linked to the enterprise level. Each CRN should correspond to one enterprise; however, the same enterprise can cover more than one CRN.

The IDBR also contains an **"ADDRESS" file** with unit name, full address, postcodes and other details at the different unit levels (around **15 million** entries in total). The "ADDRESS" file does not specify whether the address corresponds to an enterprise, local unit etc.; however, this information can be retrieved from data available in other files (enterprise, local unit etc.) by matching on the "**addressref**" identifier. All ONS surveys can be linked to the IDBR given the presence of the enterprise/local unit/reporting unit identifiers.

Matching strategy

The matching strategy was based on developing an approach to identify those entities engaging in publicly funded training (as defined by the EDS) within the IDBR (at the enterprise and local level). The matching approach may be divided in three different macro stages:

1) **Matching using Company Registration Number** and parent/ultimate CRN (CRN is available for around 30% of the EDS/ILR entities and in the IDBR);
2) **Matching using company/trading name and postcode** (using information on company details available in the EDS dataset and the IDBR "ADDRESS" file). This stage consisted of several steps and was undertaken after standardising company names in both datasets.
3) **Fuzzy matching using entity characteristics** including combinations of
   a) company name[3]/ trading name (with and without vowels)
   b) first seven letters of company name/trading name
   c) First word of company name/trading name
   d) Full postcode
   e) Postcode District
   f) Postcode Area
   g) SIC Code (2003) and/or SIC Code (2007)
   h) EDS ID[4]
   i) Name similarity[5]
   j) Third or fourth word of company name[6]
   k) Address similarity[7]
   l) Legal status[8]

[3] Match based on company name only when only one entref is available in the IDBR (or one entref covering at least 80% of cases) - and company is not a partnership or unincorporated business (based on the definition in the EDS).
[4] Match based on EDS entries sharing the EDS ID on ultimate owner
[5] The tolerance (i.e. Number of different characters) was based on the overall number of characters in the name - no differences up to 7 characters, 1 digit for strings of 8-10 characters, 2 digits for strings of 11-14 characters and 3 digits for longer strings (string length is always computed on the lower string)
[6] Whether at least one other word (excluding the first two) matched across the two strings
[7] Address similarity is based on building number and whether the other strings of the EDS address (sub_building_name, building_name, primary_thoroughfare_name) match the IDBR address.
[8] Legal status checks exclude from the match any EDS entities matched to IDBR entities classified as "Sole Proprietors" or "Partnership".

Table 1: Matching approach applied to ILR-EDS and IDBR

	CRN	Parent CRN	Ultimate CRN	Full company name	Full trading name	1st 7 letters of co. name	1st 7 letters of trading name	Co. name (no vowels)	1st word of co. name	1st word of trading name	Full postcode	Postcode district	Postcode area	SIC code 2007 (2 digit)	SIC code 2003 (2 digit)	EDS ID	Name similarity	SIC code	3rd and 4th words co. name	Address similarity	Legal Status
																	Additional Controls				
Stage 1	■																				
Stage 1a		■																			
Stage 1b			■																		
Stage 2				■							■										
Stage 3					■						■										
Stage 4				■								■									
Stage 5					■							■									
Stage 6						■					■			■							
Stage 7						■					■						■	■	■	■	
Stage 8											■						■	■	■	■	
Stage 9								■			■										
Stage 10									■		■						■		■	■	
Stage 11						■						■					■		■		
Stage 12												■								■	
Stage 13								■				■					■				■
Stage 14									■			■					■	■	■	■	
Stage 15												■					■	■	■	■	
Stage 16				■									■								■
Stage 17													■								■
Stage 18				■																	■

Note: Stage 1, Stage 16 and Stage 17 are based on Company Registration Number and were undertaken before all other matching stages

Quality Assurance: how good was the matching process?

Having implemented the various matching stages, Table 2 below illustrates the number of matches that were achieved in each stage, as well as the proportion of the EDS entities that were matched with the IDBR. The analysis indicates that of the **533,033** entities, **403,493** were matched (**75.7%**) with **129,543** unmatched (**24.3%**) in this first stage.

Table 2: Assessment of matching progress by stage

Stage	Number of Obs.	% Matched	Manual Check			Match rate (Correct + Probably/ All)	Good quality matches	Poor quality matches
			Correct	Probably	Incorrect			
Stage 1	161,427	30.3%	60	0	0	100.0%	30.3%	0.0%
Stage 1a	1,998	0.4%	60	0	0	100.0%	0.4%	0.0%
Stage 1b	12	0.0%	60	0	0	100.0%	0.0%	0.0%
Stage 2	91,967	17.3%	60	0	0	100.0%	17.3%	0.0%
Stage 3	45,529	8.5%	60	0	0	100.0%	8.5%	0.0%
Stage 4	11,522	2.2%	55	2	3	95.0%	2.1%	0.1%
Stage 5	4,369	0.8%	50	7	3	95.0%	0.8%	0.0%
Stage 6	32,769	6.1%	179	1	0	100.0%	6.1%	0.0%
Stage 7	14,239	2.7%	50	2	8	86.7%	2.3%	0.4%
Stage 8	5,457	1.0%	56	3	1	98.3%	1.0%	0.0%
Stage 9	583	0.1%	59	1	0	100.0%	0.1%	0.0%
Stage 10	11,533	2.2%	115	4	1	99.2%	2.1%	0.0%
Stage 11	3,831	0.7%	43	10	7	88.3%	0.6%	0.1%
Stage 12	1,934	0.4%	28	19	13	78.3%	0.3%	0.1%
Stage 13	225	0.0%	30	19	11	81.7%	0.0%	0.0%
Stage 14	882	0.2%	22	14	24	60.0%	0.1%	0.1%
Stage 15	751	0.1%	17	22	21	65.0%	0.1%	0.0%
Stage 16	7,183	1.3%	35	20	5	91.7%	1.2%	0.1%
Stage 17	2,119	0.4%	15	36	9	85.0%	0.3%	0.1%
Stage 18	5,163	1.0%	15	37	8	86.7%	0.8%	0.1%
Results from Manual check			***1,021***	***197***	***114***			
			76.7%	14.8%	8.6%			
Unmatched	**129,540**	**24.3%**						
Matched	**403,493**	**75.7%**	**~397,485**		**~6,008**		**74.6%**	**1.1%**
Total	**533,033**	**100.0%**						

Note: Stage 1, Stage 1a and Stage 1b are based on Company Registration Number and were undertaken before all other stages

However, to understand the quality of the matching process (other than those where a match was achieved on CRN), 60 matched cases were selected at random from each of the stages (and up to 180 in some stages) and a manual check of the accuracy of the match was undertaken. The manual check suggests that approximately **397,485** of these matches are 'good' matches (**98.5%**), with **6,008** being mistakenly matched (**1.5%**).

In more detail, the analysis indicates that the initial stages based on firm characteristics including name and detailed postal address were assessed to be of generally high quality, however, the quality of the match deteriorated in later stages of the process. Specifically, in the later stages, when information relating to postcode *district* was used instead of full postcode, and the first seven letters of an entities name (or trading name) was used

(instead of full company name), even with the inclusion of additional checks, the manual check indicated that the level of 'good' matches declined significantly. Specifically, compared to a **100%** match on company or trading name and full postcode (stage 2 or 3) or a **95%** match rate on full company name and postcode district (stage 4), the match rate in stage 7 was assessed to be **86.7%** (first seven letters of entity name, full postcode and additional checks for name and address similarity, SIC code and 3rd or 4th words of company name if they exist). Specifically, this meant that of the 60 proposed matches, 50 were considered 'good', 2 were plausible, but 8 were likely to be incorrect.

Data set adopted for analysis

Given this, it was decided to omit the matches based on the following stages of the process: Stages 7, 11, 12, 13, 14, 15, 16, 17 and 18. As a result, the total number of EDS entities matched with the IDBR stood at **367,166** corresponding to **68.9%** of all EDS entities engaged in publicly funded training. In terms of this data set, although the number of matches was **36,327** fewer that the 'fully' matched data set, the manual check suggests that approximately **366,184** of these matches are 'good' matches (**99.7%**), with only **982** being mistakenly matched (**0.3%**).

Table 3: Final matched data adopted for econometric analysis

Stage	Number of Obs.	% Matched	Manual Check			Match rate	Good quality matches	Poor quality matches
			Correct	Probably	Incorrect			
Stage 1	161,427	30.3%	60	0	0	100.0%	30.3%	0.0%
Stage 1a	1,988	0.4%	60	0	0	100.0%	0.4%	0.0%
Stage 1b	12	0.0%	60	0	0	100.0%	0.0%	0.0%
Stage 2	91,967	17.3%	60	0	0	100.0%	17.3%	0.0%
Stage 3	45,529	8.5%	60	0	0	100.0%	8.5%	0.0%
Stage 4	11,522	2.2%	55	2	3	95.0%	2.1%	0.1%
Stage 5	4,369	0.8%	50	7	3	95.0%	0.8%	0.0%
Stage 6	32,769	6.1%	179	1	0	100.0%	6.1%	0.0%
Stage 7	14,239	2.7%	50	2	8	86.7%	2.3%	0.4%
Stage 8	5,457	1.0%	56	3	1	98.3%	1.0%	0.0%
Stage 9	583	0.1%	59	1	0	100.0%	0.1%	0.0%
Stage 10	11,533	2.2%	115	4	1	99.2%	2.1%	0.0%
Stage 11	3,831	0.7%	43	10	7	88.3%	0.6%	0.1%
Stage 12	1,934	0.4%	28	19	13	78.3%	0.3%	0.1%
Stage 13	225	0.0%	30	19	11	81.7%	0.0%	0.0%
Stage 14	882	0.2%	22	14	24	60.0%	0.1%	0.1%
Stage 15	751	0.1%	17	22	21	65.0%	0.1%	0.0%
Stage 16	7,183	1.3%	35	20	5	91.7%	1.2%	0.1%
Stage 17	2,119	0.4%	15	36	9	85.0%	0.3%	0.1%
Stage 18	5,163	1.0%	15	37	8	86.7%	0.8%	0.1%
Results from Manual check			***766***	***18***	***8***			
			96.7%	2.3%	1.0%			
Unmatched	**165,867**	**31.1%**						
Matched	**367,166**	**68.9%**	**~366,184**		**~982**		**68.7%**	**0.3%**[9]
Total	**533,033**	**100.0%**						

Note: Stage 1, Stage 1a and Stage 1b are based on Company Registration Number and were undertaken before all other stages

[9] The 1% figure from the manual check relates to the proportion of the matches that were assessed as being inaccurate. However, this is unweighted (each stage accounts for a different number of matches). Most of these inaccuracies are in the later stages – where there are fewer matches – so when we weight for the number of matches in each stage this proportion drops to 0.3%

Limitations of the match and associated caveats

In Figure 2 to Figure 5, we present the comparison of the matching rate according to a number of firm level characteristics (using information from the EDS). In relation to region of incorporation, the analysis indicates that the matching rate is broadly comparable – with the exception of London, where only **58%** of EDS entities are matched (compared to **69%** overall). Similarly, in relation to industrial classification, there is reasonable comparability in general; however, the proportion of firms in Transportation and Storage (**43%**) and the Construction industries (**62%**) is significantly lower than the average.

Figure 2: Matching rate between EDS and IDBR by region

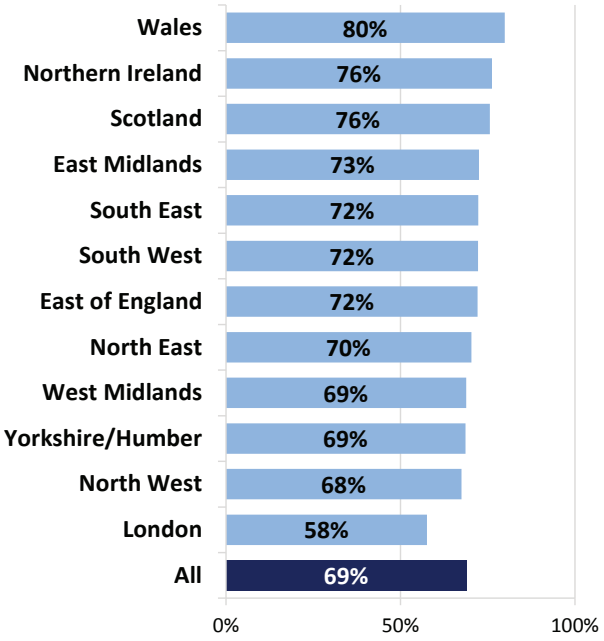


Figure 3: Matching rate between EDS and IDBR by SIC

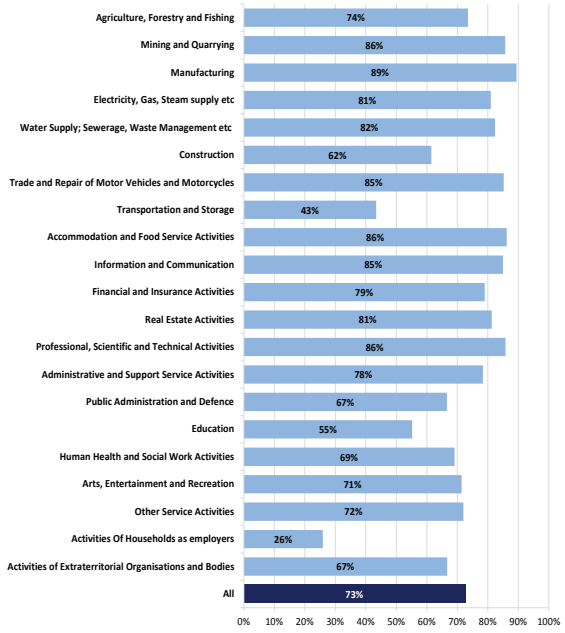


Figure 4: Matching rate between EDS and IDBR by legal status

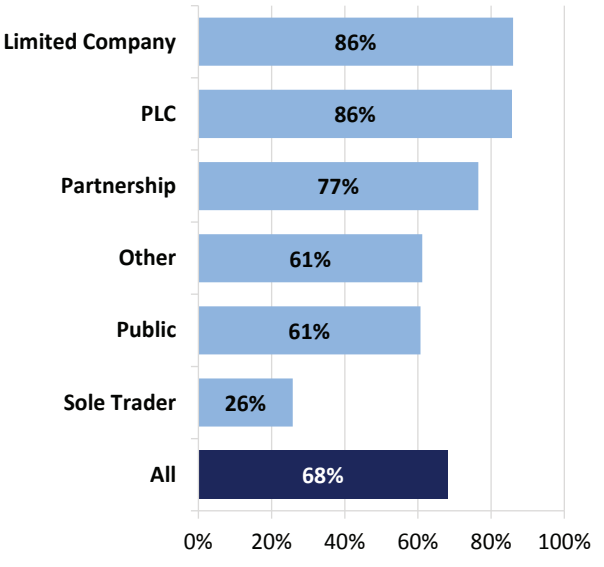
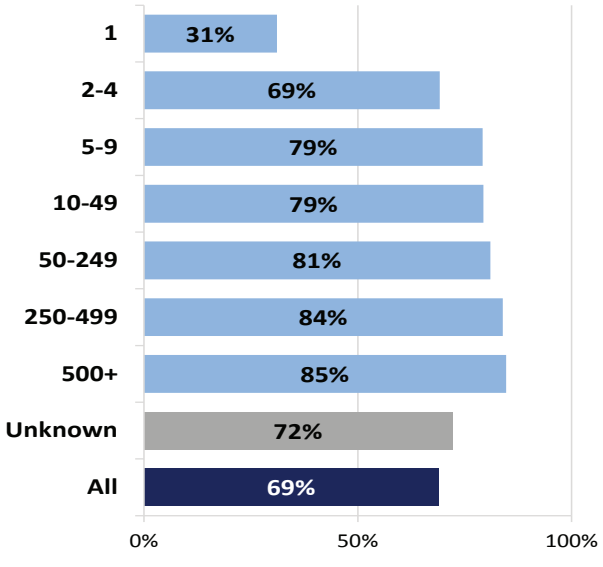


Figure 5: Matching rate between EDS and IDBR by employment (site employees)



London Economics' analysis of ILR-EDS-IDBR data. Note: The total match rate for all EDS entities was 69%. The "All" category in the Figures above (dark blue bars) reflect the total for entries with non-missing values for the relevant variable and may therefore differ from the overall match rate

However, fundamental differences in matching rates are apparent when considering the legal status of the entity and the level of employment (either site-level or group-level). Specifically, compared to approximately **86%** of Public or Private Limited Companies that were matched, the proportion of sole traders that were matched was just **26%**. Similarly, compared to a match rate of **69%** for those entities employing between just 2 and 4 individuals (at site level), the matching rate associated with entities where a single person might be employed at the site stood at just **31%**.

Given these fundamental differences in the match rate, and the particular firm level characteristics that are systematically associated with lower match rates, some caution needs to be exercised when undertaking or interpreting any analysis at the firm level. **Specifically, any firm-level analysis would need to explicitly acknowledge the issues relating to sample composition and the (likely) omission of small firms, or explicitly restrict any statistical or econometric analysis to larger firms**.

Potential for further refinement of the matching strategy

Using the sample of entries that were manually reviewed, of the **36,327** matches that were omitted from the final data set used in the econometric analysis, approximately **5,026** (**13.8%**) were assessed to be of poor quality, although **31,301** (**86.2%**) were likely to be of good quality. Below we outline a series of steps that may be considered for future refinements of the matching strategy

- Stage 7 (based on the first seven letters of company name and full postcode) and Stage 11 (based on the first seven letters of company name and postcode district) are likely to produce a high proportion of valid matches with small adjustments to the matching strategy and may therefore be explored further[10];
- Further steps in the matching process may potentially use the postcode and SIC code with checks on address and company name to identify valid matches[11];
- Another potential route to identify valid matches is to match based on the reverse company name (for example first seven letters of the reverse name) and postcode and then perform the usual checks based on other characteristics[12];

Finally, after these tweaks to the matching process based on the existing strategy, it may be informative to undertake a manual review of a random sample of unmatched cases (maybe around 1,000 EDS entries) to assess whether EDS entries are left unmatched due to differences in the information reported in the two data sources (EDS and IDBR) when a valid match exists, or because there are no seemingly valid matches in the IDBR (for example for small businesses under the VAT threshold).

[10] In particular there are probably around 10,000 valid matches that could be retained from stage 7 and 3,000 valid matches from stage 11, with a margin for error around 5%.
[11] An initial exploration of this approach produced around 9,000 valid matched with a margin for error between 0% and 7% (according to the SIC code definition used).
[12] An initial exploration of this step produced around 4,000 valid matches

Description of training variables

Training measures

Individualised Learner Record

The training measure used to assess the proportion of workers in receipt of publicly funded training leading to a formally recognised qualification (see Annex 2 for full description of steps undertaken to clean and code the data) was aggregated from the individual level (i.e. a binary variable in each calendar year) to the enterprise level (number of employees undertaking publicly funded training during the calendar year). Using information from the IDBR on SIC code, the information at the enterprise level was aggregated at the 2-digit SIC code level (with some further aggregation when needed) for the years 2010-13.

The proportion of workers engaged in publicly funded training in each sector was then generated using the total number of individuals receiving training and total number of people employed in each sector (from the Labour Force Survey) .

We also constructed two additional variables using information from the ILR and the IDBR:

- A variable capturing the **average number of days spent in training** (or training span for those in receipt of training) for each calendar year (as recorded by the ILR, i.e. notional number of days including also weekends);
- **The proportion of *enterprises* engaging in publicly funded training** in each sector

UKCES Employer Skills Survey

The UKCES Employer Skills Survey (ESS) collects information from employers on skills, training and other variables. The survey is organised at the establishment level. Information from the ESS in 2009, 2011 and 2013 was used and aggregated at the industry level for the following variables:

- **ESS training intensity**, defined as the proportion of staff trained over past 12 months, excluding the volume of Health and Safety and Induction training;
- **Average number of days of training per employee receiving training**, capturing the average length of (actual) training received per trainee.
- **Proportion of ESS establishments undertaking some form of training**, excluding establishments undertaking solely Health and Safety and Induction training;

Given that the ESS is only run every other year, for the subsequent econometric analysis, we imputed the values for 2010 and 2012 using the average of adjacent periods (t-1 and t+1) to generate a consistent time series.

ONS Labour Force Survey

The Labour Force Survey is a survey of the labour force organised at the individual (rather than employer) level. One of the key training variables in the LFS is whether the individual has undertaken any job-related training or education over the past four weeks. An industry

level measure representing the **proportion of employees undertaking job-related training in the last four weeks** was generated for the years 2009-13.

Relationship between training and industry level characteristics using alternative data sources

Publicly funded training at industry level

Using the final ILR-EDS-IDBR dataset described previously, in Figure 6, we present information on the proportion of workers in receipt of publicly funded training. The analysis illustrates that although on average approximately **4.8%** of workers are in receipt of publicly funded training, there is significant variation around the mean. In general, across manufacturing industries, the average proportion of workers in receipt of training is approximately **3.2%**; however, there are some notable exceptions to this - with approximately **7.1%** and **6.6%** of workers employed in food manufacturing and metal fabrication industries (respectively) in receipt of publicly funded training. At the other end of the spectrum, in industries related to **Employment Activities**, **Office Administration**, **Education**, **Residential Care Activities** and **Other Personal Service Activities**, the proportion of workers in receipt of publicly funded training exceeds **10%**.

Measures of training across different data sources (ILR and ESS)

In Figure 7, we provide a range of different estimates of the incidence of training alternative data sources. In particular, we provide estimates of the proportion of workers in receipt of any form of training (both public and privately funded) using the Labour Force Survey between 2009 and 2013, as well as the proportion of workers in receipt of any form of training (both publicly and privately funded excluding the volume of Health & Safety and Induction training) from the Employers Skills Survey (between 2009 and 2013). Clearly, these two alternative measures of training are wider in definition than the more restrictive definition of publicly funded training leading to a formally recognised qualification from the ILR.

The analysis again illustrates the fact that on average, approximately **10.2%** of workers were in receipt of any of training in the previous 4 weeks (ranging from **3.6%** in the **Manufacture of textiles** to **22.4%** in **Residential Care activities**). Similarly, in relation to the ESS, the average proportion of workers in receipt of training in the previous 12 months stood at **36%**, ranging from **18.3%** in the **Manufacture of leather related goods** to **57.1%** in the **Employment Activities** industry.

In Figure 8, we present (again) information on the proportion of workers in receipt of public funded training leading to a formally recognised qualification, as well as the proportion of enterprises offering such training. The analysis indicates that approximately **6.5%** of enterprises offer or facilitate publicly funded training. Again, there is substantial variation in the incidence of enterprises offering publicly funded training, with almost **50%** of enterprises in the **Residential Care** activities sector offering workers training compared less than **1%** in industries engaged in **Film, Television and Sound Recording** or **Creative Arts and Entertainment.**

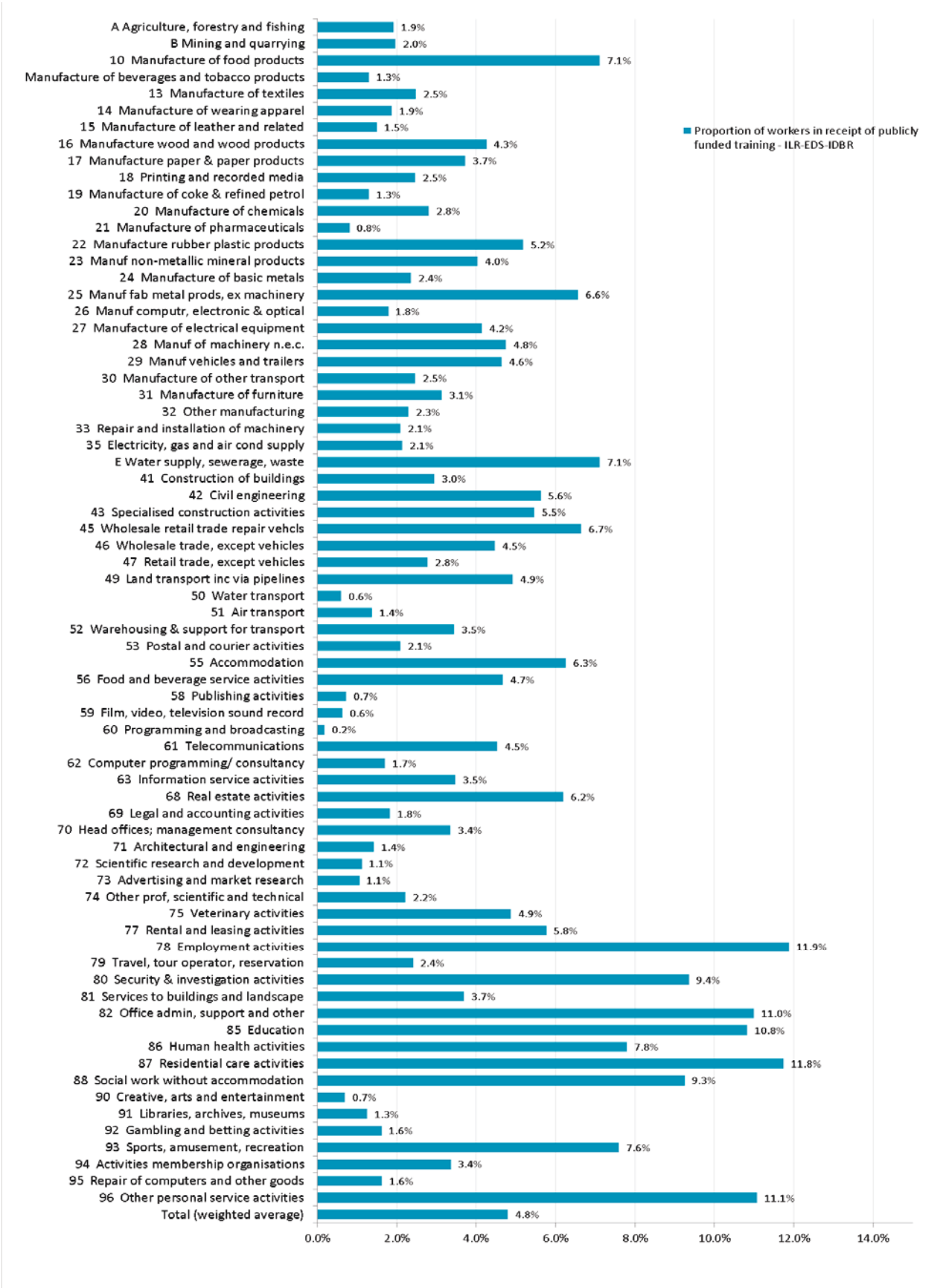


Figure 6: Proportion of employees in receipt of publicly funded training by industry

Note: The figures refer only to SIC codes covered in the ABS.

Source: London Economics based on ILR, EDS, IDBR and LFS

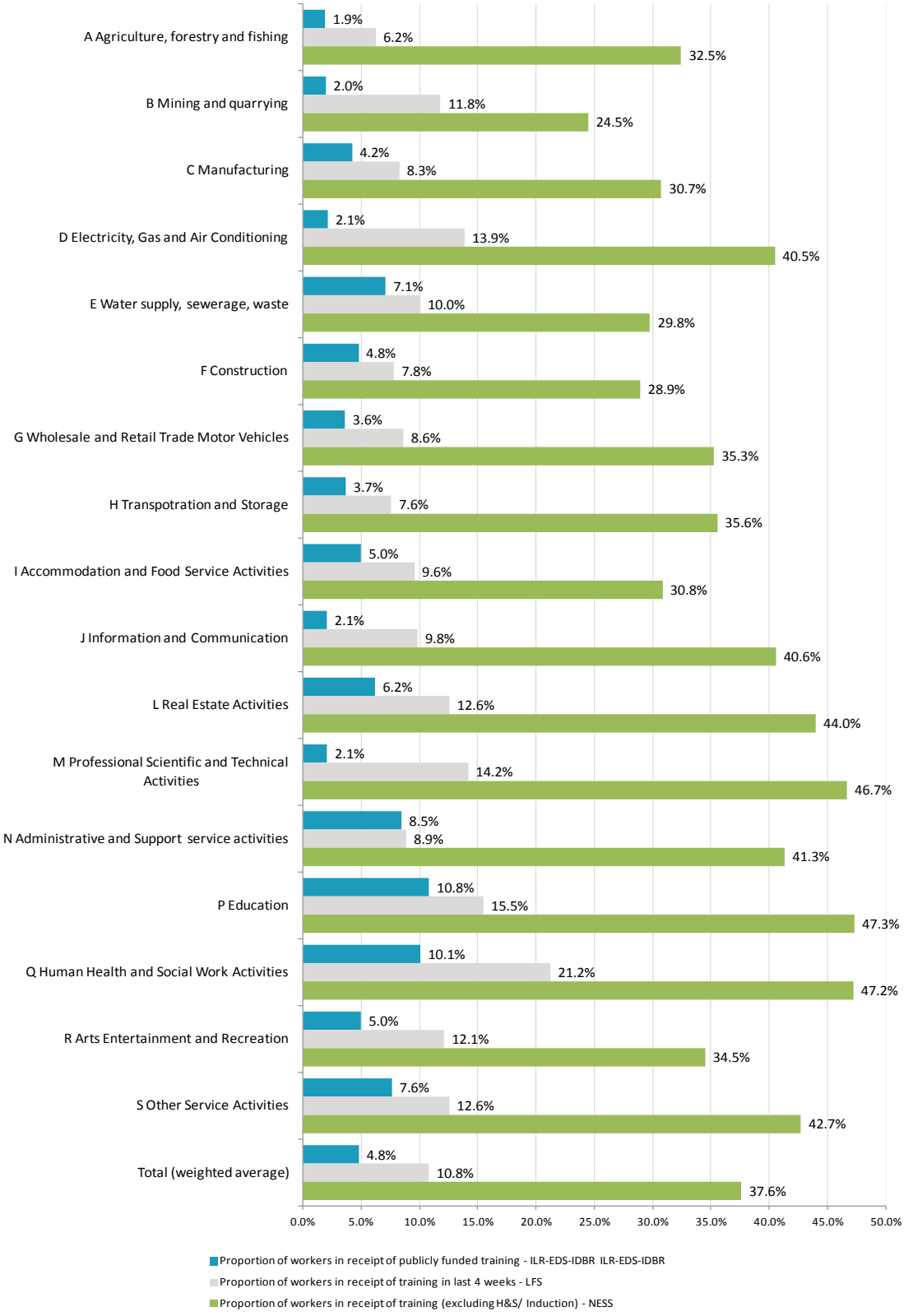


Figure 7: Proportion of employees in receipt of different types of training by industry

Note: Publicly funded training is based on the ILR definition (for entities matched to the IDBR); the LFS measure of training reports job-related training undertaken over the last four weeks; the ESS measure of training is total training undertaken excluding the volume of Health and Safety and Induction training. The figures refer only to SIC codes covered in the ABS.

Source: London Economics based on ILR, EDS, IDBR, LFS and ESS

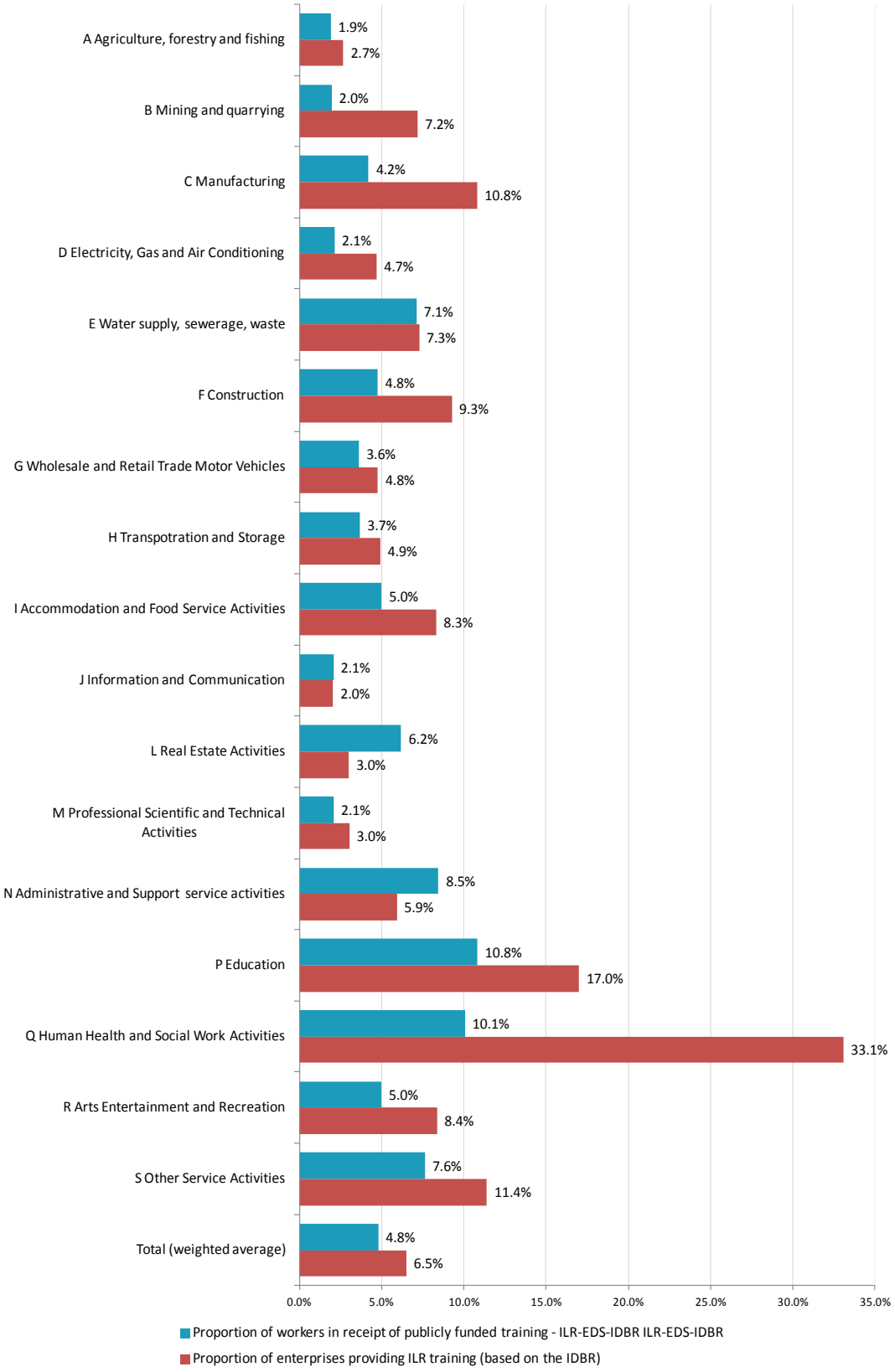


Figure 8: Proportion of employees and enterprises engaging in publicly funded training by industry

Note: Pooled 2010-13. Proportion of employees undertaking publicly funded training (ILR) and proportion of enterprises with at least one employee undertaking ILR training. The figures refer only to SIC codes covered in the ABS. ***Source: London Economics based on ILR, EDS, IDBR and LFS***

Relationship between different training variables

In Figure 9, we have also presented information on the proportion of workers in receipt of publicly funded training leading to a formally recognised qualification (from the ILR) against the proportion of workers in receipt of *any* form of training (excluding H&S and induction training) in the previous 12 months from the ESS. Furthermore, we have also classified industries predominantly engaged in **Utilities** in **Manufacturing**, **Construction**, **Transport-related** and **Storage/ distribution** industries using red circles. The blue diamonds represent more service orientated industries including **Administrative and Support services**; **Public Administration**; **Education**; **Human Health and Social Work**; **Arts, Entertainment and Recreation**, and **Other service activities**.

The analysis *broadly* suggests that within the more service orientated sectors of the economy, there is a *positive* relationship between the proportion of workers receiving publicly funded training (ILR) and the proportion of workers in receipt of any form of training (ESS), whilst the relationship is more ambiguous in the more manufacturing orientated sectors of the economy. Although there appears to be a negative relationship between the two measures of training, this is at least in part driven by two outliers (**Manufacture of Pharmaceuticals** and **Air Transport**). Removal of these outliers reverses the relationship, with a slight positive correlation between the training measures.

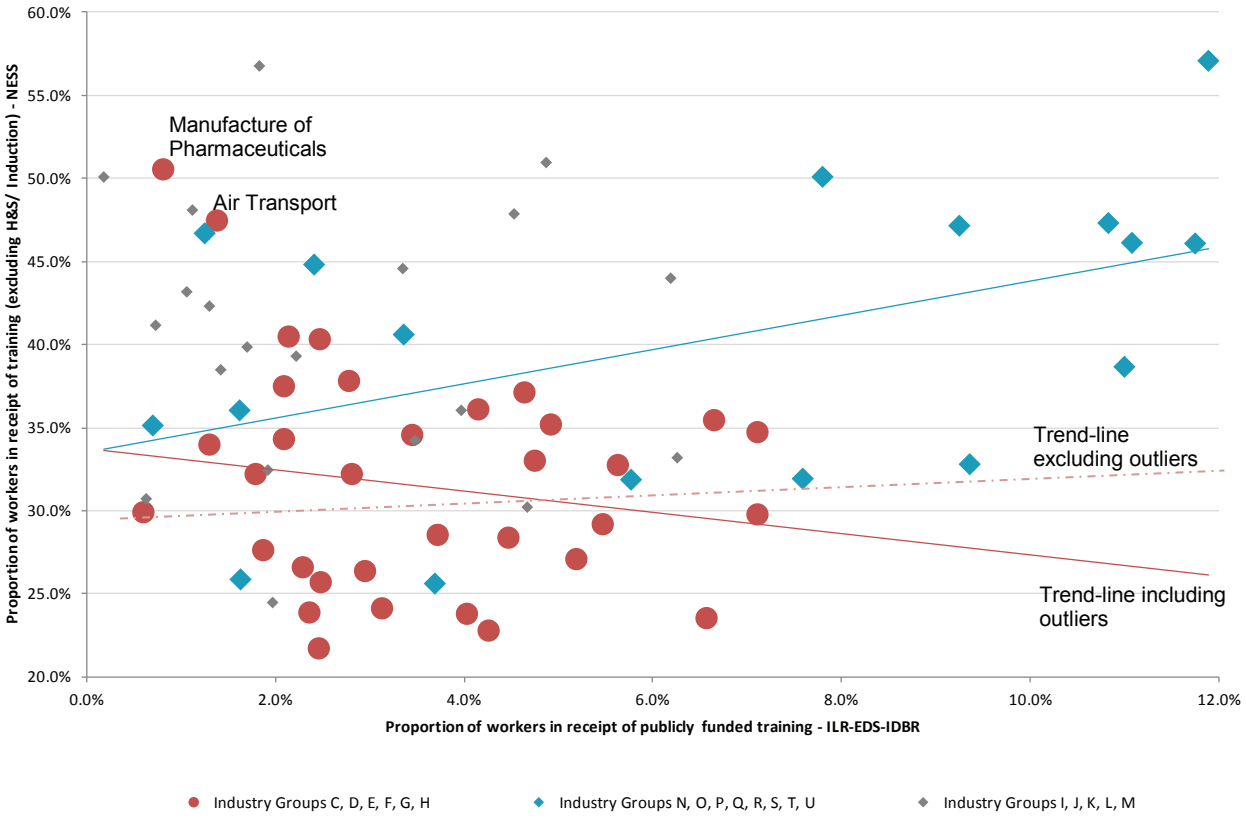


Figure 9: Proportion of workers in receipt of publicly funded training (ILR) against proportion of workers in receipt of any training (ESS) by industry

Note: Publicly funded training is based on the ILR definition (for entities matched to the IDBR); the ESS measure of training is total training undertaken excluding the volume of Health and Safety and Induction training

Source: London Economics based on ILR, EDS, IDBR, LFS and ESS

A comparison between publicly funded training and GVA per worker

In Figure 10 below, we have split the industries into two - those industries associated with an above average proportion of workers in receipt of publicly funded training (blue bubbles) - and those with a less than average proportion (pink bubbles). The **size** of the bubbles represents the industry level GVA per capita (see next section for specific details of how this variable was constructed). The analysis suggests that industries in which there is a lower proportion of workers in receipt of publicly funded training (**2.2%** in the bottom 35 industries compared to **6.8%** in the top 35 industries weighted by employment) are associated with a greater GVA per head (approximately **£50,000** (weighted by employment) compared to approximately **£40,000**).

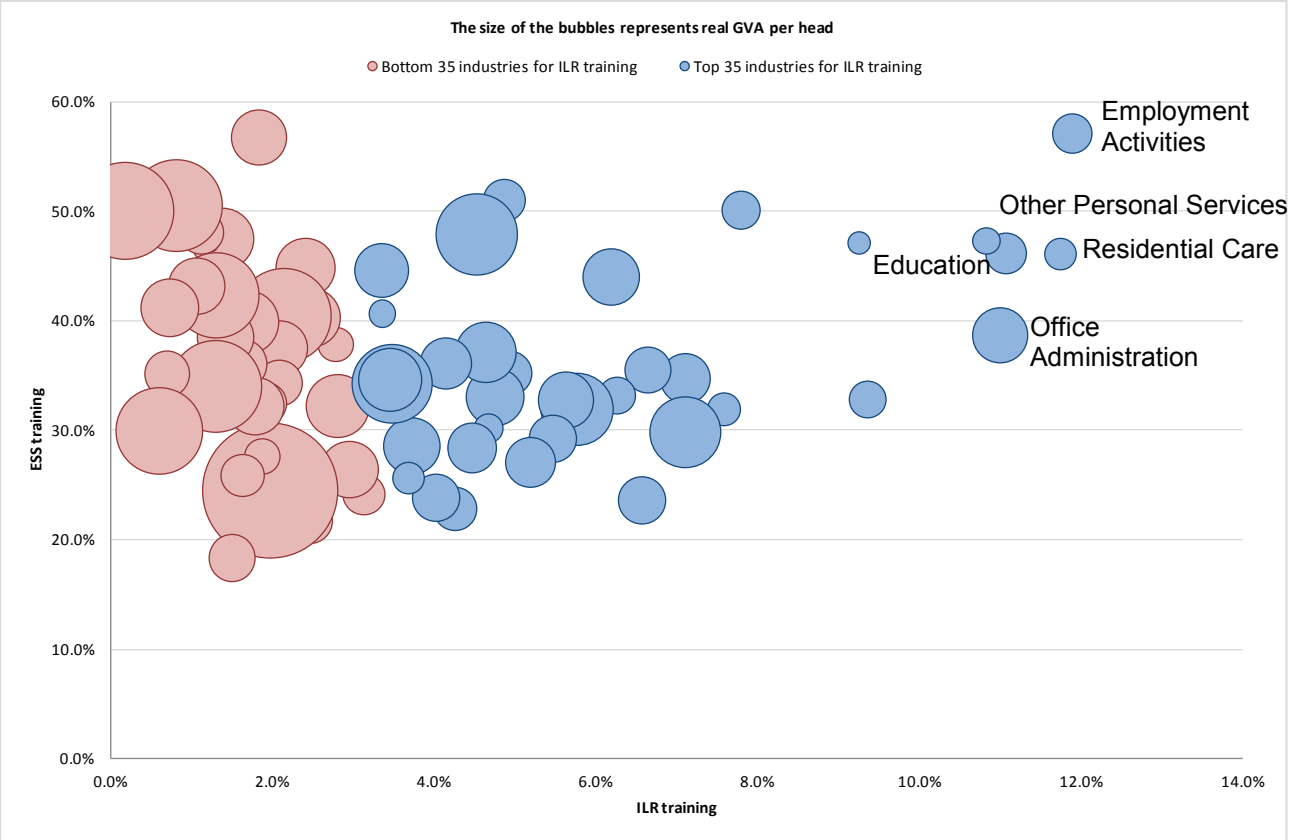


Figure 10: Proportion of workers in receipt of different types of training and GVA
Note: Publicly funded training is based on the ILR definition (for entities matched to the IDBR); the ESS measure of training is total training undertaken excluding the volume of Health and Safety and Induction training; GVA is real GVA per worker. ***Source: London Economics based on ILR, EDS, IDBR, LFS, ESS and ABS***

Of interest are the outlying industries, with the **Employment Activities**, **Other Personal Services**, **Residential Care**, **Office Administration** and **Education** industries (accounting for approximately 15% of total employment[13]) having a high incidence of workers in receipt of publicly funded training but relatively low per capita GVA.

[13] See Annex 3 for additional information on the relationship between publicly funded training, all training and employment by industry

A few points on the characteristics of publicly funded workplace training in the period considered should be taken into consideration when interpreting the data (as well as in the later analysis when considering the impact of training on industry and firm-level outcomes):

- The Train to Gain initiative was phased out starting from the end of the 2010/11 academic year while funding for Apprenticeships was increased in the period;
- Apprenticeship provision is demand-led and it is possible that lower productivity industries facing greater cost pressures and/or credit constraints, may make a more intensive use of publicly funded training than undertake privately provided training.
- As noted by Ryan (2011)[14], there may be significant variation across industries in the contents of Apprenticeship provision, meaning that publicly funded training may have significantly different quality standards.

As a result, it is important to note the potential for reverse causality to exist between publicly funded training and productivity measures and to bear this in mind when interpreting the various econometric model specifications (that control or do not control for endogeneity).

[14] "While some of the learning that is [..] supported by the Apprenticeships programme, such as craft training in engineering, combines part-time vocational education with work-based training, and thus satisfies the definition [of apprenticeship], much does not. This is because training standards vary greatly by occupation and sector, according to the decisions of individual Sector Skills Councils, and the 'frameworks' that those Councils have adopted for the **service occupations** (including those in business administration, retailing, customer service, and childcare), require little off-the-job learning and no formal education". According to Ryan, "'apprenticeship' has come to denote any publicly-funded programme of work-based learning that satisfies the (frequently undemanding) requirements for public subsidy, however limited its educational content. It is therefore important to distinguish **apprenticeship**, i.e., the functional category defined above, from **Apprenticeship**, i.e., the programme organised and funded by government. In such a situation, the use of the term 'apprenticeship' is often confusing, cosmetic and objectionable." Paul Ryan, (2011)."*Apprenticeship: between theory and practice, school and workplace*" Economics of Education Working Paper Series 0064, University of Zurich

Determining the impact of publicly funded training on industry level outcomes

Background and brief overview of the relevant literature

In the United Kingdom, **Dearden et al.** (2005)[15]combined individual-level data on training from the Labour Force Survey with industry level data from the Annual Census of Production (the predecessor of the Annual Business Inquiry and the Annual Business Survey). Their findings suggest that the overall effect of training on **productivity** at industry level is positive and robust, around twice as high as the **wage effect** and consistent across different model specifications. In fact, the coefficient associated to training varies from around **0.7** in the *Random Effect* and *Fixed Effect* specifications to **0.6** in the system *GMM* specification (which accounts for potential **endogeneity** of the training decision). For wages, the coefficient is around **0.35**, and roughly similar across different specifications. Based on these results, the authors report that *an increase in training by one percentage point at industry-level is associated with an increase in productivity (value added) of about **0.6%**, and in wages of about **0.3%***[16] .

Construction of industry characteristics

In the **industry-level** analysis, we investigated whether there is any significant impact of training on productivity using data aggregated at the **2-digit SIC code level** (with some further aggregation when required (e.g. Agriculture, Forestry and Fishing)). Data for the analysis were drawn from a variety of data sources (although not all variables were used in the final model specification). The two **training variables** used represent a measure of overall training intensity from the **UKCES Employer Skills Survey** (i.e. the proportion of workers in receipt of any form of training (excluding Health and Safety and induction training)) and/or the proportion of workers in receipt of publicly funded training leading to a formally recognised qualification from the **ILR**.

The **other variables** were drawn from a variety of data sources, including the ESS, the LFS and the ABS. In particular:

- **UKCES Employer Skills Survey**
  - *Proportion of establishments with a business plan, a training plan and a budget for training expenditure;*
  - *Proportion of staff not fully proficient (i.e. proxy for skills gap),*

[15] Dearden, L., Reed, H. and Van Reenen, J. (2005). 'The Impact of Training on Productivity and Wages: Evidence from British Panel Data', IFS Working Papers W05/16, Institute for Fiscal Studies.

[16] Conti (2005) used a similar approach for Italian firms, combining data for the period 1996-1999 from the Italian Labour Force Survey (including information on training) with accounting data on firms for the corresponding years (including information on productivity), drawn from the AIDA Database. The data from the two datasets were then aggregated at the regional (20 regions) and sectoral level (12 sectors). The main findings indicate that training has a positive and significant effect on productivity, while the effect for wages is much less robust, and smaller in size. In fact, using the results obtained in the GMM model, the author estimated that the effect of raising the stock of trained workers in an industry by one percentage point leads to a 0.4% increase in productivity and to a 0.1% increase in wages (although the latter is not statistically significant).See, Conti, G, (2005.). "Training, productivity and wages in Italy," Labour Economics, Elsevier, vol. 12(4), pages 557-576, August.

  - *Proportion of staff qualified to level 4 or above; and*
  - *Occupational structure.*
- **Labour Force Survey (pooled between 2009 and 2013 aggregated to industry level)**
  - *Highest qualification held,*
  - *Occupational structure,*
  - *proportion of Part-time workers;*
  - *number of Hours worked;*
  - *proportion in a Temporary job;*
  - *Distribution of workplace size;*
  - *Banded 'tenure' (time spent working with the same company);*
  - *Age band;*
  - *Gender (proportion female;, and*
  - *Employment estimates;*
- **Capital Stock Series** (published by the ONS)
  - *Net Capital stock series at chained volume measures* for 2009-2013
- **Business Enterprise Research and Development Survey**
  - *Expenditure on R&D* performed in UK businesses by SIC division, 2009 to 2013
- **Industry level deflators** (predominantly) drawn from ONS series
  - *Producer Price Indices* (PPI),
  - *Service Producer Price Index* (SPPI),
  - *Consumer Price Index* (CPI) and
  - *Average Weekly Earnings*),
  - *BIS/ONS Construction Output Price Indices*, and
  - Defra's *Index of Producer Prices of Agricultural Products*.

Annual Business Survey (ABS) and Business Register and Employment Survey (BRES)

The **Annual Business Survey** covers only the UK Non-Financial Business Economy, which accounts for approximately two thirds of the UK economy in terms of Gross Value Added. The industries covered are: ***Agriculture*** (support activities SIC 01.6 and hunting and trapping SIC 0.17, only), ***Forestry and Fishing*** - Section A; ***Production industries*** - Sections B-E, ***Construction industries*** - Section F; ***Distribution industries*** - Section G; ***Non-Financial Service industries*** - Sections H, I, J, L, M, N, P (private provision only), Q (private provision only in SIC 86.1 and 86.9), R and S. After removing sectors *not* covered by the ABS, we were left with **71 industries** at the **2 digit SIC code level**. The following variables were taken from the ABS and BRES 2009-13:

- *Gross Value Added;*
- *Turnover;*
- *Capital expenditure;*
- *Total purchases;*
- *Employment costs, and*
- *Employment (BRES)*

Model specification

Econometric modelling

Starting from a Cobb-Douglas production function with constant returns to scale, it is possible to derive the equation to be estimated if longitudinal data is available:

$$y_{it} = \beta x_{it} + \varepsilon_{it} \qquad (1)$$

where y is the (natural logarithm) of productivity (measured by value added per worker (or a measure of employment costs)), x identifies a series of (potentially endogenous) regressors including training, and ε is the error term (where i identifies the industry and t the year).The error term may include an individual fixed effect (α_i), an individual-invariant time-specific component, captured by time dummies (τ_i) and an idiosyncratic component (ω_{it})

$$\varepsilon_{it} = \alpha_i + \tau_t + \omega_{it} \qquad (2)$$

Different estimators are potentially available for equation (1). In particular we use the following estimators:

- Pooled Ordinary Least Squares (pooled OLS);
- Generalised Least Squares (Random Effects (RE));
- Within-groups (Fixed Effects (FE)); and
- System Generalised Method of Moments (GMM);

The **choice of the estimator will depend on the presence and nature of the biases** affecting the estimation of equation (1), which in turn depend on the assumption regarding the relationship between the training variable and the different components of the error term in equation (2). In fact we can have the following cases:

- **Training is strictly exogenous with respect to the idiosyncratic component and not correlated with the individual effect**: In this case, equation (1) can be estimated by pooled OLS or Generalised Least Squares (random effects), with the GLS being more efficient. In this case also, the cross-sectional OLS estimator would be unbiased and consistent. The pooled OLS estimator *does not* exploit the panel structure of the data (i.e. the fact that the same unit is repeatedly available over time);
- **Training is strictly exogenous with respect to the idiosyncratic component but correlated with the individual effect** (α_i): The presence of time-invariant unobservable effects means that the GLS (Random Effects) and pooled OLS estimators are biased, while the within-groups (Fixed Effect) estimator is unbiased and consistent (the estimator subtracts time averages from each variable, so that time invariant components disappear);
- **Training is not strictly exogenous with respect to the idiosyncratic component and correlated with the individual fixed effect**: In this scenario, we have to deal with not only a fixed effect, but also with endogeneity of the training decisions. We require Instrumental Variables (IV) to tackle endogeneity. The longitudinal framework allows us to take differences over time and to use past values of the endogenous regressors as valid instruments. The suitable lags (at *t-1*,

t-2 etc.) of the endogenous variables that can be used as instruments depend on the serial correlation of ω_{it} over time[17].

Two GMM estimators can be used in these circumstances: the so-called ***difference GMM*** (Arellano and Bond[18]) that estimates equation (1) in first differences using the past levels as instruments; and the ***system GMM*** (Arellano and Bover (1995)[19] and Blundell and Bond (1998)[20]), where we can combine the equation in differences and in levels (instrumented with lagged levels and lagged first differences respectively). The system GMM is generally more efficient when the lagged levels are poor predictors of the first difference.

The relationship between training and productivity is likely to present both unobservable time-invariant firm-specific characteristics and the endogeneity of training decisions, which would mean that the first three estimators presented are biased and inconsistent. Below, we briefly recap the different characteristics of the estimators available and whether they are able to deal with the different biases likely to be present in the analysis of training and productivity.

Table 4: Different modelling approaches in respect of potential biases

Estimator	Exploit panel dimension of data	Deals with biases from the presence of time-invariant unobservable effects	Deals with biases from joint determination of training decisions and productivity
Pooled OLS	No	No	No
GLS (random effects)	Yes	No	No
Within Group (fixed effects)	Yes	Yes	No
System GMM	Yes	Yes	Yes

Dependent variables

Y identifies alternatively:

- **Productivity**, expressed as the **log of real value added per worker**;
- (the log of) **real employment costs per worker**

Explanatory variables

X contains the set of explanatory variables and includes:

- Proportion of staff trained over the past 12 months (ESS[21])
- Proportion of staff undertaking publicly funded training (ILR)
- (log of) Capital Stock per worker (Capital Stock Series)
- (log of) Expenditure on R&D per worker (at t-1) (BERD)

[17] For example if ω_{it} has an autoregressive component of order one, x_{it-1} will still be endogenous, but earlier lags (starting from t-2) are suitable instruments.
[18] Arellano, M., Bond, S.R. (1991). "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations". Review of Economic Studies 58, 277—297
[19] Arellano, M., and O. Bover. (1995). "Another look at the instrumental variable estimation of error-components models". Journal of Econometrics 68: 29-51.
[20] Blundell, Richard & Bond, Stephen, 1998. "Initial conditions and moment restrictions in dynamic panel data models," Journal of Econometrics, Elsevier, vol. 87(1), pages 115-143.
[21] Note that the ESS training measure used only covers the period 2009-2013, and is generated using the three data points available for 2009, 2011 and 2013 and interpolating the values for 2010 and 2012. The variable used excluded the volume of Health & Safety/ Induction training from training intensity, and there have been some changes to the variable measuring that type of training over the years. Furthermore, the ESS only covers establishments with at least at least two staff (including both employees and working proprietors) and excludes sole traders and establishments with just one employee.

- From the ESS:
  - Proportion of establishments with a business plan, a training plan and a budget for training expenditure;
  - proportion of staff not fully proficient (at t-1);

- From the LFS:
  - (log of) usual hours worked,
  - proportion of workers in a temporary job;
  - highest qualification held (grouped as level 3 and above (including apprenticeships) or level 2 and below);
  - the proportion of managers and directors;
  - proportion working in a small workplaces (less than 25 employees);
  - proportion of workers in different age bands (four categories)[22];
  - proportion of workers in different tenure bands (less than 2 years, 2 to 10 years and more than 10 years);
  - gender split;
  - Proportion of employees in each region (by industry) and
  - yearly dummies;

Findings

In Table 6 we present the regression results for the *Random Effects* and *Fixed Effects* models, while the results from the System GMM regressions are shown in Table 7[23]. All regressions are run controlling for a) *ESS training variable only*; b) *both ESS and ILR training variables*; and c) the *ILR training variable only*. Results are presented separately for productivity and employment costs (with the results from alternative model specifications presented in the Annex).

Coefficients on ESS training variables, comparison of estimates with previous analyses and interpretation

The coefficient for ESS training in the GVA regression (*Fixed Effect* specification – Table 6) is around **0.74**, suggesting that a 1 percentage point increase in the ESS training proportion is associated with an increase in GVA of around **0.74%**. The corresponding estimate in the employment costs regression is around **0.36**, approximately half of the size of the estimate for productivity (suggesting that a 1 percentage point increase in the ESS training proportion is associated with an increase in workers' pay of around **0.36%**). The estimates for the *Random Effects* specification are slightly larger, both in the productivity and employment costs regressions (**0.778** and **0.447** respectively).

Introducing the ILR training variable has a limited effect on the estimate for the ESS training (**0.741** and **0.359** respectively in the *Fixed Effects* model and **0.795** and **0.448**

[22] In earlier model specifications, there were apparent multicollinearity issues affecting the age and qualification structure variables. To avoid this, we simplified the structure of the set of control variables and in the final specification controlled for 4 age bands, as well as 2 bands for highest qualification. Furthermore, we decided to enter only one category for the occupational structure (the proportion of directors/managers) to avoid multicollinearity with the qualification structure variables.

[23] The system GMM regressions were undertaken in Stata using the command *xtabond2* developed by Roodman, D. (2009)." How to do xtabond2: An introduction to difference and system GMM in Stata". Stata Journal 9(1): 86-136.

respectively in the *Random Effects* model). **These results are very consistent with previous research in the area** (Dearden *et al* (2005))[24].

To prove further interpretation on the meaning of the findings, these estimates (as with the Dearden *et al.* (2005) results) capture **both** the **direct effects** accruing to the firm **and** any **industry level spillovers** accruing to training and non-training firms. However, the estimates do not provide any indication on the relative size of the direct effect or the spillover effect, but only the total effect at the industry level. Similarly, the estimate of the impact of training on the wage bill covers **both** the impact on wages of those workers in receipt of training, as well as the spillover effect on workers that might not have received training.

The difference between the size of the wage and productivity returns indicate how industry level returns are shared between workers and firms – but does not provide information on the extent to which there are direct or indirect effects.

Coefficients relating to ILR training variables

In the specifications incorporating ILR training measures, industry-level estimates are more erratic, sometimes changing sign across specifications, and displaying relatively large standard errors (the coefficients are never statistically significant). This pattern is observed both in the regressions controlling for ESS and ILR training variables simultaneously and in the regressions controlling for ILR training only.

None of the coefficients associated with the training variables are statistically significant in the system GMM specification, and while the estimates for ESS training are not too dissimilar from the coefficients of the *Random Effects* and *Fixed Effects* specification (although with large standard errors), the point estimates for ILR training continue to be quite erratic. Also, when looking at the measure based on the proportion of Level 3 Apprenticeships only (presented in Annex 5), the point estimates are typically positive, but always show very large standard errors.

What do the results suggest and what do they not suggest?

At the industry level, the analysis indicates that there are **positive productivity and wage returns associated with increasing proportions of workers in the receipt of** (any form of) **training**, but failed to detect any robust relationship when considering training intensity using information from the ILR only.

However, as we explain below, **this does not necessarily indicate the absence of any impact of publicly funded training on productivity, but rather than we are unable to draw any strong conclusions due to the data currently available**. In fact, the full range of analyses suggests that there is **no consistent pattern** in the industry level coefficients associated with the ILR training variable, with the sign of the coefficient changing from negative to positive as we move from the *Random Effects* to the *Fixed Effects* specification.

[24] In terms of other explanatory variables, the coefficient for capital stock is slightly larger than **0.20** in the productivity regressions (and only significant in the RE specification), while estimates are smaller and never statistically significant in the employment costs regressions. The coefficients for lagged R&D in the Random Effects specification are around **0.037** and **0.025** for GVA and employment costs respectively, but are never statistically significant in the Fixed Effects specification. The interpretation of coefficients when both the dependent variable and one or more explanatory variables are log transformed is in terms of percentage increase in both variables. For example a coefficient of 0.2 for the capital stock indicates that an increase of 1% in the capital stock per worker is associated with an increase in productivity of around 0.2%

Furthermore, it is the deficiencies in the underpinning data that results in a lack of consistency in the direction of the effect (limiting *any* sort of conclusion), rather than a simple lack of statistical significance or the presence of coefficients close to zero (which might indicate no effect).This inability to demonstrate an impact is fundamentally related to the **data supporting** the analysis, but also in part dependent on issues relating to the exact **model specification** and the potential existence of autocorrelation in the error term and insufficient data to control for endogeneity[25].

In particular, in relation to the data difficulties surrounding the **ILR training measure**:

First, the ILR training proportion was generated using the matched sample only, so this is an **underestimate** of total training undertaken (in particular omitting some of the smaller firms). Furthermore, the training measure from the ILR is not generated from an employer survey (as is the case with the ESS), but from administrative data – and we are combining ILR data with ABS, LFS, and other data sources of completely different nature[26].

Secondly, looking at the actual distribution of the training variables, the variation for the ILR measure of training is relatively high, given that it ranges from values around 0.1% to 14.8% (see Table 5 below). In fact, the ratio of the Standard Deviation divided by the mean (Coefficient of Variation) is much higher for ILR training than for ESS training (0.781 compared to 0.262). This probably partly reflects the matching process (and the loss of some information about training undertaken) and the fact that some industries may make limited use of apprenticeships and other publicly funded training (for historical reasons or given that apprenticeship training would not suit their training requirements or perceptions in relation to the quality of training). Therefore, because the Coefficient of Variation is significantly higher for the ILR measure relative to the ESS measure, this could imply that the ILR measure may simply provide a weaker signal, reflecting the heterogeneity in the content of publicly funded training undertaken (training standards may vary significantly across industries).

Table 5: Summary of basic descriptive statistics of training variables by source

	Pooled over the period	
	ILR Training	ESS training
Mean	0.040	0.360
Standard Deviation	0.031	0.095
Coefficient of Variation (SD/mean)	0.781	0.262
Min	0.001	0.172
Max	0.148	0.663

Some further interpretation of model diagnostics

In relation to the model specification, in this paragraph we discuss the diagnostics supporting the *System GMM* estimation, reported at the bottom of Table 7. The System GMM regressions contain the p-values from the autocorrelation tests of order 1 and 2 (AR1 and AR2, see Table 7). A low p-value of the test (e.g. less than 0.1) indicates the presence of autocorrelation in the first-differenced residuals (i.e. we reject the null

[25] In Table 12, we also present alternative model specifications where alternative measures of training are incorporated into the analysis (for instance, the average number of days of training for those in receipt of training); as well as information on the proportion of enterprises engaged in training. The point estimates in this model specification suggest that there is a positive relationship between publicly funded training leading to formally recognised qualifications and both salary bill and productivity (although not statistically significant). See Annex 5 for further information.

[26] It is also possible that survey responses may be more suitable for subsequent aggregation at the industry level: this may be especially the case for enterprises with multiple local units (or establishments), and/or when the IDBR reports multiple SIC codes of activity for each enterprise (and the primary SIC code not necessarily corresponding to the one that would be reported in a survey by the establishment providing training).

hypothesis of no autocorrelation). In fact, for the industry level regressions the AR1 tests always show the presence of autocorrelation of Order 1 in the differenced residuals at the 10% level (suggesting endogeneity, which is expected given the structure of the model). If there is no evidence of second-order autocorrelation, the values at t-2 (and earlier lags) can be used as valid instruments for the potentially endogenous variables

However, for the wage regressions we cannot reject the presence of AR2 – **second order autocorrelation**. More generally, when considering the presence of endogeneity, we should take into account that the only potentially exogenous instruments are the lagged values of the endogenous variables (e.g. training). However, autocorrelation in the residuals will indicate that some lagged values cannot be treated as valid instruments (e.g. they are still endogenous), and deeper lags should be used (which are not currently available given the data periods available)[27].

Finally, the Hansen test of whether the instrument set appears exogenous fails to reject the null hypothesis of instrument validity (suggesting the "appropriateness" of the instrument set used)[28].

What does this mean for the range of estimators?

Although the *Fixed Effects* estimates will be biased in the presence of endogeneity and we are not able to quantify the direction or the extent of the bias (given that our *GMM estimates* are never statistically significant), they are still meaningful coefficients. For example in the Dearden (2005) paper, the training coefficient in the *Fixed Effect* specification overstated the unbiased return by around 15% (i.e. the coefficient was 0.7 for the FE specification vs. 0.6 for the GMM coefficient accounting for endogeneity), but to a much smaller extent in the wage regression (0.365 vs. 0.351). This means that although we are not able to identify the exact magnitude or the direction of the bias, the *Fixed Effect* coefficient may still provide valuable information on the effect of training on productivity.

Final Remarks

As such, at industry-level, given the many issues in relation to the data used, it would be fundamentally incorrect to conclude that engagement in publicly funded training has no effect on either salary bill or productivity levels. Although there are obvious methodological challenges, we believe that this analysis is important. Despite using completely different survey data to undertake the analysis (reflecting the employer rather than employee perspectives on training undertaken), and acknowledging the gaps in the data set considered; the difference in timeframe, and the industries under consideration, these results are very similar to those that were generated using individual level LFS survey data (see Dearden et al., (2005)).

[27] Similar diagnostics tests are standard in System GMM estimation: for example the diagnostics presented in Table 2 of the Dearden et al. report (2005) clearly support the presence of first-order autocorrelation, but show no sign of second-order serial correlation. However, in earlier specifications of the wage regression they found signs of second-order autocorrelation which led them to use deeper lags. Specifically, "*the wage regression uses instruments dated t-3 and before in the differenced equation (and dated t-2 in the levels equation). This is because there were some signs of significant second-order serial correlation using t-2 dated instruments in the wage equation, which invalidates the IVs*".

[28] The Hansen (or Sargan-Hansen) test of the over-identifying restrictions is commonly used to test the validity of the instrument set used: for example Dearden et al. (2005) present the Sargan statistics in Table 2 of their report. Spevifically, the authors state that "*the Sargan test of over-identifying restrictions does not reject [the null hypothesis]*").

Table 6: Results of econometric modelling (Fixed and Random Effects models)

Dependent variable	Log GVA per worker						Log employment costs per worker					
Controlling for (training variables)	*ESS only*	*ESS only*	*Both*	*Both*	*ILR only*	*ILR only*	*ESS only*	*ESS only*	*Both*	*Both*	*ILR only*	*ILR only*
Estimator	*Random Effects*	*Fixed Effects*	*Random Effects*	*Fixed Effects*	*Random Effects*	*Fixed Effects*	*Random Effects*	*Fixed Effects*	*Random Effects*	*Fixed Effects*	*Random Effects*	*Fixed Effects*
Training intensity ESS	0.778* (0.398)	0.744** (0.372)	0.795** (0.394)	0.741* (0.374)			0.447** (0.217)	0.365* (0.192)	0.448** (0.214)	0.359* (0.195)		
Trained proportion ILR			-0.946 (1.053)	0.439 (0.968)	-0.785 (0.981)	0.503 (0.853)			-0.083 (0.816)	0.718 (0.849)	0.027 (0.788)	0.749 (0.792)
Capital stock per head (log)	0.206*** (0.053)	0.222 (0.240)	0.206*** (0.050)	0.218 (0.240)	0.209*** (0.050)	0.272 (0.248)	0.041 (0.029)	-0.097 (0.106)	0.041 (0.029)	-0.104 (0.106)	0.043 (0.029)	-0.078 (0.110)
R&D expenditure per head at t-1 (log)	0.037* (0.022)	0.006 (0.022)	0.037* (0.022)	0.004 (0.023)	0.037* (0.022)	0.002 (0.022)	0.025** (0.012)	-0.002 (0.017)	0.025** (0.012)	-0.005 (0.018)	0.025** (0.012)	-0.005 (0.018)
Usual hours worked (log)	1.042 (0.721)	-0.911 (0.603)	1.115 (0.728)	-0.905 (0.605)	1.193 (0.736)	-0.956 (0.616)	0.611* (0.330)	-0.246 (0.431)	0.616* (0.330)	-0.236 (0.429)	0.655** (0.332)	-0.260 (0.430)
Proportion in a temporary job	-2.776** (1.301)	-1.951* (1.023)	-2.833** (1.315)	-1.941* (1.022)	-2.733** (1.354)	-1.851* (1.082)	-0.812* (0.444)	-0.339 (0.343)	-0.820* (0.447)	-0.324 (0.335)	-0.761 (0.473)	-0.280 (0.351)
Highest qualification:L3 and above (including apprenticeships)	-0.239 (0.354)	-0.345 (0.387)	-0.226 (0.359)	-0.344 (0.387)	-0.119 (0.366)	-0.344 (0.399)	0.037 (0.206)	-0.222 (0.225)	0.040 (0.207)	-0.221 (0.222)	0.105 (0.207)	-0.221 (0.227)
Occupation: Managers and professionals	0.172 (0.253)	0.062 (0.231)	0.154 (0.252)	0.062 (0.231)	0.128 (0.271)	-0.004 (0.236)	0.279* (0.160)	-0.068 (0.144)	0.281* (0.160)	-0.068 (0.140)	0.277* (0.167)	-0.100 (0.158)
Proportion with business plan, training plan and training budget	-0.503 (0.351)	-0.401 (0.411)	-0.496 (0.367)	-0.392 (0.409)	-0.437 (0.402)	-0.422 (0.465)	0.402* (0.232)	0.280 (0.209)	0.405* (0.237)	0.294 (0.204)	0.447* (0.235)	0.279 (0.203)
Proportion of staff not fully proficient at t-1	-0.913*** (0.291)	-0.977*** (0.275)	-0.911*** (0.304)	-0.981*** (0.276)	-0.790*** (0.286)	-0.854*** (0.264)	-0.194 (0.188)	-0.266 (0.162)	-0.193 (0.188)	-0.271 (0.171)	-0.125 (0.184)	-0.210 (0.169)
Small workplace (less than 25 employees)	-0.847*** (0.244)	-0.385 (0.422)	-0.833*** (0.241)	-0.396 (0.422)	-0.906*** (0.251)	-0.417 (0.424)	-0.721*** (0.128)	-0.128 (0.165)	-0.722*** (0.135)	-0.145 (0.163)	-0.764*** (0.134)	-0.156 (0.166)
Age band 16-24	1.693* (0.883)	1.496 (0.904)	1.762** (0.896)	1.454 (0.912)	1.962** (0.941)	1.740* (0.957)	0.498 (0.527)	0.754 (0.458)	0.502 (0.533)	0.687 (0.464)	0.607 (0.548)	0.825* (0.465)
Age band 25-39	2.595*** (0.835)	2.352*** (0.821)	2.653*** (0.848)	2.328*** (0.828)	2.826*** (0.904)	2.494*** (0.858)	0.737 (0.481)	0.849* (0.441)	0.742 (0.484)	0.810* (0.450)	0.838* (0.505)	0.891** (0.443)
Age band 40-54	1.683** (0.696)	1.519** (0.675)	1.698** (0.706)	1.515** (0.680)	1.843** (0.746)	1.681** (0.708)	0.765 (0.505)	0.727 (0.462)	0.767 (0.505)	0.722 (0.470)	0.846 (0.529)	0.802* (0.477)
Tenure 2-10 years	-0.566 (0.564)	-0.505 (0.427)	-0.554 (0.584)	-0.515 (0.428)	-0.507 (0.623)	-0.442 (0.433)	-0.143 (0.263)	-0.151 (0.215)	-0.141 (0.265)	-0.167 (0.215)	-0.110 (0.286)	-0.132 (0.223)
Tenure more than 10 years	-0.057 (0.697)	0.037 (0.496)	-0.115 (0.710)	0.040 (0.493)	-0.102 (0.742)	0.086 (0.496)	0.296 (0.370)	0.241 (0.338)	0.292 (0.368)	0.247 (0.333)	0.307 (0.394)	0.269 (0.338)
Proportion female	0.027 (0.270)	0.208 (0.301)	0.037 (0.263)	0.207 (0.304)	0.191 (0.274)	0.286 (0.335)	-0.083 (0.201)	0.242 (0.224)	-0.086 (0.199)	0.241 (0.217)	-0.003 (0.190)	0.280 (0.215)
Observations	270	270	270	270	270	270	271	271	271	271	271	271

Note: Standard errors clustered at the industry level in parentheses. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level; All regressions include regional controls and yearly dummies. Omitted category for age band is 55 and over. Omitted category for Tenure is less than 2 years. ***Source: London Economics based on data from ABS, ILR, LFS, BRES and ESS***

Table 7: Results of econometric modelling (System GMM Model)

System GMM						
Controlling for (training variables)	*ESS only*	*Both*	*ILR only*	*ESS only*	*Both*	*ILR only*
Dependent variable	*GVA*	*GVA*	*GVA*	*Employment costs*	*Employment costs*	*Employment costs*
Training intensity ESS	0.930 (1.064)	0.847 (0.840)		0.326 (0.402)	0.274 (0.359)	
Trained proportion ILR		-2.994 (2.912)	-2.749 (2.807)		0.451 (2.067)	0.324 (1.471)
Capital stock per head (log)	0.135 (0.129)	0.163* (0.089)	0.130 (0.152)	0.001 (0.053)	0.008 (0.064)	0.030 (0.052)
R&D expenditure per head at t-1 (log)	-0.101 (0.154)	-0.023 (0.093)	-0.015 (0.088)	0.017 (0.041)	0.018 (0.041)	-0.012 (0.046)
Usual hours worked (log)	2.955 (4.115)	2.958 (2.448)	3.067 (2.677)	-0.963 (1.129)	-0.431 (1.183)	-0.492 (1.112)
Proportion in a temporary job	-5.750** (2.496)	-4.350 (2.788)	-4.112** (2.054)	-2.118** (0.840)	-1.960** (0.740)	-1.736** (0.668)
Highest qualification:L3 and above (including apprenticeships)	0.934 (0.998)	0.553 (0.992)	0.918 (0.828)	0.048 (0.395)	0.054 (0.374)	0.238 (0.345)
Occupation: Managers and professionals	0.188 (1.106)	-0.225 (0.908)	-0.465 (0.754)	0.611 (0.488)	0.643 (0.427)	0.723 (0.493)
Proportion with business plan, training plan and training budget	-1.468 (0.885)	-0.654 (1.101)	-0.134 (0.704)	0.535 (0.473)	0.489 (0.428)	0.537 (0.406)
Proportion of staff not fully proficient at t-1	0.149 (3.731)	-1.159 (3.895)	-0.925 (1.671)	-0.190 (0.578)	-0.134 (0.532)	0.196 (0.774)
Small workplace (less than 25 employees)	-1.641* (0.907)	-1.056 (0.696)	-0.916 (0.560)	-0.967*** (0.321)	-0.973*** (0.293)	-0.984*** (0.291)
Age band 16-24	6.125* (3.575)	4.137* (2.243)	4.066* (2.337)	0.152 (1.493)	0.504 (1.249)	0.745 (1.194)
Age band 25-39	4.995** (2.129)	4.385** (1.966)	4.371** (1.823)	1.737** (0.855)	1.499* (0.898)	2.237** (0.902)
Age band 40-54	4.008** (1.948)	3.042 (2.155)	3.424* (1.996)	1.679** (0.750)	1.627** (0.750)	1.988** (0.848)
Tenure 2-10 years	0.049 (1.394)	0.143 (2.049)	0.006 (1.238)	-0.051 (0.692)	0.005 (0.637)	0.076 (0.463)
Tenure more than 10 years	0.859 (2.028)	-0.194 (1.669)	0.208 (1.613)	0.223 (0.823)	0.467 (0.644)	0.525 (0.548)
Proportion female	0.129 (1.327)	0.195 (0.853)	0.353 (0.695)	-0.919** (0.393)	-0.675 (0.452)	-0.794** (0.382)
Observations	270	270	270	271	271	271
AR1	0.06	0.06	0.09	0.02	0.03	0.05
AR2	0.48	0.41	0.65	0.01	0.02	0.03
Hansen test	0.70	0.80	0.53	0.55	0.70	0.38

Note: Two-step robust standard errors in parentheses. All regressions include regional controls and yearly dummies. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level; Omitted category for age band is 55 and over. Omitted category for Tenure is less than 2 years Variables treated as endogenous in the System GMM regression: ESS training, ILR training, Capital Stock, R&D expenditure, Number of Hours. The values reported for AR(1) and AR(2) are the p-values for first and second order autocorrelated disturbances in the first differences equations (null hypothesis is no autocorrelation). The Hansen test reports the p-values for the null hypothesis of instrument validity.

Source: London Economics based on data from ABS, ILR, LFS and ESS

Determining the impact of publicly funded training on firm-level outcomes

Background and brief overview of the relevant literature

In the United Kingdom, **Haskel and Galindo-Rueda** (2005)[29] used firm-level data from the Annual Business Inquiry matched with firm-level data on skills and workforce characteristics from the 2001 Employer Skills Survey. Although the analysis was restricted to a single time period due to the cross-sectional nature of the NESS, the authors found that *higher level qualifications had a positive effect on firms' productivity*, with the results varying by sector and being robust only for *full-time male workers*. However, the analysis demonstrated that *low-level qualifications had a negligible effect on productivity*. The analysis also considered the impact of qualifications on wages and compared this with the effect on productivity. Comparing the effect of skills on wages with the effect of skills on productivity, the authors found that, for *higher level* qualifications, there was a positive differential for services (i.e. a larger effect of skills on wages) and a negative differential (not statistically significant) for manufacturing.

Looking at other European countries, **Colombo and Stanca** (2008)[30] used a panel of **Italian firms** and illustrated that a one percentage point increase in training intensity boosted firms' productivity by about **0.07%**. They also found that *not* controlling for unobserved firm characteristics led to an over-estimate of the effect of training on productivity, while ignoring endogeneity led to an underestimate of the effect. Moreover, the effect was even larger when they controlled for *training duration* (available only in a subsample of cases). Training had also a positive effect on wages in the firms undertaking the training, but this was found to be significantly smaller than the effect on productivity. The impact of training by occupational groups was varied, with high returns found for blue-collar workers and negligible returns for white-collar workers.

In another analysis, **Konings and Vanormelingen** (2010)[31] used longitudinal data on Belgian firms and found that the productivity effect of training in the aggregate equation (controlling for the endogeneity of training and inputs) was around **0.24%**, while the wage effect was around **0.17%**. When estimating the production function by industry, the un-weighted average for the training coefficient was around **0.18** in the productivity equation and around **0.12** in the wage equation (indicating that a one percentage point increase in training raised firm-level productivity by approximately **0.18%** and wages by **0.12%**).

---

[29] Galindo-Rueda, F. & Haskel, J., 2005. "Skills, Workforce Characteristics and Firm-Level Productivity: Evidence from the Matched ABI/Employer Skills Survey" IZA Discussion Papers 1542, Institute for the Study of Labor (IZA).
[30] Colombo, E, & Stanca, L. (2008). 'The Impact of Training on Productivity: Evidence from a Large Panel of Firms', Working Papers 134, University of Milano-Bicocca, Department of Economics.
[31] Konings, J. & Vanormelingen, S., (2010). 'The Impact of Training on Productivity and Wages: Firm Level Evidence', IZA Discussion Papers 4731, Institute for the Study of Labor (IZA).

Construction of firm-level characteristics

The matched ILR/EDS/IDBR dataset described in the previous sections was used for further analysis at the firm level. In fact, thanks to the presence of the IDBR identifiers (at the enterprise, reporting unit and local unit level), it was possible to link information from the Office for National Statistics and other surveys with information on publicly funded training.

In particular, we used the Annual Business Survey (ABS) to construct a **firm-level** dataset containing information from the ABS on productivity (Gross Value Added), employment costs (and other variables) with information on publicly funded training from the ILR. The structure of the ABS is presented in Annex 1.

All nominal variables were deflated using industry level deflators[32] (mainly provided by the ONS and drawn from the Producer Price Index, the Service Producer Price Index and Average Weekly Earnings). Firm-level employment estimates were taken from the IDBR employment variable available in the ABS (the issues related to employment data are discussed below). Descriptive statistics for the ABS group of firms are presented in Annex 5.

Model specification

Similar to the industry level analysis we estimated the following model

$$y_{it} = \beta x_{it} + \varepsilon_{it} \qquad (3)$$

where *i* now identifies firms rather than industries and *t* identifies the years from 2010 to 2013. The methodological approach used for the firm-level analysis replicated the approach outlined for the industry-level analysis. However, it is important to note that **micro firms** (with less than 10 employees) are only sampled with low frequency and were not retained for the panel analysis.

Below we present in more detail the variable used in the firm-level analysis. All firm-level variables are drawn from the ABS/IDBR unless otherwise stated.

Dependent variables

Similarly to the industry level analysis, the dependent variables used are:

- ***Productivity***, expressed as (the log of) *real GVA per worker (at market prices)*;
- (the log of) ***real employment costs per worker***;

Independent variables

The variable identifying *the proportion of workers undertaking publicly funded training* was generated using ILR data matched to the IDBR and the ABS.

Other control variables at the firm-level covered:

- (the log of) **real capital expenditure per worker**;

[32] The majority of the deflators used were at the 2-digit SIC code level, while a small number were available at the 4-digit and 3-digit level.

- (the log of) expenditure on **computer and related services** and **advertising and marketing services** as a proportion of total purchases;
- **Foreign ownership**;
- Share of **part-time** workers (BRES);
- Enterprise with **multiple local units**;
- Company **age band** (divided in 4 categories);
- **Legal status** (7 categories);
- **Company size** (4 categories based on employment, excluding micro firms);
- **Industry** at the aggregate SIC section level (17 categories);
- **Regional** and **yearly** dummies;

A series of key variables are not currently available at the firm level, so we decided to control using the industry-level measure[33]. In particular we included the following indicators:

- From the ESS, we retained the **proportion or workers receiving training** (excluding the volume of Health and Safety and Induction training) and the **proportion of staff not fully proficient** (measured at t-1). The ESS variable were aggregated at the industry (71 groups) and company size level (4 employment bands)[34];
- From the LFS, we used the **number of hours worked** (entered in logs) and the **proportion of temporary workers** (all measures aggregated at the industry level);
- From the Capital Stock Series, we retained the **capital stock measure** at the industry level[35];
- From BERD, we used the measure of **R&D** (log expenditure per worker) at the industry level.

Caveats

Unfortunately, there are currently a number of **key weaknesses** associated to the analysis of training and productivity at the firm-level. The main ones are:

- Availability of **employment** measures at the firm level: The ABS does not collect employment level information so instead this key information needs to be taken from another source - the **Business Register and Employment Survey** (BRES) - which is a survey of approximately 80,000 businesses. If an enterprise is selected for BRES, then all its constituent local units are selected (although this may not always happen in practice)[36]. The ONS currently advises caution when combining ABS data with BRES estimates of employment due to differences in sampling

[33] Unfortunately, for the firm level analysis, there are some significant gaps in the data, for example we lack information on other training undertaken and a firm-level estimate for the capital stock. We tried to use the industry level measure for some of these key variables as an imperfect proxy for the firm level data, but clearly these measures have little variation across firms (there are only 71 industries). The alternative is to omit these potential control variables, which would be 'selective' in terms of which results to estimate and present.
[34] To determine company size we used organisation size rather than establishment size when an establishment was part of a larger organisation
[35] In previous analysis, we used capital expenditure as well as capital stock when available (on a subset of firms). The relevant measure in the production function should be the stock of capital, and not the flow (capital expenditure). We controlled for capital expenditure given that an estimate for the capital stock at the firm level was not available. However for any robust conclusions to be drawn we would need to have firm-level information on the level of capital stock, at least for a subset of firms
[36] For more information see the publication BRES quality and methodology information available here

periods, sample designs and questionnaire reference periods (the two surveys also have different estimation methodologies, data validation and quality assurance and publication timetables)[37]. In addition, BRES does not cover all IDBR enterprises or all enterprises (or reporting units) surveyed in the ABS. Due to the issues existing with the BRES estimates of employment at the enterprise level, we decided to use the IDBR estimate of employment available in the ABS[38]. However, the employment estimates in the IDBR are in turn derived from BRES, the Short Term Employment Survey, PAYE and imputation from VAT returns[39].

- Currently the ONS' VML does not hold up-to-date **capital stock series** at the firm level[40]. Although a more up-to-date version of the capital stock series may be available through other sources[41], no data transfer was made before the completion of the project and we were unable to use any estimate of the capital stock series at the firm level.
- Information on all or **other** (non-publicly funded) **training undertaken** is not available in the ABS or other ONS surveys at the firm level. The UKCES' Employer Skills Survey contains information on training and skills. However, only around 50% of ESS entries are matched to the IDBR and it is possible to attach ESS information in the ABS for only about 4,000 reporting units. The ESS 2013 is not yet available in the VML and has not been matched to the IDBR (the most recent ESS currently available is the 2011 ESS). As a consequence no usable information at the firm level was available for the analysis (not even on a restricted sample of firms).

Findings

We estimated equation (3) using *Random Effects*, *Fixed Effects* and *System GMM*[42] over the period 2010-13. To take into account the lack of some key variables at the firm level we used different sets of control variables. In particular we used the following specifications:

- A basic specification controlling for firm-level characteristics only (excluding company size and SIC section controls);
- A second specification controlling for firm-level characteristics only, but including company size and SIC section controls;
- An extended specification controlling for key industry level firm-level characteristics in addition to the firm-level characteristics from the second specification.

Regression results for the random and fixed effect specifications are presented in Table 8.

---

[37] See the ABS Technical Report - June 2014 available here

[38] ONS documentation in the VML (where the firm-level analysis was undertaken) also advises against the use of BRES employment estimates combined with ABS data.

[39] In the publication "Further information about IDBR sources, structure and updating for publications" the ONS reports that "The use of IDBR employment in publications and for employment analysis has to be considered carefully. Generally ONS would recommend using BRES employment estimates for detailed industry and geographical employment comparisons. The only time that IDBR employment should be considered is for use is for very small area, or fine cross tabulations, below the level of BRES publication."

[40] Personal communication from the VML team received in July 2015 and reiterated in November 2015: "As for the Capital Stock data (BSDI series) the most recent data years are 2000-2009.The original capital stocks file runs from 1979 to 2006; a later update exists running up to 2009, but without clear documentation to show whether different assumptions were made. New users will therefore have to use their own discretion as to which series to use, or whether they can invest the time to re-run the code and create their own. Note that it depends on the ARD and hence will be very hard to update beyond 2009. One VML user has done work on the Capital Stock data and the data may be useful to others."

[41] The UK Data Service

[42] We also estimated pooled OLS but results are not shown in the report

The findings for the different specifications show that:

- In the productivity regressions (fixed effect model) the coefficient estimating the effect of ILR training on productivity is around **0.78** in the basic specification and then drops to **0.67** when we also control for firm size and SIC sections. When we control for industry level measures, the point estimate drops further to **0.46** and is no longer statistically significant. The point estimates for the Random Effect specification are much smaller (between **0.13** and **0.27**) – but not statistically significant.
- The regressions for the wage bill show a more consistent path for the coefficient of ILR training: the point estimate is always around **0.2** in the random effect model and ranges between **0.27** and **0.37** in the fixed effect model.
- Looking at the fixed effect model, the analysis suggests that raising the proportion of publicly funded training by one percentage point is associated with a **0.67%** increase in productivity (with the effect dropping to **0.47%** and not statistically significant in the specification controlling for the industry level measures). The findings from the wage regression would suggest that raising the proportion of publicly funded training by one percentage point is associated with a **0.3%** increase in the wage bill. These results are comparable to the industry level analysis.
- Looking at the other firm level variables, *capital expenditure* is positive in both the productivity (coefficient between **0.10** and **0.14**) and wage regressions (coefficient between **0.026** and **0.06**). The coefficient for the *expenditure on computer and advertising services* is in positive in the productivity regressions (around **0.15-0.17** in the fixed effect model) while drops to zero in the wage regressions[43]. The effect of *foreign ownership* is positive only for the wage regression and only in the random effect model. The coefficient for the *share of part-time workers* is negative in the random effect model, while it is small and never statistically significant once time-invariant effects are controlled for. Finally, the coefficient associated with *multiple local unit enterprises* is generally negative in the wage regressions, and never statistically significant in the productivity regressions.
- The effect of the industry-level measures is not clear-cut, with the coefficient associated with ESS training never statistically significant and close to zero in almost all specifications (excluding the fixed effect model in the productivity regression). The coefficient for the capital stock at the industry level is never statistically significant and moves from negative to positive in the productivity regressions (while it is very small in the wage regressions).
- **The system GMM regressions exhibit strong second order autocorrelation and the Hansen test always rejected the validity of the set of instruments used**. The limited number of periods currently available and the lack of key data restrict further analysis on the GMM specification. Results for the GMM specification are shown in Annex 4[44].
- In general, all the estimated coefficients presented in Table 8 and Table 9 point towards **a positive impact of publicly funded training on firm-level wage bill and productivity**. However, the firm-level analysis presented suffers from two main weaknesses, *the length of the dataset currently available* (four years) and *the lack of*

[43] Both variables are entered in logs, so the interpretation is that a 1% increase in (for example) capital expenditure is associated with an increase in the dependent variable (e.g. productivity) of around 0.14%
[44] The system GMM regressions were estimated in Stata using the *xtabond2* command developed by Roodman (2009)

information on two key variables that may significantly affect the estimates of the effect of publicly funded training on productivity and the wage bill: the proportion of other training undertaken (ESS training) and the level of capital stock at the firm level. Also, the random effect and fixed effect models do not control for the endogeneity of training decisions[45] and it was not possible to achieve a well-defined specification for the system GMM regressions. Consequently, **these findings should be seen as exploratory and should be treated with extreme caution.**

Analysis disaggregated by firm size

We have also disaggregated the sample into large enterprises (those with employment of 250 or more) and medium size enterprises (those with employment between 10 and 249) and replicated the analysis based on the *third* model specification (controlling for both firm characteristics and industry level measures). The main regression results are presented in Table 9. Although the coefficient for publicly funded training are statistically insignificant in the productivity regression, the point estimate stand at **0.34** (large firms) and **0.51** (medium size firms) in the fixed effect model. The coefficient associated with publicly funded training in the wage regression is quite similar across the two groups and always statistically significant: in the random effect model is approximately **0.17** for large firms (**0.18** for medium size firms), while in the fixed effect model is about **0.27** for large firms and slightly larger for medium size firms (**0.30**).

Looking at the other firm-level variables, the coefficient associated with capital expenditure is higher for medium size firms in the productivity regressions and quite similar in the wage regressions, while the coefficient associated with expenditure on computer and advertising services as a proportion of total purchases is slightly higher in the productivity regressions for large firms (and generally small in the wage regressions). The coefficients associated with industry level measures (in particular, overall training undertaken and the capital stock measure) are quite erratic in terms of both size and sign. The results from the GMM regressions are presented in Annex 5.

[45] It is possible that more productive firms decide to engage in publicly funded training rather than publicly funded training has a positive effect on productivity or wages

Table 8: Results of firm-level economic modelling (Fixed and Random Effects models)

Dependent variable	Log GVA per worker						Log employment costs per worker					
Controlling for (training variables)	*Basic controls*		*Add firm size & SIC section controls*		*Add industry level controls*		*Basic controls*		*Add firm size & SIC section controls*		*Add industry level controls*	
Estimator	*Random Effects*	*Fixed Effects*	*Random Effects*	*Fixed Effects*	*Random Effects*	*Fixed Effects*	*Random Effects*	*Fixed Effects*	*Random Effects*	*Fixed Effects*	*Random Effects*	*Fixed Effects*
Training proportion (ILR)	0.197 (0.289)	**0.777** (0.358)**	0.264 (0.284)	**0.668** (0.356)*	0.126 (0.339)	**0.461 (0.432)**	0.185*** (0.049)	**0.369*** (0.055)**	0.202*** (0.048)	**0.266*** (0.048)**	0.215*** (0.052)	0.306*** (0.054)
Training intensity (ESS)					-0.008 (0.370)	0.680 (0.520)					-0.048 (0.050)	0.031 (0.056)
(log) Capital stock per worker					-0.286*** (0.081)	0.195 (0.212)					-0.003 (0.014)	-0.041 (0.028)
(log) Capital expenditure per worker	0.142*** (0.024)	0.103*** (0.030)	**0.143*** (0.025)**	**0.100*** (0.030)**	0.141*** (0.027)	0.097*** (0.034)	0.060*** (0.003)	0.035*** (0.003)	0.057*** (0.003)	0.032*** (0.003)	0.050*** (0.003)	0.026*** (0.003)
(log)R&D per worker					0.003 (0.031)	0.090 (0.057)					0.037*** (0.004)	0.005 (0.007)
(log) Expenditure on computer and advertising services as a proportion of total purchases	0.072*** (0.021)	0.156*** (0.032)	**0.094*** (0.022)**	**0.155*** (0.032)**	0.109*** (0.024)	0.172*** (0.037)	-0.007** (0.004)	0.005 (0.005)	-0.004 (0.004)	0.004 (0.005)	0.000 (0.004)	0.004 (0.005)
(log) Hours worked					0.372 (0.378)	0.933 (1.108)					1.037*** (0.070)	-0.277*** (0.106)
Proportion in a temporary job					-4.262*** (1.091)	-2.000 (1.894)					-0.606*** (0.148)	-0.127 (0.163)
Staff non proficient at t-1					-1.329 (0.840)	-1.636* (0.915)					-0.022 (0.082)	0.004 (0.080)
Foreign owned	0.077 (0.067)	-0.136 (0.121)	0.030 (0.069)	-0.135 (0.121)	0.055 (0.071)	-0.048 (0.126)	0.158*** (0.010)	-0.004 (0.013)	0.155*** (0.010)	-0.007 (0.013)	0.137*** (0.009)	-0.002 (0.010)
Share of part-time workers	-1.105*** (0.116)	-0.107 (0.159)	-0.671*** (0.131)	-0.102 (0.160)	-0.576*** (0.154)	-0.092 (0.191)	-0.672*** (0.026)	-0.015 (0.028)	-0.497*** (0.026)	-0.011 (0.027)	-0.434*** (0.025)	-0.034 (0.023)
Multiple Local Units	-0.014 (0.070)	0.122 (0.154)	0.020 (0.079)	0.161 (0.154)	0.058 (0.086)	0.287 (0.193)	-0.068*** (0.010)	-0.101*** (0.017)	0.005 (0.011)	-0.067*** (0.016)	0.006 (0.011)	-0.059*** (0.017)
Observations	37,454	37,454	37,454	37,454	32,500	32,500	37,656	37,656	37,656	37,656	32,683	32,683
Number of reporting units	11,794	11,794	11,794	11,794	11,591	11,591	11,831	11,831	11,831	11,831	11,631	11,631

Note: Standard errors clustered at the industry level in parentheses. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level; All regressions include controls for company age bands, legal status, region and year. The second and third sets of regressions also include controls for employment size bands and SIC sections. Grey shading identifies measures aggregated at the industry level.

Source: London Economics based on data from ABS, ILR, IDBR, LFS and ESS

Table 9: Results of firm-level economic modelling by firm size (Fixed and Random Effects models)

Group	Large reporting units (employment 250 or more)				Medium size reporting units (employment 10-249)			
Dependent variable	**GVA**		**Employment costs**		**GVA**		**Employment costs**	
Estimator	*Random Effects*	*Fixed Effects*	*Random Effects*	*Fixed Effects*	*Random Effects*	*Fixed Effects*	*Random Effects*	*Fixed Effects*
Training proportion (ILR)	-0.171 (0.396)	0.340 (0.446)	0.168*** (0.053)	0.266*** (0.052)	0.365 (0.513)	0.511 (0.800)	0.180** (0.073)	0.298*** (0.084)
Training intensity (ESS)	0.413 (0.412)	0.900* (0.518)	-0.096* (0.056)	0.059 (0.060)	-0.003 (0.821)	1.465 (1.841)	0.060 (0.113)	-0.118 (0.144)
(log) Capital stock per worker	-0.355*** (0.109)	-0.001 (0.186)	-0.012 (0.018)	-0.058* (0.031)	-0.201* (0.118)	1.024 (0.748)	0.001 (0.022)	0.022 (0.065)
(log)Capital expenditure per worker	0.071** (0.029)	0.042 (0.033)	0.045*** (0.003)	0.024*** (0.003)	0.268*** (0.045)	0.185** (0.076)	0.054*** (0.006)	0.025*** (0.006)
(log) R&D per worker	-0.009 (0.037)	0.079 (0.055)	0.032*** (0.005)	0.001 (0.007)	0.016 (0.054)	0.141 (0.171)	0.045*** (0.007)	0.012 (0.016)
(log)Expenditure on computer and advertising services as a proportion of total purchases	0.130*** (0.030)	0.200*** (0.043)	0.011** (0.005)	0.006 (0.007)	0.110** (0.043)	0.117 (0.075)	-0.013** (0.006)	0.006 (0.010)
(log) Hours worked	-0.083 (0.446)	1.437 (1.108)	0.956*** (0.083)	-0.168 (0.125)	1.454** (0.728)	1.609 (2.982)	1.079*** (0.123)	-0.633*** (0.242)
Proportion in a temporary job	-3.803*** (1.221)	-2.562 (1.766)	-0.663*** (0.167)	0.022 (0.180)	-4.845** (1.955)	-1.045 (5.227)	-0.657** (0.277)	-0.549* (0.327)
Staff non proficient at t-1	-1.392* (0.812)	-1.499* (0.832)	0.098 (0.090)	0.076 (0.086)	-1.052 (2.447)	-5.447 (4.290)	-0.901*** (0.215)	-0.463** (0.213)
Foreign owned	0.079 (0.072)	-0.132 (0.108)	0.095*** (0.010)	-0.003 (0.011)	-0.174 (0.149)	0.306 (0.418)	0.239*** (0.019)	-0.002 (0.029)
Share of part-time workers	-0.644*** (0.171)	-0.247 (0.195)	-0.488*** (0.030)	-0.060** (0.025)	-0.501 (0.313)	0.106 (0.464)	-0.396*** (0.043)	-0.003 (0.040)
Multiple Local Units	-0.071 (0.119)	0.169 (0.200)	0.012 (0.013)	-0.035** (0.018)	0.143 (0.112)	0.295 (0.465)	-0.016 (0.019)	-0.085** (0.035)
Observations	22,540	22,540	22,716	22,716	9,960	9,960	9,967	9,967
Number of reporting units	6,850	6,850	6,890	6,890	5,145	5,145	5,146	5,146

Note: Standard errors clustered at the industry level in parentheses. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level; All regressions include controls for company age bands, legal status, region and year, employment size bands and SIC sections. Grey shading identifies measures aggregated at the industry level.
Source: London Economics based on data from ABS, ILR, IDBR, LFS and ESS

Conclusions and recommendations

This entire analysis has been built on combining a range of very different data sources to understand the impact of publicly funded training on industry-level and firm-level productivity. As such, the **conclusions** of the analysis must consider a range of different factors, including not just the results from the econometric analysis, but also the quality and coverage of the matching process. **Recommendations** for further analysis must also consider both of these factors when assessing the feasibility of repeating or extending this analysis, but also the likelihood that any limitations that have been identified will ever be addressed sufficiently.

Quality and coverage of the matching process

The ILR-EDS information covering all publicly funded education and training in England is collected for a very specific reason – namely, to provide a track of the Further Education & Skills training that is provided by education providers, which is then linked to subsequent payment for the provision of these services. In contrast, the information contained within the Inter-Departmental Business Register is a record of all businesses, and is one of the key sources from which the Office of National Statistics samples enterprises to generate a statistical picture of business activity and behaviour. The two data sources with company information (EDS and IDBR) are quite different in terms of coverage, type of information provided, unit of analysis (site level and statistical units) and data quality. Thus, apart from direct match based on CRN (when available in the EDS), there is no automatic matching procedure available.

We implemented a hugely computationally intensive methodology to exploit as much of the information available in each data set, but also the hierarchical structures in each of the data sets. Over 20 stages of analysis, the approach involved matching entities from the EDS using a range of characteristics – starting with the most certain, uniquely identifiable firm characteristics, moving towards increasingly probabilistic criteria.

Considering **all** stages, of the **533,033** entities in the ILR-EDS data set, **403,493** were matched (**75.7%**) with **129,543** unmatched (**24.3%**). A manual check of the accuracy of the match was undertaken for a sample within each matching stage and suggested that approximately **397,485** of these matches were 'acceptable' matches (**98.5%**), with **6,008** being mistakenly matched or **false-positives** (**1.5%**). Given that this dataset would be used repeatedly and the quality of the match was of greater importance than the breadth of match, it was decided to tighten the quality threshold. As such, **36,327** matches were omitted from the data set relating to some of the fuzzy matching stages, of which approximately **5,026** (**13.8%**) were assessed to be of poor quality, although **31,301** (**86.2%**) were likely to be acceptable.

The overall result was that **367,166** of all EDS entities engaged in publicly funded training were matched with the IDBR (corresponding to **68.9%**). However, the manual check suggested that approximately **367,166** of these matches were acceptable matches (**99.7%**), with only an estimated **982** being false positives (**0.3%**).

The overall result suggests that the matching process was successful in terms of the quality of the match; however, further analysis and strategies could be

employed to increase the breadth of match without affecting quality. We recommend establishing a **regular matching exercise where the latest ILR edition is matched to the latest IDBR extract, to generate a "live" coverage of all firms engaging in publicly funded training**. We also recommend undertaking an in-depth review of unmatched cases to assess why some entries are left unmatched. For instance, are some entities left unmatched because of limited quality in the information available or provided or are some firms are likely to be too small (in terms of VAT threshold) to be covered by the IDBR? Furthermore, in terms of the overall quality of the match, what will be the cost in terms of the incidence of false positives.

Econometric analysis

There were two different strands of econometric analysis undertaken – the first to assess the impact of training on productivity and labour costs using the matched EDS-IDBR data at **industry-level** – and a second replicating the approach used at the industry-level for **firm-level** analysis.

At **industry-level**, the analysis suggests that raising overall training intensity (i.e. the proportion of workers in receipt of **any** form of training excluding Health and Safety and Induction training) by **1 percentage point** is associated with an increase on productivity of about **0.74%** and around **0.36%** on the wage bill. However, it was not possible to identify any unambiguous impact of publicly funded training leading for a formally recognised qualification on either the productivity or wage-bill measure. Although these estimates in relation to any form of training are in line with previous analyses using training intensity at the industry level (e.g. Dearden *et al* (2005)), **the result is never statistically significant in the specification controlling for the endogeneity of training decisions** (i.e. whether training leads to higher productivity or whether more productive firms are likely to engage more in training).

At **firm-level**, there are large evidence gaps in the data. Although there are a number of positive estimates of the relationship between (publicly funded) training and firm level outcomes, the lack of firm-level data for key variables (i.e. capital stock and privately funded training), as well as inconsistencies in relation to employment measures and the length of the time series currently available (4 years), make the firm-level results **very speculative** at this stage. Also, the estimates only apply to firms repeatedly available over time in the ABS (large and medium size firms).

Recommendations and future research options

Improving data

As the analysis presented here is entirely dependent on the quality of the data available, there are several recommendations in this respect

- *Evidence Gaps (Capital Stock)*: currently no estimate for the capital stock series is available for the analysis at firm-level in the ONS' VML, where data from the ABS on productivity and other key measures is held. **We recommend liaising with the ONS and other data providers (UKDS) to obtain updates on the availability of the capital stock series estimates at the firm level.**
- *Data inconsistencies (Employment)*: the Business Register and Employment Survey (BRES) is the official source of employee and employment estimates used by the

ONS. However, the ONS currently advises caution when combining ABS data with BRES estimates of employment due to differences in sample selection periods, sample designs and questionnaire reference periods. **We recommend liaising with the ONS to getter a better understanding on the extent of the bias generated by these methodological differences at the firm level (i.e. whether the bias is likely to be substantial or negligible, thereby allowing ABS data at the firm level to be combined with BRES data on employment).**

- *Lack of consistent information and metadata for employment data (imputation)*: the other issue related to employment data is related to how employment is estimated in the IDBR. Employment estimates in the IDBR are derived from BRES, the Short Term Employment Survey, PAYE and imputation from VAT returns. BRES does not cover all IDBR or ABS units. Although it is the case that when an enterprise is selected in BRES, all constituent local units should be selected, there may be gaps in the data for some local units (so that the total at the enterprise or reporting unit level may not be entirely accurate). Improved metadata reporting the source of employment estimate at the local unit level (whether it is BRES or imputed etc.) would help the assessment of the quality of the employment data at the firm level. We understand that the ONS are currently looking into the various issues associated with employment estimates at the firm level and would advise liaising with the ONS to gain a better understanding on possible improvements to employment data and metadata.
- *Lack of most recent training data and data availability (ESS)*: another crucial aspect is the availability of information on all training (or privately funded training) at the firm level. Firm level information on training, skills, skills gaps and other variable is collected by the UKCES' Employer Skills Survey. The latest edition of the ESS (2013) is not available yet in the VML for the firm level analysis. We recommend that the ESS 2013 and future editions of the ESS (the 2015 ESS is currently underway) are matched to the IDBR and subsequently to the ABS. From previous ESS editions we estimated that around 4,000 reporting units in the ABS may be linked to the ESS for each edition. The number of reporting units available in both the ESS 2009 and ESS 2011 is smaller (around 1,000).
- *Estimates of publicly funded training at the industry level*: it may be useful to explore whether there is any alternative source of information on Apprenticeships (for instance, the BIS Apprenticeship Survey) allowing industry level analysis on publicly funded training. Also, it may be worth further consideration of whether the EDS information on sector may be used (at least for entities not identified in the IDBR) to avoid the gap caused by the 30% of entities left unmatched.

Issues that are unlikely to be ever solved

When considering the potential to undertake an analysis of the impact of training on firm level productivity, the fact that small firms are rarely sampled in the ABS, as well as the fact that small firms engaged in publicly funded training are less likely to be matched with the IDBR in the first instance, means that any analysis of firm-level productivity will only ever be able to consider impacts amongst medium to large sized firms with any degree of certainty.

Future analysis

If the various improvements to the quality of the underpinning data sources can be achieved, then there are a number of useful analyses that could be achieved. In the **short**

run, rather than an aggregate analysis, it is possible to replicate the industry-level analysis to consider whether *specific types* of publicly funded education and training have an impact on industry-level productivity and salary costs. Specifically, given the fact that there is essentially complete information in relation to the nature of the learning aims undertaken by learners in the ILR, an analysis of the impact of learning aims focused on Maths or English (for instance) could be achieved.

To undertake any legitimate future analysis at the firm-level, all of the issues relating to missing or inconsistent data and the length of the available time series would need to be resolved; therefore we recommend repeating the analysis when:

- Updated estimates of the capital stock series at the firm level become available;
- There is sufficient clarity on the quality of employment estimates available from BRES and in the IDBR;
- Further ESS become available and ESS firms are matched to the IDBR and the ABS;
- Further ILR editions are available and information on publicly funded training is matched to the IDBR and the ABS;

The analysis should cover all reporting units consistently available over time (a sample of large and medium-size firms) and should be undertaken on the sample of all available reporting units in the ABS, as well as on the restricted sample with information available from the ESS on total training undertaken.

The combination of these factors would suggest that it may be worth revisiting the analysis when a longer time series for all relevant data sets become available (ABS, ESS and ILR). Although the **2015** ABS and ESS data sets and the 2015/16 ILR should all be available for analysis at the beginning of 2017, further robustness would be introduced adding also the **2017** ABS and ESS data sets (and the ILR 2017/18): however, these datasets are not likely to be available before 2018/19.

References

Arellano, M., Bond, S.R. (1991). "*Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations*". Review of Economic Studies 58, 277-297

Arellano, M., and Bover, O. (1995). "*Another look at the instrumental variable estimation of error-components models*". Journal of Econometrics 68: 29-51.

Blundell, R & Bond, S. 1998. *"Initial conditions and moment restrictions in dynamic panel data models*". Journal of Econometrics, Elsevier, vol. 87(1), pages 115-143.

Colombo, E, & Stanca, L. (2008). '*The Impact of Training on Productivity: Evidence from a Large Panel of Firms*', Working Papers 134, University of Milano-Bicocca, Department of Economics.

Conti, G, (2005). "*Training, productivity and wages in Italy*". Labour Economics, Elsevier, vol. 12(4), pages 557-576, August

Dearden, L., Reed, H. and Van Reenen, J. (2005). *"The Impact of Training on Productivity and Wages: Evidence from British Panel Data"*. IFS Working Papers W05/16, Institute for Fiscal Studies.

Galindo-Rueda, F. & Haskel, J., 2005. "*Skills, Workforce Characteristics and Firm-Level Productivity: Evidence from the Matched ABI/Employer Skills Survey*" IZA Discussion Papers 1542, Institute for the Study of Labor (IZA).

Konings, J. & Vanormelingen, S., (2010). '*The Impact of Training on Productivity and Wages: Firm Level Evidence*', IZA Discussion Papers 4731, Institute for the Study of Labor (IZA).

Roodman, D. (2009)."*How to do xtabond2: An introduction to difference and system GMM in Stata*". Stata Journal 9(1): 86-136

Ryan, P. (2011)."*Apprenticeship: between theory and practice, school and workplace*". Economics of Education Working Paper Series 0064, University of Zurich

Annex 1: Further information on data sets used in the analyses

The Annual Business Survey (ABS)

The **Annual Business Survey** (previously Annual Business Inquiry) is the main structural business survey conducted by the Office for National Statistics (ONS). The Annual Business Survey covers the production, construction, distribution and service industries, which represent about two-thirds of the UK economy in terms of Gross Value Added (GVA).

The financial variables covered include:

- Turnover,
- Purchases,
- Employment costs,
- Capital expenditure and stocks, and
- Approximate gross value added (aGVA), which is calculated as an input into the measurement of gross domestic product (GDP).

The information in the ABS is collected at **reporting unit** (RU) level, and the dataset covers slightly less than **50,000** reporting units each year. Around **98%** of the entries in the ABS are single-RU enterprises, i.e. the enterprise corresponds to the reporting unit.

In general, the sampling structure of the ABS is designed so that in any year, half of the sample will be newly-selected, and half will have been selected in the previous year as well. However, **for the largest size bands (i.e. employment of 250 or more), all enterprises are selected every year (so the ABS can be seen as a census of larger enterprises)**. At the other end of the spectrum, any business with employment of between 0 and 9 individuals that has been selected in a particular year's survey will only be selected for that single year, and will not be reselected for *at least* three years following initial selection (with few exceptions).

Respondents to the Annual Business Survey (ABS) are required to return data for a number of financial variables, ideally for the most recent calendar year (January to December). However, to reduce the burden on survey respondents, reporting units are given the option to return data covering a business year ending on any date in a specified range. For example, for the 2013 survey year, the range for acceptable business year end dates was between 6 April 2013 and 5 April 2014. Around **36%** of respondents typically return their information based on the calendar year.

The Business Register and Employment Survey (BRES)

The **Business Register and Employment Survey** (BRES) is the official source of employee and employment estimates used by the ONS and can be used to derive employment estimates at varying industrial and geographical levels. The BRES sample is addressed to approximately **80,000** businesses each year, covering around **500,000** local

units. The BRES design covers all complex and large businesses each year. For example, any business with sites in more than one region or industry is covered each year. Additionally, any business with employment of over 100 is covered each year. The employment information requested refers to a reference date in mid-September. The information from BRES is used to update the information on employment in the IDBR, together with information from PAYE and VAT (imputation based on turnover) for smaller enterprises.

The UKCES' Employer Skills Survey

The **UKCES' Employers Skills Survey (ESS)** enquires about available and required skills, skills gaps, training decisions, enterprise organisation etc. The ESS provides information on all training undertaken (publicly and privately funded) as well as information on the existing skills and occupational structure. The ESS is a biennial survey and typically covers around 80,000 establishments (some of them be part of a larger organisation). Although the ESS 2013 has already been published, the latest edition currently available to researchers in the ONS' Virtual Microdata Laboratory (where the firm level analysis took place) is the ESS 2011. The match rate to the IDBR for the ESS 2011 is around 50%.

The survey was first conducted at UK wide level in 2011, developing from a series of surveys conducted in each of the countries of the UK. The ESS explores the skills challenges that employers face both within their existing workforces and in terms of bringing in new skilled labour, the levels and nature of training investment, recruitment of young people and education leavers and the relationship between skills challenges, training activity and business strategy. The study is conducted at the establishment level, rather than at the enterprise level. Sole traders and establishments with just one employee and no working proprietors are not covered by the ESS.

Annex 2: Construction of training variables in the ILR

The Individualised Learner Record is organised at the learner and learning aim level. The dataset provides information on a variety of personal and course characteristics and, for training undertaken through the employer, there is also an employer identifier (variable A44 corresponding to the EDS identifier).For the current exercise we used ILR collections between 2010/11 and 2013/14.

After appending all different ILR editions to create a single dataset, we undertook the following **preliminary cleaning steps:**

- In total, there were around **19.7 million** observations at the learning aim level (i.e. each learner may have multiple entries) across the four years, corresponding to between **4.4 million** and **5.6 million** observations in each academic year;
- However, almost **2.9 million** observations (mostly from the 2010/11 academic year) corresponded to unusable A44 identifiers (i.e. “999999999” and “-1”) and they were removed from the dataset, leaving around **16.8 million** observations

After these preliminary steps, we undertook a series of recoding steps to generate training measures at the learner level for each calendar year[46], and then aggregated the measures at the firm level using the employer identifier. The recoding steps undertaken can be summarised as follows:

- Remove records that ended before 1st January 2010;
- Remove records with zero length (i.e. the course start and end date are the same);
- Remove qualifications at “other” level;
- Identify the earliest start date and latest end date for overlapping spells (so that we considered the total length of time the individual spent in training);
- For continuing learners in each academic year, the end date (originally coded as missing if the learner was continuing) was replaced with the end of the academic year (31st July);
- Generate the number of days spent in training for each calendar year;
- Learners with at least one training day in the calendar year were considered to be in training in that calendar year (training variable=1);
- No adjustment was made for withdrawals/non completion (i.e. training undertaken is considered irrespectively of completion and achievement);
- Apprenticeship and non-apprenticeship aims (at all levels) were merged in one training variable to avoid double counting (when the same learner has multiple apprenticeship and non-apprenticeship training spells in the same calendar year);
- We allowed for double counting of learners having training spells with two different employers (e.g. if an individual spends 3 months training with company A and then

[46] We switched from the academic to the calendar year to provide consistency with subsequent matching with ONS surveys (as the ABS is mainly organised according to the calendar year)

moves to company B and has another 6 months of training, he will be counted twice – once with company A and once with company B);

- Given the structure of the ILR, we have full information for the calendar year 2011, 2012 and 2013, but only partial information for the calendar years 2010 (the ILR 2009/10 was not part of the data we received) and 2014 (the ILR 2014/15 is not yet available);

Annex 3: Alternative training measures versus industry employment

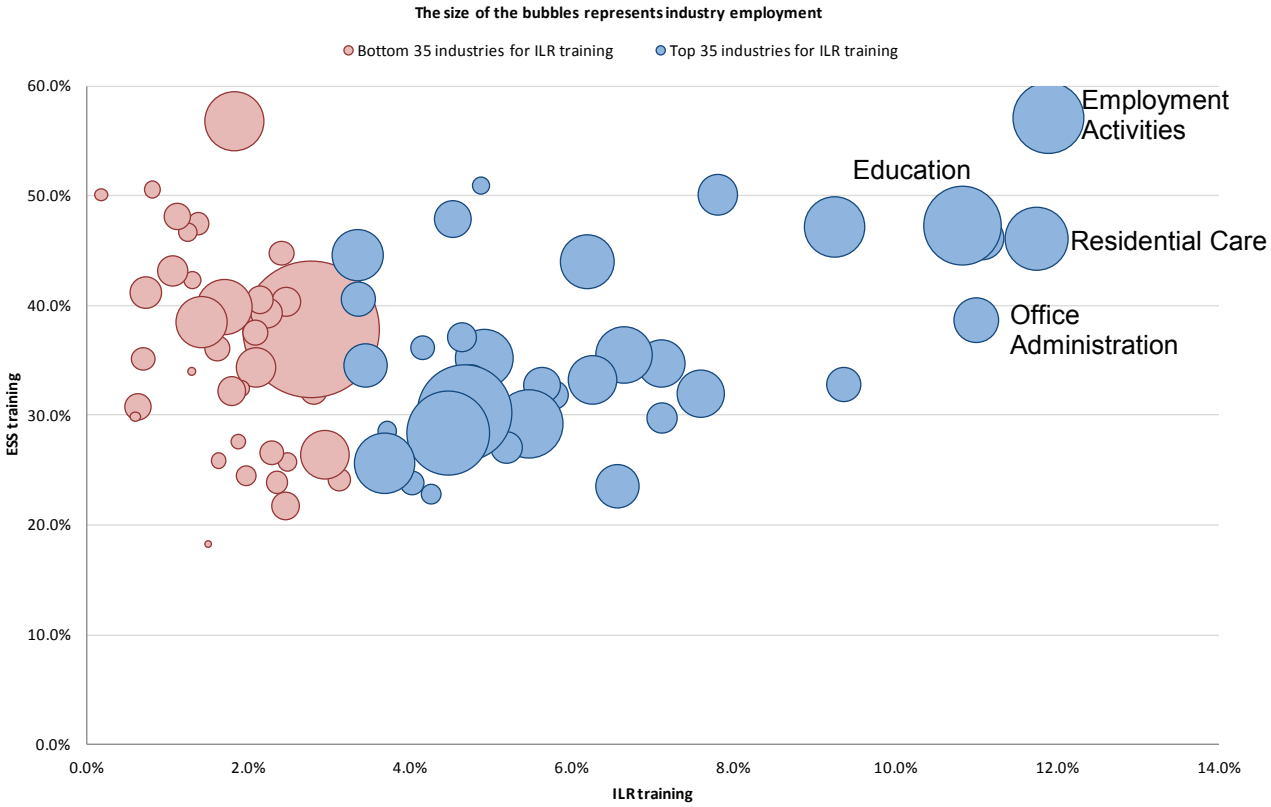


Figure 11: Proportion of workers in receipt of different types of training and employment by industry

Note: Publicly funded training is based on the ILR definition (for entities matched to the IDBR; the ESS measure of training is total training undertaken excluding the volume of Health and Safety and Induction training; employment represents the BRES estimate by industry

Source: London Economics based on ILR, EDS, IDBR, LFS, ESS and BRES

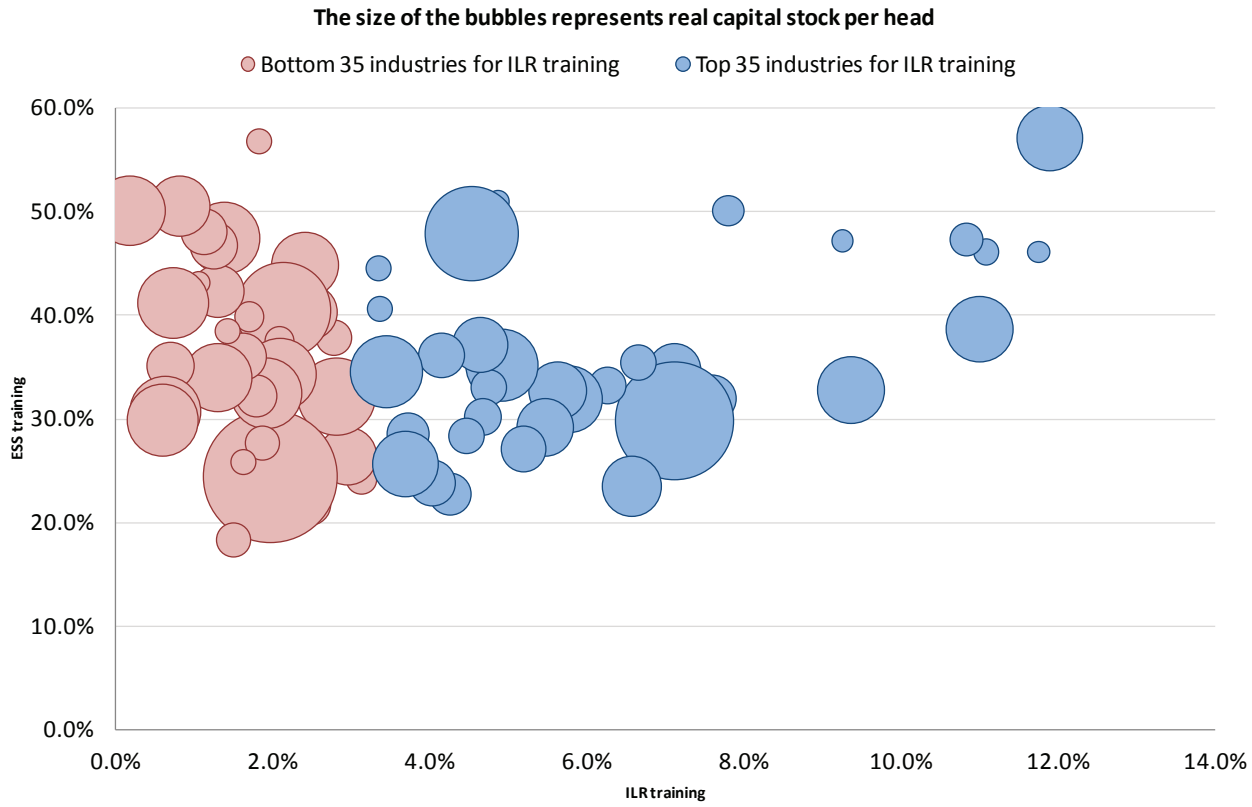


Figure 12: Proportion of workers in receipt of different types of training and capital stock by industry

Note: Publicly funded training is based on the ILR definition (for entities matched to the IDBR; the ESS measure of training is total training undertaken excluding the volume of Health and Safety and Induction training; employment represents the BRES estimate by industry

Source: London Economics based on ILR, EDS, IDBR, LFS, ESS and BRES

Annex 4: Training incidence at industry level

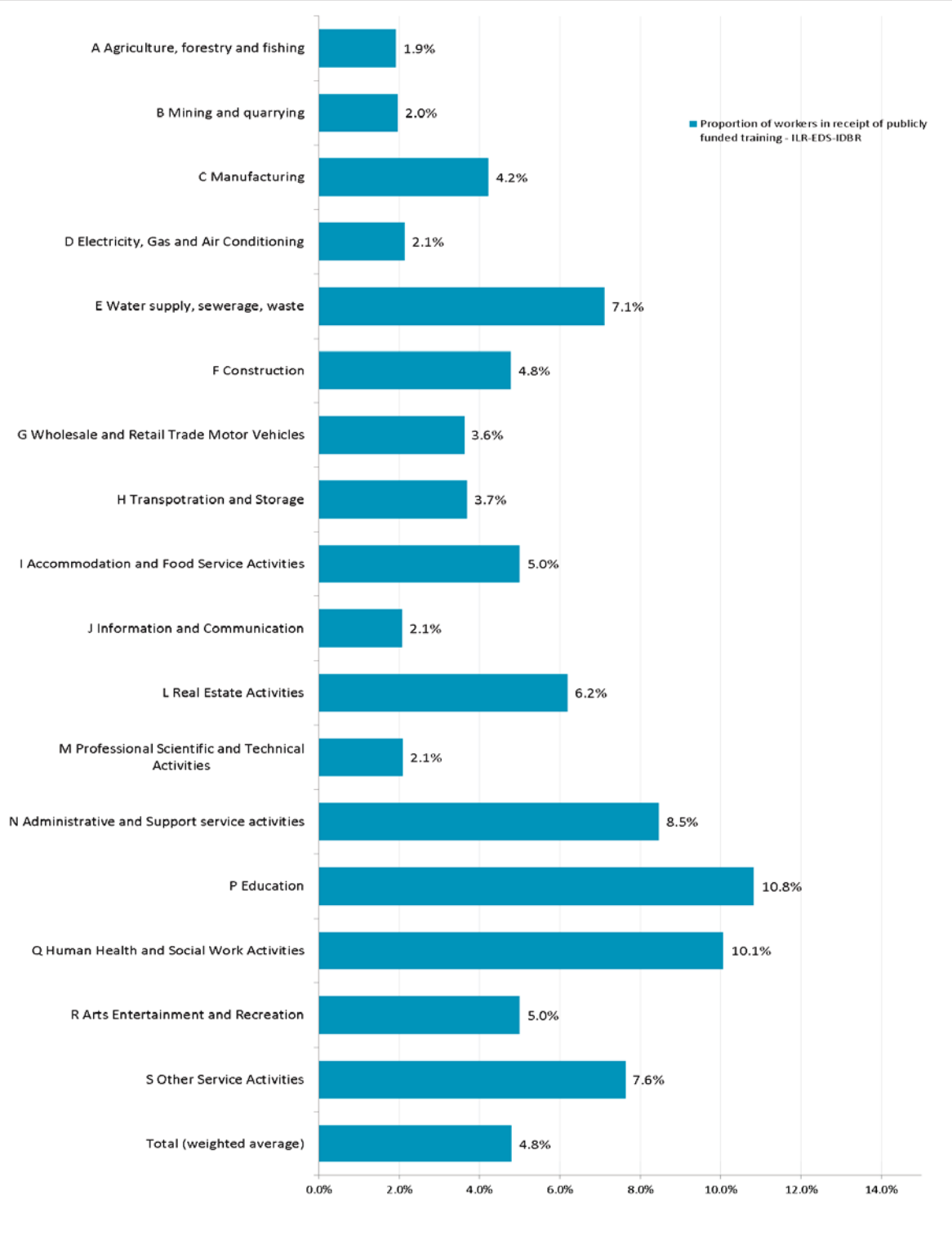


Figure 13: Proportion of employees in receipt of publicly funded training by main industry (ILR-EDS-IDBR)

Note: Publicly funded training is based on the ILR definition (for entities matched to the IDBR). Only SIC codes covered by the ABS are shown. ***Source: London Economics based on ILR, EDS, IDBR and LFS***

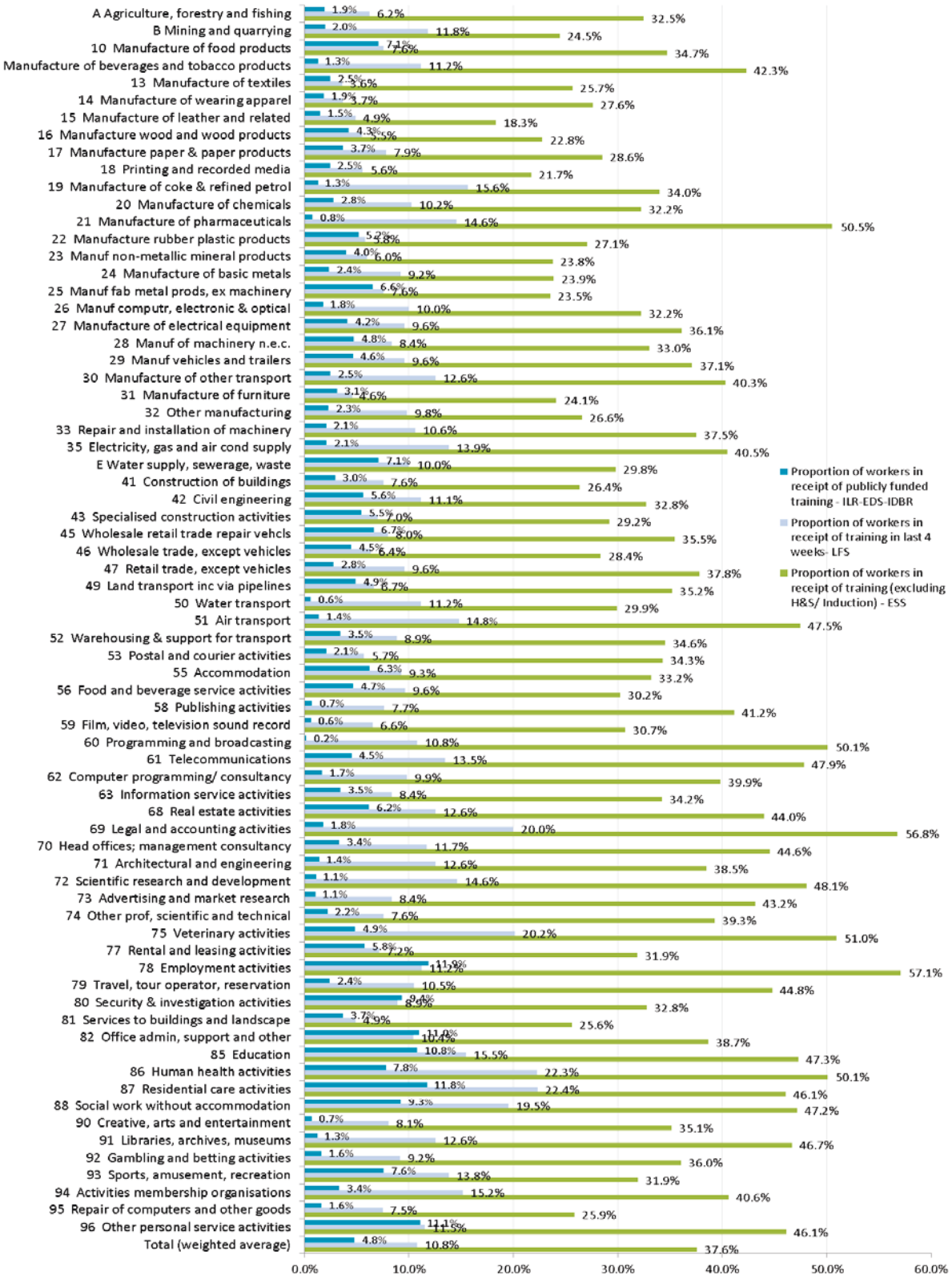


Figure 14: Proportion of employees in receipt of different types of training by main industry

Note: Publicly funded training is based on the ILR definition (for entities matched to the IDBR); the LFS measure of training reports job-related training undertaken over the last four weeks; the ESS measure of training is total training undertaken excluding the volume of Health and Safety and Induction training. Only SIC codes covered by the ABS are shown. ***Source: London Economics based on ILR, EDS, IDBR, LFS and ESS***

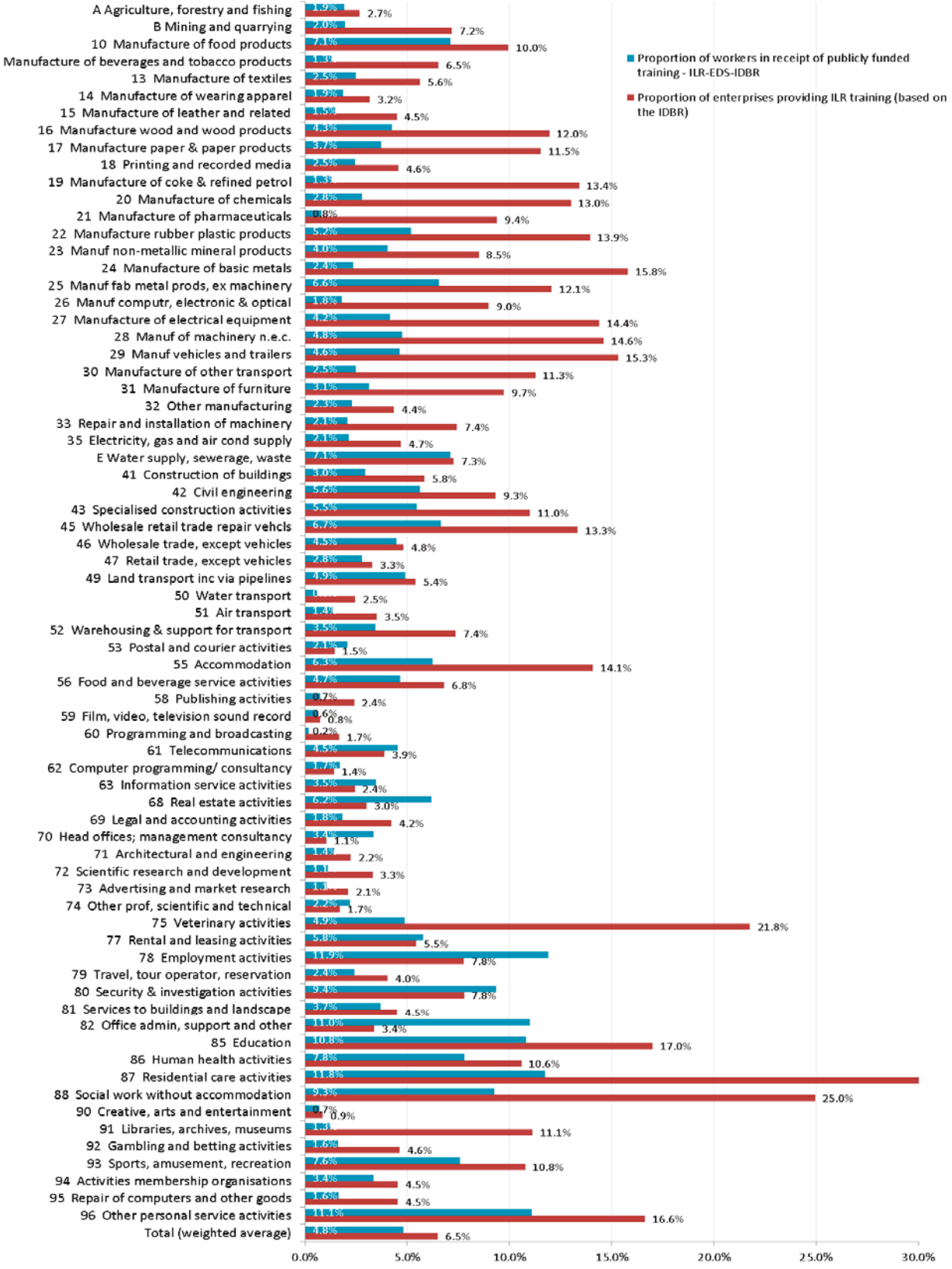


Figure 15: Proportion of employees and enterprises engaging in publicly funded training by industry

Note: Proportion of employees undertaking publicly funded training and proportion of enterprises engaging in publicly funded training. Only SIC codes covered by the ABS are shown. ***Source: London Economics based on ILR, EDS, IDBR and LFS***

Annex 5: Further analysis at the industry and firm level

Table 10: ILR training and firm level characteristics in the ABS and ILR training

Variable	Proportion of ABS firms undertaking publicly funded (ILR) training		% of workers undertaking (ILR) publicly funded training	Variable	Proportion of ABS firms undertaking publicly funded (ILR) training		% of workers undertaking (ILR) publicly funded training
Year	**Obs**	**%**	**%**	**Legal status**	**Obs**	**%**	
2010	8,149	16.9%	1.8%	Company (including Building Society)	31776	22.2%	2.4%
2011	9,463	19.9%	2.5%	Sole Proprietor	693	3.0%	1.2%
2012	9,924	20.7%	2.4%	Partnership	874	6.3%	1.3%
2013	10,024	21.3%	2.2%	Public	280	60.3%	3.5%
Total		**19.7%**	**2.2%**	Central Government	97	37.7%	0.4%
Company size	**Obs**	**%**		Local Authority	71	55.5%	2.0%
1-9	2,584	2.6%	1.3%	Non-profit body or mutual association	3769	37.2%	3.1%
10-49	7,340	18.1%	3.0%	**Government Office Region**	**Obs**	**%**	
50-249	8,931	40.3%	3.4%	East Midlands	3500	26.6%	3.3%
250-499	8,343	59.4%	3.7%	East of England	3531	19.3%	2.2%
500+	10,362	76.1%	3.3%	London	4749	15.6%	1.3%
SIC section				North East	1695	32.8%	4.7%
A - Agriculture, forestry and fishing	69	6.2%	1.1%	North West	5287	28.7%	3.7%
B - Mining and quarrying	155	18.7%	1.4%	Scotland	901	3.8%	0.2%
C - Manufacturing	8395	32.7%	2.8%	South East	5754	19.5%	1.9%
D - Electricity, gas, steam and air conditioning supply	150	35.8%	2.6%	South West	3640	22.0%	2.7%
E - Water supply, sewerage, waste management and remediation activities	331	29.9%	3.5%	Wales	482	7.7%	0.3%
F - Construction	4270	19.3%	3.7%	West Midlands	4126	26.5%	3.2%
G - Wholesale and retail trade; repair of motor vehicles and motorcycles	6882	17.0%	1.7%	Yorkshire & Humberside	3895	28.1%	3.5%
H - Transport and storage	1584	21.4%	2.1%	**Foreign**			
I - Accommodation and food service activities	1809	19.1%	1.5%	No	30556	17.4%	2.2%
J - Information and communication	918	8.2%	0.6%	Yes	7004	47.1%	2.6%
L - Real estate activities	871	13.1%	1.4%	**Company age band**			
M - Professional, scientific and technical activities	2164	8.7%	0.8%	Up to 5 years	2283	4.7%	1.4%
N - Administrative and support service	2689	17.5%	1.7%	6-10 years	3098	10.0%	2.1%
P - Education	1917	41.5%	3.3%	11-20 years	9372	20.3%	2.6%
Q - Human health and social work activities	3346	45.9%	7.3%	20+ years	22807	35.0%	2.6%
R - Arts, entertainment and recreation	1173	21.9%	2.3%				
S - Other service activities	837	12.5%	3.0%				

Note: All tables are unweighted and report the raw number of observations and proportions available in the ABS. Company size: total employment. The proportion of firms refers to the proportion of firms engaging in ILR training in each group. Company age band refers to the enterprise age.

Source: London Economics based on data from ABS, ILR, IDBR, LFS and ESS

Table 11: Firm level characteristics by ILR training status and employment size band

	Log of GVA		Log of capital expenditure per worker		Expenditure on computer and advertising as a proportion of total purchases		Share of part-time workers	
Company size (employment)	Non-training	Training	Non-training	Training	Non-training	Training	Non-training	Training
1-9	8.8	9.6	0.37	0.50	4.0%	3.5%	28.2%	27.2%
10-49	9.6	9.9	0.54	0.59	3.7%	3.1%	26.0%	23.3%
50-249	9.8	10.0	0.80	0.89	4.5%	3.5%	19.8%	17.6%
250-499	9.4	9.4	0.95	0.98	6.7%	4.6%	22.7%	21.9%
500+	9.1	9.2	1.00	1.10	7.0%	4.9%	24.7%	25.6%
Total	9.1	9.6	0.48	0.88	4.1%	4.1%	26.5%	22.5%

Note: All tables are unweighted and report the raw number of observations and proportions available in the ABS. Company size: total employment. All nominal variables are deflated using Producer Price deflators.
Source: London Economics based on data from ABS, ILR, IDBR, and BRES

Alternative model specifications at Industry Level

For completeness, we undertook a number of alternative model specifications using different measures of training to assess the sensitivity of the results.

The analysis that has been presented to date contains the proportion of employees in receipt of training at industry level, but does not control for these other training measures. Below, we have presented the findings replicating Table 6 and Table 7 from the main report, but adding also two additional training measures for each data source used (ESS and ILR) along with the proportion of employees receiving training:

- the average number of training days for employees receiving training[47];
- the proportion of enterprises engaging in training (ESS training or publicly funded training);

The proportion of employees receiving training is referred to in Table 12 as 'trained proportion ILR'. The numerator is the total number of employees undertaking ILR training aggregated at the industry level (as defined in the main report) while the denominator is total industry employment (from LFS estimates).

The implication of adding in the additional average training measures is that:

- the ESS coefficient for the proportion of employees receiving training is generally slightly higher compared to those presented in Table 6 of the main report;
- The coefficient for the ILR training proportion is also constantly larger than the corresponding coefficient presented in Table 6 in the main report and almost always positive (although never statistically significant);
- The coefficient associated with the proportion of firms engaging in training is consistently negative for both ESS and ILR training and the size can be quite large (especially for ILR training);
- Although this may suggest a negative association between the proportion of enterprises engaging in publicly funded training and productivity (although with no causal effect), the interpretation is only valid keeping constant the value of all other covariates (including the value for the proportion of employees receiving training). Furthermore, the two variables (proportion of employees and proportion of enterprises) are highly correlated (coefficient of correlation between 0.55 and 0.6)
- These measures do not cover entities that were not matched in the IDBR;
- The coefficient for the average number of days is typically quite small (due to the scale used, e.g. number of days) and never statistically significant;

The various coefficients in the GMM specifications are quite large in absolute terms in the productivity regressions and in line with the coefficients presented in Table 14 in the wage regressions (although never statistically significant).

[47] The average days in training variable at the industry level are generated as total number of days divided by total employees receiving training at the industry level (so not total employment overall, but total number of employees in training). So, if there were 100 employees in an industry, of which 40 received training, and there were 160 training days undertaken, the average industry level measure average days in training would stand at 4 (and not 1.6)

Table 12: Results of industry-level modelling – alternative model specification

Dependent variable	Log GVA per worker						Log employment costs per worker					
Controlling for (training variables	*ESS only*	*ESS only*	*Both*	*Both*	*ILR only*	*ILR only*	*ESS only*	*ESS only*	*Both*	*Both*	*ILR only*	*ILR only*
Estimator	*Random Effects*	*Fixed Effects*	*Random Effects*	*Fixed Effects*	*Random Effects*	*Fixed Effects*	*Random Effects*	*Fixed Effects*	*Random Effects*	*Fixed Effects*	*Random Effects*	*Fixed Effects*
Training intensity ESS	0.890** (0.388)	0.793** (0.365)	0.843** (0.396)	0.716* (0.378)			0.461** (0.226)	0.379* (0.197)	0.415** (0.209)	0.333* (0.190)		
Average number of training days ESS	-0.001 (0.006)	0.001 (0.006)	-0.003 (0.007)	-0.001 (0.006)			0.000 (0.004)	0.002 (0.004)	-0.002 (0.004)	0.001 (0.004)		
% of establishments undertaking training ESS	-0.698 (0.435)	-0.422 (0.358)	-0.673 (0.416)	-0.436 (0.372)			-0.099 (0.226)	-0.152 (0.228)	-0.061 (0.221)	-0.147 (0.228)		
Trained proportion ILR			-0.006 (1.211)	0.714 (0.990)	**0.182 (1.160)**	**0.842 (0.906)**			0.530 (0.831)	0.973 (0.901)	**0.601 (0.807)**	**1.066 (0.871)**
Average number of training days ILR			0.001 (0.001)	0.001 (0.001)	0.001 (0.001)	0.001 (0.001)			0.001 (0.001)	0.000 (0.001)	0.001 (0.001)	0.000 (0.001)
% of enterprises undertaking training ILR			-1.631* (0.876)	-1.379 (1.193)	-1.689* (0.897)	-1.437 (1.223)			-0.927** (0.444)	-1.032 (0.683)	-0.937** (0.448)	-1.110 (0.714)
Capital stock per head (log)	0.210*** (0.053)	0.239 (0.240)	0.182*** (0.058)	0.220 (0.248)	0.182*** (0.058)	0.257 (0.252)	0.041 (0.030)	-0.088 (0.108)	0.025 (0.031)	-0.114 (0.119)	0.028 (0.029)	-0.097 (0.118)
R&D expenditure per head at t-1 (log)	0.036 (0.022)	0.005 (0.021)	0.036 (0.022)	0.012 (0.024)	0.036 (0.022)	0.012 (0.024)	0.025** (0.012)	-0.001 (0.017)	0.024** (0.012)	0.001 (0.016)	0.024* (0.013)	0.001 (0.017)
Usual hours worked (log)	1.132 (0.735)	-0.870 (0.588)	1.217 (0.771)	-0.878 (0.622)	1.206 (0.765)	-1.000 (0.643)	0.627* (0.333)	-0.259 (0.426)	0.666* (0.346)	-0.247 (0.443)	0.669* (0.349)	-0.283 (0.450)
Proportion in a temporary job	-2.890** (1.275)	-2.034** (1.016)	-2.993** (1.270)	-2.054** (1.016)	-2.807** (1.311)	-1.883* (1.064)	-0.840* (0.446)	-0.353 (0.355)	-0.916** (0.448)	-0.344 (0.354)	-0.822* (0.463)	-0.284 (0.355)
Highest qualification:L3 and above (including apprenticeships)	-0.117 (0.343)	-0.311 (0.385)	-0.115 (0.323)	-0.329 (0.375)	-0.139 (0.335)	-0.362 (0.385)	0.060 (0.208)	-0.206 (0.224)	0.055 (0.198)	-0.220 (0.221)	0.093 (0.197)	-0.234 (0.225)
Occupation: Managers and professionals	0.253 (0.228)	0.117 (0.237)	0.113 (0.249)	0.026 (0.262)	0.024 (0.293)	-0.075 (0.266)	0.300* (0.158)	-0.036 (0.133)	0.234 (0.164)	-0.115 (0.148)	0.223 (0.172)	-0.166 (0.170)
Proportion with business plan, training plan and training budget	-0.175 (0.401)	-0.243 (0.489)	-0.031 (0.409)	-0.290 (0.491)	-0.308 (0.377)	-0.507 (0.455)	0.451* (0.260)	0.323 (0.233)	0.542** (0.247)	0.309 (0.223)	0.538*** (0.208)	0.235 (0.195)
Proportion of staff not fully proficient at t-1	-0.928*** (0.292)	-0.990*** (0.276)	-0.919*** (0.317)	-0.994*** (0.291)	-0.811*** (0.309)	-0.884*** (0.284)	-0.197 (0.192)	-0.284* (0.163)	-0.187 (0.190)	-0.288* (0.168)	-0.136 (0.183)	-0.230 (0.158)
Small workplace (less than 25 employees)	-0.828*** (0.245)	-0.322 (0.440)	-0.908*** (0.254)	-0.387 (0.435)	-1.013*** (0.260)	-0.487 (0.413)	-0.728*** (0.130)	-0.109 (0.168)	-0.791*** (0.147)	-0.158 (0.167)	-0.830*** (0.143)	-0.196 (0.167)
Age band 16-24	1.594* (0.904)	1.460 (0.902)	1.466* (0.860)	1.345 (0.856)	1.714* (0.904)	1.598* (0.920)	0.485 (0.535)	0.748 (0.467)	0.383 (0.524)	0.660 (0.460)	0.453 (0.530)	0.768* (0.459)
Age band 25-39	2.424*** (0.806)	2.264*** (0.809)	2.323*** (0.786)	2.154*** (0.767)	2.643*** (0.885)	2.381*** (0.826)	0.720 (0.477)	0.842* (0.448)	0.645 (0.479)	0.769* (0.456)	0.729 (0.499)	0.848* (0.448)
Age band 40-54	1.519** (0.713)	1.439** (0.683)	1.469** (0.699)	1.401** (0.644)	1.749** (0.737)	1.616** (0.671)	0.750 (0.512)	0.717 (0.469)	0.722 (0.504)	0.723 (0.456)	0.782 (0.515)	0.804* (0.455)
Tenure 2-10 years	-0.628 (0.570)	-0.525 (0.430)	-0.689 (0.568)	-0.595 (0.437)	-0.606 (0.592)	-0.532 (0.438)	-0.150 (0.259)	-0.166 (0.213)	-0.192 (0.243)	-0.213 (0.207)	-0.172 (0.260)	-0.184 (0.214)

Tenure more than 10 years	-0.190 (0.721)	-0.026 (0.515)	-0.297 (0.707)	-0.078 (0.513)	-0.159 (0.703)	0.030 (0.495)	0.283 (0.360)	0.228 (0.340)	0.246 (0.349)	0.201 (0.338)	0.269 (0.376)	0.236 (0.339)
Proportion female	0.042 (0.274)	0.180 (0.309)	0.173 (0.276)	0.255 (0.325)	0.309 (0.274)	0.370 (0.344)	-0.089 (0.200)	0.237 (0.227)	-0.026 (0.187)	0.281 (0.215)	0.063 (0.180)	0.331 (0.210)
Observations	270	270	270	270	270	270	271	271	271	271	271	271

Note: Standard errors clustered at the industry level in parentheses. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level; All regressions include regional controls and yearly dummies. Omitted category for age band is 55 and over. Omitted category for Tenure is less than 2 years. ***Source: London Economics based on data from ABS, ILR, LFS, BRES and ESS***

Table 13: Results of econometric modelling controlling for Level 3 Apprenticeship training (Fixed and Random Effects models)

Dependent variable	Log GVA per worker						Log employment costs per worker					
Controlling for (training variables)	*ESS only*	*ESS only*	*Both*	*Both*	*ILR only*	*ILR only*	*ESS only*	*ESS only*	*Both*	*Both*	*ILR only*	*ILR only*
Estimator	*Random Effects*	*Fixed Effects*	*Random Effects*	*Fixed Effects*	*Random Effects*	*Fixed Effects*	*Random Effects*	*Fixed Effects*	*Random Effects*	*Fixed Effects*	*Random Effects*	*Fixed Effects*
Training intensity ESS	0.778* (0.398)	0.744** (0.372)	0.784* (0.406)	0.726* (0.372)			0.447** (0.217)	0.365* (0.192)	0.419* (0.219)	0.322* (0.193)		
Trained proportion ILR (L3 Apprenticeships only)			-0.347 (2.713)	1.049 (2.521)	0.649 (2.820)	1.848 (2.668)			1.236 (2.406)	2.425 (2.322)	1.790 (2.476)	2.780 (2.351)
Capital stock per head (log)	0.206*** (0.053)	0.222 (0.240)	0.205*** (0.052)	0.223 (0.240)	0.209*** (0.052)	0.276 (0.248)	0.041 (0.029)	-0.097 (0.106)	0.042 (0.029)	-0.097 (0.107)	0.044 (0.029)	-0.073 (0.110)
R&D expenditure per head at t-1 (log)	0.037* (0.022)	0.006 (0.022)	0.037 (0.023)	0.002 (0.025)	0.036 (0.023)	-0.001 (0.025)	0.025** (0.012)	-0.002 (0.017)	0.024** (0.012)	-0.009 (0.019)	0.023* (0.012)	-0.011 (0.019)
Usual hours worked (log)	1.042 (0.721)	-0.911 (0.603)	1.060 (0.733)	-0.895 (0.600)	1.151 (0.739)	-0.933 (0.607)	0.611* (0.330)	-0.246 (0.431)	0.632* (0.330)	-0.208 (0.427)	0.673** (0.328)	-0.225 (0.429)
Proportion in a temporary job	-2.776** (1.301)	-1.951* (1.023)	-2.792** (1.309)	-1.954* (1.025)	-2.703** (1.352)	-1.871* (1.082)	-0.812* (0.444)	-0.339 (0.343)	-0.834* (0.444)	-0.347 (0.338)	-0.779* (0.469)	-0.310 (0.350)
Highest qualification:L3 and above (including apprenticeships)	-0.239 (0.354)	-0.345 (0.387)	-0.238 (0.353)	-0.331 (0.385)	-0.132 (0.361)	-0.320 (0.396)	0.037 (0.206)	-0.222 (0.225)	0.042 (0.206)	-0.189 (0.214)	0.103 (0.207)	-0.184 (0.218)
Occupation: Managers and professionals	0.172 (0.253)	0.062 (0.231)	0.172 (0.253)	0.056 (0.235)	0.143 (0.270)	-0.012 (0.237)	0.279* (0.160)	-0.068 (0.144)	0.288* (0.164)	-0.083 (0.139)	0.284* (0.171)	-0.113 (0.153)
Proportion with business plan, training plan and training budget	-0.503 (0.351)	-0.401 (0.411)	-0.504 (0.356)	-0.400 (0.410)	-0.456 (0.388)	-0.430 (0.462)	0.402* (0.232)	0.280 (0.209)	0.393* (0.232)	0.280 (0.207)	0.430* (0.228)	0.267 (0.205)
Proportion of staff not fully proficient at t-1	-0.913*** (0.291)	-0.977*** (0.275)	-0.912*** (0.292)	-0.980*** (0.275)	-0.798*** (0.271)	-0.860*** (0.262)	-0.194 (0.188)	-0.266 (0.162)	-0.198 (0.189)	-0.272 (0.166)	-0.137 (0.183)	-0.219 (0.161)
Small workplace (less than 25 employees)	-0.847*** (0.244)	-0.385 (0.422)	-0.847*** (0.245)	-0.389 (0.424)	-0.924*** (0.253)	-0.410 (0.425)	-0.721*** (0.128)	-0.128 (0.165)	-0.741*** (0.133)	-0.137 (0.161)	-0.782*** (0.131)	-0.146 (0.161)
Age band 16-24	1.693* (0.883)	1.496 (0.904)	1.709* (0.881)	1.465 (0.905)	1.900** (0.920)	1.722* (0.946)	0.498 (0.527)	0.754 (0.458)	0.489 (0.529)	0.682 (0.457)	0.582 (0.542)	0.795* (0.462)
Age band 25-39	2.595*** (0.835)	2.352*** (0.821)	2.608*** (0.836)	2.331*** (0.821)	2.776*** (0.888)	2.479*** (0.847)	0.737 (0.481)	0.849* (0.441)	0.740 (0.483)	0.801* (0.444)	0.830* (0.501)	0.866* (0.438)
Age band 40-54	1.683** (0.696)	1.519** (0.675)	1.690** (0.701)	1.510** (0.675)	1.832** (0.739)	1.663** (0.700)	0.765 (0.505)	0.727 (0.462)	0.778 (0.511)	0.707 (0.467)	0.851 (0.533)	0.775 (0.473)
Tenure 2-10 years	-0.566 (0.564)	-0.505 (0.427)	-0.562 (0.587)	-0.524 (0.418)	-0.534 (0.625)	-0.468 (0.421)	-0.143 (0.263)	-0.151 (0.215)	-0.168 (0.270)	-0.197 (0.208)	-0.149 (0.288)	-0.172 (0.213)
Tenure more than 10 years	-0.057 (0.697)	0.037 (0.496)	-0.061 (0.708)	0.033 (0.494)	-0.059 (0.739)	0.073 (0.493)	0.296 (0.370)	0.241 (0.338)	0.294 (0.371)	0.230 (0.330)	0.302 (0.392)	0.248 (0.332)
Proportion female	0.027 (0.270)	0.208 (0.301)	0.028 (0.270)	0.206 (0.302)	0.177 (0.283)	0.281 (0.331)	-0.083 (0.201)	0.242 (0.224)	-0.092 (0.200)	0.239 (0.216)	-0.016 (0.191)	0.272 (0.212)
Observations	270	270	270	270	270	270	271	271	271	271	271	271

Note: Standard errors clustered at the industry level in parentheses. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level; All regressions include regional controls and yearly dummies. Omitted category for age band is 55 and over. Omitted category for Tenure is less than 2 years. ***Source: London Economics based on data from ABS, ILR, LFS, BRES***

Table 14: Results of firm-level econometric modelling (System GMM)

Dependent variable	Log GVA per worker			Log employment costs per worker			Log GVA per worker	Log employment costs per worker	Log GVA per worker	Log employment costs per worker
Controlling for (training variables)	*Basic controls*	*Firm size & SIC section controls*	*Industry level controls*	*Basic controls*	*Firm size & SIC section controls*	*Industry level controls*	*Firm and industry controls*	*Firm and industry controls*	*Firm and industry controls*	*Firm and industry controls*
Sample	*All*	*All*	*All*	*All*	*All*	*All*	*Large firms*	*Large firms*	*Medium firms*	*Medium firms*
Estimator	*System GMM*	*System GMM*	*System GMM*	*System GMM*	*System GMM*	*System GMM*	*System GMM*	*System GMM*	*System GMM*	*System GMM*
Training proportion (ILR)	0.978 (0.775)	0.415 (0.660)	1.010 (0.696)	-0.259** (0.125)	-0.439*** (0.161)	-0.069 (0.120)	0.686 (0.613)	-0.068 (0.124)	1.417 (1.727)	0.338 (0.271)
Training intensity (ESS)			1.451* (0.818)			-0.172* (0.103)	1.464** (0.727)	-0.098 (0.103)	-3.420 (2.883)	-0.276 (0.319)
(log)Capital stock per worker			-0.012 (0.515)			-0.158 (0.098)	0.167 (0.447)	-0.144* (0.081)	-0.438 (0.976)	0.292* (0.162)
(log)Capital expenditure per worker	0.296* (0.158)	0.536*** (0.074)	0.228 (0.144)	0.083*** (0.022)	0.184*** (0.019)	0.091*** (0.019)	-0.077 (0.120)	0.077*** (0.016)	0.903*** (0.310)	0.096*** (0.036)
(log)R&D per worker			0.002 (0.088)			0.024* (0.013)	-0.003 (0.075)	0.022* (0.013)	0.053 (0.269)	0.016 (0.038)
(log)Expenditure on computer and advertising services as a proportion of total purchases	-0.143 (0.183)	0.067 (0.049)	0.154 (0.119)	-0.007 (0.026)	0.054*** (0.011)	0.012 (0.015)	0.078 (0.097)	-0.011 (0.015)	0.195 (0.295)	0.052 (0.035)
(log)Hours worked			0.743 (1.452)			0.413** (0.208)	0.457 (1.500)	0.335 (0.230)	3.017 (3.356)	0.711 (0.466)
% in a temporary job			-6.602*** (1.681)			-0.360 (0.272)	-6.796*** (1.676)	-0.513* (0.279)	-1.295 (3.321)	-1.023** (0.481)
Staff non proficient at t-1			-1.007 (1.146)			-0.166 (0.163)	-0.722 (1.101)	-0.031 (0.155)	-0.278 (3.447)	-0.202 (0.352)
Foreign	0.096 (0.078)	0.088 (0.065)	0.163** (0.074)	0.217*** (0.012)	0.213*** (0.012)	0.201*** (0.011)	0.131* (0.072)	0.140*** (0.012)	-0.094 (0.176)	0.321*** (0.023)
Share of PT workers	-1.050*** (0.230)	-1.116*** (0.135)	-0.566** (0.285)	-1.248*** (0.038)	-1.254*** (0.035)	-0.965*** (0.046)	-0.729** (0.294)	-1.012*** (0.053)	0.100 (0.564)	-0.734*** (0.078)
Multiple Local Units	-0.008 (0.076)	-0.051 (0.073)	-0.045 (0.082)	-0.052*** (0.011)	0.003 (0.013)	-0.013 (0.011)	-0.171 (0.109)	-0.001 (0.014)	0.073 (0.124)	-0.030 (0.020)
Observations	37,454	37,454	32,500	37,656	37,656	32,683	22,540	22,716	9,960	9,967
Number of reporting units	11,794	11,794	11,591	11,831	11,831	11,631	6,850	6,890	5,145	5,146
AR1	0	0	0	0	0	0	0	0.00	0.00	0.00
AR2	0.02	0.02	0.06	0.01	0.10	0.15	0.49	0.28	0.06	0.94
Hansen test	0	0	0	0	0	0	0	0	0.16	0.00

Note: Two-step robust standard errors in parentheses. *** Significant at the 1% level; ** Significant at the 5% level; * Significant at the 10% level; All regressions include controls for company age bands, legal status, region and year. The second and third sets of regressions also include controls for employment size bands and SIC sections. Grey shading identifies measures aggregated at the industry level. Variables treated as endogenous in the System GMM regression: ESS training, ILR training, Capital Stock, Capital expenditure, Expenditure on computer and advertising services, R&D expenditure, Number of Hours. The values reported for AR(1) and AR(2) are the p-values for first and second order autocorrelated disturbances in the first differences equations. The Hansen test reports the p-values for the null hypothesis of instrument validity. ***Source: London Economics based on data from ABS, ILR, IDBR, LFS and ESS***

This publication available from www.gov.uk/bis

Contacts us if you have any enquiries about this publication, including requests for alternative formats, at:

Department for Business, Innovation and Skills
1 Victoria Street
London SW1H 0ET
Tel: 020 7215 5000

Email: enquiries@bis.gsi.gov.uk

BIS/16/255