

# **R&D and Productivity: Evidence from large UK establishments with substantial R&D activities**

Stephen R Bond<sup>a</sup> and Irem Guceri<sup>b</sup>

<sup>a</sup>Nuffield College, Department of Economics, and Centre for Business Taxation, University of Oxford, UK; <sup>b</sup>Centre for Business Taxation, Saïd Business School, University of Oxford, UK

## **ARTICLE HISTORY**

Compiled June 6, 2016

## **ABSTRACT**

We use ONS micro data for large UK establishments in the production industries in the period 1997-2008 to study the relationship between their productivity and the presence of substantial R&D activities, either at the production unit itself, or at other UK reporting units owned by the same enterprise group. We estimate that total factor (revenue) productivity is on average about 14% higher at the establishments which have substantial R&D themselves, compared to those with no R&D. Among the establishments with no R&D themselves, we estimate that productivity is on average about 9% higher at those which belong to enterprise groups which do have substantial R&D elsewhere in the UK in the same sub-sector. For the establishments with substantial R&D themselves, we also estimate a significant positive relationship between current productivity and past R&D expenditure using dynamic specifications which allow for both establishment-specific ‘fixed effects’ and a serially correlated error component.

## **KEYWORDS**

R&D, productivity

## **JEL CLASSIFICATION**

D24, O30

## **1. Introduction**

This paper uses micro data from the UK Office for National Statistics’ (ONS) main surveys of research and development (R&D) and production activities to study the relationship between R&D spending and (revenue) productivity among large UK establishments in the production industries in the period 1997-2008. We focus on production units which are large enough to be surveyed each year in the Annual Business Inquiry,<sup>1</sup> and which have either no R&D activity or which have R&D activities that are substantial enough to be reported annually and in detail in the Business Expenditure on Research and Development (BERD) survey. We also use ownership information

---

Corresponding author: Stephen R Bond. Email: [steve.bond@economics.ox.ac.uk](mailto:steve.bond@economics.ox.ac.uk)

to link production units to substantial R&D activities that are conducted elsewhere in the UK within the same enterprise group. Our focus on larger production units which are surveyed annually allows dynamic models of production to be considered, thereby controlling for persistent unobserved influences on productivity which may be correlated with R&D spending. Our focus on substantial R&D operations allows R&D personnel and purchases to be excluded from the measures of labour and material inputs used in our analysis of productivity, and allows the sectoral composition of the enterprise group’s UK R&D activity to be taken into account.

Our main findings can be summarised briefly. There is a strong association between the presence of substantial R&D activity and higher productivity, with (measured) total factor (revenue) productivity being about 14% higher on average in the production units that conduct significant R&D. This association remains after controlling for differences in size, age, ownership (domestic/foreign) and sub-sector of production activity. Within the sub-sample of production units which have substantial R&D activities, the evidence of a positive association between higher R&D spending and higher productivity is less strong. For a given production unit, however, we do find evidence that an increase in R&D spending tends to be followed by an increase in productivity; this suggests that the cross-sectional association may not be driven entirely by unobserved heterogeneity or establishment-specific ‘fixed effects’. We also find that the presence of substantial R&D activity elsewhere in the UK within the same enterprise group is associated with higher productivity when that R&D is classed as being within the same sub-sector as the production unit, but not when the R&D elsewhere relates to a different sub-sector of production activity.

Earlier papers which have used these UK ONS micro datasets to study the relationship between R&D and productivity include Rogers (2006), Harris and Li (2009), Sena and Higon (2014) and Harris and Moffat (2015). Our paper differs in its focus on the behaviour of larger establishments and in its consideration of R&D conducted by the same enterprise group but at different locations. Concerning this latter aspect, the closest precedent that we are aware of is the study by Adams and Jaffe (1996) for the chemicals sector in the USA.<sup>2</sup> We share the goals of the classic contribution by Crepon, Duguet and Mairesse (1998), using micro data on establishments and econometric methods which account for some sources of simultaneity bias; lacking data from innovation surveys, however, we consider the relationship between R&D and productivity using a more traditional reduced form approach.

Sections 2 and 3 describe the data and the econometric framework that we use in our analysis. Section 4 presents our main findings, first considering the relationship

between productivity and substantial R&D activity at the same production unit, and then extending the analysis to consider substantial R&D activities by other reporting units that are owned by the same enterprise group. An accompanying online working paper provides more detail about both our sample and our results.

## 2. Data

Our sample is obtained from two ONS micro datasets, the Annual Respondents' Database (ARD) which provides information on production inputs and outputs, and the BERD survey. These are linked using the Inter-Departmental Business Register (IDBR), which also allows us to construct, for each production unit, measures of UK R&D conducted by the same enterprise group but at different locations. We use data for ONS *reporting units* ('establishments'), which is the most disaggregated level at which data on gross output and all production inputs is available in the ARD.<sup>3</sup> Multiple reporting units may be owned by an *enterprise*, and multiple enterprises may be owned by an *enterprise group*. The enterprise group is the most aggregated level to which reporting units can be matched using the IDBR, and corresponds to the UK component of corporate groups.<sup>4</sup>

### 2.1. Production data

We use annual observations on the production variables during the period 1997-2008, in which the source for the ARD was the Annual Business Inquiry. During this period, both the survey questionnaire and the sampling scheme remained constant. Reporting units with 250 or more employees were surveyed annually in all production industries, while the sampling probabilities for smaller reporting units depended on the sub-sector of production activity. We use data on reporting units which provide information on the production variables for 3 or more consecutive years,<sup>5</sup> which ensures that most of the reporting units in our sample are large enough to be surveyed annually. We also restrict our sample to reporting units in the production industries. Table 1 summarises the main variables that we use in this paper. Nominal variables are expressed in constant prices using ONS price indices for the most disaggregated sub-sectors available. Observations with extreme high or low values for the growth rates or the ratios of any of these production variables are dropped from the sample used for estimation. Bond and Guceri (2016) provides further details.

[TABLE 1 HERE]

## **2.2. *R&D data***

The BERD survey is an annual census only for reporting units with at least 400 employees or those with substantial R&D activities. A subset of reporting units with the largest R&D expenditures are asked to provide more detailed information using a ‘long-form’ questionnaire. The number of long-form recipients increased over our sample period, from 352 in 1997 to 785 in 2008. These units accounted for around 80% of total UK BERD, and around 85% of manufacturing sector BERD. The number of these long-form recipients which also reported production data in the ARD was more stable over our sample period, fluctuating between 131 in 2003 and 159 in 2001; these long-form production units accounted for around 30% of total UK BERD. The number in our estimation sample fluctuated between 87 in 2000 and 114 in 2001.

In addition to total R&D spending, the long-form BERD survey reports data for the number and total salaries of R&D employees, and total purchases of R&D materials. We use this information to exclude R&D personnel and R&D purchases from the measures of employment and material inputs used in our analysis of production, as suggested by Schankerman (1981), and to exclude R&D employees from the average wage variable that we use as a measure of labour quality. Since we found that these ‘double-counting’ corrections had an important impact on our results, we omit from our sample production units which have R&D activities but do not report this more detailed information (i.e. they appear in the BERD survey but are not long-form recipients). For the remaining production units which have substantial R&D activities, we require that they report long-form R&D information for at least 4 consecutive years.<sup>6</sup>

R&D expenditure is converted to constant prices using an index which gives equal weight to the GDP deflator and the average earnings of scientists. For each production unit, we construct measures of UK R&D conducted by other members of the same enterprise group by summing over all reporting units which belong to the same group and which report long-form R&D information. The sectoral breakdown of R&D expenditure available for long-form recipients then allows us to restrict this measure to other UK R&D which relates to the same sub-sector as the production unit’s own production.<sup>7</sup>

## **2.3. *Sample***

Our resulting sample contains 14,242 annual observations on 2,895 distinct establishments which are observed for 3 or more consecutive years between 1998 and 2008. Of

these 14,242 observations, 603 were on production units which themselves reported detailed information on substantial R&D activities in the previous year, and 3,631 were on production units which belonged to enterprise groups in which at least one member reported detailed information on substantial R&D activities in the previous year.

The median age of the establishments in our sample is 23 years, with the lower quartile and upper quartile of the age distribution being 14 and 29 years respectively. 41.5% of our observations have more than 250 employees, and 8.6% have more than 1000 employees. 22.5% of the establishments are foreign-owned. Table 2 reports the sectoral distribution of our full sample of 14,242 observations on production units (93.3% of which are within the manufacturing sector) and our sub-sample of 603 observations on long-form BERD recipients (almost all of which are in manufacturing). Table 3 reports some basic descriptive statistics for the variables that are used in our empirical analysis. The production establishments with substantial R&D activities tend to be much larger than those with no R&D activity; they account for less than 5% of our total number of observations, but almost one quarter of the total value added produced by the establishments in our sample. They also tend to employ more skilled workers than the non-R&D establishments, as suggested by average annual real wages that are about 50 percent higher at either the mean or the median.

[TABLE 2 HERE]

[TABLE 3 HERE]

### 3. Empirical framework

Our starting point is a 3 factor Cobb-Douglas production function which can be written in log-linear form as

$$y_{it} = \beta_L l_{it} + \beta_K k_{i,t-1} + \beta_M m_{it} + a_{it} \quad (1)$$

in which  $Y_{it}$  is a measure of the value of gross output for reporting unit  $i$  in year  $t$ ,  $L_{it}$  is a measure of the number of employees during year  $t$ ,  $K_{i,t-1}$  is a measure of the capital stock at the end of year  $t - 1$ ,  $M_{it}$  is a measure of total purchases of materials during year  $t$ , and corresponding lower case notation denotes the natural logarithms of these variables. The residual term ( $a_{it}$ ) then denotes the log of total factor (revenue) productivity, or the variation in the value of gross production which is not accounted for by the measured variation in these observed inputs (given the imposed functional

form).

We allow the log of total factor (revenue) productivity to depend on measures of previous R&D activity. In the empirical analysis reported below we consider various functional forms for this relationship. Letting  $r_{i,t-1}$  denote the log of R&D expenditure in year  $t - 1$ , one simple example would be a linear relationship as in

$$a_{it} = \beta_R r_{i,t-1} + e_{it} \quad (2)$$

which gives the augmented production function

$$y_{it} = \beta_L l_{it} + \beta_K k_{i,t-1} + \beta_M m_{it} + \beta_R r_{i,t-1} + e_{it}. \quad (3)$$

We also include the log of the average wage ( $W_{it}$ ) paid to employees of reporting unit  $i$  during year  $t$ , which controls for variation in labour quality and/or hours worked, giving the extended specification

$$y_{it} = \beta_L l_{it} + \beta_K k_{i,t-1} + \beta_M m_{it} + \beta_R r_{i,t-1} + \beta_W w_{it} + u_{it}. \quad (4)$$

The restriction  $\beta_W = \beta_L$  would imply that variation in labour quality/hours is fully reflected in the cost of labour, in which case the appropriate measure of labour input would be the wage bill rather than the number of employees; we investigate but do not impose this restriction.

Our empirical specifications all include a full set of year dummies and a full set of sector dummies, at approximately the two digit SIC level.<sup>8</sup> In some specifications, the R&D variable is a binary dummy which indicates whether the reporting unit had either substantial R&D expenditure or no R&D expenditure in year  $t - 1$ . In other specifications, we use the level of R&D spending in year  $t - 1$ . Since the level of R&D spending is zero for the majority of the reporting units in our sample, this precludes the use of the standard logarithmic transformation. Instead we use the inverse hyperbolic sine transformation, so that our lower case notation denotes<sup>9</sup>

$$x_{it} = \sinh^{-1} X_{it} = \ln(X_{it} + \sqrt{X_{it}^2 + 1}). \quad (5)$$

The choice between using natural logarithms or inverse hyperbolic sines made no noticeable difference to the results for baseline specifications with no R&D variables or for specifications with R&D dummies.<sup>10</sup>

The use of flow data on R&D expenditure also raises a different issue, since it is unlikely that productivity this year depends only on last year's R&D spending. A

more plausible interpretation draws on the knowledge stock model of Griliches (1979), in which productivity this year depends on the stock of knowledge at the end of the previous year ( $G_{i,t-1}$ ), given by cumulated R&D expenditures with the perpetual inventory formulation

$$\begin{aligned} G_{i,t-1} &= R_{i,t-1} + (1 - \delta)R_{i,t-2} + (1 - \delta)^2 R_{i,t-3} + \dots \\ &= (1 - \delta)G_{i,t-2} + R_{i,t-1} \end{aligned} \quad (6)$$

in which  $R_{i,t-1}$  is R&D spending in year  $t - 1$ , and  $\delta$  is the rate at which knowledge acquired through R&D depreciates. Along a steady state path in which both knowledge and R&D spending are growing at the constant rate  $g$ , we have the relation

$$R_{i,t-1} = \left( \frac{g + \delta}{1 + g} \right) G_{i,t-1}. \quad (7)$$

Close to such a steady state path, the log (or inverse hyperbolic sine) of the measured flow variable provides an approximation to the log (or inverse hyperbolic sine) of the unobserved stock of knowledge.<sup>11</sup>

### 3.1. *Some econometric issues*

There are numerous reasons why ordinary least squares (OLS) estimates of production function models like equation (4) above may be inconsistent. With panel data on reporting units, the residual component of (transformed) total factor productivity ( $u_{it}$ ) may have time-invariant, persistent and transitory components. Explanatory variables like the capital stock at the end of last year, and last year's R&D spending, which are quasi-fixed in the sense of being determined before the start of year  $t$ , may nevertheless be correlated with time-invariant or persistent components of  $u_{it}$ . Explanatory variables like labour and material inputs, which are flexible in the sense that they can be adjusted during year  $t$ , may also be correlated with transitory components of  $u_{it}$ . Intuitively, establishments with higher productivity are likely to be larger, and the levels of more flexible inputs are likely to be more sensitive to temporary fluctuations in productivity. Measurement error may be a further source of inconsistency, and this may also be more important for some of the inputs. The net effect of these different sources of simultaneity bias on the behaviour of the OLS estimates is difficult to characterise.<sup>12</sup>

To investigate the reliability of our OLS estimates of the relationship between R&D and productivity, we also consider generalised method of moments (GMM) estimates of a dynamic production function specification, following the approach suggested by

Blundell and Bond (2000). We allow the error term ( $u_{it}$ ) to have a time-invariant component ( $\eta_i$ , i.e. an establishment-specific ‘fixed effect’) and an autoregressive component ( $v_{it}$ )

$$\begin{aligned} u_{it} &= \eta_i + v_{it} \\ v_{it} &= \alpha v_{i,t-1} + \varepsilon_{it}. \end{aligned} \tag{8}$$

This formulation allows for both permanent and persistent components of (transformed) total factor productivity, with productivity shocks represented by  $\varepsilon_{it}$ . Quasi-differencing equation (4) then gives the dynamic specification

$$y_{it}^* = \beta_L l_{it}^* + \beta_K k_{i,t-1}^* + \beta_M m_{it}^* + \beta_R r_{i,t-1}^* + \beta_W w_{it}^* + \eta_i^* + \varepsilon_{it} \tag{9}$$

in which  $x_{it}^* = x_{it} - \alpha x_{i,t-1}$ ,  $v_{it}^* = \varepsilon_{it}$  and  $\eta_i^* = \eta_i(1 - \alpha)$ . Conditional on the value of the autoregressive parameter  $\alpha$ , this is a linear panel data model with explanatory variables which may be correlated with the time-invariant component of the error term ( $\eta_i^*$ ) and with current or past values of the productivity shocks ( $\varepsilon_{it}$ ). The parameters  $(\beta_L, \beta_K, \beta_M, \beta_R, \beta_W)$  can be estimated consistently using GMM. Assuming that the  $\varepsilon_{it}$  are serially uncorrelated,  $(l_{i,t-2}, k_{i,t-3}, m_{i,t-2}, r_{i,t-3}, w_{i,t-2})$  and longer lags of these variables are valid instruments for the equations in first-differences

$$\Delta y_{it}^* = \beta_L \Delta l_{it}^* + \beta_K \Delta k_{i,t-1}^* + \beta_M \Delta m_{it}^* + \beta_R \Delta r_{i,t-1}^* + \beta_W \Delta w_{it}^* + \Delta \varepsilon_{it}. \tag{10}$$

Assuming that the first-differenced variables  $(\Delta l_{i,t-1}, \Delta k_{i,t-2}, \Delta m_{i,t-1}, \Delta r_{i,t-2}, \Delta w_{i,t-1})$  are uncorrelated with the time-invariant unobserved heterogeneity ( $\eta_i$ ), these lagged first-differences are also available as instruments for the equations in levels (9). Both sets of moment conditions can be combined using the ‘system GMM’ estimator developed by Arellano and Bover (1995) and Blundell and Bond (1998), and implemented in Stata by Roodman (2009). Standard tests of instrument validity can then be used to check the maintained assumptions.

The dynamic model (9) can also be written in the form

$$\begin{aligned} y_{it} &= \alpha y_{i,t-1} + \beta_L l_{it} - \alpha \beta_L l_{i,t-1} + \beta_K k_{i,t-1} - \alpha \beta_K k_{i,t-2} + \beta_M m_{it} \\ &\quad - \alpha \beta_M m_{i,t-1} + \beta_R r_{i,t-1} - \alpha \beta_R r_{i,t-2} + \beta_W w_{it} - \alpha \beta_W w_{i,t-1} + \eta_i^* + \varepsilon_{it} \end{aligned} \tag{11}$$

and can be estimated without imposing the non-linear ‘common factor’ restrictions which link the coefficients on the current and lagged values of each explanatory variable to that on the lagged dependent variable.<sup>13</sup> Estimates of this unrestricted specification



are useful for assessing the validity of the dynamic structure implied by (8), and for providing consistent estimates of the persistence parameter  $\alpha$ .

## 4. Results

### 4.1. *Static specifications*

We begin by presenting OLS estimates of static production function specifications for our sample of large UK establishments, initially without any R&D variables. The first column of Table 4 reports the results for a specification with only the basic employment, capital and material inputs, as in equation (1). The estimated partial elasticity parameters for each of these inputs is positive and significantly different from zero. The three parameters sum to 1.016, suggesting that the production technology is very close to displaying constant returns to scale.<sup>14</sup> For comparison, the average shares of labour costs and material costs in gross output are 0.25 and 0.59 in our sample.

The second column of Table 4 reports the results for an extended specification which includes the average wage paid to employees as a control for variation in labour quality and/or hours worked. The estimated coefficient on the average wage is very close to the estimated coefficient on the number of employees in this and subsequent specifications.<sup>15</sup> This suggests that workers who were 10% more (less) productive than the average were also paid 10% more (less) than the average in these large UK establishments during our sample period. Given this, the effective labour input could be measured more parsimoniously using the log (or inverse hyperbolic sine) of the wage bill.<sup>16</sup>

The remaining columns of Table 4 report the results for augmented specifications with various measures of R&D activity, as illustrated in equation (4). Column (iii) includes a binary dummy variable which takes the value one (in year  $t$ ) for production units which had substantial R&D activity in the previous year (i.e. they were long-form recipients in the BERD survey), and zero (in year  $t$ ) for production units which had no R&D activity in the previous year (i.e. they were not included in the sampling framework for the BERD survey). On average, we estimate that total factor (revenue) productivity was 14.4% higher in the sub-sample of large establishments which themselves had substantial R&D activity. This result is not found if we do not correct the measures of labour and materials at the production units with R&D activity for the inclusion of R&D personnel and purchases in the standard measures of employment and material costs.<sup>17</sup> This contrast suggests that spending on labour and materials engaged in R&D on average generated a similar contribution to revenue as spending

the same amount on labour and materials engaged directly in production. This in turn suggests that the typical allocation of resources between R&D and production activities may have been broadly efficient, at least within these establishments.

Column (iv) replaces the binary indicator of substantial R&D activity by the inverse hyperbolic sine of total R&D expenditure reported by the long-form recipients in year  $t - 1$ . This specification also indicates that total factor (revenue) productivity is significantly higher in the sub-sample of establishments with significant R&D activities. For this sub-sample, the mean of the transformed R&D expenditure variable is 9.95, so this specification also suggests that their productivity is about 14% higher on average compared to the production units with no R&D.

The two specifications considered in columns (iii) and (iv) of Table 4 differ in their implications for whether higher spending on R&D is associated with higher productivity, within the sub-sample of establishments which have large R&D operations. Our sample provides only weak evidence in favour of this association. The specification reported in column (v) includes both the dummy variable and the continuous measure of R&D spending. In this case, the hypothesis that the two coefficients on these R&D variables are both zero can be strongly rejected, but neither of the individual coefficients is significantly different from zero. Thus we have strong evidence that establishments with substantial R&D activity tend to have a productivity advantage over establishments with no R&D activity, but not that this productivity advantage increases with the level of R&D expenditure.

To explore this further, the specification reported in column (vi) divides the observations with substantial R&D activity (in year  $t - 1$ ) into 4 quartiles according to their level of R&D spending (also in year  $t - 1$ ), and includes 4 dummy variables which indicate whether or not an observation is in the corresponding quartile of the R&D expenditure distribution.<sup>18</sup> This specification indicates that establishments in each quartile of the R&D spending distribution have a significant productivity advantage compared to establishments with no R&D activity. Formally we do not reject the restriction that these 4 coefficients are all equal, which again indicates no statistically significant difference between the productivity of the highest spenders and the lowest spenders within our sample of establishments with substantial R&D activities.<sup>19</sup> However the point estimates here do suggest some tendency for the productivity advantage to increase with the level of R&D expenditure, at least up to the third quartile. Moreover none of these R&D quartile dummies has a coefficient which is significantly different from zero when added as an additional explanatory variable to the specification in column (iv), with our continuous R&D spending measure.<sup>20</sup> Thus neither of

the more parsimonious functional forms considered in columns (iii) and (iv) is rejected by this consideration of more flexible specifications.

The finding that production units with substantial R&D activity have higher total factor (revenue) productivity than production units with no R&D activity could of course reflect other differences between these sub-samples of large UK establishments which happen to be correlated with the presence of R&D. Among the observable candidates, we can rule out size, age and ownership differences as being important contributors to the relationship between R&D and productivity that we have reported in this section. In Bond and Guceri (2016), we report results for extended specifications, based on both column (iii) and column (iv) of Table 4, with additional controls for size (using employment size band dummies), age (using age quartile dummies) and foreign ownership. We find evidence that productivity tends to be somewhat higher in the larger establishments, lower in the older establishments, and higher in the establishments owned by foreign firms.<sup>21</sup> In all cases, however, we continue to find a highly significant difference between the establishments with and without substantial R&D activities, and the average size of this difference is stable at around 14%.

[TABLE 4 HERE]

#### 4.2. *Dynamic specifications*

We now consider specifications which allow for unobserved differences between establishments. The approach outlined in section 3.1 allows for both time-invariant and persistent unobserved components of total factor (revenue) productivity, with the latter assumed to follow a first-order autoregressive process. Any approach which allows observed explanatory variables to be correlated with establishment-specific ‘fixed effects’ relies on variation over time within the set of observations on a given production unit. Since we observe almost no variation over time in the binary R&D status indicator for our sample of establishments, we focus in this section on models which use the continuous measure of R&D spending, as in column (iv) of Table 4.

Table 5 presents our system GMM estimates of quasi-differenced production functions of the form of equation (9), both with and without the R&D expenditure variable. The instruments used for the equations in first-differences are the second, third and fourth lags of each of the included explanatory variables, and the instruments used for the equations in levels are the first lag of the first difference of each of these variables. We impose the value of the autoregressive parameter ( $\alpha$ ) to be 0.5, based on preliminary estimates of the unrestricted version of equation (11).<sup>22</sup> These GMM estimates

of the coefficients for the labour, capital and material inputs are broadly similar to the OLS estimates in, for example, column (ii) of Table 4. The estimated coefficient on the R&D expenditure variable is positive and significantly different from zero at the 10% level. The point estimate here is larger than the OLS estimate reported in column (iv) of Table 4, although we caution that the GMM approach estimates this parameter imprecisely. This imprecision reflects the limited variation over time in the level of R&D spending reported by individual establishments. The main conclusion that we draw from this section is that there is no indication that the association between R&D and productivity suggested by the OLS estimates of static specifications is likely to be seriously biased upwards by the neglect of either permanent or autoregressive differences in productivity between establishments.

In Bond and Guceri (2016) we report the corresponding system GMM estimates of the unrestricted version of equation (11). We also find a significant positive long-run effect of R&D expenditure on total factor (revenue) productivity using this more general dynamic specification. Individually, none of the non-linear ‘common factor’ restrictions, which are imposed in the specifications reported in Table 5, are rejected for any of the included explanatory variables.<sup>23</sup> This suggests that the persistence in the residual component of total factor (revenue) productivity can be approximated well using the combination of time-invariant and first-order autoregressive components considered in equation (8).

[TABLE 5 HERE]

#### 4.3. *R&D at affiliated establishments*

Our sample also allows us to investigate the relationship between productivity at large UK establishments and substantial R&D activity conducted by other UK reporting units within the same enterprise group. The latter include R&D facilities which themselves have no significant production operations.

In the first column of Table 6, we augment the specification from column (iii) of Table 4 by including an additional binary dummy which takes the value one (in year  $t$ ) for production units which themselves have no substantial R&D activity (i.e. they are not long-form recipients in the BERD survey) but which belong to enterprise groups which have substantial R&D activity elsewhere in the UK (i.e. they own one or more reporting units that were long-form recipients in the BERD survey, in year  $t - 1$ ). This dummy takes the value zero (in year  $t$ ) both for production units owned by enterprise groups which have no substantial R&D activity in the UK (i.e. they have no members

that are long-form recipients in the BERD survey) and by enterprise groups whose only substantial UK R&D activity is conducted at the production unit itself.<sup>24</sup> We find a significant association between higher productivity and the presence of substantial R&D activity elsewhere in the UK within the same enterprise group, although the average productivity gain associated with the presence of substantial R&D at affiliated UK establishments is lower than the average productivity gain associated with the presence of substantial R&D at the production unit itself. The estimated coefficient on the dummy for the presence of substantial R&D at the production unit itself suggests an average difference of 15.7% between the total factor (revenue) productivity of the establishments which have substantial R&D activities and the establishments which have no R&D activities and which belong to enterprise groups with no substantial R&D activities in the UK.<sup>25</sup>

In the second column of Table 6 we distinguish between the presence of substantial R&D at affiliated UK establishments which is reported to be in the same sub-sector as the production unit, and the presence of substantial R&D at affiliated UK establishments which is reported to be in different sub-sectors, or outside the production industries. We find a significant positive association only with substantial R&D by other members of the same enterprise group which is related to the production unit's sub-sector of production activity, and not with substantial R&D by other members of the same enterprise group which is related to other sub-sectors of production industries, or to non-production or service sector activities.

Columns (iii) and (iv) of Table 6 report the results for otherwise similar specifications which use continuous measures of R&D spending. In column (iii), we add to the specification from column (iv) of Table 4 the inverse hyperbolic sine of total R&D expenditure in year  $t - 1$  reported by all other long-form recipients owned by the same enterprise group. In column (iv), we decompose this measure into R&D spending by affiliated establishments which is related to the production unit's sub-sector of production industries, and that which is not. The results here are qualitatively and quantitatively similar to those found for the specifications in columns (i) and (ii) of Table 6, using the binary R&D indicators.<sup>26</sup> Similar results were also found in extended specifications with additional controls for the size, age and ownership (domestic/foreign) of the production unit.<sup>27</sup>

[TABLE 6 HERE]

## 5. Conclusions

We have used ONS micro data for large UK establishments in the production industries in the period 1997-2008 to study the relationship between their productivity and the presence of substantial R&D activities, either at the production unit itself, or at other reporting units in the UK that are owned by the same enterprise group. Correcting the measures of labour and material inputs for the inclusion of R&D personnel and purchases at the establishments which have substantial R&D activities, we estimate that total factor (revenue) productivity is on average about 14% higher at these establishments, compared to those which have no R&D activity. This estimate increases to about 16% when we compare establishments with substantial R&D to establishments which have no R&D and which belong to enterprise groups which have no substantial R&D activity within the UK. Among the establishments which have no R&D activity themselves, we estimate that productivity is on average about 9% higher at those which belong to enterprise groups which have substantial R&D activity elsewhere in the UK in the production unit's sub-sector of production activity. For the establishments which have substantial R&D activity themselves, we also estimate a significant positive relationship between current productivity and past R&D expenditure using dynamic specifications which allow for both establishment-specific 'fixed effects' and for a serially correlated unobserved component of total factor (revenue) productivity.

There are several ways in which the research presented here can be refined and extended. Information on the location of production units and the location of affiliated establishments with substantial R&D activity can be used to investigate whether geographical proximity has an important effect on the relation between productivity and substantial R&D conducted by other members of the same enterprise group within the UK. More accurate measures of the connections between the nature of the production activity and the nature of the R&D activities at affiliated establishments can also be considered, perhaps using the distance measures suggested by Bloom, Schankerman and Van Reenen (2013). The rich information on the composition of R&D expenditure for all UK establishments with substantial R&D activities that is available for long-form recipients in the BERD survey will also allow the nature of spillover effects from substantial R&D spending by other firms to be explored.

## Acknowledgement(s)

We thank Eric Bartelsman, Kit Baum, Bronwyn Hall, Hans Lööf, Jacques Mairesse, Pierre Mohnen and other participants at the CDM Workshop III (MERIT, University

of Maastricht, 23-24 March 2015) and two anonymous referees for helpful comments.

## Disclosure statement

This work contains statistical data from the Office for National Statistics which is Crown copyright and reproduced with the permission of the controller of HMSO and Queen’s Printer for Scotland, made available for use by the UK Data Archive. The Office for National Statistics and the UK Data Archive bear no responsibility for the analysis of the statistical data or the opinions presented in this paper.

## Funding

We thank the ESRC for financial support under award ES/L000016/1. Guceri thanks Argela for financial support under the doctoral studentship agreement between the University of Oxford and Argela.

## Notes

<sup>1</sup>This was formerly known as the Census of Production.

<sup>2</sup>There is of course a much larger literature on externalities or ‘spillover effects’ from the R&D expenditures of other firms. See Hall, Mairesse and Mohnen (2010) for a recent review, and Bloom, Schankerman and Van Reenen (2013) for a recent example.

<sup>3</sup>For reporting units which have operations at more than one site, information on turnover and employment may be available at the more disaggregated *local unit* level.

<sup>4</sup>The IDBR definitions of enterprise and enterprise group follow the EU Regulation on Statistical Units (EEC 696/93). An enterprise group is an association of enterprises bound together by legal and/or financial links. The ONS User Guide for the Business Structure Database provides further details.

<sup>5</sup>We require a minimum of 3 consecutive observations to estimate the dynamic production models reported in section 4.2 below.

<sup>6</sup>This arises because we relate production in year  $t$  to R&D expenditure in year  $t - 1$ , but also require information in year  $t$  to correct the measures of labour and material inputs for ‘double-counting’.

<sup>7</sup>We combine the sub-sectors chemicals and pharmaceuticals for this purpose. The available classification of product groups is reported in Table 2 below, corresponding to the two digit SIC level. Further details are provided in Bond and Guceri (2016).

<sup>8</sup>Table 2 reports the full set of sector classifications that we consider.

<sup>9</sup>Thus we have  $x = 0$  for  $X = 0$ , and  $x$  approaches  $\ln 2X = \ln 2 + \ln X$  for large  $X$ .

<sup>10</sup>See Bond and Guceri (2016).

<sup>11</sup>Construction of a knowledge stock measure from the flow data on R&D spending requires an initialisation and an assumption about the appropriate rate of depreciation. We experimented with standard approaches, finding broadly similar results to those obtained with the flow measure of R&D that are reported here. These results are discussed further in Bond and Guceri (2016).

<sup>12</sup>See Griliches and Mairesse (1998) for a thorough discussion of these and other issues in the estimation of production functions.

<sup>13</sup>For example, denoting the unrestricted coefficient on the lagged labour input ( $l_{i,t-1}$ ) by  $\gamma_L$ , equation (11) implies the restriction  $\gamma_L = -\alpha\beta_L$ .

<sup>14</sup>Although the constant returns to scale restriction is strongly rejected by a standard F test.

<sup>15</sup>The restriction that these two coefficients are equal is not rejected by a standard F test at conventional significance levels.

<sup>16</sup>Denoting the wage bill by  $B$  and the average wage by  $W = B/L$ , this follows from  $\ln L + \ln W = \ln L + \ln B - \ln L = \ln B$ .

<sup>17</sup>Using input measures which are not corrected for double-counting, the estimated coefficient on the R&D dummy falls to 0.026, with a standard error of 0.023. These results are reported in detail in Bond and Guceri

(2016).

<sup>18</sup>Observations with the lowest R&D spending are classed as being in the first quartile (Q1), and observations with the highest R&D spending are classed as being in the fourth quartile (Q4). All four dummy variables are zero for observations with no R&D spending. Note that R&D spending in the bottom quartile is only ‘low’ in the context of establishments whose R&D activity is substantial enough to be among the long-form recipients for the BERD survey.

<sup>19</sup>The F test of the restriction that these four coefficients are equal has a p-value of 0.53.

<sup>20</sup>These results are reported in Bond and Guceri (2016).

<sup>21</sup>These differences are all quite modest. For example, we estimate the average difference between foreign-owned and UK-owned establishments to be about 4%.

<sup>22</sup>Similar results were obtained using values of  $\alpha$  in the range 0.4-0.6. For comparison, the OLS estimates of the coefficient on the lagged dependent variable in this specification are about 0.85, while the Fixed Effects (Within) estimates are about 0.2. The former are expected to be biased upwards in the presence of establishment-specific ‘fixed effects’, while the latter are expected to be biased downwards due to the small number of time series observations available for individual establishments (Nickell, 1981).

<sup>23</sup>Although Wald tests formally reject the joint validity of these non-linear common factor restrictions.

<sup>24</sup>For the production units which themselves have substantial R&D activity, we found no further productivity advantage associated with the presence of additional substantial R&D activity elsewhere in the UK by other members of the same enterprise group. Indeed we found some weak evidence that additional substantial R&D within the same group at other locations may have a small negative effect on the productivity of such reporting units. These results are discussed further in Bond and Guceri (2016).

<sup>25</sup>This estimate is naturally higher than we found in Table 4, since the specification used in Table 6 removes establishments owned by enterprise groups with substantial R&D activity elsewhere in the UK from the comparison group.

<sup>26</sup>The sample mean of the transformed R&D expenditure variable by other members of the same enterprise group in the production unit’s sub-sector, for production units which themselves have no R&D, is 10.289.

<sup>27</sup>These results are reported in Bond and Guceri (2016).

## References

- Adams, James D. and Adam B. Jaffe. 1996. “Bounding the Effects of R&D: an Investigation using Matched Establishment-Firm Data.” *RAND Journal of Economics* 27 (4): 700-721.
- Arellano, Manuel and Stephen R. Bond. 1991. “Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations.” *Review of Economic Studies* 58 (2): 277-297.
- Arellano, Manuel and Olympia Bover. 1995. “Another look at the instrumental-variable estimation of error component models.” *Journal of Econometrics* 68 (1): 29-52.
- Bloom, Nicholas, Mark Schankerman and John M. Van Reenen. 2013. “Identifying Technology Spillovers and Product Market Rivalry.” *Econometrica* 81 (4): 1347-1393.
- Blundell, Richard W. and Stephen R. Bond. 1998. “Initial Conditions and Moment Conditions in Dynamic Panel Data Models.” *Journal of Econometrics* 87 (1): 115-143.
- Blundell, Richard W. and Stephen R. Bond. 2000. “GMM Estimation with Persistent Panel Data: an Application to Production Functions.” *Econometric Reviews* 19 (3): 321-340.
- Bond, Stephen R. and Irem Guceri. 2016. “Supplementary Material for ‘R&D and Productivity: Evidence from Large UK Establishments with Substantial R&D Activities.’” Working Paper WP16/06 Oxford University Centre for Business Taxation 2016.
- Crepon, Bruno, Emmanuel Duguet and Jacques Mairesse. 1998. “Research, Innovation and Productivity: an Econometric Analysis at the Firm Level.” *Economics of Innovation and New Technology* 7 (2): 115-158.
- Griliches, Zvi. 1979. “Issues in Assessing the Contribution of Research and Development to Productivity Growth.” *Bell Journal of Economics*, 10 (1): 92-116.
- Griliches, Zvi and Jacques Mairesse. 1998. “Production Functions: the Search for Identification.” In *Econometrics and Economic Theory in the 20th Century*, edited by Steinar Strøm, 169-203. Cambridge: Cambridge University Press.
- Hall, Bronwyn H., Jacques Mairesse and Pierre Mohnen. 2010. “Measuring the Returns to R&D.” In *Handbook of the Economics of Innovation* edited by Bronwyn H. Hall and Nathan Rosenberg, Volume 2, 1033-1082. Amsterdam: North-Holland.



- Harris, Richard and Qian C. Li. 2009. "Exporting, R&D and Absorptive Capacity in UK Establishments." *Oxford Economic Papers* 61 (1): 74-103.
- Harris, Richard and John D. Moffat. 2015. "Plant-Level Determinants of Total Factor Productivity in Great Britain, 1997-2008." *Journal of Productivity Analysis* 44 (1): 1-20.
- Nickell, Stephen J. 1981. "Biases in Dynamic Models with Fixed Effects." *Econometrica* 49 (6): 1417-1426.
- Rogers, Mark. 2006. "Estimating the Impact of R&D on Productivity using the BERD-ARD Data." Technical report, Department of Trade and Industry.
- Roodman, David. 2009. "How to Do xtabond2: an Introduction to 'Difference' and 'System' GMM in Stata." *The STATA Journal* 9 (1): 86-136.
- Schankerman, Mark. 1981. "The Effects of Double-Counting and Expensing on the Measured Returns to R&D." *Review of Economics and Statistics* 63 (3): 454-458.
- Sena, Vania and Dolores A. Higon. 2014. "Productivity, R&D Spillovers and Educational Attainment." *Oxford Bulletin of Economics and Statistics* 76 (1): 1-23.

#### Datasets

Office for National Statistics, Business Expenditure on Research and Development, 1994-2008: Secure Data Service Access. Colchester, Essex: UK Data Archive, March 2011. SN: 6690.

Office for National Statistics, Business Structure Database, 1997-2011: Secure Data Service Access. 3rd Edition. Colchester, Essex: UK Data Archive, October 2012. SN: 6697.

Office for National Statistics, Annual Respondents Database, 1973-2009: Secure Data Service Access. 3rd Edition. Colchester, Essex: UK Data Archive, June 2012. SN: 6644.

**Table 1.** Production variables.

Gross output (Y)	Turnover plus change in stocks of finished goods and raw materials
Number of employees (L)	Average of quarterly survey reports within year $t$
Average wage (W)	Total salaries/Number of employees
Capital stock (K)	Constructed from the data on capital expenditures, using a perpetual inventory formula with a 6% annual depreciation rate; initial values allocate a share of the ONS capital stock estimate for each two digit sub-sector to individual reporting units, in proportion to total purchases of materials (including energy)
Material inputs (M)	Total purchases of materials (including energy)

Note: Nominal values are converted to constant prices using deflators for the most disaggregated SIC sector available:

Gross output: ONS producer price indices (output)

Materials: ONS producer price indices (inputs)

Capital: ONS implicit price indices for capital services

Wages: labour input deflators constructed using the ONS Annual Survey of Hours and Earnings.

Further details are provided in Bond and Guceri (2016).

**Table 2.** Sectoral composition of full sample and long-form BERD sub-sample.

Two digit sector	Description	All observations	Share of total value added	Long-form BERD observations	Share of total value added
A	Agriculture, forestry, fishing	75	0.1%		
B	Extractive industries	657	2.5%		
C	Food, beverages, tobacco	2,280	19.2%	46	2.6%
D	Textiles and clothes	1,183	1.6%		
E	Wood, paper, publishing	2,358	10.9%		
F	Refined petroleum	49	0.9%	15	0.8%
G	Chemicals	899	12.8%	} 141	4.6%
H	Pharmaceuticals	65	0.6%		
I	Rubber and plastics	594	2.6%		
J	Nonmetallic minerals	829	2.4%		
K	Basic iron ore, ferro alloys	608	2.9%		
L	Nonferrous metals	249	1.0%		
M	Fabricated metal products	902	1.7%		
N	Machinery and equipment	849	4.8%	115	2.9%
O	Office machinery, computers	146	1.9%	13	0.7%
P	Electrical machinery	272	0.9%		
Q	Radio, TV, communication	253	2.7%	55	1.0%
R	Precision instruments	206	1.0%	87	0.7%
S	Motor vehicles	463	16.9%	29	5.1%
T	Railway, rolling stock	173	0.4%		
U	Shipbuilding	96	0.6%		
V	Aircraft, spacecraft	204	6.2%	59	3.7%
W	Furniture	608	1.5%	12	0.1%
X	Recycling	86	0.2%		
Y	Utilities	138	4.0%		
Other <sup>a</sup>	n/a	n/a	n/a	31	1.4%
Total		14,242	100%	603	23.7%

<sup>a</sup> Sectors with less than 10 long form BERD observations cannot be separately identified

**Table 3.** Descriptive statistics.

a) Full sample (14,242 observations).

Variable	Mean	Std Dev	Lower Quartile	Median	Upper Quartile
Gross output	72,780	269,660	5,927	17,926	48,410
No of employees	420	876	71	176	429
Average wage	20	16	14	19	24
Capital stock	68,354	349,190	4,304	12,958	34,848
Material inputs	46,767	189,259	3,279	10,759	30,261

b) Long-form BERD sub-sample (603 observations).

Variable	Mean	Std Dev	Lower Quartile	Median	Upper Quartile
Gross output	406,747	818,719	82,407	164,159	537,011
No of employees	1,612	2,066	558	891	1,683
Average wage	31	39	23	28	33
Capital stock	324,620	652,439	50,600	104,319	359,654
Material inputs	261,784	576,240	39,630	100,253	292,074
R&D spending	24,312	59,214	4,695	8,528	18,487

Note: All variables (except number of employees) are measured in thousands of 1997 pounds.

For the long-form BERD sub-sample, the measures of employment, average wage and material inputs are corrected for double-counting of R&D.

**Table 4.** Static production functions.

	(i)	(ii)	(iii)	(iv)	(v)	(vi)
$\beta_L$ : labour	0.247 (.008)	0.331 (.009)	0.328 (.009)	0.328 (.009)	0.328 (.009)	0.328 (.009)
$\beta_W$ : wage		0.338 (.031)	0.331 (.031)	0.331 (.031)	0.331 (.031)	0.331 (.031)
$\beta_K$ : capital	0.114 (.013)	0.079 (.010)	0.075 (.010)	0.075 (.010)	0.075 (.010)	0.075 (.010)
$\beta_L$ : materials	0.655 (.013)	0.595 (.015)	0.597 (.015)	0.597 (.015)	0.597 (.015)	0.597 (.015)
$\beta_R$ : R&D dummy			0.144 (.024)		0.072 (.170)	
$\beta_R$ : R&D spending				0.014 (.002)	0.007 (.017)	
$\beta_R$ : R&D Q1						0.107 (.033)
$\beta_R$ : R&D Q2						0.142 (.031)
$\beta_R$ : R&D Q3						0.187 (.050)
$\beta_R$ : R&D Q4						0.139 (.037)
R squared	0.968	0.972	0.972	0.972	0.972	0.972
Observations	14,242	14,242	14,242	14,242	14,242	14,242
Obs with R&D	603	603	603	603	603	603

Note: OLS estimates.

Standard errors reported in brackets are asymptotically robust to heteroskedasticity and serial correlation (cluster-robust at the reporting unit level).

All specifications include year dummies and two-digit sector dummies.

For column (v), an F test rejects the hypothesis that the two coefficients on the two R&D variables are both zero, with a p-value of 0.000.

**Table 5.** Dynamic production functions.

	(i)	(ii)
$\beta_L$ : labour	0.381 (.058)	0.366 (.053)
$\beta_W$ : wage	0.378 (.135)	0.362 (.113)
$\beta_K$ : capital	0.061 (.050)	0.049 (.052)
$\beta_L$ : materials	0.625 (.044)	0.634 (.042)
$\beta_R$ : R&D spending		0.037 (.021)
Observations	11,288	11,288
Obs with R&D	492	492
1 <sup>st</sup> order autocorrelation	-12.37	-12.69
2 <sup>nd</sup> order autocorrelation	1.34	1.46
Hansen (p): all instruments	0.169	0.269
Hansen (p): levels instruments	0.025	0.060

Note: System GMM estimates, using the instruments described in the text, implemented in Stata using Roodman's (2009) `xtabond2` command.

Standard errors reported in brackets are asymptotically robust to heteroskedasticity and serial correlation (cluster-robust at the reporting unit level).

All specifications include year dummies and two-digit sector dummies.

Test statistics for no 1<sup>st</sup>/2<sup>nd</sup> order serial correlation in the first-differenced residuals are Arellano-Bond (1991) tests, asymptotically standard normal under the hypothesis of no serial correlation.

p-values are reported for the Hansen tests of overidentifying restrictions.

**Table 6.** Static production functions with R&D of affiliated establishments.

	(i)	(ii)	(iii)	(iv)
$\beta_L$ : labour	0.327 (.009)	0.327 (.009)	0.327 (.009)	0.326 (.009)
$\beta_W$ : wage	0.328 (.031)	0.328 (.031)	0.328 (.031)	0.328 (.031)
$\beta_K$ : capital	0.072 (.010)	0.072 (.010)	0.072 (.010)	0.072 (.010)
$\beta_L$ : materials	0.599 (.015)	0.600 (.015)	0.599 (.015)	0.600 (.015)
$\beta_R$ : own R&D dummy	0.157 (.025)	0.159 (.025)		
$\beta_R$ : group R&D dummy	0.066 (.016)			
$\beta_R$ : group R&D dummy: same sector		0.088 (.020)		
$\beta_R$ : group R&D dummy: other sectors		0.007 (.020)		
$\beta_R$ : own R&D			0.016 (.003)	0.016 (.003)
$\beta_R$ : group R&D			0.007 (.002)	
$\beta_R$ : group R&D: same sector				0.009 (.002)
$\beta_R$ : group R&D: other sectors				-0.001 (.002)
R squared	0.972	0.972	0.972	0.972
Observations	14,242	14,242	14,242	14,242
Obs with own R&D	603	603	603	603
Obs with group R&D	3,631	3,631	3,631	3,631

Note: OLS estimates.

Standard errors reported in brackets are asymptotically robust to heteroskedasticity and serial correlation (cluster-robust at the reporting unit level).

All specifications include year dummies and two-digit sector dummies.

‘Group’ R&D measures refer to R&D of affiliated establishments for the production units which themselves have no R&D.