THE MEDIUM-TERM EFFECT OF R&D ON FIRM GROWTH

Marco Capasso

UNU-MERIT and School of Business and Economics, Maastricht University

Bart Verspagen

UNU-MERIT and School of Business and Economics, Maastricht University
Centre for Technology, Innovation and Culture (TIK), University of Oslo

Abstract

This study analyses the medium-term effect of R&D expenditure on firm employment growth. Four cross-sectional waves of an innovation survey conducted in the Netherlands have been used to evaluate the effect on firm growth in the five years following the investment. Panel data fixed effect techniques, also allowing for selection bias corrections, indicate a positive influence of R&D on growth. Limited dependent variable models have been used throughout the whole analysis to consider explicitly the cases of firms exiting the market in the analyzed medium term.

1. Introduction

This paper examines the relation between R&D expenditure and firm medium-term growth. The recent literature has confirmed that firm-level analysis is necessary to capture the heterogeneity of the economy (Reichstein et al., 2010). In particular, the dynamics of some groups of firms, like gazelles and high growth firms, seem to be driven by different factors than the ones characterizing the dynamics of more moderately growing firms (Coad and Rao 2008; Hölzl 2009; Stam and Wennberg 2009). A current challenge of economic researchers is then expanding the Gibrat's Law approach (Gibrat, 1931) to understand how such heterogeneity can be explained (Stam, 2010). Innovation is one of the usual suspects in defining differences in performance (and especially sustained performance) among firms, and connecting more explicitly innovation patterns with what is known about firm growth is thus a challenge for current research (Cefis and Orsenigo, 2001). In particular, Coad and Rao (2008) have observed that "innovation is of crucial importance for a handful of 'superstar' fast-growth firms".

In this paper, growth in terms of firm employment is the dependent variable. As Coad (2010) states, "employment growth can be seen not only as an input (in the production process) but also as an output if, for example, the policy maker is interested in the generation of new jobs", and Henrekson and Johansson (2010) have shown the particular role played by high growth firms in job creation.

We implicitly assume an ideal pattern linking, unidirectionally, R&D to innovation to productivity to employment growth (where direct intermediate steps linking not adjacent variables are also possible). However, our study will take into account only the first and the last ring of this chain, and our independent variable of interest will be only R&D expenditure. Of course, alternative approaches would be possible that consider at the same time three or more rings of the same chain, as in the multistep procedure by Crépon, Duguet, and Mairesse (1998), or that take into account multi-directional causation processes, as in the panel VAR approach followed by Coad and Rao (2010). Instead, we estimate the relation between R&D expenditure and firm growth without considering the intermediate logical steps in terms of innovation success and productivity changes, and without considering the potential "feedback" effect of the influence of firm growth on R&D. By doing so, we are technically close to the recent works by Hölzl (2009) and Hölzl and Friesenbichler (2010).

Our original contribution to the literature comes from the time span we consider after the R&D expenditure, that is the amount time allowed to R&D for exerting an influence on firm growth: we analyze the medium-term effect of R&D, and in particular the effect after five years from the investment. Such decision is due to the suspicion that many of the previous cited firm-level works fail to find appreciable influences of R&D on growth, "in contrast to aggregate evidence which clearly shows that R&D and innovation lead to higher growth at the country level" (Hölzl, 2009), because firm-level growth is often measured only one or two years after the R&D expenditure (e.g. Klomp and van Leeuwen, 2001, and Coad and Rao, 2008), while a "long time lag [is] required for a commercially valuable discovery to finally materialize in terms of growth of sales or profits" and "successful R&D may even entail further short-term costs (e.g. costs related

to product development) before yielding long-term benefits" (Coad and Rao, 2010). However, many technical problems arise when considering medium-term performance, as shown in Section 3 (on methodology). Notably, only a few studies on growth performance have considered a medium or long term, the main exceptions being an analysis by Brouwer et al. (1993) and some recent works on the effect of firm strategies on growth (e.g. Pelham and Wilson, 1996, Leitner and Gueldenberg, 2010).

2. Data and variables

For our research, we use the data from the Community Innovation Survey (CIS) that refer to the Netherlands. The CIS is a firm level survey conducted every 4 years in all EU member states (plus son non-EU countries like Norway and Iceland). We consider the four waves of the innovation survey conducted in 1996, 1998, 2000 and 2002. As we will explain later, the original data will be used for applying panel data techniques (exploiting the presence of firms in more than one wave), while a reduced version of the database, in which double counting of the same firms is avoided, will be used when pooling the four waves in a unique cross-section. Table 1 provides an idea of the relative amount of firms available in the four waves, respectively for the original data (after erasing observations or which relevant information is missing) and for the pooled cross-section cases.

Table 1: Structure of the panel.
In the Pattern column, each digit corresponds to a survey year, where the years are 1996, 1998, 2000 and 2002; a dot indicates that the firm is missing in the corresponding survey, while a 1 indicates that the firm is observed in the corresponding survey. In the right part of the table, firms that were present in more than one survey year have been considered only for the oldest year. Firms having missing information on R&D expenditure, turnover, size, and affiliation to a group, have been excluded from the data.

Original data (used for panel data techniques)					Without double counting (used for the pooled cross-section)			
Frequency	Percent	Cumul.	Pattern		Frequency	Percent	Cumul.	Pattern
7325	44.37	44.37	..1.		7945	48.12	48.12	..1.
3187	19.30	63.67	.1..		5223	31.64	79.76	.1..
1272	7.70	71.37	.11.		2168	13.13	92.89	1...
1174	7.11	78.49	...1		1174	7.11	100.00	...1
771	4.67	83.16	1...		16510	100.00		xxxx
620	3.76	86.91	..11					
506	3.06	89.98	.111					
409	2.48	92.45	11..					
259	1.57	94.02	111.					
987	5.98	100.00	other					
16510	100.00		xxxx					

For each firm, we computed R&D intensity as the ratio between R&D expenditure (survey variables *uitota*, *uitota*, *rtot* and *rtot* for 1996, 1998, 2000 and 2002 respectively) and turnover (survey variables *omztot96*, *omz98imp*, *turn*, *turn02* for 1996, 1998, 2000 and 2002 respectively). We exclude from the analysis all the firms for which the ratio between R&D and turnover is higher than one, and we use a logarithmic transformation to obtain the variable *RD* that will be used in the rest of the analysis as our measure of R&D intensity. Figure 1 shows the (unconditional) distribution of *RD* when pooling all the observations. Apart from the right-truncation in zero (due to our exclusion of firms having R&D expenditure higher than turnover), the distribution resembles a Gaussian.

If we name *t* each year in which the survey has been conducted, the corresponding medium-term firm performance is computed as the firm growth between *t+1* and *t+5*, where firm size is proxied by firm employment plus one, and the data on employment have been retrieved by matching the CIS data with the data of the Business Register, a census of the whole Dutch firm population conducted every year. By matching with the Business Register, that contains yearly information on the whole population of firms registered for fiscal purposes in the Netherlands, we are able to check the survival of firms, and to measure the growth rate of surviving firms, during the five years following the CIS survey wave in which the same firms were surveyed.

To define firm growth for each firm *i* and year *t* (where *t*=1996, 1998, 2000, 2002), we start from the expression of relative firm growth (subsequent to the R&D expenditure):

$$relgrowth_{it} = \frac{size_{it+5} - size_{it+1}}{size_{it+1}}$$

which can have values between -1 and +∞, and we transform it in the following way:

$$g_{it} = \log\left(relgrowth_{it} + 2\right) \tag{1}$$

Such measure of growth can take only values included between 0 and +∞ (zero in case of exit). Figure 2 shows that the resulting (unconditional) distribution (obtained when pooling all the observations and not considering exits) resembles a Laplace (in line with the findings of Stanley et al., 1996, and Axtell, 2001, who use a log size difference approximation of growth) and looks symmetric in the body, while its left tail is truncated in zero, and its right tail is very long to include some episodes of outstandingly high growth. The measure obtained in (1) will be the growth proxy used in the rest of our study.

A descriptive summary of the main variables used in our analysis is reported in Table 2 (firm size and growth) and in Table 3 (R&D). In Table 4, descriptive statistics are reported for growth rates and R&D intensity after erasing observations that concern exits and cases of R&D expenditure equal to zero. This allows for a direct comparison with the bottom part of Table 2 (statistics on growth, excluding exits but including cases of zero

R&D) and the bottom part of Table 3 (statistics on R&D intensity, excluding cases of zero R&D but including exits). By means of such comparison, it is possible to find out that not only the mean, but also the first and third quartile, of the growth rate distribution become higher after excluding the firms that have zero R&D, which could signal a positive effect of R&D expenditure on the medium-term growth of surviving firms. Analogously, the mean, as well as the first and third quartile, of the R&D intensity distribution are shifted upward when excluding exits, indicating that surviving firms seem to be more likely than exiting firms to have previously had a positive R&D expenditure.

Figure 1: Density plot of R&D (computed as logarithm of R&D expenditure over turnover, firms with R&D equal to zero have not been considered)
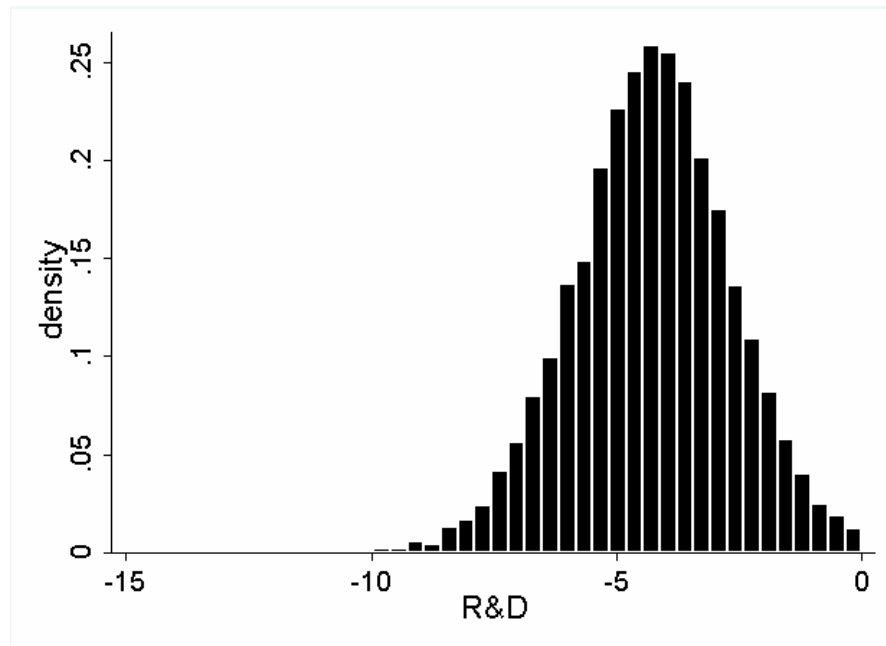

Figure 2: Density plot of growth, excluding exits (computed as logarithm of relative growth plus two).
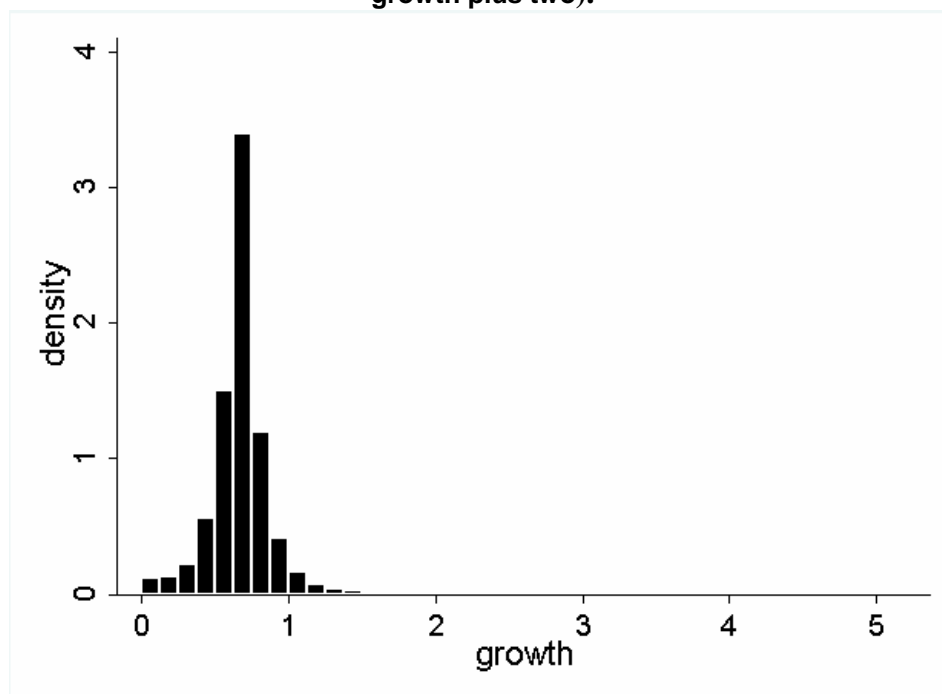
Table 2: Descriptive statistics on firm size and growth.

In the right part of the table, firms that were present in more than one survey year have been considered only for the oldest year.

	Original data (used for panel data techniques)				Without double counting (used for the pooled cross-section)		

Statistics on firm log size
(computed as logarithm of firm employment plus one)

quantiles				quantiles			
0.01	0.6931	mean	3.9605	0.01	0.6931	mean	3.6981
0.05	1.3863	variance	1.9649	0.05	1.0986	variance	1.9286
0.10	2.0794	skewness	0.0413	0.10	1.7918	skewness	0.1003
0.25	3.0910	kurtosis	3.6974	0.25	2.8332	kurtosis	3.5392
0.50	4.0775			0.50	3.7612		
0.75	4.7958			0.75	4.5433		
0.90	5.6058			0.90	5.3327		
0.95	6.2285			0.95	5.9610		
0.99	7.5000	no. obs.	22024	0.99	7.2218	no. obs.	16510

Statistics on growth, excluding exits
(computed as logarithm of relative growth plus two)

quantiles				quantiles			
0.01	0.0755	mean	0.6879	0.01	0.0741	mean	0.6940
0.05	0.3413	variance	0.0630	0.05	0.3280	variance	0.0710
0.10	0.4580	skewness	3.3208	0.10	0.4479	skewness	3.3251
0.25	0.5947	kurtosis	36.0475	0.25	0.5947	kurtosis	33.9296
0.50	0.6931			0.50	0.6931		
0.75	0.7563			0.75	0.7630		
0.90	0.8873			0.90	0.9057		
0.95	1.0116			0.95	1.0474		
0.99	1.5404	no. obs.	18303	0.99	1.6094	no. obs.	13660

Table 3: Descriptive statistics on R&D.

In the right part of the table, firms that were present in more than one survey year have been considered only for the oldest year.

Original data (used for panel data techniques)					Without double counting (used for the pooled cross-section)				

Statistics on the ratio between R&D expenditure and turnover, excluding cases of R&D equal to zero

quantiles					quantiles				
0.01	0.0002	mean	0.0443		0.01	0.0003	mean	0.0482	
0.05	0.0009	variance	0.0089		0.05	0.0010	variance	0.0101	
0.10	0.0017	skewness	5.0585		0.10	0.0018	skewness	4.7600	
0.25	0.0049	kurtosis	35.4893		0.25	0.0051	kurtosis	31.3011	
0.50	0.0142				0.50	0.0152			
0.75	0.0397				0.75	0.0440			
0.90	0.1057				0.90	0.1156			
0.95	0.1866				0.95	0.2005			
0.99	0.5143	no. obs.	13694		0.99	0.5559	no. obs.	9559	

Statistics on R&D intensity, excluding cases of R&D equal to zero
(computed as logarithm of the ratio between R&D expenditure and turnover)

quantiles					quantiles				
0.01	-8.3111	mean	-4.2989		0.01	-8.2373	mean	-4.2280	
0.05	-7.0092	variance	2.6339		0.05	-6.9417	variance	2.6881	
0.10	-6.3771	skewness	-0.2692		0.10	-6.3392	skewness	-0.2555	
0.25	-5.3284	kurtosis	3.5949		0.25	-5.2827	kurtosis	3.4675	
0.50	-4.2530				0.50	-4.1893			
0.75	-3.2268				0.75	-3.1246			
0.90	-2.2470				0.90	-2.1577			
0.95	-1.6786				0.95	-1.6071			
0.99	-0.6649	no. obs.	13694		0.99	-0.5872	no. obs.	9559	

Table 4: Descriptive statistics R&D and growth, when excluding both exits and firms having no R&D expenditure.

In the right part of the table, firms that were present in more than one survey year have been considered only for the oldest year.

	Original data (used for panel data techniques)				Without double counting (used for the pooled cross-section)		

Statistics on growth, excluding exits and cases of R&D equal to zero
(computed as logarithm of relative growth plus two)

quantiles				quantiles			
0.01	0.0741	mean	0.6926	0.01	0.0741	mean	0.7024
0.05	0.3589	variance	0.0580	0.05	0.3448	variance	0.0669
0.10	0.4700	skewness	3.3015	0.10	0.4661	skewness	3.3911
0.25	0.6022	kurtosis	40.8928	0.25	0.6061	kurtosis	39.0552
0.50	0.6931			0.50	0.6931		
0.75	0.7650			0.75	0.7765		
0.90	0.8920			0.90	0.9163		
0.95	1.0116			0.95	1.0488		
0.99	1.4758	no. obs.	11500	0.99	1.5464	no. obs.	7975

Statistics on R&D intensity, excluding exits and cases of R&D equal to zero
(computed as logarithm of the ratio between R&D expenditure and turnover)

quantiles				quantiles			
0.01	-8.1925	mean	-4.2812	0.01	-8.1180	mean	-4.2092
0.05	-6.9620	variance	2.5583	0.05	-6.9064	variance	2.5996
0.10	-6.3392	skewness	-0.2279	0.10	-6.2969	skewness	-0.1916
0.25	-5.3041	kurtosis	3.4298	0.25	-5.2601	kurtosis	3.1693
0.50	-4.2399			0.50	-4.1847		
0.75	-3.2233			0.75	-3.1246		
0.90	-2.2521			0.90	-2.1518		
0.95	-1.6806			0.95	-1.6090		
0.99	-0.6841	no. obs.	11500	0.99	-0.6216	no. obs.	7975

3. Methodology

Of the 22024 firms observed in the four CIS survey waves and matched with ABR data (including firms present in more than one wave), 3721 have exited during the five years following the survey. Given the medium-term span on which we measure performance, the decision of balancing the panel, and thus exclude from the analysis the exiting firms, would result empirically into a strong reduction of the amount of data used, and theoretically into neglecting the influence that R&D (and in general the whole innovation process) has on firm survival, an influence already shown on similar data by Cefis and Marsili (2005).

We face two problems of variable left-limitation: the one of the dependent variable (firm growth) and the other of the independent variable of interest (R&D intensity). The typical way of dealing with such problem is through the limited variable regression models named Tobit, and in particular either the original Tobit model (Tobit type I, introduced by Tobit, 1958) or its alternate version usually employed for correcting possible selection biases (Tobit type II, also known as Heckit, introduced by Heckman, 1979, and homogenized in the Tobit framework by Amemiya, 1984). The choice between Tobit type I and Tobit type II should be based on the assumptions made about the variable limitation: is the limit value observed for some individuals (censored observations) deriving from the same process that causes the non-limit value for other individuals (noncensored observations)? Rephrasing for our two cases of left-limitation, the question becomes respectively: "Are the firm exits from the market deriving from the same process that defines the growth of surviving firms?" and "Is the decision of declaring no R&D expenditure deriving from the same process that defines the amount of money spent on R&D by firms that declare an R&D expenditure?".
We will deal with the growth variable limitation and with the R&D variable limitations in two different ways.

For the growth variable, we assume that firms exiting the market are firms that have experienced strong negative growth rates (relative growth rates lower than -100%, i.e. values of our growth measure lower than zero). In other words, we assume that exit from the market and growth rates of surviving firms are governed by the same process (i.e. by the same relation with the independent variables). The natural consequence of our assumption is adopting a Tobit type I model for explaining exit and growth; we thus distance ourselves from the studies of Hall (1987) and Evans (1987) who instead choose a Tobit type II model. When analyzing growth with panel data techniques (with fixed effects), we will use the Tobit type I panel data version introduced by Honoré (1992, with the corresponding function *pantob* of the Stata software package) instead of the traditional Tobit type I.

For the R&D variable, we assume that the process leading firms to have (and declare) an R&D expenditure higher than zero is different from the (logically subsequent) process leading firms to decide how much money shall be spent on R&D (and declared in the CIS survey). Therefore, we have two separate equations (respectively, selection and outcome equation of a Tobit type II model) defining respectively the decision of having (and

declaring) an R&D expenditure and the amount of money declared to be invested in R&D, where the two equations are allowed to have different independent variables, and different parameters for the same independent variables. Such solution was already used by Hall, Lotti and Mairesse (2009), who estimate a Heckman selection model in which first a probit model is estimated to evaluate the influence of some selection variables on the probability of having an observed R&D higher than zero, and then regress the observed R&D intensity on some control variables plus the inverse Mills' ratio and other byproducts of the probit estimation (to correct for the selection bias). After obtaining an estimation of the parameters of this second equation, it is possible to build a theoretical value of the predicted R&D intensity which is able to proxy the R&D effort also for firms whose observed R&D expenditure is equal to zero. This is particularly useful in cases like ours where R&D expenditure is equal to zero for many firms, especially the smallest ones for which a proper assessment of the R&D effort is not easy. Using the predicted R&D intensity rather than the observed R&D intensity can partially solve such problem (Hall, Lotti and Mairesse, 2009). Unfortunately, the additional variables we use to explain the probability of having an observed R&D higher than zero (selection equation) are available only for a subset of observations, and therefore we will use, in different models, both the observed R&D intensity and the estimated one.

Summing up, we will estimate four different models:

Model 1: pooled cross-section, Tobit type I model for growth, considering only firms having an observed R&D intensity higher than zero.

We pool the four waves in a unique cross-section, and we assume that a latent variable y_{it} is, for each firm, linearly related to the independent variables, and is linked to the observed firm growth g_{it}, computed as in (1), as in the following:

$$y_{it} = \alpha + \beta' x_{it} + u_{it}$$
$$g_{it} = \begin{cases} y_{it} & \text{if} \quad y_{it} > 0 \\ 0 & \text{if} \quad y_{it} \leq 0 \end{cases}$$

This is tantamount to saying that exiting firms (i.e. firms for which $g_{it} = 0$) are firms for which the latent variable y_{it} assumes nonpositive values. The vector of independent variables is structured as

$$x_{it} = \begin{bmatrix} logsize_{it} & RD_{it} & group_i & sector_i & wave_i \end{bmatrix}$$

where, for each firm i and time t, $logsize_{it}$ is equal to the logarithm of firm employment plus one, RD_{it} is the R&D intensity defined as in Section 2, $group_i$ is a dummy variable taking value equal to one if the firm i is part of a bigger industrial group and zero otherwise, and $sector_i$ is a vector of 51 dummy variables, each one associated to a given

2-digit sector, assuming value equal to one if the firm *i* belongs to the given sector and zero otherwise, $wave_i$ is a vector of dummy variables, each one associated to the survey wave to which the observation belongs. To avoid double counting of the same firms in the pooled cross-section, for firms that were present in more than one survey wave, only the observations pertaining to the oldest wave are kept, thus reducing the number of observations to 16510 (i.e. exactly the total number of firms present in the database after cleaning the data).

Model 2: pooled cross-section, Tobit type I model for growth, considering for all firms the estimated R&D intensity instead of the observed one. Estimated R&D intensity is predicted by a Tobit type II model in which observed R&D intensity is the dependent variable of the outcome equation.

In Model 2, the difference with respect to Model 1 lies in the vector of independent variables

$$x_{it} = \left[logsize_{it} \quad \widehat{RD}_{it} \quad group_i \quad sector_i \quad wave_i \right]$$

which does not contain the observed log R&D intensity RD_{it}, and contains instead an estimation $\widehat{RD}_{it}$ obtained *á la* Heckman (Heckman, 1979; also named Tobit type II by Amemiya, 1984) following the procedure by Hall, Lotti and Mairesse (2009). In particular, we first estimate a probit selection equation in which the dependent variable is the probability of having an observed R&D higher than zero

$$RDP_{it} = \begin{cases} 1 & \text{if} \quad RDP_{it}^* = w_{it}\gamma + \varepsilon_{it} > \overline{c} \\ 0 & \text{if} \quad RDP_{it}^* = w_{it}\gamma + \varepsilon_{it} \leq \overline{c} \end{cases}$$

where RDP_{it} is an (observable) indicator function that takes value 1 if firm *i* reports positive R&D expenditures, RDP_{it}^* is a latent indicator variable such that firm *i* decides to perform R&D expenditures if RDP_{it}^* is above a given threshold $\overline{c}$, ε_{it} is the error term, and w_{it} is a set of explanatory variables affecting R&D, including firm log size and the CIS subjective variables that refer to the problems encountered by the firms in implementing innovation strategies. Namely, the additional variables are *per* (not sufficiently qualified personnel), *kno* (insufficient knowledge), *lfi* (lack of financing), *cos* (high costs), *mar* (uncertain market), *reg* (restrictive regulation), *org* (not flexible organization), *ris* (financial risk).[1]

---

[1] All these hampering factors are listed in the Dutch CIS metadata as "knelpunten" (English: "bottlenecks"). Their coding, as well as the values they can assume, change across different survey waves, so that some simple homogenization was necessary to obtain the final eight additional dummy variables used in the selection equation.

We then consider only the observations for which the observed R&D is higher than zero, and estimate the outcome equation

$$RD_{it} = z_{it}\theta + e_{it} \qquad \text{if} \qquad RDP_{it} = 1 \tag{2}$$

where z_{it} is a set of determinants of R&D expenditures, including size and the following byproducts of the previous probit estimation: the obtained predicted probability of having R&D higher than zero (*eventprob*) and the corresponding Mills' ratio (*imr*), as well as their squares and their product. After obtaining an estimation of the coefficient θ, and given that we know z_{it} for all the observations (including the ones for which the declared R&D was zero), we can then compute a predicted value of the R&D intensity (say, the R&D effort) $\widehat{RD_{it}} = z_{it}\theta$ for all the observations. The predicted R&D intensity $\widehat{RD_{it}}$ will then be included in the vector x_{it} of the independent variables of the growth equation (in the place of the observed *RD* used in Model 1).

Notice that the two stages (selection and outcome) which compose this extended version of the Heckman procedure combine the information about declaring or not an observed R&D expenditure higher than zero, on the one hand, and the information on the amount of R&D expenditure, on the other hand, in order to infer the R&D effort of all the firms (see Hall, Lotti and Mairesse, 2009, for details). Such estimated R&D effort is then used as the independent variable of interest in the final equation, where growth is the dependent variable.

In both the probit selection model and the outcome equation, firm sector dummies and survey wave dummies have been included. As in model 1, only the observations pertaining to the oldest wave has been kept, thus reducing the number of observations to 16510 (i.e. exactly the number of firms present in the database after cleaning the data). The number of observations is further reduced to 4445, because of the additional variables needed for the selection equation (and not available for all the firms). The outcome equation is estimated on the 3794 observations for which the observed R&D is higher than zero, and allows the estimation of the R&D effort for all the 4445 observations used in the selection equation. The same 4445 observations are eventually used in the final equation where the estimated R&D effort is the independent variable of interest.

Model 3: panel data, Tobit type I model with fixed effects, considering only firms having an observed R&D intensity higher than zero.

In Model 3, the only difference with respect to Model 1 lies in the fact that now the four cross-sectional waves are not pooled in a unique cross-section, but are analyzed as panel data with fixed effects. In particular, the model is estimated by using the Stata function *pantob* that refers to the model by Honoré (1992), which allows for fixed effects in truncated regression models. Such model allows to explore the "within" effect of R&D

on growth, instead of the "between" effect estimated in the pooled cross-sections of the previous models. Neither sectoral nor survey wave dummy variables have been included in Model 3, as the fixed effects method is aimed at erasing idiosyncracies at firm-level (thus, no need to care for sectoral idiosincracies), and is examining changes over time in the firm performances (thus, no need to care for cross-correlations). To allow the model for estimating fixed effects, we use only the data pertaining firms that have been surveyed in more than one CIS wave. The panel is reduced to 6593 observations from the original 22024, corresponding to 2747 firms from the original 16510.

Model 4: panel data, Tobit type I model with fixed effects, considering for all firms the estimated R&D intensity instead of the observed one. Estimated R&D intensity is predicted by a Tobit type II model in which observed R&D intensity is the dependent variable of the outcome equation.

In Model 4, the only difference with respect to Model 2 lies in the fact that now the four cross-sectional waves are not pooled in a unique cross-section, but are analyzed as panel data with fixed effects. As in model 3, the model is estimated by using the Stata function *pantob* that refers to the model by Honoré (1992), which allows for fixed effects in truncated regression models to explore the "within" effect of R&D on growth. As in model 2, estimated R&D is used in the place of observed R&D. The estimation of the R&D effort works in the same way of model 2, that is applying the Heckman procedure to a pooled cross-section in order to obtain a predicted value of R&D intensity for all the observations. However, now the pooled cross-section is including double-counted firms (firms surveyed in more than one wave) for consistency with the fixed effects procedure that will take place in the final step of the model. 6117 observations are thus used in the selection equation (i.e. the observations for which the information on the additional variables requested by the Heckman procedure is available, including firms present in more than one wave), and 5284 of them are employed for the outcome equation (i.e. the ones in which the firm has declared a R&D expenditure higher than zero). This allows to estimate the R&D effort for all the 6117 observations used in the selection equation. However, only 1951 observations pertain to firms that have been surveyed in more than one wave (883 firms) and will be used in the final equation with fixed effects, where the estimated R&D effort is the independent variable of interest, and growth is the dependent variable. As for Model 3, neither sectoral nor survey wave dummy variables have been included in Model 4.

4. Results

The regression results obtained for the four models are shown in Table 5.[2] In Model 1, the parameter referring to *RD* is not significant. The effect of firm size is negative and

---

[2] All regressions were run by using the Stata software package. In particular, for Models 1 and 2, the *tobit* function has been used with the option suffix ", ll(0) vce(bootstrap, rep(100))"; and, for Models 3 and 4, the *pantob* function has been used with the option suffix ", bootstrap". No bootstrap was instead implemented for the OLS and Probit estimations.

strongly significant, in line with the literature (since Hymer and Pashigian, 1962). Belonging to a group seems also to exert a significant negative effect on growth.

The influence of *RD* is still nonsignificant after using its predicted value instead of its observed value, i.e. after estimating in the analysis also the innovation effort of firms that do report zero R&D expenditure, and controlling for the selection bias. In Table 5 only the results on the final equation are reported; see Table 6 (selection equation) and Table 7 (outcome equation) for the results of the Heckman procedure preliminary to the estimation of Model 2. The probability of declaring an R&D expenditure higher than zero is positively related to the declaration of having an insufficiently trained personnel and of facing financial risks of flexibility in the firm organization (variables *per* and *ris*, Table 6). This information is used by inserting in the outcome equation the variables *imr* and *eventprob* derived from the selection equation. However, the selection bias do not seem to be strong (see the nonsignificant coefficients in the left part of Table 7) nor does the R&D become significantly effective when being proxied by its predicted value (second column of Table 5). The information jointly provided by the first two columns of Table 5, corresponding to the results for respectively Model 1 and Model 2, suggests that R&D intensity does not seem to be a good variable for discerning between firms that survive and grow, on the one hand, and firms that do not grow or do not survive, on the other hand. In other words, the pooled cross-section results do not assign particular value to the variable *RD*, which means that *RD* does not explain performance differences across firms.

The third column of Table 4 shows the results for Model 3, where the effect of *RD* becomes positive and significant. While the results of Models 1 and 2 have shown that firms having a higher R&D intensity are not likely to have a higher medium-term growth (as shown by Models 1 and 2), the result of Model 3 (a fixed-effect model) indicates that a firm which decides to raise the level of R&D will be likely to raise its medium-term growth ("within" effect). The effect of size is still negative and significant (as in the cross-sectional analysis), while belonging to a group is now shown to be a factor that can hamper growth. Once we consider the estimated R&D effort instead of the declared R&D intensity, the main result does not change (Model 4, last column of Table 5): the R&D coefficient is significant for the fixed-effect model, i.e. an increase in the R&D intensity positively affects the performance path of a firm along time.

Table 5: Estimation results.

In Model 2, predicted R&D is used instead of observed R&D. The estimation procedure for model 3 with export among the regressors did not converge. Dummy variables relating to 2-digit sectors have been included in all models, and dummy variables relating to the cross-sectional waves have been included in models 1 and 2, i.e. in the pooled cross-section models. Standard errors in brackets below the parameter estimates; * significant at 10%, ** significant at 5%, significant at 1%.

	Model 1	Model 2	Model 3	Model 4
RD	0.00402	-0.00005	0.00727**	0.14582**
	(0.00297)	(0.00246)	(0.00339)	(0.05987)
logsize	-0.02002***	-0.00526	-0.33967***	-0.26545***
	(0.00421)	(0.00592)	(0.03939)	(0.07200)
group	-0.03038***	-0.04112***	0.04227***	0.07632**
	(0.01061)	(0.01529)	(0.01101)	(0.03237)
no. units	9559	4445	2747	883
no. observ.	9559	4445	6593	1951
of which:				
censored	1584	827	733	230
uncensored	7975	3618	5860	1721
chi2-stat	9065.74	338.33	99.06	60.47
prob>chi2	0.00	0.00	0.00	0.00

Table 6: Heckman selection equation probit estimates (dependent variable: probability of having an R&D expenditure higher than zero).
See the Appendix for the definition of the variables. Standard errors in brackets on the right side of the parameter estimates; * significant at 10%, ** significant at 5%, significant at 1%.

	selection equation (dep. variable: probability of declaring a R&D expend. higher than zero)	
	without double counting (for Model 2 - pooled cross)	with double counting (for Model 4 - panel data)
logsize	0.16485 ***	0.22543 ***
	(0.02322)	(0.01984)
group	0.16376 ***	0.09472 *
	(0.06390)	(0.05534)
per	0.13891 *	0.13384 *
	(0.08037)	(0.07567)
kno	0.02721	0.01341
	(0.02721)	(0.07781)
lfi	0.10811	0.10472 *
	(0.06587)	(0.05586)
cos	0.01178	0.01954
	(0.07091)	(0.06079)
mar	-0.08318	-0.06587
	(0.06520)	(0.05493)
reg	0.09012	0.06254
	(0.06785)	(0.05733)
org	-0.01631	-0.01569
	(0.06415)	(0.05385)
ris	0.11423 *	0.07720
	(0.06568)	(0.05647)
no. observ.	4445	6117
chi2-stat	1131.87	1315.09
prob>chi2	0.00	0.00

Table 7: Heckman outcome equation estimates (dependent variable: observed R&D expenditure).

Eventprob = predicted probability of.having R&D expenditure higher than zero; imr = inverse Mill's ratio; crossprod = eventprob times imr (eventprob and imr resulting from the selection equation estimates shown in table A2). Standard errors in brackets below the parameter estimates; * significant at 10%, ** significant at 5%, significant at 1%.

	outcome equation (dep. variable: observed R&D intensity)			
	without double counting (for Model 2 - pooled cross)		with double counting (for Model 4 - panel data)	
logsize	-0.41702	***	-0.37910	***
	(0.03282)		(0.03668)	
group	-0.06219		0.01038	
	(0.06507)		(0.05185)	
eventprob	-1823.068		-540.22	
	(1344.12)		(807.77)	
imr	-1208.417		-358.59	
	(842.426)		(500.14)	
eventprob^2	885.029		261.58	
	(663.401)		(398.79)	
imr^2	387.747		114.88	
	(260.547)		(152.75)	
eventprob*imr	1172.375		345.56	
	(831.164)		(493.64)	
no. observ.	3794		5284	
F-stat	22.31		27.83	
prob>F	0.00		0.00	

5. Conclusion

Our analysis shows that R&D expenditure exerts a positive influence on firm employment in the medium-term (five years after the investment). However, the influence appears only when considering fixed effects methods, that allow for firm idiosyncracies in the firm growth process. The economic implication is: R&D intensity does not explain performance differences across firms, but explains, instead, changes along time of the performance of a given firm. An increase in the R&D intensity will make a firm deviate upward in its performance path, where performance is meant to be not only growth after survival, but also the probability of survival itself. Indeed, throughout the whole analysis, the problem of firm exits has been explicitly addressed by means of truncated regression models. At the same time, our results are confirmed when using a predicted value of R&D intensity, instead of the observed one, to assess a measure of R&D effort also for firms that report zero R&D expenditure.

Our study suggests that the influence of R&D on employment growth can easily be underestimated for two intertwined reasons. On the one hand, R&D needs some time to show its effects, and a short-term analysis based on the observation of employment growth during only one or two years after the investment could not be sufficient to capture all the causal links connecting R&D to employment. On the other hand, the effect that R&D exerts on firm survival cannot be ignored, especially when using firm-level analyses to predict the aggregate outcome of innovation policies at regional or country scale.

References

Acs, Z.J. & Audretsch, D.B., 1990. The Determinants of Small-Firm Growth in U.S. Manufacturing. *Applied Economics*, 22(2), 143-53.

Acs, Z.J. & Audretsch, D.B., 1988. Innovation in Large and Small Firms: An Empirical Analysis. *The American Economic Review*, 78(4), 678-690.

Amemiya, T., 1984. Tobit models: A survey. *Journal of Econometrics*, 24(1-2), 3-61.

Axtell, R.L., 2001. Zipf Distribution of U.S. Firm Sizes. *Science*, 293(5536), 1818-1820.

Bernard, A.B. & Jensen, J.B., 2004. Why Some Firms Export. *Review of Economics and Statistics*, 86(2), 561-569.

Brouwer, E. & Kleinknecht, A. & Reijnen, J.O.N., 1993. Employment Growth and Innovation at the Firm Level. *Journal of Evolutionary Economics*, Springer, 3(2), 153-159.

Cassiman, B., Golovko, E. & Martínez-Ros, E., 2010. Innovation, exports and productivity. *International Journal of Industrial Organization*, 28(4), 372-376.

Cefis, E. & Orsenigo, L., 2001. The persistence of innovative activities A cross-countries and cross-sectors comparative analysis. *Research Policy*, 30(7), 1139-1158.

Cefis, E. & Marsili, O., 2005. A matter of life and death: innovation and firm survival. *Industrial and Corporate Change*, 14(6), 1167-1192.

Coad, A. & Rao, R., 2008. Innovation and firm growth in high-tech sectors: A quantile regression approach. *Research Policy*, 37(4), 633-648.

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

Crépon, B., Duguet, E. & Mairessec, J., 1998. Research, Innovation And Productivi[Ty: An Econometric Analysis At The Firm Level. *Economics of Innovation and New Technology*, 7(2), 115-158.

Evans, D.S., 1987. The Relationship Between Firm Growth, Size, and Age: Estimates for 100 Manufacturing Industries. *The Journal of Industrial Economics*, 35(4), 567-581.

Geroski, P.A., 2005. Understanding the implications of empirical work on corporate growth rates. *Managerial and Decision Economics*, 26(2), 129-138.

Gibrat, R., 1931. *Les Inégalités économiques*, Paris: Recueil Sirey.

Hall, B.H., 1987. The Relationship Between Firm Size and Firm Growth in the US Manufacturing Sector. *The Journal of Industrial Economics*, 35(4), 583-606.

Hall, B.H., Lotti, F. & Mairesse, J., 2009. Innovation and productivity in SMEs: empirical evidence for Italy. *Small Business Economics*, 33(1), 13-33.

Heckman, J.J., 1979. Sample Selection Bias as a Specification Error. *Econometrica*, 47(1), 153.

Henrekson, M. & Johansson, D., 2009. Gazelles as job creators: a survey and interpretation of the evidence. *Small Business Economics*, 35(2), 227-244.

Hölzl, W., 2009. Is the R&D behaviour of fast-growing SMEs different? Evidence from CIS III data for 16 countries. *Small Business Economics*, 33(1), 59-75.

Hölzl, W. & Friesenbichler, K., 2010. High-growth firms, innovation and the distance to the frontier. *Economics Bulletin*, 30(2), 1016-1024.

Honoré, B.E., 1992. Trimmed Lad and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects. *Econometrica*, 60(3), 533-565.

Hymer, S. & Pashigian, P., 1962. Firm Size and Rate of Growth. *Journal of Political Economy*, 70.

Klomp, L. & Van Leeuwen, G., 2001. Linking Innovation and Firm Performance: A New Approach. *International Journal of the Economics of Business*, 8(3), 343-364.

Kox, H.L.M. & Rojas-Romagosa, H., 2010. Exports and Productivity Selection Effects for Dutch Firms. *De Economist*, 158(3), 295-322.

Leitner, K. & Güldenberg, S., 2009. Generic strategies and firm performance in SMEs: a longitudinal study of Austrian SMEs. *Small Business Economics*, 35(2), 169-189.

Pelham, A.M. & Wilson, D.T., 1996. A longitudinal study of the impact of market structure, firm structure, strategy, and market orientation culture on dimensions of small-firm performance. *Journal of the Academy of Marketing Science*, 24(1), 27-43.

Reichstein, T., Dahl, M., Ebersberger, B. & Jensen, M., 2010. The devil dwells in the tails. *Journal of Evolutionary Economics* 20(2), 219-231.

Stam, E., 2010. Growth beyond Gibrat: firm growth processes and strategies. *Small Business Economics*, 35(2), 129-135.

Stam, E. & Wennberg, K., 2009. The roles of R&D in new firm growth. *Small Business Economics*, 33(1), 77-89.

Tobin, J., 1958. Estimation of Relationships for Limited Dependent Variables. *Econometrica*, 26(1), 24.