Economic impacts of internal and external R&D processes over time
How and when sales of new products are increased *

Thor Donsby Noe † [a]

[a]Universitat de Barcelona, Spain

January 7, 2019

Abstract

In this paper Spanish manufacturing firms are analyzed to access the timing and impact size of different Research & Development processes. Lagged and persistent effects are accessed using leads of the share of new products in total turnover of the firm as the dependent variable. The Between, Fixed Effects, and Random Effects estimators are applied and compared to provide robustness. Internal R&D spending is found to have the larger and more persistent economic impact. However, for firms without internal R&D expenses investment in external R&D processes can be a quick fix that helps introduce new products into the market even within the same year, though without building up a know how.

*I am thankful to Vicente Royuela Mora for his helpful supervision and inspiring teaching of the course Panel Data. All errors remain my own.
†*E-mail address:* thornoe@gmail.com (T. D. Noe).

1 Introduction

Innovation driven growth is a topic of high interest both academically and when firms are to decide on their innovation strategies. Using Spanish firm level data from 2005-2016 the objective of this paper is to analyze how different R&D strategies can contribute differently to increase the sales of new products in the short- or medium-term. Namely the difference in ramp-up time and persistence of internal and external R&D expenses is analyzed using the different-length leads of sales of new products and Between estimation as well as Fixed Effects and Random Effects estimation[1]. Internal R&D activity is found to have the larger persistent effect while external R&D spending can be quite efficient when looking for an immediate impact.

A few main findings and the in the literature is summed up as an empirical and theoretical background in section 2 before describing the data, variables and connections in section 3. The results are presented in section 4 before concluding in 5.

2 Background

Innovation driven growth has been a contested topic throughout the era of modern economics. Sternberg and Arndt (2001) carried out a comparative study in ten different regions show that firm level innovations are both determined by internal characteristics of the individual firm and by outside determinants including both intra-regional and extra-regional factors. However, the PITEC dataset that we use has a very low level of disaggregated regional information on location of firms. On the bright side external factors has been found to be of less importance than the firm-level determinants (Sternberg and Arndt, 2001).

Segarra and Teruel (2014) uses the PITEC to quantify the impact of R%D processes not only on the probability of innovating but on economic growth of the firm. They find a significant presence of heterogeneity. Applying quantile regression internal R&D activity mainly has a sizeable effect for high-growth firms while external R&D expenses mainly benefit below-median-growth firms.

3 Data

3.1 The PITEC panel of Spanish firms

The final 2016-version of the PITEC panel contains yearly responses till 2016 from firms with at least one employee. 7,283 firms are observed since the first year of the panel survey, 2003. This first wave was constructed by joining two panels of which the first is a panel of 7,264 firms with 200 or more employees, covering 86% of all firms with 200 or more employees at the time; the other panel contains 3,794 small and medium-sized enterprises (SMEs) according to 56% of firms with less than 200 employees that

[1]The STATA do-files can be accessed from github.com/thornoe/ub/tree/master/Panel_data/paper/stata_code

carry out internal R&D activities (Vega-Jurado et al., 2009). New to the 2004 wave of the survey is the introduction of 3,040 new firms to the panel of which 437 were SMEs with external R&D expenses only and about 1,000 were SMEs without R&D expenditure.[2] Additionally 2,480 SMEs with internal R&D activities are added to the panel in the 2005 wave of the panel.

Thus, the panel is unbalanced and to include the higher number of firms I limit the time span to 2005-2016 providing a balanced panel of 12,803 firms. 46 firms are added in later waves, but these and 705 others are dropped due to missing information in 2005. Due to firms stopping to respond to the survey I need to drop 200-500 firms from the panel each of the following waves, but for 2014 and 2016 where 2,000 and 700 firms are dropped respectively, leaving us with a total of 5,662 firms with full information, 44% of the original 2005-sample. The data set does not allow us to convincingly distinguish between firms ceasing to exist, deliberate non-responses, and random non-responses. However, that 60% of firms still answered the survey in 2016 indicates a sizeable random non-response rate in the previous years. Most of the firms failing to answer the survey in one or more years are smaller firms of which half of them have 30 employees at most, even more so, heterogeneity in the response rate is also present between the different industries, especially outside of the service sector. [3] Due to low sample sizes and a expectedly much different innovation processes I drop a total of 305 firms mainly engaged in agriculture, resource extraction, recycling, energy, water, sanitation, or construction. There can be fundamental differences in the innovative processes even between the manufacturing and service sector (Hoffman et al., 1998) as well as in the outcomes (Harrison et al., 2014), thus, of the remaining 5,357 firms only the subsample of 3,013 manufacturing firms are analyzed, leaving out the 2,344 firms in the service sector.

3.2 Indicator of innovation success

The dependent variable is new products the share of new products unique to the firm measured as the share of total revenue. Thus, it is not only an indicator of innovation in products but of the commercial success of product innovations. The variable is worked out as the sale of products introduced within the same financial year as well as products introduced within the two prior years. Therefore different length of leads $k \in 0, 1, 2, 3$ is tried out in the regression such that new_{t+k} is determined by the independent variables at period t.

[2]Description of the methodology, the full questionnaires etc. in PITEC is available by the Spanish Foundation of Science & Technology at icono.fecyt.es/pitec

[3]The range is from 46% of the firms having missing information for one or more years in both the "chemicals, pharmaceuticals, plastics, ceramics & petrol" and "foods, beverages & tobacco" industries to 66% in construction, 69% of the firms in "furniture, games, toys, & other manufacturing" and 91% within "agriculture and resource extraction".

3.3 Determinants of innovation

The main time-varying determinants used are the percentage of total employees engaged in internal R%D activities $pidp_t$, a dummy $idin_t$ for having expenses on internal R&D activities, and a dummy $exter_t$ for only having external R&D expenditures.

Furthermore, for the Random Effects estimation industry dummies and regional effects dummies for Madrid, Catalonia, and Andalusia are included.

3.4 Descriptive statistics

Table 1 show mean and standard deviation of the key variables sorted by the different industries within manufacturing. Heterogeneity is present throughout including the row showing the correlations $corr(new_{t+2}, pidp_t)$.

Table 1: Descriptive statistics

	1 Foods	2 Textiles	3 Chemicals	4 Metal	5 Machinery	6 Furniture	Total
new_0	12.85 (26.61)	14.38 (28.82)	13.91 (26.93)	13.35 (27.42)	17.87 (28.35)	18.19 (31.01)	15.10 (27.82)
pidp	5.606 (10.56)	6.000 (11.40)	10.05 (13.45)	6.535 (10.45)	13.91 (18.39)	6.121 (10.95)	9.600 (14.63)
idin	0.598 (0.490)	0.515 (0.500)	0.744 (0.436)	0.643 (0.479)	0.755 (0.430)	0.570 (0.495)	0.683 (0.465)
exter	0.0348 (0.183)	0.0360 (0.186)	0.0343 (0.182)	0.0340 (0.181)	0.0383 (0.192)	0.0524 (0.223)	0.0364 (0.187)
Obs.	5,112	3,996	10,164	4,524	11,196	1,164	36,156
Corr.	0,0841	0,0997	0,0677	0,0953	0,0431	0,0985	0,0752

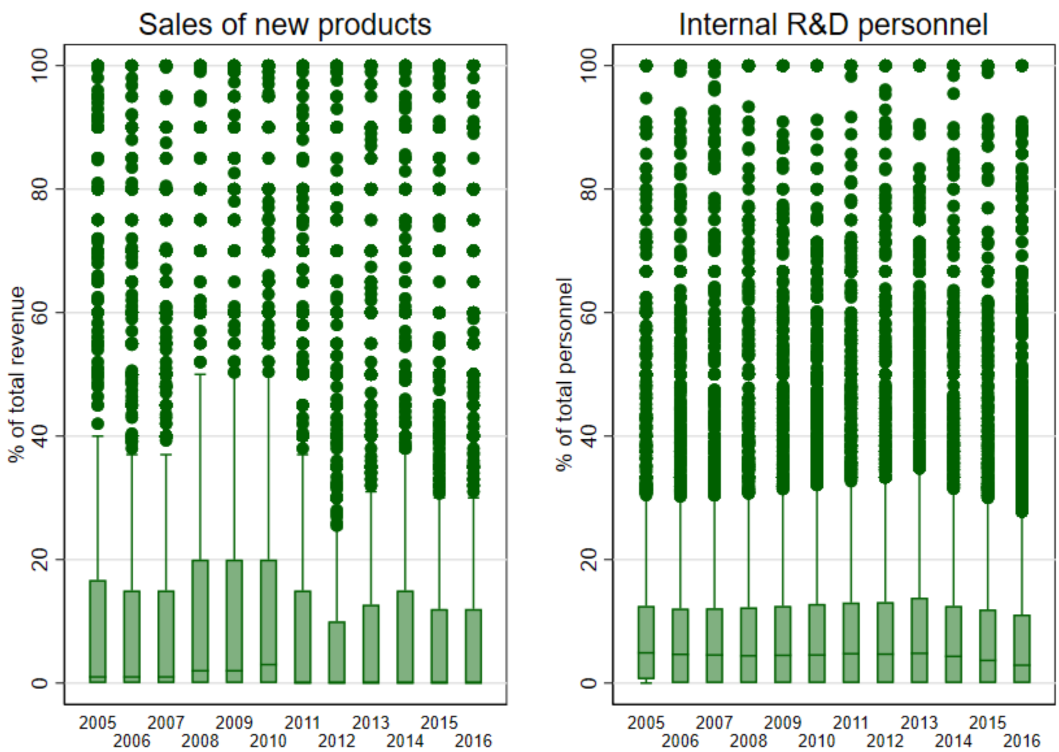
mean coefficients; sd in parentheses

Furthermore it should be noted that a stable share of sales of new products is rare, that is, the variation within firms over time is higher than the average variation between firms when it comes to new_0. As expected the between variation is the higher for $pidp_t, idin_t$ showing that the internal R&D activity of the firm is less volatile across industries.

3.5 The time aspect

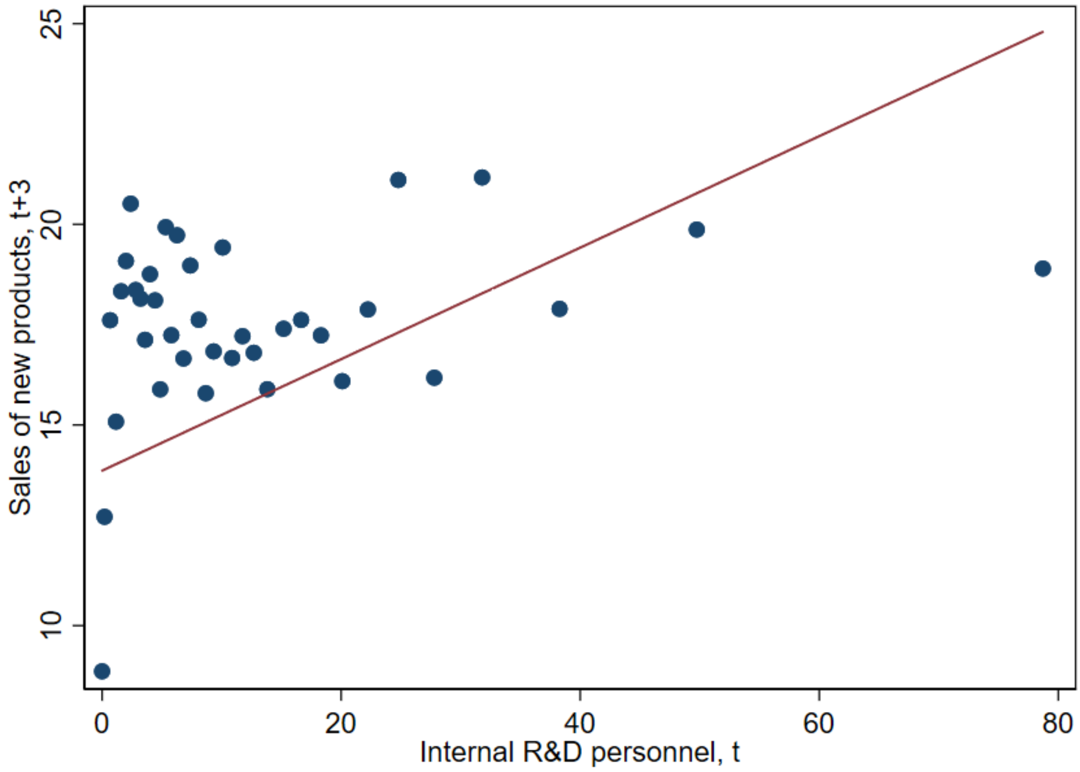
Looking at the boxplots in figure 1 it comes to mind that there exists a contrast between the sales of new products peaking in 2008-2010 and the share of internal R&D personnel being quite stable throughout the time period, just for a small decrease in 2015-2016.

Figure 1: Boxplots of aggregate changes over time



This is in accordance with the low degree of correlation shown in table 1. Furthermore, the bin scatterplot in figure 2 does not show a convincing relation either.

Figure 2: Scatter plot of internal R&D personnel and sales of new products



4 Results

4.1 Baseline estimation results

The estimated coefficients of the baseline model is shown in table 2. In the 1st panel Fixed Effects (FE) estimation is used, the 2nd panel shows the Between (BE) estimation, and the 3rd shows the Random Effects (RE) estimation.

Table 2: Different regression estimates for the base model

	(1) fe, k=0 b/se	(2) fe, k=1 b/se	(3) fe, k=2 b/se	(4) fe, k=3 b/se
pidp	-0.016	-0.005	-0.001	0.042**
	(0.018)	(0.019)	(0.020)	(0.021)
idin	6.437***	7.027***	6.576***	3.371***
	(0.506)	(0.536)	(0.576)	(0.626)
exter	6.001***	5.311***	4.644***	0.695
	(0.903)	(0.935)	(0.978)	(1.038)
cons	10.580***	8.843***	8.955***	13.009***
	(0.567)	(0.580)	(0.601)	(0.628)
r2	0.008	0.009	0.008	0.004
N	36080	33074	30070	27065
	(1) be, k=0 b/se	(2) be, k=1 b/se	(3) be, k=2 b/se	(4) be, k=3 b/se
pidp	0.023	0.018	0.014	0.005
	(0.025)	(0.025)	(0.025)	(0.026)
idin	12.115***	12.243***	12.708***	12.880***
	(0.876)	(0.893)	(0.911)	(0.937)
exter	10.947***	10.565***	10.581***	10.087***
	(2.433)	(2.414)	(2.407)	(2.419)
cons	-348.384**	-265.493**	-281.846**	-294.354**
	(143.233)	(131.524)	(127.018)	(118.921)
r2	0.086	0.085	0.081	0.076
N	36080	33074	30070	27065
	(1) re, k=0 b/se	(2) re, k=1 b/se	(3) re, k=2 b/se	(4) re, k=3 b/se
pidp	0.004	0.010	0.010	0.035**
	(0.014)	(0.015)	(0.016)	(0.016)
idin	7.763***	8.333***	8.251***	6.216***
	(0.437)	(0.458)	(0.486)	(0.519)
exter	6.766***	6.161***	5.680***	2.468***
	(0.847)	(0.872)	(0.906)	(0.954)
cons	9.356***	7.687***	7.537***	10.843***
	(0.589)	(0.598)	(0.612)	(0.631)
N	36080	33074	30070	27065

Standard errors are in parentheses. * p<0.10, ** p<0.05, *** p<0.01
Time dummies are estimated but not shown.

The estimated coefficients using the FE model show that the presence of internal R&D activities $idin_t$ is important and the sales of new products new_{t+k} are effected significantly stronger for $k = 1$ periods ahead. The effect of having external R&D expenses only $exter_t$ as opposed to no R&D expenses at all has a high effect within the same year on new_{t+0} while the size of the effect drops a little off and falls towards zero for $k = 3$ which shows that external R&D spending only have an immediate effect. On the other hand, the specification with new_{t+3} is the only one for which internal R&D personnel as a share of total employees has a significant effects which shows that the effect of internal knowledge production is lasting but might be more neglectable in the short term. However, the estimates of the Between model indicate that a high average presence of either internal or external R&D is more important in the long-run as well while the effects of the average of internal R&D personnel as a share of total personnel is insignificant. The estimates of the RE model show a pattern similar to that of the FE model.

4.2 Specification tests

First of all the Breusch Pagan LM test for homoscedasticity of random effects is clearly rejected for all lengths of leads $k \in 0, 1, 2, 3$. That is, we reject the H_0 hypothesis that the variance of the individual time-invariant random component should be zero, $Var(\hat{\psi}_i) = 0$. This indicates a heteroscedasticity issue and we should consider using heteroscedasticity robust standard errors.

The Hausman test that the RE and FE estimates is clearly rejected. This means that the baseline model is either misspecified or the endogeneity assumption of $cov(\psi_i, x_{it})$ is violated which means that the RE estimates are inconsistent. In order to reduce the endogeneity the parametrization of the RE model is improved by including a series of time-invariant controls which lowers the value of the χ^2 tests. But even when including all reasonable controls, significant or not, the Hausman is still borderline rejected with 99% confidence for the model with leads $k = 1$ which is the better throughout the specifications.

4.3 Including time-invariant controls

Though the Hausmann test points at the RE model being inconsistent, the estimates should nonetheless be more consistent. The RE approach also allows us to estimate the effects of several time-invariant variables of interest, including a surprising negative effects of being located in Madrid as opposed to any other part of Spain. The significance of the coefficients of these estimates are determined using heteroscedasticity-robust standard errors that take the clustering on firm ID over time into account. The estimated model is shown in table 3. Though all the model specifications are relevant for different time frames, the one with a one period lead $k = 1$ is the stronger model judged by the Hausman test statistic.

Table 3: Estimates with time invariant-controls

	(1) re, k=0 b/se	(2) re, k=1 b/se	(3) re, k=2 b/se	(4) re, k=3 b/se
pidp	-0.006	0.001	0.001	0.025
	(0.019)	(0.019)	(0.020)	(0.022)
idin	7.865***	8.442***	8.350***	6.291***
	(0.565)	(0.561)	(0.586)	(0.659)
exter	6.701***	6.102***	5.616***	2.407**
	(1.139)	(1.095)	(1.132)	(1.142)
Rest of Spain	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)
Madrid	-1.911**	-1.908**	-1.858**	-1.892**
	(0.959)	(0.962)	(0.946)	(0.962)
Cataluña	-0.284	-0.291	-0.065	0.254
	(0.621)	(0.634)	(0.653)	(0.688)
Andalucía	2.424*	3.186**	3.291**	3.729**
	(1.340)	(1.398)	(1.440)	(1.512)
Foods, beverages & tobacco	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)
Textiles, footwear, lumber, cardboard, paper & graphic arts	2.510**	2.711**	2.530**	2.259*
	(1.081)	(1.108)	(1.135)	(1.197)
Chemicals, pharmaceuticals, plastics, ceramics & petrol	0.195	0.208	0.111	0.205
	(0.823)	(0.849)	(0.874)	(0.915)
Metallurgy & metal manufacturing	0.262	0.530	0.590	0.659
	(0.976)	(1.009)	(1.048)	(1.105)
Electronics & machinery	4.095***	4.002***	3.944***	3.973***
	(0.844)	(0.875)	(0.900)	(0.943)
Furniture, games, toys & other manufacturing	5.490***	5.598***	5.352***	4.771***
	(1.771)	(1.836)	(1.807)	(1.814)
cons	7.743***	5.998***	5.845***	9.067***
	(0.872)	(0.884)	(0.899)	(0.958)
N	36080	33074	30070	27065

Cluster robust standard errors are in parentheses. * p<0.10, ** p<0.05, *** p<0.01
Time dummies are estimated but not shown.

5 Conclusion

The preferred model in table 3 imply that internal R&D activity in year t should raise the share of new products with around eight percentage points of total turnover not only within the same year but especially in the coming 1-2 years. In the absence of internal R&D activities external R&D expenditure can raise the share of new products significantly, but the effects is obviously less persistent. The RE estimates are as is often the case inconsistent, though more efficient than the FE estimates. Including even more regional controls and other corrections for heteroscedasticity could possibly fix this issue. However, the direction and the overall significance of the effects is recurring across the FE, RE, and to some degree the BE estimator.

References

Harrison, Rupert et al. (2014). "Does innovation stimulate employment? A firm-level analysis using comparable micro-data from four European countries." In: *International Journal of Industrial Organization* 35, pp. 29–43.

Hoffman, Kurt et al. (1998). "Small firms, R&D, technology and innovation in the UK: a literature review." In: *Technovation* 18.1, pp. 39–55.

Segarra, Agustí and Mercedes Teruel (2014). "High-growth firms and innovation: an empirical analysis for Spanish firms." In: *Small Business Economics* 43.4, pp. 805–821.

Sternberg, Rolf and Olaf Arndt (2001). "The firm or the region: what determines the innovation behavior of European firms?" In: *Economic Geography* 77.4, pp. 364–382.

Vega-Jurado, Jaider, Antonio Gutiérrez-Gracia, and Ignacio Fernández-de Lucio (2009). "Does external knowledge sourcing matter for innovation? Evidence from the Spanish manufacturing industry." In: *Industrial and corporate change* 18.4, pp. 637–670.