Firm-level contributions to the R&D intensity distribution: evidence and policy implications

Sebastiano Cattaruzzo*, Agustí Segarra-Blasco* and Mercedes Teruel*

Abstract:

This paper decomposes the Spanish aggregate R&D distribution to disentangle the contributions of R&D public financing, gazelle firms, and financial constraints. Applying the Chernozhukov, Fernández-Val and Melly (2013) distribution regression approach, we estimate the contributions of these components at each point of the distribution. The analysis is carried out for two periods, pre-crisis 2004–2008 and post-crisis 2009–2014. We thereby introduce a comparative perspective that allows us to consider possible business cycle effects. Our findings are that the main causes of the significant post-crisis drop in Spanish aggregate R&D are changes in the public financing scheme and in the decreased contribution of gazelles. Our results provide a rigorous analysis of Spanish R&D, hint at a possible transmission channel for reduced business dynamism and offers interesting insights for policymaking.

Keywords: R&D, decomposition, distribution, intensity, gazelles

JEL codes: O30, O33, L20

Corresponding author email: sebastiano.cattaruzzo@urv.cat

(*) Universitat Rovira i Virgili, Department of Economics, GRIT & ECO-SOS, Av. Universitat, 1, 43204, Reus, Spain

1. Introduction

Many studies have analysed the determinants of firms' internal R&D expenditure—see, for example, (Montresor and Vezzani, 2015) or (Coad, 2019). Such studies typically find the heterogeneity of firm-level R&D investment (Coad et al., 2020), the incidence of non-observable characteristics on the R&D effort (Cohen and Klepper, 1992), and the non-homogenous nature of R&D activity (Czarnitzki and Hottenrott 2011a; Barge-Gil and López 2014). The topic is important since R&D investment expands a country's technological frontier. Thus, we need to understand how the distribution of R&D investments is influenced by different firms' characteristics and by the economic cycle.

To the best of our knowledge, there is a gap in the literature on how common firm-level factors such as receiving public subsidies, or being financially constrained, affect a country's aggregate R&D distribution and the varying effects of such factors in different phases of the economic cycle. Some recent multi-level studies have stressed the importance of considering aggregate impacts using micro-data (Di Giovanni et al., 2014).[1] Here, we propose a decomposition method based on distribution regression to explore Spanish firms undertaking R&D (henceforth, R&D firms).

Using the *Panel de Innovación Tecnológica* database (henceforth PITEC) that covers Spanish R&D firms and applying the decomposition method developed by Chernozhukov et al. (2013), we investigate the determinants of change in the distribution of R&D intensities. The focus on intensities, rather than pure quantities, is both to correct for obvious scale effects, but also to follow the model developed in Cohen and Klepper (1992). Specifically, we quantify to what extent the variations are attributable to changes of observable characteristics in relation to the total observed change. Instead of standard estimation methods centred around the mean, a precise reckoning for the whole distribution is preferable. Our approach entails estimation and inference procedures to compute appropriately chosen counterfactuals that allow the decomposition of differences (change over time in our case) in a given distribution.

[1] Over the last decade a wide number of investigations has highlighted the importance of multi-level studies and proposed estimations based on micro (or sectoral) data, but which can be quantified in relation to the most common aggregate measures (Gabaix, 2011; Acemoglu et al., 2012; Di Giovanni & Levchenko, 2010; Di Giovanni et al., 2014; Carvalho & Gabaix, 2013; Foerster et al., 2011). These studies tend to rely on either input-output matrices or quite long-term micro data.

This technique allows us to quantify both the levels and the changes in the contribution of given factors to the composition and evolution of Spanish internal R&D. In particular, we isolate three determinants: the level of R&D public financing, the share of "gazelles,"[2] and the extent of financial constraints. Also, given the 2003–2014 data coverage, we split the sample in two phases, one increasing and the other decreasing. Thus, the analysis sheds lights on the role that the above factors play in medium-term R&D aggregate fluctuations in relation to business cycle movements. Our results suggest that public financing and the contribution of gazelles are the two main determinants of the distribution's shape. The contribution from gazelles decreases considerably over time. Curiously, financial constraints impact negatively in the period of expansion, while they are unimportant in the period of contraction.

For many reasons, Spain is a perfect candidate for this analysis (COTEC, 2018). Among the four major EU economies (Italy, Spain, France, and Germany), Spain was the only one to experience a continuous contraction in investment between 2009 and 2014. Until 2008, its R&D expenditure was converging to EU levels, but it then started to diverge again. Spanish aggregate R&D expenditure has historically been a balanced mix of public and private expenditure, but public expenditure has contributed less over time. Furthermore, the EU recommends that private sector expenditure should be approximately two-thirds of the total, but it is currently only 47% in Spain.

This paper makes several contributions. Methodologically, we introduce a compelling approach for distribution decomposition. Although widely applied in labour economics, as far as we know this to be its first application in innovation studies and that it constitutes an improvement on the previous quantile regression approaches. At the theoretical level, we provide an analysis framework whose aim is to isolate the most consistent determinants of R&D investment. Empirically, we show how several factors combine in shaping the levels and evolution of aggregate R&D intensity in different phases of the business cycle. Finally, from a policy perspective, the results are relevant in shaping innovation policy toward the long-pursued 3% R&D target proposed in the 2000 Lisbon Agenda and incorporated in the policies of the Horizon 2020 program (Veugelers and Cincera, 2015). Also, they possibly point towards solutions for a

[2] Small fast-growing enterprises, c.f. David Birch, Job Creation in America: How Our Smallest Companies Put the Most People to Work.

reduction of the long-standing EU-US gap in R&D intensities. Our analysis captures the role of public subsidies, high-growth firms (henceforth HGFs), and young innovative companies (henceforth YICs) in diminishing such gaps.

The rest of the work proceeds as follows. Section 2 summarizes the most relevant literature on R&D heterogeneity and on its determinants with the aim of contextualizing our research questions. Section 3 contains a description of the data and relevant statistics, while Section 4 explains the econometric methodology. Section 5 shows our main empirical results and Section 6 concludes and discusses the policy implications.

2. Conceptual framework

Internal R&D intensity depends on characteristics that have been largely analysed in the literature (Hall and Hayashi, 1989; Hall, 1993; Aghion et al., 2005; Breschi and Malerba, 1997; Griffiths and Webster, 2010). Although firm-level structural variables (i.e., industry concentration, market share, ownership structure, or lagged performance) explain well the levels of R&D intensity (Crépon et al., 1998; Jefferson et al. 2006), this study opts for a parsimonious selection of policy-relevant variables.

Since Cohen and Levin (1989), it has been clear that firm conclusions on firm-specific determinants are intrinsically elusive. Probably, the most robust findings are the procyclical nature of R&D (Barlevy, 2007) and the fact that the R&D distribution directly affects countries' performance (Falk, 2007). This section reports the evidence on R&D heterogenous firm-level behaviours, presents the variables that may most affect the distribution, and connects these concepts to recent literature, particularly that related to reduced business dynamism.

2.1. Heterogeneity of the R&D intensity distribution

The literature on R&D intensity is extremely rich in both empirical and theoretical contributions. Although Cohen and Klepper (1992) proposed a reliable probabilistic approach to model its distribution, most studies still rely on the use of production functions augmented to accommodate knowledge capital as input. Montresor and Vezzani (2015) demonstrate clearly how this approach is fragile and how its estimations vary significantly for each point of the distribution. Indeed, one of the main conclusions reached by Cohen and Klepper (1992) was that the heterogeneity in firm-level R&D intensities could not be explained by observable firm characteristics, but rather by unobserved ones, which the authors interpreted as R&D-related expertise.

Coad (2019) shows clearly how heterogeneity in firm-level R&D intensities is an empirical fact, and that this property is resistant to sectoral disaggregation, meaning that even in in high-tech sectors there are firms whose intensity can be increased (and vice versa for low-tech sectors). Also, Evangelista (2006) presented evidence on the widespread heterogeneity in both services and manufacturing sectors, with the former surprisingly dominating the latter. This has important implications for R&D policies, as the objective should be to correctly target the policy effort and to increase overall R&D intensities, rather than trying to pursue deeper and more complicated transformations which favour only the high-tech sectors, leaving low- and medium-tech ones behind.

Factors like the supply of different product lines, the presence of varying innovation opportunities, the diversity in knowledge bases due to cumulativeness, and varying exogenous rates of technological progress, have a direct impact on firm-level R&D intensities, even within the same sector. Only by assuming that all of these have similar patterns across rival firms, would it be possible to imagine convergence in R&D intensities, but this is not the case in the real world and heterogeneity is a widespread phenomenon.

2.2. Firm-level determinants: public subsidies, gazelles, and financial constraints

Although empirical approaches largely recognize the role of un-observables (or heuristics) in the explanation of firm-level R&D investment, some factors can be identified. For instance, it is a stylized fact that investment in R&D is conducted at a sub-optimal level, thus requiring public intervention to provide support to firms. As reported by the OECD (2016), subsidies are the main tool of public R&D policy for SMEs. Good design and implementation of the public financing scheme are essential for making it an effective tool (Appelt et al., 2016; Soete et al., 2021). Furthermore, there have been extensive debates among scholars to establish whether public support may lead to crowding-out effects (Zúñiga-

Vicente et al. (2014) for a survey on the evidence).[3] Generally, R&D subsides exert a positive effect on R&D investments. However, it is likely that the effects are greater for firms with a high R&D intensity. Another relevant factor is the presence of gazelles, firms characterized by their dynamism influencing countries' innovation (and employment) patterns (Brown et al., 2017). When considering gazelles, various definitions and groupings can be found. Here, we consider possibly the broadest possible definition that includes HGFs (both in terms of employment and sales) and young innovative companies. Despite their very broad definition (Delmar et al., 2003), the main characteristics of HGFs are that they are younger, a quasi-homogenous presence across sectors, a tendency to be more innovative and a greater involvement in international markets (Moreno & Coad, 2015; Teruel et al., 2021). YICs are younger than 6 years, with fewer than 250 employees, and operating at least at 15% of R&D intensity. Almost by definition, these are the companies that should foster aggregate productivity growth thanks to their large innovation focus and disruptive approaches (Schneider & Veugelers, 2010; Czarnitzki & Delanote, 2013). Thanks to their organizational profiles, HGFs and YICs are ideally located to be major influencers of the distribution of R&D intensities in a given country.

Thirdly, a third factor consists of financial constraints (henceforth FC). As argued in Dosi (1990), finance is directly connected to the possibilities and ways of conducting innovative activities. FCs influence all firms to some extent, but especially innovative ones due to both uncertainty and information asymmetries between borrowers and lenders. Consequently, we expect that highly R&D intensive firms may be more harmed than those with lower intensity. Also, García-Quevedo et al. (2018) find that the innovation concept stage is the one most hindered by internal and external financial constraints. This factor tends to become increasingly influential during economic downturns and for SMEs, especially R&D intensive ones (Brown and Petersen, 2015; Lee et al., 2015).

[3] This question, however, is outside the scope of this paper. Our aim is to disentangle the effect that the public financing scheme has on each part of the R&D distribution.

Finally, to clarify the definition of firm-level variables and then also ease their interpretation in the subsequent context of distributional decomposition, in Table 1 we report each firm-level determinant under consideration and its corresponding concept at distribution level.

TABLE 1 HERE

2.3 Reduced business dynamism: R&D intensity as a likely transmission channel

It is well known that firms located at the tails, particularly at the right tail, are largely responsible for aggregate dynamics. In this context, HGFs or gazelles are of utmost importance (Bijnens & Konings, 2020). Recent economic evidence points at a widespread reduced business dynamism in advanced economies. The decline in dynamism has been robustly identified for countries which include the USA (Decker et al., 2014, 2016a, 2016b and 2018; Guzman and Stern, 2020), Australia (Bakhtiari, 2017), Belgium (Bijnens and Konings, 2020), Turkey (Akcigit et al., 2020), Canada (Macdonald, 2014), and Portugal (Sarmento and Nunes, 2010). More recently, a multi-country analysis for 18 countries and 22 industries with data covering the last two decades demonstrated how the phenomenon is both common and globally diffuse (Calvino et al., 2020). Still lacking a stylized explanation, researchers have proposed various theories. Among these, the reduced presence of gazelles is one effective candidate.

When looking at the causes, it is virtually impossible to isolate a specific one. Nevertheless, technological effort, as proxied by R&D intensity, is a key component of market dynamics, since it allows starting and follower firms to challenge the industry leaders, thus fostering competition and faster growth. In this process, tools aimed at providing support such as targeted R&D subsidies can be very valuable (Akcigit et al. 2020). Both empirical and theoretical evidence point at a reduction in knowledge diffusion, especially between frontier and laggard firms, as a coherent explanation for the reduction in business dynamism (Akcigit & Ates, 2021). Therefore, the role of right-tail firms (i.e., gazelles) and of R&D subsidies constitute direct embodiments of this.

3. Data

3.1. Database and Statistics

Our database is PITEC, *Panel de Innovación Tecnológica*, a yearly project conducted by the Spanish Statistical Office and the Spanish Foundation for Science and Technology. Based on the Community Innovation Survey framework, thanks to its representativeness and structure it is one of the most analysed data sources in innovation studies (De Marchi, 2012; Barge-Gil & López, 2014; Segarra & Teruel, 2014; Audretsch et al., 2014; Coad et al., 2020).[4]

We focus on all firms performing R&D activity, regardless their sector. Following Cohen and Klepper (1992), the R&D intensity distribution shows extremely regular patterns across sectors, thus suggesting the existence of a common stochastic process within industries. Also, as highlighted by Tether (2005), and contrary to common belief, the service sector is a locus of innovation. We only remove firms reporting abnormal R&D intensities,[5] and firms from critical sectors such as Oil, Agriculture, and Extractive Industries. Our sample comprises 4,366 observations in the years 2004 and 2008, and 3,289 observations in the years 2009 and 2014—in total 15,310 year-firm pairs.

Table A-2 shows how the main variables of interest fluctuated over the period of observation. The R&D investment roughly followed the Spanish business cycle trend, meaning that it had a high and steady growth rate during the period 2004–2008 and then it contracted in the period 2009–2014. In parallel, the governmental R&D financing scheme followed roughly the same trend. There was, however, little change in the share of gazelles. Finally, financial constraints seem to have slightly increased during the recovery of Spanish economy.

As a measure of R&D, we focus on intensities in terms of sales, rather than pure investment. We correct for likely scale effects by following the extensive literature on the topic (Leonard, 1971; Grabowski and Baxter, 1973; Dosi, 1988). Figure 1 shows the empirical differences along the distributions of interest. Further, given that testing differences on the mean (or other points of the distribution) using traditional

[4] See Table A-1 for a description of the representativeness of the total national investment in internal R&D.

[5] Specifically, firms declaring 1€ of sales while putting large amounts on the R&D item. and firms declaring R&D intensities higher than 100%

approaches (e.g., the t-test or the Kolmogorov-Smirnov test) would generate inevitable biases due to the different sample sizes across the years, and due to the cross-time dependence of incumbent firms' observations, we opt for a different approach. We set up a simple, quantile panel regression model where the dependent variable is *R&D intensity*, while as explanatory variables we include only the categorical variable *year* (see Table 2). Through this, we can explore difference between years across the whole distribution, while considering serial correlations and distributional specificities.

TABLE 2 HERE

FIGURE 1 HERE

First, looking at the kernel densities, it is quite clear that from 2004 there has been a small location shift toward the right, such that the aggregate distributions relative to 2009 and 2014 exhibit a higher mean value. In addition, the unconditional analysis shows that the changes that took place in the expansionary period are quite concentrated around average and slightly above-average R&D performing firms. On the contrary, looking at the contractionary period, the movements also involve the right-tail which is populated by highly technological enterprises. The quantiles with low R&D intensive firms seem to be largely unaffected by time-variations, suggesting a smaller role for them in the distribution.

3.2. Further Motivation(s)

In terms of R&D intensity, Spain was converging to European levels, but started to diverge in 2009. This sudden interruption coinciding with the global economic downturn. Between 2008 and 2016, the Spanish economy experienced a loss of 30% in its public R&D budget and a reduction of 43% in the number of enterprises performing R&D activities. Also, private R&D expenditure experienced the effects of the crisis, making Spain the only one of the four major EU economies whose R&D investment decreased continuously between 2009 and 2014 (Xifré, 2018). Trivially, this had a dramatic impact on the path to convergence.

FIGURE 2 HERE

Figure 2 shows how the Spanish R&D expenditure followed two clear trends: considerable growth during the period 2004–2008, followed by a non-trivial decline dictated by the consequences of the economic crisis in the period 2009–2014. Public financing, which has been sustaining private expenditures, was a major contributor. Up to 2007, more than 90% of the allocated budget was invested, but since then the trend has been decreasing, reaching the historical minimum of only 46.6% in 2017 (COTEC, 2018).

4. Econometric approach

4.1. Methodology and Types of Counterfactuals

Methods such as the Kitagawa-Blinder-Oaxaca decomposition have a long tradition in economics.[6] The seminal works by Oaxaca (1973) and Blinder (1973) paved the way to the development of methods aimed at going beyond the simple distributional mean by decomposing a distribution according to appropriately chosen components.[7] Recently, Chernozhukov et al. (2013) overcame the apparent inapplicability of quantile regressions by developing distribution regressions (Fortin et al., 2011). Although distributional and quantile regression return equivalent representations of the data, the two approaches have differences. The most relevant of these are greater computational efficiency, less bias in presence of mass points, and no requirement of smoothness of the conditional density function, as the approximation is conducted pointwise. Furthermore, distributional regressions return cleaner estimates in comparison to those returned by quantile regression.

As reported in Chernozhukov et al. (2013), it is possible to identify three cases of counterfactual effects (henceforth CE). First, when the covariates do not vary, but the conditional distribution does (Type 1, or coefficient effect). Second, when the covariate distribution varies but the conditional distribution does not (Type 2, or characteristic effect). And finally, when both the conditional and the covariate distribution vary (Type 3). In this work, we first focus on Type 2, looking at how different values of covariates would

[6] Typical applications of the original method in the industrial literature regard the decomposition of productivity (i.e. Fariñas and Ruano, 2004).

[7] Evelyn M. Kitagawa (1955) was the first to propose this type of decomposition.

have affected the distribution, and then, we appraise the reliability of the empirical model by introducing Type 3 effects.[8]

4.2. Identification Strategy and our Decomposition

Following Chernozhukov et al. (2013) and Fortin et al. (2011), we develop an identification strategy to decompose the changes in the R&D distribution for Spain. We conduct the analysis for two opposed phases of the economic cycle.

TABLE 3 HERE

In line with the literature, the factors suspected to influence the evolution of R&D dynamics are governmental financing, the role of gazelles, financially constrained firms and, finally, certain firms' characteristics. We include firm size in terms of employees and their level of sales. Additionally, we control for their sales growth rates normalized by sector. Doing so, we control for demand sectoral variations, but we also remove biasing factors such as inflation, while controlling for individual firms' growth (Coad & Grassano, 2019; Bianchini et al., 2017). This argument holds even more strongly for young and small companies (García-Quevedo et al., 2014). Finally, this variable also controls for firms' relative performances, as well as for differences in opportunities between sectors. Table 3 presents the definitions of the three explanatory factors, and of the control vector.Starting from the estimation of Type 2 CE, we develop the decomposition relative to the years 2004–2008, the expansionary period.[9] This implies the estimation of how the 2008 internal R&D distribution would look, requiring that the covariates of interest assume the 2004 values.

Suppose that $F_{RD\_int,\,(a,b,c,d)}$ corresponds to the counterfactual distribution of log R&D intensities, when: a) the public financing scheme is as in year *a*, b) the gazelles are those in year *b*, c) financial constraints apply as in year *c*, and, d) firms' characteristics are as in year *d*. Thanks to the law of iterated probabilities, it is

[8] Despite its advantages, the approach has one main limitation: counterfactuals may not be enough for causal insights. Thus, the results are interpreted in an associative manner.
[9] It is possible to obtain the same decomposition for the contractionary period by substituting the appropriate time indexes.

possible to decompose the observed change in the distribution of R&D intensity between two years (2004, year 0, and 2008, year 1) into the sum of the above four effects, as follows:[10]

(1) $F_{RD_{int\_1}|(1,1,1,1)} - F_{RD_{int\_1}|(0,0,0,0)}$

$$= \left[F_{RD_{int\_1}|(1,1,1,1)} - F_{RD_{int\_1}|(0,1,1,1)}\right] + \left[F_{RD_{int\_1}|(0,1,1,1)} - F_{RD_{int\_1}|(0,0,1,1)}\right]$$

$$+ \left[F_{RD_{int\_1}|(0,0,1,1)} - F_{RD_{int\_1}|(0,0,0,1)}\right] + \left[F_{RD_{int\_1}|(0,0,0,1)} - F_{RD_{int\_1}|(0,0,0,0)}\right]$$

In more detail, this requires the identification and estimation of the following counterfactuals:

- $F_{RD_{int\_1}|\,(0,1,1,1)}(y) = \int F_{RD_{int\_1}|\,(1,1,1,1)}(y) \cdot dF_{RD_{int}|\,(0,1,1,1)}(x)$, corresponding to the distribution of R&D intensities that would prevail in 2008 if firms were subject to the public financing scheme of 2004;[11]
- $F_{RD_{int\_1}|\,(0,0,1,1)}(y) = \int F_{RD_{int\_1}|\,(1,1,1,1)}(y) \cdot dF_{RD_{int}|\,(0,0,1,1)}(x)$, corresponding to the distribution of R&D intensities that would prevail for firms in 2008 if firms were subject to the public financing scheme of 2004 and the gazelles were those of 2004.
- $F_{RD_{int\_1}|\,(0,0,0,1)}(y) = \int F_{RD_{int\_1}|\,(1,1,1,1)}(y) \cdot dF_{RD_{int}|\,(0,0,0,1)}(x)$, corresponding to the distribution of R&D intensities that would prevail in 2008 if gazelles were the ones of 2004 and firms were subject to the public financing scheme and financial constraints of 2004.

Following these estimations, we aim to appraise how much of the observed differences between the two conditional distributions is explained by actual variations of the covariates, and how much is left to unobservability. This falls under the Type 3 CE:

(2) $F_{RD_{int\_1}|(1,1,1,1)} - F_{RD_{int\_0}|(0,0,0,0)}$

$$= \left[F_{RD_{int\_1}|(1,1,1,1)} - F_{RD_{int\_1}|(0,0,0,0)}\right] + \left[F_{RD_{int\_1}|(0,0,0,0)} - F_{RD_{int\_0}|(0,0,0,0)}\right]$$

[10] The results of these decompositions can be order independent. We run the estimations for each possible order of the decomposition. Table A-4 shows the results remain unchanged.
[11] Firm-level public financing data is available only for 2003 and 2005. For the baseline version, we use 2003 public financing data but, in the robustness checks, we explore the sensitivity of the findings to alternative choices. All relevant results show correlations higher than 90%, thus we can safely conclude that this does not affect our findings.

From Equation (2), Type 3 CE consists of the sum of Type 1 and Type 2 effects, where the latter is the subject of the first part of the analysis. Indeed, it corresponds to the total differences emerging through the proposed conditional model and consists of i) a "characteristics effect" due to changes in our selected determinants, and ii) a "coefficient effect," which is imputable to changes in the estimated parameters. While the characteristic effect can be interpreted quite straightforwardly, the same does not hold for the coefficient effect, which may hide simple intercept shifts or an altered coefficient in any of our covariates.

5. Results

The empirical results first address how the distribution would have changed if its covariates had been those of the starting year (type 2 CE). This is replicated for both the pre- and post-crisis, respectively in Subsections 5.1 and 5.2. Then, in Subsection 5.3 we quantify how much the observed change was actually explained by variation in the chosen covariates (type 3 CE). Finally, Subsection 5.4 summarizes our results under a comparative perspective.

5.1. Pre-crisis (2004–2008): the Expansionary Phase

Starting our decomposition analysis, Table 4 and Figure A-1 show the results. The first terms correspond to the effects that governmental financing had on shaping the R&D distribution. An equivalent question would be: how would the distribution of R&D in 2008 look if public financing were of the same level and given to the same firms as they were in 2003? As expected, the answer is lower. There are three clear patterns. First, there is a positive association between the new financing scheme and firms' technological intensity. Second, the progressiveness with which these subsidies influenced the distribution is also clear. Finally, most of these effects are concentrated among top R&D firms.

Similarly, our results regarding the influence of gazelles hint at a progressive lessening of these "hyper-firms" contribution to R&D. It appears that past gazelles were able to contribute more to the aggregate technological effort of the country. Curiously, these effects are concentrated not only among the more

R&D intensive, but also in the less intensive, quantiles. The magnitude of the effects differs significantly, the 90/50 ratio being roughly 7-fold, but it is interesting to see how gazelles contribute quite extensively to R&D intensity.

Concerning the financial constraints, the (negative) impact of these constraints on the R&D distribution decreased during the expansionary period, it being highly likely that this was also due to the impact of R&D subsidies. This is interesting since Table A-2 shows how the number of financially constrained firms increased during the expansionary period. Specifically, the top R&D intensive firms were the most affected by financial constraints. This is in line with the results by Czarnitzki and Hottenrott (2011a), who found cutting-edge research (as opposed to routine research) more impacted by these constraints.

TABLE 4

Finally, firms' characteristics are mainly for controlling for possible standard cofounding factors. Particularly, firms' main characteristics (such as size, sales, and sector-normalized growth) played a significant role in shaping the aggregate distribution. Interestingly, the effects show a strong asymmetry. The upper part of the distribution would show lower performance if firms' characteristics were those of 2004, while the lower quantiles would be greater. All of this suggests that Spanish firms became more R&D-intensive during the expansionary phase, increasing their presence in the top performing quantiles.

5.2. After-crisis (2009–2014): the Contractionary Phase

This section presents the changes in the effect of our key factors during a contraction phase (Table 5). The global economic downturn hit Spain with great strength and led to a considerable contraction in both R&D investment and subsidies (Cruz-Castro & Sanz-Menéndez, 2016). Hence, it is particularly interesting to assess what might have happened if the Spanish government had been able to maintain the pre-crisis public financing scheme intact. It appears that it would have generated a higher performance among firms' R&D intensity. This is also in line with the general countercyclical effects found in the literature (Aristei et al., 2017). Again, the progressive effect of the scheme that emerged in the previous estimation is confirmed, the top performing quantiles are those that would have benefited the most.

Section 5.1. showed that the gazelles' contribution slowed down considerably in 2004 as compared to 2008. The same is true when comparing 2009 gazelles with those of 2014. This evidences a general decrease in the capacity of these firms to contribute to aggregate R&D which, is now likely to be sustained by other types of firms. Interestingly, focusing on the 90/50 ratio as a proxy for (positive) outliers' dispersion, this is about twice as big as the same ratio measured during the expansionary phase (from 7 to 14). This hits at a major concentration of the effects on the right part of the distribution. Besides the observed decline in gazelles' relevance for R&D dynamics, it is also clear that gazelles show low resiliency during the crisis, struggling hard to maintain their usual contributions to aggregate R&D distribution.

TABLE 5 HERE

Financial constraints had a distinct trend during the expansionary phase, but the corresponding results for the contractionary period differ considerably. Unexpectedly, financial constraints seemed to have almost no role in affecting R&D intensities of Spanish firms in the period 2009–2014. Indeed, some negative effects are seen, but the absence of statistical significance under several different statistical tests suggests large levels of dispersion in how financial constraints impacted firm-level contributions (and decisions). These results are in line with Mulkay et al. (2001), Cincera (2003) and Audretsch and Weigand (2005) who find that R&D investments are less subject to financial constraints than physical capital investments are. This lack of impact is characteristic of R&D projects whose long-term nature and high adjustment costs cause insensibility between financial constraints and the decision to reduce the investment. Particularly, Cincera (2003) and García-Quevedo et al. (2018) show that FCs affect the decision to invest in R&D rather than on ongoing R&D projects. However, sectors highly dependent of external financial sources, or with a low degree of tangible assets, may be more affected (Aghion et al., 2012).

Finally, during the contractionary phase, the asymmetric effect that emerged through the expansion vanishes. Instead, we find a negative impact, but more strongly concentrated around two quantiles (0.8 and 0.9). This is interpreted as a deterioration of firms' characteristics, becoming less technologically prone and showing a greater propensity to invest intensively in research and development, over the period from 2009 to 2014.

5.3. Aggregate Model Decomposition (Type 3 CE) and the Role of Unobservables

This subsection explores the decomposition from a more aggregate approach and quantifies how much of the observed changes are explained by our conditional model and how much of it is due to unobserved variation.

Recalling that in 2008–2009 the distribution was at its peak; the left side of Figure 3 presents the Type 3 CE that emerges changing both the covariates and the conditional distribution of 2008 from the ones of 2004. Trivially, the effect would be negative, and the distribution would be lower. On the right side of the figure, the same difference emerges regarding 2014 by considering the covariates and conditional distribution of 2009. The impact of the decreased contribution of gazelles to the right-tail of the distribution is already clear.

FIGURE 3 HERE

Nevertheless, it is possible to go into considerable detail, decomposing these overall differences in Type 1 and Type 2 CE, recalling that Type 2 corresponds to what done in the previous two subsections (coefficient effect), while Type 1 (characteristics effect) reports on the effect attributable to unobserved characteristics. Figure 4 and Table 6 present these estimations.

FIGURE 4 HERE

Furthermore, we observe differences in the relative magnitudes of these effects between the two phases under analysis. The effect of our selected determinants in the contractionary phase dominates the same effect estimated for the expansionary phase. This is also clear from Table 4, where the relative magnitudes are roughly two-fold.

TABLE 6 HERE

Generally, the fact that the coefficient effect is so strong in both periods suggests the influence that unobservable firm-level characteristics have on R&D intensities. This portion corresponds to what Cohen and Klepper (1992) defined as R&D-related expertise. Further, these characteristics weigh more in the expansionary period, where an increase in intensity was expected, rather than in the contractionary one. Finally, in both specifications, this effect decreases across quantiles, making the top R&D intensive quantiles less subject to these economic fluctuations. Economically, this can be explained by the fact that these quantiles are mostly composed of science-based firms, or R&D specialists (Cattaruzzo, 2020), whose main (and sometimes only) business focus is introducing novelties to the market.

5.4. Comparison and interpretation

Table 7 summarizes the empirical findings regarding the effects of the selected determinants (Type 2 CE). In addition to the estimated coefficients, we also report the associated percentages of the total estimated change.

The most interesting asymmetries relate to the role that public financing and gazelles had in shaping the distribution of Spanish internal R&D. Indeed, in the expansionary period, public R&D financing from the years 2008 and 2009 was strongly positive and progressive in incentivizing firms' technological intensity. Additional evidence for this is that in 2007 the usage of public R&D financing reached its peak, and then started a monotonic decline that led to only half of the allocated national budget for R&D financing being actually taken up by enterprises. This is in line with previous findings and likely due to changes in the application and format of the public financing, which led to much greater inefficiency (Cruz-Castro & Sanz-Menéndez, 2016).

Similarly, and especially in the expansionary period, gazelles played a leading role but their contribution to R&D declined from 2004 to 2014. Here, we refer to Decker et al. (2016a) who noticed an impoverishment in the tails of the growth rate distribution, thus hinting at a progressive decline in the contribution to aggregate productivity from YICs and HGFs. In partial confirmation of this, our results show that gazelles' contribution to Spanish firms' R&D intensity declined strongly. This can potentially explain their declining contribution to productivity growth through the usual transmission channels

(R&D → innovation → productivity). The results of the aggregate decomposition between characteristics and coefficient effects further reinforce this interpretation.

Finally, if financial constraints were indeed impacting firms at the beginning of the expansionary period, what emerged is that the same did not happen during the strong contraction to the Spanish industrial system. Recalling that these estimations derive from a counterfactual framework, a likely explanation is that the public financing scheme employed in the expansionary period alleviated the financial needs of most firms. Thus, the constraints in place in 2009 are not relevant in explaining variations of R&D intensities.

TABLE 7 HERE

6. Concluding remarks and implications

Cohen and Klepper (1992) offered a seminal way to look at the R&D intensity distribution, where much of its anatomy is explained by non-observables, managerial skills, a phenomenon that is common across industries. In a complementary fashion, our work identifies and quantifies specific firm-level contributions, also looking at how these dynamically change over time and according to distinct phases of the business cycle. Given the wide coverage and timing of PITEC, the Spanish R&D aggregate distribution before and after the 2007 global economic downturn constitutes an insightful research setting. There, the detailed decomposition shows robust emerging patterns in terms of the chosen firm-level determinants and modelling approach.

We apply the technique developed by Chernozhukov et al. (2013), which offers comprehensive explanations of the individual factors' contribution to each part of the R&D distribution. This allows us to consider its skewed nature and largely heterogeneous composition. We analyse how four determinants (public financing, gazelles, financial constraints, and firms' characteristics) contribute to shaping the overall distribution of R&D. In line with Cohen and Klepper (1992), our results show a dominance of unobservables for all quantiles other than the top R&D intensive firms. This implies that, for non-R&D specialists, the determinants under study are not the main drivers of their investment decision, rather they are driven by other economic fluctuations. Looking at the individual determinants, we identify public financing and the gazelles' contribution as the main determinants in shaping the distribution of interest. In particular, the contribution from gazelles decreases considerably over the period from 2004 to 2014. There are several potential explanations for this decreased contribution but can be mostly identified in structural factors currently affecting economies, such as the rising importance of intangible and digital assets, which in turn can lead to increases in market shares and power of the best performing firms, posing additional barriers to growth. Additionally, the effect of globalization and the predominance of global value chains cannot be neglected since they deeply affect firm dynamics and entry mechanisms. This not only extends the existing empirical findings on reduced business dynamism across developed economies (Decker et al., 2016a, 2016b; Decker et al. 2017; Akcigit and Ates, 2021), but also constitutes a possible transmission channel. Our evidence is in line with the steady declines in business dynamism showed in Calvino et al. (2020) over the last two decades, even after accounting for the role of the business cycle.

6.1 Policy implications

Our results point clearly to two to different policy actions. First, gazelles are considered as important players in innovative and dynamic markets but due to their quasi-random growth process, supporting them has always been difficult. Policies aimed at promoting, not only their presence, but also their contribution in terms of innovative performance, should be pursued. Second, Spain is the EU country that has cut public R&D spending the most after the 2007 crisis. Since 2017, public investment in R&D has been recovering, but its level has not reached the levels before the 2008 crisis. Efforts should be made by institutions and governmental agencies to discover why individual firms do not apply for their allocated R&D budget.

Both actions should aim for the convergence of R&D intensities across EU countries. According to the 2021 European Innovation Scoreboard, the innovative capacity of the EU-27 economies grew at an annual rate of 8.9% between 2012 and 2019. Despite this, the erratic evolution of innovation policies in countries such as Spain has limited the ability of the EU to reverse the decline in R&D spending and innovation in the Euro area. To foster the EU innovative capacity and making it fitter to face sustainability and digitalization challenges, two elements are necessary: 1) better coordination of policies and 2) better tailoring of these policies. Recent evidence shows clearly that R&D firms react to macroeconomic expansions and contractions somehow heterogeneously and in this regard, there exist different innovative search profiles that can be systematized (Arvanitis and Woerter, 2014). The evidence presented here is an additional and complementary piece of information for a bottom-up tailoring of public support policy design.

In conclusion, R&D spending is concentrated in a limited number of companies, industries, and countries. Our analysis shows that gazelle-like companies are undervalued, and their centrality suggests a re-evaluation (Kuhlmann and Rip, 2018). This study offers evidence for the size of their impact on the whole innovative capacity. Although these young and dynamic companies are not the ones that invest the most in R&D in absolute terms, they are often the ones that generate the most disruptive innovations, thereby affecting the rest of the productive network. These actors require better access to financial resources and their needs should be considered by policymakers. Their actions not only generate large knowledge and technology spillovers, but also exhibit greater transformation potential than governments and established companies.

Acknowledgement: we thank participants of the 26th SPRU PhD forum (Brighton, 14–15 May 2020), of the Workshop on Industrial and Public Economics (Reus, 4–5 February 2021), of the XXIII Encuentro de Economía Aplicada (online, 3-4 June 2012) and of EU-SPRI conference (online, 9-11 June 2021) for useful feedback.

Declaration of competing interests: nothing to declare.

References

Acemoglu, D., V.M. Carvalho, A. Ozdaglar and A. Tahbaz-Salehi (2012), "The network origins of aggregate fluctuations", *Econometrica*, 80(5), 1977-2016.

Aghion, P., N. Bloom, R. Blundell, R. Griffith and P. Howitt (2005), "Competition and innovation: An inverted-U relationship", *The quarterly journal of economics*, *120*(2), 701-728.

Aghion, P., P. Askenazy, N. Berman, G. Cette and L. Eymard (2012), "Credit constraints and the cyclicality of R&D investment: Evidence from France", *Journal of the European Economic Association*, 10(5), 1001-1024.

Akcigit, U., Y. E. Akgunduz, S. M. Cilasun, E. Ozcan-Tok and F. Yilmaz (2020), "Facts on business dynamism in Turkey", *European Economic Review*, *128*, 103490.

Akcigit, U. and S. T. Ates (2021), "Ten Facts on Declining Business Dynamism and Lessons from Endogenous Growth Theory", *American Economic Journal: Macroeconomics*, 13(1), 257-98.

Appelt, S., M. Bajgar, C. Criscuolo and F. Galindo-Rueda (2016), "R&D Tax Incentives: Evidence on design, incidence and impacts", *OECD Science, Technology and Industry Policy Papers*.

Aristei, D., Sterlacchini, A., & Venturini, F. (2017). Effectiveness of R&D subsidies during the crisis: firm-level evidence across EU countries. *Economics of Innovation and new Technology*, *26*(6), 554-573.

Arvanitis, S. and M. Woerter (2014), "Firm characteristics and the cyclicality of R&D investments", *Industrial and Corporate Change*, *23*(5), 1141-1169.

Audretsch, D.B., and J. Weigand (2005), "Do knowledge conditions make a difference?: Investment, finance and ownership in German industries", *Research Policy*, 34(5), 595-613.

Audretsch, D.B., A. Segarra and M. Teruel (2014), "Why don't all young firms invest in R&D?", *Small Business Economics*, 43(4), 751-766.

Bakhtiari, S. (2017), “Entrepreneurship Dynamics in Australia: Lessons from Micro-Data”, *Research Paper No. 5/2017*, Commonwealth of Australia.

Barge-Gil, A. and A. López (2014), “R&D determinants: Accounting for the differences between research and development”, *Research Policy*, 43(9), 1634-1648.

Barlevy, G. (2007), “On the cyclicality of research and development”, *American Economic Review*, 97(4), 1131-1164.

Bianchini, S., G. Bottazzi and F. Tamagni (2017), “What does (not) characterize persistent corporate high-growth?”, *Small Business Economics*, *48*(3), 633-656.

Bijnens, G. and J. Konings (2020), “Declining business dynamism in Belgium. Small Business Economics”, 54(4), 1201-1239.

Blinder, A. S. (1973), “Wage discrimination: reduced form and structural estimates”, *Journal of Human Resources*, 436-455.

Breschi, S. and F. Malerba (1997), “Sectoral innovation systems: technological regimes, Schumpeterian dynamics, and spatial boundaries”, *Systems of innovation: Technologies, institutions and organizations, 1*, 130-156.

Brown, J.R. and B.C. Petersen (2015), “Which investments do firms protect? Liquidity management and real adjustments when access to finance falls sharply”, *Journal of Financial Intermediation*, 24(4), 441-465.

Brown, R., S. Mawson and C. Mason (2017), “Myth-busting and entrepreneurship policy: the case of high growth firms”, *Entrepreneurship & Regional Development*, 29(5-6), 414-443.

Calvino, F., C. Criscuolo and R. Verlhac (2020), “Declining business dynamism: Structural and policy determinants”, *OECD Science, Technology and Industry Policy Papers*.

Carvalho, V. and X. Gabaix (2013), “The great diversification and its undoing”, *American Economic Review*, 103(5), 1697-1727.

Cattaruzzo, S. (2020), “*On R&D sectoral intensities and convergence clubs”,* JRC Working Papers on Corporate R&D and Innovation No 01/2020, Joint Research Centre

Chernozhukov, I. Fernández-Val and B. Melly (2013), “Inference on counterfactual distributions”, *Econometrica*, *81*(6), 2205-2268.

Cincera, M. (2003), Financing constraints, fixed capital and R&D investment decisions of Belgian firms, in: P. Butzen and C. Fuss (Eds) *Firms' Investment and Finance Decisions: Theory and Empirical Methodology*, pp. 129–147 (Cheltenham: Edward Elgar).

Coad, A. (2019), “Persistent heterogeneity of R&D intensities within sectors: Evidence and policy implications”, *Research Policy*, 48(1), 37-50.

Coad, A. and N. Grassano (2019), “Firm growth and R&D investment: SVAR evidence from the world’s top R&D investors”, *Industry and Innovation*, 26(5), 508-533.

Coad, A., A. Segarra and M. Teruel (2020), “A bit of basic, a bit of applied? R&D strategies and firm performance”, The *Journal of Technology Transfer*, doi.org/10.1007/s10961-020-09826-1.

Cohen, W. M. and R. C. Levin (1989), “Empirical studies of innovation and market structure”, In R. Schmalensee and R. Willig (Eds.) Handbook of Industrial Organization – Vol.2, 1059–1107. Amsterdam, Netherlands: North Holland.

Cohen, W. M. and S. Klepper (1992), “The anatomy of industry R&D intensity distributions”, *The American Economic Review*, 773-799.

COTEC (2018) “Tecnología e innovación en España: informe COTEC 2018”, Madrid: Fundación COTEC para la innovación.

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

Cruz-Castro, L. and L. Sanz-Menéndez (2016), “The effects of the economic crisis on public research: Spanish budgetary policies and research organizations", *Technological Forecasting and Social Change*, 113, 157-167.

Czarnitzki, D. and J. Delanote (2013), “Young innovative companies: the new high-growth firms?”, *Industrial and Corporate Change*, 22(5), 1315-1340.

Czarnitzki, D. and H. Hottenrott (2011a), “Financial constraints: Routine versus cutting edge R&D investment”, *Journal of Economics & Management Strategy*, 20(1), 121-157.

De Marchi, V. (2012), “Environmental innovation and R&D cooperation: Empirical evidence from Spanish manufacturing firms”, *Research Policy*, 41(3), 614-623.

Decker, R., J. Haltiwanger, R. Jarmin and J. Miranda (2014), “The role of entrepreneurship in US job creation and economic dynamism”, *Journal of Economic Perspectives*, *28*(3), 3-24.

Decker, R. A., J. Haltiwanger, R. S. Jarmin and J. Miranda (2016a), “Where has all the skewness gone? The decline in high-growth (young) firms in the US”, *European Economic Review*, 86, 4-23.

Decker, R. A., J. Haltiwanger, R. S. Jarmin and J. Miranda (2016b), “Declining business dynamism: What we know and the way forward”, *American Economic Review*, 106(5), 203-07.

Decker, R. A., J. Haltiwanger, R. S. Jarmin and J. Miranda (2017), "Declining dynamism, allocative efficiency, and the productivity slowdown", *American Economic Review*, 107(5), 322-26.

Delmar, F., P. Davidsson and W.B. Gartner (2003), "Arriving at the high-growth firm", *Journal of Business Venturing*, *18*(2), 189-216.

Di Giovanni, J. and A.A. Levchenko (2010), "Putting the parts together: trade, vertical linkages, and business cycle comovement", *American Economic Journal: Macroeconomics*, 2(2), 95-124.

Di Giovanni, J., A.A. Levchenko and I. Mejean (2014), "Firms, destinations, and aggregate fluctuations", *Econometrica*, 82(4), 1303-1340.

Dosi, G. (1988), "Sources, procedures, and microeconomic effects of innovation", *Journal of Economic Literature*, 1120-1171.

Dosi, G. (1990), "Finance, innovation and industrial change" *Journal of Economic Behavior & Organization*, 13(3), 299–319.

Evangelista, R. (2006), "Innovation in the European service industries", *Science and Public Policy*, 33(9), 653-668.

Falk, M. (2007), "R&D spending in the high-tech sector and economic growth", *Research in economics*, *61*(3), 140-147.

Fariñas, J. C. and S. Ruano (2004), "The dynamics of productivity: A decomposition approach using distribution functions", *Small Business Economics*, *22*(3), 237-251.

Foerster, A.T., P.-D. G. Sarte and M.W. Watson (2011), "Sectoral vs. Aggregate Shocks: A Structural Factor Analysis of Industrial Production", *Journal of Political Economy*, 119 (1), 1-38.

Fortin, N., T. Lemieux and S. Firpo (2011), "Decomposition methods in economics", In *Handbook of labor economics* (Vol. 4, 1-102). Elsevier.

Gabaix, X. (2011), "The granular origins of aggregate fluctuations", *Econometrica*, *79*(3), 733-772.

García-Quevedo, J., G. Pellegrino and M. Vivarelli (2014), "R&D drivers and age: Are young firms different?", *Research Policy*, 43(9), 1544-1556

García-Quevedo, J., A. Segarra and M. Teruel (2018), "Financial constraints and the failure of innovation projects", *Technological Forecasting & Social Change*, 127, 127-140.

Grabowski, H.G. and N.D. Baxter (1973), "Rivalry in industrial research and development: An empirical study", *Journal of Industrial Economics*, 209-235.

Griffiths, W. and E. Webster (2010), "What governs firm-level R&D: Internal or external factors?", *Technovation*, 30(7-8), 471-481.

Guzman, J. and S. Stern (2020), "The State of American Entrepreneurship: New Estimates of the Quantity and Quality of Entrepreneurship for 32 US States, 1988–2014", *American Economic Journal: Economic Policy*, *12*(4), 212-43.

Hall, B., and F. Hayashi (1989), "*Research and Development as an Investment* (No. w2973)", National Bureau of Economic Research.

Hall, B. (1993), "The stock market's valuation of R&D investment during the 1980's", *American Economic Review*, 83(2), 259-264.

Jefferson, G. H., Huamao, B., Xiaojing, G., & Xiaoyun, Y. (2006). R&D performance in Chinese industry. *Economics of innovation and new technology*, 15(4-5), 345-366.

Kitagawa, E. M. (1955), "Components of a difference between two rates", *Journal of the American statistical association*, 50(272), 1168-1194.

Kuhlmann, S. and A. Rip (2018), "Next-generation innovation policy and grand challenges", *Science and public policy*, *45*(4), 448-454.

Lee, N., H. Sameen and M. Cowling (2015), "Access to finance for innovative SMEs since the financial crisis", *Research Policy*, 44(2), 370-380.

Leonard, W.N. (1971), "Research and development in industrial growth", *Journal of Political Economy*, *79*(2), 232-256.

Macdonald, R. (2014), "Business Entry and Exit Rates in Canada: A 30-year Perspective", *Economic Insights* No. 38, Statistics Canada.

Machado, J. A. and J. S. Silva (2019), "Quantiles via moments", *Journal of Econometrics*, *213*(1), 145-173.

Montresor, S. and A. Vezzani (2015), "The production function of top R&D investors: Accounting for size and sector heterogeneity with quantile estimations", *Research Policy*, 44(2), 381-393.

Moreno, F. and A. Coad (2015), "High-growth firms: Stylized facts and conflicting results", *Entrepreneurial growth: individual, firm, and region*, *17*, 187-230.

Mulkay, B., B. H. Hall and J. Mairesse (2001), "Investment and R&D in France and the United States, in: H", Heinz & R. Strauch (Eds) *Investing Today for the World of Tomorrow*, pp. 229–273 (Berlin: Springer Verlag).

Oaxaca, R. (1973), "Male-female wage differentials in urban labor markets", *International economic review*, 693-709.

OECD (2016), "OECD Science, Technology and Innovation outlook", 2016.

Schneider, C. and R. Veugelers (2010), "On young highly innovative companies: why they matter and how (not) to policy support them", *Industrial and Corporate Change*, *19*(4), 969-1007.

Sarmento, E. and A. Nunes (2010), "Entrepreneurship performance indicators for employer enterprises in Portugal", *Temas Económicos*.

Segarra, A. and M. Teruel (2014), "High-Growth Firms and innovation: an empirical analysis for Spanish firms", *Small Business Economics*, 43(4), 805-821.

Soete, L., B. Verspagen and T. H. Ziesemer (2021), "Economic impact of public R&D: an international perspective", *Industrial and Corporate Change*.

Tether, B. S. (2005), "Do services innovate (differently)? Insights from the European innobarometer survey", *Industry & Innovation*, *12*(2), 153-184.

Teruel, M., A. Coad, C. Domnick, F. Flachenecker, P. Harasztosi, M.L. Janiri and R. Pal (2021), "The birth of new HGEs: internationalization through new digital technologies", *Journal of Technology Transfer*.

Veugelers, R. and M. Cincera (2015). The impact of Horizon 2020 on innovation in Europe. *Intereconomics - Forum*, 50, 2-9.

Xifré, R. (2018), "Spanish investment in R&D+I in the wake of the crisis: Public versus private sector", *Spanish Economic and Financial Outlook (SEFO)*, 7(4), 67-79.

Zúñiga-Vicente, J.Á., C. Alonso-Borrego, F. J. Forcadell and J. I. Galán (2014), "Assessing the effect of public subsidies on firm R&D investment: a survey", *Journal of Economic Surveys*, 28(1), 36-67.

Appendix

TABLE A-1 HERE

TABLE A-2 HERE

TABLE A-3 HERE

FIGURE A-1 HERE

FIGURE A-2 HERE

Appendix - Robustness checks

We explore the sensitivity of our results to the specification-related choices. First, we lag the public financing scheme of the contractionary period, as we had to do for the expansionary one.[12] Second, we impute the firm-level values of R&D public financing in 2004 using the mean of the values in 2003 and 2005. Further, we estimate the expansionary period using exclusively 2005–2008 data. Finally, as the applied methodology and relative estimates might be order-dependent, following Chernozhukov et al. (2013), we re-estimate the baseline specification in reverse order.

Using the baseline model as reference, we observe similar estimates. This holds for the most significant effects from our analysis. Obviously, the non-statistically significant effects and the controls (i.e., financial constraints in the contractionary period and firms' characteristics) show low correlations, as the large error estimates make them fluctuate considerably. Oppositely, we stress the high degree of correlation among the identified, statistically significant effects.

TABLE A-4 HERE

[12] PITEC lacks data for public financing in 2004.

Table 1 - Correspondences between firm-level and distributional variables

Firm-level variables	Distribution-wise correspondent
Binary gazelle status	Share and location of gazelle firms
Binary financial constraint status	Share and location of financially constrained firms
Public subsidy amount	Quantity and recipients of subsidies

Note: we use the term "location" to refer to the firms' location across the R&D intensity distribution.

Table 2 - Statistical differences across the distributions

	Location	Scale	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
2008	**Y**	**Y**	N	N	N	N	**Y**	**Y**	**Y**	N	N
2014	**Y**	**Y**	N	N	N	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**

Note: the refence category is the R&D intensity distribution in 2004, estimations are conducted using xtqreg package, which follows the method of Machado & Santos-Silva (2019). Differences are reported as such if statistically significant with p-value less than 5%.

Table 3 – Variable definitions	
Factor name	**Definition**
Internal R&D intensity	Ratio of internal R&D expenditure to sales level (in logs)
R&D public financing	Internal R&D expenditure financed with public funding (in €)
Gazelles - HGF sales - HGF employment - YIC	Either one of the three conditions below holds: - Quantile-based definition including the top decile of the unconditional sales growth distribution. - Quantile-based definition including the top decile of the unconditional employment growth distribution. - Firms with fewer than 250 employees, less than six years old and at least 15% of R&D intensity.
Financial constraints	Firms declaring lack of either internal or external funding as highly relevant.
Firm-level characteristics - sales level - size - sector-normalized sales growth	 Volume of turnover (in €). Number of employees. Year sales growth minus 2-digit sector average growth.[13]

[13] Sectors follow the CNAE-2009 classification, which can be directly linked to the more detailed NACE classification.

Table 4 – Results relative to the impact that each factor of choice in 2004 would have had on the 2008.

Quantile	R&D Public financing	Gazelle firms	Financial constraints	Firms' characteristics
0.1	-0.023***	-0.009	-0.019***	0.072***
	(0.007)	(0.008)	(0.005)	(0.01)
0.2	-0.027***	-0.004	-0.031***	0.056***
	(0.007)	(0.007)	(0.005)	(0.007)
0.3	-0.031***	0.007	-0.036***	0.042***
	(0.007)	(0.008)	(0.006)	(0.007)
0.4	-0.035***	0.019**	-0.038***	0.031***
	(0.007)	(0.008)	(0.006)	(0.005)
0.5	-0.040***	0.029***	-0.040***	0.024***
	(0.007)	(0.009)	(0.006)	(0.005)
0.6	-0.043***	0.041***	-0.040***	0.017***
	(0.007)	(0.01)	(0.006)	(0.006)
0.7	-0.051***	0.067***	-0.039***	0.007
	(0.008)	(0.014)	(0.007)	(0.008)
0.8	-0.066***	0.106***	-0.041***	-0.01
	(0.01)	(0.017)	(0.008)	(0.011)
0.9	-0.095***	0.190***	-0.038***	-0.044**
	(0.015)	(0.024)	(0.01)	(0.017)

*Number of observations: 4,366. Significance levels corresponding to * p<0.1, ** p<0.05, and *** p<0.01. Standard errors are computed on 100 bootstrap repetitions.*

Table 5 – Results relative to the impact that each factor of choice in 2009 would have had on the 2014 low.				
Quantile	**R&D Public financing**	**Gazelle firms**	**Financial constraints**	**Firms' characteristics**
0.1	0.026***	-0.002	-0.001	-0.009
	(0.008)	(0.008)	(0.006)	(0.012)
0.2	0.023***	0.000	0.002	-0.006
	(0.008)	(0.007)	(0.005)	(0.009)
0.3	0.024***	-0.001	-0.001	-0.002
	(0.008)	(0.007)	(0.004)	(0.007)
0.4	0.032***	0.004	0.000	-0.003
	(0.009)	(0.008)	(0.005)	(0.006)
0.5	0.036***	0.007	0.000	0.001
	(0.011)	(0.008)	(0.005)	(0.006)
0.6	0.042***	0.015	-0.001	-0.004
	(0.012)	(0.01)	(0.006)	(0.008)
0.7	0.057***	0.029**	-0.002	-0.012
	(0.015)	(0.012)	(0.008)	(0.01)
0.8	0.087***	0.044***	-0.002	-0.021*
	(0.021)	(0.015)	(0.009)	(0.013)
0.9	0.122***	0.097***	-0.003	-0.029*
	(0.03)	(0.022)	(0.009)	(0.017)

*Number of observations: 3,308. Significance levels corresponding to * p<0.1, ** p<0.05, and *** p<0.01. Standard errors are computed on 100 bootstrap repetitions.*

Table 6 – Aggregate decomposition results

Quantiles	Characteristics effect - explained by determinants		Coefficient effect - explained by coefficients	
	"2004–08"	"2009–14"	"2004–08"	"2009–14"
0.1	17.1%	34.9%	82.9%	65.1%
0.2	17.7%	40.8%	82.3%	59.2%
0.3	21.0%	39.0%	79.0%	61.0%
0.4	18.1%	37.9%	81.9%	62.1%
0.5	20.5%	42.7%	79.5%	57.3%
0.6	24.0%	41.5%	76.0%	58.5%
0.7	33.1%	42.2%	66.9%	57.8%
0.8	43.2%	45.8%	56.8%	54.2%
0.9	52.9%	86.6%	47.1%	13.4%

Table 7 – Results summary table

	Expansionary period					Contractionary period				
		Effect of					Effect of			
Quantile	Total estimated change	Public financing	Gazelle firms	Financial constraint	Firms' chars.	Total estimated change	Public financing	Gazelle firms	Financial constraint	Firms' char.
0.1	0.021	-0.023	-0.010	-0.019	0.072	0.013	0.026	-0.002	-0.001	-0.009
		-106.9%	-45.5%	-87.3%	339.7%		201.8%	-19.1%	-9.4%	-73.2%
0.3	-0.017	-0.031	0.007	-0.036	0.042	0.021	0.024	-0.001	-0.001	-0.002
		177.6%	-40.9%	208.0%	-244.6%		118.2%	-6.9%	-2.5%	-8.8%
0.5	-0.028	-0.040	0.029	-0.040	0.024	0.044	0.036	0.007	0.000	0.001
		145.9%	-104.4%	145.6%	-87.1%		82.7%	15.7%	0.4%	1.2%
0.7	-0.016	-0.051	0.067	-0.039	0.007	0.072	0.057	0.029	-0.002	-0.012
		310.9%	-408.6%	239.3%	-41.6%		80.0%	39.8%	-2.7%	-17.1%
0.9	0.014	-0.095	0.190	-0.038	-0.044	0.188	0.122	0.097	-0.003	-0.029
		-694.1%	1394.1%	-278.2%	-321.8%		64.8%	51.9%	-1.4%	-15.4%

Note: The coefficient estimates are reported from the above estimations, while percentage contributions to the total variation is computed and reported in the second line of each cell. Lastly, the total estimated change has a rough correspondence to the estimated "characteristics effect," as different decomposition orderings would return different pointwise estimates for the non-statistically significant effects.

Table A-1 – Coverage of Spanish INTERNAL R&D expenditure of PITEC and INE			
Year	**PITEC (m €)**	**INE (m €)**	**% coverage**
2003	3.82	4.44	86%
2004	4.44	4.86	91%
2005	5.08	5.48	93%
2006	5.68	6.55	87%
2007	6.21	7.45	83%
2008	6.90	8.07	85%
2009	6.28	7.56	83%
2010	5.67	7.50	76%
2011	5.78	7.39	78%
2012	5.2	7.09	73%
2013	5.11	6.90	74%
2014	5.02	6.78	74%

Data are expressed in millions (m €). Source: INE and PITEC

Table A-2 – Descriptive statistics– Spanish R&D firms								
Year	Internal R&D	Gazelles	Public financing	FC	Sales	Size	Sector-normalized sales growth	
2004	775,896.40	0.19	125,066.90	0.40	65.1×10^{5}	251.54	0.01	mean
	5,682.00	5,682.00	4,096.00	5,682.00	5,682.00	5,682.00	4102.00	N
	4.4×10^{8}	1,065.00	512×10^{6}	2,261.00	37×10^{10}	1,429,226.00	53.48	sum
	123×10^{8}	1.00	92.3×10^{8}	1.00	$5{,}79\times10^{7}$	20,155.00	6.63	max
	151.00	0.00	0.00	0.00	1,623.00	1.00	-8.16	min
	1.6×10^{13}	0.15	2.9×10^{12}	0.24	1.3×10^{17}	1.2×10^{6}	0.22	variance
	17.98	1.60	45.25	0.42	12.61	11.98	0.74	skewness
	441.02	3.57	2,285.63	1.17	181.85	173.38	63.34	kurtosis
2008	1,288,035.00	0.21	237,293.50	0.45	88.4×10^{6}	308.27	0.03	mean
	5,296.00	5,296.00	5,296.00	5,296.00	5,296.00	5,296.00	5261.00	N
	6.82×10^{7}	1,112.00	1.26×10^{7}	2,394.00	468×10^{9}	1,632,608.00	165.85	sum
	498×10^{6}	1.00	56.5×10^{6}	1.00	9.25×10^{7}	41,168.00	5.69	max
	921.70	0.00	0.00	0.00	1,661.00	1.00	-6.86	min
	1.2×10^{14}	0.17	1.9×10^{12}	0.25	1.9×10^{17}	2.3×10^{6}	0.24	variance
	31.51	1.42	21.34	0.19	12.17	15.97	-0.89	skewness
	1,207.59	3.03	684.55	1.04	186.69	341.81	40.72	kurtosis
2009	1,299,279.00	0.20	268,185.70	0.48	87.5×10^{6}	321.05	0.02	mean
	4,782.00	4,782.00	4,782.00	4,782.00	4,782.00	4,782.00	4741.00	N
	6.21×10^{7}	961.00	1.28×10^{7}	2,279.00	418×10^{9}	1,535,265.00	107.08	sum
	365×10^{6}	1.00	61.4×10^{6}	1.00	9.03×10^{7}	40,504.00	5.69	max
	2,129.00	0.00	0.00	0.00	1,484.00	1.00	-4.53	min
	7.1×10^{13}	0.16	2.9×10^{12}	0.25	2.3×10^{17}	2.6×10^{6}	0.23	variance
	26.70	1.49	20.45	0.09	12.98	15.41	-0.13	skewness
	965.69	3.23	567.85	1.01	206.13	308.55	25.03	kurtosis
2014	1,472,686.00	0.18	153,202.70	0.44	101×10^{6}	336.28	0.03	mean
	3,340.00	3,340.00	3,340.00	3,340.00	3,340.00	3,340.00	3318.00	N
	4.9×10^{9}	595.00	512×10^{6}	1,455.00	338×10^{9}	1,123,167.00	113.97	sum
	188.0×10^{6}	1.00	38×10^{6}	1.00	$9{,}680\times10^{6}$	37,835.00	5.72	max
	2,330.80	0.00	0.00	0.00	4,279.00	1.00	-5.37	min
	5.4×10^{13}	0.15	1.3×10^{12}	0.25	2.6×10^{17}	$2{,}0\times10^{6}$	0.12	variance
	14.14	1.68	21.21	0.26	13.00	13.47	1.12	skewness
	256.87	3.83	576.94	1.07	208.47	247.88	76.24	kurtosis

Table A-3 – Pairwise correlation matrix							
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
(1) Internal R&D	1						
(2) Public financing	0.05*	1					
(3) Gazelles	0.04*	0.01*	1				
(4) FC	0.10*	0.01*	0.05*	1			
(5) Sales	-0.02*	0.12*	0.00	-0.06*	1		
(6) Size	-0.05*	0.07*	0.00	-0.07*	0.63*	1	
(7) Sector-normalized sales growth	0.05*	0.01*	0.33*	-0.02*	0.05*	0.02*	1

* corresponds to the 0.05 significance level.

Table A-4 - Robustness checks recap

EXPANSIONARY PERIOD (2004–2008)					CONTRACTIONARY PERIOD (2009–2014)			
	(1)	(2)	(3)	(4)		(1)	(2)	(3)
	Effect of R&D public financing					**Effect of R&D public financing**		
(1) Baseline	1				**(1) Baseline**	1		
(2) Imputation	0.99*	1			**(2) Lag on RD_publi**	0.99*	1	
(3) 2005–2008	0.99*	0.99*	1		**(3) Reverse order**	0.98*	0.98*	1
(4) Reverse order	**0.95***	**0.96***	**0.95***	**1**				
	Effect of gazelles					**Effect of gazelles**		
(1) Baseline	1				**(1) Baseline**	1		
(2) Imputation	0.99*	1			**(2) Lag on RD_publi**	0.99*	1	
(3) 2005–2008	0.99*	0.99*	1		**(3) Reverse order**	0.44*	0.43*	1
(4) Reverse order	**0.95***	**0.95***	**0.92***	**1**				
	Effect of financial constraints					**Effect of financial constraints**		
(1) Baseline	1				**(1) Baseline**	1		
(2) Imputation	0.97*	1			**(2) Lag on RD_publi**	0.72*	1	
(3) 2005–2008	0.83*	0.80*	1		**(3) Reverse order**	0.23*	0.19	1
(4) Reverse order	0.74*	0.74*	0.84*	1				
	Effect of firms' characteristics					**Effect of firms' characteristics**		
(1) Baseline	1				**(1) Baseline**	1		
(2) Imputation	0.99*	1			**(2) Lag on RD_publi**	0.95*	1	
(3) 2005–2008	0.96*	0.96*	1		**(3) Reverse order**	-0.53*	-0.41*	1
(4) Reverse order	-0.38*	-0.37*	-0.57*	1				

Significance stars correspond to the 5% significance level. The quantile estimates can be shared on a request to the authors. For conciseness, we report the pairwise correlations for each of our four effects estimated along 100 percentiles of the distribution under each of the above specifications.

Figure 1

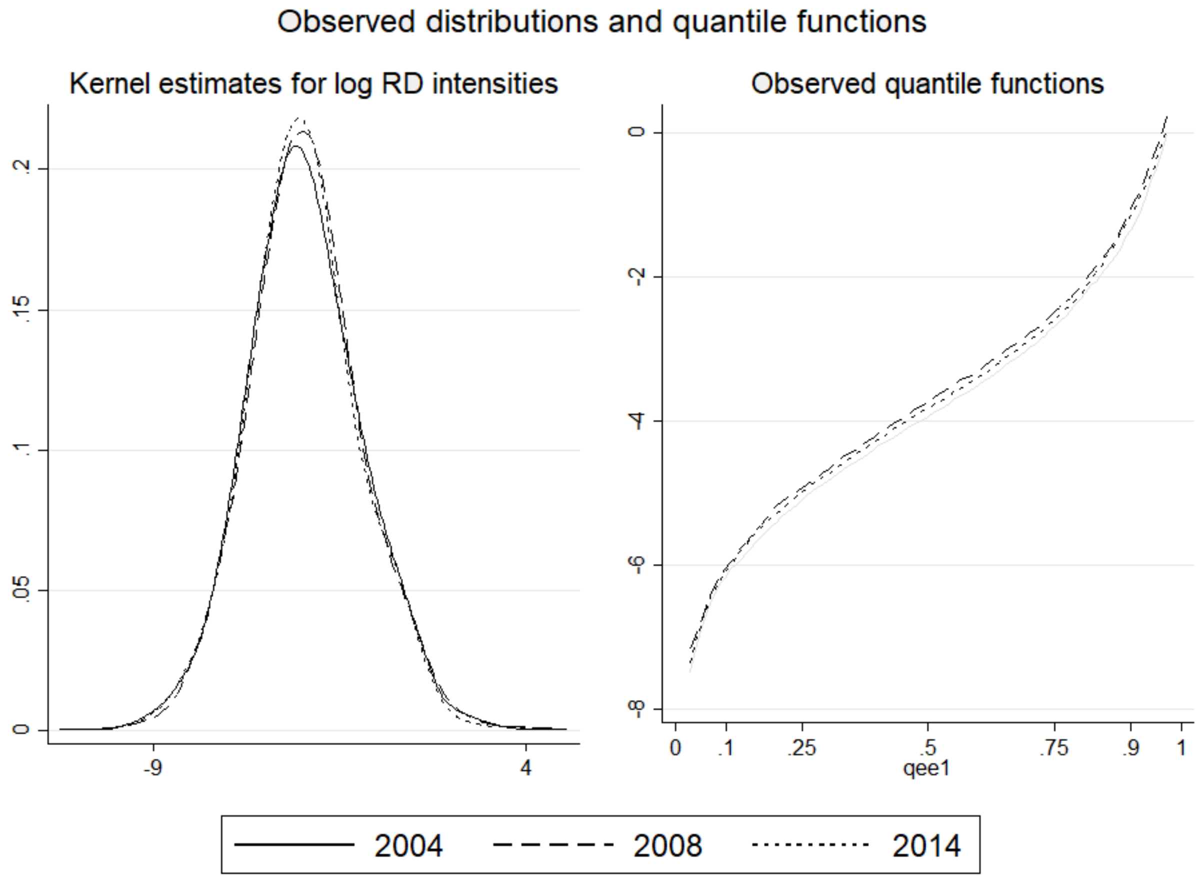


Figure 2

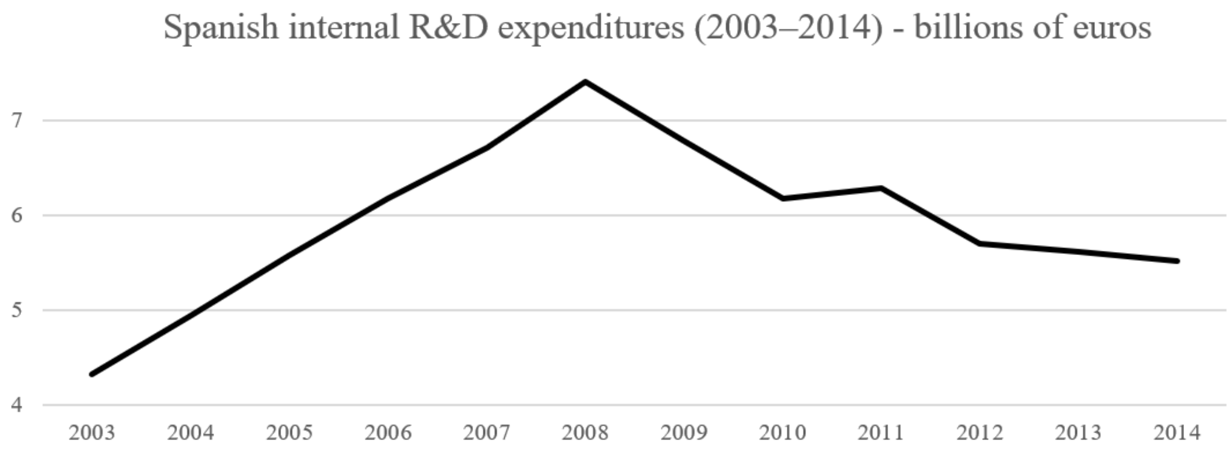
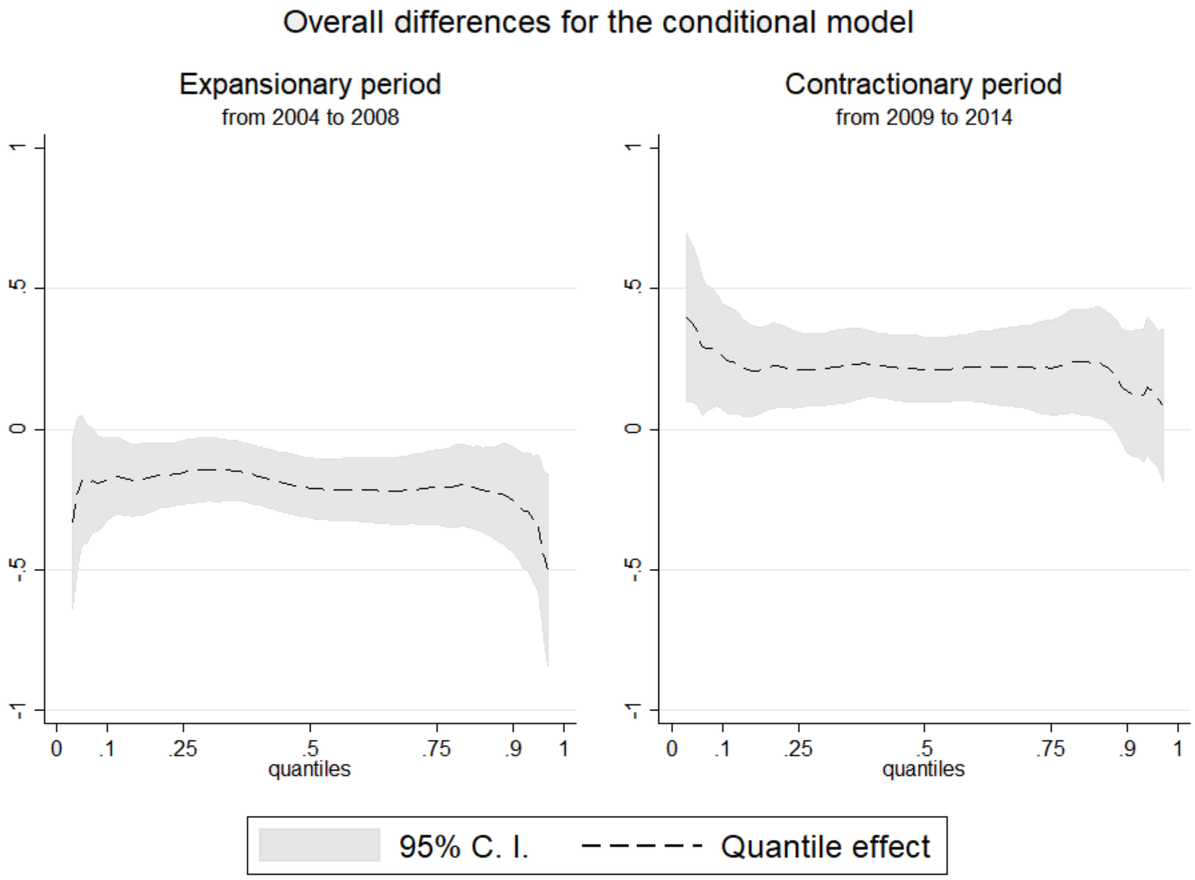
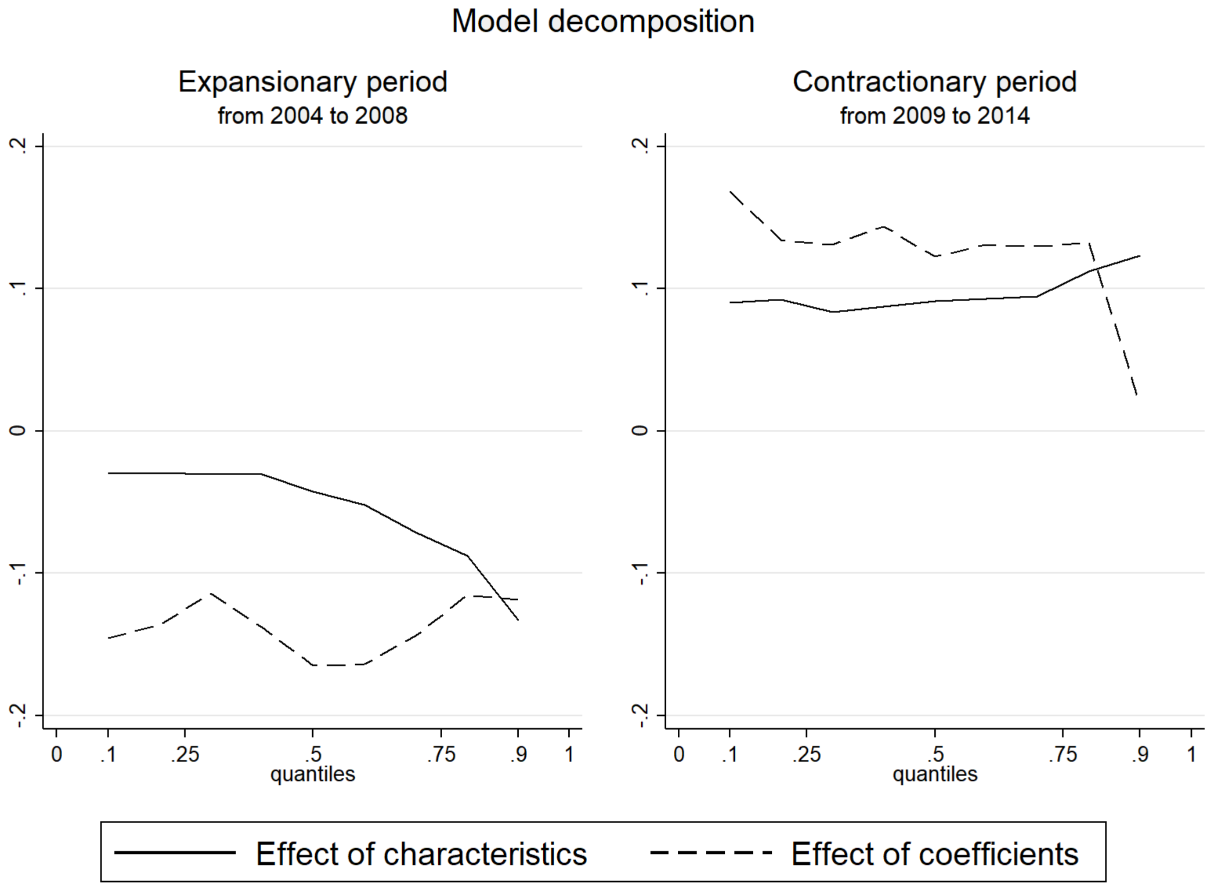


Figure 3



Overall differences for the conditional model
Expansionary period
from 2004 to 2008
Contractionary period
from 2009 to 2014
1
.5
0
-.5
-1
0
.1
.25
.5
.75
.9
1
quantiles
95% C. I.
Quantile effect

Figure 4



Model decomposition
Expansionary period
from 2004 to 2008
Contractionary period
from 2009 to 2014
.2
.1
0
-.1
-.2
0
.1
.25
.5
.75
.9
1
quantiles
Effect of characteristics
Effect of coefficients

Figure A-1

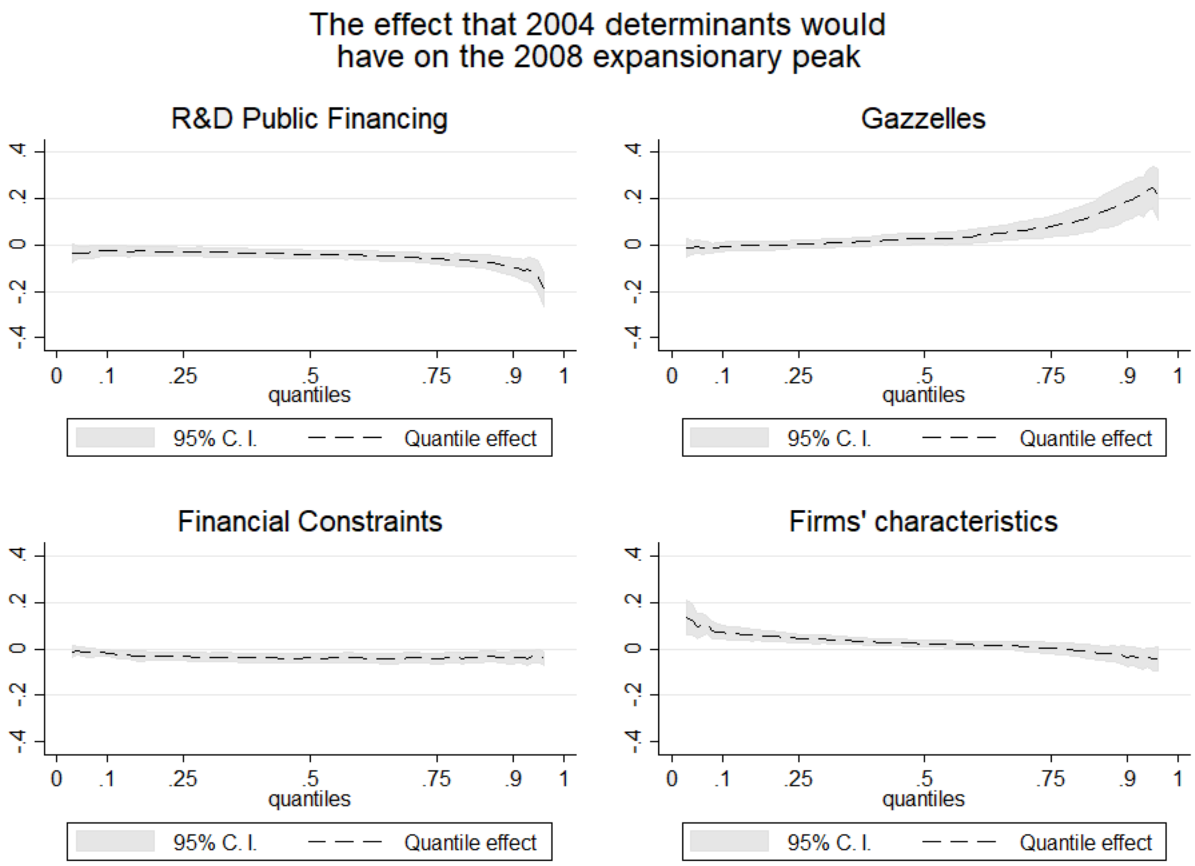
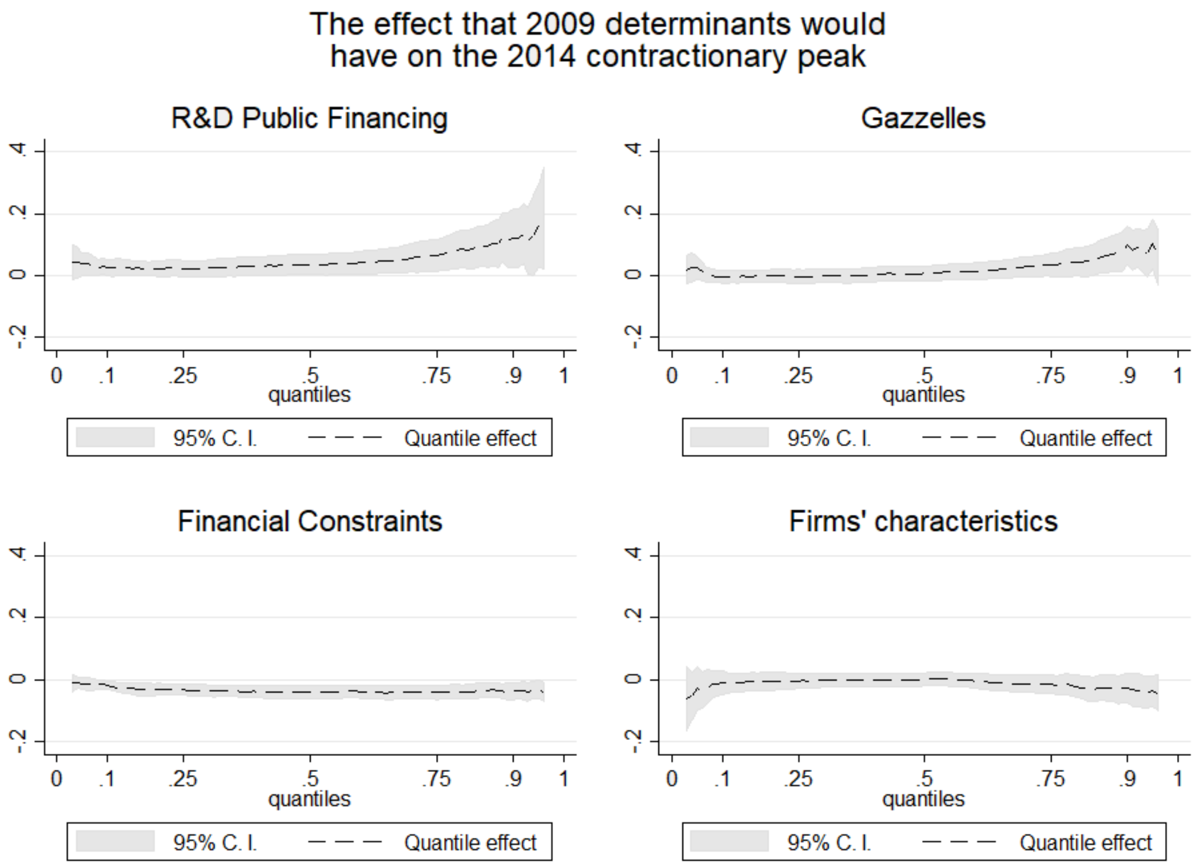

The effect that 2004 determinants would
have on the 2008 expansionary peak
R&D Public Financing
Gazzelles
Financial Constraints
Firms' characteristics
-.4
-.2
0
.2
.4
0
.1
.25
.5
.75
.9
1
quantiles
95% C. I.
Quantile effect

Figure A-2


The effect that 2009 determinants would have on the 2014 contractionary peak
R&D Public Financing
Gazzelles
Financial Constraints
Firms' characteristics
-.2
0
.2
.4
0
.1
.25
.5
.75
.9
1
quantiles
95% C. I.
Quantile effect

Caption list: