Nesta Working Paper No. 13/12

Innovation and UK High-Growth Firms

Vania Sena
Mark Hart
Karen Bonner

Innovation and UK High-Growth Firms

Vania Sena
University of Essex

Mark Hart
Aston University

Karen Bonner
Aston University

Nesta Working Paper 13/12
November 2013

www.nesta.org.uk/wp13-12

Abstract

In spite of the emphasis given to the role of innovation as a driver of high-growth episodes, very little is known about the characteristics of the process which leads to the production of innovation among HGFs. For instance, what are the most common sources of information for innovation among HGFs? What types of organisations do HGFs tend to collaborate with for the purpose of innovation? Do they differ from those used by more slower growing firms? Do HGFs benefit from specific types of knowledge spillovers? Do innovative HGFs produce more output than non-HGFs? Using linked ONS business datasets we undertake a series of econometric analyses to provide answers to these questions. We find that that growth at national and local level can be stimulated by policy interventions designed to foster greater levels of innovation among HGFs. This can happen in two broad ways: first, government can help create an environment where knowledge spillovers from both patents and investment in R&D can freely circulate. Our results show the importance of the spillovers generated by nearby firms (of either type: HGFs or non-HGFs) for HGFs and this suggests that creating the conditions for the clustering of the economic activities still makes sense. Second, the government can facilitate firms' investment in intangible assets which can help trigger high-growth episodes. Typically, governments tend to support investment in R&D though tax breaks but in reality the changing structure of the economy suggests that other types of intangible assets may be more important than investment in R&D to trigger innovation among HGFs.

JEL Classification: This classification system is prepared and published by the Journal of Economic Literature, see www.aeaweb.org/journal/jel_class_system.html for detailed codes.

Corresponding Author: Vania Sena. Southend Campus. Elmer Approach. Southend-on-Sea. SS1 1LW. United Kingdom. Email: vsena@essex.ac.uk

EXECUTIVE SUMMARY

The UK policy debate on job creation revolves around the performance of the high-growth firms (HGFs), that is, firms with ten or more employees which experience above average growth over three years (Anyadike-Danes et al., 2009; Mason et al., 2009). Although not much is known about the drivers of high-growth, innovation is usually considered to be one of the most likely candidates with the existing empirical evidence suggesting that product innovation is usually positively correlated with employment growth.

In spite of the emphasis given to the role of innovation as a driver of high-growth episodes, very little is known about the characteristics of the process which leads to the production of innovation among HGFs. For instance, what are the most common sources of information for innovation among HGFs? What types of organisations do HGFs tend to collaborate with for the purpose of innovation? Do they differ from those used by more slower growing firms? Do HGFs benefit from specific types of knowledge spillovers? Do innovative HGFs produce more output than non-HGFs?

The first part of this report tries to fill this gap in our knowledge about HGFs and addresses the above questions by estimating a set of innovation production functions and a production function on a sample of British HGFs in the first instance and then on a pooled sample of British HGFs and non-HGFs. In both cases, product and process innovation are estimated simultaneously as a function of the R&D input (which is allowed to be zero as some firms may have not invested in R&D and still be capable of producing some innovation outputs) and other firm's characteristics. Our independent variables of interest are: a) the sources of information for innovation, b) the types of organisations firms collaborate for innovation and c) the knowledge spillovers generated by the patents belonging to both types of firms, the investment in R&D of the upstream firms (of both types) and the investment in R&D of the neighbouring firms.

The empirical analysis employs the business datasets made available by the ONS through the Virtual Microdata Lab (VML), namely the Business Structure Database (BSD), the Community Innovation Survey (CIS), the Business Enterprise R&D Survey (BERD) and the Annual Respondents Database (ARD). We have first identified the HGFs in the Business Structure Database (BDS) and matched the resulting dataset to several releases of the Community Innovation Survey (CIS) – containing information on firms innovative outputs and inputs - the Business Enterprise R&D survey (BERD) and the Annual Respondents Database (ARD) – with information on firms' outputs, capital and labour. Our final dataset covers the period from 1998 to 2006 and is made of 1,248 HGFs and of 7,189 Non-HGFs.

The results of this part of the empirical analysis show that:

- Knowledge spillovers from neighbouring firms' patents are negatively associated to the probability of a HGF to be either a product or process innovator while R&D spillovers from high-growth neighbouring firms are positively associated to the propensity of HGFs to be product innovators.

- HGFs which collaborate with suppliers are more likely to be process innovators while those which collaborate with competitors are less likely to be product innovators.

- HGFs which source information from internal sources and from suppliers are more likely to be either product or process innovators while those which source information from universities are less likely to be either product or process innovators.

- HGFs do benefit more than non-HGFs from the knowledge spillovers generated from the investment in R&D of the neighbouring HGFs.

- HGFs which source information from higher education (HE) establishments/competitors/suppliers/internal sources are not more likely to be product/process innovators than non-HGFs.

- Equally, HGFs which collaborate with higher education (HE) establishments/competitors/suppliers/internal sources are not more likely to be product/process innovators than non-HGFs.

The second aim of the report is to quantify the impact of the HGFs' innovative activities on the propensity to innovate of those firms which do not experience high-growth; more specifically, we focus on the knowledge spillovers generated by the investment in R&D and the patents applied for by HGFs and test whether these are correlated with the propensity to innovate of the non-HGFs. We use the same dataset as in the first part of the report and the results suggest that knowledge spillovers generated by HGFs are not significantly correlated with the propensity to innovate of non-HGFs.

Finally, the report investigates whether there is a correlation between investment in intangible assets and the propensity to innovate of HGFs. In this part of the report, we focus on the role that intangible assets play in contributing to the HGFs' propensity to innovate and test whether HGFs benefit more from their investment in intangible assets than more slower growing firms. To this purpose, we estimate an innovation production function where a binary indicator of whether the firm has invested in some type of intangible assets appears now among the regressors and is interacted with the dummy for the high-growth status.

For this part of the report, we have used a different dataset. More specifically, we have merged the Sixth Community Innovation Survey (CIS) with the NESTA 2009 Intangible Asset Survey (IAS) - surveying firms' investment in different types of intangible assets. The results show that HGFs which invest in training are more likely to introduce a product or process innovation. Also investment in software among HGFs is positively associated to their propensity to introduce a product innovation (but not a process innovation).

The key policy implication of this report is that growth at national and local level can be stimulated by policy interventions designed to foster greater levels of innovation among HGFs. This can happen in two broad ways: first, government can help create an environment where knowledge spillovers from both patents and investment in R&D can freely circulate. Our results show the importance of the spillovers generated by nearby firms (of either type: HGFs or non-HGFs) for HGFs and this suggests that creating the conditions for the clustering of the economic activities still makes sense. Second, the government can facilitate firms' investment in intangible assets which can help trigger high-growth episodes. Typically, governments tend to support investment in R&D though tax breaks but in reality the changing structure of the economy suggests that other types of intangible assets may be more important than investment in R&D to trigger innovation among HGFs.

1. INTRODUCTION

Much of the UK policy debate on firms' growth and job creation tends to revolve around the performance of the high-growth firms (HGFs), that is, firms with ten or more employees which experience above average growth over three years (i.e., the OECD definition of a HGF). This interest has been prompted by recent research showing that around 6 percent of UK businesses with the highest growth rates have generated half of the new jobs in all businesses employing at least 10 employees in consecutive three-years periods – 2002-05 and 2005-08 (Anyadike-Danes et al., 2009).

What are the characteristics of HGFs? Although high-growth is a temporary stage of a firm's life and can be potentially attained by any firm, researchers have found that HGFs tend to share some common features. For instance, they can be found across a wide range of different size groups, sectors and regions (Anyadike-Danes et al., 2009); they are relatively young firms (as firms which are aged five years or less are over-represented among HGFs). Although not much is known about the drivers of high-growth, innovation and innovation success are usually considered to be the most likely factors. The existing empirical evidence suggests that product innovation is usually positively correlated with employment growth with innovation success being the main driver of firms' high-growth episodes (Coad and Rao, 2008; Holzl, 2009; Mason et al., 2009).

In spite of the emphasis given to the role of innovation as a driver of high-growth episodes, very little is known about the characteristics of the process which leads to the production of innovation among HGFs. At the moment, the existing evidence is very fragmented and as a result there are many unanswered questions. For instance, what are the most common sources of information for innovation among HGFs? What types of organisations innovative HGFs tend to collaborate with for the purpose of innovation? Do they differ from those used by the remaining firms? Do HGFs benefit from specific types of knowledge spillovers? Do innovative HGFs produce more output than non-HGFs?

The first part of this report tries to fill this gap in our knowledge about HGFs and addresses the above questions by estimating a set of innovation production functions and a production function on a sample of HGFs first and then on a pooled sample of high-growth (HG) and non-high-growth (NHG) firms. In the former case, the control group is made up of the non-innovative HGFs (this allows us to control for an additional source of heterogeneity in our sample, namely the high-growth dimension) while in the latter case, we compare directly the innovative HGFs to the non-HGFs. Our empirical specification is based on a smaller version of the well- known Crepon, Duguet, and Mairesse (CDM) model for innovation survey data which allows to model the innovation production process and then to estimate the impact that the innovation outputs have on the level of output. So, the two innovation equations are estimated simultaneously as a function of the R&D input (which is allowed to be zero as some firms may have not invested in R&D and still be capable of producing some innovation outputs) and other firm's characteristics while the predicted values of the innovation outputs are then added to the production function to control for their potential endogeneity. Our independent variables of interest are of three types: a) the sources of information for innovation, b) the types of organisations firms collaborate for innovation and c) a set of proxies for knowledge spillovers. In the model estimated on the pooled sample of HGFs and non-HGFs, these variables are interacted with a dummy variable taking the value of one for firms which have experienced high-growth to test whether their impact on the propensity to innovate differs between HGFs and non-HGFs.

In this report, we focus on knowledge spillovers generated by: a) the patents applied for by neighbouring firms, b) the investment in R&D performed by the upstream firms (either high growth or non high growth) and c) the investment in R&D performed by the neighbouring non-HGFs. Patents and investment in R&D are a standard source of knowledge spillovers and are widely used in the literature on knowledge spillovers while the requirement makes patents an established source of knowledge spillovers.

The second aim of the report is to quantify the impact of HGFs' innovative activities on the propensity to innovate of those firms which do not experience high-growth; more specifically, we focus on the knowledge spillovers generated by the investment in R&D of the neighbouring HGFs and the patents applied for by HGFs and we try to understand whether these are correlated to the propensity to innovate of the non-HGFs.

The notion that HGFs' activities can indirectly contribute to the performance of their neighbouring firms is not new. For instance, Mason et al. (2009) have quantified the indirect impact of the HGFs to the economic performance of 45 UK city-regions and found that local employment can be boosted thanks to the presence of HGFs. However, HGFs are also likely to contribute to the production of innovation among the slow-growing firms (although this has not been tested yet): as HGFs innovate more often, they can generate knowledge spillovers from which other firms can take advantage from with the result that their innovation outputs can be boosted. These results can potentially be quite relevant from a policy perspective as they show that supporting innovation among HGFs may be beneficial to the wider population of firms. Therefore, we estimate an innovation production function for non-HGFs where three measures of knowledge spillovers are introduced among the regressors in addition to the usual set of controls. Again, we consider three sources of knowledge spillovers: the R&D investment of the upstream firms, the R&D investment of neighbouring HGFs and the patents applied for by HGFs. The use of patents as a source of spillovers may provide useful insights to policy makers. Suppose for instance that the patenting activities of HGFs can stimulate the innovative activities among non-HGFs; if so, changes to the legislation which would allow HGFs to patent their innovations faster may help to increase of the innovation rate across the economy.

Lastly, the report investigates whether there is a correlation between investment in intangible assets and the propensity to innovate of HGFs. Some authors have suggested that some firms can be more innovative than others because they own some distinctive resources, which include the physical, technological, commercial assets used by firms to develop their new products (Barney, 1991). Here, in this part of the report, we focus on the role that intangible assets play in contributing to the HGFs' propensity to innovate and test whether HGFs benefit more from their investment in intangible assets than slow-growth firms. To this purpose, we estimate an innovation production function where a binary indicator of whether the firm has invested in any type of intangible assets appears now among the regressors and is interacted with the dummy for the high-growth status.

For the empirical analysis, we have used the business datasets made available by the ONS through the Virtual Microdata Lab (VML). For the first and the second parts of the empirical analysis focusing on the innovation production functions for high-growth firms and the spillovers generated by their activities, we use the Business Structure Database (BSD), the Community Innovation Survey (CIS), the Business Enterprise R&D Survey (BERD) and the Annual Respondents Database (ARD). So, we have first identified the HGFs in the Business Structure Database (BDS) and matched the resulting dataset to several releases from the Community Innovation Survey (CIS) – containing information on firms innovative outputs and inputs - the Business Enterprise R&D survey (BERD) and the Annual Respondents Database (ARD) – with information on firms' outputs, capital and labour. Our final dataset covers the period from 1998 to 2006 and is made of 1248 HGFs and of 7189 Non-HGFs. For the third part of the analysis (i.e. the one focusing on the investment in intangible assets and HGFs), we have merged the Sixth Community Innovation Survey with the NESTA 2009 Intangible Asset Survey (IAS) surveying firms' investment in different types of intangible assets.
Our results suggest that:

- Knowledge spillovers from neighbouring firms' patents are negatively associated to the probability of a HGF to be either a product or process innovator while R&D spillovers from high-growth neighbouring firms are positively associated to the propensity of HGFs to be product innovators.

- HGFs which collaborate with suppliers are more likely to be process innovators while those which collaborate with competitors are less likely to be product innovators.

- HGFs which source information from internal sources and from suppliers are more likely to be either product or process innovators while those which source information from universities are less likely to be either product or process innovators.

- HGFs do benefit more than non-HGFs from the knowledge spillovers generated by the investment in R&D of neighbouring HGFs.

- HGFs which source information from higher education establishments/competitors/suppliers/internal sources are not more likely to be product/process innovators than non-HGFs.

- Equally, HGFs which collaborate with higher education establishments/competitors/suppliers/internal sources are not more likely to be product/process innovators than non-HGFs.

- Knowledge spillovers generated by the HGFs are not significantly correlated with the propensity to innovate of the non-HGFs.

- HGFs which invest in training are more likely to introduce a product or process innovation. Also investment in software among HGFs is positively associated to their propensity to introduce a product innovation (but not a process innovation).

The paper is structured as follows. Section 2 reviews the existing evidence on the link between high-growth firms and innovation. Section 3 outlines our empirical specification while Section 4 discusses the data. Section 5 presents the results of our empirical analysis. Finally some conclusions are offered in Section 6.

2. LITERATURE REVIEW

HGFs are defined by the OECD as those firms with ten or more employees that have average annual growth rates of 20 percent or more (in terms of employment or sales) over a three-year period. Sometimes the term HGFs is used interchangeably with the term "gazelles" which are defined as a sub-set of HGFs born five years (or less) before the end of a three-year period.

Although the policy interest is relatively recent, researchers have been interested in HGFs for a while, mostly because of their contribution to job creation (Delmar et al., 2003). Research has mostly tried to establish some empirical facts about HGFs. Parker et al. (2010) and Henrekson and Johansson (2010) compiled a survey on HGFs and found that they can be in all sectors, they tend to be younger than the rest of firms and may be more R&D intensive than other growing firms. They may overlap with the SMEs even if in reality high-growth can be experienced by large firms as well. They also confirmed that for the US and UK, this segment of the firms' population is the central driver of the aggregate job creation.

However, not much is known about the drivers of high growth. The capability of firms to achieve high-growth status is usually attributed to innovation (either product or process or organisational innovation). However, the empirical literature has not been able to identify a clear link between innovation and firms' growth (see Coad and Rao, 2008). For instance, it is usually found that product innovation is positively correlated with employment growth but the results are more ambiguous in the case of process innovation. Some authors have suggested that innovation success is more important than innovation *per se* but they still get some mixed results (Mason et al., 2009). For instance, Hozl (2008) finds that high-growth firms in Southern European countries do not differ greatly from non-high growth firms in terms of innovation success while the opposite is true for the Northern EU countries.

Evidence from the UK is actually clearer in showing the centrality of innovation for high-growth. Mason et al. (2009) use the Community Innovation Survey (CIS), 2002-04, and match them to the firm-level data in the Business Structure Database (BSD). The descriptive analysis suggests that HGFs tend to be more innovative than other firms in the economy. Also, the econometric analysis suggests that HGFs with a higher share of sales from new products experiences an improvement to its employment growth rate.

Evidence on the characteristics of the process which leads to the production of innovation among HGFs is limited. HGFs are less likely to experience barriers to innovation than their slow-growth counterparts (Holzl and Janger, 2011) which is consistent with the fact that HGFs are more likely to be either product or process innovators. There is also some limited evidence on the types of collaborators HGFs prefer to work for the purpose of innovation. For instance, Holzl and Friesenbichler (2008) find that high-growth SMEs in Northern Europe are more likely to collaborate on innovation with universities and other firms than high-growth SMEs from Southern Europe.

In terms of the resources used by HGFs, some authors have focused on the investment in R&D and tested formally whether HGFs are more R&D intensive than slower growing firms. For instance, Holzl (2008) has used the Community Innovation Survey (CIS), 1998-2000, for several European countries and found that within Northern EU countries, high-growth SMEs tend to be more R&D intensive than their Southern European counterparts.

Finally, there exists some research showing that HGFs can improve the performance of their local area. In general, there are several channels through which this can happen: first, fast growing firm can help reduce the unemployment rate in an area; second, they can generate positive spillovers which in turn can help local firms to grow. Last but not the least, as innovation is an important output of the activities undertaken by HGFs, they may generate knowledge spillovers which can boost the local firms' capability to innovate. However, research has only focused on points 1) and 2). Mason et al. (2009) used a dataset of 45 UK city-regions and found that HGFs have a positive impact on local employment.

All in all, this quick review of the existing literature on innovation among HGFs shows that our knowledge of the innovation production process among HGFs is very patchy and, therefore, further investigation is needed. In particular, for policy purposes it would be interesting to know whether innovative HGFs take advantage of specific types of knowledge spillovers, whether they use specific sources of information and whether there are specific internal resources that innovative HGFs invest in for the purpose of innovation.

3. EMPIRICAL FRAMEWORK

Our key empirical specification is based on a smaller version of the well- known Crepon, Duguet, and Mairesse (CDM) model for innovation survey data where product and process innovation are estimated simultaneously as a function of the R&D input (which is allowed to be zero), knowledge spillovers, sources of information, types of collaborators and other firm's characteristics. The CDM model does not describe a set of causal relationships, because of the lack of appropriate instruments and therefore the resulting estimates describe the correlations in the data.

To allow for the fact that the R&D input is potentially endogenous, the R&D input is replaced by the predicted value of the investment in R&D whose estimate is computed from a system of two equations - the firm's decision to invest in R&D and the resulting R&D intensity measured as R&D per employee. Consistent with Griffiths et al. (2006), we estimate the CDM model for all firms and not only those which report a positive investment in R&D as some firms may not invest in R&D and still be able to produce either a product or a process innovation.

Formally, the first two equations model simultaneously the firm's decision to invest in R&D and its R&D intensity using a standard Tobit type II or sample selection model. The decision to invest in R&D is governed by the following equation:

$$rd_i = \begin{cases} 1 \text{ if } rd^* = w_i\alpha + \varepsilon_i > 0 \\ 0 \text{ if } rd^* = w_i\alpha + \varepsilon_i \leq 0 \end{cases} \qquad i=1,\dots,N \qquad (1)$$

Where rd^* is an unobservable latent variable whose value determines whether the firm invests in R&D, rd is an observed indicator variable that is equal to zero for firms that do not invest in R&D and equal to one for R&D-investing firms, w is a vector of variables associated to the R&D investment decision, α is a vector of parameters to be estimated and ε_i is an error term. In the model estimated on the pooled sample of HGFs and non-HGFs, we add a dummy variable taking the value of one for HGFs.

Conditional on firms investing in R&D, we observe the amount of resources invested in R&D (modelled as R&D intensity, the logarithm of R&D per employee):

$$r_i = \begin{cases} z_i\beta + e_i \text{ if } rd_i \neq 0 \\ 0 \qquad\qquad \text{if } rd_i = 0 \end{cases} \qquad (2)$$

where z_i is a vector of variables affecting the R&D intensity, β is the vector of coefficients and e_i is an error term. Again in the model estimated on the pooled sample of HGFs and non-HGFs, we control for the high-growth status of firms by introducing a dummy variable taking the value of one for HGFs. Assuming that the two error terms are distributed as a bivariate Normal with zero mean, variances $\sigma_{\varepsilon}^{2}=1$ and σ_{e}^{2} and a correlation coefficient ρ, the system of equations (1) and (2) can be estimated as a generalised Tobit model.

The next equations in our model are the two innovation production functions where we distinguish between two types of innovation outcomes (product and process innovations). It is usually assumed that there may be unobservables that drive the production of both types of innovation with the result that the two innovation production functions are correlated and, therefore, estimated in a simultaneous fashion. Each type of innovation is measured by a dummy variable (INN) indicating whether the firm has introduced at least one product/ process innovation:

$$INN_{i}=\chi r_{i}^{*}+x\delta+\lambda z+d_{s}+d_{r}+u_{i} \qquad (3)$$

where *INN* is the measure of innovation, r^* is the predicted value of the R&D intensity (this way we can control to some extent for the fact that the investment in R&D is endogenous), x is a vector of variables that affect the firms' propensity to innovate, z are our variables of interest (information sources, types of collaborators and knowledge spillovers), d_s and d_r are industry and region dummies and u^1 is the residual. In the case of the model estimated on the pooled sample of HGFs and non-HGFs, the z variables are interacted with the dummy for HGFs. We estimate (3) simultaneously as a bivariate probit system where the two disturbances are assumed to be correlated.

Firms produce output using a Cobb-Douglas production function where labour (l), capital (k) and the predicted value of the innovation output appear as inputs (in turn these are interacted with the dummy for HGFs in the case of the model estimated on the pooled sample of HGFs and non-HGFs). Finally, we also include the usual set of local and industry dummies to control for unobserved characteristics that affect the output level:

$$y_{i}=a+bk_{i}+el_{i}+\pi\widehat{INN}_{i}+d_{s}+d_{r}+v_{i} \qquad (4)$$

4. VARIABLES AND DATA

The data we use for our empirical analysis has been created by merging four different datasets at the Office of National Statistics Virtual Microdata Laboratory (ONS VML): 1) the Business Structure Database (BSD); 2) the Annual Respondents Database (ARD); 3) the UK Community Innovation Surveys (CIS), releases 3, 4, 5; 4) Patent data from the UK Intellectual Property Office (IPO). In linking the different datasets, we focus on the sample of firms covered by the CIS. It is important to notice that there is not too much variation across the different CIS waves and this suggests that most of the variation is cross-sectional. We conduct the analysis at the 'firm' level as we can identify whether a firm has experienced an episode of high-growth only through the BSD (which contains information at the firm level). In principle, the linked dataset is a firm-level panel containing detailed information on firm characteristics, innovative activities as well as patent and trademark filings over the nine year period 1998-2006.

In the empirical implementation of the structural model outlined in Section 2, we follow Griffith *et al*. (2006) and we distinguish between two different kinds of innovation outcomes, product and process innovation. Each innovation indicator is proxied by a dummy variable taking the value of 1 in case the firm has introduced at least one product or one process innovation (either new to the market or to the firm). The specification of the innovation production function captures the different motivations behind the propensity to develop innovations, motivations which we assume to be common to both HGFs and non-HGFs. Traditionally, the innovation literature distinguishes between technology push and demand pull determinants of innovation. In the first case, the production of innovation is driven by the development of internal capabilities and routines (typically proxied by the R&D investment) which drives the identification of technological opportunities; in the second case, innovations are developed to address the demand (from either consumers or other firms) for new goods or services. In the empirical specification, the technology push factors are proxied by the R&D intensity while the demand pull factors are proxied by the proportion of firms in the 3-digit industry claiming that meeting regulations or addressing environmental concerns is of high, medium, or low importance for innovation (as opposed to no importance).

We also control for the degree of competition a firm faces. Competition may drive innovation forward as firms belonging to more competitive sectors are forced to introduce new processes and products so to maintain their market share. In the empirical specification, the degree of competition a firm faces is proxied by the Herfindal index.

Our variables of interest include the sources of information firms use to innovate, their preferred collaborators and the knowledge spillovers they can eventually benefit from. We focus on four main sources of information and types of collaborators, namely internal sources (and collaborators), competitors, suppliers and universities.

We focus on the knowledge spillovers generated by: a) the patenting activities of the neighbouring firms (measured as the number of patents' applications by postcode weighted by the distance between postcodes); b) the investment in R&D of the upstream firms (either HGFs or non-HGFs) and c) the R&D investment of the non-HGF neighbouring firms. Both patents and investment in R&D are well established sources of knowledge spillovers: indeed patents allow firms to access technical knowledge which could help them to innovate further while investment in R&D produces new ideas and knowledge which can help either downstream or neighbouring firms to improve their existing products or to introduce new ones. Finally, the geographical dimension of the knowledge spillovers has been identified as important for innovation (Baptista and Swan, 1998).

To test whether sources of information and types of collaborators affect differently the propensity to innovate of HGFs and non-HGFs, we interact them with a binary indicator taking the value of one for HGFs. We add a set of firm-level controls which include the size of the firm, its age, whether it belongs to a high-tech sector and whether it is foreign–owned. Finally, we control for the two-digit industrial sector, the CIS wave the firm was sampled from and the region where the firm is located.

In the value-added equation, we measure output as the firm's value added while labour and capital are measured respectively by the number of employees and by the stock of capital at time *t*. As mentioned above, we introduce among the regressors the predicted values of the two innovation outputs and interact them with the high-growth status indicator to test whether the impact of the each type of innovation on the value added is different between HGFs and non-HGFs.

Table 1 gives a quick overview of the main characteristics of the HGFs and Non-HGFs in our sample, with the last column indicating whether the differences between the means are significant. On average, HGFs are 17 years old and 40% of HGFs are foreign-owned. Two-fifths (42%) of HGFs have introduced product innovations, and 30% of HGFs have introduced a process innovation during the same period. Over two-thirds (70%) of HGFs are exposed to international competition while 20% of HGFs collaborate with other organisations. They use a variety of sources of information where suppliers (80%) and customers (81%) seem to be the most important sources of information for innovation. Finally, 40% of HGFs source information for innovation from an higher education institution. As for the non-HGFs, on average they have the same age as their high-growth counterparts although 35% (cf 25% of HGFs) of them are likely to introduce a product (process) innovation. One-third (36%) of non-HGFs are owned by a foreign company and 64% of non-HGFs are exposed to international competition. Only 16% of non-HGFs collaborate with another organisation (either private or public). In terms of sources of information for innovation, they source information mostly from suppliers and customers while only 37% of non-HGFs source information from a Higher Education Institution (HEI). Only 6% of non-HGFs have applied for a patent.

5. RESULTS

5.1 Main results

Table 2 shows the results of the innovation production functions for the high growth firms only[1]. First of all, the two equations appear to be positively correlated (conditional on the observables) with most correlations being significant and positive. In terms of our variables of interest, the estimates show that:

a) Knowledge spillovers from neighbouring firms' patents are negatively associated to the probability of a HGF to be either a product or a process innovator.

b) R&D spillovers from high growth neighbouring firms are positively associated to the propensity of HGFs to be a product innovator.

c) HGFs which collaborate with suppliers are more likely to be process innovators.

d) HGFs which collaborate with competitors are less likely to be product innovators.

e) HGFs which source information from internal sources and from suppliers are more likely to be either product or process innovators.

f) HGFs which source information from universities are less likely to be either product or process innovators.

As for the other control variables, the estimates show that the innovators among the HGFs tend to be older than the non-innovators and to have a higher R&D intensity. Demand pull factors are significantly associated with the propensity of HGFs to innovate although the direction of the effect is the opposite expected. For instance, the proportion of firms innovating to meet regulatory constraints is positively associated with the propensity of HGFs to introduce either a product or a process innovation; however, the proportion of firms innovating for environmental concerns is negatively correlated to the propensity of HGFs to introduce either a product or a process innovation. Finally, Table 3 presents the estimates of the production function for HGFs where HGFs which introduce either a product or a process innovation experience an increase in value-added. In an additional set of regressions, we have estimated the direct impact of the

[1] The results from the first stage are not presented here but can be made available.

different types of collaboration on the value-added of HGFs and the results suggest that these do not have a significant impact on the value-added of HGFs. Our estimates of the innovation production functions for the pooled sample of HGFs and non-HGFs are presented in Table 4. First of all, our estimates tell us that product and process innovation are positively correlated (conditional on the observables), with most correlations being positive (between 0.60 and 0.67) and significant. Second, the estimates seem to be robust across the different specifications.

In terms of our variables of interest, the results suggest that:

a) HGFs do not benefit more (in terms of propensity to be either a product or a process innovator) than non-HGFs from the patenting activities of their neighbouring firms.

b) In a similar fashion, HGFs do not benefit more than non-HGFs from the knowledge spillovers generated from the investment in R&D of the neighbouring firms or the upstream firms.

c) HGFs do benefit more than non-HGFs from the knowledge spillovers generated from the investment in R&D of the neighbouring HGFs.

d) HGFs which source information from higher education establishments/competitors/suppliers/internal sources are not more likely to be a product/process innovator than non-HGFs.

e) Equally, HGFs which collaborate with higher education establishments/competitors/suppliers/internal sources are not more likely to be a product/process innovator compared than non-HGFs.

As for the other variables, both product and process innovators tend to be older than the non-innovators. They are more likely to belong to the high-tech sectors and the R&D intensity is positively correlated with the propensity to innovate. Also, foreign firms are more likely to be process innovators. The level of the concentration of the product market is significant but with a negative sign. Among the demand-pull factors, only the share of firms in the sector innovating for environmental concerns is significant although with a negative sign.

Table 5 presents the estimates of the production functions for our pooled sample. The estimates are consistent across the different specifications and they suggest that HGFs which introduce either a product or a process innovation experience an increase in their value-added.

Table 6 presents the estimates of the innovation production functions for non-HGFs where the knowledge spillovers from HGFs are introduced as an additional set of independent variables. Overall, the estimates suggest that the knowledge spillovers generated by HGFs are not significantly correlated with the propensity to innovate of the non-HGFs. As for the other control variables, older and foreign firms are more likely to be either process or product innovators. Demand pull factors are significantly associated with the likelihood of being either a product or a process innovator, although with the opposite signs as firms which belong to sectors where innovation is driven by the need to address regulatory requirements are more likely to innovate while the opposite is true for firms belonging to sectors where innovation is driven by environmental concerns. As for collaboration, collaborating with suppliers is positively associated with the likelihood of introducing a process innovation, but not a product innovation. Collaborating with competitors and universities is negatively associated with the likelihood of being either a product or a process innovator. As for the sources of information, sourcing information from internal sources and from suppliers is positively associated with the likelihood of being either a product or a process innovator while the opposite is true for firms sourcing information from competitors and higher education institutions.

5.2 Intangible assets and HG firms

Recent empirical research on the determinants of firm performance has focused on the role that investments in intangible assets can play in fostering growth. Given the obvious policy interest in the topic, we have decided to investigate whether the data suggest the possibility of an empirical association between investment in intangible assets and the innovative activities of HGFs. None of the datasets used for the above empirical analysis contains information on investment in intangible assets and, therefore, we use a different data-set for this part of the empirical analysis. In 2009 NESTA commissioned a survey on the investment in intangible assets among the general population of firms. The survey – known as Intangible Asset Survey (IAS) – explores the level of spending and the life lengths of private sector investments in intangible assets. The survey draws on a statistically representative sample of 2004 UK private sector firms, using the UK business register, and was conducted between October 2009 and January 2010. The total number of observations is 838. As well as asking about R&D expenditure, the survey asked firms to detail the expenditure on a wider range of intangible assets: training, software, branding, design and business process. We have matched the data from the IAS to the BSD to identify the HGFs first and then to the sixth wave of the Community Innovation Survey (CIS6) to get information about their innovation activities. We decided to match with the CIS6 as this covers the period 2007-2009, broadly consistent with the years the IAS survey refers to. In the matching process, we have taken into account the fact that the IAS collects the data at the level of establishment while the BSD allows the identification of HGFs at firm (or enterprise) level. Because of this, we have collapsed the establishment data to the firm level so that the new dataset can identify whether a firm has experienced high-growth and whether it has invested in any type of intangible assets. The match between the BSD and the IAS has produced a dataset of 669 observations and only 68 of those have experienced high-growth. Table 7 shows the proportion of HGFs and non-HGFs which invest in each type of intangible assets based on the matched BSD and IAS dataset. Overall, HGFs seem more likely to invest in most types of intangible assets (apart from investment in R&D), although the means are not significantly different.

The empirical specification of the model which relates the investments in intangible assets of HGFs with their propensity to innovate is much simpler. First of all, given the exploratory nature of this analysis, we estimate directly the two innovation production functions without controlling for the endogeneity of the investment in intangible assets. Second, the empirical analysis is carried out on both HGFs and non-HGFs – that is, the control group is made of the sub-set of firms which do not experience high-growth. Therefore, we introduce in our specifications a dummy variable for HGFs and interact it with the investment in intangible assets. Finally, we control for the concentration in the 3-digit industry and introduce the two demand pull factors among the regressors. We have tried to control for the sources of information but the quality of the matching prevents us to get information on all possible sources of information and, therefore, we have decided to drop this group of variables altogether. In the estimation, we have tried to control for additional investments in intangibles but some of these dummy variables were dropped by Stata because of multicollinearity. Finally we also introduce a set of dummy variables for the size and the industry. The results we get from the estimation of the two innovation production functions (see Table 8) show that HGFs which invest in training are more likely to introduce a product or process innovation. Investment in software among HGFs is positively associated with their propensity to introduce a product innovation (but not a process innovation). As for the other types of investment in intangible assets, they do not seem to be associated to the likelihood of HGFs to introduce any type of innovation.

6. CONCLUSIONS

The purpose of this report was threefold. First, we want to shed some light on the characteristics of the innovation production process among HGFs. More specifically, we wanted to address the following questions: what are the favourite sources of information for innovation among HGFs? What types of organisations do innovative HGFs collaborate with for the purpose of innovation? Do they differ from those used by the remaining firms? Do HGFs benefit from specific types of knowledge spillovers? Do innovative HGFs produce more output than non-HGFs?

Second, the report investigates where there is a correlation between the knowledge spillovers generated by the innovative activities of HGFs and the propensity to innovate of those firms which do not experience high-growth. Third, the report tests whether there is a correlation between investment in intangible assets and the innovative propensity of HGFs.

The empirical analysis has been conducted on a sample of HGFs and non-HGFs which have been constructed by using different business databases made available from the ONS VML. Our results suggest that:

- Knowledge spillovers from neighbouring firms' patents are negatively associated with the probability of a HGF to be either a product or process innovator while R&D spillovers from high-growth neighbouring firms are positively associated with the propensity of HGFs to be product innovators.

- HGFs which collaborate with suppliers are more likely to be process innovators while those which collaborate with competitors are less likely to be product innovators.

- HGFs which source information from internal sources and from suppliers are more likely to be either product or process innovators while those which source information from universities are less likely to be either product or process innovators.

- HGFs do benefit more than non-HGFs from the knowledge spillovers generated by the investment in R&D of the neighbouring HGFs.

- HGFs which source information from higher education establishments/competitors/suppliers/internal sources are not more likely to be a product/process innovator than the non-HGFs.

- Equally, HGFs which collaborate with higher education establishments/competitors/suppliers/internal sources are not more likely to be a product/process innovator than non-HGFs.

- Knowledge spillovers generated by HGFs are not significantly correlated with the propensity to innovate of non-HGFs.

- HGFs which invest in training are more likely to introduce either a product or a process innovation. Also investment in software by HGFs is positively associated to their propensity to introduce a product innovation (but not a process innovation).

References

Anyadike-Danes, M., Bonner, K., Hart, M. and Mason, C., (2009), Measuring Business Growth: High-growth firms and their contribution to employment in the UK, NESTA, London.
Baptista, R. and Swann, P. (1998), Do firms in cluster innovate more?, Research Policy, 27, 525-540.
Barney, J. (1991), Firm resources and sustained competitive advantage, Journal of Management, 17, 99-120.
Birch, D.L. (1981), Who Creates Jobs?, The Public Interest, 65, 3-14.
Coad, A. and Rao, R. (2008), Innovation and high-growth in high-tech sectors: a quantile regression approach, Research Policy, 37, 633-648.
Crepon, B., Duguet, E. and Mairesse, J., 1998, Research, Innovation and Productivity : an Econometric Analysis at the firm level, Economics of Innovation and New Technology, 7, 115-158.
Demar, F., Davidsson, P. and Gartner, W. B. (2003), Arriving at the high growth firm, Journal of Business Venturing, 18, 189-216.
Griffith, R., E. Huergo, Jacques Mairesse & B. Peters, 2007, Innovation and Productivity across Four European Countries, Oxford Review of Economic Policy, 22(4), 483-498
Henrekson, M. and Johansson, D. (2010), Gazelles as Job Creators: A Survey and Interpretation of the Evidence, Small Business Economics, 35, 227-244.
Holzl, W. (2008), Is the R&D behaviour of Fast Growing SMEs Different? Evidence from CIS III Data for 16 Countries? WIFO Working Papers, 327.
Holzl, W. and Friesenbichler, K., (2008), Final Sector Report: Gazelles". Europe Innova Sector Report, Vienna: WIFO.
Mason, G., Bishop, K. and Robinson, C., (2009), Business Growth and Innovation, NESTA, London.
Parker, S. C., Storey, D. J. and Witteloostuijn, A van, (2010), What happens to Gazelles? The Importance of Dynamic Management Strategy, Small Business Economics, 35, 203-226.

Table 1. *Descriptive Statistics – High-Growth Firms and Non High-Growth Firms*

	HGFs	Non-HGFs	t-test
Product Innovators (1/0)	42%	35%	Significant at 5%
Process Innovators (1/0)	30%	25%	Significant at 5%
Foreign ownership (1/0)	40%	36%	Significant at 5%
International Competition (1/0)	70%	64%	Significant at 5%
Collaboration (1/0)	20%	16%	Significant at 5%
R&D	69%	48%	Not significant
Source of Information: Internal	74%	66%	Significant at 5%
Source of Information: Suppliers	80%	76%	Significant at 5%
Source of Information: Customers	81%	76%	Significant at 5%
Source of Information: Competitors	72%	67%	Significant at 5%
Source of Information: Higher Education	40%	37%	Not Significant
High-tech firms	0.9%	0.024%	Not significant
Concentration index	0.02%	0.02%	Significant at 5%
Formal IP (1/0)	40%	43%	Significant at 5%
Informal IP (1/0)	55%	59%	Significant at 5%
Regulation (1/0)	64%	10%	Significant at 5%
Age	17	18	Not significant
Has applied at least one UK/EPO patent (number of observations)	114	437	Significant at 5%

Note: HGF stands for high-growth while Non-HGF stands for non-high-growth firms. The cells indicate the percentage of non-HGF/HGFs with the relative attribute (for instance, product innovation). Source: Authors' calculations based on the BSD-CIS combined dataset.

Table 2. *Innovation Production Functions – High Growth firms*

Process Innovation	***Coeff.***	***Coeff.***
Spillovers from patents of neighbouring firms	-0.048*	-0.05*
R&D spillovers along the value chain from non high growth firms	0.0001	0.0001
R&D spillovers from high growth firms	0.00003	0.00004
R&D spillovers from non high growth neighbours	0.00001	0.00002
Concentration index	1.0454**	0.963**
Regulation	1.4347***	1.768***
Environmental regulation	-0.6728***	-0.631**
Foreign dummy	0.05951	0.231**
Predicted R&D	0.3575***	0.316***
Age	0.0091***	0.019***
Collaboration: Internal	-	0.28
Collaboration: Suppliers	-	0.411***
Collaboration: Competitors	-	-0.0009
Collaboration: Universities	-	-0.391
Sources of Information: Internal	0.4878***	0.618***
Sources of Information: Suppliers	0.832***	0.289***
Sources of Information: Competitors	-0.0965	-0.129***
Sources of Information: Universities	-0.2872***	-0.291***
High tech dummy	0.8588	0.798*
Product Innovation		
Spillovers from patents of neighbouring firms	-0.055**	-0.054**
R&D spillovers along the value chain from non high growth firms	-0.00001	-0.0001
R&D spillovers from high growth firms	0.00004*	0.00004*
R&D spillovers from non high growth neighbours	-0.00001	-0.00001
Concentration index	0.6299	0.342
Regulation	1.7799***	1.991***
Environmental regulation	-0.5272	-0.407**
Foreign dummy	0.1271***	0.271***
Predicted R&D	0.7083***	0.831***
Age	0.0168***	0.034***
Collaboration: Internal	-	-0.139
Collaboration: Suppliers	-	0.241
Collaboration: Competitors	-	-0.982**
Collaboration: Universities	-	-0.518
Sources of Information: Internal	0.4811***	0.418***
Sources of Information: Suppliers	0.3564***	0.319***
Sources of Information: Competitors	0.0942	0.05

Sources of Information: Universities	0.39***	-0.318***
High tech dummy	-0.9732*	-0.798*
Correlation Coefficient	0.608***	0.678***
N	1208	

Notes: Marginal effects in the table. Observations are weighted to give nationally representative results. t-ratios are computed by using standard errors that are clustered around the firm. Industry, size and time effects are controlled for but the estimates are not reported. Source: ONS.

Table 3. *Production Function – High-Growth Firms*

Value Added	
Labour	0.7***
Capital	0.17***
Age	0.013**
Product Innovation (predicted value)	0.71***
Process Innovation (predicted value)	0.65***
Adjusted R-squared	0.71
N	1208

Notes: Observations are weighted to give nationally representative results. t-ratios are computed by using standard errors that are clustered around the firm. Industry, size and time effects are controlled for but the estimates are not reported. Source: ONS.

Table 4. *Innovation Production Functions with interactions.*

	Process Innovation	***Product Innovation***	***Process Innovation***	***Product Innovation***
	Coeff.	***Coeff.***	***Coeff.***	***Coeff.***
HGFs (1/0)	0.037	0.092	0.0006	0.0308**
Concentration index	-2.45**	-1.53**	-2.42**	-1.47
Regulation	0.49	0.013	0.46	0.010
Environmental regulation	0.082	-0.96**	0.10	-0.95**
Predicted R&D	0.97***	1.55***	0.97***	1.55
Foreign	0.090**	0.003	0.91**	0.004
Age	0.027***	0.040***	0.027***	0.040***
Collaboration: Internal	-0.174	-0.17	0.152	-0.091
Collaboration: Suppliers	0.228	-0.43	0.30	-0.32
Collaboration: Competitors	-0.493	0.01	-0.44	-0.134
Collaboration: Universities	-0.264	-0.047	-0.36	-0.103
Source of Information: Internal	0.43***	0.54***	0.39***	0.53***
Source of Information: Suppliers	1.15***	0.93***	1.17***	0.97***
Source of Information: Competitors	-0.437***	-0.32***	-0.43***	-0.30**
Source of Information: Higher Education	-0.372***	-0.52***	-0.35***	-0.524***
Collaboration: Internal*HGF	-0.138	0.615	-	-
Collaboration: Suppliers*HGF	0.29	0.56	-	-
Collaboration: Competitors*HGF	0.315	-1.23	-	-
Collaboration: Universities*HGF	-0.040	-0.011	-	-
Source of Information: Internal*HGF	-	-	0.34**	0.085
Source of Information: Suppliers*HGF	-	-	-0.17	-0.289
Source of Information: Competitors*HGF	-	-	-0.029	-0.64
Source of Information: Higher	-	-	-0.125	0.00032

Education*HGF				
High tech firm	0.96***	1.00***	0.97***	0.99***
Correlation coefficient	0.62**		0.62***	
N	24128		24128	

Notes: Marginal effects in the table. Observations are weighted to give nationally representative results. t-ratios are computed by using standard errors that are clustered around the firm. t-ratios for the interaction terms have been computed by using standard errors computed according to the Ai and Norton formula. Industry, size and time effects are controlled for but the estimates are not reported. Source: ONS.

Table 5. *Production functions*

	Model 1	***Model 2***
Value Added	***Coeff.***	***Coeff.***
Labour	0.69***	0.65***
Capital	0.17***	0.20***
Age	0.010**	0.010**
Product Innovation (predicted value)	0.59***	0.58***
Process Innovation (predicted value)	0.45***	0.48***
High growth dummy (1/0)	0.14***	0.13***
Product Innovation * High growth dummy	0.12***	0.15***
Process Innovation * High growth dummy	0.10***	0.09***
Adjusted R-squared	0.67	0.67
N	24128	24128

Notes: Observations are weighted to give nationally representative results. Industry and time dummies are introduced among the regressors in both equations. t-ratios are computed with standard errors that are clustered around the firm. Source: ONS.

Table 6. *Innovation production functions - Non high growth firms*

Process Innovation	***Coeff.***	***Coeff.***
Spillovers from patenting activities of the high-growth firms	-0.009	0.005
R&D spillovers from neighbouring high-growth firms	0	0
R&D spillovers along the value chain from all firms	0.0004	0.00038
Concentration index	0.063	0.092
Regulation	0.9568***	1.476***
Environmental regulation	-0.507**	-0.68***
Foreign	0.043*	0.010*
Predicted R&D	0.3362***	0.623***
Age	0.007***	0.010***
Collaboration: Internal	-	0.1816
Collaboration: Suppliers	-	0.3127**
Collaboration: Competitors	-	-0.419*
Collaboration: Universities	-	-0.52**
Source of Information: Internal	0.362***	0.38***
Source of Information: Suppliers	0.7660***	0.83***
Source of Information: Competitors	-0.062*	-0.11***
Source of Information: Higher Education	-0.281***	-0.32***
High tech firm	0.315*	0.25
Product Innovation		
Spillovers from the patenting activities of the high-growth firms	0.018	0.028
R&D spillovers from neighbouring high-growth firms	0	0
R&D spillovers along the value chain from non high-growth firms	0.0004	0.0003
Concentration index	0.055	0.089
Regulation	1.246***	1.658***
Environmental regulation	-0.410**	- 0.565***
Predicted R&D	0.6862***	0.922
Foreign	0.085**	0.13***
Age	0.013***	0.020***
Collaboration: Internal	-	0.009
Collaboration: Suppliers	-	0.092
Collaboration: Customers	-	-
Collaboration: Competitors	-	-0.410*
Collaboration: Universities	-	- 0.818***
Source of Information: Internal	0.392***	0.418***
Source of Information: Suppliers	0.402***	0.475***
Source of Information: Customers	-	-
Source of Information: Competitors	0.035	0.034

Source of Information: Universities	-0.391***	-0.42***
High-tech firm	-0.26	-0.29
Correlation Coefficient	0.61***	0.643***
N	22920	

Notes: Marginal effects in the table. Observations are weighted to give nationally representative results. t-ratios are computed by using standard errors that are clustered around the firm. Industry, size and time effects are controlled for but the estimates are not reported. Source: ONS.

Table 7. *Descriptive Statistics for Intangible Assets*

	Non-HGFs	**HGFs**	**t-test on the difference between the means**
Investment in Training (1/0)	44% (n=601)	54% (n=68)	Not significant
Investment in Software (1/0)	38% (n=601)	42% (n=68)	Not significant
Investment in Reputation (1/0)	25% (n=601)	25% (n=68)	Not significant
Investment in R&D (1/0)	12% (n=601)	10% (n=68)	Not significant
Investment in Design (1/0)	10% (n=601)	11% (n=68)	Not significant
Investment in Improvements (1/0)	15% (n=601)	19% (n=68)	Not significant
Investment in Training (Expenditure over employees)	0.51 (n=205)	0.59 (n=27)	Not significant
Investment in Software (Expenditure over employees)	0.97 (n=158)	1.10 (n=18)	Not significant
Investment in Reputation (Expenditure over employees)	2.38 (n=119)	1.64 (n=15)	Not significant
Investment in R&D (Expenditure over employees)	4.53 (n=59)	Nd	Not significant
Investment in Design (Expenditure over employees)	1.09 (n=54)	Nd	Not significant
Investment in Improvements (Expenditure over employees)	0.71 (n=69)	Nd	Not significant

Notes: HG stands for high-growth while NHG stands for non high-growth firms. The cells indicate the percentage of non high-growth firms/high growth firms with the relative attribute (for instance, investment in training). Source: Authors' calculations based on the BSD-IAS combined dataset.

Table 8. *Innovation production functions- All firms*

Process Innovation	***Coeff.***
Investment in training (1/0)*HGF dummy	0.17**
Investment in software (1/0)* HGF dummy	-0.25**
Investment in training (1/0)	-0.00073*
Investment in software (1/0)	0.12
Concentration index	0.75
Regulation	-4.21
Environmental regulation	0.36
HG Firm (1/0)	0.46*
Product Innovation	***Coeff.***
Investment in training (1/0)*HGF dummy	0.21**
Investment in software (1/0)* HGF dummy	0.95**
Investment in training (1/0)	0.15**
Investment in software (1/0)	0.16**
Concentration index	-3.53***
Regulation	-8.39**
Environmental regulation	6.25
HG Firm (1/0)	0.81**
N	145

Notes: Marginal effects in the table. The coefficient associated to the interaction terms are calculated by using the formula of Norton and Ai. t-ratios are computed by using standard errors that are clustered around the firm. Industry and size effects are controlled for but the estimates are not reported. The two equations have not been estimated in a simultaneous way and therefore there is no correlation coefficient. Source: ONS.

Appendix A: Construction of the dataset

For this study we have constructed an ad hoc dataset by using the following five components available at the ONS Virtual Microdata Laboratory. These are all linked by the unique reference number:

Business Structure Database (BSD): the dataset is derived from the Inter Departmental Business Register (IDBR) and provides longitudinal business demography information for the population of businesses in the UK.

Annual Respondents Database (ARD2): the ARD2 is constructed from the microdata collected in the Annual Business Inquiry (ABI) conducted by the ONS. The ARD covers both the production (including manufacturing) and the non-production sector (services). However, the time series dimension varies across the two sectors; while for the production sector it is possible to have information available up to 1980 (and early 70s for some industries), the data for the services sector is available only after 1997. The information is assembled from the replies to the Census forms; as this is a mandatory requirement for UK-based business, the response rates to the ARD are rather high and this makes it highly representative of the underlying population. Each establishment has got a unique reference number that does not change over time and so allows us to build up a panel dataset. The ARD is a stratified random sample where sampling probabilities are higher for large establishments; indeed for establishments with more than 250 employees, the sampling probability is equal to one. The ARD contains all the basic information (namely the inputs and output variables) needed to estimate the production function. Output is measured by the deflated added value. Employment is measured by the total number of employees. As for capital, it is well known that the ARD does not contain information on capital stock. However, stock of capital has been constructed at the ONS by using the perpetual inventory method.

UK Community Innovation Survey (CIS) 3, 4, 5 and 6: the CIS is a stratified sample of firms with more than 10 employees drawn from the IDBR. The CIS contains detailed information on firms' self-reported innovative activities. This covers firms' innovation activities over a three-year window, targeting firms with more than ten employees. The CIS is a survey carried out by national statistical agencies in all 25 EU member states under the coordination of Eurostat. The sampling frame for the UK CIS was developed from the Interdepartmental Business Register (IDBR) with the survey being conducted by post. Weights were used to make the sample representative of the British services sector. Firms are asked whether they have produced any innovation in the reference period (i.e. the three years before the survey starts) and if so, what type of innovation they have introduced. In turn innovation can be of three types: product innovation, process innovation and wider (or organisational) innovation. The CIS provides information on what external sources of information a firm uses and whether it collaborates with other companies, suppliers, customers, competitors, laboratories and universities to develop innovation. For the main analysis, we use three surveys: CIS 3 which covers the period 1998-2000, CIS 4 which covers 2002-2004, and CIS 5 which covers 2004-2006. For the analysis on the intangible assets, we use the CIS 6 which covers the period 2007-2009.

Patent data: we use a match of UK patents obtained from Optics and EPO patents (designating the UK and obtained from EPO's Patstat database, version April 2010) with the IDBR. The patents-IDBR match was carried out by the ONS/ UKIPO using firms' names as patent documents lack unique firm identifiers.

Appendix B

This is the list of independent variables used in our empirical analysis (it includes the variables used to model both the investment in R&D and the innovation production functions):

- *Information Sources*: this is set of categorical variables reflecting different sources of information for innovation. These take the value of 1 if information from internal sources (customers/ suppliers/ competitors/ universities) was of high and medium importance.

- *High-tech indicator*: this is a dummy variable that takes the value of 1 for firms that belong to the high-tech sectors (according to the OECD definition[2]).

- *Foreign ownership indicator:* this is a dummy variable that takes the value of 1 if the firm is foreign-owned.

- *Size*: we use the following size categories: 2-49, 50-99, 100-249, 250-999, >1000 employees.

- *Age*: this is measured by the number of years the firm has been active.

- *Demand-pull factors*: these are proxied by: a) the share of firms in the three-digit industry for which meeting regulations or standards is of high, medium, or low importance for innovation (as opposed to no importance)[3] and b) the share of firms in three-digit industry according to which environmental concerns were of high, medium or low importance for innovation (as opposed to no importance).

- *Collaboration partners*: this is a set of categorical variables taking the value of 1 if collaboration with internal departments (or customers/ or suppliers/or competitors/or universities) is reported to be of high and medium importance.

- *Knowledge spillovers:* we focus on the knowledge spillovers generated by: a) the patenting activities of the neighbouring firms (measured as the number of

[2] The OECD definition of high tech is the following: pharmaceuticals SIC 2423; aircraft & spacecraft SIC 353; medical, precision & optimal instruments SIC 33; radio, television & communication equipment SIC 32; office, accounting & computing machinery SIC 30.

[3] Note that because we also include 2-digit industry dummies in the regressions, the demand pull effects are measured relative to the average for the relevant industry.

patents' application by postcode weighted by distance between postcodes); b) the firms' investment in R&D in the upstream sectors (measured as the total stock of R&D performed by firms weighted by the coefficients of the input-output table) and c) the R&D investment of firms in the same postcode area (measured again as the stock of R&D of firms weighted by the distance between postcodes).