Preliminary, please do not quote. Comments welcome.

Innovations and Productivity Growth in the UK: Evidence from CIS2 and CIS3*

Chiara Criscuolo

University College London and Centre for Research into Business Activity (CeRiBA),

Jonathan Haskel

Queen Mary, University of London, Centre for Research into Business Activity (CeRiBA) and CEPR

Abstract

We use matched innovation survey and Census data to investigate the link between innovation and productivity growth and the factors that are linked with high innovation. Our results are very preliminary but in our sample we have significant effects from (a) co-operating with other firms on technological change and (b) in CIS3 obtaining information from suppliers, customers and competitors. This suggests that policymakers might examine the conditions under which firms can and cannot co-operate on innovation projects. Regarding innovation and productivity growth we have rather mixed results with process and product innovations being important in CIS2 and CIS3 respectively.

---

* Contact: Chiara Criscuolo, CeRiBA, Centre for Research Into Business Activity, Office for National Statistics, Zone D4-16, 1 Drummond Gate, London SW1V 2QQ, Email: chiara.criscuolo@ons.gov.uk. This paper presents preliminary results, comments welcome. Work on this paper has been funded by a grant from the Evidence Based Policy Fund supported by HM Treasury, DTI and ONS, and carried out at CeRiBA at the Business Data Linking Branch at the ONS; we are grateful to all institutions concerned for their support. Results in this paper use the ONS ARD data set which is a research tool and may not entirely match official published figures.

CeRiBA Discussion Paper

Centre for Research into Business Activity –www.CeRiBA.org.uk

1 Introduction

This paper explores two questions:

a) What is the link between innovation and productivity growth?

b) What factors are linked with high innovation?

To approach (a), many studies look at correlations between productivity growth and R&D and / or patents. The strengths and weaknesses of this are well known; measures of R&D are reasonably well codified but R&D is an input to the innovation process and not an output. Also, firms might generate technological advance outside formal R&D laboratories which R&D expenditure might not capture. Patents are an output of the innovation process, but by no means all innovations are patented and there are great differences in the relative importance of patenting as a barrier to imitation both between sectors and among different type of innovations. The approach to question (b) is to try to study the incentives to undertake innovative activity. Many studies regress R&D or patents on size, measures of concentration etc. (see e.g. Griliches, 1998 for a survey).

To attempt to overcome these problems, the OECD has developed guidelines for company surveys that measure innovations directly. Such surveys set out a definition of innovations and ask companies to report the output of the innovation process (introduction of innovative new products, new processes, percentage of sales arising from new and improved products, and "soft" innovations, such as organisational change), the inputs to innovations (R&D, scientists, sources of knowledge) and the obstacles to innovation (finance, bad luck etc.). The Oslo manual (1992) codifies such survey models and the Community Innovation Survey (CIS), carried out in EU countries in the early, mid and late 1990s implemented the questions. For the UK there have been three CIS surveys, CIS1 (covering the period 1991-3), CIS2 (1994-6) and CIS3 (1998-2000).

A significant limitation of the CIS questionnaire is that it has no productivity/TFP information on it. Thus, one cannot examine the relation between the various innovation measures and productivity without linking the CIS survey with other data sources which can provide information on productivity. This could allow analyses of obvious interest since it would quantify the performance consequences of the innovation processes that the CIS describes. In this paper we implement this linking for the UK. The UK CIS was carried out using the same sampling frame as the UK Census of Production. We have therefore obtained the raw CIS2 and CIS3[1] and corresponding Census data and matched them. The Census provides a wealth of information on output, employment, material use, capital etc. Thus the matching allows us to compare productivity/TFP from the Census with innovations from the CIS. To the best of our knowledge we are the first ones to use this matched information for the UK.[2]

Since these data have not been analysed before our main objective in this paper is to understand better the data. The few other papers using data from other countries have developed systems of equations and adopted instrumental variable methods to attempt to uncover causal relations in the data. We eschew this approach in this paper in favour of simply trying to understand the data set and some key correlations. For a framework to examine correlations, we set out an output production function to look at innovation and productivity growth and an ideas production function to look at the determinants of innovation.

The remainder of this paper is as follows. In section 2 we describe the data. Section 3 sets out an organising framework for the econometric results that we report in section 4. Section 5 concludes.

[1] We cannot use CIS1 since it had a very low response rate.

[2] Harris (2001) has matched the UK CIS with the UK Census of production, but has not analysed the matched data. Examples of matched CIS/Census data are, for France, Crepon, Duguet, Mairesse (2000); for Holland, Klomp and van Leeuwen (2002) ; for Sweden, Lööf and Heshmati (2001); and for Finland, Leiponen (2002).

2 Data

2.1 The ARD

The ONS Annual Respondents Database (ARD) is described in some detail in Haskel and Martin (2002), so only a brief description is included here. The ARD consists of the micro data from the *Annual Census of Production* (ACOP) up to 1997 and the *Annual Business Inquiry* (ABI) thereafter. The micro data are the replies to the Census forms, response to which is mandatory under the 1947 Statistics of Trade Act. These forms request information on inputs and outputs. Information is also collected on plants' industry, region, and nationality of ownership. Each unit who replies is assigned a unique identification number, which allows units to be linked over time into a panel. Units also have another identification number corresponding to the entity that owns them (the firm) so units under common ownership share the same firm identifier.

The surveys are drawn from the current business register held by the ONS. The ACOP until 1994 was drawn from a register maintained by the Office for National Statistics (ONS), in 1994 this register was re-organised and combined with others to become the *Inter Departmental Business Register* (IDBR). Both stages of the register make use of VAT records, historical information and other surveys including commercial data (such as Dun and Bradstreet). The IDBR additionally uses PAYE tax data.

To limit reporting burdens on businesses, the ACOP / ABI are both stratified sample surveys. The sampling works as follows. All larger enterprises over a threshold number of employees are surveyed, but a sample is taken of smaller enterprises (with the sampling rules changing every so often, Oulton, 1997, Barnes and Martin, 2002). These surveyed businesses form what is called the "selected" sample and they account for over 80% of total employment in manufacturing (Oulton, 1997). The rest of the units on the register are not sampled (the "non-selected" sample), and their information on industrial classification, region and employment comes from the business register. In addition to those units that were "non-selected" for the survey, some units that did not respond also have only register data available. Register employment information comes from separate inquiries and may, for firms, below 10 employees, be imputed from turnover data (Perry 1995).

A decision has to be made as to what level of aggregation to work at. Surveys are conducted at the "reporting unit" level. This may be an individual "local unit", where a "local unit" (in the manufacturing survey) is a production facility at a single mailing address, which corresponds to a "production unit" or "plant" (in retailing for example it would be a shop). However, in multi-plant or multi local unit firms, the "reporting unit" may be a group of local units, where the grouping is agreed by the ONS and the firm (along similar product lines for example). As data are not collected at the local unit level in the UK, a choice has to be made between estimating local unit values from those for the reporting unit they are a part of or using data at the reporting unit level. This work is at the reporting unit level.

There are a number of important data issues that arise. First, with the introduction of the IDBR in between 1993 and 1994, a complete recoding of local unit, reporting unit and enterprise identification numbers was undertaken. To minimise measurement error we use a cleaned[3] dataset that excludes all unrealistic growth rates for output and inputs of the production process.

The second issue is the capital stock. The Census does not ask plants to report capital stocks, so as in previous work we used plant investment data to calculate capital stocks. We chose industry-level starting capital-stock values and depreciation rates for buildings, plant and machinery, and vehicles and deflated each component of investment by ONS industry-year investment deflators. See Martin (2000) for more details.

2.2 The Community Innovation Survey

The Community Innovation Survey (CIS) is a voluntary postal survey carried out by ONS on behalf of the DTI. Eurostat proposes an initial questionnaire and the DTI adds questions. ONS randomly selects a stratified sample of firms with more than 10 employees drawn from the Inter-Departmental Business Register (IDBR) by SIC92 2-digit class and 8 employment size bands. The IDBR excludes agriculture, fishing and forestry, public administration and defence, education, health and social work. The survey covers both the production (manufacturing; mining; electricity, gas and water; construction) and the service sectors.

There are of course a host of issues regarding the sampling procedure, composition of the questionnaire, and matching procedure. Starting first with sampling, the CIS is a voluntary

[3] Details of the cleaning procedure are given in the appendix.

and postal survey. One of the main problems for this sort of surveys is the risks of low-response and non-response bias. To boost response, enterprises are sent the survey, posted a reminder, posted a second reminder (with the survey again) and finally telephoned.

A second problem is that the survey was conducted at the enterprise level; where enterprise is defined as "the smallest combinations of legal units which have a certain degree of autonomy within an enterprise group". This corresponds neatly with the reporting unit level at which the Census is carried out thus facilitating matching.

The third issue is that the answers to a survey are necessarily subjective (Tether, 2001). The essential idea of the survey is to try to get enterprises to report separately technological change/innovations as opposed to organisational innovations. In turn, technological innovations are split into process and product innovations. Since companies are asked about products or processes that are "technologically new" there is obvious scope for differences in interpretation of "technological" and "new". The questionnaire does however give extensive guidance on both these terms, but there is always the problem of misinterpretation or indeed reporting yes to all questions for fear of a company being seen as in some way backward. All this introduces measurement error into the level of innovations, biasing us against finding a significant relation between productivity growth and innovations if firms are randomly mis-reporting or biasing down the expected relation if firms report positively innovations when they are not in fact innovating.

2.2.1 The Second Community Innovation Survey (CIS2)

Fieldwork for the CIS2 survey took place between August 1997 and March 1998 and firms were asked to complete data referring to the period 1994 to 1996. The sample and response rates are set out in Table 1.

Table 1
CIS2 Sample and response rates

	Manufacturing	Services	Total
Enterprises in population covered by survey	61,250	94,000	155,520
Enterprise Sampled	3,906	1,986	5892
Enterprises answering the Survey	1,596	743	2339
Response rate (%)	41	37	40

Source: Craggs and Jones, and authors' calculations.

At the time of selection, the IDBR had details of around 155,000 enterprises (column 3, top row). 5,892 were sampled and 2,339 responded; accounting for bankruptcy, this was a response rate of 43.2%. Manufacturers and service sector companies were sent very slightly different questionnaires[4]; replies and response rates were 1,596 and 743, 41% and 37% respectively.

CIS2 data custodians report finding no bias in favour of innovators in the returned sample based on a random sample of 317 non-responders (Craggs and Jones, 1998, p.57). To explore this a bit further, we used the matched survey universe (i.e. the firms chosen to be surveyed regardless of whether they reply or not). The results are set out in Table 2, which shows that non-respondents have on average significantly larger turnover and employment than respondents.

Table 2

Characteristics of respondents and non-respondents in CIS2

	Manufacturing	Services	Total
Mean employment respondents	376	932	553
Mean employment non- respondents	535	1,333	812
Mean turnover respondents	46,114	232,690	105,780
Mean turnover non-respondents	79,812	978,629	392,994

Source: Authors' calculations.

Table 3, column 1, sets out some details of the CIS2. As the top row, first column, shows, we have data on 2,339 reporting units. We cannot analyse the service sector at present[5], this leaves us with a sample of 1,596 companies. Also we cannot include in the analysis the non-manufacturing sectors of productions (i.e. electricity, gas and water; construction) and the manufacturing of coke, refined petroleum products and nuclear fuel (Sector 23 of the 2-digit ISIC92 classification) this leaves us with a sample 1,453 reporting units.

[4] The main difference is the definition of innovation.

[5] Service sector data is available in the ARD since 1996 onwards.

Table 3

CIS 2 and CIS3 reporting unit profiles

		CIS 2	CIS3
1	Number of Reporting Units	2,339	8,172
2	Number of Reporting Units in Services	743	3, 605
3	Number of Reporting Units in production	1,596	4, 567
4	Number of Reporting Units in Manufacturing excluding sector 23	1,453	3, 425
	Panel		
5	Number of reporting units in both surveys	787	
6	Number of reporting units in manufacturing in CIS2 and in CIS3	509	

Note: *selected means that the reporting units are in the ARD sample with full Census information.
Source: Authors' calculations.

2.2.2 The Third Community Innovation Survey (CIS 3)

The Third Community Innovation Survey (CIS 3) was in the field twice. The first wave sampled 13,340 enterprises, the second top-up covered 6,285 to make the sample representative at the regional level. The CIS 3 covers the period 1998-2000. Of the total 19,625 enterprises to which the survey was sent, 8,172 responded (Table 3, column 2), achieving a response rate of 42%. We are in the process of investigating respondents characteristics.

2.3 Matching the CIS 2 and CIS 3

One of our aims is to construct an innovation panel, merging the CIS2 and CIS3. The results of this exercise are reported in table 3. 787 enterprises are in both surveys. Some of them are recorded in the manufacturing sector in CIS2 and in the service sector in CIS3. Since in CIS3 the survey is the same for both sector we decided to consider these enterprises as being part of the manufacturing sector if they are in the manufacturing sector according to the ARD. We exclude from the present analysis those enterprises that are in the service sector in the CIS2 and CIS3, or in CIS2 alone. The main reason for this choice is that the CIS2 questionnaires differ between service and production sector and at present we are only analysing the manufacturing sector.

What information can we get from the panel? Although the gist of questionnaire for CIS2 and CIS3 is the same , the two questionnaires differ in several respects, so that the construction of an innovation panel needs some caution[6].

First, the main question regarding product innovation differ. In CIS2 the question was: “Has your enterprise introduced onto the market any technologically new or improved products?”. In CIS3, the question reads “did your enterprise introduce any technologically new or significantly improved products (goods or services) *which were new to your firm*?”.

Secondly, the CIS 3 questionnaire is the same for both production and service sector, and the firms are asked directly about their main product (CIS2 had two different questionnaires for services and production).

Thirdly, there are no filter questions in CIS3. The issue here is as follows. In CIS2 companies can skip a part of the questionnaire by declaring that they have not engaged in any innovative activity and do not have any intention to start innovative projects in the next five years. The answers to this “filter question” show that 21.7% of respondents (508 firms, of which 316 in the production sector and 192 in services) have not engaged in any innovative activity and do not have any intention to start innovative projects in the next five years[7].

It is worth noting however that the UK version of the CIS presents a notable advantage in respect to the surveys carried out in other countries, e.g. France; Germany and Finland. The UK questionnaire allows the collection of information on innovative activity also on companies which have not been successful innovators during the period 1994-96, but had been engaged, successfully or not, in the previous ten years or have projects to innovate in the next five years. In the other countries no further information is available for non-innovating firms.

Fourth, the ordering and the general layout/editing of the questions in the survey differs. This matters if we believe that respondents get tired of answering the questionnaire and become less attentive towards the end of the questionnaire.

---

[6] See also Frenz (2002) .

Lastly, the question regarding public support for innovation in CIS3 distinguishes between the source of the support (regional, central or European government), whereas in CIS2 the question entails a yes/no answer on whether the enterprise has received any central government financial support. Also, the classification of innovation-related public programs slightly differs among the two surveys.

2.4 Matching the CIS2 and ARD

Turning to the matching issues, businesses are asked to provide reporting unit level information for the CIS and we therefore matched at the reporting unit level, merging the ARD panel for the 1991-1999 period to the CIS2.

Table 4

CIS 2 and ARD

		CIS 2	Successfully merged with ARD 1991-1999
1	Number of Reporting Units	2,339	1,502
2	Number of Reporting Units in Service	743	30
3	Number of Reporting Units in production	1,596	1,472
4	Number of Reporting Units in Manufacturing excluding sector 23	1,453	1,443
5	Number of Reporting Units in Manufacturing (excl. sec23) selected* at least once between 1991-1999		1,344
6	Number of Reporting Units in Manufacturing (excl. sec23) selected at least twice between 1991-1999		707
7	Number of reporting units in the cleaned sample		578

Note: *selected means that the reporting units are in the ARD sample with full Census information.

Table 4 shows the results of the matching between CIS2 and the ARD Panel. The first row shows that of the 2,339 reporting units in CIS2, we could potentially match 1,502 reporting units, 1,472 in production and 30 in services. We decided to drop those observations that were in the ARD manufacturing panel but were coded in the service sector by the CIS2. Row 4 shows the number of successful matched reporting units that are in the manufacturing sector excluding the manufacturing of coke, refined petroleum products and nuclear fuel. Only for a subset of these 1,443 successful matches we are able to build a productivity

[7] Tether notes that perhaps this could partly explain the high presence of non-innovators in high-technology sectors.

growth profile. Only 1,344 were in the selected sample at least once. Row 7 shows that 707 were present in the selected file of the ARD at least twice during the period. When we look at the productivity growth/innovation relation we can only do so for firms that were selected at least twice during 1991-1999. Also, for some enterprises relevant information on output and inputs of the production function are missing or productivity growth rates are unbelievable; accounting for these problems leaves us with a sample of 578 observations.

2.5 Matching the CIS 3 and ARD

As we did for the CIS2 we matched at the reporting unit level, merging the ARD panel for the 1996-2000 period to the CIS3. We describe the results of the matching in table 5.

Table 5
CIS 3 and ARD

		CIS 3	Successfully merged with ARD
1	Number of Reporting Units	8,172	3,397
2	Number of Reporting Units in distribution and Services	3,605	98
3	Number of Reporting Units in production (Mining, manufacturing and Construction)	4,567	3,299
4	Number of Reporting Units in Manufacturing excluding sector 23	3,425	3,277
5	Number of Reporting Units in Manufacturing (excl. sec23) selected* in 2000		1,593
6	Number of Reporting Units in Manufacturing (excl. sec23) selected* in 1996 or 1997 or 1998 and in 2000		827
7	Number of Reporting Units in Manufacturing (excl. sec23) in the cleaned sample		716

Note: *selected means that the reporting units are in the ARD sample with full Census information.

Of the 8,172 enterprises which responded to the survey, 3,397 were successfully matched with the ARD manufacturing data. In 98 cases there was a discrepancy between the industrial classification in the Innovation survey and that of the Production survey. In these ambiguous cases, since the innovation survey is the same for both sectors, we decided to include these enterprises in the sample, using also the direct information from the enterprises available in CIS3 (“please briefly describe your enterprise’s main product “).

The number of reporting units that are in the manufacturing sector excluding sector 23 according to the ARD are 3,277 as shown in the second column of row 4. Row 5 shows that

1,593 were surveyed in 2000. As for the analysis using CIS2 we need longitudinal information to be able to draw growth profiles. We therefore report in rows 6 the number of reporting units that are selected in 2000 and in previous years. In the last row we report the number of firms for which we can construct an average annual growth rate after having cleaned the dataset. In the analysis we choose to use this cleaned sample.

3 Productivity growth and innovation.

3.1 Basic correlations.

Table 6 sets out the characteristics of our matched samples. Looking at the top row, the median firm in either CIS2 or CIS3 does no patenting and spends nothing on R&D. This is as expected, since R&D and patenting is so concentrated in a very small number of firms (Bloom and van Reenen, 2000, report that the 12 largest UK firms account for 72% of patenting and 80% of R&D expenditure). The median firm, in our sample at least, does report having performed some innovation expenditure (i.e. some spending on R&D, new machines related to innovation, or external consultants). Looking at the final two columns, in CIS2 the median firm reported non-zero sales due to new products and being a process innovator, although this was not the case in CIS3.

Two features of Table 6 are worth commenting on. First, whilst R&D and patents are skewed, innovations are rather less so, suggesting that our data does provide information on the outputs of the innovative process over and above R&D. Second, as is found in the overall CIS sample (recall we are using just that sample that we can match with our productivity growth data here), the innovation rate has dropped in CIS3 relative to CIS2 (in the last row for example, 54% of firms were process innovators in CIS2 but only 37% in CIS3).

Table 6

Characteristics of innovation measures in the regression samples

	CIS3			CIS2		
	Median in sample	Mean in sample	Mean for prod. inn.	Median in sample	Mean in sample	Mean for prod. inn.
Number of patents	0	3	5	0	4	6
R&D intensity	0	0.63%	1.34%	0	0.33%	0.47%
Total inn. expenditure	0.24%	2.35%	4.36%	1.82%	3.97%	5.31%
% sales new products	0	9.48%	24.07%	0.1	19.37%	30.17%
Process innovator	0	37.25%	56.59%	1	54.15%	70.08%

Source: authors' calculations.

.

Table 7 shows mean productivity growth for the firms reporting no and yes to the row questions in Table 6. Looking at the top row for example, mean productivity growth for firms reporting no patents is about the same as that for patenting firms in CIS2 but slightly lower in CIS3. For the innovation measures by contrast, the positive productivity growth relation shows up a bit more clearly: innovators are more likely to have faster productivity growth than non-innovators, especially for process innovators on CIS2.

Table 7

Productivity growth and innovation measures in our matched samples.

Productivity growth: D(VA/Emp)	CIS3		CIS2	
	no	yes	no	yes
Patentors	3.26%	4.99%	3.16%	3.24%
R&D doers	3.14%	4.86%	2.04%	10.12%
Spendors in innovation activity	3.26%	3.88%	1.20%	3.92%
Product innovators	2.91%	4.90%	2.62%	3.51%
Process innovators	3.19%	4.37%	0.94%	5.07%

Source: authors' calculations.

3.2 Innovations and productivity growth: econometric model.

We make no claim to contribute to the theory of innovation and productivity growth. The purpose of this section is merely to set out a framework, based on existing work, around which to organise the data. We suppose there is a conventional output production function which relates real physical output Y to real physical inputs X for a given state of knowledge K

$$Y_i = K_i X_i^{\ a} \qquad (1)$$

Changes in knowledge thus raise productivity growth given inputs X. Letting lower case letters denote logs and differencing gives that

$$Dy_i = Dk_i + a\,Dx_i \qquad (2)$$

Letting changes in the knowledge stock be measured by our innovations measures gives

$$D(y-l)_i = a_1 D(c-l)_i + a_2 D(m-l)_i + a_6 PdctInn + a_7 PrcsInn + l_I + e_i \qquad (3)$$

where changes in X are measured by the ARD data using data on employment, capital and materials[8] and λ is an industry dummy.

3.3 Factors affecting innovations.

Changes in knowledge *DK*, are to measured in our work by innovations. We suppose that just as output of physical goods arises from inputs, so the output of ideas arises from inputs. Thus changes in the knowledge stock are captured by the ideas, or innovation production function

$$DK = f(Z^{PHY}, Z^{KNOW})e \qquad (4)$$

In (2) Z^{PHY} are the physical inputs into the ideas process (i.e. the numbers of scientists, laboratories, test tubes, the efficiency of the ideas production organisation). Z^{KNOW} are the knowledge inputs into the ideas process (e.g. the flow of ideas generated from existing knowledge in the firm, or the flow of ideas from visiting a trade fair) and ε are any other factors affecting the conversion of ideas inputs into innovation output (e.g. luck, technological opportunity).

---

[8] We are unfortunately unable to distinguish between capital in physical production and that in ideas production and hence bias our results away from finding an effect of ΔK, to the extent that ΔK uses capital in ideas production (e.g. laboratories), except when R&D takes place in a separate physical location.

An important feature of knowledge inputs Z^{KNOW} are that they are potentially transferable across organisations (unlike a laboratory). Thus we may write the knowledge inputs as those originating from company *i* itself and those from outside company *i*

$$DK_i = f(Z^{PHY}, Z_i^{KNOW}, Z_{\_i}^{KNOW})e \qquad (5)$$

Knowledge inputs presumably derive from the existing knowledge stock. There is for example presumably a knowledge stock at a trade fair or in a university. The firm visiting the trade fair or co-operating with a university may or may not gain ideas. Equally, firms themselves presumably have a knowledge stock built up (the knowledge underlying previous innovations for example) from which new ideas flow. However, not all ideas are of equal importance and thus, as is well-known, we wish to measure not only the stock but the weight in the knowledge flow that that particular source has for a given firm (just as citation-weighted patents have been used). In terms of notation, we may substitute the Z^{KNOW} in (3) in terms of the knowledge stock times the flow of ideas from that stock

$$DK_i = f(Z^{PHY}, q_{1i}K_i, q_{2i}K_{\_i})e \qquad (6)$$

where K_i and $K_{\_i}$ are the knowledge stocks inside and outside the firm and θ_{1i} and θ_{2i} convert these stocks to flows of ideas to the firm.[9]

Before proceeding to measurement issues, (9) highlights the R&D question that we have previously raised. We know that a very small proportion of firms do R&D. Denote an element of qK_i in (9) the own R&D of the firm. Does this mean that the vast majority of firms who do no R&D do not innovate? Equation (9) highlights the reason why this is not the case: they presumably get information from elsewhere.

3.3.1 Measurement of the variables in (9)

Z^{PHY}, physical inputs to ideas production. We have a number of potential measures of this. We have data on acquisition of machinery and equipment, other external technology, industrial

[9] The θ weights may of course be endogenous; firm *i* with talented engineers for example may get more ideas from visiting a trade conference than firm *j* who has no such engineers, in which case $K_{\_i}$ is the same but $\theta_{2i} > \theta_{2j}$.

design and training all of which are linked to innovations.[10] Firms are also asked their expenditure on intra- and extra-mural R&D. For simplicity we combined all these expenditures together and expressed them as a fraction of sales.

$q_{1i}K_i$, $q_{2i}K_{-i}$, the flow of ideas generated from existing knowledge inside and outside the firm. Firms are asked "how important to your enterprise are the following as sources of information for new technological innovation projects or the completion of existing projects". A number of information sources are provided (sources within enterprise, fairs & exhibitions for example) and firms are asked to grade their importance on a scale of 0 to 3. Cassiman and Veuglers argue that this provides data on the flow of ideas. Consider the response to the question about trade fairs for example. There is presumably a stock of ideas at a trade fair, which is of course very hard to measure. We are interested however in the flow of ideas to firm i from that stock. A firm reporting trade fairs as an important source of information against one reporting them as being an unimportant source presumably captures the different flows of ideas from trade fairs accruing to the two firms, which is what we require.
Our precise measures are set out in the Data Appendix, but described briefly here.

To measure $q_{1i}K_i$ we use the reported importance of ideas from sources *within* the enterprise and other enterprises within the enterprise group.

To measure $q_{2i}K\_{i}$, we use the reported importance of ideas from various sources *outside* the enterprise. Since there are 18 sources we group them together for convenience as follows. First, vertical sources or market sources of information (from clients, customers, suppliers or competitors). Second, commercial sources of information (from consultancy enterprises, research association or commercial R&D laboratories), Third, universities; fourth free (from professional conferences, meetings, journals etc.). Fifth, regulatory (from health and safety regulations; product standards; environmental regulations), and finally government (from government institutes; training and enterprise councils; Business links).

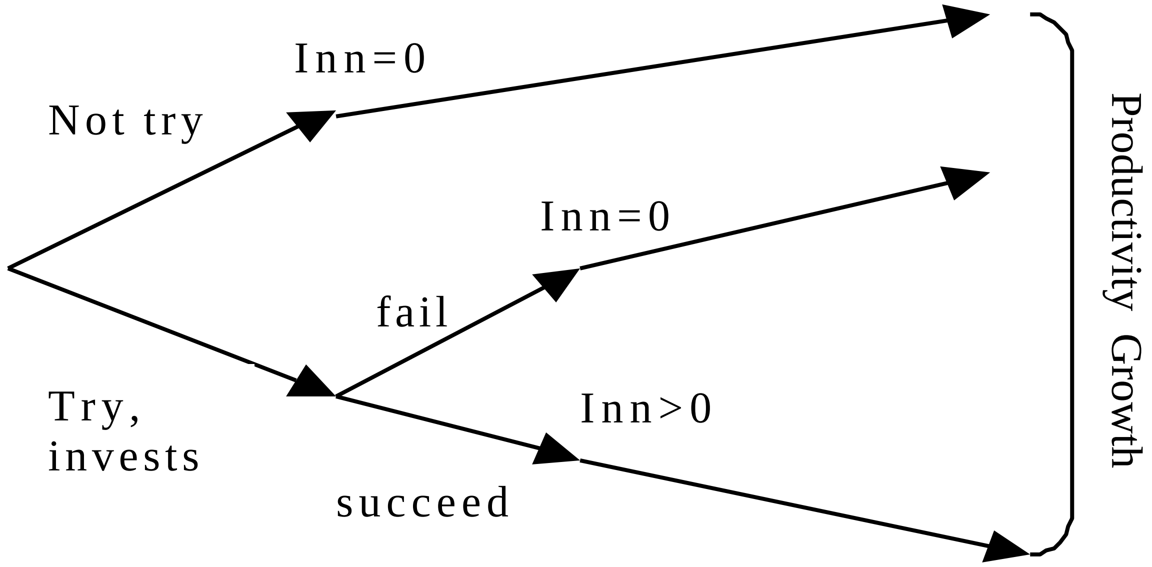
[10] Firms are instructed to confine machinery acquisition for example to that directly related to the innovation process rather then, say, a computer that uses an innovation (the microchip).

e, other factors. Finally, to measure random factors such as luck, we add a random error term. We also add industry dummies to proxy differing technological opportunity that might affect the flow of new ideas.

3.4 Methodology

Before proceeding, a word on econometric method. This is informed by the following diagram.

Figure 1: An innovation decision tree



Modelling stages

Choice	IPF	OPF
$Z^{PHY} = f(y)$	$DK_i = f(Z^{PHY}, Z_i^{KNOW}, Z_{\_i}^{KNOW})e$	$DY = DKf(DX)$

Note: IPF and OPF are ideas and output production function respectively.

The firm illustrated above starts by deciding whether to try innovating or not (the decision to try being, for example, whether to invest in innovation inputs such as R&D). Of those firms investing some succeed and some fail. All of the firms subsequently have productivity growth. Below the diagram we show the technological processes involved. At the first stage of the process, we might model the sorting of firms into those trying (i.e. undertaking innovation investment) and those not. At the second stage, we might model the relation between innovation outputs and inputs by an innovation production function. At the final stage we model the relation between innovation and productivity by an output production function.

A number of points are worth making. First, the decision whether to invest or not in inputs to innovation is presumably endogenous. Hence the modelling of the innovation production function in the second stage is for a sample who may or may not succeed in innovating, but have self-selected into a group who have at least put some inputs into innovating. Ignoring this selection process might be problematic. The provision of free information by government, or competition, for example might make innovation more likely to succeed once the decision has been made to try, but they might also make trying more likely. To identify this selection process convincingly, one needs variables that affect the probability of trying but not the probability of success (conditional on inputs) conditional on trying. An R&D tax credit for example, might affect the probability of trying, but not of success controlling for the amount of R&D spent. On the other hand, if talent of the firm is a determinant of both success and trying and more talented firms get the grants, then this might not be a good selection variable.

Second, as explained above in the UK firms are asked to fill in their innovation inputs whether they succeed or fail in innovation. In other countries firms who do not succeed do not fill in their innovation inputs. Hence a reading of zero innovation in those surveys does not distinguish between those firms who have tried but not succeeded and those who did not try in the first place. Thus our selection problem is different.

Third, as well as the selection process involved, the diagram shows some of the possible endogeneity issues. For example, if unobservably better firms (better managed for example) both try to innovate more readily, are more likely to succeed and have higher productivity growth, then this might influence our correlations.

At this stage of our work we do not investigate all three stages. Rather we investigate the IPF and the OPF stages. This will help us to understand the basic correlations in the data before moving to a more structural approach. To estimate the OPF we use OLS. To estimate the IPF we use only the firms who tried to innovate. Thus our regressors should be understood as the partial effects on innovation, conditional upon trying to innovate. If there is a positive correlation between the inputs to innovation and trying to innovate, our regressors will underestimate the total effect of the input on innovation. We also need to account for the fact that the innovation data is 0/1 for process innovation and 0 and less than or equal to 1 for product innovation. Thus we suppose that the propensity of an establishment to innovate in various ways can be expressed as a function of the variables in (4).

$$INNOV_i^* = b_1 Z_i^{PHY} + b_2(q_{1i}K_i)_i + b_3(q_{2i}K_{\_i})_i + l_I + e_i \tag{7}$$

Since process innovation is binary, if we further assume that

$$\begin{aligned} PrcsInn_i &= 0 \qquad if\ INNOV^* < 0 \\ PrcsInn_i &= 1 \qquad if\ INNOV^* \geq 0 \end{aligned} \tag{8}$$

We can then estimate the β parameters in (4) for process innovation using probit or logit. For product innovation, we use the fraction of sales reported by the firm accounted for by sales of the new product innovation. This varies between 0 and 1, with a number of firms reporting 0. Hence we use a tobit estimator.

4 Results

Our objective in this section is to understand some robust correlations in the data. Thus we look for consistency of results across CIS2 and CIS3. Questions of causation cannot of course be answered in these results.

4.1 Estimates of the output production function

The results of estimating (2) are set out in Table 8. Columns 1 and 2 shows the results using the CIS2 and CIS3 respectively. The results are somewhat uneven; ProcInn raises productivity growth positively and significantly in CIS2, but not in CIS3, where as ProdInn raises productivity growth positively and significantly in CIS3, but not in CIS2.

Table 8

Productivity Growth Regressions

	(1)	(2)	(3)	(4)
	Delta(GO/L)	Delta(VA/L)	Delta(GO/L)	Delta(VA/L)
	CIS3		CIS2	
Delta(M/L)	0.4226		0.3205	
	(0.0457)***		(0.1004)***	
Delta(K/L)	0.1202	0.2074	0.1379	0.2105
	(0.0351)***	(0.0709)***	(0.0421)***	(0.0675)***
product innovations	0.0673	0.1008	-0.0174	-0.0367
	(0.0287)**	(0.0525)*	(0.0178)	(0.0333)
process innovations	-0.0064	-0.0016	0.0223	0.0363
	(0.0075)	(0.0155)	(0.0082)***	(0.0171)**
Observations	649	642	568	566
R-squared	0.49	0.11	0.43	0.11

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

4.2 Estimates of the ideas production function

Table 9 sets out the results of estimating probits and tobits for ProcInn and ProdInn on CIS2 and CIS3. Based on the variables that are consistently significant we conclude that:

(a) Co-operation (an element of Z_i^{KNOW}) is positively correlated with innovation in all cases and significantly so.

(b) knowledge flows from internal sources such as the plant, other plants in the firm etc (an element of Z_i^{KNOW}) is positively correlated with innovation in all cases, strongly significant in CIS3 and significant for process innovation in CIS2.

(c) Expenditure on innovation (an element of Z_i^{PHY}) is positively correlated with innovation in all cases and significantly so, except for ProdInn in CIS3.

Other variables are only sporadically significant.

Note this is on the basis of the sample for which we have productivity growth information. In other results (not reported) we found the same pattern of signs and significance for a sample of all manufacturing firms (2,164 for CIS3 and 823 for CIS2). The additional findings here are a positive and strongly significant effect in CIS3 from external vertical information and a positive effect of government support in CIS3. In addition, the total expenditure variable is significant in all cases.

To investigate these results further, we undertook the same regressions for the panel sample. Table 10 shows that co-operation is consistently significant in this case, with internal information significant CIS3. There are two interpretations of this result. First, it could be that the sources of information genuinely changed between CIS2 and CIS3. Alternatively, it could be that the questionnaire in CIS2 was interpreted in a different way.

Table 9

Innovation output Regressions

	CIS3		CIS2	
	Tobit	Probit[+]	Tobit	Probit[+]
	%innovative sales	process innovation	%innovative sales	process innovation
Cooperation	0.1223	0.1247	0.0671	0.1230
	(0.0358)***	(0.0550)**	(0.0317)**	(0.0488)**
Internal Info	0.0904	0.0716	0.0309	0.0498
	(0.0219)***	(0.0314)**	(0.0196)	(0.0286)*
Vertical Info	0.0277	0.1178	-0.0165	-0.0454
	(0.0255)	(0.0354)***	(0.0285)	(0.0414)
Commercial Info	0.0264	0.0101	0.0108	-0.0245
	(0.0244)	(0.0323)	(0.0210)	(0.0309)
free Info	0.0416	0.0168	-0.0321	-0.0443
	(0.0251)*	(0.0345)	(0.0237)	(0.0353)
Regulatory Info	-0.0444	-0.0219	0.0030	0.0345
	(0.0224)**	(0.0307)	(0.0178)	(0.0270)
University Info	0.0302	0.0027	-0.0100	0.0504
	(0.0257)	(0.0364)	(0.0206)	(0.0294)*
Gov. Info	-0.0439	-0.0183	0.0405	-0.0117
	(0.0311)	(0.0423)	(0.0197)**	(0.0304)
Gov. fin. Support	-0.0404	0.1980	0.0163	0.0587
	(0.0533)	(0.0809)**	(0.0466)	(0.0677)
Innovation exp.	0.3819	1.2378	0.9150	2.4004
	(0.4122)	(0.6218)**	(0.2295)***	(0.6944)***
Observations	525		475	

Robust standard errors in parentheses
(+) marginal effects reported
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 10

Innovation output Regressions: "panel" sample

	(1)	(2)	(3)	(4)
panel sample	CIS3		CIS2	
	%new product	process	%new product	process
Co-op	0.2023	0.2136	0.1238	0.1399
	(0.0576)***	(0.0791)***	(0.0441)***	(0.0736)*
Internal Info	0.1297	0.1282	0.0057	0.0501
	(0.0350)***	(0.0459)***	(0.0248)	(0.0394)
Vertical Info	0.0281	0.0772	0.0558	-0.0375
	(0.0000)	(0.0555)	(0.0372)	(0.0596)
Commercial Info	0.0090	-0.0448	0.0201	0.0544
	(0.0349)	(0.0459)	(0.0296)	(0.0455)
free Info	0.0416	-0.0107	0.0004	-0.0995
	(0.0407)	(0.0530)	(0.0000)	(0.0535)*
Regulatory Info	-0.0413	-0.0433	0.0271	0.0227
	(0.0398)	(0.0508)	(0.0289)	(0.0393)
University Info	0.0287	-0.0076	-0.0425	0.0002
	(0.0362)	(0.0543)	(0.0270)	(0.0446)
Gov. Info	-0.0123	-0.0615	-0.0077	-0.0607
	(0.0409)	(0.0596)	(0.0000)	(0.0444)
Gov. fin. Support	0.0676	0.1160	0.0056	0.0922
	(0.0742)	(0.1337)	(0.0622)	(0.0984)
Innovation exp.	0.3001	1.6461	0.2586	1.0218
	(0.3379)	(1.1108)	(0.3076)	(0.4904)**
Observations	235	229	235	233

Robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

5 Conclusions

Our conclusions are rather tentative. First, regarding innovation and productivity growth we have rather mixed results with process and product innovations being important in CIS2 and CIS3 respectively. Second, regarding the correlates of innovation output, co-operation, expenditure and learning from internal sources are consistently significant and positive.

6 References

Arrow, K. J., (1962) “Economic Welfare and the Allocation of Resources for Invention.” In *The Rate and Direction of Innovative Activity: Economic and Social Factors*. Washington: National Bureau of Economic Research.

Barnes, M. and Martin, R., (2002), “Business Data Linking: An Introduction.” *Economic Trends*, 581, April; 34-41.

Bartelsman, E., and Doms, M., (2000), “Understanding Productivity: Lessons from Longitudinal Microdata”, *Journal of Economic Literature,* 38, 3, 569-594.

Bloom, N, and van Reenen, J., (2000), “Patents, productivity and market value in a panel of British firms”, IFS working paper W00/21, *Economic Journal*, forthcoming.

Cabagnols, A. and Le Bas, C. (2002) “Differences in the determinants of process and product innovations: The French Case”, in A. Kleinecht and P. Mohnen (eds.) *Innovation and Firm Performance*, Palgrave, London, pp.112-149.

Cassiman, B. & Veugelers, R. (2001), “R&D Cooperation and Spillovers: some empirical evidence from Belgium”, forthcoming in *American Economic Review,* CEPR Discussion Paper 2330.

Crepon, B., Duguet, E. and Mairesse, J., (2000), Research innovation and Productivity: An Econometric Analysis at Firm Level”, *National Bureau of Economic Research Working Paper*, No.6696.

Frenz, M. (2002) “Innovation activities: a sustained or sporadic phenomenon? An analysis of the second and third Community Innovation Survey” *DTI Project* mimeo.

Griliches, Z., (1998), *R&D and Productivity. The Econometric Evidence*. The University of Chicago Press. Chicago.

Griliches, Z., and Regev, H., (1992), “Productivity and Firm Turnover in Israeli Industry 1979-88”, *National Bureau of Economic Research Working Paper*, No. 4059.

Hall, B., (1987) “The Relationship Between Firm Size and Firm Growth in the US Manufacturing Sector” *Journal of Industrial Economics*, Vol. 35, No. 4, June, pp. 583-606.

Haskel, J and Martin, R., (2002), “The UK Manufacturing Productivity Spread” *CeRiBA*, mimeo.

Hulten, C., (2000), “Total Factor Productivity: a short Biography”” forthcoming in *New Directions in Productivity Analysis*, C. R. Hulten, E. R. Dean, and M. J. Harper (eds.), Studies in Income and Wealth, The University of Chicago Press for the National Bureau of Economic Research, Chicago.

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

van Leeuwen, G. (2002) “Linking Innovation to Productivity Growth Using Two Waves of CIS”, STI Working Paper 2002/8, OECD.

Leiponen, A. (2000) “Competencies, innovation and profitability of firms” *Economics of Innovation and New Technology* 9(1):1-24.

Lööf and Heshmati (2001) "On the relationship between innovation and performance: a sensitivity analysis" SSE / EFI Working Paper Series in Economics and Finance No. 446.

Mairesse, J. and Mohnen; P., (2001), "To Be or Not To Be Innovative: An Exercise in Measurement" *National Bureau of Economic Research Working Paper*, No.8644.

Martin, R., (2002), "Building the Capital Stock." *CeRiBA*, mimeo.

Mohnen, P. and Dagenais, M. (2002) "Towards an Innovation Intensity index: The Case of CIS1 in Denmark and Ireland", in A. Kleinecht and P. Mohnen (eds.) *Innovation and Firm Performance,* Palgrave, London, pp.3-30

OECD (1992) *Oslo Manual – Proposed Guidelines for Collecting and Interpreting Technological Innovation Data*, (First Edition) OECD, Paris.

Oulton, N., (1997), "The ABI respondents database: A new resource for industrial economics research", *Economic Trends*, 528, November, 46-57.

Patel, P. and Pavitt, K., (1995) "Patterns of Technological activity: their Measurement and Interpretation" in Stoneman, P. (ed.) *Handbook of the Economics of Innovation and Technological Change*, pp.14-51,

Perry, J., (1995), "The Inter-Departmental Business Register." *Economic Trends*, 505, November, 27-30.

Tether, B. (2001) "Identifying Innovation, Innovators and Innovative Behaviours: A critical assessment of the Community Innovation Survey (CIS)" CRIC Discussion Paper No 48, December.

7 Data Appendix

7.1 Data Cleaning

This section provides detail on how we cleaned the dataset and provide definitions for relevant variables. We cleaned the dataset according to the following criteria (partly following Hall and Mairesse, 1995).

First, we have removed all observations for which growth of value added, employment, capital and material inputs is missing.

Second, we drop any observation for which the average annual growth rate in value added, gross output or material inputs was more than 300 percent or less than -90%.

Third, we adopt a similar criterion for observations for which the average annual growth in labour and in capital is more than 200% or less than –50%.

7.2 Variables definition

- Output and Input growth rates:
    - CIS2: variable in 1996 – average of the variable between 1994 and 1992
    - CIS3: variable in 2000 – average of the variable between 1998 and 1996

- Information flows:

the following variables are categorical variables that takes values 0 to 3. Each of the them summarises information from the CIS according to the following criterion: the variables takes the maximum values between those reported in that particular group. The groups considered are the following. Note that given minor differences between the two surveys we have to slightly modify the definitions of our variables accordingly.

- Internal sources of information: within the enterprise; other enterprises within the enterprise group
- Vertical sources (Market sources) of information: Clients or Customers; suppliers Competitors
- Commercial sources of information:

- CIS2: consultancy enterprises; research associations or independent Research and Technology Organisations.
- CIS3: consultants; commercial laboratories / R&D enterprises
- University as a source of information:
  - CIS2: universities or other higher education institutes; private non-profit research institutes
  - CIS3: universities or other higher education institutes; private research institutes
- Free:
  - CIS2: professional conferences, meetings, journals; computer based information networks; trade associations; fairs and exhibitions; patent disclosures
  - CIS3: professional conferences, meetings, journals; computer based information networks; trade associations; fairs and exhibitions
- Regulatory: health and safety regulations; product standards; environmental regulations
- Government:
  - CIS2: government institutes; training and enterprise councils; Business links.
  - CIS3 Government research organisations; other public sector (e.g. business links, government offices)