Centre of Excellence for Science and Innovation Studies

CESIS Electronic Working Paper Series

Paper No. 159

Innovation, R&D and Productivity

- assessing alternative specifications of CDM-models

Börje Johansson and Hans Lööf

(CESIS)

January 2009

The Royal Institute of technology
Centre of Excellence for Science and Innovation Studies (CESIS)
http://www.cesis.se

Innovation, R&D and Productivity:
Assessing alternative Specifications of CDM-Models

Börje Johansson[1] and Hans Lööf[2]

Centre of Excellence for Science and Innovation Studies (CESIS) [1,2], Royal Institute of Technology [1,2] and Jönköping International Business School (JIBS) [1]

December, 2008

Abstract

This paper applies a CDM-model framework to depict the successive links (correlations) between (i) innovation expenditure, (ii) innovation output, and (iii) firm productivity. The CDM model has become popular in many countries among scholars using data from the Community Innovation Survey (CIS). First, the study contrasts a general structural OECD version of the model against a model with country-specific design. Second, the study examines the gains from separating the labour force into ordinary and knowledge labour – as a means to avoid double counting of R&D investments. Third, the paper examines the difference between recognising a firm as a member of an unspecified company group versus a multinational group. Fourth, the paper explores how well sales per employee serves as a proxy for labour productivity proper. Fifth, the paper scrutinises the quality of CIS information by comparing key variables from the voluntary CIS survey with the same variables (for the same firms) recorded in the compulsory and audited register data in Sweden.

Keywords: Productivity, Innovation, R&D, CDM-model

JEL Codes**: 031, O32**

[2] Corresponding author Hans Lööf, CESIS, Royal Institute of Technology, 100 44 Stockholm, Sweden. Email:hansl@infra.kth.se, phone: +46 8 790 80 12.

1. Introduction

In 1980 Ariel Pakes and Zvi Griliches published their seminal paper “Patent and R&D at the Firm Level: A First Look”. Their paper reported on the relationship between patent applications and R&D expenditures, based on data for 121 large U.S. companies covering an 8-year period. The study showed that there is a statistically significant relationship between a firm’s R&D expenses and the number applied and granted patents. They described the link between R&D and patent as a “knowledge production function”.

Using refined econometric methodology and richer micro data sets over the past decades, many applied studies have confirmed this basic findings made by Pakes and Griliches (1980). One branch of this literature employs the successive waves of Community Innovation Survey (CIS) data and a 4-equation structural model originally developed by Crépon, Duguet and Mairesse (1998), here referred to as the CDM model framework. This framework has become popular for estimating an extended knowledge production model. Lööf and Heshmati (2002, 2006) showed that the CDM model can be estimated in simple framework including a heckman selection model and an instrumental variables regression.

Recently the OECD initiated an international project with the aim to investigate the possibility of applying general CDM models on the internationally harmonized micro-data from the CIS-surveys in order to compare factors determining innovation and productivity in different member countries (OECD 2008).

The first and prime aim of this paper is to investigate whether or not it is possible to use a common structural model for estimating micro-data in individual countries, with the implicit question: is it necessary to make a specific model design for each individual country? In the context of specific model designs the paper tests the effects of separating total employment into ordinary and knowledge labour, as a way to avoid counting R&D worker inputs twice, first as a labour input and second as wages included in reported R&D spending.

Another specification issue has to do with firms that belong to a multi-firm group, which may be international or totally domestic (uninational). Does it matter to distinguish multinational firms as members of an MNE from uninational firms. For a member of an MNE one may conjecture that knowledge transfer between firms belonging to the same multinational group has the potential of containing a richer set of novelties.

Still another specification issue has to do with a frequently used performance variable, namely sales per employee. Such a measure is often thought of as a proxy form labour productivity, and should ideally be observed as value added per employee. The question is: how do results compare when the two different productivity measures are employed in otherwise similar model specifications?

Our final effort is to examine the quality of CIS data. To what extent can they be appreciated as accurate when matched against similar information from other sources. In the present study it is possible to shed light on this issue, because information about individual firms is available from both the CIS survey and from register data at Statistics Sweden. The latter data set is collected from a compulsory and quality-controlled (audited) survey, whereas the CIS data com from a voluntary survey, with less systematic control. The objective is to reach the conclusion that the CIS data can be trusted. The empirical analysis is based on 2,841 firm level observations from Swedish CIS-data for the period 2002-2004 (CIS IV) merged with register data from Statistics Sweden.

This study's results with regard to a general multi-country may be summarised as follows. First, we find it problematic to apply OECD's " general structural model" to the CIS data from Sweden, and conjecture that the same is true for other European countries (OECD 2008). Such a communal model is a feasible approach for simple model formulations, but not for econometric exercises that aim to capture a spectrum of particularities of knowledge production as a part of a CDM model. The OECD model is compared with a model labelled IGNORE (Knell et al. 2008) and the model SENSE, partly based on Johansson and Lööf (2008).

With regard to model design, the study provides evidence favouring a separation of the labour force into inputs of ordinary and knowledge-intensive labour. The latter category can be associated with a firm's R&D efforts, it should not be included among the determinants in the R&D equation.

As a third result, we find that MNE-membership is neutral with respect to innovation output among innovative firms but it exerts a positive impact on the innovative firm's labour productivity when productivity is approximated by sales. Fourth, considering the descriptive statistics, we find a high degree of agreement between the information on key variables in the

CIS-survey and the register data. Moreover, we provide strong evidence that sales per employee is a good approximation of labour productivity using the CDM-model.

In the next section of the paper we discuss the production function as a general way of assessing the correlation between tangible and intangible investment and firm performance.. Productivity 3 introduces the empirical model. Section 4 presents the data. In section 5 we compare the results from different model specifications and different variable definitions. Section 6 concludes.

2. Literature

It is a commonly held view that R&D and investments in machinery and equipment together with knowledge labour and ordinary labour makes a main contribution to firms' performance (Griliches, 1988; Romer, 1990; Geroski, Machin and Van Reenen, 1993; Jones, 1995; Van Reenen, 1997). Firms invest in knowledge and capital in order to enhance their competitiveness and capability to earn profits. Ericson and Pakes (1995) show that the stochastic outcome of a firm's own investments in R&D together with physical capital, human capital, marketing and the competitive pressure from other firms within or outside the industry determine the sales performance, profitability and growth of the firm.

Already three decades ago, Griliches (1979) stated that the production function approach focusing on total factor productivity or labour productivity as a function of past R&D-investments, physical capital, human capital, firm size and industry specific factors, is the only available general way of trying to answer questions about contribution of R&D to growth. Thirty years later, we still have no better methodology.

Two problems in econometric inference using the production function for estimating the relationship between R&D, innovation and productivity results from selection bias and simultaneity bias. The first issue derives from the fact that the R&D performing firms are a self-selected group, such that firms in the group choose to make commitments to R&D programmes in response to different investment decisions of other companies (Bond et. al., 1999). The second issue, relates to the fact that R&D investments tend to be affected by past productivity, and both these variables tend to move together with other variables of interest. The Frascati Manual (OECD, 1993) defines R&D as creative work undertaken on systematic basis in order to increase the stock of knowledge and the use of this stock of knowledge for new applications. Since it is difficult to finance R&D in a similar way as other investments, the firm's previous economic performance is of fundamental importance. A comprehensive discussion of this phenomenon can be found in Hall (2002). When productivity and R&D, and other variable of interest, tend to move together, the regression results are affected by simultaneity bias.

In an attempt to correct for undesirable effects of selectivity and simultaneity bias and the complexity of innovation processes that have affected many past R&D and patent studies, Crépon, Duguet and Mairesse (1998) specify a four equation model similar to the basic idea outlined in the Pakes-Griliches model (1980). Schematically, the latter can be described as a three-equation model with the three sequential steps. First, the generation of additional knowledge is described by the relation (i) $\Delta K = R + u$, where the growth of economically valuable knowledge, ΔK, is a function of research expenditures, R, plus an error term, u. The authors labeled this relation the "knowledge production function". The second step is given by the equation (ii) $P = \Delta K + v$, where P is the patent outcome of the R&D process and v is an error term. The variable P indicates an invention output which potentially has a clear commercial value. The third step is captured by the equation (iii) $Z = \Delta P + e$, where Z represents the realized economic benefits from the increase, ΔP, in the patent stock, and where e is an error term.

The main contribution of Crépon et al was the inclusion of a selection equation in the model and the design of a structural model appropriate for studies based on information on non-innovative firms and extensive data on innovative firms. This new model is the CDM model, which is applied in the sequel with three different specifications: the OECD model, the IGNORE model, and the SENSE model.

3. The three empirical models

3.1 The structure of the CDM-model

A common empirical approach for analyzing the relationship between R&D, innovation and productivity is a parametric model of Cobb-Douglas form. Many recent versions of this standard model include techniques to correct for selection bias. When only the innovation sample is used in some parts of the model, the firms are not randomly drawn from the larger population, and selection bias may arise. The innovation literature has also suggested that, due to the complicated process from new ideas to innovation output or productivity growth, a knowledge production function should be estimated not as a single equation but as a system of equations. However, when several links of the process of transforming new ideas to productivity are considered in a simultaneous equation framework, simultaneity bias will emerge if the variables of interest tend to move together.

Crépon, Duguet and Mairesse (1998) launched an empirical model (labelled as "CDM" in Lööf and Heshmati 2002), which (i) relates innovation input to innovation output, while it at the same time (ii) takes both selectivity and simultaneity issues into account. Lööf and Heshmati (2002) present a simplification of the original CDM-model in which the general structure of the empirical model can be interpreted as a multi-step model consisting of four equations. At the first step, firms decide whether or not to engage in innovation activities (selection equation), and then a selective group of the firms decide how much they will invest in R&D. This is specified by a Heckman selection model. The second part of the model can be formulated as an instrumental variable equation. All three model specifications in this paper apply an instrumental variable approach, which relates innovation input to innovation output, and innovation output to productivity. More specifically, the model is given by the following four equations:

$$y_{0i} = \begin{cases} 1 & \text{if} \quad y^*_{0i} = X_{0i}\beta_0 + \varepsilon_{0i} > 0 \\ 0 & \text{if} \quad y^*_{0i} = X_{0i}\beta_0 + \varepsilon_{0i} \le 0 \end{cases} \tag{1}$$

$$y_{1i} = y^*_{1i} = X_{1i}\beta_1 + \varepsilon_{1i} \quad \text{if} \quad y_{0i} = 1 \tag{2}$$

$$y_{2i} = \alpha_{21} y_{1i} + \alpha_{23} y_{3i} + X_{2i}\beta_2 + \varepsilon_{2i} \quad \text{if} \quad y_{0i} = 1 \tag{3}$$

$$y_{3i} = \alpha_{32} y_{2i} + X_{3i}\beta_3 + \varepsilon_{3i} \quad \text{if} \quad y_{0i} = 1 \tag{4}$$

where y^*_{0i} is a latent innovation decision variable measuring the propensity to innovate, y_{0i} is the corresponding observed binary variable being 1 for innovative firms. y_{1i}, y_{2i} and y_{3i} describe innovation input, innovation output and productivity, respectively. X_{0i}, X_{1i}, X_{2i} and X3i are vectors of various variables explaining innovation decision, innovation input, innovation output and labour productivity. The predicted inverse Mills' ratio (Heckman,

1979) is included in X_{2i} and/or X_{3i} to correct for possible selection bias. The β's and α's are the unknown parameter vectors. ε_{0i}, ε_{2i} and ε_{3i} are i.i.d. drawings from a multivariate normal distribution with zero mean.

We start with the 2,843 observations in equation (1), and the number of observations are restricted to the 1,042 innovative firms (37% of the observed firms) in equations 2-4. In order to make a proper inference of the results, the standard errors are bootstrapped in equation 4.

3.2 Specification of the model

Equation 1: Selection equation, first step in the Heckman model

In equation 1, we want to explain whether a firm practises innovation activities or not. We use 2 843 observations on all firms in the sample and the dependent variable is

$$y_i = 1 \quad \textit{if} \;\; \textit{firm i is engaged in innovation activities}$$
$$y_i = 0 \quad \textit{if} \;\; \textit{firm i is not engaged in innovation activities}$$

We specify three different versions of the model labelled as (i) OECD, (ii) IGNORE and (iii) SENSE. The three specifications are described in the following table.

Independent variables explaining the propensity to innovate in three specifications

OECD, formulation (1.1)	IGNORE, formulation (1.2)	SENSE, formulation (1.3)
$\ln E$ = ln employment	$\ln E$ = ln employment	$\ln E$ = ln employment
G1 = company group	*G1* = company group	*G2* = MNE
FM = foreign market	*FM* = foreign market	*FM* = foreign market
H1 = knowledge hamp.	-	*H1* = knowledge hamp.
H2 = market hamp.	-	*H2* = market hamp.
H3 = cost hampering	-	*H3* = cost hampering
-	*H4* = total hampering	-
D = industrial dummies	*D* = industrial dummies	*D* = industrial dummies

As described by variable list (1.1), the variables in the "OECD" specification comprise the variables $\ln E$, *G1* which is a dummy variable indicating whether or not the firms belongs to a company group, *FM* which is a dummy variable informing about whether the firm has foreign customers or not, *H1* which informs whether or not any of the following factors during the past three years have hampered innovation activities: (i) lack of qualified personnel (ii) lack of information about technology and (iii) difficulties in finding cooperation partners for innovation, *H2* which is a dummy for (i) lack of information about potential markets (ii) the presence of established enterprises that dominate the relevant market(s), and (iii) uncertainty with regard to demand for innovative goods or services as a factor which hampers innovation efforts, *H3* which is a dummy variable that indicates whether there have been innovation obstacles during the past three years due to (i) lack of funds within the own enterprise or group, (ii) lack of finance from sources outside the enterprise or (iii) too high innovation costs. Finally, *D* represents industry dummies.

The next specification, called IGNORE in (1.2), is essentially the same as (1.1) with the exception that the three hampering variables, *H1*, *H2* and *H3*, have been aggregated to one index-dummy variable, labelled *H4*. The third specification in (1.3) is identical to the "OECD" specification with one exception: the corporate ownership-structure variable, *G2*, is restricted to refer to international company groups, i.e., domestically and foreign-owned multinational enterprises (MNEs).

Equation 2: Outcome equation, second stage in the Heckman model

Given that a firm is classified as innovative, equation (2) has the logarithm of R&D expenditures and other innovation cost per employee as dependent variable, denoted $\ln(I/E)$. The explanatory variables in the three regressions are presented below:

Three specifications of explanatory factors where $\ln(I/E)$ is *dependent variable*

OECD, formulation (2.1)	IGNORE, formulation (2.2)	SENSE, formulation (2.3)
G1 = company group	*G1* = company group	*G2* = MNE
FM = foreign market	*FM* = foreign market	*FM* = foreign market
C1 = external collaboration	*C2* = cooperation index	*C1* = external cooperation
-	-	$\ln(\hat{E})$ = log ordinary labour
-	-	PA = patent application
D = industrial dummies	*D* = industrial dummies	*D* = industrial dummies

The OECD-specification in (2.1) contains indicators for group membership *G1*, foreign market access *FM*, and external innovation partners *C1*. A set of industry dummies is included. The IGNORE specification in (2.2) substitutes a cooperation index, *C2*, for the binary cooperation variable, *C1*, where the cooperation index captures the frequency of external collaboration. The third specification in (2.3) includes log ordinary labour, $\ln(\hat{E})$, as a size variable. It should be noticed that the SENSE specification separates the labour force into ordinary and knowledge-intensive labour, where the latter category can be associated with a firm's R&D efforts. By distinguishing between knowledge workers and ordinary workers, we eliminate or reduce the possible endogeneity of the employment variable in the R&D-equation. Moreover, we introduce a dummy variable for patent application, denoted *PA*. Otherwise the variables are the same as in the OECD-specification.

Equation 3: First step in the two stage least square estimator

The instrumental variables estimator is used in order to correct for possible inconsistency in the parameter estimations due to endogenous regressors, meaning that changes in *x* are associated not only with changes in *y*, but are also associated with changes in the error term. Technically, we introduce a (set of) variable(s) *z* that has the property that changes in *z* are associated with changes in *x* but does not lead to change in *y* other than indirectly via *x*. In this paper we do this by employing the two-stage least square (2SLS) estimator. The potentially endogenous variable (innovation output as a right-hand side variable in the productivity regression) that we would like to be instrumented is estimated in the first stage. The including instruments are both the exogenous covariates in stage one and the full set of variables in stage two. In the specifications below (3.1-3.3) we present the exogenous covariates in stage two, where the dependent variable is $\ln(\hat{S}/E)$, reflecting innovation sales per employee.

As can be seen for equation 3, the OECD and the SENSE specifications employ the same types of variables, whereas the IGNORE specification lack knowledge indicators and has instead a set of hampering variables, *H1*, *H2* and *H3*. The estimation of $\ln(\hat{S}/E)$, where $\hat{S}$ denotes innovation sales and E total employment, has been regressed on innovation sales per employee, $\ln(I/E)$, as predicted from equation 2. This part of the regression is the same for all three formulations. Equations 3.1 and 3.3 use the same cooperation indicators, *C7*, *C8*, *C9* and *C10*. These two equations also include information about knowledge labour, $\hat{\hat{E}}/E$ and $\ln\hat{\hat{E}}$, respectively. In addition, these two equations have physical investments $\ln(M/E)$ and $\ln M$, respectively.

Equation 3.3 is special by using two labour input variables, ordinary labour, $\hat{E}$, and knowledge labour $\hat{\hat{E}}$. By introducing this distinction, we hopefully have eliminated the endogeneity that may obtain when employment is a variable in the R&D function. Finally, all three equations include an inverted mills ratio in order to correct for possible selection bias.

Three specifications of independent factors explaining $\ln(\hat{S}/E)$ *in equation 3*

OECD, formulation (3.1)	IGNORE, formulation (3.2)	SENSE, formulation (3.3)
G1 = company group	*G1* = company group	*G2* = MNE group
ln*E* = log employment	ln*E* = log employment	$\ln \hat{E}$ = log ordinary labour
$(\hat{\hat{E}}/E)$ = knowledge intensity	-	$\ln \hat{\hat{E}}$ = log knowledge labour
$\ln(M/E)$ = log machinery investment per employee	-	$\ln M$ = log machinery investment
$\ln(I/E)$ = log innovation investment per employee	$\ln(I/E)$ = log innovation investment per employee	$\ln(I/E)$ = log innovation investment per employee
PR = process innovation dummy	*PR* = process innovation dummy	*PR* = process innovation dummy
C7 = client cooperation	*C3* = inward cooperation	*C7* = client cooperation
C8 = supplier cooperation	*C4* = outward cooperation	*C8* = supplier cooperation
C9 = private cooperation	*C5* = international cooperation	*C9* = private cooperation
C10 = public cooperation	*C6* = national cooperation	*C10* = public cooperation
-	*SM* = small and medium-sized	-
-	*H3* = cost hampering	-
-	*H1* = knowledge hampering	-
-	*H2* = market hampering	-
Inverted mills ratio	Inverted mills ratio	Inverted mills ratio
D = industrial dummies	*D* = industrial dummies	*D* = industrial dummies

Equation 4: Second step in the two stage least square estimator

In equation 4 , the regression make use innovation sales, as predicted by the pertinent version of equation 3, when estimating the productivity effects of R&D efforts. In this way, we may think of the predicted value of $\ln(\hat{S}/E)$ as the major variable for explaining the productivity effect as measured by $\ln(S/E)$, i.e., log sales per employee. We may also not that *S/E* is a proxy for labour productivity proper, and its capacity to replace value added per employee, *V/E*, is evaluated in table 6.

The OECD specification, (4.1), has a similar structure as the SENSE specification, (4.3). The representation labour inputs differ as (4.1) uses total employment, $\ln E$, and knowledge intensity, $\hat{\hat{E}}/E$, whereas (4.3) uses ordinary labour, $\ln \hat{E}$, and knowledge labour, $\ln \hat{\hat{E}}$, where we observe that E = $\hat{E}+\hat{\hat{E}}$. Another scale difference is that physical investments are represented by $\ln(M/E)$ in (4.1) and by $\ln M$ in (4.3). Finally, the two equations differ by having *G1* as company group variable in (4.1), whereas *G2* is used in (4.3), where *G2* refers to multinational company groups. The IGNORE equation (4.2) does not include information about physical investment and knowledge labour.

Independent variables explaining $\ln(S/E)$ *in three specifications of equation 4*

OECD, formulation (4.1)	IGNORE, formulation (4.2)	SENSE, formulation (4.3)
$\ln(\hat{S}/E)$ = ln innovation sales per employee	$\ln(\hat{S}/E)$ = ln innovation sales per employee	$\ln(\hat{S}/E)$ = ln innovation sales per employee
$\ln E$ = ln employees	$\ln E$ = ln employees	$\ln \hat{E}$ = ln ordinary labour
$\hat{\hat{E}}/E$ = knowledge intensity	-	$\ln \hat{\hat{E}}$ = ln knowledge labour
$\ln(M/E)$ = ln machinery investment per employee	-	$\ln M$ = ln machinery investment
G1 = company group	G1 = company group	G2 = company group
PR = process R&D	PR = process R&D	PR = process R&D
Inverted mills ratio	Inverted mills ratio	Inverted mills ratio
D = industrial dummies	D = industrial dummies	D = industrial dummies

Before concluding this section, some important difference between the three models should be stressed. First, only the OECD and SENSE specifications contain information on human capital and physical capital. Second, the SENSE specification distinguishes between ordinary and skilled labour in equations 2-4. Third, the information on obstacles to innovation and external cooperation on innovation has been exploited in a slightly different way in the IGNORE specifications compared to the two other specifications. Fourth, only the SENS-specification contains a patent variable. We now turn to the data presentation.

4. Data Description

This section describes the data used in the subsequent analysis. The data we use is a combination between the Swedish Community innovation survey IV data (covering the period 2002-2004) and register data on the observed firms. In total 2,843 firm level observations are included in the study. All variables are presented in Appendix I in Table AI, BI and CI.

From Table 1, Panel A, we can see that the median values of employment and turnover are almost identical between the two different data sources, while the CIS-data report slightly larger mean values than the register data. The proportion of those firms which report both innovation investments and income from new product innovation is 37%. In the subsequent analysis we define this group of firms as "innovative." The CIS-survey is censored to cover only firms with 10 or more employees. However, the register data indicates that also firms with fewer than 10 employees are included in the survey.

A prominent feature in the sample is that the innovative firms are significantly larger than the other firms. Moreover, they have higher value added per employee and they are more human capital intensive (knowledge intensive), as can be seen from Table 1, Panel B.

Table 2 shows firm characteristics for all firms and for innovative firms. Looking first at corporate ownership structure we see that the majority of the firms belong to a group and that 55% of the innovative firms also have affiliates outside Sweden, i.e., they are classified as multinationals. Although we know from other studies (Anderson et al 2007) that most Swedish firms are mainly oriented towards the domestic market, close to 60 percent of all firms and nearly 80% of the innovative firms have foreign markets. The hampering variables indicate that innovative firms suffer more from obstacles to innovate than non-innovative firms.

The lower part of table 2 report characteristics only for innovative firms. 56% of the firms cooperate on innovation with external partners and the additional information on cooperation on innovation and the demand-pull variable as well the cooperation variables are based on various index-calculations based on the CIS-data. The bottom of the table report that the typical firm in the analysis is an SME with less than 250 employees.

Appendix II describes the data treatment. Panel A informs that the original matched data consisted of 3,108 observations with 10 or more employees according to the CIS-survey. From this we dropped 266 observations due to the following reasons: (i) 74 observed firms outside the Nace 2 classes 15-37 or 50-74, (ii) 68 observations with zero or negative value added, (iii) 20 observations with less than 5 employees according to the register data and (iv) 21 observations with zero turnover in the CIS-data. Moreover, the following censoring rules have been applied: Innovation expenditures larger than turnover times 2 is censored to turnover times 2 (8 changes); Physical investment larger than turnover times 2 is censored to turnover times 2 (10 changes); Value added (from the register data) larger than turnover times 0.8 is censored to turnover times 0.8.

5. RESULTS

Our key variables are the performance variables productivity and innovation output, the innovation-effort variables R&D and other innovation investment variables, and the production-function variables labour force, human capital and physical capital. The first conceptual issue is the definition of productivity. The CIS data only contain observations on sales, while value added would have been a more appropriate measure. Second, labour force, can be separated into ordinary labour and knowledge-intensive labour, where the latter can be associated with a firm's R&D effort. In the SENSE-specification of the model, we suggest that the size of R&D should not be explained by total labour but by ordinary labour alone in order to avoid simultaneity. A closely related issue is how to define knowledge intensive labour or human capital. In the CIS-data human capital is defined as employees with a university background. The SENSE-model applies three years university education as a minimum criterion for classifying an employee as knowledge labour. A final conceptual issue concerns the corporate ownership structure. The CIS-data informs whether or not a firm belongs to a group. However, recent research suggest that international transmission of technology provides a potential to increase the revenues from firms' investment in research and development and that multinational enterprises, MNEs, are in a special position to handle knowledge transfer (see, for instance, Gupta and Govindarajan, 1994). In view of this, we find it of special importance to examine whether a firm's membership of an international company group is a more informative measure than an unspecified membership of any type of company group .

The main objective of the paper is (i) to investigate how a firm's productivity is correlated with innovation, and how its innovation is correlated with R&D, and (ii) to examine how these correlations are affected by different specifications of the CDM model. This type of assessment is motivated by the frequent used of the CDM framework among scholars using the Community Innovation Survey. Three specifications are compared. The basic specification, called OECD, has been developed and negotiated by an international research team, with a specific objective to lay the ground for a consensus approach to analyse CIS micro data (OECD 2008). The first alternative to the OECD specification is a formulation suggested by the Nordic IGNORE project[1].

The OECD and the IGNORE formulations are thus two recently proposed specifications for international comparison studies based on CIS data collected in different countries. The third model version, called SENSE, is a specification which has the same structure as the OECD-specification, though with some important modifications: (i) we separate the labour force into "ordinary labour" and "knowledge labour", and (ii) firms are categorised into the two groups MNE members and other firms.

Our estimates of the Heckman model are presented in Table 3 and the Instrumental variable regressions are shown in Table 4. In order to make some sensitivity test, the innovation output and productivity estimates from the structural CDM-model is compared with OLS estimates. These results are presented in table 5. In the second sensitivity test, we use an identical specification (SENSE) and investigate the impact of using sales per employee as a proxy for labour productivity (value added per employee). This latter comparison is presented in Table 6.

5.1 Selection equation in the Heckman model

Starting with the selection equation, the left part of Table 3 displays the coefficient estimates for determinants to the firm's decision to engage in R&D activities. The most important finding here is that the point estimates for group membership (column 1 and column 2) is about of the same order as the as the MNE-indicator (column 3). Both variables are positive and significant. A second finding is that the overall results are in line with the literature and no significant difference between the three specifications can be found. As could be

[1] The main objective of the IGNORE (Innovation and Growth in the Nordic Economies) project is to suggest an empirical model for analyzing the relationship between innovation and productivity growth in the Nordic and Baltic countries using firm level data.

expected, the propensity to be engaged in innovation activities is an increasing function of firms' size, foreign market and various obstacles to innovation activities. The latter seemingly paradoxical results signal that the typical innovative firm is not running out of new ideas. A down-to-the-earth interpretation would be that obstacles become relevant to firms which try to make innovation efforts.

5.2 Outcome equation in the Heckman model

Not surprisingly, the outcome equation in the right part of Table 3 reports that the extent of a firm's R&D correlates with foreign-markets contacts and external cooperation on innovation. Given that a firm has decided to be engaged in R&D activities, group membership is neutral with respect to the size of the R&D-expenditures, according to the OECD and the IGNORE models. Likewise, the MNE indicator in the SENSE-regression is just outside the 10% level of significance.

Only the SENSE alternative includes the labour force among the covariates and it is restricted to ordinary labour in order to avoid R&D as both left-hand side and right-hand side variables. Column 6 reports that the number of ordinary labour is negatively associated with R&D, in a highly significant way, even when controlling for industry classification. Column 6 also informs that the size of R&D-expenditures correlates positively with patent-application.

The Chi2-statitics in the bottom of the table provides test information about the correlation between the two error terms in the selection equation (1) and the outcome equation (2). If the error terms from the two equations are uncorrelated, the innovation expenditures equation can be estimated in a consistent way by ordinary least squares without the selection equation. However, if the error terms are correlated ($\sigma_{12} \neq 0$) the OLS estimator will result in a sample selection bias. Interestingly, the statistics only suggest presence of correlation in the SENSE specification. Hence, the specification of the first step of the OECD model and the IGNORE model is not satisfactory. Only for the SENSE specification the Heckman model is motivated.

5.3 Innovation output equation in the instrument al variable regression

The left part of Table 4 shows the innovation output estimate, where the dependent variable is log innovation sales per employee. Previous literature as presented in Appendix III, suggest that the elasticity of productivity with respect to R&D should be in the range 0.01-0.40 with a concentration around 0.10. Recent research on the correlation between innovation output and R&D suggest estimates that are somewhat higher with a concentration of the estimates in the range 0.20-0.40. Looking at the top of Table 4, we see that the OECD and SENS point estimates for the correlation between log innovation sales and log innovation R&D and other innovation expenditures are of plausible size, 0.20 and 0.35 respectively. However, only the SENSE estimate is significant. The estimate associated with innovation input is not statistically different from zero in the OECD-regression. Moreover, the IGNORE result is unreasonably large (1.21). In order to control for possible selection bias of the estimate, three specifications contain a predicted mills ratio (from the Heckman model). Table 4 reports that it is non-significant in all regressions.

In an attempt to reduce the risk of explaining the size of R&D expenditures with R&D employees in the outcome equation (2), the SENSE specification has eliminated skilled labour from the employment variable. In equation (3) both categories of the labour force are included among the determinants to innovation output. Table 4 reports that their estimated effect have different signs, with a negative and insignificant parameter for ordinary labour. The log of knowledge labour is positive (0.06) and significant at the 10% level. In the OECD-specification, the employment variable is non-significant and the dummy variable for knowledge labour is significant and quite sizeable (1.15). The coefficient for knowledge labour in the SENS-estimation appears to be very small as compared to the knowledge-intensity coefficient in the OECD regression. However, it should be noted that the two variables are defined differently. While the knowledge-labour variable expresses log of employment with a university education three years or more, the knowledge-intensity variable informs about the fraction of employment with university education.

In contrast to the other two specifications, the IGNORE has no human capital variable among the covariates. Therefore, the only labour force variable is total employment and the table shows a negative association with innovation output at the 10% level of significance.

One of the objectives of the paper is to test whether there is different information in the dummy variable for company-group membership compared to membership of a multinational company. The underlying hypothesis is that firms that are part of a company group with foreign affiliates have better access to global knowledge than firms that belong to a company group with only domestic affiliates. The group variable does not distinguish between firms that are part of a group with only domestic firms and firms that also have foreign affiliates. The MNE-indicator, however, informs that the firm has foreign affiliates.

Looking at the estimates displayed in the left part of Table 4 it is shown that the group-variable (G1) is significant in the OECD and IGNORE regressions, whereas the MNE-estimate (G2) is insignificant in the SENSE alternative. Hence, we can establish than the two variables contain different information and the tentative interpretation is that domestic groups stimulate their members in the process of transferring R&D to new products.

In concordance with established theory, current change in physical capital is positively related to R&D investment in the OECD and SENSE regressions, supporting the idea that innovation in form of ideas must be embodied in new machinery and equipment in order to generate new products and innovation sales. It should be noted that the physical capital is defined different in the two equations. In the OECD regression, this variable is defined as log per employee, while it is log of total physical capita in the SENSE regression.

The results of the innovation output equation point to the importance sensitivities of process innovation to product innovation, indicating that process and product innovations typically are interrelated. However, in contrast to the OECD regression, the estimates are only weakly significant in the two other cases. As noted in previous literature on CIS-data analyses, the cooperation variables are not associated with innovation output in any evident and systematic way (See Kleinknecht, 1998).

5.4 Labour productivity equation in the instrumental variable regression

The results of the productivity equations are presented in the right part of table 4. Looking at the key variable for the purpose of the study, the table reports that point elasticities for innovation output (predicted from equation 3) are highly significant and have expected signs and sizes (OECD 0.51, IGNORE 0.52 and SENSE 0.36), as can be seen from Appendix III.

Turning to the coefficients on firm size and labour force, columns 4 and 5 report that productivity is an increasing function of firm size according to the OECD and IGNORE specifications. However, when the labour force is split into two different categories, as in column 6 for SENSE, we can see that productivity is an increasing function only of knowledge labour. The point estimate for ordinary labour is outside any acceptable degree of significance.

The results from equation 3 show that MNE membership is neutral with respect to a firm's innovation output. The important factor is whether or not a firm belongs to a group. Interestingly, the SENSE estimation presented in column 6 indicates that MNE membership actually exert a positive influence on firm performance, but directly through its production function (equation 4). The somewhat more important magnitude of the MNE-coefficient in the SENSE regression (0.18 and highly significant), indicates that MNEs are more efficient than other firm in transforming R&D, human capital, physical capital, and other inputs to higher sales per employee. The *G1* coefficient, referring to company-group membership, is much lower, with the value 0.08, and only significant at the 10% level in the OECD regression. In the IGNORE regression the *G1* coefficient is not significant at all.

5.5 Ordinary least square estimates

Table 5 provides some sensitivity tests of the estimates by presenting OLS regression results for the innovation output equation and the productivity equations, respectively. We focus the discussion on two estimates: R&D and innovation output

Starting with R&D (log innovation expenditures per employee), the left part of Table 5 shows almost identical point estimates for the correlation with innovation output (log innovation sales per employee). The magnitudes of the estimates are close to 0.2 and highly significant. This shall be compared with the CDM-results above that showed insignificant OECD estimates, questionably large magnitude of the IGNORE estimate and twice as high SENSE estimates.

The main difference between these OLS regression results and those presented in table 4 is that we here use the observed R&D-expenditures and not the predicted values in the OLS together with the inverted Mills ratio. Further, the instrumental-variables part of the CDM model also includes all the covariates from the productivity equation among the determinants in the innovation output equation, while the OLS only contains those reported in the left part of Table 5.

Of course, it can be expected that the OLS estimates reported in Table 4 are biased. But in the cases of OECD and IGNORE they also indicate some problem with these specifications in the CDM model. The OLS results are closer to the literature than the CDM-resultats. Moreover, it can be assumed that (the relatively small) difference between the two estimates using the SENSE specification is explained by biased OLS-estimates.

In correspondence to the parameter estimates of R&D input reported in the left part of the table, the differences in the point estimates of the innovation output are small between the three regressions. In fact, the elasticity of productivity with respect to innovation output is 0.3 and highly significant in both the OECD and the SENSE regressions. The IGNORE estimate is 0.22 and significant at the 1% level.

Our concluding finding from the OLS-regressions is that they suggest that innovation output is an increasing function of innovation input and that labour productivity is closely associated with innovation output. Only the SENSE specification produces estimates for the two relationships of a similar magnitude as the OLS-regression. Some miss-specifications of the other models can be suspected and this shows up in the innovation output equation.

5.6 Definition of labour productivity

The CIS-data do not allow for a proper productivity analysis, since no information on value added is collected in the survey. An important issue is therefore how well labour productivity can be approximated by the sales variable. Table 6 shows selected variables from the last step in the four-equation CDM-model using the SENSE specification with the two different performance measures. The covariates are identical to those included in equation 4.3.

Column 1 reports the estimates using sales information from the CIS-survey. These results are the same as the ones reported in Table 4, column 6. In contrast, column 2 presents the results when when we have substituted value added per employee for sales per employee. The first row of column 2 provides strong evidence that the CIS-reported innovation sales is a good proxy for labour productivity proper. The elasticity of log value added per employee with respect to innovation output is 0.38 and the result is highly significant. This estimate is only marginally larger than the elasticity of sales with respect to innovation output (0.36).

6. CONCLUSIONS

This paper investigates how the sequence of correlations between productivity, innovation and R&D is affected by different specifications of a general model that recently has become widely popular among scholars using the Community Innovation Survey (CIS) information.

In addition, the paper examines the quality of the CIS-data. It demonstrates the problems with using a common structural model for estimating micro-data in individual countries and indicates that a country-specific design may be preferable. The paper shows the benefits from separating the labour force into ordinary and knowledge labour in order not to count R&D-investments twice. With regard to the distinction between multi-unit firms, the analysis does not manage to show that MNE membership is a more appropriate variable than an unspecified company-group membership. Especially, the paper shows that sales value per employee is a feasible substitute for value added per employee. Finally, the paper also assesses the quality of the CIS-information by comparing key variables from the voluntary the innovation survey with information in the compulsory and audited register data. The two sources of information are found to be consistent.

Referenses

Andersson, M., Lööf, H. and Johansson, S. (2008) Productivity and International Trade - firm-level evidence from a small open economy. *Review of World Economics.* Vol 144.

Bond, S., Elston, J.A., Mairesse, J., Mulkay, B., (1999) "Financial Factors and Investment in Belgium, France, Germany and the UK: a Comparison using Company Panel Data," Working Papers 9964, Centre de Recherche en Economie et Statistique.

Crépon, B., Duguet, E., Mairesse, J., (1998) "Research, Innovation, and Productivity: An Econometric Analysis at the Firm Level," *Economics of Innovation and New Technology*, 7(3): 115-156.

Ebersberger, B., Lööf, H,. (2005) "Multinational Enterprises, Spillover, Innovation and Productivity," *International Journal of Management Research.*

Gupta, A. K., Govindarajan, V., (1994) "Organizing for knowledge flows within MNCs," *International Business Review*, 3 (4), 443-57.

Ericson, R. E., Pakes, A., (1995) "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *Review of Economic Studies*, 62, 53–82

Geroski, P., Machin, S.,Van Reenen, J, (1993) "The Profitability of Innovating Firms," RAND Journal of Economics, The RAND Corporation, vol. 24(2), 198-211.

Griliches, Z., (1988) "Productivity Puzzles and R&D: Another Nonexplanation," *Journal of Economic Perspectives*, American Economic Association, vol. 2(4), 9-21.

Hall, B. H., (2002) "The Financing of Research and Development," *Oxford Review of Economic Policy*, Oxford University Press, vol. 18(1), 35-51.

Heckman, J., (1979) "Sample Selection Bias as a Specification Error", *Econometrica*, vol47, 153–161.

Janz, N., Lööf, H., Peters, B., (2004) "Innovation and Productivity in German and Swedish Manufacturing Firms: Is there a Common Story? "*Problems& perspectives in management*, 2, 184-204.

Johansson, B., Lööf, H., (2008). "The Impact of Firm's R&D Strategy on Profit and Productivity," Working Paper Series in Economics and Institutions of Innovation 156, Royal Institute of Technology, CESIS.

Jones, C. I., (1995) "R&D-Based Models of Economic Growth," *Journal of Political Economy*, University of Chicago Press, vol. 103(4), 759-784.

Kleinknecht, A., (1998) "Is Labour Market Flexibility Harmful to Innovation?" *Cambridge Journal of Economics*, 22 (3), 387-96.

Knell, M., Nås, S.O., Bloch, C., Graversen, E., Lööf, H., Finnbjörnsson, T. Vahter, P., Massao, J., Lehtoranta, O., Squicciarini, M., Hahnel, P and Mohnen, P., (2008). "Innovation and growth in the Nordic Economics ," Nordic Innovation Centre.

Lööf, H. and Heshmati, A., (2002) "Knowledge capital and performance heterogeneity: A firm-level innovation study," *International Journal of Production Economics*. 76 (1), 61-85.

Lööf, H., (2004) "A Comparative Perspective on Innovation and Productivity in Manufacturing and Service Firms," in Uwe Cantner, Elias Dinopoulos Robert F. Lanzillotti (eds.) *Entrepreneurship, the New Economy and Public Policy, Schumpeterian Perspectives*. Springer, 181-202.

Lööf, H., Heshmati, A., (2006). "On the Relationship Between Innovation and Performance: A Sensitivity Analysis," *Economics of Innovation and New Technology*. Vol 15, Issues 4/5, 317-345.

OECD (1993) "The Measurement of Scientific and Technological Activities. Proposed Guidelines for Collecting and Interpreting Innovation Data," Edition, Paris: OECD.

OECD (2008). "Innovation in Firms : Findings from a comparative analysis of innovation surveys microdata. Background and highlights," DSTI/STP(2008)7/CHAP5. Paris:OECD.

Pakes, A., Griliches, Z., (1980). "Patents and R and D at the Firm Level: A First Look," NBER Working Papers 0561, National Bureau of Economic Research.

Romer, P. M., (1990) "Endogenous Technological Change," *Journal of Political Economy*, University of Chicago Press, vol. 98(5), 71-102.

Van Reenen, J., (1997) "Why has Britain had slower R&D growth?," *Research Policy*, Elsevier, vol. 26(4-5), 493-507.

Table section

Table 1: Summary descriptive statistics quantitative variables

Panel A: All firms. Number of observations 2, 843

	Data	Mean	Std Dev	Median	Min	Max
Employment	CIS	207	1,153	27	10	39,333
Log turnover [a]	CIS	4,95	0.84	4.88	-1.96	8.54
Log innovation costs [b]	CIS	0.040	0.152	0.000	0	2
Log innovation income [b]	CIS	0.092	0.182	0.000	0	1
Firm with innovation costs and income	CIS	0.370	0.482	0.000	0	1
Employment	ADM	188	1,037	26	5	34,299
Log turnover [a]	ADM	4,94	0.79	4.88	-1.80	8.49
Log value added [b]	ADM	3.99	0.51	3.96	-2.93	4.98
Log physical investments [c]	ADM	0.119	0.209	0.053	0	2
Human capital I [d]	ADM	0.25	0.23	0.16	0	1
Human capital II [e]	ADM	0.13	0.18	0.06	0	1

Panel B: Innovative firms: firms with both innovation costs and innovation income.
Number of observations 1,042

	Data	Mean	Std Dev	Median	Min	Max
Employment	CIS	287	1,115	42	10	19,456
Log turnover [a]	CIS	5.08	0.77	5.02	-1.96	8.25
Log innovation costs [b]	CIS	0.084	0.213	0.027	0.01	2
Log innovation income [b]	CIS	0.227	0.226	0.15	0.01	1
Employment	ADM	262	1,054	41	5	18, 141
Log turnover [a]	ADM	5.07	0.72	5.02	-1.80	7.88
Log value added [b]	ADM	4.05	0.55	4.04	-2.94	6.12
Log physical investments [b]	ADM	0.113	0.192	0.053	0.01	1
Human capital I [d]	ADM	0.31	0.25	0.21	0	1
Human capital II [e]	ADM	0.16	0.29	0.08	0	0.91

Notes: (a) In 1000 Euros, (b) Per employee, (s) Fraction of turnover, (d) Employment with a university education as a fraction of total employment, (e) employees with a university education three years or more as a fraction of total employment. (CIS) Survey data from Community Innovation Survey 4, referring to year 2004. (ADM) Administrative data from Statistics Sweden on firms observations in the CIS survey. The data refers to year 2004. (CIS) is data from Community Innovation Survey IV.

Table 2: Summary descriptive statistics indicator variables

Panel B: Descriptive statistics

	All firms: n=2,841		Innovative firms: n=1,052	
	Mean	Std. dev	Min	Max
G1 (company group)	0.60	0.48	0.71	0.45
G2 (MNE)	0.42	0.49	0.55	0.49
PA (patent application)	0.09	0.29	0.19	0.39
FM (foreign market)	0.58	0.49	0.76	0.42
H1 (hampering: knowledge)	0.13	0.34	0.18	0.38
H2 (hampering: market)	0.28	0.45	0.29	0.45
H3 (hampering: cost)	0.25	0.43	0.32	0.46
H4 (hampering total)	0.35	0.24	0.44	0.19
C1 (external cooperation)	-	-	0.56	0.49
C2 (cooperation index)	-	-	0.11	0.16
C3 (inward cooperation)	-	-	0.14	0.19
C4 (outward cooperation)	-	-	0.09	0.16
C5 (international cooperat.)	-	-	0.13	0.22
C6 (national cooperation)	-	-	0.24	0.28
C7 (demand -pull)	-	-	0.60	0.22
SME	-	-	0.77	0.41

Table 3: Heckman equation

	Selection equation Dependent variable: Innovative firm.			Outcome equation Dependent variable: Log innovation expenditures per employee		
	OECD	IGNORED	SENS	OECD	IGNORED	SENS
$\ln E$	0.128*** (0.022)	0.118*** (0.023)	0.113*** (0.021)			
$\ln \hat{E}$						-0.225*** (0.045)
G1	0.219*** (0.058)	0.200*** (0.060)		0.100 (0.124)	-0.054 (0.124)	
G2 (MNE)			0.184** (0.063)			0.209 (0.132
PA (patent)						0.746*** (0.138)
FM (foreign market)	0.596*** (0.056)	0.580*** (0.057)	0.585*** (0.057)	0.674*** (0.163)	0.501*** (0.164)	0.389* (0.224)
H1 (knowl.)	0.296*** (0.070)		0.303*** (0.066)			
H2 (market)	-0.074 (0.057)		-0.057 (0.058)			
H3 (cost)	0.297*** (0.060)		0.319*** (0.060)			
H4 (total)		1.361*** (0.117)				
C1 (coop)				0.650*** (0.105)		0.616*** (0.105)
C2 (index)					2.541*** (0.333)	
CONST	Included	Included	Included	Included	Included	Included
D (industry)	Included	Included	Included	Included	Included	Included
Prob>chi2				0.126	0.709	0.000
OBS	2,841	2,841	2,841	1,052	1,052	1,052

Notes: Innovative firm is a dummy variable indicating whether or not a firm has reported both innovation expenditures year *t* and sales income from innovations year *t* launched on the market during the period t $_{t-2}$
Standard errors within parentheses. ***significant at 1%; ** significant at 5% and * significant at 10% level.

Table 4: Instrumental variable regression

	Innovation output Equation Dependent variable LISPE			Productivity Equation Dependent variable LLPPE		
	OECD	IGNORED	SENS	OECD	IGNORED	SENS
$\ln(I/E)$	0.194	1.213***	0.348***			
	(0.151)	(0.526)	(0.111)			
$\ln(\hat{S}/E)$				0.513***	0.521***	0.363***
				(0.172)	(0.079)	(0.093)
$\widehat{MILLS}$	-0.002	0.894	0.313	0.083	0.197**	0.138
	(0.260)	(0.668)	(0.219)	(0.099)	(0.091)	(0.089)
PROCESS	0.183**	0.141*	0.163*	-0.067	-0.057	-0.038
	(0.083)	(0.078)	(0.084)	(0.057)	(0.038)	(0.045)
G1 (group)	0.201**	0.307**		0.076*	0.032	
	(0.095)	(0.132)		(0.057)	(0.042)	
G2 (MNE)			0.159			0.177***
			(0.103)			(0.052)
$\ln E$	-0.050	-0.120*		0.089***	0.100***	
	(0.039)	(0.067)		(0.021)	(0.027)	
$\ln \hat{E}$			-0.061			0.035
			(0.053)			(0.024)
$\ln(\hat{\hat{E}}/E)$	1.145**			-0.052		
	(0.223)			(0.218)		
$\ln \hat{\hat{E}}$			0.055*			0.025*
			(0.029)			(0.014)
$\ln(M/E)$	0.076***			0.014		
	(0.019)			(0.019)		
$\ln M$			0.042**			0.018
			(0.019)			(0.012)
C7 (clients)	-0.103		-0.091			
	(0.114)		(0.115)			
C8 (suppliers	0.220*		0.148			
	(0.117)		(0.114)			
C9 (private)	-0.093		-0.131			
	(0.110)		(0.112)			
C10 (public)	0.020		-0.031			
	(0.112)		(0.113)			
H3 (cost)		-0.038				
		(0.255)				
H1 (knowl.)		0.796***				
		(0.301)				
H2 (market)		0.184				
		(0.278)				
C3 (inward)		-0.007				
		(0.695)				
C4 (outward)		-0.407				
		(0.704)				
C5 (internat.)		-0.118				
		(0.201)				
C6 (nationell)		-0.961***				
		(0.262)				
SM (SMEs)		-0.309**			0.046	
		(0.152)			(0.071)	
D (industry)	Included	Included	Included	Included	Included	Included
CONST	Included	Included	Included	Included	Included	Included
OBS	1,052	1,052	1,052	1,052	1.052	1,052

Notes: Standard errors within parentheses. ***significant at 1%; ** significant at 5% and * significant at 10% level

Table 5: Ordinary least square

	Innovation output Equation Dependent variable LISPE			Productivity Equation Dependent variable LLPPE		
	OECD	IGNORED	SENS	OECD	IGNORED	SENS
$\ln(I/E)$	0.169***	0.167***	0.182***			
	(0.024)	(0.023)	(0.024)			
$\ln(\hat{S}/E)$				0.301***	0.223***	0.299***
				(0.028)	(0.034)	(0.028)
PR (process)	0.098	0.082	0.063	-0.020	-0.011	-0.030
	(0.083)	(0.077)	(0.077)	(0.032)	(0.033)	(0.037)
G1	0.186**	0.127		0.115***	0.041	
	(0.093)	(0.084)		(0.040)	(0.034)	
G2			0.216**			0.171***
			(0.096)			(0.043)
$\ln E$	0.006	-0.133***		0.073***	0.056**	
	(0.030)	(0.048)		(0.013)	(0.022)	
$\ln \hat{E}$			-0.043			0.021
			(0.046)			(0.020)
$\ln(\hat{\hat{E}}/E)$	0.745***			0.192*		
	(0.2228)			(0.108)		
$\ln \hat{\hat{E}}$			0.031			0.025*
			(0.028)			(0.013)
$\ln(M/E)$	0.059***			0.031***		
	(0.019)			(0.012)		
$\ln M$			0.032*			0.032
			(0.023)			(0.023)
C7 (client)	-0.108		-0.079			
	(0.114)		(0.114)			
C8 (supplier)	0.253**		0.226**			
	(0.109)		(0.014)			
C9 (private)	-0.137		-0.145			
	(0.104)		(0.104)			
C10 (public)	0.031		0.011			
	(0.104)		(0.103)			
H3 (cost)		-0.392**				
		(0.152)				
H1 (knowl.)		0.534**				
		(0.221)				
H2 (market)		-0.111				
		(0.196)				
C3 (inward)		1.235**				
		(0.359)				
C4 (outward)		0.664*				
		(0.393)				
C5 (internat.)		-0.133				
		(0.187)				
C6 (national)		-0.847***				
		(0.278)				
SM (SMEs)		-0.358**			-0.034	
		(0.159)			(0.069)	
D (industry)	Included	Included	Included	Included	Included	Included
CONST	Included	Included	1,052	Included	Included	Included
OBS	1,052	1,052	Included	1,052	1,052	1,052

Notes: Standard errors within parentheses. ***significant at 1%; ** significant at 5% and * significant at 10% level

Table 6: Labour productivity equation, selected variables

	Productivity Equation Dependent variable LLPPE	
	Sales	Value added
$\ln \hat{S}/E)$ and $\ln(\hat{V}/E)$	0.363*** (0.093)	0.378*** (0.092)
Predicted mill ratio	0.138 (0.089)	0.150* (0.081)

Notes: The table shows selected variables from the last step in the four-equation CDM-model. The covariates are identical to those included in equations 4.3. Standard errors within parentheses. ***significant at 1%; ** significant at 5% and * significant at 10% level

APPENDIX I: Variable definitions

Panel AI: Original variables

Output variables	Specification	Input variables	Specification
S	Sales value	E	Employment (total labour)
$\hat{S}$	Innovation sales	$\hat{E}$	Ordinary labour, with less than 3 years university education
V	Value added	$\hat{\hat{E}}$	Knowledge labour with 3 years university education or more
$\hat{V}$	Value added based on innovations	M	Investment in machinery equipment
		I	Innovation expenditure (investment)

Notes: (a) Employment figures from the CIS survey in all three specifications

Panel BI: Quantitative variables

Variable	**Definition**
$\ln(\hat{S}/E)$	Log [(fraction of sales related to innovation*total sales)/(total employment)]
$\ln(\hat{V}/E)$	Log [(fraction of value added related to innovation*total value added)/(total employment)]
$\ln(S/E)$	Log sales per employee
$\ln(V/E)$	Log value added per employee
$\ln E$	Log employment (firms with 10 or more employees according to the CIS-survey)
$\ln \hat{E}$	Ordinary labour: Log employment with les than 3 years university education
$\ln \hat{\hat{E}}$	Knowledge-intensive labour : Log employment with a university education 3 years or more
$\ln(\hat{\hat{E}}/E)$	Fraction of employment with a university education, 3 years or more
$\ln M$	Log gross investment in machinery and equipment
$\ln(M/E)$	Log gross investment in machinery and equipment per employee
$\ln(I/E)$	Log Innovation expenditures per employeee

Notes: (a) Employment figures from the CIS survey in all three specifications

Panel CI: Indicator variables

Variable	**Definition**
G1	Belongs to a domestically owned or a foreign owned enterprise group
G2 (MNE)	Belongs to a domestically or foreign owned multinational enterprise group
PA (Patent application)	Patent application to the domestic patent office
FM (Foreign market)	Exports to foreign customers
H1 (Hampering knowledge) [a]	Composite variable: (i) Lack of qualified personnel; (ii) Lack of information on technology; (iii) Difficulty in finding cooperation partners for innovation
H2 (Hampering market) [a]	Composite variable: (i) Lack on information about the markets; (ii) Market dominated by established enterprises; (iii) Uncertain demand for innovative goods or services.
H3 (Hampering cost) [a]	Composite variable: (i) Lack of funds within the enterprise or group; (ii) Lack of finances outside the enterprise: (iii) Innovation costs too high.
H4 (Hampering total)	Composite variable: (i) Knowledge; (ii) Market; (iii) Costs
C1 (Cooperation)	External collaboration on innovation
C2 (Cooperation index)	Composite variable reflecting the frequency if external collaboration on innovation (i) Within the own firms, or with (ii) Suppliers; (iii) Clients; (iv) Competitors; (v) Consultants, (vi) Universities; (vii) Government.
C3 (Inward cooperation)	Composite variable: (i) Suppliers; (ii) Clients; (iii) Competitors
C4 (Outward cooperation)	Composite variable: (i) Consultants, (ii) Universities and (iii) Government
C5 (International cooperation)	Composite variable: International cooperation partners as a fraction of all innovation partners
C6 (National cooperat.)	Composite variable: National collaboration partners
C7 (Client cooperation)	External collaboration on innovation with clients
C8 (Supplier cooperat.)	External collaboration on innovation with suppliers
C9 (Private cooperat.)	External collaboration on innovation with private actors
C10 (Public cooperat.)	External collaboration on innovation with public actors
PR (process innovation)	Dummy indicating process innovation
SM (SME)	The firm has less than 250 emploees

Notes: (a) Employment figures from the CIS survey in all three specifications. (a) factors hampering innovation activities or projects or influencing a decision not to innovate with a high or medium degree of importance.

Appendix II: Data treatment

Panel AII: The original data has been treated in the following way:

Treatment of the sample size	Number of observations
1. Original matched data	3108
2. Dropped if Nace2<15	-35
3. Dropped if Nace2 = 38-49	-116
4. Dropped if Nace2>74	-2
5. Dropped if ADM value added=0	-68
6. Dropped if ADM employment <6	-20
7. Dropped if CIS turnover=0	-21
Used data	**2,842**

Panel BII: Treatment of the data variables

Treatment of the data variables	Number of changes
Innovation expenditures> turnover x 2 =turnover x 2	8
Physical investment> turnover x 2 = turnover x 2	10
Value added>turnover=turnover x 0.8	104

Appendix III: Selected cross sectional estimates from the literature

Panel AIII: The Elasticity of productivity with respect innovation output, selected studies

Study	Elasticity	Sample (Innovative firms in the last step of the CDM-model)
Crépon, Duguet and Mairesse (1998)	0.10	4,164 French manufacturing firms observed in national innovation surveys 1986-1990
Lööf and Heshmati (2002)	0.13	354 Swedish manufacturing firms observed in Community Innovation Survey 2 (1997)
Lööf (2004)	0.16	330 Swedish knowledge intensive firms, observed in a national innovation survey (1999)
Lööf (2004)	0.14	220 Swedish business services observed in a national innovation survey (1999)
Ebersberger and Lööf 2005	0.40	429 Danish manufacturing firms belonging to a group observed in Community Innovation Survey 3 (2001)
Ebersberger and Lööf 2005	0.20	515 Finnish manufacturing firms belonging to a group observed in Community Innovation Survey 3 (2001)
Ebersberger and Lööf 2005	0.06	1,119 Norwegian manufacturing firms belonging to a group observed in Community Innovation Survey 3 (2001)
Ebersberger and Lööf 2005	0.22	694 Swedish manufacturing firms belonging to a group observed in Community Innovation Survey 3 (2001)
Janz, Lööf and Peters (2004)	0.27	352 German manufacturing firms observed in Community Innovation Survey 3, (2001)
Janz, Lööf and Peters (2004)	0.29	206 Swedish manufacturing firms observed in Community Innovation Survey 3, (2001)

Panel BIII: The Elasticity of innovation output with respect to innovation input , selected studies

Study	Elasticity	Sample (Innovative firms in the last step of the CDM-model)
Crépon, Duguet and Mairesse (1998)	0.43	4,164 French manufacturing firms observed in national innovation surveys 1986-1990
Lööf and Heshmati (2002)	0.29	354 Swedish manufacturing firms observed in Community Innovation Survey 2 (1997)
Lööf (2004)	0.41	330 Swedish knowledge intensive firms, observed in a national innovation survey (1999)
Lööf (2004)	0.53	220 Swedish business services observed in a national innovation survey (1999)
Ebersberger and Lööf 2005	0.44	429 Danish manufacturing firms belonging to a group observed in Community Innovation Survey 3 (2001)
Ebersberger and Lööf 2005	0.33	515 Finnish manufacturing firms belonging to a group observed in Community Innovation Survey 3 (2001)
Ebersberger and Lööf 2005	-0.36	1,119 Norwegian manufacturing firms belonging to a group observed in Community Innovation Survey 3 (2001)
Ebersberger and Lööf 2005	0.53	694 Swedish manufacturing firms belonging to a group observed in Community Innovation Survey 3 (2001)
Janz, Lööf and Peters (2004)	0.49	352 German manufacturing firms observed in Community Innovation Survey 3, (2001)
Janz, Lööf and Peters (2004)	0.61	206 Swedish manufacturing firms observed in Community Innovation Survey 3, (2001)