<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper published in *Economics of Innovation and New Technology*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the original published paper (version of record):

Lööf, H. (2017)

A new approach to estimation of the R&D-innovation-productivity relationship

Economics of Innovation and New Technology, 26(1-2): 121-133

<https://doi.org/10.1080/10438599.2016.1202515>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-240106>

A new approach to estimation of the R&D–innovation–productivity relationship

Christopher F. Baum, Hans Lööf, Pardis Nabavi & Andreas Stephan

To cite this article: Christopher F. Baum, Hans Lööf, Pardis Nabavi & Andreas Stephan (2017) A new approach to estimation of the R&D–innovation–productivity relationship, *Economics of Innovation and New Technology*, 26:1-2, 121-133, DOI: [10.1080/10438599.2016.1202515](https://doi.org/10.1080/10438599.2016.1202515)

To link to this article: <https://doi.org/10.1080/10438599.2016.1202515>

Published online: 14 Jul 2016.

Submit your article to this journal [↗](#)

Article views: 473

View Crossmark data [↗](#)

Citing articles: 5 View citing articles [↗](#)

A new approach to estimation of the R&D–innovation–productivity relationship

Christopher F. Baum^a , Hans Lööf^b , Pardis Nabavi^b and Andreas Stephan^c

^aDepartment of Economics, Boston College and Department of Macroeconomics, DIW Berlin, Germany; ^bDepartment of Industrial Economics and Management, Royal Institute of Technology, Stockholm, Sweden; ^cJönköping International Business School and Ratio Institute, Stockholm, Sweden

ABSTRACT

We apply a generalized structural equation model approach to the estimation of the relationship between R&D, innovation and productivity that focuses on the potentially crucial heterogeneity across sectors. The model accounts for selectivity and handles the endogeneity of this relationship in a recursive framework which allows for feedback effects from productivity to future R&D investment. Our approach enables the estimation of the different equations as one system, allowing the coefficients to differ across sectors, and also permits us to take cross-equation correlation of the errors into account. Employing a panel of Swedish manufacturing and service firms observed in three consecutive Community Innovation Surveys in the period 2008–2012, our full-information maximum likelihood estimates show that many key channels of influence among the model's components vary meaningfully in their statistical significance and magnitude across six different sectors based on the OECD classification on technological and knowledge intensity. These results cast doubt on earlier research which does not allow for sectoral heterogeneity.

ARTICLE HISTORY

Received 31 May 2015
Accepted 30 May 2016

KEYWORDS

R&D; innovation; productivity; generalized structural equation model; community innovation survey

JEL CLASSIFICATION

C23; L6; O32; O52

1. Introduction

In the empirical area of economics of innovation, a large and growing number of papers study the R&D–innovation–productivity (RIP) relationship at the micro level. Many of the challenges of estimating the RIP relationship with any reasonable precision are well known and related to issues such as specification of the empirical model, econometric methods, the data-generating process, and the reliability of the observed variables. A largely overlooked challenge, however is to accommodate the large degree of heterogeneity within the economy.

This study aims to contribute to a deeper understanding of the RIP relationship by applying a new approach to estimation of the process from the decision to engage in innovation to its final impact on productivity while allowing for sectoral heterogeneity.

The paper ‘Patents and R&D at the Firm Level: A First Look’ by Pakes and Griliches (1984) represents an important milestone in the modern research on the link between R&D, innovation and productivity by introducing a general model for the relationship. Crepon, Duguet, and Mairesse (1998) advance the Pakes and Griliches approach by formulating a recursive econometric approach that describes the process that goes from new ideas to economic growth. This approach is commonly labeled as the CDM model, incorporating a generalized Tobit model to handle the selectivity issue of observing innovation output only for a subsample of firms and a two-stage estimation procedure

to account for simultaneity in the system of equations. First, the paper estimates reduced form coefficients in each of the equations separately using a maximization approach, and in the second stage it applies the minimum distance asymptotic least-squares estimator to retrieve the structural parameters. The model is estimated as a system allowing for arbitrary cross-correlation among the disturbances.

A drawback of the original CDM paper is the cross-sectional nature of the data and estimates. This prevents the possibility of studying possible feedback effects and dynamics in the RIP linkages. More recently, Aw, Roberts, and Xu (2011) propose an approach that focuses on the firm's demand for R&D and the expected long-run payoff to the firm of undertaking R&D investment. Peters et al. (2013) and Peters, Roberts, and Vuong (forthcoming) show that this approach has potential to further develop the CDM literature by linking the firm's R&D investment, innovation and productivity in a dynamic framework.

This paper estimates the link between the key variables of the model as a recursive system of equations using maximum likelihood. Specifically, we estimate the RIP relationship in the context of a generalized structural equation model (GSEM) using the full-information maximum likelihood (FIML) estimator. This enables the estimation of the entire CDM model as one system, allowing the coefficients to differ across technology and knowledge sectors, and also permits us to take cross-equation correlation of the errors into account. We consider the importance of dynamics in this relationship and the potential for allowing firm performance to feed back to the level of future R&D investment.

One of the empirical attractiveness of the CDM approach is that it accommodates the issues of selectivity and endogeneity in estimating the R&D–innovation–productivity relationship. Following this tradition, we account for the first issue by adding a selection equation to the system of equations. This selection equation models the decision of firms to innovate, and this decision might be correlated with innovation outcomes. We formulate the original CDM model as a recursive system of equations using GSEM, and add a latent variable that accounts for correlations due to unobserved factors across equations which resolves the issues of endogeneity.

During the past decade, the CDM model has become a workhorse for micro-econometric productivity analysis based on Community Innovation Survey (CIS) data and similar firm-level information. CIS surveys contain information that lends itself unusually well to being analysed with a CDM approach. Studies based on CIS data on more than 40 countries over the last decades have contributed to a deeper insight into the micro-foundations of innovation. The potential of the survey data significantly increases when it is merged with official register data to produce a broader set of firm and employee characteristics for the observed units.

In this paper, we use more than nearly 12,000 firm-level observations from three consecutive Swedish CIS surveys over the period 2008–2012. Applying a unique firm identifier, the survey data are merged with data provided by Statistics Sweden and Eurostat on accrual accounts, domestic market share, imports, human capital, location and patents to implement the estimation framework.

Our measures of the influence of R&D investment on innovation sales and of innovation sales on labor productivity are generally in line with the empirical literature. At the same time, we find significant evidence of heterogeneity across technology and knowledge sectors in their magnitudes and significance. The impact of other explanatory factors on the key variables also differs across sectors, with significant effects in some sectors and not others. These results cast doubt on earlier research which does not allow for sectoral heterogeneity.

The rest of the paper is organized as follows. Section 2 briefly describes the methodology, while Section 3 presents the empirical data. Section 4 deals with the model specification and Section 5 reports the estimation results. Section 6 concludes and suggests areas for further research.

2. Estimation methodology

Our estimation approach is based on the GSEM of Rabe-Hesketh, Skrondal, and Pickles (2004). This framework allows for several features which are applicable in the context of our research. A detailed

discussion of these aspects of the GSEM framework is provided by Roodman (2011) in relation to his `cmp` routine, an earlier implementation of GSEM. These models are based on the generalized linear model (GLM) framework. Stata's GSEM extends that framework to incorporate multiple equation systems and latent variables.

We formulate the original CDM model as a recursive system of equations using GSEM including latent variable created by factor analytical methods representing covariances between observed dependent variables. The latent variable captures a set of unobserved factors that may be relevant in each equation where it appears, mitigating omitted-variables bias. IV techniques are used to deal with not only endogeneity but omitted-variables bias and measurement error in the regressors (e.g. arising from proxies). In that sense the presence of the latent variable deals with the fact that there are likely to be additional factors, beyond the observables, to be considered in the model. The coefficient of the latent variable is normalized to 1 in one equation, as we cannot otherwise define its scale, and estimated in the other.

First, we implement a selection equation which evaluates the likelihood that a firm will engage in innovative activity, and combine it with three linear regression equations in what has been termed a 'mixed-process' model, incorporating both continuous and censored responses. The data entering the selection equation comprise the full sample, while the data in subsequent equations are limited to those firms for which we have measures of innovation. The GSEM framework allows different observations to enter each equation in the model, with the available observations being used to estimate each equations' parameters.

The three subsequent equations involve endogeneity, but of a particular nature which may be expressed as a recursive or triangular equation system. The FIML estimates produced by GSEM are capable of handling this form of simultaneity. A maximum likelihood estimator of a seemingly unrelated equation (SUR) system 'can consistently estimate parameters in an important subclass of mixed-process simultaneous systems: ones that are recursive, with clearly defined stages, and that are fully observed, meaning that endogenous variables appear on the right-hand side only as observed' (Roodman 2011, 174). This is precisely the context of our research question, in which a firm's current R&D intensity is hypothesized to influence its level of innovation sales, which is in turn hypothesized to influence its labor productivity.

The approach to estimate coefficients for the including sectors simultaneously within the recursive system using FIML allowing for cross-correlation of the disturbances is new to the CDM literature. While prior studies within this tradition have accounted for heterogeneity by industry dummies (OECD 2009) and separate regressions (Hall and Sena, *forthcoming*) our approach provides a methodology for estimation of all parameters – including those allowed to vary across sectors – in a single process. This potentially is a more efficient approach, as it allows estimation of the possible cross-equation correlations for each including sector between the R&D equation and the productivity equation, or between the selection part (Heckman model) and the IV part (innovation and productivity).

In contrast to the strategy of Crepon, Duguet, and Mairesse (1998) of including firms without R&D and/or innovation output in all stages of the model, many later CDM-studies using Community Innovation Survey apply the full sample only in the first stage of the model. The reason for this deviation from the original CDM is that only innovative firms are responding on the bulk of the questions in the questionnaire. They include extensive information on firm's product and process innovations, R&D and other knowledge investments, co-operation on innovation, the relative importance of various knowledge flows, obstacles to innovation and sales income from innovative products. Typically the innovative firms represent only the minority of the surveyed firms. In this study we only consider firm with both positive input and output in the second stage of the model.

Another difference between our study and the original CDM paper is that latter uses an asymptotic least-square (ALS) approach. While the ALS approach is basically a GMM estimator which is asymptotically efficient and normally distributed, but does not exploit the cross-equation correlations, GSEM is estimated using full-information maximum-likelihood. As the name GSEM indicates, it is a generalization of structural equation models in the same way as generalized linear models are a

generalization of simple linear models. It allows specifying both the link function between the dependent and the explanatory variables as well as specifying a distribution family for the response variable. Notably, each equation can have different link and distribution family specifications. In that respect, GSEM is similar to Roodman's CMP procedure, which models recursive systems of equations with mixed distributions of dependent variables (continuous, binary, censored, etc).

3. Data and summary statistics

We employ 11,923 Swedish firm-level data observations on 7083 unique firms from three consecutive CIS surveys, 2008, 2010 and 2012, covering the period 2006–2012. For all observed firms, we have access to supplementary information concerning both internal firm characteristics, human capital, patent applications, the local milieu of the firms and foreign trade relations. The data are provided by Statistics Sweden and the European Patent Office database, PATSTAT. From 2008, the CIS surveys are compulsory in Sweden and the response rate is around 85%. Only firms with 10 or more employees in the year they are surveyed are included in the study.

In order to specify the equations, we consider a number of factors that potentially affect firms' RIP relationship. The variables of main interest are R&D investment, innovation sales, and labor productivity. The variables are measured in intensity form, i.e. per worker. The definitions of the variables used are presented in Table 1.

In our data, the largest fraction of the observations (50.8%) refers to firms with no R&D and no sales income from innovative products. A smaller fraction (29.5%) consists of firms with positive R&D and positive innovation sales during the same year. The remainder of our sample corresponds to observations with positive R&D and no innovation sales (4.8%) and observations with no R&D but positive innovation sales (4.9%).

Table 2 presents the sample averages of the dependent and explanatory variables for the total of 7083 firms and the subsample of 2,487 firms that have both R&D expenditures and sales income from innovative products in the same year. We refer to this subsample as plus-two (P2) firms, as both their innovation inputs and outputs are positive. Among the P2 firms, 48.2% are observed in one CIS-

Table 1. Variable definitions.

Variable	Definition
PRP2	Dummy for 'plus-two': positive R&D and positive innovation sales
$\log rd$	Research and development per employee
$\log is$	Innovation sales per employee
$\log ip$	Value added per employee
$\log tfp$	TFP calculated by Levisohn and Petrin method
$\log L$	Total number of employees
$\log K$	Physical capital
$\log k$	Physical capital to labor ratio
Hc	Human capital (share with at least 3 years of university education)
Pat	Dummy for patents granted or patent applications filed
Mf	Dummy for selling in foreign market
Ms	Market share
$\log im$	Imports per employee
ImG7	Share of imports from G7 countries
Smr	Dummy for Stockholm metro region
DI	Domestic non-affiliated, independent firm
DG	Domestic group-affiliated firm
DMNE	Domestic MNE, members of a domestic multinational group
FMNE	Foreign MNE, members of a foreign multinational group
Sector	High-technology manufacturing (HT), medium-high technology (HMT) manufacturing,
dummy	
variables ^a	medium-low technology manufacturing (LMT), low technology manufacturing (LT), knowledge-intensive services (KIS) and other services (OS)

Notes: All variables are measured at the firm level. Lower case variables imply divided by employees.

^aAccording to Eurostat classification, http://ec.europa.eu/eurostat/cache/metadata/Annexes/htec_esms_an3.pdf.

Table 2. Summary statistics.

	(1) All firms		(2) Plus-two firms	
	Mean	SD	Mean	SD
Log rd	4.53	5.23	10.39	1.74
Log is	4.24	5.92	12.31	1.38
Log lp	13.17	0.84	13.26	0.55
lnTFP	13.40	0.57	13.45	0.59
rd ^a	51.97	481.93	126.53	816.40
is ^a	1299.64	43883.62	525.88	1873.84
lp ^a	645.09	1230.49	699.49	1547.16
tfp ^a	813.92	1998.65	898.17	3035.91
PRP2	0.29	0.46	1.00	0.00
Log L	3.80	1.32	4.23	1.46
L	158.17	708.87	279.72	1125.65
Log K	14.78	2.35	15.27	2.44
K ^a	33808.60	115498.45	53792.13	151162.45
hc	0.17	0.21	0.23	0.23
Pat	0.04	0.18	0.09	0.29
Mf	0.64	0.48	0.81	0.39
Ms	0.04	0.11	0.07	0.15
Log im	6.87	5.79	8.78	5.25
lmG7	0.24	0.34	0.32	0.35
Smr	0.22	0.41	0.23	0.42
DI	0.24	0.42	0.15	0.36
DG	0.31	0.46	0.25	0.43
DMNE	0.22	0.42	0.30	0.46
FMNE	0.23	0.42	0.30	0.46
HT	0.05	0.23	0.08	0.28
HMT	0.13	0.34	0.20	0.40
LMT	0.15	0.36	0.14	0.35
LT	0.21	0.41	0.19	0.40
KIS	0.21	0.40	0.24	0.43
OS	0.24	0.43	0.14	0.34
Observations	11,923		3511	
Unique firms	7083		2487	

^ain thousand Swedish krona. For variable definitions, see Table 1.

survey, 32.2% in two surveys and 19.6% in all three Swedish CIS surveys conducted during the period 2008–2012.

The plus-two firms are larger, with a higher intensity of physical and human capital, more patent applications, larger market share, more presence on foreign markets, higher imports and a larger import fraction from the G7 countries. Plus-two companies are more likely to be members of a multinational group, and they are also more likely to operate in the high-technology (HT) and knowledge-intensive sectors of the economy. No differences can be found in their propensity to be localized in metropolitan areas.

Table 3 breaks down the plus-two companies into six different sectors based on the Eurostat classification on technological and knowledge intensity.¹ These sectors are HT manufacturing, HMT manufacturing, medium-low technology manufacturing (LMT), low-technology (LT) manufacturing, KIS and other services (OS). The most striking findings in the summary statistics are a great uniformity in terms of the average value of innovation sales per employee as well as large differences in human capital intensity and patent applications. It is also notable that two out of three innovative service firms operate in foreign markets.

4. Model specification

In the empirical analysis, we first estimate the probability that the observed firm has both innovation input and innovation output. Innovation input is measured as R&D expenditures and

Table 3. Summary statistics by sector.

	HT	HMT	LMT	LT	KIS	OS
$\log rd$	11.37 (1.479)	10.55 (1.446)	10.08 (1.517)	9.956 (1.670)	10.83 (1.812)	9.728 (1.971)
$\log is$	12.38 (1.156)	12.39 (1.462)	12.24 (1.243)	12.20 (1.274)	12.24 (1.445)	12.53 (1.492)
$\log lp$	13.32 (0.592)	13.22 (0.485)	13.21 (0.432)	13.22 (0.495)	13.33 (0.658)	13.25 (0.543)
$\log tfp$	13.38 (0.611)	13.44 (0.501)	13.44 (0.498)	13.39 (0.559)	13.56 (0.708)	13.43 (0.587)
rd^a	340.8 (2617.7)	91.70 (162.8)	57.66 (87.06)	76.13 (243.9)	187.5 (528.1)	82.23 (212.6)
is^a	404.7 (431.3)	530.9 (1067.3)	418.9 (787.6)	428.9 (1057.3)	615.9 (3369.7)	682.0 (1175.0)
lp^a	711.5 (387.8)	617.3 (314.4)	590.7 (248.1)	641.8 (588.9)	895.5 (3033.0)	658.0 (478.3)
tfp^a	763.3 (454.2)	781.4 (633.5)	845.0 (2056.1)	796.5 (1027.7)	1204.2 (5812.6)	803.8 (575.6)
$\log L$	4.130 (1.503)	4.548 (1.485)	4.284 (1.333)	4.257 (1.451)	3.894 (1.390)	4.304 (1.536)
L	399.6 (1904.5)	364.0 (1210.7)	213.9 (615.2)	215.1 (439.2)	249.5 (1440.0)	300.8 (749.9)
$\log K$	15.03 (2.233)	15.78 (2.217)	16.04 (2.095)	15.84 (2.401)	13.92 (2.362)	15.42 (2.408)
K^a	42996.3 (135042.3)	55167.5 (140079.6)	54860.6 (136691.2)	77516.8 (177624.2)	29636.2 (127091.5)	66494.5 (179889.4)
Hc	0.29 (0.17)	0.14 (0.13)	0.08 (0.085)	0.12 (0.14)	0.48 (0.25)	0.17 (0.18)
Pat	0.19 (0.40)	0.15 (0.36)	0.093 (0.29)	0.057 (0.23)	0.06 (0.24)	0.025 (0.16)
Mf	0.95 (0.22)	0.93 (0.25)	0.88 (0.32)	0.80 (0.40)	0.74 (0.44)	0.60 (0.49)
Ms	0.048 (0.11)	0.077 (0.14)	0.089 (0.18)	0.11 (0.20)	0.026 (0.08)	0.048 (0.11)
$\log im$	11.50 (2.69)	11.14 (3.81)	10.71 (4.23)	9.05 (5.20)	4.63 (4.55)	8.75 (5.91)
$ImG7$	0.46 (0.31)	0.39 (0.31)	0.30 (0.31)	0.25 (0.31)	0.32 (0.41)	0.24 (0.32)
Smr	0.23 (0.42)	0.12 (0.32)	0.079 (0.27)	0.16 (0.37)	0.41 (0.49)	0.33 (0.47)
DI	0.13 (0.33)	0.11 (0.32)	0.15 (0.36)	0.16 (0.37)	0.17 (0.38)	0.16 (0.37)
DG	0.17 (0.38)	0.16 (0.37)	0.23 (0.42)	0.30 (0.46)	0.30 (0.46)	0.27 (0.44)
$DMNE$	0.36 (0.48)	0.33 (0.47)	0.33 (0.47)	0.26 (0.44)	0.30 (0.46)	0.27 (0.45)
$FMNE$	0.35 (0.48)	0.40 (0.49)	0.28 (0.45)	0.27 (0.44)	0.23 (0.42)	0.30 (0.46)
Observations	292	690	507	683	856	483
Unique Firms	191	451	369	515	637	400

Notes: HT is the high-technology manufacturing. HMT is the high-medium technology manufacturing. LMT is the low-medium technology manufacturing. LT is the low-technology manufacturing. KIS is the knowledge-intensive service. OS is the other services.

^ain thousand Swedish krona. For variable definitions, see Table 1.

innovation output is measured as sales income from product innovation. Those firm-year observations with positive innovation input and innovation output (the P2 firms) are then used to estimate the relationship between R&D and its determinants, how much of the firm differences in innovation can be attributed to R&D, and the relationship between labor productivity and innovation sales.

In the model, $PRP2$ is the observed dichotomous indicator for plus-two firms (the probability to be a P2 firm). The other dependent variables rd (innovation input), is (innovation output) and lp (labor

productivity) are measured per employee, with subscript i referring to firm, s to sector and t to time. Variables rd , is and lp are available only for the plus-two firms.

To analyze the link between R&D, innovation and productivity, we use the following econometric specification:

$$PRP2_{it} = \beta_{0s} + \beta_{1s} \log L_{it} + \beta_{2s} \log k_{it} + \beta_{3s} Ms_{it} + \beta_{4s} Mf_{it} + \beta_{5s} Smr_{it} + \beta_{6s} \log im_{it} + \beta_{7s} SD_{it} + \mathcal{L}_{it} + \varepsilon_{it}, \quad (1)$$

$$\log rd_{it} = \gamma_{0s} + \gamma_{1s} \log lp_{i,t-1} + \gamma_{2s} \log k_{it} + \gamma_{3s} Pat_{i,t-1} + \gamma_{4s} Ms_{it} + \gamma_{5s} Mf_{it} + \gamma_{6s} Smr_{it} + \gamma_{7s} ImG7_{it} + \gamma_8 \mathcal{L}_{it} + \gamma_{st} + \epsilon_{it}, \quad (2)$$

$$\log is_{it} = \delta_{0s} + \delta_{1s} \log rd_{it} + \delta_{2s} \log k_{it} + \delta_{3s} Ms_{it} + \delta_{4s} Smr_{it} + \delta_5 \mathcal{L}_{it} + \delta_{st} + \nu_{it}, \quad (3)$$

$$\log lp_{it} = \lambda_{0s} + \lambda_{1s} \log is_{it} + \lambda_{2s} \log L_{it} + \lambda_{3s} \log K_{it} + \lambda_{4s} hc_{it} + \lambda_{5s} Ms_{it} + \lambda_{6s} Smr_{it} + \lambda_{7s} Own_{it} + \lambda_{st} + \zeta_{it}, \quad (4)$$

where L is the firm size (number of employees), k is the physical capital per employee, Ms is the market share, Mf is a dummy variable for presence in foreign markets, Smr is a dummy variable for location in Stockholm, the capital metropolitan region in Sweden, im is imports per employee, SD are sector indicators, and $\mathcal{L}$ is a latent variable capturing unobserved factors.² In the second equation, rd is the research and development expenditures using the broad CIS definition, lp is the labor productivity, Pat is an indicator of positive number of patent applications in each year, and $ImG7$ is the import fraction from G7 countries.³ In Equation (3), is is innovation sales, and in Equation (4) hc is human capital, K is physical capital, OWN is corporate ownership structure, consisting of four different categories which can be domestic independent non-affiliated (DI), domestic group affiliated (DG), domestic MNE ($DMNE$), or foreign MNE ($FMNE$). The iid errors of the equations are denoted as ε , ϵ , ν , and ζ , respectively, and are assumed to follow a multivariate normal distribution. We also allow for contemporaneous correlation between the errors (ϵ , ν) and (ϵ , ζ). Equations (2)–(4) contain sector-year fixed effects, denoted as γ_{st} , δ_{st} , and λ_{st} . It should be noted that Equation (2) includes lagged labor productivity, which represents the feedback from firm performance (Equation (4)) to the firm's innovation efforts. The latent variable $\mathcal{L}$ in Equations (1)–(3) addresses the issue of omitted-variables bias in this triangular equation system by capturing an unobserved common component between those equations, which allows for contemporaneous correlations among these three disturbance processes.⁴ It should also be noted that $\mathcal{L}$ is constrained to have the same coefficients across sectors in Equations (2) and (3).

5. Regression results

Section 5.1 presents our main estimation results. The GSEM coefficients are reported in Tables 4–7, while a robustness check of the importance of the particular measurement of the productivity variable (labor productivity versus total factor productivity) is provided in the appendix (<http://fmwww.bc.edu/cfb/EINTappendix.pdf>). Section 5.2 provides a formal test of the homogeneity of coefficients for plus-two firms across sectors, with results are presented in Table 8.

5.1. GSEM estimates

The probit model (Equation (1)) results in Table 4 show that the likelihood of being a plus-two firm is positively associated with firm size, market share, foreign market presence, and imports. No difference in the probability of being a plus-two firm can be found between firms in the Stockholm metropolitan region and firms in other regions. The sector coefficients suggest that the propensity to be a plus-two firm is largest in HT manufacturing (the reference category), high-medium manufacturing and KIS.

Table 4. GSEM selection equation (Probit).

	(1)
PRP2	
Log L	0.04*** (0.00)
Log (K/L)	−0.00 (0.00)
Ms	0.12*** (0.04)
Mf	0.12*** (0.01)
Smr	0.01 (0.01)
Log im	0.01*** (0.00)
HMT ^a	−0.02 (0.02)
LMT ^a	−0.12*** (0.02)
LT ^a	−0.12*** (0.02)
KIS ^a	0.01 (0.02)
OS ^a	−0.18*** (0.02)
Latent variable <i>L</i>	1.0 (constrained)
Observations	11,923
Unique firms	7,083

Notes: Marginal effects reported. Year dummies and latent variable *L* included. For variable definitions, see Table 1. Robust standard errors reported.

^aThe reference category is HT firms.

* $p < .10$, ** $p < .05$, *** $p < .01$.

Table 5 reports the results from the research and development equation (Equation (2)). The effect of lagged labor productivity is positive across all six sectors, but the coefficient is significantly different from zero only for the other services sector category. The elasticity of research intensity with respect to physical capital per employee is significant or weakly significant in all but the low-tech and other services sectors. Firms' R&D expenditures are an increasing function of lagged patents in all sectors. One interpretation is that assurance of property rights increases the firm's willingness to invest in innovation projects. However, it can also be assumed that patents reflect knowledge stock due to past R&D investments. The importance of domestic market share (*Ms*) differs across sectors, whereas presence in foreign markets has uniformly positive impacts.

Recent micro-econometric studies suggest that internal R&D efforts and external knowledge are indispensable inputs into the generation of new technological knowledge. Antonelli and Colombelli (2015) show that external knowledge can induce complementarities and a self-reinforcing feedback leading to increased R&D investments. In Equation (2), we test for the importance of potential incoming knowledge on firms' R&D engagement in two ways. The first possible source of knowledge

Table 5. GSEM R&D equation.

Log rd	HT	HMT	LMT	LT	KIS	OS
$\log p_{t-1}$	0.26 (0.27)	0.25 (0.19)	0.08 (0.12)	0.20 (0.16)	0.44 (0.29)	0.92*** (0.27)
$\log k$	0.12* (0.07)	0.22*** (0.06)	0.13* (0.08)	0.06 (0.05)	0.14*** (0.05)	0.06 (0.05)
Pat_{t-1}	0.59*** (0.19)	0.87*** (0.12)	0.73*** (0.17)	1.08*** (0.23)	1.10*** (0.18)	0.87** (0.42)
<i>Ms</i>	0.20 (0.65)	0.77** (0.35)	0.28 (0.31)	−0.28 (0.34)	−2.16*** (0.70)	−0.30 (0.73)
<i>Mf</i>	1.11*** (0.36)	0.87*** (0.26)	1.00*** (0.23)	0.53*** (0.18)	0.73*** (0.15)	1.04*** (0.19)
<i>Smr</i>	0.42** (0.19)	0.22 (0.16)	0.19 (0.24)	0.14 (0.17)	−0.18 (0.13)	−0.18 (0.19)
<i>ImG7</i>	0.09 (0.25)	0.30* (0.17)	−0.25 (0.22)	0.19 (0.23)	0.26* (0.15)	0.02 (0.25)
Latent variable <i>L</i>	1.02*** (0.11)	1.02*** (0.11)	1.02*** (0.11)	1.02*** (0.11)	1.02*** (0.11)	1.02*** (0.11)
Observations	292	690	507	683	856	483
Unique firms	191	451	369	515	637	400

Notes: Sector-Year dummies and latent variable *L* included. Robust standard errors reported. For variable definitions, see Table 1.

* $p < .10$, ** $p < .05$, *** $p < .01$.

Table 6. GSEM Innovation sales equation.

Log is	HT	HMT	LMT	LT	KIS	OS
Log rd	0.33*** (0.13)	0.41*** (0.11)	0.33*** (0.13)	0.36*** (0.12)	0.47*** (0.12)	0.37*** (0.12)
Log (K/L)	-0.09 (0.06)	-0.02 (0.06)	0.02 (0.06)	0.07* (0.04)	0.02 (0.04)	-0.14*** (0.05)
Ms	1.64*** (0.46)	-0.03 (0.47)	0.50* (0.30)	0.85*** (0.27)	1.19 (0.74)	0.52 (0.48)
Smr	-0.07 (0.17)	-0.20 (0.20)	0.45** (0.19)	-0.16 (0.15)	0.29*** (0.10)	0.14 (0.14)
Latent variable <i>L</i>	0.18* (0.09)	0.18* (0.09)	0.18* (0.09)	0.18* (0.09)	0.18* (0.09)	0.18* (0.09)
Observations	292	690	507	683	856	483
Unique Firms	191	451	369	515	637	400

Notes: Sector-Year dummies and latent variable *L* included. Robust standard errors reported. For variable definitions, see Table 1.
 * $p < .10$, ** $p < .05$, *** $p < .01$.

spillovers is local knowledge in the Stockholm area, which is Sweden's most important learning environment in terms of business diversity, workforce education and concentration of public and private R&D efforts. The second possible source of external knowledge is global spillovers, as captured by imports from the G7 countries.⁵ The estimated partial elasticities for location in the Stockholm metro region are positive and significantly different from zero only among HT manufacturing firms. The parameter estimate for this sector classification suggests that location in Stockholm is associated with about 50% more R&D per worker. Imports from the G7 countries have a positive effect on R&D expenditures among high-medium technology (HMT) manufacturing firms and KIS. On average we estimate that the marginal effect is about 30%, significant at the 10% level. The latent variable's coefficient capturing unobservable factors is positive and significantly different from zero in Equation (2), but cannot be distinguished from its value of unity in Equation (1).

Table 6 reports the GSEM estimates for Equation (3), innovation sales. In accordance with the original CDM estimates, the elasticity estimates for R&D expenditures are positive and highly significant across the six sectors, varying between 0.33 and 0.47. The effect of capital intensity is positive and significant (at the 10% level) only for LT manufacturing, and negative and significant for other

Table 7. GSEM Labor productivity equation.

log <i>lp</i>	HT	HMT	LMT	LT	KIS	OS
log <i>is</i>	0.13*** (0.03)	0.04** (0.02)	0.07*** (0.02)	0.05** (0.02)	0.10*** (0.03)	0.05*** (0.02)
log <i>L</i>	-0.24*** (0.06)	-0.21*** (0.03)	-0.10*** (0.04)	-0.19*** (0.04)	-0.15*** (0.04)	-0.10*** (0.03)
log <i>K</i>	0.11*** (0.02)	0.10*** (0.02)	0.05** (0.02)	0.13*** (0.02)	0.09*** (0.02)	0.06*** (0.02)
Hc	0.07 (0.20)	0.60*** (0.21)	0.85** (0.35)	1.01*** (0.21)	0.29*** (0.09)	0.62*** (0.22)
Ms	0.29 (0.32)	0.68*** (0.14)	0.31** (0.14)	0.43*** (0.12)	0.17 (0.24)	0.22 (0.23)
Smr	-0.04 (0.08)	0.03 (0.06)	0.14*** (0.05)	0.07 (0.05)	0.14*** (0.04)	-0.02 (0.05)
DG ^a	0.04 (0.12)	0.09 (0.06)	0.10** (0.05)	0.13*** (0.05)	0.12** (0.05)	0.18*** (0.07)
DMNE ^a	0.23** (0.12)	0.18** (0.07)	0.11* (0.06)	0.21*** (0.06)	0.06 (0.07)	0.19** (0.08)
FMNE ^a	0.47*** (0.13)	0.24*** (0.07)	0.16** (0.08)	0.17*** (0.07)	0.26*** (0.07)	0.27*** (0.08)
Observations	292	690	507	683	856	483
Unique Firms	191	451	369	515	637	400

Notes: Year dummies included. Robust Standard errors reported. For log(*L*) is equal to $\epsilon_L - 1$, where ϵ_L denotes the output elasticity of labor. For variable definitions, see Table 1.

^aReference category is non-affiliated firms (DI). Observe that the estimates.

* $p < .10$, ** $p < .05$, *** $p < .01$.

Table 8. Significant effects and tests of sectoral homogeneity: Tables 5–7.

Sector	HT	HMT	LMT	LT	KIS	OS	p-value
<i>R&D equation</i>							
$lp(t-1)$						+	0.047
$\log k$		+			+		0.023
$Pat(t-1)$	+	+	+	+	+	+	0.353
Ms		+			–		0.007
Mf	+	+	+	+	+	+	0.296
Smr	+						0.067
$lmG7$							0.426
<i>Innovation sales equation</i>							
R&D	+	+	+	+	+	+	0.032
$\log k$						–	0.016
Ms	+			+			0.164
Smr			+		+		0.025
<i>Labor productivity equation</i>							
is	+	+	+	+	+	+	0.013
$\log(L)$	–	–	–	–	–	–	0.065
$\log(K)$	+	+	+	+	+	+	0.041
Hc		+	+	+	+	+	0.005
Ms		+	+	+			0.287
DG			+	+	+	+	0.895
DMNE	+	+		+		+	0.503
FMNE	+	+	+	+	+	+	0.347
Smr	+	+		+		+	0.084

Notes: + significant positive at the 5%, – significant negative at the 5%. Observe that the estimate for $\log(L)$ is equal to $\epsilon_L - 1$.

services sectors. Innovation sales increase significantly with domestic market share for all manufacturing sectors except HMT. No link between market size and innovation output intensity can be found among service firms. Stockholm has by far the largest concentration of KIS companies in Sweden. In our sample, 4 out of 10 companies in this sector are located in the capital region. The estimation results suggest that KIS firms in Stockholm are significantly more innovative than their counterparts in other regions. Among the other sectors, we find no corresponding location effect except for low-medium technology firms. The latent variable's coefficient is significantly different from zero at the 10% level and clearly different from its value of unity in Equation (1), suggesting that the omitted factors have a smaller impact on innovation sales.

The final link in the CDM model is captured by Equation (4), with the estimation results presented in Table 7. In contrast to the original CDM approach, where innovation sales were measured as a share of total sales, our coefficients represent the impact of an increase of innovation sales per worker. The estimates are positive and significant across the six different sectors, while the magnitudes of these elasticity estimates are largest in the most knowledge-intensive sectors of high-tech manufacturing and KIS. The estimates for the factors of production L (labor) and K (physical capital) are in accordance with the Schumpeterian literature using the Cobb–Douglas technology. The human capital coefficient is positive and highly significant for all sectors except high-tech manufacturing. A tentative interpretation of the hc estimate is that its variance within the HT sector is relatively small, as most firms would require a highly educated workforce in that sector. Thus, there is little variation in hc that can explain differences in labor productivity, particularly given that this sector has the smallest number of firms in our data. Market share is more closely linked to productivity in manufacturing sectors, however not significantly positive for high-tech firms. Location has mixed effects, while foreign multinationals are uniformly more productive than domestic firms.

Our estimation approach allows us to model cross-equation covariances among equations' errors. One of those covariances, between R&D and innovation sales, is significantly different from zero, corresponding to a correlation of -0.34 . The other modeled covariance, between R&D and labor productivity, is negative but not significantly different from zero. These cross-equation effects could not be analyzed in a single-equation approach, and illustrate the potential importance of common shocks across elements of the R&D–innovation–productivity relationship.

As a robustness check, we replace labor productivity with total factor productivity (TFP) using the approach suggested by Levinsohn and Petrin (2003). The main results are consistent with the estimates reported in Tables 4–7, although there are some variations which can be linked to the different construction of the two measures. In order to save space, we do not report these results, but they are available upon request.

5.2. Test of sectoral homogeneity

The main objective of our study is to test for homogeneity of the effects of the key variables in the CDM model across the four manufacturing sectors and the two service sectors in our study. If the hypothesis of homogeneity is rejected, the results cast doubt on earlier research which does not allow for heterogeneity across sectors with different levels of technology and knowledge intensity.

Table 8 is organized in three sections. The first section tests homogeneity of the determinants of R&D-intensity, the second section tests the homogeneity assumption of the determinants to innovation sales, and the third section tests the homogeneity of the explanatory factors to productivity. The plus (+) and minus (–) signs give an overview of the significant effects at the 5% level presented in Tables 5–7, and the p -value is that of an F -test for equality of the coefficients across sectors, providing a formal significance test of the homogeneity hypotheses.

Looking first at the seven determinants of R&D reported in Table 5, it is shown that the null hypothesis of homogeneity of coefficients across sectors is rejected for lagged labor productivity, physical capital, market share and metropolitan location, with all but the last significant at the 5% level.

The GSEM estimates in Table 6 suggest that the elasticity of innovation sales is a positive and highly significant increasing function of R&D intensity. Despite the similarities of the estimates, a formal test of homogeneity across sectors is rejected in the second section of Table 8. The mixed results for capital intensity and location presented in Table 6 also reject the hypothesis on homogeneity across sectors.

The productivity equation reported in Table 7 shows that the elasticity of labor productivity with respect to innovation output is positive and highly significant across sectors. The size of the estimates is within the range 0.04–0.13. A formal test of homogeneity across sectors provided in Table 8 shows that homogeneity across sectors is rejected for innovation output, the coefficients of labor (at 10%), capital, human capital, and location (at 10%). In contrast to the other productivity determinants, we find that foreign multinationals are uniformly more productive than domestic firms.

In summary, in each of the CDM equations, we find strong and convincing evidence of sectoral heterogeneity in the key coefficients linking components of the model, as well as in other explanatory factors. This implies that constraining the estimates across sectors would be a clear misspecification of these relationships.⁶

6. Concluding remarks

During the past decade, the CDM model has become a workhorse for micro-econometric productivity analysis based on Community Innovation Survey (CIS) data and similar firm-level information. The model formulates the process from decision to engage in R&D to productivity performance as a recursive system of equations allowing for arbitrary correlations among disturbances in the four equations system.

While prior implementations of the CDM model have generally estimated the model in several steps, our approach estimates all parameters in a single process. We apply a general structural equation model (GSEM), estimating approach that considers the issues of selectivity of endogeneity in line with the CDM literature. We also take into account heterogeneity in the research–innovation–productivity relation across various branches of the economy, and the model allows for feedback effects from productivity to future R&D investment.

We first implement a selection equation which evaluates the likelihood that a firm will engage in innovative activity, and combine it with three linear regression equations in a mixed process model, incorporating both continuous and censored responses. The three subsequent equations involve endogeneity, but of a particular nature which may be expressed as a recursive or triangular equation system. The FIML estimates produced by GSEM are capable of handling this form of simultaneity.

The approach to estimate coefficients for the including sectors simultaneously within the recursive system using FIML allowing for cross-correlation of the disturbances is new to the CDM literature. While prior CDM literature has accounted for heterogeneity by industry dummies and separate regressions our approach provides a methodology for estimation of all parameters – including those allowed to vary across sectors – in a single process. This potentially is a more efficient approach, as it allows estimation of the possible cross-equation correlations for each including sector between the R&D equation and the productivity equation, or between the selection part (Heckman model) and the IV part (innovation and productivity).

We apply the GSEM approach on about 12,000 observations on more than 7,000 individual firms from three consecutive Community Innovation Surveys in Sweden in the period 2008–2012, complemented with register information from Statistics Sweden.

By estimating a sectoral equation system including all elements of the relationship between R&D, innovation and productivity, we test hypotheses on the importance of sectorial differences of the effects of explanatory factors. In the R&D equation, the results show that the null hypothesis of homogeneity of coefficients across six broad sectors is rejected for lagged labor productivity. In the innovation equation, the GSEM estimates suggest that the elasticity of innovation sales is a positive and highly significant increasing function of R&D intensity. Despite the similarities of the estimates, a formal test of homogeneity across sectors is rejected. The productivity equation reports that the elasticity of labor productivity with respect to innovation output is positive and highly significant across sectors. Similar to the other equations, a formal test of their homogeneity across sectors clearly rejects that hypothesis.

In summary, in each of the CDM equations, we find strong and convincing evidence of sectoral heterogeneity in the key coefficients linking components of the model, as well as in other explanatory factors. The results show that many key channels of influence among the model's components meaningfully differ in their statistical significance and magnitude across sectors defined by different technology and knowledge levels. This implies that constraining the estimates across sectors would be a clear misspecification of these relationships. The results cast doubt on earlier research which does not allow for sectoral heterogeneity.

An overall policy conclusion from our study is that policy makers should account for heterogeneity across sectors when implementing programs and instruments to address potential market failures regarding R&D and innovation. In future research, cross-country comparisons at the aggregate and industry levels, incorporating dynamics, in this methodological framework should prove fruitful.

Notes

1. http://ec.europa.eu/eurostat/cache/metadata/Annexes/htec_esms_an3.pdf
2. As the scale of the latent variable is arbitrary, its coefficient and variance must be normalized in one equation to allow its magnitude to be estimated. Its coefficient in Equation (1) is set to 1.0, and its coefficients in Equations (2) and (3) is constrained to be equal across sectors.
3. In the PRP2 equation, we include import intensity among the determinants of R&D engagement, while the R&D equation consider the relationship between imports from the G7 countries and the magnitude of R&D. Both variables are assume to capture global knowledge spillovers. Following prior literature, we assume that imports from G7 countries represent a particular category of knowledge spillovers since this group of countries accounts for the vast majority of global R&D. In line with ? we assume that knowledge spillovers induce complementarities in firms' R&D efforts. Cohen and Levinthal showed that knowledge spillovers may increase equilibrium R&D investment.

4. The latent variable was not included in Equation (4) due to convergence issues.
5. Andersson and Löf (2009) find that the value per kilogram (a rough indicator of the quality of trade flows) of Swedish import flows from the G7 area by firms in the same industry is higher than the average for the firms' import flows from all countries.
6. Thanks to international harmonization of the CIS data and international standards of business register data, our GSEM approach and the particular empirical specification used in this paper are applicable on both individual country studies as well as on cross-country comparisons. Although we think that our main results can be generalized to other countries, idiosyncratic factors might influence sign, size and statistical significance of the estimated relationships differently across countries.

Acknowledgments

We thank seminar participants at London, Paris and Maastricht workshops for helpful comments and suggestions. Furthermore, we thank participants at conferences: Economics of Innovation and Patents at ZEW Mannheim 2015, EARIE Muenchen 2015, EcoMod Boston 2015 and Deutsche Statistische Gesellschaft Hamburg 2015 for insightful comments. We thank four reviewers for comments. All remaining errors are our own.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by VINNOVA (Cesis Etapp II).

ORCID

Christopher F. Baum <http://orcid.org/0000-0003-4766-3699>

Hans Löf <http://orcid.org/0000-0002-5871-8571>

Pardis Nabavi <http://orcid.org/0000-0003-2420-5660>

Andreas Stephan <http://orcid.org/0000-0001-5776-9396>

References

- Andersson, Martin, and Hans Löf. 2009. "Learning-by-Exporting Revisited: the Role of Intensity and Persistence." *The Scandinavian Journal of Economics* 111 (4): 893–916.
- Antonelli, Cristiano, and Alessandra Colombelli. 2015. "External and Internal Knowledge in the Knowledge Generation Function." *Industry and Innovation* 22 (4): 273–298.
- Aw, Bee Yan, Mark J. Roberts, and Daniel Yi Xu. 2011. "R&D Investment, Exporting, and Productivity Dynamics." *American Economic Review* 101 (4): 1312–1344.
- Crepon, Bruno, Emmanuel Duguet, and Jacques Mairesse. 1998. "Research, Innovation And Productivity: An Econometric Analysis At The Firm Level." *Economics of Innovation and New Technology* 7, 115–158. doi:10.1080/10438599800000031.
- Hall, Bronwyn H., and Vania Sena. *Forthcoming*. "Appropriability Mechanisms, Innovation and Productivity: Evidence from the UK." *Economics of Innovation and New Technology*. doi:10.1080/10438599.2016.1202513.
- Levinsohn, James, and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *The Review of Economic Studies* 70 (2): 317–341.
- Pakes, Ariel, and Zvi Griliches. 1984. *Patents and R&D at the Firm Level: A First Look*. Edited by Zvi Griliches. University of Chicago Press.
- Peters, Bettina, Mark J. Roberts, and Van Anh Vuong. *Forthcoming*. "Dynamic R&D Choice and the Impact of the Firm's Financial Strength." *Economics of Innovation and New Technology*.
- Peters, Bettina, Mark J. Roberts, Van Anh Vuong, and Helmut Fryges. 2013. "Estimating Dynamic R & D Demand: An Analysis of Costs and Long-Run Benefits Introduction - Linking R & D and Firm Performance." NBER Working Paper No. w19374.
- Rabe-Hesketh, Sophia, Anders Skrondal, and Andrew Pickles. 2004. "Generalized Multilevel Structural Equation Modeling." *Psychometrika* 69, 167–190.
- Roodman, David. 2011. "Fitting Fully Observed Recursive Mixed-Process Models with cmp." *Stata Journal* 11 (2): 159–206.