CESIS Electronic Working Paper Series

Paper No. 408

A New Approach to Estimation of the R&D-Innovation-Productivity Relationship

Christopher F Baum
Hans Lööf
Pardis Nabavi
Andreas Stephan

June, 2015

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

A New Approach to Estimation of the R&D-Innovation-Productivity Relationship

Christopher F Baum*, Hans Lööf†, Pardis Nabavi‡, Andreas Stephan§

May 27, 2015

Abstract

We evaluate a Generalized Structural Equation Model (GSEM) approach to the estimation of the relationship between R&D, innovation and productivity that focuses on the potentially crucial heterogeneity across technology and knowledge levels. The model accounts for selectivity and handles the endogeneity of this relationship in a recursive framework. Employing a panel of Swedish firms observed in three consecutive Community Innovation Surveys, our maximum likelihood estimates show that many key channels of influence among the model's components differ meaningfully in their statistical significance and magnitude across sectors defined by different technology levels.

Keywords: R&D, Innovation, Productivity, Generalized Structural Equation Model, Community Innovation Survey

JEL: C23, L6, O32, O52

*Department of Economics, Boston College and Department of Macroeconomics, DIW Berlin

†Department of Industrial Economics and management, Royal Institute of Technology, Stockholm

‡Department of Industrial Economics and management, Royal Institute of Technology, Stockholm

§Jönkoping International Business School

1 INTRODUCTION

There is a broad agreement in the literature that firms' productivity is driven by technological change. A large number of productivity studies at the micro level focus on the R&D–innovation–productivity relationship, accounting for both observable and unobservable factors. Shortcomings associated with available data, statistical and econometric methods, and theoretically founded economic models make it difficult to estimate the relationship with any reasonable precision. Another challenging issue in the empirical area of economics of innovation studies is to accommodate the large degree of heterogeneity across sectors.

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 and a GMM approach to account for simultaneity. Most recently, Aw et al. (2011) propose a dynamic approach to the CDM framework that models firms' R&D investment taking into account market demand.

In this paper, we are using a unified estimation methodology which allows to model both the propensity to engage in innovation activities and the observable consequences for engaged firms. In contrast to much of the existing literature, we allow for complete flexibility of the estimated relationships across sectors with different technology and knowledge intensity. This allows us to identify meaningful differences across technology levels in the way that firms employ innovation inputs and generate innovation sales. Specifically, we estimate the R&D–innovation–productivity relationship in the context of a generalized structural equation model (GSEM) using a full-information maximum likelihood estimator. This enables the estimation of the entire CDM model as one system, allowing the coefficients to differ across sectors, and also allows 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 R&D investment.

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 analyzed 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 rises significantly when it is merged with official register data to produce a broader set of firm and employee characteristics for the observed units.

In some countries, the CIS surveys are mandatory, with the opportunity to study the same company over time based on a unique firm identifier. As the surveys have a set of questions that are similar across time, the CIS data are suitable for a panel data approach, which is capable of identifying effects that are not detectable in a pure cross-section. Depending on the stratification of sample and rate of response, CIS surveys may offer a possibility to compare firms across industries and regions. In this paper, we use the Swedish CIS survey with supplementary information concerning firm characteristics to implement the estimation framework.

Our main results provide support for the key elements of the CDM approach, yielding measures of the influence of R&D investment on innovation sales and of innovation sales on labor productivity generally in line with the original CDM values. At the same time, we find significant evidence of heterogeneity across technology and knowledge sectors in their magnitudes. The impact of other explanatory factors on the key variables also exhibits considerable differences across sectors, with significant effects in some sectors and not others. These results cast doubt on earlier research which does not allow for this heterogeneity.

The rest of the paper is organized as follows. Section 2 describes the methodology. Section 3 presents the empirical data and estimation results. Section 4 concludes and suggests areas for further research.

2 ESTIMATION METHODOLOGY

Our estimation approach is based on the generalized structural equation model (GSEM) of Rabe-Hesketh, S., and Pickles. (2004). This framework allows for several features which are applicable to 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.

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. This approach stands in contrast to earlier two-step methods of modeling selectivity. Second, 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.

Third, the three subsequent equations involve endogeneity, but of a particular nature which may be expressed as a recursive, or triangular equation system. The full-information maximum likelihood (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, p. 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.

Finally, by estimating a single equation system encompassing all elements of the research question, we are able to perform hypothesis tests which evaluate the importance of sectoral differences of the effects of explanatory factors. The test 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.

3 DATA AND RESULTS

3.1 Data and Summary statistics

We employ Swedish firm-level data 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, the local milieu of the firms and foreign trade relations. From 2008, the CIS surveys are compulsory in Sweden and the response rate is around 85 percent. 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' R&D-innovation-productivity relationship. Using the European Patent Office database, PATSTAT, we match information on patents with the firm identifier for the survey firms over the period 2006–2012. 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 definition of the variables used are presented in Table 1.

Table 2 presents the sample averages of the dependent and explanatory variables for the total of 7,083 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 firms, as both their innovation inputs and outputs are positive. 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 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 based on technological and knowledge intensity.[1] These sectors are high technology manufacturing (HT), medium-high technology manufacturing (HMT), medium-low technology manufacturing (LMT), low technology manufacturing (LT), knowledge intensive services (KIS) and other services (OS). The most striking findings in the summary statistics are a great

[1] http://ec.europa.eu/eurostat/cache/metadata/Annexes/htec_esms_an3.pdf

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 service firms operate in foreign markets.

3.2 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 (rd), and innovation output is measured as sales income from product innovation (is). Both variables are expressed in intensity form (per employee). Those firm-year observations with positive innovation input and innovation output are then used to estimate the relationship between rd and its determinants, how much of the sectoral differences in is can be attributed to rd, and the relationship between labor productivity (lp) and is.

The CDM approach addresses the two important issues of selectivity and endogeneity. We account for the first issue by adding the selection equation in the system estimator. Our GSEM approach encompasses a linear triangular systems with unobserved components, which resolves the issues of endogeneity. The estimated coefficients are allowed to differ across sectors defined above. In contrast to the original CDM model, we also investigate the possibility that the prior period's productivity could influence the level of R&D investment.

The CIS survey is structured in a way such that a filter question separates firms into innovators and non-innovators. In our paper, we use this filter to select firms into the plus-two category of those that have both positive innovation input and positive innovation output. In the model, $PRP2$ is the observed dichotomous indicator for plus-two firms. The other dependent variables rd (innovation input), is (innovation output) and lp (labor productivity) are measured as per-employee, with subscript i referring to firm, s to sector and t time:

$$
\begin{aligned}
PRP2_{it} &= \beta_0 + \beta_1 \log L_{it} + \beta_2 \log(K/L)_{it} + \beta_3 Ms_{it} + \beta_4 Mf_{it} + \\
&\quad \beta_5 Smr_{it} + \beta_6 \log Im_{it} + \beta_7 SD_{it} + \mathcal{L} + \varepsilon_{it} \qquad (1)
\end{aligned}
$$

$$
\begin{aligned}
\log rd_{ist} &= \gamma_0 + \gamma_1 \log lp_{is,t-1} + \gamma_2 \log(K/L)_{ist} + \gamma_3 Pat_{is,t-1} + \\
&\quad \gamma_4 Ms_{ist} + \gamma_5 Mf_{ist} + \gamma_6 Smr_{ist} + \gamma_7 ImG7_{ist} + \\
&\quad \gamma_8 \mathcal{L} + \gamma_i + \epsilon_{it} \qquad (2)
\end{aligned}
$$

$$
\begin{aligned}
\log is_{ist} &= \delta_0 + \delta_1 \log rd_{ist} + \delta_2 \log(K/L)_{ist} + \delta_3 Ms_{ist} + \\
&\quad \delta_4 Smr_{ist} + \delta_5 \mathcal{L} + \delta_i + \nu_{it} \qquad (3)
\end{aligned}
$$

$$
\begin{aligned}
\log lp_{ist} &= \lambda_0 + \lambda_1 \log is_{ist} + \lambda_2 \log L_{ist} + \lambda_3 \log K_{ist} + \lambda_4 hc_{ist} + \\
&\quad \lambda_5 Ms_{ist} + \lambda_6 Smr_{ist} + \lambda_7 OWN_{2-4,ist} + \lambda_i + \zeta_{it} \qquad (4)
\end{aligned}
$$

where L is firm size, K is physical capital, Ms is 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, SD are sector indicators, and $\mathcal{L}$ is a latent variable capturing unobserved factors. In the second equation, rd is research and development expenditures using the broad CIS definition, lp is labor productivity, Pat is a dummy for positive number of patent applications in each year, and $ImG7$ is the import fraction from G7 countries. In equation (3), is is innovation sales, and hc in equation (4) is human capital, OWN consists of four different ownership categories which can be Non-affiliated ($NAFF$), Domestic Affiliated ($DAFF$), Domestic MNE ($DMNE$), or Foreign MNE ($FMNE$). The idiosyncratic errors of the equations are denoted as ε, ϵ, ν, and ζ, respectively. We also allow for contemporaneous correlation between the errors (ϵ, ν) and (ϵ, ζ). The fixed effects of equations (2) to (4) are denoted as γ, δ, and λ. 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. $\mathcal{L}$ in equations (1), (2), and (3) addresses the issue of selectivity, as $\log rd$ and $\log is$ are measured only for the plus-two firms.

3.3 Results

In this section we present our estimation results. The probit model 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. The sector dummies suggest that the propensity to be a plus-two firm is largest in high technology manufacturing, high-medium manufacturing and knowledge-intensive services.

Table 5 reports the results from the research and development equation, with significant findings reported in Table 8. The effect of lagged labor productivity is positive across all six sectors, but significant only for other services. The effect of capital intensity is significant in all but the low-tech and other services. Firms' R&D expenditures are an increasing function of lagged patents in all sectors. The effects of market share (Ms) differ across sectors, whereas presence in foreign markets has uniformly positive effects. Location in the Stockholm metro region is only important in the high-tech sector, while import share from G7 countries has varying effects. The latent variable's coefficient is positive and significant, indicating the importance of unobserved factors. Formal tests of the homogeneity of coefficients across sectors are rejected for lagged labor productivity, market share, and location.

Table 6 reports the GSEM estimates for equation (3), innovation sales, with significant findings reported in Table 9. In accordance with the original CDM estimates, the elasticity estimates for R&D are positive and highly significant across the six sectors, varying between 0.33–0.47. Despite their similarities, a formal test of homogeneity across sectors is rejected. The effect of capital intensity is negative and significant for high-tech and other services sectors, while market share has a positive effect for high-tech and low-tech sectors. Location only appears important for low-medium tech and knowledge intensive services. Homogeneity across sectors is also rejected for capital intensity and location. The latent variable's coefficient is positive and weakly significant.

The final link in the CDM model is captured by equation (4), with the estimation results presented in Table 7 and significant findings summarized in Table 10. In contrast to the original CDM approach where the innovation sales were measured as a share of total sales, our coefficients represent the impact of an increase of innovation sales per worker. The magnitudes of these elasticity estimates are largest in the most knowledge-intensive sectors of high tech manufacturing and knowledge-intensive services. A formal test of their homogeneity across sectors clearly rejects that hypothesis. The estimates for the factors of production L (labor) and K (physical capital) are in accordance with the Schumpeterian literature using Cobb–Douglas technology.[2] The human capital coefficient is positive and highly significant for all sectors except high-tech manufacturing. Market share is more linked to productivity in manufacturing sectors, however not significantly positive for high-tech firms. Location has mixed effects. Homogeneity across sectors is also rejected for the coefficients of labor, capital, human capital, and location. In accordance with previous literature, we find that 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 significant, corresponding to a correlation of -0.35. The other modeled covariance, between R&D and labor productivity, is negative but not significantly

[2]Observe that in equation (4) the interpretation of the coefficient of $logL$ is $(\lambda_2 - 1)$, as the dependent variable is expressed in per employee terms (VA/L).

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.

In summary, in each of the CDM equations, we find strong evidence of 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.

4 CONCLUDING REMARKS

We evaluate a Generalized Structural Equation Model (GSEM) approach to the estimation of the relationship between R&D, innovation and productivity that focuses on the potentially crucial heterogeneity across technology and knowledge levels. We find that the key estimates are qualitatively similar to those reported in the seminal paper by Crepon, Duguet, and Mairesse (1998). Our empirical approach offers attractive possibilities to analyze micro data on firms' innovation activities in the context of selectivity and endogeneity. It is well designed to account for the particular nature of the measurements of innovation inputs, outputs and firm performance, capturing the key linkages between these key economic variables. In future research, cross-country comparisons at the aggregate and industry levels, incorporating dynamics, in this methodological framework should prove fruitful.

References

Aw, B. Y., M. J. Roberts, and D. Y. Xu (2011). R&D Investment, Exporting, and Productivity Dynamics. *American Economic Review 101*(4), 1312–44.

Crepon, B., E. Duguet, and J. Mairesse (1998). Research, Innovation And Productivity: An Econometric Analysis At The Firm Level. *Economics of Innovation and New Technology 7*, 115–158.

Pakes, A. and Z. Griliches (1984). *Patents and R & D at the Firm Level : A First Look*, Volume I. University of Chicago Press.

Rabe-Hesketh, A. S. S., and A. Pickles. (2004). Generalized multilevel structural equation modeling. *Psychometrika 69*, 167–190.

Roodman, D. (2011). Mixed-process, Estimating fully observed recursive Cmp, models with. *Stata Journal 11*(2), 159–206.

TABLES

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 lp	Labor productivity
Log L	Total number of employees
Log K	Physical capital
hc	Human capital (share with at least 3 years of university education)
Pat	Dummy for patents granted or patent applications filed
Mf	Dummy for foreign market presence
Ms	Market share
Log im	Imports per employee
ImG7	Share of imports from G7 countries
Smr	Dummy for Stockholm metro region

Table 2: Summary statistics

	(1)		(2)	
	All firms		Plus-two firms	
	mean	sd	mean	sd
Log rd	2.59	6.99	10.39	1.74
Log is	1.89	7.67	12.31	1.38
Log lp	13.17	0.84	13.26	0.55
PRP2	0.44	0.50	1.00	0.00
Log L	3.80	1.32	4.23	1.46
Log K	14.79	2.38	15.29	2.51
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
ImG7	0.24	0.34	0.32	0.35
Smr	0.22	0.41	0.23	0.42
NAFF	0.24	0.42	0.15	0.36
DAFF	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
High-Tech Manuf (HT)	0.05	0.23	0.08	0.28
Medium-High Tech Manuf (HMT)	0.13	0.34	0.20	0.40
Medium-Low Tech Manuf (LMT)	0.15	0.36	0.14	0.35
Low-Tech Manuf (LT)	0.21	0.41	0.19	0.40
Knowledge Intensive Services (KIS)	0.21	0.40	0.24	0.43
Other Services (OS)	0.24	0.43	0.14	0.34
Observations	11,923		3,511	
Unique Firms	7,083		2,487	

Table 3: Summary statistics by sector

	HT	HMT	LMT	LT	KIS	OS
Log rd	11.37	10.55	10.08	9.96	10.83	9.73
	(1.48)	(1.45)	(1.52)	(1.67)	(1.81)	(1.97)
Log is	12.38	12.39	12.24	12.20	12.24	12.53
	(1.16)	(1.46)	(1.24)	(1.27)	(1.45)	(1.49)
Log lp	13.32	13.22	13.21	13.22	13.33	13.25
	(0.59)	(0.48)	(0.43)	(0.49)	(0.66)	(0.54)
Log L	4.13	4.55	4.28	4.26	3.89	4.30
	(1.50)	(1.48)	(1.33)	(1.45)	(1.39)	(1.54)
Log K	15.04	15.81	16.06	15.88	13.95	15.46
	(2.27)	(2.29)	(2.15)	(2.48)	(2.45)	(2.49)
hc	0.29	0.14	0.08	0.12	0.48	0.17
	(0.17)	(0.13)	(0.085)	(0.14)	(0.25)	(0.18)
Pat	0.19	0.15	0.093	0.057	0.06	0.025
	(0.40)	(0.36)	(0.29)	(0.23)	(0.24)	(0.16)
Mf	0.95	0.93	0.88	0.80	0.74	0.60
	(0.22)	(0.25)	(0.32)	(0.40)	(0.44)	(0.49)
Ms	0.048	0.077	0.089	0.11	0.026	0.048
	(0.11)	(0.14)	(0.18)	(0.20)	(0.08)	(0.11)
Log im	11.50	11.14	10.71	9.05	4.63	8.75
	(2.69)	(3.81)	(4.23)	(5.20)	(4.55)	(5.91)
ImG7	0.46	0.39	0.30	0.25	0.32	0.24
	(0.31)	(0.31)	(0.31)	(0.31)	(0.41)	(0.32)
Smr	0.23	0.12	0.079	0.16	0.41	0.33
	(0.42)	(0.32)	(0.27)	(0.37)	(0.49)	(0.47)
NAFF	0.13	0.11	0.15	0.16	0.17	0.16
	(0.33)	(0.32)	(0.36)	(0.37)	(0.38)	(0.37)
DAFF	0.17	0.16	0.23	0.30	0.30	0.27
	(0.38)	(0.37)	(0.42)	(0.46)	(0.46)	(0.44)
DMNE	0.36	0.33	0.33	0.26	0.30	0.27
	(0.48)	(0.47)	(0.47)	(0.44)	(0.46)	(0.45)
FMNE	0.35	0.40	0.28	0.27	0.23	0.30
	(0.48)	(0.49)	(0.45)	(0.44)	(0.42)	(0.46)
Observations	292	690	507	683	856	483
Unique Firms	191	451	369	515	637	400

Table 4: GSEM selection equation

PRP2	(1)
Log L	0.22***
	(0.02)
Log (K/L)	-0.00
	(0.01)
Ms	0.43**
	(0.17)
Mf	0.59***
	(0.05)
Smr	0.05
	(0.05)
Log im	0.03***
	(0.00)
HMT[a]	-0.12
	(0.08)
LMT[a]	-0.54***
	(0.08)
LT[a]	-0.53***
	(0.08)
KIS[a]	0.00
	(0.08)
OS[a]	-0.82***
	(0.08)
Latent	Constraint
Observations	11,923
Unique Firms	7,083

Robust Standard errors reported

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[a] The reference category is HT.

Table 5: GSEM R&D equation

Log rd	HT	HMT	LMT	LT	KIS	OS
Log lp_{t-1}	0.27	0.26	0.08	0.20	0.44	0.92^{***}
	(0.27)	(0.19)	(0.13)	(0.17)	(0.30)	(0.27)
Log (K/L)	0.14^{**}	0.23^{***}	0.14^{*}	0.06	0.14^{***}	0.05
	(0.07)	(0.05)	(0.08)	(0.05)	(0.05)	(0.06)
Pat_{t-1}	0.57^{***}	0.86^{***}	0.73^{***}	1.07^{***}	1.07^{***}	0.86^{**}
	(0.19)	(0.12)	(0.17)	(0.23)	(0.19)	(0.42)
Ms	0.16	0.76^{**}	0.28	-0.28	-2.17^{***}	-0.34
	(0.65)	(0.35)	(0.31)	(0.34)	(0.70)	(0.74)
Mf	1.11^{***}	0.87^{***}	1.00^{***}	0.53^{***}	0.75^{***}	1.06^{***}
	(0.36)	(0.26)	(0.23)	(0.18)	(0.15)	(0.19)
Smr	0.42^{**}	0.22	0.18	0.14	-0.18	-0.17
	(0.19)	(0.16)	(0.24)	(0.17)	(0.13)	(0.19)
ImG7	0.08	0.29^{*}	-0.25	0.19	0.27^{*}	0.03
	(0.25)	(0.17)	(0.22)	(0.23)	(0.15)	(0.25)
Latent	1.05^{***}	1.05^{***}	1.05^{***}	1.05^{***}	1.05^{***}	1.05^{***}
	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)
Observations	292	690	507	683	856	483
Unique Firms	191	451	369	515	637	400

Robust Standard errors reported

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: GSEM Innovation sales equation

Log is	HT	HMT	LMT	LT	KIS	OS
Log rd	0.33***	0.41***	0.33***	0.36***	0.47***	0.38***
	(0.13)	(0.11)	(0.13)	(0.12)	(0.12)	(0.12)
Log (K/L)	-0.11*	-0.02	0.03	0.06	0.03	-0.14***
	(0.06)	(0.06)	(0.06)	(0.04)	(0.04)	(0.05)
Ms	1.69***	-0.03	0.49	0.87***	1.17	0.59
	(0.46)	(0.47)	(0.30)	(0.26)	(0.73)	(0.49)
Smr	-0.06	-0.20	0.45**	-0.18	0.28***	0.14
	(0.17)	(0.20)	(0.19)	(0.14)	(0.10)	(0.14)
Latent	0.18*	0.18*	0.18*	0.18*	0.18*	0.18*
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
Observations	292	690	507	683	856	483
Unique Firms	191	451	369	515	637	400

Robust Standard errors reported

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: GSEM Labor productivity equation

Log lp	HT	HMT	LMT	LT	KIS	OS
Log is	0.13***	0.05***	0.07***	0.05***	0.10***	0.05***
	(0.03)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Log L	-0.23***	-0.19***	-0.10***	-0.19***	-0.15***	-0.10***
	(0.06)	(0.03)	(0.04)	(0.04)	(0.03)	(0.02)
Log K	0.10***	0.08***	0.04**	0.13***	0.08***	0.05**
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
hc	0.07	0.59***	0.81**	0.97***	0.29***	0.62***
	(0.20)	(0.21)	(0.36)	(0.21)	(0.09)	(0.22)
Ms	0.31	0.71***	0.32**	0.46***	0.19	0.21
	(0.32)	(0.15)	(0.14)	(0.12)	(0.24)	(0.23)
Smr	-0.05	0.03	0.14***	0.06	0.14***	-0.02
	(0.08)	(0.06)	(0.05)	(0.05)	(0.04)	(0.05)
DAFF[a]	0.04	0.08	0.10**	0.13***	0.12**	0.18***
	(0.12)	(0.06)	(0.05)	(0.05)	(0.05)	(0.07)
DMNE[a]	0.24**	0.18***	0.11*	0.21***	0.06	0.19**
	(0.12)	(0.07)	(0.06)	(0.06)	(0.07)	(0.08)
FMNE[a]	0.48***	0.25***	0.17**	0.18***	0.27***	0.27***
	(0.13)	(0.07)	(0.08)	(0.07)	(0.07)	(0.08)
Observations	292	690	507	683	856	483
Unique Firms	191	451	369	515	637	400

Robust Standard errors reported

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[a] The reference category is non-affiliated firms (NAFF).

Table 8: R&D Equation

Sector	HT	HMT	LMT	LT	KIS	OS
LaborProductivity(t-1)						+
log(K/L)	+	+			+	
Patents(t-1)	+	+	+	+	+	+
MktShare		+			−	
ForMktShare	+	+	+	+	+	+
Location	+					
ImportsG7						

Table 9: Innovation Sales Equation

Sector	HT	HMT	LMT	LT	KIS	OS
R&D/L	+	+	+	+	+	+
log(K/L)						−
MktShare	+			+		
Location			+		+	

Table 10: Labor Productivity Equation

Sector	HT	HMT	LMT	LT	KIS	OS
InnovSales/L	+	+	+	+	+	+
log(L)	−	−	−	−	−	−
log(K)	+	+	+	+	+	+
HumanCapital		+	+	+	+	+
MktShare		+	+	+		
DomAffiliated			+	+	+	+
DomMNE	+	+		+		+
ForMNE	+	+	+	+	+	+
Location	+	+		+		+