Innovative Sales, R&D and Other Innovation Expenditures: Are there Lags? Estimating from Dynamic Panel Data Sample Selection Models

Wladimir Raymond, Pierre Mohnen, Franz Palm and Sybrand Schim van der Loeff[1]

May 23, 2009

Abstract

This paper studies the dynamic relationship between input and output of innovation in Dutch manufacturing using an unbalanced panel of enterprise data from five waves of the Community Innovation Survey during 1994-2004. We estimate by maximum likelihood a dynamic panel data bivariate tobit with double-index sample selection accounting for individual effects. We find persistence of innovation input and innovation output, a lag effect of the former on the latter and a feedback effect of the latter on the former. The lag effect remains significant in the high-tech sector even after four years. Firm and industry effects are also important.

[1]The empirical part of this study has been carried out using the remote access facility of the Centre for Research of Economic Microdata of Statistics Netherlands. The authors wish to thank Statistics Netherlands for helping them in accessing and using the Micronoom data set. The views expressed in this paper are solely those of the authors. The first author acknowledges financial support from METEOR. The first and third author acknowledge financial support by the Royal Netherlands Academy of Arts and Sciences (KNAW). Please address correspondence to: Wladimir Raymond, University of Maastricht, P.O. Box 616, 6200 MD Maastricht, The Netherlands; Tel.: (+31) 43 388 4962; Fax: (+31) 43 388 4874; w.raymond@maastrichtuniversity.nl.

1 Introduction

This paper studies the dynamic relationship between innovation input and innovation output in Dutch manufacturing using an unbalanced panel of enterprise data from five waves of the Community Innovation Survey (CIS) during the period 1994-2004. More specifically, the dynamics that we analyze consists of the persistence of innovation input as measured by the ratio of R&D or total innovation spending over total sales, the persistence of innovation output as measured by the share in total sales accounted for by sales of new or improved products (innovative sales), the lag effect of innovation input on innovation output and the feedback effect of innovation output on innovation input. The analysis pertains to the literature on panel data knowledge production functions (KPF) that are central to the economics of technical change and the persistence of innovation that plays an important role in endogenous growth and industrial dynamics.

Panel data KPF is pioneered in the work of Pakes and Griliches (1980a) who define a theoretical model relating innovation input to innovation output, and derive a distributed lag regression where the number of patents (innovation output) is regressed on current and five lags of R&D (innovation input) and firm individual effects.[2] They estimate a log-log specification of their model that ignores the discreteness of patent using the 'within' estimator to account for the individual effects. They find simultaneity between innovation input and innovation output in the sense that current R&D affects positively and significantly patents, and a lag truncation which means that the coefficient of the last lag is significant but nothing between current and the last lagged R&D is significant. Pointing out the limitations of the study of Pakes and Griliches, Hausman et al. (1984) propose several panel data models to estimate the patents-R&D relationship that take into account the discreteness of patent, namely fixed- and random-effects Poisson and negative binomial (NegBin) regressions. Using similar data to Pakes and Griliches' they find that whenever the individual effects are allowed to be correlated with R&D there is only evidence of simultaneity between R&D and patents, in particular no lag effect of R&D on patents can be ascertained. This result is confirmed by the study of Hall et al. (1986) who use similar data, albeit for a larger number of firms over a smaller period, to estimate fixed-effects Poisson, NegBin and GMT Poisson regressions.[3]

Like in the literature on panel data KPF, empirical studies that investigate the persistence of innovation have for a long time used exclusively patent data. With the exception of Crépon and Duguet (1997), all these studies conclude that there is no clear-cut evidence of strong persistence in innovation activities, regardless of the methodology. The use of patent data is, however, too

[2] See also Pakes and Griliches (1980b).

[3] We call a GMT Poisson regression, in reference to Gouriéroux-Montfort-Trognon, a Poisson regression that is estimated using a quasi-generalized pseudo maximum likelihood method devised by the authors (see Gouriéroux et al., 1984).

demanding because analyzing the persistence of innovation using patent data amounts to analyzing the persistence in "winning the patent race", which is even harder than coming up with a new product. With the advent of innovation survey in the early 90s, less demanding data can be used to study the persistence of innovation, a recent instance being the study of Peters (2009) who finds strong persistence in innovation activities. While innovation survey data provide researchers with several qualitative and quantitative measures of innovation input and innovation output, empirical studies based on such data mostly investigate the persistence of innovation using qualitative measures of innovation input or innovation output. Thus, such studies overlook important other components of the dynamics, e.g. the lag effect of innovation input on innovation output or the feedback effect of the latter on the former, that require the use of quantitative measures of innovation. The main reason for this overlooking is that, in the innovation surveys, the quantitative variables of innovation input and innovation output are truncated because they are observed only for a certain type of enterprise that we define later as innovative and censored because, even for innovative enterprises, some innovation input variables (e.g. R&D) and innovation output may take on zero or very small values. Hence, studying the dynamics of innovation accounting for its above-mentioned components requires the use of dynamic panel data censored-truncated regressions with multiple equations that are rather difficult to implement given the additional issues of unobserved individual effects and initial conditions.

The contribution of this paper to the literature is to consider a unified framework where all four components of the dynamics of innovation mentioned above are studied and the various features of the CIS data are accounted for as much as possible. More specifically, we analyze jointly the lag structure between innovation input and innovation output, the persistence of innovation input and innovation output, and the the feedback effect of innovation output on innovation input using a dynamic panel data bivariate tobit with time-varying double index sample selection and accounting for unobserved individual effects. A selection rule based on double indices is used to correct for sample selection bias that would occur if we restricted the analysis to the sample of continuously-innovative enterprises for which we have a complete set of data. This is explained in detail in the body of the paper. Another feature of the CIS data that is worth mentioning is its very little 'within' variation, which makes the use of semiparametric fixed-effects techniques that rely on 'within' or time differences unfeasible. As a result, we take the random-effects route by making distributional assumptions on both the idiosyncratic errors and the individual effects that are integrated out using two-step Gauss-Hermite quadrature along the lines of Raymond (2007), and using Wooldridge's (2005) "simple solutions" to the initial conditions problem. The results suggest the presence of significant dynamics in the innovation process even after controlling for individual effects correlated with the initial values of the variables of interest. In other words, we find persistence of innovation input and innovation output, a lag effect of the former on the latter and a feedback effect of the latter on the former. Impulse response functions suggest that the lag effect remains significant in the high-tech sector after four years. Firm and industry effects play an important role in the relationship.

The remainder of the paper is organized as follows. We describe the data in Section 2, present the model in Section 3 and its estimation in Section 4. In Section 5 we discuss the results, and in Section 6 we summarize and conclude.

2 Data

The data are collected by the *Centraal Bureau voor de Statistiek* (CBS) and stem from five waves of the Dutch CIS, namely CIS 2 (1994-1996), CIS 2.5 (1996-1998), CIS 3 (1998-2000), CIS 3.5 (2000-2002) and CIS 4 (2002-2004), merged with data from the Production Survey (PS). Only enterprises in Dutch manufacturing (SBI 15.1-37.2) are included in the analysis.[4] We consider enterprises with at least ten employees and positive sales at the end of each period covered by the innovation survey. Furthermore, we remove from the sample enterprises whose total innovation expenditures count for more than 50% of total sales.

The CIS and PS data are collected at the enterprise level. A combination of a census and a stratified random sampling is used for each wave of the CIS and PS. A census is used for the population of enterprises with at least 50 employees, and a stratified random sampling is used for enterprises with less than 50 employees. The stratum variables are the economic activity and the number of employees of an enterprise. The same cut-off point of 50 employees is applied to each wave of the CIS and PS resulting in about 3000 enterprises in each wave of the merged data of our sample.

In the CIS questionnaire, enterprises are first asked some general questions about their identity, economic activity, exports, total sales, number of employees and whether they belong to a group or not. Then come three crucial questions regarding 1) whether an enterprise has introduced new or improved products into the market 2) whether it has introduced new or improved processes and 3) whether it has ongoing or abandoned innovation activities during the period under review. If an enterprise answers "yes" to either of these three questions, then it has to fill out the whole questionnaire where information on expenditures, sources and effects of innovation as well as innovation cooperation has to be provided. Such an enterprise is called innovative and has to

[4] SBI stands for the Dutch standard industrial classification and gives the enterprise economic activity.

report a positive value of total innovation expenditures at the end of the period under review.[5] If an enterprise answers "no" to all three questions, then only the general questions are answered, and the information on innovation expenditures, information sources, effects, and cooperation is missing. In order to cope with this feature of the data, we estimate a sample selection model where we correct for the bias that occurs if the analysis is restricted to the sub-sample of enterprises for which we have a complete set of data.

2.1 Patterns

Table 1 shows the patterns of the unbalanced panel of enterprises for which the dynamics of innovation can be potentially studied.[6] There are 3144 such enterprises for which descriptive statistics regarding size (number of employees), market share defined as the ratio of the sales of an enterprise over the total sales of the 3-digit industry it belongs to (see the classification of Eurostat, 1992),[7] and the proportion of occasionally and continuously-innovative enterprises are shown for each pattern where an occasionally-innovative enterprise is defined as one that has innovative activities in any wave of the CIS between 1994 and 2004 and a continuously-innovative enterprise is defined as one that has such activities in at least two consecutive waves. For instance, the first pattern consists of 632 enterprises (20% of the unbalanced panel) that were sampled only over the periods 1994-1996 and 1996-1998. These enterprises have on average 145 employees (a median of 35 employees) and an average market share of 0.25% (a median of 0.04%) over 1994-1998; 67% of them have some innovation activities in either period and 26% have continuous innovation activities, i.e. they are innovative over the whole period 1994-1998 of the pattern. The second pattern represents a balanced panel of 338 enterprises that were sampled from 1994 till 2004. It counts only for 11% of the whole unbalanced panel of the analysis and consists of a significantly larger proportion of continuously-innovative enterprises that have on average a significantly larger market share. Hence, restricting the analysis to the sole balanced panel would miss a lot of information out of the unbalanced panel, and would yield results that are biased towards continuously-innovative enterprises with a large market share. A result of the table that is worth mentioning is the decrease of the proportion of occasional and continuously-innovative enterprises as time passes coupled with

[5]In addition to R&D, innovation expenditures comprise the purchase of rights and licenses to use external technology, the purchase of advanced machinery and computer hardware devoted to the implementation of product and process innovations, expenditures for technical preparations to realize the actual implementation of product and process innovations, expenditures for marketing activities aimed at market introduction of product innovations, and expenditures for staff training aimed at the development and/or introduction of new product and process innovations. Only indicators of the last three components of innovation expenditures are provided in CIS 4. As a result, total innovation expenditures of this analysis consist of R&D and the first two components.

[6]The sample of enterprises that take part in at least two waves of the Dutch CIS and that can be potentially included in the dynamic analysis of innovation will be referred to as the "feasible" sample.

[7]Total sales of a 3-digit industry is obtained by adding up the sales of all the firms in our sample that belong to that industry after multiplying them by the appropriate raising factor.

an increase of the average market share. For instance, comparing the patterns of enterprises that take part in only two waves of the CIS, we observe a decrease of the proportion of occasionally and continuously-innovative enterprises from 67% and 26% respectively in the period 1994-1998 to 48% and 17% in the period 2000-2004, while at the same time market share increases on average from 0.25% to 0.40%. This seems to indicate that as time passes a small core of innovative enterprises emerge and hold a significantly larger market share.

Table 1: Size, market share and the proportion of innovative enterprises in each pattern of the unbalanced data for Dutch manufacturing: CIS 2, CIS 2.5, CIS 3, CIS 3.5 and CIS 4

Pattern	# firms	%	Size		Market share† (%)		Innovative	
			Mean	Median	Mean	Median	Occasional	Continuous‡
11000	632	20.10	145	35	0.247	0.035	0.668	0.261
11111	338	10.75	263	119	0.706	0.126	0.758	0.536
00011	298	9.48	187	50	0.396	0.038	0.478	0.168
11110	245	7.79	161	75	0.419	0.071	0.691	0.443
11100	231	7.35	160	70	0.297	0.071	0.732	0.407
00110	184	5.85	81	48	0.166	0.032	0.563	0.220
00111	153	4.87	364	115	0.493	0.065	0.588	0.294
01100	145	4.61	126	50	0.445	0.037	0.617	0.228
11010	133	4.23	84	60	0.305	0.051	0.689	0.218
11011	115	3.66	189	110	0.518	0.119	0.722	0.322
11001	110	3.50	116	45	0.339	0.059	0.715	0.203
01111	107	3.40	332	102	0.557	0.109	0.666	0.418
01110	102	3.24	127	74	0.318	0.064	0.637	0.346
01011	76	2.42	326	71	0.455	0.069	0.684	0.158
10111	70	2.23	157	121	0.362	0.107	0.689	0.289
10110	59	1.88	126	70	0.552	0.071	0.644	0.181
11101	58	1.84	174	93	0.337	0.120	0.716	0.323
10011	50	1.59	159	106	0.527	0.104	0.713	0.180
01101	38	1.21	186	76	0.432	0.074	0.711	0.228
Total	3144	100.00	191	75	0.433	0.072	0.677	0.339

†in the domestic market. ‡A continuously-innovative enterprise is one that has innovation activities in at least two successive waves of the CIS.

Table 2 shows the proportion of non-innovative, occasionally- and continuously-innovative enterprises of the feasible sample. For instance 20% of the sample have no innovation activities in periods t-1 and t, 14% have innovation activities in period t-1 but have no innovation activities in period t, and 10% have no innovation activities in period t-1 but have innovation activities in period t. Neither of these three types of enterprises can be included in the dynamic analysis of innovation. Indeed, a complete set of data is available only for innovative enterprises and the dynamic analysis requires a complete set of data in at least two consecutive periods. As a result, an enterprise has to be innovative in at least two consecutive periods if it is to be included in the dynamic analysis. In other words, we have a complete set of data to study the dynamics of innovation for less than 60% of the feasible sample. Carrying out the analysis using the sub-sample of continuously-innovative firms is likely to suffer from sample selection bias. As a result, we use the full feasible sample in a dynamic model correcting for sample selection bias, which results in a sample selection model with a selection rule based on time-varying double indices.[8]

Table 2: Percentage of non-innovative, occasionally- and continuously-innovative enterprises in the feasible sample

	Period t					
Period t-1	Non-innovative		Innovative		Total→	
	# obs.	%	# obs.	%	# obs.	%
Non-innovative	1100	19.674	538	9.622	1638	29.297
Innovative	756	13.522	3197	57.181	3953	70.703
Total↓	1856	33.196	3735	66.803	5591	100.000

2.2 Dependent variables

The variables of interest of the analysis are innovation input measured either by the ratio of total (intramural and extramural) R&D expenditures over total sales or the ratio of total innovation expenditures over total sales, and innovation output measured by the share in total sales accounted for by sales of new or improved products. Innovation input is constructed from R&D or total innovation expenditures stemming from the CIS and total sales stemming from the PS, while innovation output is directly reported as a ratio in the CIS. Both variables of innovation input and the variable of innovation output are measured for the last year of the period under review and are logit transformed in the estimation of the model so as to make them lie within the set of real numbers.[9]

Table 3: Innovation input and innovation output of continuously-innovative enterprises‡

Variable	Statistics						
	Mean	(Std. Dev.)	P_{10}	P_{18}	Q_1	Median	Q_3
R&D expenditures/sales	0.022	(0.041)	0.000	0.001	0.002	0.008	0.023
Total innov. expenditures/sales	0.032	(0.051)	0.002	0.004	0.006	0.015	0.036
Share of innovative sales	0.240	(0.242)	0.000	0.000	0.050	0.200	0.350

‡P_{10}, P_{18}, Q_1 and Q_3 denote the 10^{th} and 18^{th} percentiles, and the first and the third quartiles respectively.

Table 3 shows descriptive statistics of the variables of interest for continuously-innovative enterprises. It suggests that 10% of such enterprises have no R&D expenditures and have total

[8] In the econometric literature, models with selection rules based on two indices are often referred to as "double hurdle" models (see e.g. Cragg, 1971; Blundell and Meghir, 1987). However these models are different from the one considered in this analysis in that the double indices are defined according to two different latent variables measured at the same time period in the double hurdle model, while in our model they are defined according to a single latent variable but taken at two different time periods.

[9] Innovation input and innovation output may take the value 0, and innovation output may take the value 1. For instance, among the continuously-innovative enterprises are non-product innovators and newly-established product innovators. The share of innovative sales takes on the value 0 for the former and 1 for the latter. We replace respectively these values by ϵ_1 between 0 and the lowest positive value of the corresponding variable, and by ϵ_2 between the largest value (smaller than 1) of the variable and 1.

innovation expenditures less than or equal to 0.2%, while 18% are not successful in achieving product innovations. The mean and median figures of innovation input are rather small and 75% of the continuously-innovative enterprises have R&D or total innovation expenditures over sales no larger than 0.04. Similarly, the mean and median share of innovative sales is about 0.20 and 75% of the continuously-innovative enterprises have a share of innovative sales no greater than 0.35. The zero and small values of the variables of interest for a significant proportion of continuously-innovative enterprises condition the choice of the empirical model. These values can be seen as inducing measurement error in the model which is to be explicitly controlled for.[10] We deal with this by using tobit-type models which censor these zero and small values and hence lessen their influence in the model. For instance, if we used a type 1 tobit (according to Amemiya's (1984) terminology) to model separately R&D input and innovation output with a censoring threshold equal to 0, 10% and 18% of the sample of continuously-innovative enterprises would be censored. However, we model jointly innovation input and innovation output using bivariate tobit models that allow for a correlation between the two processes. Table 4 shows the four types of continuously-innovative enterprises that enter the bivariate tobit model for both measures of innovation input and for different censoring thresholds. For instance, 4% of the sample of continuously-innovative enterprises are neither R&D performers nor product innovators, 14% perform R&D but are not successful in achieving product innovations, 6% do not perform R&D but are somehow successful in achieving product innovations, and 75% perform R&D and are successful in achieving product innovations.

Since we control for sample selection bias, an additional dependent variable is to be considered. This variable is binary indicating whether an enterprise is innovative, where being innovative is defined as being 1) a product innovator or 2) a process innovator, or having 3) ongoing or abandoned innovation activities. The dependent variable of the selection equation is constructed from these three dummy variables stemming from the CIS. As we explained before, the selection rule is based on a time-varying double indices. Hence, the dependent variable has to be considered at two consecutive periods t-1 and t resulting in being continuously innovative.

2.3 Explanatory variables

We explain current innovation input and current innovation output by their lagged counterparts,[11] size, market share, and indicators of cooperation, innovation sources of information and effects of innovation. We also include as regressors three dummy variables of category of industry according

[10] As a matter of fact, we have in earlier stages of the analysis estimated a panel VAR model that ignores this issue and encountered problems of convergence. This is explained in detail later in the analysis.

[11] It is to be noted that a one period lag actually corresponds to two years (since the Dutch CIS is held on a bi-annual term).

Table 4: Degree of censoring of innovation input and innovation output of continuously-innovative enterprises

(Input, Output)	R&D expenditures/sales		Total innov. expenditures/sales	
	# obsv.	%	# obsv.	%
		c_1=0, c_2=0		
($\leq c_1$,$\leq c_2$)	140	4.379	21	0.657
($>c_1$,$\leq c_2$)	457	14.295	576	18.017
($\leq c_1$,$>c_2$)	193	6.037	52	1.627
($>c_1$,$>c_2$)	2407	75.289	2548	79.700
		c_1=0.002, c_2=0.05		
($\leq c_1$,$\leq c_2$)	332	10.385	140	4.379
($>c_1$,$\leq c_2$)	473	14.795	665	20.800
($\leq c_1$,$>c_2$)	468	14.639	224	7.007
($>c_1$,$>c_2$)	1924	60.181	2168	67.814
		c_1=0.05, c_2=0.002		
($\leq c_1$,$\leq c_2$)	576	18.017	523	16.359
($>c_1$,$\leq c_2$)	33	1.032	86	2.690
($\leq c_1$,$>c_2$)	2278	71.254	2108	65.937
($>c_1$,$>c_2$)	310	9.697	480	15.014
		c_1=0.05, c_2=0.05		
($\leq c_1$,$\leq c_2$)	763	23.866	698	21.833
($>c_1$,$\leq c_2$)	42	1.314	107	3.347
($\leq c_1$,$>c_2$)	2091	65.405	1933	60.463
($>c_1$,$>c_2$)	301	9.415	459	14.357

to the OECD (2007) classification (see Appendix A) where the low-tech category is the reference, and three time dummy variables where the period 2002-2004 is the reference.

Size and market share are included as determinants of innovation on the grounds of the Schumpeterian tradition (Schumpeter, 1934; 1942) where firm size and market share are expected to have an impact on both the amount of innovational effort and innovational success (see Kamien and Schwartz, 1975; Acs and Audretsch, 1987). As mentioned earlier, size is measured by the number of employees, and domestic market share is defined as the ratio of the sales of an enterprise over the total sales of the 3-digit industry it belongs to. The number of employees and sales stem from the PS and are measured for the last year of the period under review. Size and domestic market share are log-transformed in the estimation.

Enterprises that undertake innovative activities in cooperation are expected to benefit from knowledge spillovers, hence to perform better technologically (D'Aspremont and Jacquemin, 1988). The dummy variable for cooperation indicating whether an enterprise undertakes its innovative activities in cooperation during the period under review is directly reported in the CIS.

The CIS data also provide information regarding the importance of information sources of innovation on a 0-3 Likert scale. Three dummy variables of internal sources (from the enterprise or the enterprise group), institutional sources (from universities, public or private research institutes) and market sources (from customers, competitors or suppliers) are constructed as taking the value one if the corresponding information sources are deemed very important (i.e. take the value 3 on the scale), and zero otherwise. Similarly, the CIS data provide information regarding the importance of innovation effects on a 0-3 Likert scale. Three dummy variables of product-oriented effects (i.e. increase the range of goods, improve their quality, increase market share or enter new markets), process-oriented effects (i.e. improve flexibility of production, increase its capacity or reduce labor costs, materials or energy per unit output), and environment-oriented effects (i.e. reduce environmental impacts) are constructed as taking the value one if the corresponding effects of innovation are deemed very important (i.e. take the value 3 on the scale), and zero otherwise. We include the lagged dummy variables of effects of innovation in the equations of innovation input and innovation output.

Finally, we explain the probability of an enterprise to be continuously innovative (selection equation) by size, market share and a dummy variable, directly reported in the CIS, for being part of a group as defined in the Oslo manual (OECD, 2005). We also include dummy variables of time and category of industry. These variables are the few ones that are available for both innovative and non-innovative enterprises.

Table 5: Descriptive statistics of the explanatory variables for the feasible sample and for continuously-innovative enterprises

Variable	Mean	(Std. Dev.)	Mean	(Std. Dev.)
	All enterprises		Continuously innovative	
Size	208.511	(782.143)	257.930	(897.512)
Market share (%)	0.480	(1.762)	0.667	(2.131)
Part of a group	0.725	-	0.781	-
Cooperation	-	-	0.406	-
Information sources				
Internal sources	-	-	0.524	-
Market sources	-	-	0.347	-
Institutional sources	-	-	0.055	-
Effects of innovation	-	-		
Product-oriented	-	-	0.673	-
Process-oriented	-	-	0.369	-
Environment-oriented	-	-	0.220	-
# observations	5591		3197	

Table 5 shows descriptive statistics of the explanatory variables for all enterprises and continuously-innovative enterprises of the feasible sample. Due to the design of the CIS, there are very few variables that enable us to discriminate between innovative and non-innovative enterprises. The table suggests that continuously-innovative firms have on average a significantly larger market share than that of all firms of the feasible sample, and a significantly larger proportion of continuously-innovative enterprises is part of a group. Furthermore, less than 42% of continuously-innovative enterprises have innovation cooperation, and a significantly larger percentage of them deems internal innovation sources and product-oriented innovation effects the most important.

3 Model

The model consists of two parts, namely a selection equation based on two indices explaining the probability of being continuously innovative and a dynamic bivariate tobit regression that explains innovation input and innovation output given continuously-innovative enterprises. The dynamics of innovation that we study includes the persistence of innovation input and innovation output, the lag effect of innovation input on innovation output and the feedback effect of innovation output on innovation input. As we explained earlier, we consider a selection rule based on two indices in order to correct for sample selection bias that occurs because the dynamics of innovation can be studied only for enterprises that are innovative in (at least) two consecutive periods.

3.1 Double index selection

Let d_{it}^{*} be a latent variable that represents firm's i incentive to carry out innovation activities at period t ($i = 1, ...N$; $t = 1, ...T_i$). This innovation incentive can be expressed as a function of firm, market and industry characteristics $\mathbf{w}_{it}$ taken at period t, unobserved individual effects $\eta_{i,}$ and other unobserved time-varying variables u_{it}. Formally d_{it}^{*} is written as

$$d_{it}^{*} = \boldsymbol{\delta}'\mathbf{w}_{it} + \eta_i + u_{it}, \tag{1}$$

where $\boldsymbol{\delta}$ is a vector of parameters to be estimated. The incentive to carry out innovation activities is not observed, instead we observe d_{it} that takes on the value one if the enterprise is actually innovative, which is the case if the incentive to perform innovation activities is sufficiently large (i.e. if it crosses a certain threshold, say 0), and zero otherwise. Formally, d_{it} is written as

$$d_{it} = \mathbf{1}[d_{it}^{*} > 0], \tag{2}$$

where $\mathbf{1}[\ ]$ is the indicator function that takes on the value one if the condition between squared brackets is satisfied, and zero otherwise. A continuously-innovative enterprise is defined as one for which the incentive to carry out innovation activities crosses the threshold at two consecutive periods, i.e. $d_{it}^{*} > 0$ and $d_{i,t-1}^{*} > 0$, hence the double index selection rule

$$d_{it}d_{i,t-1} = \mathbf{1}[d_{it}^{*} > 0 \cap d_{i,t-1}^{*} > 0], \tag{3}$$

where $\cap$ is the intersection operator.

3.2 Dynamic bivariate tobit

Let y^*_{1it} and y^*_{2it} denote two latent variables modeling innovation input and innovation output of firm i at period t. They are expressed as functions of past observed innovation input $y_{1i,t-1}$ and innovation output $y_{2i,t-1}$, current and past supposedly exogenous explanatory variables $\mathbf{x}_{it}$, unobserved individual effects α_i and other time-varying unobserved variables ϵ_{1it} and ϵ_{2it}. Formally, innovation input and innovation output are written as

$$y^*_{1it} = \gamma_{11} y_{1i,t-1} + \gamma_{12} y_{2i,t-1} + \boldsymbol{\beta}'_1 \mathbf{x}_{it} + \alpha_i + \epsilon_{1it}, \tag{4}$$

$$y^*_{2it} = \gamma_{21} y_{1i,t-1} + \gamma_{22} y_{2i,t-1} + \boldsymbol{\beta}'_2 \mathbf{x}_{it} + \lambda\alpha_i + \epsilon_{2it}, \tag{5}$$

where γ_{jk}, $\boldsymbol{\beta}_j$ $(j, k = 1, 2)$ and λ are parameters to be estimated. A few remarks are worth making when considering eqs. (4) and (5). First, we denote innovation input and innovation output with a "$*$" as a superscript to emphasize that they are only partially observed. More specifically, the conditions in equation (3) must be satisfied for innovation input and innovation output to be observed. Secondly, the same term α_i enters both equations, which means that we assume the observed and unobserved variables that proxy individual effects to be the same across equations but with a different effect on each equation, hence the presence of the factor loading λ in equation (5). Finally, the same vector of explanatory variables $\mathbf{x}_{it}$ enters both equations of our analysis although it be may different across equations.[12]

Let y_{1it} and y_{2it} denote the observed counterparts to y^*_{1it} and y^*_{2it}. They are fully observed for enterprises that satisfy the conditions of equation (3). However, even when they are fully observed, y_{1it} and y_{2it} are censored in the sense that they take on zero or very small values for a significantly large percentage of continuously-innovative enterprises (cf. Table 3), hence the choice of tobit-type models to study the dynamics of innovation. We use a dynamic bivariate tobit so as to estimate jointly both equations allowing for a cross-equation correlation between the idiosyncratic errors ϵ_{1it} and ϵ_{2it}.[13]

The observed dependent variables of innovation input and innovation output are defined as

[12] Unless economic theory suggests otherwise, there is no reason why the vector of explanatory variables should be different across equations.

[13] The use of common factor individual effects in the two equations implicitly assumes a cross-equation correlation of one between the individual effects.

follows

$$(y_{1it}, y_{2it}) = \begin{cases} (.,.) \text{ if } d^*_{it} \leq 0 \cup d^*_{i,t-1} \leq 0 \\ (c_1, c_2) \text{ if } d^*_{it} > 0 \cap d^*_{i,t-1} > 0 \cap y^*_{1it} \leq c_1 \cap y^*_{2it} \leq c_2 \\ (y^*_{1it}, c_2) \text{ if } d^*_{it} > 0 \cap d^*_{i,t-1} > 0 \cap y^*_{1it} > c_1 \cap y^*_{2it} \leq c_2 \\ (c_1, y^*_{2it}) \text{ if } d^*_{it} > 0 \cap d^*_{i,t-1} > 0 \cap y^*_{1it} \leq c_1 \cap y^*_{2it} > c_2 \\ (y^*_{1it}, y^*_{2it}) \text{ if } d^*_{it} > 0 \cap d^*_{i,t-1} > 0 \cap y^*_{1it} > c_1 \cap y^*_{2it} > c_2 \end{cases} \tag{6}$$

where $\cup$ denotes the union operator.

Equation (6) identifies five categories of enterprises. A first category consists of enterprises that are not continuously innovative, i.e. that do not satisfy the conditions of equation (3). This category consists of three sub-categories of enterprises that are never innovative, those that are innovative at period t but are not at period t-1, or those that are not innovative at period t-1 but are innovative at period t (cf. Table 2). According to equations (4) and (5), innovation input and innovation output are missing for enterprises that belong to the first category because of missing values in their current or lagged determinants. A second category consists of continuously-innovative enterprises that have zero or very small values of innovation input and innovation output, i.e. the corresponding latent variables do not cross the censoring thresholds c_1 and c_2. A third category consists of continuously-innovative firms with sufficiently large innovation input, i.e. $y^*_{1it} > c_1$, but with zero or very small values of innovation output, i.e. $y^*_{2it} \leq c_2$. A fourth category consists of continuously-innovative enterprises that are in the opposite situation to that of enterprises of the third category. Finally, a fifth category consists of continuously-innovative enterprises whose innovation input and innovation output are sufficiently large, i.e. $y^*_{1it} > c_1$ and $y^*_{2it} > c_2$.

Choice of c_1 and c_2

The censoring thresholds c_1 and c_2 determine the degree of censoring of innovation input and innovation output. For instance, if equations (4) and (5) are estimated separately, Table 3 suggests that innovation input measured by R&D expenditures over total sales and innovation output are censored for 10% and 18% of continuously-innovative enterprises respectively if both censoring thresholds are equal to 0. Table 4 shows the four types of continuously-innovative enterprises that enter the bivariate tobit analysis, as identified by equation (6), for both measures of innovation input and for different values of the censoring thresholds. These values are chosen according to a trial and error method. More specifically, we have in earlier stages of the analysis ignored the issue of zero and small values of innovation input and innovation output estimating by ML a bivariate panel VAR model using the sample of continuously-innovative enterprises. We have also estimated a dynamic bivariate tobit model with both censoring thresholds equal to 0. Neither approach was satisfactory because we could not achieve convergence in maximizing the log-likelihood and obtain reliable estimates. One problem that we faced was that we could not obtain the standard errors of the estimates because the Hessian matrix or the outer product of gradients could not be inverted. We suspected the zero and small values of innovation input and innovation output to "contaminate" the estimation by inducing measurement errors in the probability distribution functions of the log-likelihood of the panel VAR model. Censoring only the zero values of both dependent variables was unfortunately not sufficient to obtain reliable estimates. Hence, we had to censor more small values of innovation input and innovation output in order to achieve reliable estimates, which led us to choose c_1 equal to 0.002 and c_2 equal to 0.05. Other choices of $c_1 = 0.05$ and $c_2 = 0.002$, and $c_1 = c_2 = 0.05$ are made to study the robustness of the analysis to different censoring thresholds.

4 Estimation

The maximum likelihood estimation of the dynamic bivariate tobit model with double index sample selection is described as follows. We first solve the initial conditions problem that occurs in the second stage of the model using Wooldridge's (2005) "simple solutions", and make distributional assumptions on the individual effects and the idiosyncratic errors.

4.1 Initial conditions

The Wooldridge treatment of the initial conditions consists in projecting the individual effects of equations (4) and (5) on the initial period values of the dependent variables y_{1i0} and y_{2i0}, and on each time period values of sufficiently time-varying regressors or on their within mean $\bar{\mathbf{x}}_i$ so as to allow for individual effects that are correlated with exogenous explanatory variables. Formally,

$$\alpha_i = b_0 + b_1 y_{1i0} + b_2 y_{2i0} + \boldsymbol{b}_3' \bar{\mathbf{x}}_i + \mu_i, \tag{7}$$

where μ_i is independent of $\boldsymbol{\epsilon}_{it}$=$(\epsilon_{1it}, \epsilon_{2it})$, y_{1i0}, y_{2i0} and $\mathbf{x}$, and b_0, b_1, b_2 and $\boldsymbol{b}_3$ are additional parameters to be estimated.[14] The assumption of common factor individual effects implies that the additional parameters b_0, b_1, b_2 and $\boldsymbol{b}_3$ are different across equations only up to the factor loading λ.

[14]If $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ include an intercept parameter, it is not separately identified from b_0.

4.2 Distributional assumptions

In order to specify the likelihood function, we make the following assumptions on the individual effects and the idiosyncratic errors. First, conditional on η_i, u_{it} is identically and independently distributed across individuals and over time so that the bivariate probability of being selected in the estimation sample is the product of two univariate probabilities. In other words, a special form of serial correlation referred to as equicorrelation in the econometric literature is assumed in the error terms $\eta_i + u_{it}$. A similar assumption is made for ϵ_{1it} and ϵ_{2it} conditional on α_i. Secondly, the sample selection effect is assumed to operate only through the individual effects. This assumption results in a simpler likelihood expression but does not harm the analysis by restricting the model. Indeed, it is a common assumption made in the econometric literature on panel data sample selection models. For instance, by making a similar assumption, Kyriazidou (1997) takes kernel-weighted time difference of observations that eliminates not only the individual effects but also the sample selection effect (see also Charlier et al., 2001). Furthermore, we estimate in another paper a dynamic type 2 tobit model using the same data and find that such an assumption is plausible because the correlation between the idiosyncratic errors of the selection equation and the regression equation is not significantly estimated unlike that of the individual effects of the two equations once we use a proper treatment of the initial conditions (see Raymond et al., forthcoming).

To summarize, (η_i, μ_i) and $(u_{it}, u_{i,t-1}, \epsilon_{1it}, \epsilon_{2it})$ are mutually independent, and identically and independently normally distributed with mean zero and covariance matrices

$$\Sigma_{\eta\mu} = \begin{pmatrix} \sigma_\eta^2 & \\ \rho_{\eta\mu}\sigma_\eta\sigma_\mu & \sigma_\mu^2 \end{pmatrix}, \; \Sigma_{\boldsymbol{u\epsilon}} = \begin{pmatrix} 1 & & & \\ 0 & 1 & & \\ 0 & 0 & \sigma_1^2 & \\ 0 & 0 & \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

4.3 Likelihood

According to equation (6) five different contributions to the likelihood function are to be distinguished. The individual contribution to the likelihood of a firm that is not selected in the sample on which the estimation of the dynamic bivariate tobit is based is given by

$$L_{0i} = \prod_{i=1}^{T_i} [\Phi_1(-M_{it} - \eta_i) + \Phi_1(-M_{i,t-1} - \eta_i) - \Phi_1(-M_{it} - \eta_i)\Phi_1(-M_{i,t-1} - \eta_i)]^{1-d_{it}d_{i,t-1}}, \tag{8}$$

where Φ_1 denotes the univariate cumulative distribution function (cdf) of the standard normal distribution and $M_{it} = \boldsymbol{\delta}'\mathbf{w}_{it}$.

Let $D_{1it} = \mathbf{1}[y^*_{1it} > c_1]$ and $D_{2it} = \mathbf{1}[y^*_{2it} > c_2]$ where $\mathbf{1}[\ ]$ is the indicator function explained previously, and to save space define

$$N_{1it} = \gamma_{11}y_{1i,t-1} + \gamma_{12}y_{2i,t-1} + \boldsymbol{\beta}_1'\mathbf{x}_{it} \tag{9}$$

$$N_{2it} = \gamma_{21}y_{1i,t-1} + \gamma_{22}y_{2i,t-1} + \boldsymbol{\beta}_2'\mathbf{x}_{it}, \tag{10}$$

the contribution to the likelihood of a firm that is selected in the estimation sample but whose measures of innovation input and innovation output are zero or very small, i.e. the corresponding latent variables are below the thresholds c_1 and c_2, is given by

$$L_{1i}^{c_1c_2} = \prod_{i=1}^{T_i}\left[\left(\int_{-M_{it}-\eta_i}^{\infty}\int_{-M_{i,t-1}-\eta_i}^{\infty}\int_{-\infty}^{c_1-N_{1it}-\alpha_i}\int_{-\infty}^{c_2-N_{2it}-\lambda\alpha_i} f_4\left(u_{it}, u_{i,t-1}, \epsilon_{1it}, \epsilon_{2it}\right) du_{it}du_{i,t-1}d\epsilon_{1it}d\epsilon_{2it}\right)^{(1-D_{1it})(1-D_{2it})}\right]^{d_{it}d_{i,t-1}}, \tag{11}$$

where f_4 denotes the quadrivariate probability distribution function (pdf) of the normal distribution. However, according to the assumption on the idiosyncratic errors conditional on the individual effects and the assumption on the sample selection effect, f_4 can be written as the product of $f_1(u_{it})$, $f_1(u_{i,t-1})$ and $f_2(\epsilon_{1it}, \epsilon_{2it})$ where f_1 and f_2 denote the univariate and the bivariate pdf of the normal distribution respectively. Hence, equation (11) can be written as

$$L_{1i}^{c_1c_2} = \prod_{i=1}^{T_i}\left\{\left[\Phi_2\left(\frac{c_1 - N_{1it} - \alpha_i}{\sigma_1}, \frac{c_2 - N_{2it} - \lambda\alpha_i}{\sigma_2}, \rho_{12}\right) \Phi_1\left(M_{it} + \eta_i\right)\Phi_1\left(M_{i,t-1} + \eta_i\right)\right]^{(1-D_{1it})(1-D_{2it})}\right\}^{d_{it}d_{i,t-1}}, \tag{12}$$

where Φ_2 denotes the bivariate cdf of the standard normal distribution.

The contribution to the likelihood of a firm that is selected in the estimation sample with sufficiently large innovation input but with zero or very small values of innovation output is given by

$$L_{1i}^{y_1c_2} = \prod_{i=1}^{T_i}\left[\left(\int_{-M_{it}-\eta_i}^{\infty}\int_{-M_{i,t-1}-\eta_i}^{\infty}\int_{-\infty}^{c_2-N_{2it}-\lambda\alpha_i} f_4\left(u_{it}, u_{i,t-1}, y_{1it}, \epsilon_{2it}\right) du_{it}du_{i,t-1}d\epsilon_{2it}\right)^{(1-D_{1it})(1-D_{2it})}\right]^{d_{it}d_{i,t-1}},$$

which can also be written as

$$L_{1i}^{y_1c_2} = \prod_{i=1}^{T_i}\left[\left(\Phi_1\left(M_{it}+\eta_i\right)\Phi_1\left(M_{i,t-1}+\eta_i\right)\int_{-\infty}^{c_2-N_{2it}-\lambda\alpha_i} f_2(y_{1it},\epsilon_{2it})d\epsilon_{2it}\right)^{D_{1it}(1-D_{2it})}\right]^{d_{it}d_{i,t-1}}$$
$$= \prod_{i=1}^{T_i}\left[\left(\Phi_1\left(M_{it}+\eta_i\right)\Phi_1\left(M_{i,t-1}+\eta_i\right)f_1(y_{1it})\int_{-\infty}^{c_2-N_{2it}-\lambda\alpha_i} f_1(\epsilon_{2it}|y_{1it})d\epsilon_{2it}\right)^{D_{1it}(1-D_{2it})}\right]^{d_{it}d_{i,t-1}}.$$

The final expression of $L_{1i}^{y_1c_2}$ is given by

$$L_{1i}^{y_1c_2} = \prod_{i=1}^{T_i}\Bigg[\Bigg(\Phi_1\left(M_{it}+\eta_i\right)\Phi_1\left(M_{i,t-1}+\eta_i\right)\phi_1\left[(y_{1it}-N_{1it}-\alpha_i)/\sigma_1\right]/\sigma_1$$
$$\Phi_1\left(\frac{c_2-N_{2it}-\lambda\alpha_i-\rho_{12}\frac{\sigma_2}{\sigma_1}(y_{1it}-N_{1it}-\alpha_i)}{\sigma_2\sqrt{1-\rho_{12}^2}}\right)\Bigg)^{D_{1it}(1-D_{2it})}\Bigg]^{d_{it}d_{i,t-1}}, \tag{13}$$

where ϕ_1 denotes the univariate pdf of the standard normal distribution.

Following a similar approach we show that the contribution to the likelihood of a firm that is selected in the estimation sample with sufficiently large innovation output but with zero or very small values of innovation input is given by

$$L_{1i}^{c_1y_2} = \prod_{i=1}^{T_i}\Bigg[\Bigg(\Phi_1\left(M_{it}+\eta_i\right)\Phi_1\left(M_{i,t-1}+\eta_i\right)\phi_1[(y_{2it}-N_{2it}-\lambda\alpha_i)/\sigma_2]/\sigma_2$$
$$\Phi_1\left(\frac{c_1-N_{1it}-\alpha_i-\rho_{12}\frac{\sigma_1}{\sigma_2}(y_{2it}-N_{2it}-\lambda\alpha_i)}{\sigma_1\sqrt{1-\rho_{12}^2}}\right)\Bigg)^{(1-D_{1it})D_{2it}}\Bigg]^{d_{it}d_{i,t-1}}. \tag{14}$$

Finally, the contribution to the likelihood of a firm selected in the estimation sample with sufficiently large innovation input and innovation output is given by

$$L_{1i}^{y_1y_2} = \prod_{i=1}^{T_i}\Bigg\{\Bigg[\Phi_1\left(M_{it}+\eta_i\right)\Phi_1\left(M_{i,t-1}+\eta_i\right)\phi_1[(y_{2it}-N_{2it}-\lambda\alpha_i)/\sigma_2]/\sigma_2$$
$$\frac{1}{\sigma_1\sqrt{1-\rho_{12}^2}}\phi_1\left(\frac{y_{1it}-N_{1it}-\alpha_i-\rho_{12}\frac{\sigma_1}{\sigma_2}(y_{2it}-N_{2it}-\lambda\alpha_i)}{\sigma_1\sqrt{1-\rho_{12}^2}}\right)\Bigg]^{D_{1it}D_{2it}}\Bigg\}^{d_{it}d_{i,t-1}}. \tag{15}$$

The overall individual likelihood $L_i(...|\eta_i,\mu_i)$ of the dynamic bivariate tobit with double index sample selection conditional on the individual effects is obtained, after replacing α_i by its expression (eq. (7)), by multiplying the expressions of equations (8), (12), (13), (14) and (15), i.e.

$$L_i(...|\eta_i,\mu_i) = L_{0i}L_{1i}^{c_1c_2}L_{1i}^{y_1c_2}L_{1i}^{c_1y_2}L_{1i}^{y_1y_2}. \tag{16}$$

We then obtain the unconditional individual likelihood by integrating the individual effects out of $L_i(...|\eta_i, \alpha_i)$, i.e.

$$L_i(..., \eta_i, \mu_i) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} L_i(...|\eta_i, \mu_i) g_2(\eta_i, \mu_i) d\eta_i d\mu_i, \tag{17}$$

where g_2 denotes the bivariate normal density of η_i and μ_i. We evaluate the double integral in equation (17) using two-step Gauss-Hermite quadrature along the lines of Raymond (2007, chapter 3). The overall unconditional likelihood $L(..., \eta_i, \mu_i)$ is obtained by taking the product over i of the evaluated expression of equation (17).

5 Results

In subsection 5.1 we discuss the estimation results of the dynamic bivariate tobit model with double index sample selection (full model). More specifically, we discuss the effects of past innovation input on current innovation input and innovation output, those of past innovation output on current innovation input and innovation output, and test for their equality across categories of industry. Furthermore, we discuss the role of firm and industry effects, the effects of supposedly exogenous determinants of innovation, the issue of sample selection bias and the robustness of the estimates to different censoring thresholds. We also present in the subsection (orthogonalized) impulse response functions, obtained from the tobit estimates, that analyze the response over time of innovation input and innovation output to shocks to innovation input and innovation output. In subsection 5.2 we contrast the tobit estimates with those of a panel VAR that does not correct for sample selection bias nor take account of the censoring feature of both innovation input and innovation output.

5.1 Tobit estimates

Table 6 shows ML estimation results of the full model with both measures of innovation input and censoring thresholds $c_1 = 0.002$ and $c_2 = 0.05$.[15] Panel A of the table shows the parameter estimates of the selection equation. Panel B shows the parameter estimates of the innovation input equation where we use, as measures of innovation input, R&D expenditures over total sales in the middle two columns and total innovation expenditures over total sales in the last two columns.

[15] In order to save space, we report only the estimates of the coefficients of the selection equation, the lagged dependent variables, the industry and individual effects and the initial conditions. The coefficient estimates associated with the exogenous explanatory variables are not reported but can be obtained upon request. The estimation results of the model ignoring sample selection and using different censoring thresholds are not reported either but can be obtained upon request.

Panel C shows the parameter estimates of the innovation output equation, and panel D shows those of the additional parameters (individual effects and initial conditions) of the model.

Dynamics, firm and industry effects

The dynamics of innovation is expected to be industry-specific because of product-life cycle that varies across industries (see e.g. Cefis and Orsenigo, 2001). Furthermore, Aghion et al. (2005) and Acemoglu et al. (2006) find that industries that are closer to the "technological frontier" are more competitive and that the competitive pressure pushes firms to innovate. Following this logic, firms in high-tech industries are in general closer to the "technological frontier" and therefore they are more likely to display persistence in innovation. In order to allow for industry-specific innovation dynamics, we estimated the model by interacting the lagged dependent variables ($y_{1i,t-1}$, $y_{2i,t-1}$) of each equation with four dummies of industry category as defined in the OECD classification (see Appendix A), namely high-tech, medium-high-tech, medium-low-tech and low-tech. Then we performed Wald tests on the equality of the coefficients of the lagged dependent variables, the dynamic parameters, across industry categories. The results of these tests are reported in Table 7 and suggest the presence of industry category heterogeneity with regards to the dynamics of innovation. In other words, the joint null hypothesis of equality across industry categories of the coefficients of the persistence of innovation input and innovation output, the lag effect of innovation input on innovation output, and the feedback effect of innovation output on innovation input can be rejected at 0.5% level of significance (see the lower part of Table 7). The main differences in the dynamics across industry categories stem from the persistence of innovation input and its lag effect on innovation output. The behavior of the industry categories with regards to these two components of the dynamics depends on whether R&D or total innovation expenditures is used as innovation input. For both dynamic components and both innovation input measures, two groups of industry categories can be identified where the dynamic component is similar within each group but different across groups. More specifically, the persistence of R&D is similar within the group of high- and medium-high-tech categories, and within the group of medium-low- and low-tech categories and the lag effect of R&D on innovation output in the high-tech category is different from that of the group of medium-high-, medium-low- and low-tech categories. Similarly, the persistence of total innovation spending in the medium-low-tech category is different from that of the group of high-, medium-high and low-tech categories, and the lag effect of total innovation spending on innovation output in the low-tech category is different from that of the group of high-, medium-high- and medium-low-tech categories.

Table 6: ML estimates of the dynamic bivariate tobit with double index sample selection: c_1=0.002, c_2=0.05‡

Variable	Coefficient	(Std. Err.)	Coefficient	(Std. Err.)
	R&D expenditures/sales		Total innov. expenditures/sales	
A)	Being continuously innovative			
Size (log)	0.411**	(0.046)	0.405**	(0.046)
Market share (log)	0.231**	(0.025)	0.241**	(0.025)
Part of a group	0.338**	(0.093)	0.324**	(0.095)
High-tech	0.838**	(0.170)	0.836**	(0.175)
Medium-high-tech	1.394**	(0.115)	1.418**	(0.115)
Medium-low-tech	0.463**	(0.097)	0.466**	(0.097)
Intercept	0.482	(0.384)	0.560	(0.382)
B)	Innovation input (logit)			
Persistence of innov. input				
High-tech	0.096**	(0.031)	0.094*	(0.041)
Medium-high-tech	0.074**	(0.020)	0.137**	(0.026)
Medium-low-tech	0.021	(0.018)	0.056*	(0.022)
Low-tech	0.042*	(0.017)	0.111**	(0.024)
Feedback of innov. output				
High-tech	0.089**	(0.032)	0.105**	(0.029)
Medium-high-tech	0.129**	(0.018)	0.113**	(0.016)
Medium-low-tech	0.122**	(0.019)	0.119**	(0.017)
Low-tech	0.110**	(0.018)	0.080**	(0.016)
Industry effects				
High-tech	1.346**	(0.168)	0.681**	(0.178)
Medium-high-tech	0.966**	(0.131)	0.440**	(0.146)
Medium-low-tech	0.200	(0.142)	-0.099	(0.146)
C)	Innovation output (logit)			
Lag effect of innov. input				
High-tech	0.143**	(0.042)	0.169**	(0.061)
Medium-high-tech	0.027	(0.025)	0.104**	(0.036)
Medium-low-tech	-0.005	(0.022)	0.090**	(0.031)
Low-tech	0.006	(0.022)	0.038	(0.033)
Persistence of innov. output				
High-tech	0.244**	(0.045)	0.230**	(0.043)
Medium-high-tech	0.204**	(0.025)	0.190**	(0.025)
Medium-low-tech	0.153**	(0.026)	0.144**	(0.026)
Low-tech	0.156**	(0.024)	0.146**	(0.024)
Industry effects				
High-tech	0.861**	(0.227)	0.832**	(0.259)
Medium-high-tech	0.405**	(0.172)	0.774**	(0.209)
Medium-low-tech	-0.038	(0.187)	0.310	(0.211)
D)	Extra parameters			
Initial innovation input (y_{1i0})	0.085**	(0.011)	0.062**	(0.013)
Initial innovation output (y_{2i0})	0.029*	(0.012)	0.023*	(0.012)
σ_η	1.579**	(0.072)	1.598**	(0.075)
σ_μ	0.525**	(0.042)	0.425**	(0.036)
σ_1	1.312**	(0.016)	1.083**	(0.015)
σ_2	1.550**	(0.036)	1.512**	(0.041)
λ	-1.034**	(0.181)	-1.468**	(0.244)
$\rho_{\eta\mu}$	0.131†	(0.071)	-0.251**	(0.083)
ρ_{12}	0.810**	(0.008)	0.822**	(0.008)
Number of observations	5591		5591	
Log-likelihood	-12485.809		-12716.204	

‡Note: the low-tech category is the reference, three time dummies are included in each equation.

Significance levels : † : 10% $*$: 5% $**$: 1%

Table 7: Wald tests of equality of dynamics of innovation parameters across categories of industry[‡]

R&D expenditures/sales	Total innov. expenditures/sales
Persistence of innovation input	
H$_0$: $\gamma_{11}^{H}=\gamma_{11}^{MH}=\gamma_{11}^{ML}=\gamma_{11}^{L}$	H$_0$: $\gamma_{11}^{H}=\gamma_{11}^{MH}=\gamma_{11}^{ML}=\gamma_{11}^{L}$
chi2(3)=8.31; p-value=0.040	chi2(3)=7.07; p-value=0.070
H$_0$: $\gamma_{11}^{H}=\gamma_{11}^{MH}$; $\gamma_{11}^{ML}=\gamma_{11}^{L}$	H$_0$: $\gamma_{11}^{H}=\gamma_{11}^{MH}=\gamma_{11}^{L}$
chi2(2)=1.24; p-value=0.537	chi2(2)=0.79; p-value=0.675
Feedback of innovation output	
H$_0$: $\gamma_{12}^{H}=\gamma_{12}^{MH}=\gamma_{12}^{ML}=\gamma_{12}^{L}$	H$_0$: $\gamma_{12}^{H}=\gamma_{12}^{MH}=\gamma_{12}^{ML}=\gamma_{12}^{L}$
chi2(3)=1.42; p-value=0.700	chi2(3)=4.31; p-value=0.230
Lag effect of innovation input	
H$_0$: $\gamma_{21}^{H}=\gamma_{21}^{MH}=\gamma_{21}^{ML}=\gamma_{21}^{L}$	H$_0$: $\gamma_{21}^{H}=\gamma_{21}^{MH}=\gamma_{21}^{ML}=\gamma_{21}^{L}$
chi2(3)=11.28; p-value=0.010	chi2(3)=5.07; p-value=0.167
H$_0$: $\gamma_{21}^{MH}=\gamma_{21}^{ML}=\gamma_{21}^{L}$	H$_0$: $\gamma_{21}^{H}=\gamma_{21}^{MH}=\gamma_{21}^{ML}$
chi2(2)=1.15; p-value=0.562	chi2(2)=1.41; p-value=0.494
Persistence of innovation output	
H$_0$: $\gamma_{22}^{H}=\gamma_{22}^{MH}=\gamma_{22}^{ML}=\gamma_{22}^{L}$	H$_0$: $\gamma_{22}^{H}=\gamma_{22}^{MH}=\gamma_{22}^{ML}=\gamma_{22}^{L}$
chi2(3)=5.46; p-value=0.141	chi2(3)=5.35; p-value=0.148
Joint test	
H$_0$: $\gamma_{11}^{H}=\gamma_{11}^{MH}=\gamma_{11}^{ML}=\gamma_{11}^{L}$; $\gamma_{12}^{H}=\gamma_{12}^{MH}=\gamma_{12}^{ML}=\gamma_{12}^{L}$; $\gamma_{21}^{H}=\gamma_{21}^{MH}=\gamma_{21}^{ML}=\gamma_{21}^{L}$; $\gamma_{22}^{H}=\gamma_{22}^{MH}=\gamma_{22}^{ML}=\gamma_{22}^{L}$	H$_0$: $\gamma_{11}^{H}=\gamma_{11}^{MH}=\gamma_{11}^{ML}=\gamma_{11}^{L}$; $\gamma_{12}^{H}=\gamma_{12}^{MH}=\gamma_{12}^{ML}=\gamma_{12}^{L}$; $\gamma_{21}^{H}=\gamma_{21}^{MH}=\gamma_{21}^{ML}=\gamma_{21}^{L}$; $\gamma_{22}^{H}=\gamma_{22}^{MH}=\gamma_{22}^{ML}=\gamma_{22}^{L}$
chi2(12)=28.80; p-value=0.004	chi2(12)=33.60; p-value=0.001
H$_0$: $\gamma_{11}^{H}=\gamma_{11}^{MH}$; $\gamma_{11}^{ML}=\gamma_{11}^{L}$; $\gamma_{12}^{H}=\gamma_{12}^{MH}=\gamma_{12}^{ML}=\gamma_{12}^{L}$; $\gamma_{21}^{MH}=\gamma_{21}^{ML}=\gamma_{21}^{L}$; $\gamma_{22}^{H}=\gamma_{22}^{MH}=\gamma_{22}^{ML}=\gamma_{22}^{L}$	H$_0$: $\gamma_{11}^{H}=\gamma_{11}^{MH}=\gamma_{11}^{L}$; $\gamma_{12}^{H}=\gamma_{12}^{MH}=\gamma_{12}^{ML}=\gamma_{12}^{L}$; $\gamma_{21}^{H}=\gamma_{21}^{MH}=\gamma_{21}^{ML}$; $\gamma_{22}^{H}=\gamma_{22}^{MH}=\gamma_{22}^{ML}=\gamma_{22}^{L}$
chi2(10)=13.70; p-value=0.187	chi2(10)=17.86; p-value=0.057

[‡]The superscripts H, MH, ML and L stand for high-tech, medium-high-tech, medium-low-tech, and low-tech respectively.

The estimation results suggest the presence of significant dynamics in the innovation process even after controlling for individual effects correlated with the initial values of the variables of interest. There is evidence of persistence of innovation input and innovation output in all four industry categories in the sense that lagged innovation input and lagged innovation output affect positively and significantly their current counterpart. One exception is the insignificant coefficient of the persistence of R&D in the medium-low-tech category. Furthermore, we find evidence of a significant lag effect of R&D on innovation output in the high-tech category, and a significant lag effect of total innovation spending on innovation output in all four but the low-tech category. Non-R&D innovation expenditures, e.g. the purchase of advanced machinery and computer hardware, play an important role in the generation of innovation output at least in the medium-high- and the medium-low-tech categories. Finally, we find a significant feedback effect of innovation output on innovation input in all four industry categories and regardless of how innovation input is measured.

As we mentioned earlier, the persistence of innovation output and its feedback effect on innovation input are similar across all four industry categories, but there are some differences across categories regarding the persistence of innovation input and its lag effect on innovation output.

The results also show evidence of significant firm effects, as shown by the significant estimated standard deviations of the random-effects, and significant industry effects. We already mentioned the differences across industry categories of two components of the dynamics, namely the the persistence of innovation input and its lag effect on innovation output. The results suggest further that, *ceteris paribus*, firms that belong to the high- and medium-high-tech categories of industry spend a significantly larger percentage of their sales in R&D and other innovation input components, and have a significantly larger share of innovative sales than those that belong to the medium-low- and low-tech categories.

Impulse response functions

Figures 1 and 2 show for all four industry categories impulse response functions (IRFs) derived from the tobit estimates of Table 6 with respectively R&D and total innovation spending as measures of innovation input. The IRFs are estimated for the population of continuously-innovative enterprises i.e. on the basis of equations (4) and (5), hence ignoring the selection equation. As a result, we drop the superscript "$*$" in the estimation of the IRFs as innovation input and innovation output are fully observed for continuously-innovative enterprises. Furthermore, while the censoring feature of y_{1it} and y_{2it} is accounted for in the estimation of the coefficients of the model, it is discarded in the estimation of the IRFs. To summarize, we rewrite equations (4) and (5) using matrix notation and dropping the individual subscript as

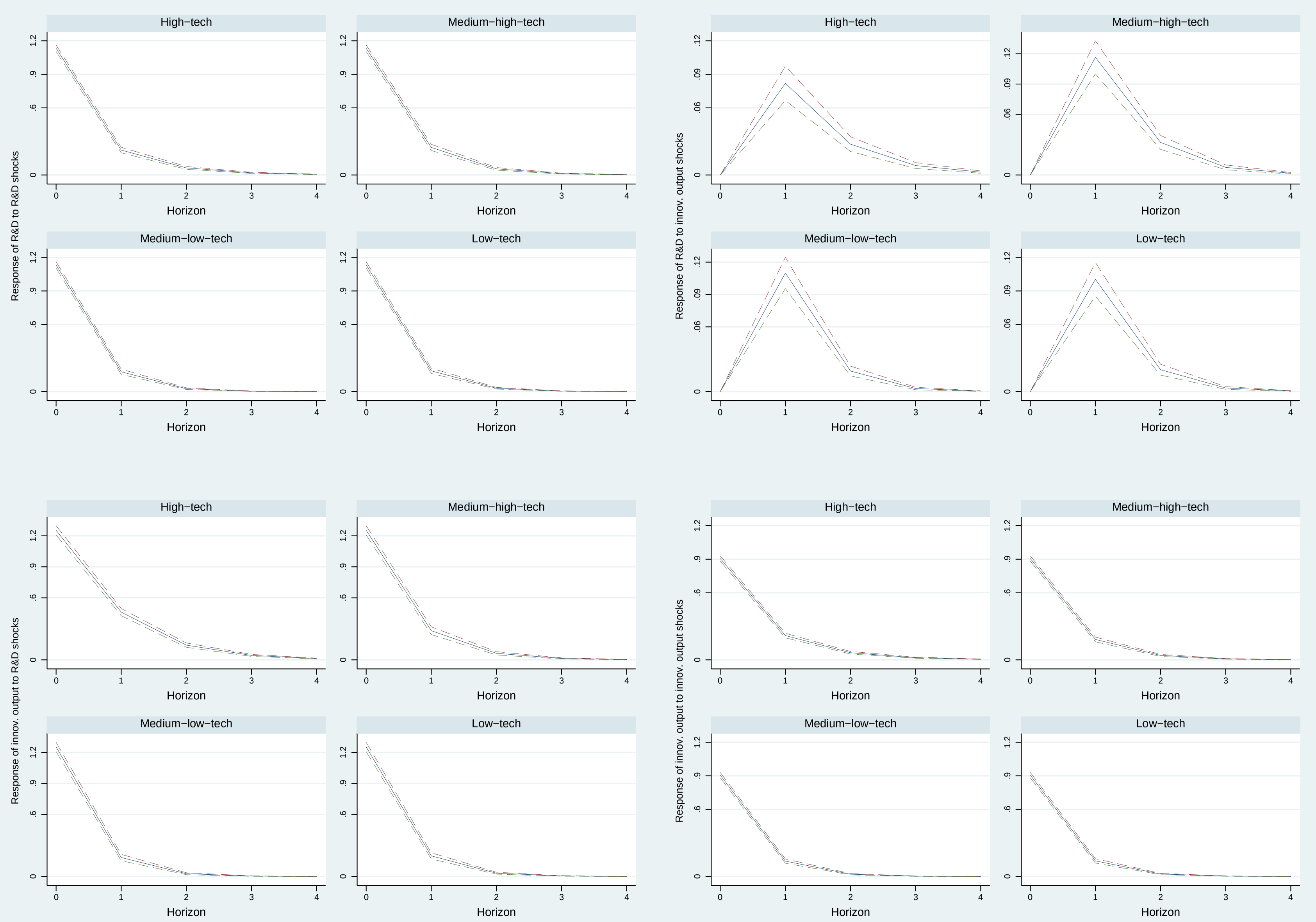
$$\mathbf{y}_t = \mathbf{\Gamma}\mathbf{y}_{t-1} + \mathbf{B}\mathbf{x}_t + \boldsymbol{\alpha} + \boldsymbol{\epsilon}_t, \tag{18}$$

where

$$\mathbf{y}_t = \begin{pmatrix} y_{1it} \\ y_{2it} \end{pmatrix}, \quad \mathbf{\Gamma} = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix}, \ \mathbf{B} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}, \quad \boldsymbol{\alpha} = \begin{pmatrix} \alpha_i \\ \lambda\alpha_i \end{pmatrix} \text{ and } \boldsymbol{\epsilon}_t = \begin{pmatrix} \epsilon_{1it} \\ \epsilon_{2it} \end{pmatrix} \rightsquigarrow N(0, \mathbf{\Sigma})$$

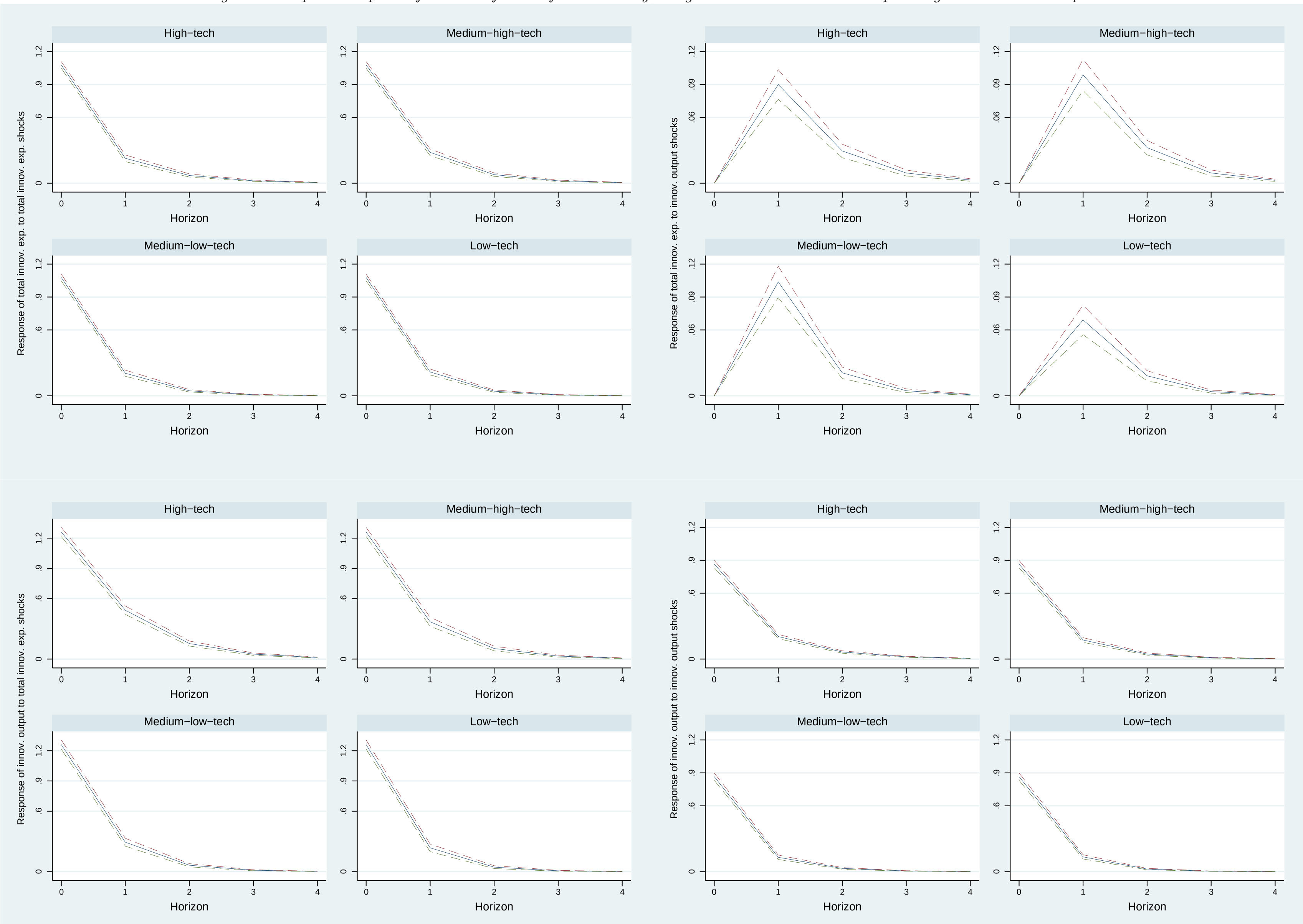
with $\mathbf{\Sigma} = \begin{pmatrix} \sigma_1^2 & \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$. Provided that $\mathbf{y}_t$ is covariance stationary (we also assume that the exogenous regressors are covariance stationary), we can invert equation (18) so as to have a vector

Figure 1: Impulse response functions for the four industry categories: R&D as innovation input measure‡



‡The confidence bands are the 5^{th} and 95^{th} percentiles of the distribution of the IRFs obtained from Monte Carlo simulations.

Figure 2: Impulse response functions for the four industry categories: Total innovation spending as innovation input measure‡



‡The confidence bands are the 5^{th} and 95^{th} percentiles of the distribution of the IRFs obtained from Monte Carlo simulations.

moving-average (VMA) representation written as

$$\mathbf{y}_t = \sum_{i=0}^{\infty} \mathbf{\Gamma}^i \boldsymbol{\epsilon}_{t-i} + \mathbf{B} \sum_{i=0}^{\infty} \mathbf{\Gamma}^i \boldsymbol{x}_{t-i} + \boldsymbol{\alpha} \sum_{i=0}^{\infty} \mathbf{\Gamma}^i, \tag{19}$$

where $\mathbf{\Gamma}^i$ denotes the simple impulse response functions at horizon i and measures the response of dependent variable j after i periods to a unit shock to dependent variable k holding everything else constant, $(j, k = 1, 2)$. Since the disturbances $\boldsymbol{\epsilon}_t$ are contemporaneously correlated, we cannot assume that everything else is held constant, i.e. equation (19) cannot provide a causal interpretation. As a result, we orthogonalize the disturbances as $\boldsymbol{\epsilon}_t = \mathbf{A}\mathbf{D}^{1/2}\mathbf{u}_t$ using the decomposition $\mathbf{\Sigma} = \mathbf{A}\mathbf{D}\mathbf{A}'$ where $\mathbf{u}_t$ denotes the orthogonalized disturbances and

$$\mathbf{A} = \begin{pmatrix} 1 & 0 \\ \frac{\sigma_{12}}{\sigma_1^2} & 1 \end{pmatrix}, \mathbf{D} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 - \left(\sigma_{12}^2/\sigma_1^2\right) \end{pmatrix}, \text{ and } \sigma_{12} = \rho_{12}\sigma_1\sigma_2. \tag{20}$$

The resulting orthogonalized impulse response functions are derived as

$$\mathbf{\Psi}^i = \mathbf{\Gamma}^i \mathbf{A}\mathbf{D}^{1/2} \tag{21}$$

and measures the response of dependent variable j after i periods, holding everything else constant, to one standard deviation shock to dependent variable k where innovation input is assumed to be determined prior to innovation output and hence comes first in the ordering. In order to obtain confidence intervals for the orthogonalized IRFs we randomly draw 500 sets of values for $\mathbf{\Gamma}$ from a normal distribution with the estimated mean and standard error and calculate the orthogonalized IRFs for each set. The 95% confidence interval is then given by the 5^{th} and 95^{th} percentiles of the simulated distribution of the orthogonalized IRFs.

The IRFs reported in Figures 1 and 2 suggest that the response of innovation input and innovation output to shocks of innovation input and innovation output does not go beyond four periods, which represents eight years. The most striking difference in the IRFs across industry categories lies in the response of innovation output to shocks to R&D which is significant after two periods (four years) in the high-tech and to a lesser extent in the medium-high-tech industry categories while it vanishes after two periods in the medium-low- and low-tech industry categories. More generally, the effects of the shocks on the variables of interest somehow vanish a bit faster in the medium-low- and low-tech categories than in the high- and medium-high-tech categories, which is particularly true for the response of R&D to shocks to R&D and innovation output.

Exogenous determinants

The effects of the supposedly exogenous explanatory variables (not reported) on innovation input are similar regardless of how it is measured. More specifically, *ceteris paribus* larger continuously-innovative enterprises do not spend more on R&D or other innovation input components, but those that have a smaller market share and innovation activities in cooperation have significantly larger R&D or total innovation expenditures. All three types of information sources of innovation namely internal, market and institutional sources condition R&D or total innovation expenditures that seem to be driven by market demand (product-oriented effects) rather than cost reduction (process-oriented effects) or environmental sustainability. As for innovation output, a similar pattern is observed with the exception of market share that is now positively and significantly correlated with innovation output. In other words, *ceteris paribus* larger continuously-innovative enterprises do not have a larger share of innovative sales, the three types of information sources of innovation are positively and significantly correlated with firm innovation success, and market demand is more important to innovation output than cost reduction and environmental sustainability.

Sample selection bias

The sample selection bias is assessed by the magnitude of the correlation ($\rho_{\eta\mu}$) between the individual effects of the selection equation and the dynamic bivariate tobit: the larger the correlation (in absolute value), the larger the bias. The estimated value of the correlation reported in Table 6 is rather small in absolute value and hardly significant when R&D is used as a measure of innovation input, hence indicating a rather small selection bias. Furthermore, the estimation results (not reported) of the dynamic bivariate tobit given continuously innovative enterprises, i.e. ignoring the sample selection effect, are very similar to those of the full model corroborating the evidence of a small selection bias. This result is not surprising given the specification of the full model. Indeed, if sample selection bias is present and significant in magnitude but uncontrolled for in the model, it will affect only the parameters of the explanatory variables of the regression(s) that enter significantly both the selection equation and the regression(s). However, because of the sampling scheme of the CIS (see Section 2), we have at our disposal very few explanatory variables that can enter both the selection equation and the dynamic bivariate tobit regression. While size, market share and being part of a group have a positive and significant effect on the probability of being continuously innovative, only market share matters in the equations of innovation input and innovation output. As a result, the lack of strong evidence of presence of sample selection may be partly due to a lack of explanatory power in the selection equation.

Robustness analysis

We have carried out a robustness analysis by estimating the model using different censoring thresholds namely $c_1 = 0.05$, $c_2 = 0.002$ and $c_1 = c_2 = 0.05$. The estimation results (not reported) suggest that the patterns of the dynamics of innovation remain unchanged. More specifically, we still find evidence of significant dynamics in the innovation process that differs across categories of industry mainly in terms of the persistence of innovation input and the lag effect of innovation input on innovation output. Similar groupings of the industry categories to those of Table 7, with regards to the components of the dynamics of innovation, are observed with these different censoring thresholds. The results also suggest that similar patterns of the firm and industry effects are observed. The standard deviations of the random effects are significantly estimated, and the estimates associated with the industry dummies show that, *ceteris paribus*, firms that belong to the high- and medium-high-tech categories of industry spend a significantly larger percentage of their sales in R&D and other innovation input components, and have a significantly larger share of innovative sales than those that belong to the medium-low- and low-tech categories.

As for the effects of the exogenous explanatory variables and the sample selection bias, the results are less robust to the new censoring thresholds than are those of the dynamics. Cooperation continues to influence innovation input, regardless of how it is measured, and innovation output. Both are still driven by market demand rather than cost reduction or environmental issues, and the three types of information sources of innovation affect positively and significantly innovation output. The effects of size, market share and the three types of information sources on innovation input are rather different than those of Table 6 and vary across measures of innovation input. More specifically, size now has a positive and significant effect on R&D intensity and market share has an even more negative effects on both measures of innovation input. Only institutional sources matter to innovation input when R&D is used, while market and institutional sources are important to total innovation spending. Finally, the magnitude of the sample selection bias increases in absolute value where the new estimated value of $\rho_{\eta\mu}$ is about -0.6. This finding seems to corroborate the lack of explanatory power of the selection equation as one explanation of the small selection bias that we have previously found. Indeed, we now have two explanatory variables that enter both the selection equation and the regression(s) namely size and market share, the effects of the latter being stronger (in absolute value) than previously. Had we more explanatory variables that explain significantly both the probability of being continuously innovative and innovation input and innovation output, we would expect the sample selection bias to be more precisely estimated and of significantly large magnitude.

5.2 VAR estimates

In order to emphasize the importance of accounting as much as possible for the features of the data, namely its (partial) truncation and censoring, we contrast the results of the full model that accounts for such features with those of a panel VAR that ignores at least the censoring characteristic of the data. In a first attempt to study the dynamics of innovation described in equations (4) and (5), we have estimated by maximum likelihood a panel VAR conditional on firms being continuously-innovative and ignoring the censoring feature of R&D, total innovation expenditures and the share of innovative sales. The individual likelihood of that model conditional on the individual effects is given by equation (15) where the outer exponent is equal to 1 and the inner exponent and the product of the two cdfs are removed from the expression. The resulting unconditional individual likelihood is obtained by replacing the new modified expression of equation (15) into equation (17). This approach to studying the dynamics of innovation was not successful because we could not obtain reliable estimates of the model. In particular, we could not obtain the standard errors of the estimates because the Hessian matrix or the outer product of gradients could not be inverted. That problem was due to too many zero and small values of innovation input and innovation output that "contaminated" the estimation by inducing measurement errors in the probability distribution functions of the log-likelihood of the panel VAR. In a second attempt, we have modified the panel VAR so as to correct for sample selection bias resulting in a sample selection panel VAR. The individual conditional likelihood of this second version of the panel VAR is given by the product of the expressions of equations (8) and (15) where the inner exponent is removed from the latter equation. This second approach was not successful either for the same reasons mentioned earlier. In order to achieve reliable estimates of the panel VAR, we had to restrict further the sample of continuously-innovative enterprises so as to have observed positive values of innovation input and innovation output. Hence, we estimated the panel VAR using the sample of continuously-innovative firms with innovation input and innovation output greater than or equal to 0.002 and 0.05 respectively, which represents about 60% of the sample of continuously-innovative firms when R&D is used as a measure of innovation input and 67% of that sample when total innovation spending is used. The individual conditional likelihood of this third version of the panel VAR is given by equation (15) where the inner and outer exponents are equal to 1 and the product of the two cdfs is removed from the expression. Table 8 shows ML estimation results of the third version of the panel VAR that suggest that only the persistence of innovation input and innovation output is significant in the dynamic structure of innovation, i.e. only the diagonal parameters of matrix $\mathbf{\Gamma}$ are positively and significantly estimated, but the firm and industry effects remain significant.

Evidently, the comparison *per se* between the estimates of Tables 6 and 8 is not meaningful because the full model and the third version of the panel VAR are estimated using two different samples. Nevertheless, we report the estimation results of the latter model to emphasize the costs of not taking account of the features of the data. Such costs are among others the inability to use the full "feasible" sample and the wrong inference made about the dynamics of the innovation process.

Table 8: ML estimates of the VAR given continuously-innovative enterprises with innovation input ≥0.002 and innovation output ≥0.05‡

Variable	Coefficient	(Std. Err.)	Coefficient	(Std. Err.)
	R&D expenditures/sales		Total innov. expenditures/sales	
A)	Innovation input (logit)			
Persistence of innov. input				
High-tech	0.161^{**}	(0.041)	0.235^{**}	(0.045)
Medium-high-tech	0.119^{**}	(0.018)	0.228^{**}	(0.030)
Medium-low-tech	0.008	(0.018)	0.074^{**}	(0.025)
Low-tech	0.049^{*}	(0.019)	0.166^{**}	(0.028)
Feedback of innov. output				
High-tech	-0.012	(0.031)	-0.025	(0.034)
Medium-high-tech	0.001	(0.016)	-0.016	(0.018)
Medium-low-tech	0.010	(0.020)	-0.024	(0.020)
Low-tech	0.017	(0.021)	-0.027	(0.019)
Industry effects				
High-tech	1.856^{**}	(0.194)	1.245^{**}	(0.191)
Medium-high-tech	0.982^{**}	(0.141)	0.554^{**}	(0.159)
Medium-low-tech	-0.095	(0.153)	-0.381^{*}	(0.159)
B)	Innovation output (logit)			
Lag effect of innov. input				
High-tech	$0.103^{\dagger}$	(0.055)	$0.111^{\dagger}$	(0.057)
Medium-high-tech	-0.005	(0.025)	0.021	(0.035)
Medium-low-tech	-0.024	(0.025)	-0.015	(0.030)
Low-tech	-0.027	(0.026)	-0.026	(0.034)
Persistence of innov. output				
High-tech	0.250^{**}	(0.045)	0.240^{**}	(0.045)
Medium-high-tech	0.146^{**}	(0.023)	0.140^{**}	(0.023)
Medium-low-tech	0.074^{**}	(0.028)	0.067^{**}	(0.025)
Low-tech	0.124^{**}	(0.029)	0.086^{**}	(0.024)
Industry effects				
High-tech	0.989^{**}	(0.247)	1.008^{**}	(0.246)
Medium-high-tech	$0.350^{\dagger}$	(0.182)	0.515^{*}	(0.204)
Medium-low-tech	-0.033	(0.203)	0.052	(0.206)
C)	Extra parameters			
Initial innovation input (y_{1i0})	0.045^{**}	(0.011)	0.038^{**}	(0.014)
Initial innovation output (y_{2i0})	0.018	(0.012)	0.013	(0.012)
σ_μ	0.602^{**}	(0.033)	0.455^{**}	(0.054)
σ_1	0.702^{**}	(0.021)	0.881^{**}	(0.027)
σ_2	1.229^{**}	(0.025)	1.164^{**}	(0.033)
λ	0.555^{**}	(0.118)	1.186^{**}	(0.255)
ρ_{12}	-0.062	(0.041)	-0.089^{**}	(0.039)
Number of observations	1924		2168	
Log-likelihood	-5656.516		-6594.099	

‡Note: the low-tech category is the reference, three time dummies are included in each equation.
Significance levels : † : 10% ∗ : 5% ∗∗ : 1%

6 Conclusion

This study gives insights into the dynamic relationship between innovation input and innovation output in Dutch manufacturing using five waves of the Community Innovation Survey. We estimate a dynamic bivariate tobit with time-varying double index sample selection and find evidence of significant dynamics in the innovation process even after controlling for individual effects correlated with the initial values of the variables of interest. In other words, there is persistence of innovation input and innovation output, a lag effect of innovation input on innovation output that remains significant after four years in the high-tech industry, and a feedback effect of innovation output on innovation input. The result on the persistence of innovation input and innovation output is in accordance with Peters (2009) and Duguet and Monjon (2002) respectively who also use CIS data but consider qualitative measures of innovation input and innovation output. The result on the lag effect of innovation input, measured by R&D or total innovation spending, on innovation output, measured by the share of innovative sales, contrasts with that of Hausman et al. (1984) and Hall et al. (1986) who only find simultaneity between R&D and patents. Like those authors, we find that innovation input and innovation output are jointly determined as shown by the significant estimate of the correlation between the idiosyncratic errors of the two processes. Our findings also show that observed industry effects play an important role in the relationship. For instance, R&D has a lag effect on the share of innovative sales only in the high-tech sector, and the lag effect of innovation input (for both measures) on innovation output lasts longer in the high- and to a lesser extent in the medium-high-tech industry than in the medium-low- and the low-tech industry. Differences in innovation behavior cannot, however, be solely attributed to observable differences across firms (e.g. high-tech versus low-tech). Unobserved heterogeneity, through firm effects, plays a crucial role in accounting for differences in innovation behavior and must be modeled.

The main caveat of this study stems from the features of the CIS data. First, the truncated-censored feature makes it difficult to study the dynamics of the innovation process as dynamic tobit-type models with multiple equations must be used in order to achieve reliable estimates. In particular, a more parsimonious panel VAR cannot be used as it involves the costs of not being able to use the full "feasible" sample and making wrong inference about the dynamics. Secondly, the CIS data exhibits very little 'within' variation, which makes the use of distribution-free semiparametric techniques that rely on 'within' or time differences unfeasible. This involves the cost of making distributional assumption about the individual effects and the idiosyncratic errors which may yield an inconsistent ML estimator if the distributional assumptions are violated. However, given the features of the data, the study provides an alternative to the instrumental variable quasidifferenced method of Holtz-Eakin et al. (1988) to estimate panel VAR models. Finally, related to the truncated-censored feature of the data, we have at our disposal very few variables that allow to discriminate between innovative and non-innovative enterprises. This results in selection equation that is not very well specified and explains in part the lack of sample selection bias found in the analysis.

Appendix A Industry classification

Table 9: Descriptive statistics per category of industries: The OECD classification based on 3-digit SBI

Category	SBI	# obsv.	%	Size		Innovative	
Industry				Mean	Median	Occasional	Continuous
High-tech							
Aircraft, spacecraft	35.3	11	0.20	395	173	0.727	0.636
Pharmaceuticals	24.4	77	1.38	303	80	0.740	0.649
Office machinery	30	34	0.61	623	115	0.853	0.853
Radio, TV equip.	32	55	0.98	178	55	0.855	0.818
Medical, optical instr.	33	190	3.40	194	70	0.684	0.600
Whole category		367	6.56	260	75	0.738	0.668
Medium-high-tech							
Elect. machinery nec	31	162	2.90	178	73	0.815	0.778
Motor vehic., trailers	34	163	2.92	385	100	0.767	0.706
Chemicals excl. pharmaceuticals	24 excl. 24.4	368	6.58	242	110	0.807	0.758
Railroad, transport equipment nec	35.2, 35.4 and 35.5	24	0.43	120	96	0.750	0.750
M&E nec	29	703	12.57	160	80	0.762	0.704
Whole category		1420	25.40	208	85	0.780	0.727
Medium-low-tech							
Ships, boats	35.1	117	2.09	126	75	0.624	0.487
Rubber & plastic	25	318	5.69	124	85	0.774	0.711
Coke, petrol & fuel	23	26	0.47	364	49	0.769	0.654
Non-metallic minerals	26	247	4.42	156	88	0.644	0.518
Metals	27-28	899	16.08	147	74	0.636	0.529
Whole category		1607	28.74	146	76	0.666	0.563
Low-tech							
NEC, recycling	36-37	391	6.99	368	100	0.550	0.437
Wood & paper	20-22	918	16.42	180	79	0.562	0.425
Food & tobacco	15-16	646	11.55	311	100	0.649	0.526
Textiles & leather	17-19	242	4.33	128	65	0.562	0.471
Whole category		2197	39.30	246	85	0.585	0.462
Whole manufacturing	15-37	5591	100.00	209	81	0.668	0.572

References

Acemoglu, D., P. Aghion and F. Zilibotti, "Distance to Frontier, Selection, and Economic Growth," *Journal of the European Economic Association* 4 (2006), 37–74.

Acs, Z. J. and D. B. Audretsch, "Innovation, Market Structure, and Firm Size," *Review of Economic and Statistics* 69 (1987), 567–574.

Aghion, P., N. Bloom, R. Blundell, R. Griffith and P. Howitt, "Competition and Innovation: An Inverted-U Relashionship," *Quaterly Journal of Economics* 120 (2005), 701–728.

Amemiya, T., “Tobit Models: A Survey,” *Journal of Econometrics* 24 (1984), 3–61.

Blundell, R. and C. Meghir, “Bivariate Alternatives to the Tobit Model,” *Journal of Econometrics* 34 (1987), 179–200.

Cefis, E. and L. Orsenigo, “The Persistence of Innovative Activities A Cross-Countries and Cross-Sectors Comparative Analysis,” *Research Policy* 30 (2001), 1139–1158.

Charlier, E., B. Melenberg and A. van Soest, “An Analysis of Housing Expenditure using Semiparametric Models and Panel Data,” *Journal of Econometrics* 101 (2001), 71–107.

Cragg, J. G., “Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods,” *Econometrica* 39 (1971), 829–844.

Crépon, B. and E. Duguet, “Estimating the Innovation Function from Patent Numbers: GMM on Count Panel Data,” *Journal of Applied Econometrics* 12 (1997), 243–263.

D’Aspremont, C. and A. Jacquemin, “Cooperative and Noncooperative R&D in Duopoly with Spillovers,” *American Economic Review* 78 (1988), 1133–1137.

Duguet, E. and S. Monjon, “Creative Destruction and the Innovative Core: Is Innovation Persistent at the Firm Level?,” UCL Discussion Paper 02-07, 2002.

Eurostat, “Nace Rev.1,” Classifications, Statistical Office of the European Communities, 1992.

Gouriéroux, C., A. Montfort and A. Trognon, “Pseudo Maximum Likelihood Methods: Applications to Poisson Models,” *Econometrica* 52 (1984), 701–720.

Hall, B. H., Z. Griliches and J. A. Hausman, “Patents and R and D: Is There a Lag?,” *International Economic Review* 27 (1986), 265–283.

Hausman, J., B. H. Hall and Z. Griliches, “Econometric Models for Count Data with an Application to the Patents-R&D Relationship,” *Econometrica* 52 (1984), 909–938.

Holtz-Eakin, D., W. Newey and H. S. Rosen, “Estimating Vector Autoregressions with Panel Data,” *Econometrica* 56 (1988), 1371–1395.

Kamien, M. I. and N. L. Schwartz, “Market Structure and Innovation: A Survey,” *Journal of Economic Literature* 13 (1975), 1–37.

Kyriazidou, E., “Estimation of a Panel Data Sample Selection Model,” *Econometrica* 65 (1997), 1335–1364.

OECD, *Oslo Manual, Guidelines for Collecting and Interpreting Innovation Data*, 3rd edition (Paris: OECD Publishing, 2005).

———, *OECD Science, Technology and Industry Scoreboard 2007: Innovation and Performance in the Global Economy* (Paris: OECD; 2007 edition, 2007).

Pakes, A. and Z. Griliches, "Patents and R&D at the Firm Level: A First Look," NBER Working Paper No. 561, 1980a.

———, "Patents and R&D at the Firm Level: A First Report," *Economics Letters* 5 (1980b), 377–381.

Peters, B., "Persistence of Innovation: Stylized Facts and Panel Data Evidence," *Journal of Technology Transfer* 34 (2009), 226–243.

Raymond, W., *The Dynamics of Innovation and Firm Performance: An Econometric Panel Data Analysis*, Ph.D. thesis, Maastricht University (2007).

Raymond, W., P. Mohnen, F. Palm and S. Schim van der Loeff, "Persistence of Innovation in Dutch Manufacturing: Is it Spurious?," *Review of Economics and Statistics* (forthcoming).

Schumpeter, J. A., *The Theory of Economic Development* (Cambridge, MA: Harvard University Press, 1934).

———, *Capitalism, Socialism and Democracy* (New York: Harper and Brothers, 1942).

Wooldridge, J. M., "Simple Solutions to the Initial Conditions Problem in Dynamic Nonlinear Panel Data Models with Unobserved Heterogeneity," *Journal of Applied Econometrics* 20 (2005), 39–54.