

# An Empirically-Based Taxonomy of Dutch Manufacturing: Innovation Policy Implications

Wladimir Raymond\*, Pierre Mohnen†, Franz Palm‡ and  
Sybrand Schim van der Loeff §

June 2, 2004

## Abstract

The paper studies the degree of homogeneity of innovative behavior in order to determine empirically an industry classification of Dutch manufacturing that can be used for policy purposes. We use a two-limit tobit model with sample selection, which explains the decisions by business enterprises to innovate and the impact these decisions have on the share of innovative sales. The model is estimated for eleven industries based on the Dutch Standard Industrial Classification (SBI 1993). A likelihood ratio test (LR) is then performed to test for equality of the parameters across industries. We find that Dutch manufacturing consists of three groups of industries in terms of innovative behavior, a high-tech group, a low-tech group and the industry of wood, where firms seem to have a rather different innovative behavior from the remaining industries. The same pattern shows up in the three Dutch Community Innovation Surveys.

**Acknowledgements:** The empirical part of this study has been carried out at the Centre for Research of Economic Microdata at Statistics Netherlands. The authors wish to thank Statistics Netherlands, and in particular Bert Diederer, for helping us in accessing and using the Micronoom data set. The views expressed in this paper are solely those of the authors. The authors also wish to thank François Laisney, Patrick Waelbroek and participants at presentations in Maastricht, Strasbourg, Leuven and Lille for their helpful comments. The first author acknowledges financial support from METEOR.

**Keywords:** Generalized tobit, High-tech industry, Homogeneity, Innovation policy, Likelihood ratio test, Model of friction, Sample selection, Two-limit tobit model, TPP innovator

**JEL classification:** C34, C51, O33, O38

---

\*University of Maastricht, P.O. Box 616 6200 MD Maastricht, The Netherlands  
W.Raymond@KE.unimaas.nl

†University of Maastricht, MERIT and CIRANO; P.Mohnen@MERIT.unimaas.nl

‡University of Maastricht and CESifo fellow, F.Palm@KE.unimaas.nl

§University of Maastricht, S.Loeff@KE.unimaas.nl

# 1 Introduction

Industry classification is of great importance for innovation policy. In order to encourage R&D or innovation, governments are interested in targeting their support towards certain types of firms or industries. It is easier to work with a few well-defined industries than with a multitude of heterogeneous firms. But these industries should be as homogeneous as possible. Therefore, in order to avoid misclassification costs, industries have to be classified as accurately as possible. The objective of this paper is to study the degree of homogeneity of innovative behavior in order to determine empirically an industry classification of Dutch manufacturing that can be used for policy purposes.

Two well-known and frequently used taxonomies are the Pavitt (1984) and the Organization for Economic Cooperation and Development (OECD (1999)) industry classifications. Using data on about 2000 significant innovations in Britain from 1945-79, Pavitt (1984) derives a classification where three categories of innovating firms are identified: supplier-dominated, production intensive, and science-based. The OECD uses the ratio of R&D expenditures over gross output as an indicator to classify manufacturing industries of the OECD countries in the categories of high-technology, medium-high-technology, medium-low-technology and low-technology. In contrast, Baldwin and Gellatly (2000) argue against the use of too rigid industry classifications. They perform a principal component analysis (PCA) on different innovation indicators on new technology based firms in Canada. They find that industry classification is highly sensitive to the choice of the underlying indicators, and that high-technology firms can be found in the industries that are commonly labeled low-technology and vice versa.

The above classifications all suffer from the absence of a model of innovative behavior. The OECD classification relies on a single indicator, and is therefore too incomplete to be useful for innovation policies. The PCA approach is firm-based, therefore, unless very detailed information at the firm level is available to the policy maker, innovation policies based on this approach are almost impossible. The Pavitt classification, which is the most useful for innovation policies, is constructed more on *a priori* knowledge and assumptions about industry characteristics and interrelationships than on empirical findings regarding their innovative behavior. Like the OECD classification and the PCA approach, the Pavitt classification is essentially descriptive. He even reports regressions results that contradict his theory. The main contribution of our study is that we use, to set up an industry classification, an econometric model that explains the decisions by business enterprises in Dutch manufacturing to innovate and the impact of these decisions on the share of innovative sales. Our model belongs to the class of the generalized tobit models used among others by Crépon et al. (1996), Felder et al. (1996), Brouwer and Kleinknecht (1996), Mohnen and Dagenais (2001), and Mairesse and Mohnen (2001) to explain innovation performance in the presence of censored data.<sup>1</sup> It combines the two-limit tobit and the generalized tobit models. In order to study homogeneity among the manufacturing industries, we shall estimate the model for each industry as described in Appendix A. Then, we shall estimate the model for categories formed

---

<sup>1</sup>The first two studies use a generalized tobit model to study the determinants of innovation input (as measured by R&D or total innovation expenditures) while the other three study the determinants of innovation output (as measured by the share of innovative sales in total sales).

by grouping industries, and test whether industries included in a category are similar in terms of innovative behavior. A category is called homogeneous if the parameters of the model for each industry in that category are identical. To implement our model, we use firm-level data stemming from three Dutch Community Innovation Surveys (CIS) pertaining to the periods 1994-96, 1996-98 and 1998-2000, and merged with production (PS) and finance (SFO) survey data for the periods 1996, 1998 and 2000.

The paper is organized as follows. Section 2 describes existing taxonomies and implications for innovation policies. The model and the likelihood ratio (LR) test are explained in Section 3. In Section 4 we describe the data sources, the variables constructed from the data samples, and some descriptive statistics. Empirical results for the model and the industry classification resulting from the LR test, and its policy implications are presented and discussed in Section 5. In Section 6 we summarize and conclude.

## 2 Existing taxonomies and innovation policy

In the literature on technical change, existing taxonomies are often driven by the need to identify technology-intensive industries. Indeed, firms that belong to these industries behave differently: they tend to innovate more, conquer new markets, use available resources more efficiently and offer higher remuneration to their employees. It is also admitted that, in the knowledge-based economy, innovation is the key to firms' performance in terms of survival and productivity. Thus, the characteristics and the innovative behavior of an industry are important to innovation policy makers. In order to encourage innovation, they try to allocate, as efficiently as possible, resources to these industries. As a consequence, a misclassification of industries can have important social costs resulting from a suboptimal allocation of resources. In this section some existing taxonomies and the implications for innovation policies are explained.

### 2.1 The Organization for Economic Cooperation and Development (OECD) classification

There are two OECD industry classifications. In the first one, the OECD uses the direct R&D intensity expressed as the ratio of R&D expenditures over gross output or over value-added. The first classification consists of three categories of industries, namely the high-technology, the medium-technology, and the low-technology industries. This classification has several drawbacks, for instance, it is not stable across the two measures used for R&D intensity.<sup>2</sup> The revised indicator is the overall R&D intensity, which is the intensity of an indirect R&D. This latter measure is defined as the ratio of R&D expenditures plus technology embodied in intermediate and capital goods over gross output. The revised classification, which is currently used, consists of four categories of industries: the high-technology, the medium-high-technology, the medium-low-technology and the low-technology industries.<sup>3</sup>

The OECD taxonomy is based on one single innovation-input indicator, namely the R&D intensity. R&D is an important input to the innovation pro-

---

<sup>2</sup>See Hatzichronoglou (1997) for more details.

<sup>3</sup>Industries that belong to these categories are described in details in OECD (1999).

cess. Calvert et al. (1996) find that, for the European countries included in their analysis, enterprises with the highest R&D intensities have the highest returns to innovation as measured by the percentages of sales from new or improved products. Nevertheless, R&D is not the only important innovation-input indicator to innovation, and there are many other innovation policy instruments besides R&D support. For instance, a few of the wide-ranging policy reforms suggested by the OECD (1998) are ‘*improving the efficiency and the leverage of industrial R&D support*’, or ‘*strengthening technology diffusion mechanisms by encouraging more competition in product markets*’. Thus, the OECD taxonomy, because of its unidimensional feature, is not suitable for innovation policies and ignores various other facets of the innovation process.

## 2.2 A principal component analysis approach (PCA)

Baldwin and Gellatly (2000) use data from *The Survey of Operating and Financing Practices* provided by Statistics Canada for their classification. The population of interest in their study is the population of small new technology-based firms (NTBFs) that entered the commercial sector (both goods and services) during the period 1983-86, and survived to 1993. They state that, in existing taxonomies (such as the OECD taxonomy), small firms are underrepresented, since the aggregate measures at the industry level are often determined by large firms. They also point out that existing industry taxonomies have a unidimensional characteristic and are therefore too narrow and incomplete. To set up their classification, they perform a principal component analysis (PCA) by accounting for innovation indicators, such as worker skills and technology use, that are commonly used to classify industries. Their framework constitutes a multidimensional approach to industry classification, and reveals two findings: first, industry classifications are highly sensitive to the indicators that are used to construct them; and high-technology firms are found in many industries, especially in the ones that are labeled low-technology.

Policy implications depend strongly on the characteristics of the PCA approach to industry classification. As seen earlier, the PCA framework has a multidimensional feature. ‘High-tech’ firms may focus on developing innovations, e.g. by introducing new or improved products onto the market, integrating existing technologies into current production methods (technology use), or investing in improving worker skills (human capital development). These three activities may be disparate and, therefore, ‘high-tech’ firms may have different objectives. As a result, any innovation policy should account for these three objectives. The multidimensional facet of the PCA framework implies multidimensionally faceted innovation policies.

Another motivation of the authors is to deal with the underrepresentation of small firms, when industry aggregates are used to classify industries. Small new technology-based firms (NTBFs) are often ignored by policy makers. They suggest that, in the future, such firms should be given great considerations in innovation policies, as they are ‘*...spearheading the new technological revolution in many industries*’.

Finally, as we saw earlier, one of the findings of the authors is that high-tech firms may be found in low-tech industries and vice versa. That means industries consist of heterogeneous firms, as opposed to the assumption of homogeneity underlying industry classifications. Thus, a large number of firms

may be penalized when innovation policies are carried out at the industry level. Such firms are advanced firms that belong to the low-tech industries. These firms are ‘*often more sophisticated, ..., than ‘average firms’ in a ‘high-tech’ sector*’. Consequently, this type of firms should be treated differently from the remaining firms in the low-tech industries by policy makers. Similarly, low-tech firms belonging to the high-tech industries may be unfairly rewarded by innovation policy makers, and should be treated differently from the remaining firms in the high-tech industries.

According to the PCA approach to industry classification, innovation policies may favor certain indicators of technological performance over others, but they should be discretionary, and should not view the high-tech and low-tech divide between firms or industries as schematic. Specifically, the high-tech firms that belong to the low-tech industries should be identified by innovation policy makers. However, it is a difficult task, as policies are usually carried out at the industry level. Consequently, the PCA approach to industry classification may not be very useful for innovation policies.

### 2.3 The Pavitt classification

Pavitt (1984) uses data on about 2000 ‘significant innovations’ in Britain from 1945-79 for his classification. A significant innovation is qualified as such by ‘experts’, who for each sector, define the threshold of significance for an innovation. Pavitt (1984) starts from these significant innovations, identifies the innovating firms and their sectors of economic activity. He also identifies the type of the innovations (product or process innovations) according to whether or not they are produced and used in the same sector, and the sources of knowledge inputs to them (e.g. universities). On the basis of some *a priori* assumptions coming from previous theories and possibly supported by the data, he sets up an industry classification where similarities and differences among sectors in the sources, nature and impact of innovations, as well as in the characteristics of the innovating firms are explained. Pavitt’s classification consists of three categories, namely the ‘supplier-dominated’, the ‘production intensive’ and the ‘science-based’ categories. The middle category consists of two sub-categories: scale intensive and specialized-suppliers. At a two-digit level of the Minimum List Heading (MLH),<sup>4</sup> the distribution of industries across categories is as follows: the supplier-dominated category consists of innovating firms in textiles, leather and footwear; the scale intensive sub-category of the production intensive category consists of the industries of food, metal manufacturing, shipbuilding, motor vehicles and glass and cement; the specialized-suppliers sub-category consists of the industries of mechanical engineering and instrument engineering; and the science-based category consists of innovating firms in the industries of chemicals, electrical, and electronic engineering.

The first policy implication of the Pavitt classification is the relative importance of ‘science and technology push’ and ‘demand pull’ in innovative activities. Pavitt finds that in supplier-dominated and production intensive firms demand pull stimulates innovative activities. More specifically, he finds that in these categories of firms demand pull, as measured by investment activities in user industries, is likely to stimulate ‘*innovative activities in both the production en-*

---

<sup>4</sup>The MLH is the UK equivalent of the Standard Industry Classification (SIC).

*gineering departments of user firms, and the upstream firms supplying capital goods*'. No such evidence is found for science-based firms.

Another implication of his classification is the relative importance of product and process innovation. It suggests that, in product groups with a high number of science-based firms, i.e. firms with a high R&D intensity, it may be expected that a high proportion of product/market opportunities are generated outside the group products. In other words, R&D intensity influences strongly and positively product innovation (as defined by Pavitt in his study) in science-based firms. In such firms, patent intensity is expected to have a stronger influence on product innovation. Additionally, in industries with production intensive firms, Pavitt's classification suggests that a high proportion of resources may be expected to be devoted to process innovations, and that these firms have high capital intensities, plant size and industrial concentration.

Pavitt (1984) also examines the extent to which firms develop their own process innovations, or buy them from suppliers. Supplier-dominated firms, with small production plants, tend to buy innovations from suppliers. On the other hand, production intensive and science-based firms, with generally large production plants, develop in-house process innovations. In other words, the Pavitt taxonomy suggests that there is a positive relationship between the '*proportion of a sector's process technology generated in-house, ..., and the size of firms and production plant*'.

## 2.4 Aim of the study

As the OECD taxonomy relies on a single innovation input indicator, R&D intensity, it is too narrow and will be incomplete for policy purposes. R&D expenditures cannot explain by itself the complicated innovation process. Baldwin and Gellatly (2000) point out that the OECD taxonomy is likely to be biased, and therefore cannot be the basis for wide-ranging innovation policies.

The PCA approach considered by Baldwin and Gellatly (2000), overcomes the problem of incompleteness by accounting for various innovation indicators that are commonly used to classify industries, leading to a multidimensional approach. Despite interesting findings from the PCA approach, it is not fully satisfactory. First, it is based on a principal component analysis, letting the data speak without an underlying model. The PCA approach is therefore essentially descriptive. Furthermore, it is not of great help in recommending innovation policies, since it is based on firm level characteristics, while policies are very often industry-based. Its main contribution is that innovation policy measures should be discretionary.

The Pavitt taxonomy is the most useful of the taxonomies considered here for innovation policies. It accounts for various criteria of the innovation process to classify industries, and its policy implications are based on a 'theory' derived from experts opinions about industries and empirical evidence. However, the Pavitt taxonomy is based on the population of innovating firms. There is no information regarding the classification of non-innovating firms. Therefore, it provides no guidance for policy measures to encourage non-innovating firms to become innovators. Furthermore, significant innovations are defined by 'experts', and may therefore suffer from subjectiveness. The significant threshold fixed by the innovation experts is not necessarily the same for all sectors. Finally, Pavitt also calls his taxonomy a 'theory', and no formal tests of the theory

could take place. The resulting taxonomy is therefore dependent on the validity of the *a priori* assumptions and the experts' opinions. Indeed, it is not clear that his theory is always in accordance with the facts.<sup>5</sup>

Our study proposes an empirically-based taxonomy of the Dutch manufacturing industries with a view to innovation policies. We account for various innovation (input and output) indicators and study the innovative behavior of eleven manufacturing industries. We estimate a model of the determinants of innovation and test for homogeneity among industries. A category of industries is considered as homogeneous if we cannot reject the hypothesis that the parameters of the model for all the industries included in that category are identical.

### 3 The model

The model we consider in this study explains the decisions by Dutch manufacturing enterprises to innovate and the impact of these decisions on the share of innovative sales. It is derived from the literature on the determinants of innovation, and a few studies on this type using innovation survey data for the manufacturing industries of some European countries are described in Table 1.

Modelling the determinants of innovation strongly depends on the design of the Community Innovation Survey. In order to correct for selection bias, tobit-type models are commonly used in empirical studies. Our model resembles the generalized tobit model studied by Crépon et al. (1996), Felder et al. (1996), Brouwer and Kleinknecht (1996), Mohnen and Dagenais (2001), and Mairesse and Mohnen (2001). It consists of two equations, a selection equation that discriminates between TPP innovators and non-innovators, and a regression equation explaining the share of innovative sales for innovators. A TPP innovator is an enterprise that has implemented new or improved products, or new or improved processes during the period under review. Formally, our model can be written as:

$$y_1^* = \beta_1 x_1 + \varepsilon_1 \quad (1)$$

$$y_2^* = \beta_2 x_2 + \varepsilon_2, \quad (2)$$

where the dependent variable is defined as:

$$y_2 = \begin{cases} 0 & \text{if } y_1^* \leq 0 \text{ (non-innovators)} \\ c_1 & \text{if } y_1^* > 0 \text{ and } y_2^* \leq c_1 \text{ (process-only innovators)} \\ y_2^* & \text{if } y_1^* > 0 \text{ and } c_1 < y_2^* \leq c_2 \text{ (product innovators)} \\ c_2 & \text{if } y_1^* > 0 \text{ and } y_2^* > c_2 \text{ (large product innovators)} \end{cases} \quad (3)$$

As it stands in equations (1), (2) and (3), our model is a combination of the two-limit tobit model and the generalized tobit model.<sup>6</sup> More specifically, it is a two-limit tobit model with sample selection. It is different from the former in that it consists of two decisions (two latent variables), and from the latter in that it has two additional thresholds  $c_1$  and  $c_2$ .

<sup>5</sup>In an appendix, Pavitt (1984) reports regressions that contradict his theory.

<sup>6</sup>The two-limit tobit model is described in details in Maddala (1983) and Thomas (2000). A variant of this model is the so-called model of friction used by Rosett (1959).

Table 1: Empirical studies on the determinants of innovation in manufacturing using the Community Innovation Survey

Study	Country	Data	Def. of innovator	Heterogeneity treatment	Econometric technique
1. Brouwer and Kleinknecht, 1996	The Netherlands	CIS 1, 1990-1992 + other data	Product innovator	Dummy for high-technology opportunity industries <sup>‡</sup>	Generalized tobit
2. Crépon et al., 1998	France	French Innovation Survey, 1986-1990 + other data	Product, process, organizational, or marketing innovator	Industry dummies; sample of all manufacturing firms	Asymptotic least squares (ALS)
3. Mohnen and Dagenais, 2001	Denmark, Ireland	CIS 1, 1990-1992	Firm with innovative activities	Industry dummies; sample of all manufacturing firms	Generalized tobit
4. Mairesse and Mohnen, 2001	France, Belgium, Denmark, Germany, Ireland, Italy, The Netherlands, Norway	CIS 1, 1990-1992; CIS 2, 1994-1996	Firm with innovative activities	Sample split: high- and low-R&D industries; industry dummies in both groups	Generalized tobit
5. Janz and Peters, 2002	Germany	CIS 3, 1998-2000	Product innovator with positive innovation expenditures	Industry dummies; sample of all manufacturing firms	3-Stage least squares (3SLS)
6. Janz et al., 2003	Germany, Sweden	CIS 3, 1998-2000	Product innovator with positive innovation expenditures	Industry dummies; sample of R&D intensive firms	Generalized tobit + 2-stage least squares (2SLS)

<sup>‡</sup>the reference is the group of low technological industries consisting of the ‘supplier-dominated’ industries of the Pavitt classification



### *Equation (1)*

The latent variable  $y_1^*$  captures the expected post-innovation return to innovation. It is a function of explanatory variables included in the vector  $x_1$ . The latent variable  $y_1^*$  pertains to both non-innovating and innovating enterprises. Because of the way the innovation survey is designed, very few explanatory variables are available for both types of enterprises.<sup>7</sup> Besides the constant term, we have included into  $x_1$  two variables: size and relative size.<sup>8</sup> The former is measured by the number of employees (in natural logarithms), while the latter is the ratio of an enterprise's turnover over total industry's turnover (in natural logarithms). According to the Schumpeterian hypotheses, size and relative size (as measured by the market share) are expected to have a positive effect on innovation activities (input side of innovation). It is often argued that larger firms have better access to finance, therefore are more likely to engage in risky projects, and benefit from economies of scale. Several recent empirical studies find evidence that size affects positively the decision by firms to invest in R&D (see Crépon et al. (1996) and Felder et al. (1996)) and the probability to be product innovators (see Brouwer and Kleinknecht (1996), Mohnen and Dagenais (2001), and Mairesse and Mohnen (2001)). When  $y_1^*$  is sufficiently large, i.e. the expected return to innovation is high enough, an enterprise has enough incentive to innovate, then the observed  $y_1$  equals one, meaning that the enterprise is a TPP innovator. Finally,  $\varepsilon_1$  is the error term that captures the effects unaccounted for by the model, and  $\beta_1$  is the vector of parameters to be estimated.<sup>9</sup>

### *Equation (2)*

The second equation pertains only to TPP innovating firms. Information included in  $x_2$  in addition to size is available in the CIS only for innovators. The vector  $x_2$  comprises, besides the constant term, the variables size, demand pull, technology push, subsidies, partnership, R&D intensity, continuous R&D and non-R&D performer as defined in Section 4.2. All the studies mentioned in Table 1 account for a demand pull variable according to the Schmooklerian tradition. Technology push is included in  $x_2$  on the basis of the Schumpeterian tradition. We expect enterprises that receive subsidies to be more innovative, on the grounds that they do more R&D which may also be expected to affect positively innovation output. While continuous R&D is very often found to affect positively innovation output, there is mixed evidence about the effect of developing innovation in partnership. It is often argued that, experience and knowledge accumulated from past R&D influences positively innovation output, while developing innovation in partnership is relevant only to firms lacking of knowledge ('weak' innovators).<sup>10</sup> The dependent variable is the share of innovative sales in total sales. It is expressed as the ratio of sales from new or improved products over total sales. This figure is directly reported by the surveyed enterprises. As the original share of innovative sales lies within

---

<sup>7</sup>See Section 4.

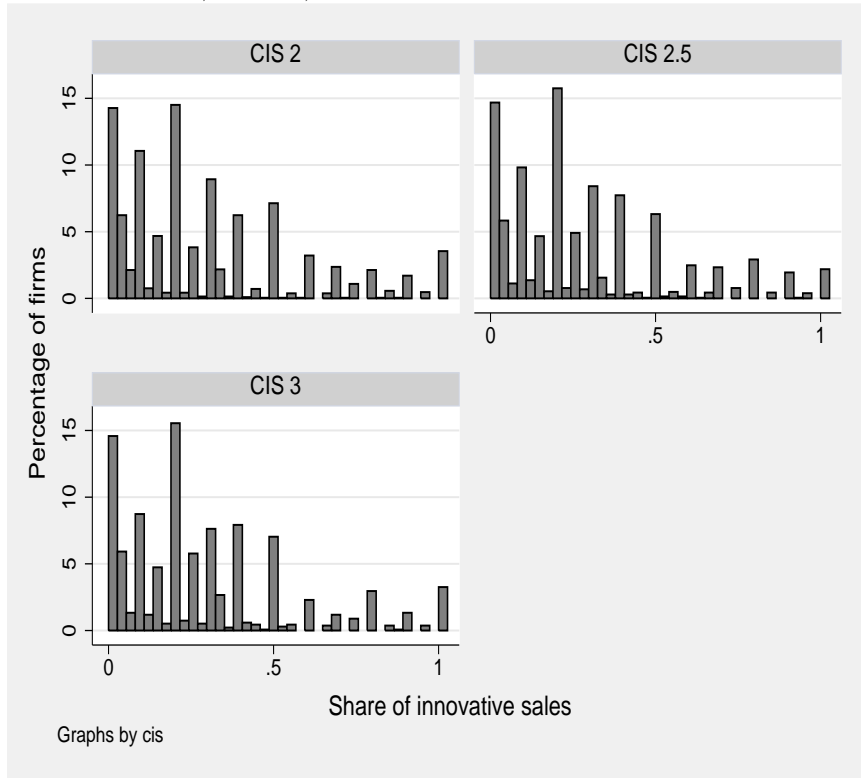
<sup>8</sup>In the innovation survey enterprises report whether they belong to a group or not. This potentially important information was, however, not available in the data provided by CBS.

<sup>9</sup>In order to simplify the notation, the enterprise and industry subscripts have been dropped.

<sup>10</sup>See Brouwer and Kleinknecht (1996) for details.

the unit interval, we take the logit transformation that may take on any real value. In addition, the logit transform makes the distribution of the share of innovative sales more symmetric than the distribution in Figure 1.  $\beta_2$  is the vector of parameters to be estimated, and  $\varepsilon_2$  is the error term capturing the effects influencing the share of innovative sales unaccounted for by the model.

Figure 1: Histogram of the share of innovative sales for the sub-sample of innovative firms: CIS 2, CIS 2.5, and CIS 3 The Netherlands



### *The thresholds $c_1$ and $c_2$*

TPP innovators comprise product-only innovators, process-only innovators, and product-and-process innovators. As there is no quantitative measure for a process innovation in the CIS, process-only innovators have a zero share of innovative sales during the period under review. Therefore, the logit transform of the share of innovative sales cannot be applied to process-only innovators. Mairesse and Mohnen (2001) choose to substitute, for the zero shares of process-only innovators, the smallest positive value of the original share of innovative sales, namely 0.01. Klomp and Van Leeuwen (2001) use a similar approach where the smallest positive value of the share of innovative sales is 0.001. In our model, we choose to censor the zero values for process-only innovators, by fixing a threshold  $c_1$  and treat all the zero values as values that belong to the censored

region.<sup>11</sup> The difference between our approach and that of Mairesse and Mohnen (2001), and Klomp and Van Leeuwen (2001) is that, for observations with share of innovative sales below the smallest positive value, the probability that the share is below this value has been used rather than the value of the density function evaluated at this value. The same ‘technical’ explanation holds for the category of large product innovators. In this case, values equal to unity are censored, by fixing the threshold  $c_2$  to 0.95, which is the value of the largest share less than one. Again, the probability that the share is above 0.95 has been used rather than the value of the density function evaluated at a value close to 1.

A further reason to introduce the thresholds  $c_1$  and  $c_2$  is that, according to the Oslo manual, process-only innovators are to be treated differently from product innovators and non-innovators. They are not to be treated as non-innovators, given that they have implemented new or improved processes. Actually, they produce unchanged products with changed methods, as opposed to non-innovators that produce unchanged products with unchanged methods. They are different from product-innovators in the sense that they have not implemented new or improved products (during the period under review), but they might do so in the future. Finally, the category of large product innovators, as measured by the share of innovative sales, mostly consists of firms that got established during the period studied by the innovation survey. Again, according to the Oslo manual, these firms are to be treated differently, as they may be different from the remaining firms in terms of innovation activities, objectives and characteristics.

Before we turn to the estimation of the model, it is to be noted that the distribution of the original share of innovative sales in total sales is rather similar in the three waves of innovation surveys included in this study. Figure 1 reveals that, innovating firms in the Dutch manufacturing during the period 1994-2000, have a rather small or medium share of innovative sales in total sales.

### 3.1 Estimation

In order to estimate the model, for each industry, the error terms are assumed to follow a bivariate normal distribution with zero mean and covariance matrix  $\Sigma = \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix}$ , where  $\rho$  and  $\sigma$  are the correlation between the error terms and the standard deviation of  $\varepsilon_2$  respectively. The model is estimated by the method of maximum likelihood, and the log-likelihood function is given by:<sup>12</sup>

<sup>11</sup>For the estimation, logit transforms of the thresholds  $c_1$  and  $c_2$  have been used.

<sup>12</sup>See Appendix B for the derivation of this likelihood function.

$$\begin{aligned}
\ln L = & \sum_0 \ln \Phi_1(-\beta_1 x_1) \\
& + \sum_{c_1} \ln \left[ \Phi_1 \left( \frac{\ln \left( \frac{c_1}{1-c_1} \right) - \beta_2 x_2}{\sigma} \right) - \Phi_2 \left( -\beta_1 x_1, \frac{\ln \left( \frac{c_1}{1-c_1} \right) - \beta_2 x_2}{\sigma}, \rho \right) \right] \\
& + \sum_{y_2^*} \ln \left[ \frac{1}{\sigma} \phi_1 \left( \frac{\ln \left( \frac{y_2}{1-y_2} \right) - \beta_2 x_2}{\sigma} \right) \times \Phi_1 \left( \frac{\beta_1 x_1 + \frac{\rho}{\sigma} \left[ \ln \left( \frac{y_2}{1-y_2} \right) - \beta_2 x_2 \right]}{\sqrt{(1-\rho^2)}} \right) \right] \\
& + \sum_{c_2} \ln \left[ 1 - \Phi_1(-\beta_1 x_1) - \Phi_1 \left( \frac{\ln \left( \frac{c_2}{1-c_2} \right) - \beta_2 x_2}{\sigma} \right) + \Phi_2 \left( -\beta_1 x_1, \frac{\ln \left( \frac{c_2}{1-c_2} \right) - \beta_2 x_2}{\sigma}, \rho \right) \right], \tag{4}
\end{aligned}$$

where  $\Phi_1$  and  $\phi_1$  are the univariate standard normal cumulative distribution and density functions, and  $\Phi_2$  is the bivariate standard normal cumulative distribution.<sup>13</sup> Furthermore, we assume the vectors of disturbances to be independent across industries such that, the log-likelihood function for the Dutch manufacturing sector is the sum of the log-likelihood functions for each industry in this sector. The disturbance covariance matrix  $\Sigma$  is allowed to be different across industries.

#### *A likelihood ratio (LR) test for homogeneity among industries*

All the empirical studies on the determinants of innovation mentioned in Table 1 implicitly assume that all industries in manufacturing have the same innovative behavior (in terms of the parameters of the model) and different intercepts. The only exception is the study by Mairesse and Mohnen (2001) that assumes that firms in the high-R&D industries have a different innovative behavior from firms in the low-R&D industries. However, the authors do not test this difference in innovative behavior. In our study, we account for the difference in innovative behavior across industries by estimating our model for each of the eleven Dutch manufacturing industries as described in Appendix A. We shall be referring to this as the unrestricted model. Let a category, for instance, consist of  $k$  industries. The null hypothesis is that the  $p$  parameters of the model, i.e. the slope parameters, the two intercepts and the parameters  $\rho$  and  $\sigma$ , for each of the  $k$  industries are equal versus the alternative that at least for one industry a parameter is different from the corresponding parameters of the  $k - 1$  other industries. The resulting test statistic must be compared to the critical value for a  $\chi^2$  distribution with  $(k - 1)p$  degrees of freedom. In the case that we allow the two intercepts in the  $k$  industries to differ, the number of degrees of freedom is  $(k - 1)(p - 2)$ .

## 4 Data

The data used to implement the model are collected by the *Centraal Bureau voor de Statistiek* (CBS). They stem from three waves of the Dutch Community

<sup>13</sup>The expressions  $\ln \left( \frac{c_1}{1-c_1} \right)$ ,  $\ln \left( \frac{c_2}{1-c_2} \right)$  and  $\ln \left( \frac{y_2}{1-y_2} \right)$  are the logit transforms of the thresholds and the original share of innovative sales respectively.

Innovation Survey, CIS 2, 2.5 and 3, merged with data from production and finance surveys (respectively PS and SFO). Only enterprises in Dutch manufacturing (SBI 15-37) are included in the analysis. The population of interest consists of enterprises with at least ten employees and a positive turnover at the end of the period covered by the innovation survey.

The Community Innovation Survey data are collected at the enterprise level. A combination of a census and a stratified random sampling is used. A census is used for the population of large enterprises, and a stratified random sampling is used for small- and medium-size enterprises. The size of an enterprise is measured by the number of employees, and the stratum variables are the economic activity and the size of an enterprise, where the economic activity is given by the Dutch standard industrial classification (SBI 1993).<sup>14</sup> Finally, the cut-off point used by CBS to choose between a census and a sampling is 100 employees.

Like in the CIS, the statistical unit of analysis within the Production Survey is the enterprise. A census is also used for large enterprises, and a stratified random sampling for small- and medium-size enterprises. The stratum variables and the cut-off points are the same as in the CIS. However, unlike the CIS, non-financial activities are not covered by the PS, implying that enterprises in the institutions of finance, insurance and pension funds (SBI 65-67) are not surveyed. The PS data included in this study replace the CIS equivalents and pertain to the periods 1996, 1998 and 2000. The corresponding CIS data have been removed from the CIS data base.

Finally, unlike the two surveys mentioned above, the statistical unit within the SFO is the ‘group of enterprises’. A group of enterprises (which we refer to as ‘company’) results from consolidating the activities of a collection of legally connected enterprises. The SFO, which consists of the SFGO (Finance Statistics for Large Companies) and the SFKO (Finance Statistics for Small Companies),<sup>15</sup> collects detailed information on balance sheet and income statement items for the population of non-financial companies established in the Netherlands.<sup>16</sup> As a result, data for companies whose ultimate benefit owners are Dutch, but are established abroad, are not available in the SFO. Additionally, like in the PS, enterprises in the institutions of finance, insurance and pension funds are not included in the consolidation of companies’ finances. Finally, the SFO data included in this study pertain to the same periods as for the PS, namely 1996, 1998 and 2000. They are used to replace financial variables that have been removed from the CIS data bank by CBS.

We explain now the construction of the variables included in this analysis.

## 4.1 Dependent variables

The variables of interest in this study are the incidence and the magnitude of innovation.

In the innovation survey, an enterprise is asked whether or not it has implemented at least one new or improved product, and whether or not it has implemented at least one new or improved process during the period under re-

---

<sup>14</sup>See Appendix A.

<sup>15</sup>A company is deemed to be large if its balance sheet is larger than a certain threshold. In the Netherlands, from 1977 until 1994, this threshold was 10 million guilders. From 1995 on, it became 25 million guilders.

<sup>16</sup>The ultimate benefit owner (UBO) may be Dutch or foreign.

view. The variable capturing the success of innovation is then constructed by assigning the value one if at least one of the questions is answered affirmatively. An enterprise having ongoing or abandoned innovative projects (without any successfully achieved innovative projects) is considered as a non-innovator for the time period under consideration.

The CIS questionnaire provides information regarding the share of innovative sales in total sales. This is the measure for the success of product innovation used in this study. A logit transformation of this measure is used for reasons mentioned earlier in the analysis. A non-TPP innovator has automatically a zero innovation output intensity. Because there is no quantitative measure of the success of process innovations in the innovation survey, some TPP innovators also have a zero innovation output intensity. They are labeled as process-only innovators.<sup>17</sup>

## 4.2 Explanatory variables

The explanatory variables can be classified into three categories: the enterprise characteristics, the enterprise activities and the industry characteristics. In the paragraphs below, we explain whether they are nominal or continuous, and whether they are constructed or directly extracted from one of the data sources mentioned earlier.

### 4.2.1 Enterprise characteristics

#### *Size*

The size of an enterprise used in this study is measured by the number of employees. This variable stems from the PS and is log-transformed.

#### *Relative size*

This figure, defined as the ratio of the turnover of an enterprise over the turnover of the industry that the enterprise belongs to, is constructed from the original turnover of an enterprise provided by the SFO data base. A log-transformation of this ratio is used.

#### *Demand pull*

In the CIS questionnaire, an enterprise is asked about the importance of the objectives of innovation, ‘open-up new markets’, ‘extend product range’ and ‘replace products phased out’, on the basis of a 0-3 Likert scale. A dummy variable proxying demand pull equals one for an enterprise if at least one of the above objectives of innovation is given the highest mark (i.e. very important), and zero otherwise.

#### *Technology push*

---

<sup>17</sup>In fact enterprises that report to have implemented at least one new or improved product but report a zero value for the share of innovative sales in total sales are considered as non-product innovators in the analysis.

Other innovation indicators available in the CIS are the sources of innovation. They may be internal (from the enterprise or other enterprises within the enterprise group), from the market (e.g. clients or customers), from public or private institutions (e.g. universities) or from other locations. Technology push is proxied by a dummy variable constructed from the indicators stating the importance of the sources of innovation from public or private institutions. This proxy takes on the value one if at least one of these institutions are deemed to be important or very important to an enterprise (i.e. at least one of the sources of innovation stemming from public or private institutes is given the values 2 or 3), and zero otherwise.

#### *Subsidy*

If an enterprise answers that it has been granted at least a subsidy during the period under review, the variable ‘subsidy’ takes on the value one and zero otherwise.

#### *Cooperation*

The dummy variable ‘cooperation’ takes on the value one if the enterprise reports that it undertook its innovative activities in cooperation of any kind and zero otherwise.

### **4.2.2 Enterprise activities**

#### *R&D intensity*

R&D intensity is the ratio of total (intramural and extramural) R&D expenditures (from the CIS) over total sales (from the SFO). The logarithmic transform of this variable is used when the variable takes on a positive value.

#### *Continuous R&D*

The dummy variable ‘continuous R&D’ takes on the value one if the enterprise reports that it performed intramural R&D continuously during the period under review and zero otherwise.

#### *Non-R&D performer*

The dummy variable ‘non-R&D performer’ takes on the value one if the enterprise had no (intramural or extramural) R&D expenditures during the period under review and zero otherwise.<sup>18</sup>

### **4.2.3 Industry characteristics**

The only industry characteristics available in the CIS and included in this study are the industry dummy variables. The manufacturing industries are defined according to a two-digit classification of the SBI 1993.

---

<sup>18</sup>In addition to the variable ‘R&D intensity’, the dummy variable ‘non-R&D performer’ is included in the analysis to ‘compensate’ for the fact that the log transform of the original R&D intensity is set up to zero for observations with original R&D intensity equal to zero.

### 4.3 Descriptive statistics

Tables 2, 3 and 4 of descriptive statistics show that, on average, the characteristics and the activities of the enterprises in Dutch manufacturing are rather similar in the three waves of the CIS. The ‘Total’ columns reveal that, during the three periods studied by the three innovation surveys, 64 per cent of the Dutch manufacturing enterprises are TPP innovators. Furthermore, the innovative sales of the Dutch manufacturing innovators account for about 29 per cent of their total sales during the three periods. A little over half of the Dutch manufacturing innovators are granted a subsidy, roughly one third work in partnership, and about 20 per cent consider that innovation sources from public and private institutions are important or very important (technology push) during the three periods. Also, among the TPP innovators, more than 75 per cent are R&D performers and the average amounts of R&D expenditures expressed as intensities are very similar for the three periods. Finally, the percentage of continuous R&D performers is rather similar, around 75 per cent, during the three periods. The only variables that vary from one wave to another are demand pull, size and relative size as measured by the market share. The values of these variables show a decrease from CIS 2 to CIS 2.5 and an increase in CIS 3.

It should be noticed that, for the three CIS, the enterprises in the industry of wood and paper (column ‘wood’) seem to be rather different from the enterprises in the manufacturing industries. These differences are particularly pronounced for the dependent variables, the variables pertaining to the firm activities and the variable subsidy. 54 per cent of the enterprises in the industry of wood are innovators while there are 64 per cent of such enterprises in the manufacturing sector as a whole. Moreover, the average return to innovation for TPP innovators of the enterprises in the industry of wood is much lower than the average return of the enterprises in total manufacturing. One explanation for this is that the industry of wood consists of a high percentage of process-only innovators. The percentage of TPP innovators in the industry of wood that are granted a subsidy is roughly 30 per cent while this percentage is more than 50 per cent in total manufacturing. Finally, looking at the variables of the enterprise activities, we notice that in the industry of wood, the percentage of R&D performers among TPP innovators, the percentage of continuous R&D performers among R&D performers and the actual average amount of R&D expenditures expressed as intensities are rather different from the equivalent figures for the manufacturing sector as a whole.

From the innovation indicators presented in Tables 2, 3 and 4, a tentative conclusion can be drawn with regard to the grouping of the industries (excluding the industry of wood). For one group, containing the manufacturing of chemicals, electrical products, plastic, and machinery and equipment, the indicators of innovation input and output and the subsidy indicator take on higher than average values. These indicators take on lower than average values in a second group containing the manufacturing of food, metallic and non-metallic products, textiles, and products not elsewhere classified.

In the following section it will be investigated empirically whether this grouping is purely descriptive or whether this grouping is the result of similar behavior described by an econometric model.



Table 2: Descriptive statistics: Means and standard deviations in parentheses for continuous variables CIS 2 The Netherlands

Variable	Chemicals	Elect- rical	Food	M&E	Metals	NEC	Non- metallic	Plastic	Textile	Vehicle	Wood	Total
Number of firms	225	285	424	384	460	212	142	170	234	217	541	3294
Dependent variables												
Percentage of innovators	0.84	0.72	0.66	0.77	0.59	0.55	0.65	0.76	0.49	0.61	0.54	0.64
Intensity of innovation for innovators	0.29 (0.25)	0.39 (0.29)	0.23 (0.23)	0.38 (0.27)	0.24 (0.24)	0.30 (0.26)	0.22 (0.26)	0.31 (0.25)	0.27 (0.25)	0.36 (0.31)	0.22 (0.26)	0.29 (0.27)
Independent variables												
Dummy for demand pull (for innovators)	0.69	0.68	0.68	0.65	0.52	0.56	0.46	0.62	0.57	0.58	0.54	0.60
Dummy for technology push (for innovators)	0.33	0.25	0.27	0.18	0.21	0.17	0.19	0.15	0.21	0.23	0.12	0.21
Dummy for subsidies (for innovators)	0.69	0.65	0.59	0.68	0.55	0.51	0.44	0.56	0.42	0.52	0.23	0.54
Dummy for innovation cooperation (for innovators)	0.46	0.37	0.37	0.25	0.36	0.26	0.34	0.42	0.23	0.34	0.24	0.33
Number of employees (in log)	4.70 (1.32)	3.97 (1.17)	4.37 (1.17)	4.17 (0.92)	4.04 (0.96)	3.65 (0.94)	4.04 (1.09)	4.03 (0.90)	3.68 (0.88)	3.97 (1.15)	4.04 (1.06)	4.08 (1.09)
Market share (in log)	-7.04 (1.71)	-7.11 (1.51)	-7.58 (1.78)	-6.59 (1.05)	-7.15 (1.19)	-6.11 (1.16)	-5.59 (1.16)	-5.70 (1.06)	-6.23 (1.30)	-6.91 (1.45)	-7.23 (1.30)	-6.85 (1.48)
Dummy for continuous R&D, if performing R&D	0.89	0.79	0.74	0.77	0.64	0.63	0.50	0.73	0.71	0.67	0.55	0.71
Dummy for non-R&D performers	0.07	0.15	0.16	0.09	0.25	0.17	0.23	0.14	0.25	0.23	0.41	0.20
R&D intensity, if performing R&D (in log)	-3.40 (1.50)	-3.17 (1.47)	-4.62 (1.71)	-3.47 (1.29)	-4.00 (1.37)	-3.56 (1.22)	-4.19 (1.55)	-3.63 (1.17)	-3.93 (1.59)	-3.70 (1.47)	-4.20 (1.60)	-3.81 (1.53)

Table 3: Descriptive statistics: Means and standard deviations in parentheses for continuous variables CIS 2.5 The Netherlands

Variable	Chemicals	Elect- rical	Food	M&E	Metals	NEC	Non- metallic	Plastic	Textile	Vehicle	Wood	Total
Number of firms	226	259	389	401	537	186	142	171	167	187	555	3220
Dependent variables												
Percentage of innovators	0.77	0.75	0.62	0.75	0.58	0.58	0.61	0.76	0.52	0.64	0.54	0.64
Intensity of innovation for innovators	0.32 (0.25)	0.37 (0.28)	0.23 (0.22)	0.36 (0.27)	0.25 (0.23)	0.30 (0.26)	0.18 (0.17)	0.30 (0.23)	0.22 (0.20)	0.37 (0.29)	0.21 (0.25)	0.28 (0.25)
Independent variables												
Dummy for demand pull (for innovators)	0.68	0.60	0.52	0.53	0.48	0.43	0.44	0.55	0.47	0.53	0.47	0.52
Dummy for technology push (for innovators)	0.23	0.22	0.19	0.20	0.14	0.13	0.29	0.15	0.11	0.22	0.11	0.18
Dummy for subsidies (for innovators)	0.69	0.61	0.50	0.69	0.50	0.41	0.44	0.64	0.45	0.57	0.28	0.53
Dummy for innovation cooperation (for innovators)	0.47	0.37	0.24	0.27	0.28	0.18	0.31	0.30	0.17	0.27	0.22	0.28
Number of employees (in log)	4.33 (1.31)	3.96 (1.17)	4.22 (1.25)	4.09 (0.95)	3.93 (0.99)	3.72 (0.92)	4.18 (1.10)	3.97 (0.93)	3.75 (0.91)	4.08 (1.15)	3.98 (1.10)	4.02 (1.09)
Market share (in log)	-7.26 (1.77)	-7.14 (1.51)	-7.70 (1.93)	-6.80 (1.13)	-7.41 (1.27)	-5.95 (1.14)	-5.66 (1.17)	-5.77 (1.12)	-6.01 (1.41)	-6.98 (1.59)	-7.33 (1.35)	-6.98 (1.56)
Dummy for continuous R&D, if performing R&D	0.84	0.80	0.73	0.78	0.64	0.63	0.66	0.73	0.67	0.71	0.58	0.71
Dummy for non-R&D performers	0.13	0.19	0.22	0.19	0.27	0.23	0.30	0.17	0.24	0.21	0.43	0.24
R&D intensity, if performing R&D (in log)	-3.44 (1.54)	-2.88 (1.52)	-4.72 (1.58)	-3.51 (1.28)	-4.16 (1.57)	-3.96 (1.47)	-4.41 (1.81)	-4.09 (1.26)	-4.00 (1.49)	-3.56 (1.59)	-4.29 (1.64)	-3.89 (1.60)

Table 4: Descriptive statistics: Means and standard deviations in parentheses for continuous variables CIS 3 The Netherlands

Variable	Chemicals	Elect- rical	Food	M&E	Metals	NEC	Non- metallic	Plastic	Textile	Vehicle	Wood	Total
Number of firms	200	134	255	292	347	124	87	104	80	95	386	2104
Dependent variables												
Percentage of innovators	0.81	0.73	0.63	0.70	0.59	0.69	0.57	0.76	0.51	0.68	0.52	0.64
Intensity of innovation for innovators	0.29 (0.24)	0.35 (0.27)	0.24 (0.25)	0.40 (0.29)	0.26 (0.26)	0.29 (0.24)	0.19 (0.17)	0.30 (0.19)	0.25 (0.20)	0.40 (0.28)	0.20 (0.23)	0.29 (0.26)
Independent variables												
Dummy for demand pull (for innovators)	0.75	0.73	0.78	0.82	0.72	0.78	0.80	0.80	0.71	0.75	0.67	0.75
Dummy for technology push (for innovators)	0.31	0.31	0.22	0.22	0.17	0.20	0.26	0.18	0.24	0.26	0.15	0.22
Dummy for subsidies (for innovators)	0.69	0.72	0.55	0.69	0.54	0.47	0.46	0.58	0.49	0.68	0.25	0.55
Dummy for innovation cooperation (for innovators)	0.51	0.47	0.29	0.29	0.29	0.26	0.32	0.32	0.37	0.38	0.26	0.33
Number of employees (in log)	4.66 (1.36)	4.60 (1.20)	4.69 (1.16)	4.26 (0.96)	4.30 (0.93)	4.20 (0.92)	4.38 (1.04)	4.41 (0.90)	4.17 (0.93)	4.79 (1.25)	4.27 (1.02)	4.41 (1.07)
Market share (in log)	-6.89 (1.70)	-6.07 (1.42)	-6.92 (1.76)	-6.71 (1.18)	-6.89 (1.19)	-5.47 (1.10)	-5.14 (1.18)	-5.16 (1.08)	-5.07 (1.25)	-6.15 (1.47)	-6.84 (1.28)	-6.46 (1.50)
Dummy for continuous R&D, if performing R&D	0.91	0.82	0.79	0.81	0.65	0.70	0.69	0.73	0.58	0.82	0.57	0.75
Dummy for non-R&D performers	0.15	0.14	0.25	0.16	0.30	0.30	0.22	0.24	0.20	0.22	0.44	0.25
R&D intensity, if performing R&D (in log)	-3.34 (1.59)	-3.21 (1.38)	-4.65 (1.70)	-3.26 (1.36)	-4.17 (1.66)	-3.71 (1.30)	-4.44 (1.18)	-3.96 (1.28)	-4.52 (1.38)	-3.77 (1.39)	-4.38 (1.74)	-3.87 (1.60)

## 5 Empirical results

This section discusses empirical results derived from our model. Our industry taxonomy stems from likelihood ratio tests performed on our model. The characteristics of the different categories in the classification are explained. We also present possible implications for innovation policies on the basis of the parameters of the model for each category.

Table 5: Likelihood ratio test results for the two-limit tobit model with sample selection: CIS 2 The Netherlands

Model	Number of firms	Log-likelihood	Number of parameters
Unrestricted model			
Chemicals	225	-482.38	14
Electrical	285	-579.37	14
M&E	384	-761.74	14
Plastic	170	-344.98	14
Vehicle	217	-429.35	14
Food	424	-845.27	14
Metals	460	-892.06	14
Non-metallic	142	-285.00	14
NEC	212	-374.76	14
Textile	234	-408.92	14
Wood	541	-1001.42	14
Restricted model 1			
High-tech industry	1281	-2625.41	22
Low-tech industry	1472	-2831.79	22
Wood	541	-1001.42	14
$\chi^2(96) = 106.72$ $p - value = 0.214$			
Restricted model 2			
High-tech industry	1281	-2625.41	22
Low-tech industry	2013	-3861.69	24
$\chi^2(108) = 163.68$ $p - value = 0.000$			
Restricted model 3			
High-tech industry	1822	-3716.48	24
Low-tech industry	1472	-2831.79	22
$\chi^2(108) = 286.03$ $p - value = 0.000$			

### 5.1 An econometrically-based industry classification

Tables 5, 6 and 7 report the relevant statistics to perform the likelihood ratio test. The model is estimated separately for eleven industries as described in Appendix A. The unrestricted model consists of eleven sub-models, where the log-likelihood is the sum of the eleven log-likelihoods of the sub-models. The restricted model assumes that the parameters are identical for industries within each category (allowing for different industry dummies), and different

Table 6: Likelihood ratio test results for the two-limit tobit model with sample selection: CIS 2.5 The Netherlands

Model	Number of firms	Log-likelihood	Number of parameters
Unrestricted model			
Chemicals	226	-465.44	14
Electrical	259	-528.18	14
M&E	401	-808.46	14
Plastic	171	-334.67	14
Vehicle	187	-366.98	14
Food	389	-733.76	14
Metals	537	-997.11	14
Non-metallic	142	-257.89	14
NEC	186	-346.73	14
Textile	167	-281.35	14
Wood	555	-1028.36	14
Restricted model 1			
High-tech industry	1244	-2529.39	22
Low-tech industry	1421	-2642.83	22
Wood	555	-1028.36	14
$\chi^2(96) = 103.30$ $p - value = 0.287$			
Restricted model 2			
High-tech industry	1244	-2529.39	22
Low-tech industry	1976	-3715.34	24
$\chi^2(108) = 191.56$ $p - value = 0.000$			
Restricted model 3			
High-tech industry	1799	-3658.28	24
Low-tech industry	1421	-2642.83	22
$\chi^2(108) = 304.35$ $p - value = 0.000$			

across categories. Tables 5, 6 and 7 show that the restricted model 1, i.e. the classification into three categories, is not rejected at the conventional 5% level of significance with a  $p$ -value ranging from almost 10% to almost 30%. When the industry of wood is classified into either the low-tech category (restricted model 2) or the high-tech category (restricted model 3), the restricted model is rejected with  $p$ -values below 1%.

Tables 5, 6 and 7 also reveal that our empirically-based industry classification holds for the three Dutch Innovation Surveys. This result is in accordance with similar descriptive statistics shown in Tables 2, 3, 4, and with similar distributions of the share of innovative sales in total sales reported in Figure 1 across the three waves of innovation survey. This is a rather strong result meaning that our model is supported by the data for the three innovation surveys.

Hence, an analysis of the innovative behavior of the enterprises in the Dutch manufacturing sector during 1994-2000 reveals that this sector can be classified into three categories. The categories consist of the industries of i) chemicals, electrical, M&E, plastic and vehicle (that we refer to as high-tech), ii) food,

Table 7: Likelihood ratio test results for the two-limit tobit model with sample selection: CIS 3 The Netherlands

Model	Number of firms	Log-likelihood	Number of parameters
Unrestricted model			
Chemicals	200	-426.05	14
Electrical	134	-274.85	14
M&E	292	-595.09	14
Plastic	104	-187.60	14
Vehicle	95	-194.44	14
Food	255	-496.08	14
Metals	347	-676.31	14
Non-metallic	87	-149.41	14
NEC	124	-257.93	14
Textile	80	-119.72	14
Wood	386	-680.55	14
Restricted model 1			
High-tech industry	825	-1707.72	22
Low-tech industry	893	-1727.00	22
Wood	386	-680.55	14
$\chi^2(96) = 114.47$ $p - value = 0.096$			
Restricted model 2			
High-tech industry	825	-1707.72	22
Low-tech industry	1279	-2423.21	24
$\chi^2(108) = 145.82$ $p - value = 0.009$			
Restricted model 3			
High-tech industry	1211	-2443.38	24
Low-tech industry	893	-1727.00	22
$\chi^2(108) = 224.67$ $p - value = 0.000$			

metals, non-metallic products, textiles, and products not elsewhere classified (that we refer to as low-tech); and iii) wood that seems to be different from the remaining industries in Dutch manufacturing. All the industries in each category have the same innovative behavior in the sense that, the parameters of the model for the industries in that category are identical with the exception of the intercepts.

We also test for homogeneity of the innovative behavior of the three categories across the three waves of innovation survey. Pooling the data of the three waves for each category, we are able to jointly test for equality of the parameters of the model for each category across waves, allowing for different industry and time intercepts. Assuming that the vector of disturbances are independent over time, we find that the innovative behavior of the three categories in CIS 2 is similar to that in CIS 2.5, with a  $\chi^2(36) = 46.72$  and a  $p-value = 0.109$ ; while the innovative behavior of the three categories in CIS 3 is different from that in CIS 2.5, with a  $\chi^2(36) = 64.72$  and a  $p-value = 0.002$ .

In Tables 8, 9 and 10 the descriptive statistics are shown of the three cate-

Table 8: Descriptive statistics for the high-tech, low-tech, and wood categories: Means and standard deviation in parentheses for continuous variables: CIS 2

The Netherlands			
Variable	High-tech industries	Low-tech industries	Wood
Number of firms	1281	1472	541
Dependent variables			
Percentage of innovators	0.74	0.59	0.54
Intensity of innovation for innovators	0.35 (0.28)	0.25 (0.25)	0.22 (0.26)
Independent variables			
Dummy for demand pull (for innovators)	0.64	0.56	0.54
Dummy for technology push (for innovators)	0.23	0.22	0.12
Dummy for subsidies (for innovators)	0.62	0.52	0.23
Dummy for innovation cooperation (for innovators)	0.35	0.32	0.24
Number of employees (in log)	4.17 (1.12)	4.02 (1.06)	4.04 (1.06)
Market share (in log)	-6.72 (1.43)	-6.83 (1.55)	-7.23 (1.30)
Dummy for continuous R&D, if performing R&D	0.78	0.66	0.55
Dummy for non-R&D performers	0.14	0.21	0.41
R&D intensity, if performing R&D (in log)	-3.44 (1.40)	-4.14 (1.59)	-4.20 (1.60)

gories based on the classification suggested by the testing procedure described above. The difference in the averages of the innovative activities and innovative output show up very clearly with the lowest values for the industry of wood. Firms that belong to the high-tech category have on average larger innovation output intensities, larger R&D intensities, receive more often innovation subsidies and innovate more frequently than firms in the low-tech category and in the industry of wood. The first category of firms also perform R&D more frequently and more often on a continuous basis. However, the three categories of firms are not on average statistically different in size and have a rather similar market share.

## 5.2 Comparison with existing taxonomies

At this stage, it is worthwhile comparing our classification with the OECD and the Pavitt classifications.

When Dutch manufacturing industries are defined at a two-digit level like in Appendix A, and when the two-category version of the OECD classification is considered, two main differences between our classification and the OECD classification are to be noted. The first one is that the industry of plastic belongs to the high-tech category of our classification, while it belongs to the low-tech category of the OECD classification. The second one is that the industry of wood has a different innovative behavior from the remaining industries in Dutch manufacturing. Thus, this industry cannot be classified into either the

Table 9: Descriptive statistics for the high-tech, low-tech, and wood categories: Means and standard deviation in parentheses for continuous variables: CIS 2.5

The Netherlands			
Variable	High-tech industries	Low-tech industries	Wood
Number of firms	1244	1421	555
Dependent variables			
Percentage of innovators	0.74	0.59	0.54
Intensity of innovation for innovators	0.35 (0.27)	0.24 (0.23)	0.21 (0.25)
Independent variables			
Dummy for demand pull (for innovators)	0.57	0.46	0.47
Dummy for technology push (for innovators)	0.21	0.17	0.11
Dummy for subsidies (for innovators)	0.63	0.46	0.28
Dummy for innovation cooperation (for innovators)	0.34	0.26	0.22
Number of employees (in log)	4.09 (1.10)	3.98 (1.08)	3.98 (1.10)
Market share (in log)	-6.84 (1.49)	-6.96 (1.67)	-7.33 (1.35)
Dummy for continuous R&D, if performing R&D	0.77	0.65	0.58
Dummy for non-R&D performers	0.19	0.27	0.43
R&D intensity, if performing R&D (in log)	-3.47 (1.48)	-4.28 (1.61)	-4.29 (1.64)

high- and low-tech category of our classification, while it belongs to the low-tech category of the OECD classification. However, our classification could have been more different from the OECD classification had we considered a finer industrial aggregation.

There are many more differences between our classification and the Pavitt classification. First of all, our low-tech category comprises both the supplier-dominated and the scale intensive categories of the Pavitt classification. More specifically, the industries of food, metallic and non-metallic products of our low-tech category fall into the scale intensive category, while the industries of textile and products not elsewhere classified fall into the supplier-dominated category. Secondly, the industry of wood, which falls into the supplier-dominated category of the Pavitt classification, again behaves differently from the remaining manufacturing industries. Finally, our high-tech category consists essentially of industries that are called science-based in the Pavitt classification with two exceptions regarding the industries of M&E and vehicle that fall respectively into the supplier-dominated and the scale intensive categories of the Pavitt classification.

### 5.3 Innovation policy implications

The three categories of firms have different innovative behavior. Hence, innovation policy makers use have different instruments to stimulate innovation according to the category that a firm belongs to. In order to derive possible



Table 10: Descriptive statistics for the high-tech, low-tech, and wood categories: Means and standard deviation in parentheses for continuous variables: CIS 3

<u>The Netherlands</u>			
Variable	High-tech industries	Low-tech industries	Wood
Number of firms	825	893	386
Dependent variables			
Percentage of innovators	0.74	0.61	0.52
Intensity of innovation for innovators	0.34 (0.26)	0.25 (0.24)	0.20 (0.23)
Independent variables			
Dummy for demand pull (for innovators)	0.72	0.70	0.63
Dummy for technology push (for innovators)	0.25	0.20	0.14
Dummy for subsidies (for innovators)	0.67	0.51	0.26
Dummy for innovation cooperation (for innovators)	0.37	0.28	0.25
Number of employees (in log)	4.49 (1.14)	4.39 (1.02)	4.27 (1.02)
Market share (in log)	-6.39 (1.49)	-6.37 (1.57)	-6.84 (1.28)
Dummy for continuous R&D, if performing R&D	0.82	0.68	0.57
Dummy for non-R&D performers	0.19	0.28	0.43
R&D intensity, if performing R&D (in log)	-3.42 (1.45)	-4.31 (1.59)	-4.37 (1.78)

innovation policies, we shall discuss the estimated coefficients of the explanatory variables which are to be understood as innovation policy instruments.

In the industry of wood, the reference group is an enterprise that performs R&D occasionally, having no innovation cooperation of any kind, for which neither demand pull nor technology push are important, and not being granted innovation subsidies. In the high-tech and low-tech categories, the reference group is an enterprise having the same characteristics as above, and that belongs to the industries of chemicals products and products not elsewhere classified respectively. Findings derived from our study are compared with findings derived from previous empirical studies. Such findings can be found in Table 11.

### 5.3.1 The selection equation

#### *Size*

According to the Schumpeterian tradition, the size of an enterprise influences positively the decision by this enterprise to engage in innovation activities, e.g. R&D, as well as the probability to be a product innovator (Brouwer and Kleinknecht (1996)). All empirical studies considered in Table 11 find such an evidence. Size always affects positively and significantly the probability to innovate in our low-tech category. Hence, from an innovation policy standpoint, to encourage non-innovating firms to become innovators, letting firms grow in size is good for innovation. The effect of size is less evident for the high-tech category and the industry of wood.

Table 11: Empirical findings from studies on the determinants of innovation in manufacturing using the Community Innovation Survey

Study	Size	Relative size	Dem. pull	Tech. push	Cooperation	Performing R&D	R&D intensity	Contin. R&D	Subsidy
Probability to innovate									
1. Brouwer and Kleinknecht, 1996	+ effect	n.a	+ effect	n.a	no effect	n.a	+ effect	+ effect	n.a
2. Crépon et al. 1998	n.a	n.a	n.a	n.a	n.a	n.a	n.a	n.a	n.a
3. Mohnen and Dagenais, 2001	+ effect	n.a	+ effect**	n.a	n.a	n.a	n.a	n.a	n.a
4. Mairesse and Mohnen, 2001	+ effect	n.a	n.a	n.a	n.a	n.a	n.a	n.a	n.a
5. Janz and Peters, 2002	+ effect	no effect	n.a	n.a	n.a	n.a	n.a	n.a	n.a
6. Janz et al., 2003	+ effect	n.a	n.a	n.a	n.a	n.a	n.a	n.a	n.a
7. Raymond et al., 2004	+ effect <sup>†</sup>	+ effect* <sup>†</sup>	n.a	n.a	n.a	n.a	n.a	n.a	n.a
Innovation output: share of innovative sales									
1. Brouwer and Kleinknecht, 1996	- effect	n.a	+ effect	n.a	no effect	n.a	+ effect	+ effect	n.a
2. Crépon et al. 1998	no effect	n.a	+ effect	+ effect	n.a	n.a	+ effect	n.a	n.a
3. Mohnen and Dagenais, 2001	no effect	n.a	+ effect**	n.a	no effect	n.a	no effect	+ effect	n.a
4. Mairesse and Mohnen, 2001	+ effect	n.a	+ effect <sup>†</sup>	no effect	no effect	+ effect <sup>†</sup>	+ effect <sup>‡</sup>	+ effect <sup>‡</sup>	n.a
5. Janz and Peters, 2002	+ effect	n.a	no effect	no effect	n.a	n.a	n.a	n.a	n.a
6. Janz et al., 2003	- effect*	n.a	no effect	no effect	n.a	n.a	n.a	n.a	- effect <sup>§</sup>
7. Raymond et al., 2004	- effect	n.a	+ effect	no effect	+ effect <sup>‡§</sup>	+ effect <sup>§§</sup>	+ effect* <sup>†</sup>	+ effect* <sup>†</sup>	+ effect <sup>‡</sup>

\*\*only for Ireland; <sup>†</sup>only for low-tech industries; <sup>‡</sup>only for high-tech industries; \*only for Germany; <sup>§</sup>only for Sweden, \*<sup>†</sup>only for high- and low-tech industries; <sup>‡§</sup>there is some evidence for the three categories; <sup>§§</sup>there is some evidence for high- and low-tech industries;

### *Market share*

The variable ‘relative size’, as measured by the market share of an enterprise, plays a positive role, except for CIS 3, in the probability to innovate for the high-tech and low-tech categories. The effect of this variable on the probability to innovate is stronger for the high-tech category than for the low-tech category. This variable never matters for the industry of wood.

It should be noticed that, when size is included together with the market share as explanatory variables, the effect of size may turn out not to affect the probability to innovate since both variables are highly correlated. Furthermore, when controlling for technological opportunities and appropriability, the size or market share effect on the probability to innovate may be insignificant (Crépon et al. (1996)). We assume that technological opportunities and appropriability are captured by industry dummies, in the sense that it is easier to innovate in some industries than in others, and that the way of appropriability of innovation differs with different industries.

### **5.3.2 The regression equation**

#### *Size*

The literature on the determinants of innovation reveals that there is mixed evidence with regard to the effect of size on the share of innovative sales in total sales given that a firm is an innovator. Our finding is in accordance with Brouwer and Kleinknecht (1996), and Janz et al. (2003), i.e., for the three CIS and for each category, smaller innovators are more successful than the larger counterparts. In other words, smaller innovators introduce more new or improved products onto the market than the larger counterparts. Another explanation is that, new or improved products introduced onto the market by smaller firms undergo a more rapid diffusion than new or improved products introduced by larger firms. Hence, from a policy standpoint, in order to stimulate innovation, innovative SMES should be promoted as opposed to monolithic innovators. While size is good for getting firms onto the innovation board wagon, at least in low-tech industries, it rather seems to stand in the way of getting innovators less successful and should therefore not be pursued in this regard. This holds for all industries.

#### *Demand pull*

On the basis of the Schumpeter tradition, many empirical studies on the determinants of innovation include the variable ‘demand pull’ as an explanatory variable of innovation. Regardless of the proxy used for ‘demand pull’, a majority of these studies find a positive impact on innovation output intensity. Our finding is in accordance with these studies. In other words, for each category and during the three periods under review, the variable ‘demand pull’ has a positive impact on the share of innovative sales in total sales.

#### *Technology push*

Schumpeter’s tradition reveals that ‘technology push’ enhances innovation. However, many empirical studies find no evidence with regard to the positive

Table 12: Maximum likelihood estimation results: CIS 2 The Netherlands

Variable	Coefficient	(Std. Err.)	Coefficient	(Std. Err.)	Coefficient	(Std. Err.)
Intensity of innovation						
	High-tech		Low-tech		Wood	
Size	-0.19**	(0.07)	-0.69**	(0.12)	-1.19**	(0.29)
Demand pull	0.26 <sup>†</sup>	(0.14)	0.75**	(0.18)	1.08**	(0.40)
Technology push	-0.07	(0.16)	-0.09	(0.22)	-0.34	(0.61)
Cooperation in innovation	0.26 <sup>†</sup>	(0.15)	0.27	(0.20)	0.99*	(0.50)
Non-R&D performers	-0.47	(0.31)	-1.34**	(0.37)	-0.28	(0.82)
Continuous R&D	0.83**	(0.18)	0.50*	(0.21)	0.45	(0.52)
R&D intensity (in log)	0.16**	(0.05)	0.17**	(0.07)	-0.04	(0.16)
Subsidies	0.45**	(0.15)	0.15	(0.20)	0.73	(0.51)
Electrical	0.81**	(0.22)	-	-	-	-
M&E	0.83**	(0.20)	-	-	-	-
Plastic	0.48 <sup>†</sup>	(0.25)	-	-	-	-
Vehicles	1.14**	(0.25)	-	-	-	-
Food	-	-	-0.38	(0.35)	-	-
Metals	-	-	-0.08	(0.34)	-	-
Non-metallic	-	-	-0.56	(0.43)	-	-
Textile	-	-	0.31	(0.39)	-	-
Intercept	-0.74	(0.45)	2.81**	(0.61)	4.56**	(1.34)
Standard deviation of $\varepsilon_2$	2.19**	(0.08)	3.40**	(0.13)	5.57**	(0.39)
Probability to innovate						
Size	0.07	(0.09)	0.21**	(0.06)	0.09	(0.11)
Market share (in log)	0.21**	(0.07)	0.10**	(0.04)	0.07	(0.09)
Electrical	-0.37**	(0.14)	-	-	-	-
M&E	-0.39**	(0.15)	-	-	-	-
Plastic	-0.58**	(0.21)	-	-	-	-
Vehicles	-0.72**	(0.15)	-	-	-	-
Food	-	-	0.25 <sup>†</sup>	(0.15)	-	-
Metals	-	-	0.13	(0.12)	-	-
Non-metallic	-	-	0.12	(0.14)	-	-
Textile	-	-	-0.14	(0.12)	-	-
Intercept	2.19*	(0.86)	0.01	(0.45)	0.22	(1.05)
Standard deviation of $\varepsilon_1$			1 (Assumed)			
Correlation of the error terms	-0.67**	(0.07)	-0.88**	(0.02)	-0.96**	(0.01)
Number of firms	1281		1472		541	
Log-likelihood	-2625.41		-2831.79		-1001.42	
Significance levels :	<sup>†</sup> : 10%	* : 5%	** : 1%			

Table 13: Maximum likelihood estimation results: CIS 2.5 The Netherlands

Variable	Coefficient	(Std. Err.)	Coefficient	(Std. Err.)	Coefficient	(Std. Err.)
Intensity of innovation						
	High-tech		Low-tech		Wood	
Size	-0.21**	(0.07)	-0.79**	(0.11)	-0.41	(0.26)
Demand pull	0.74**	(0.13)	0.57**	(0.16)	1.10**	(0.37)
Technology push	0.04	(0.16)	0.32	(0.23)	0.77	(0.65)
Cooperation in innovation	0.28†	(0.14)	0.42*	(0.20)	0.30	(0.49)
Non-R&D performers	-0.42	(0.28)	-0.57†	(0.31)	0.54	(0.52)
Continuous R&D	0.81**	(0.18)	0.83**	(0.20)	0.01	(0.18)
R&D intensity (in log)	0.13*	(0.05)	0.15**	(0.06)	-0.09	(0.15)
Subsidies	0.44**	(0.15)	0.31†	(0.18)	0.41	(0.47)
Electrical	0.22	(0.23)	-	-	-	-
M&E	0.38†	(0.21)	-	-	-	-
Plastic	0.09	(0.26)	-	-	-	-
Vehicles	0.72**	(0.26)	-	-	-	-
Food	-	-	-0.52	(0.35)	-	-
Metals	-	-	-0.55†	(0.33)	-	-
Non-metallic	-	-	-0.87*	(0.43)	-	-
Textile	-	-	-0.44	(0.42)	-	-
Intercept	-0.56	(0.42)	3.21**	(0.56)	0.94	(1.31)
Standard deviation of $\varepsilon_2$	2.30**	(0.08)	3.31**	(0.12)	5.55**	(0.38)
Probability to innovate						
Size	0.01	(0.08)	0.20**	(0.06)	0.27**	(0.10)
Market share (in log)	0.24**	(0.06)	0.13**	(0.04)	-0.09	(0.07)
Electrical	-0.10	(0.13)	-	-	-	-
M&E	-0.20	(0.12)	-	-	-	-
Plastic	-0.45*	(0.18)	-	-	-	-
Vehicles	-0.44**	(0.14)	-	-	-	-
Food	-	-	0.22	(0.15)	-	-
Metals	-	-	0.17	(0.13)	-	-
Non-metallic	-	-	-0.06	(0.14)	-	-
Textile	-	-	-0.19	(0.13)	-	-
Intercept	2.50**	(0.75)	0.25	(0.46)	-1.68†	(0.92)
Standard deviation of $\varepsilon_1$			1 (Assumed)			
Correlation of the error terms	-0.83**	(0.03)	-0.93**	(0.01)	-0.96**	(0.01)
Number of firms	1244		1421		555	
Log-likelihood	-2529.39		-2642.83		-1028.36	
Significance levels :	† : 10%	* : 5%	** : 1%			

Table 14: Maximum likelihood estimation results: CIS 3 The Netherlands

Variable	Coefficient	(Std. Err.)	Coefficient	(Std. Err.)	Coefficient	(Std. Err.)
Intensity of innovation						
	High-tech		Low-tech		Wood	
Size	-0.26**	(0.09)	-0.87**	(0.16)	-0.85*	(0.35)
Demand pull	0.48*	(0.21)	1.52**	(0.27)	2.23**	(0.46)
Technology push	-0.19	(0.21)	0.70*	(0.29)	0.90	(0.65)
Cooperation in innovation	0.51**	(0.19)	0.46 <sup>†</sup>	(0.27)	1.03 <sup>†</sup>	(0.56)
Non-R&D performers	-0.94*	(0.43)	-0.64	(0.52)	-0.52	(0.96)
Continuous R&D	0.23	(0.26)	0.38	(0.30)	-0.27	(0.60)
R&D intensity (in log)	0.21**	(0.07)	0.09	(0.09)	-0.03	(0.18)
Subsidies	0.48*	(0.21)	0.26	(0.25)	0.05	(0.55)
Electrical	0.64*	(0.30)	-	-	-	-
M&E	0.99**	(0.25)	-	-	-	-
Plastic	0.44	(0.33)	-	-	-	-
Vehicles	1.38**	(0.35)	-	-	-	-
Food	-	-	0.54	(0.46)	-	-
Metals	-	-	0.51	(0.43)	-	-
Non-metallic	-	-	0.28	(0.59)	-	-
Textile	-	-	0.78	(0.62)	-	-
Intercept	0.05	(0.61)	1.95*	(0.91)	2.65	(1.79)
Standard deviation of $\varepsilon_2$	2.46**	(0.10)	3.62**	(0.18)	5.47**	(0.50)
Probability to innovate						
Size	0.31**	(0.10)	0.40**	(0.08)	0.13	(0.13)
Market share (in log)	-0.03	(0.08)	-0.03	(0.05)	-0.01	(0.10)
Electrical	-0.23	(0.17)	-	-	-	-
M&E	-0.26 <sup>†</sup>	(0.14)	-	-	-	-
Plastic	-0.13	(0.23)	-	-	-	-
Vehicles	-0.41*	(0.18)	-	-	-	-
Food	-	-	-0.44*	(0.18)	-	-
Metals	-	-	-0.34*	(0.15)	-	-
Non-metallic	-	-	-0.41*	(0.18)	-	-
Textile	-	-	-0.50**	(0.18)	-	-
Intercept	-0.74	(0.96)	-1.35*	(0.58)	-0.58	(1.20)
Standard deviation of $\varepsilon_1$			1 (Assumed)			
Correlation of the error terms	-0.80**	(0.05)	-0.90**	(0.03)	-0.97**	(0.02)
Number of firms	825		893		386	
Log-likelihood	-1707.72		-1727.00		-680.55	
Significance levels :	† : 10%	* : 5%	** : 1%			

effect of ‘technology push’ on innovation. Except for the category of low-tech industries in CIS 2.5, ‘technology push’ does not seem to play a role in the innovation output intensity. It is often argued that, when controlling for technological opportunities, appropriability and demand conditions, as we do using industry dummies, pro-Schumpeterian results turn out non-significant (Crépon et al. (1996)).

#### *Cooperation*

Unlike the empirical studies considered in Table 11, we find some evidence with regard to the role of cooperation in innovation output intensity. In some cases, cooperating innovators seem to be more successful than non-cooperating innovators with the same characteristics. This is particularly so for the high-tech category in the three waves of the CIS. However, the effect of cooperation is not strong, which is likely to be due to the complexity of the innovation process. Not all innovation partners may be relevant for the innovation process. From a policy standpoint, encouraging cooperation of any kind is a rather weak innovation policy instrument. May be some well-defined types of cooperation should be stimulated instead. This would be worth exploring in future research.

#### *R&D*

As shown in Table 11, the three variables pertaining to R&D usually have a positive and significant effect on innovation output intensity.

There is some evidence that, among the innovators in the low-tech and high-tech categories, non-R&D performers are less successful than R&D performers with the same characteristics. This dummy variable never matters for the industry of wood. Furthermore, except for CIS 3, in both the high-tech and low-tech categories, innovators that perform R&D on a continuous basis are more successful than those that perform R&D occasionally with the same characteristics. Again, there is no apparent effect of performing R&D continuously in the industry of wood. Finally, R&D intensity, as measured by total R&D expenditures over total turnover, always plays a positive and significant role in the innovation output intensity for the high-tech and for the low-tech (except for CIS 3) categories. This variable never matters for the industry of wood.

To stimulate innovation, R&D should be encouraged. In both the high-tech and low-tech categories, R&D performers should be motivated to perform R&D continuously. In the industry of wood, R&D is not a relevant policy variable.

#### *Subsidy*

In the three waves of the CIS, innovators in the high-tech category that are granted subsidies are more successful than those that are not granted subsidies with the same characteristics. In the low-tech category and in the industry of wood, innovation subsidies do not seem to make a difference. The results suggest a positive correlation between subsidies and innovation in high-tech industries. It may be, however, that subsidies are given primarily to firms in high-tech industries because they are known to be innovative.

## 6 Conclusion

Existing industry taxonomies are commonly based on descriptive statistics and hence suffer from the absence of an empirical model. The contribution of our study was to set up an industry classification of Dutch manufacturing that is econometrically-based. We have considered a model of the determinants of innovation that explains the decisions by business enterprises to innovate, and the impact of these decisions on the share of innovative sales in total sales. We estimated a two-limit tobit model with sample selection for each industry and performed likelihood ratio tests to classify industries into categories. Each category consists of industries for which the parameters of the model are identical. Our taxonomy suggests that industries in the Dutch manufacturing sector can be classified into three categories, namely the high-tech category, the low-tech category, and the industry of wood. The high-tech category consists of the industries of chemicals, electrical, M&E, plastic and vehicle. The industries of food, metals, non-metallic products, textiles, and products not elsewhere classified form the low-tech category. Our empirically-based industry classification of Dutch manufacturing shows the same pattern for the three Dutch Community Innovation Surveys. Furthermore, the innovative behavior of the high-tech category is the same in the three CIS, and that of the low-tech category is the same in CIS 2 and CIS 2.5.

Our main innovation policy suggestions are summarized as follows. To encourage non-innovating firms to become innovators, policy makers should let firms grow in size in the low-tech category. There is mixed evidence as for the variable ‘market share’. While larger firms are likely to innovate more often, given that they are innovators, they are less successful than the smaller counterparts. In the three categories and during the three periods of study, innovative SMES should be promoted as opposed to monolithic innovators. Demand pull enhances innovation in the three categories, where its role is the most important in the industry of wood. Technology push is not a good policy instrument, and cooperation should be used more precisely. Not all partners may be relevant for the innovation process. R&D remains one of the most important innovation input indicators. To promote innovation, R&D should be stimulated in the high-tech and low-tech categories, and never plays a role in the industry of wood. Finally, while subsidized firms in the high-tech category are more successful than their non-subsidized counterparts, these two types of firms are equally successful in the low-tech category and in the industry of wood.

Although our classification is rather robust, it is not to be seen as the ultimate one. Indeed, when classifying industries, existing taxonomies use certain characteristics of the firms in these industries. Thus, different characteristics yield different classifications. Furthermore, the classification may be sensitive to the choice of the model, and even to the choice of the explanatory variables used in the model. Therefore, before being considered as definitive, our classification should be tested using other models of innovation. We have considered a static model estimated for three cross-sectional data sets. We have also assumed that R&D is determined exogenously, and the observed innovation input (R&D) and output (innovative sales) are measured at the same period of time, namely at the end of the period under review. But, more and more recent empirical studies assume that there is a decision to engage in R&D, hence endogenize R&D. These studies are the variants of the Crépon-Duguet-Mairesse (CDM) model.



However, as far as we know, all these studies endogenize R&D in a static framework.<sup>19</sup> We believe that there may be a time lag between innovation input (e.g. R&D) and innovation output as measured by the share of innovative sales. A natural way to extend our model is to consider a dynamic panel data framework.

Finally, we have found some homogeneity in the innovative behavior across certain industries in the Dutch manufacturing sector. However, we implicitly assume that industries are homogeneous. But, within the categories of industries, and even within the industries, there may be some heterogeneity among firms. For instance, high-tech firms may be found in low-tech industries and vice versa. The result of Baldwin and Gellatly (2000) is still valid and has not been tested in this study. It would be worthwhile in future research accounting for unobservable heterogeneity by considering, for instance, a random parameter model.

---

<sup>19</sup>An exception is the study by Van Leeuwen (2002) who considers a dynamic variant of the CDM model.

## Appendix

### A The Dutch standard industrial classification

Table 15: The Dutch standard industrial classification 1993

Industry <sup>20</sup>	SBI code	Industry definition
Food	15-16	Manufacture of food, beverages and tobacco.
Textile	17-19	Manufacture of textiles, wearing apparel, dressing and dyeing of fur, tannings and dressing of leather, luggage , handbags, saddlery, harness and footwear.
Wood	20-22	Manufacture of wood, products of wood and cork, except furniture, manufacture of straw and plaiting materials, pulp, paper and paper products, publishing, printing, and reproduction of recorded media.
Chemicals	23-24	Manufacture of coke, refined petroleum products and nuclear fuel, manufacture of chemicals and chemical products.
Plastic	25	Manufacture of rubber and plastic products.
Non-metallic	26	Manufacture of other non-metallic mineral products.
Metals	27-28	Manufacture of basic metals, fabricated metal products, except machinery and equipment.
M&E	29	Manufacture of machinery and equipment not elsewhere classified.
Electrical	30-33	Manufacture of office machinery and computers, electrical machinery and apparatus, radio, television and communication equipment and apparatus, medical, precision and optical instruments, watches and clocks.
Vehicle	34-35	Manufacture of motor vehicles, trailers semi-trailers and other transport equipment.
NEC	36-37	Manufacture of furniture and not elsewhere classified.

### B The likelihood function of the two-limit tobit model with sample selection

Consider the following model:

$$y_1^* = \beta_1 x_1 + \varepsilon_1 \quad (5)$$

$$y_2^* = \beta_2 x_2 + \varepsilon_2, \quad (6)$$

---

<sup>20</sup>The industries are defined at a two-digit level of the SBI 1993 in order to have enough observations for our inference on the one hand, and to preserve the confidentiality of firm-level information on the other.

where the dependent variable is defined as

$$y_2 = \begin{cases} 0 & \text{if } y_1^* \leq 0 (\text{non-innovators}) \\ c_1 & \text{if } y_1^* > 0 \text{ and } y_2^* \leq c_2 (\text{process-only innovators}) \\ y_2^* & \text{if } y_1^* > 0 \text{ and } c_1 < y_2^* \leq c_2 (\text{product innovators}) \\ c_2 & \text{if } y_1^* > 0 \text{ and } y_2^* > c_2 (\text{large product innovators}). \end{cases} \quad (7)$$

Furthermore, suppose that the error terms  $\varepsilon_1$  and  $\varepsilon_2$  are jointly and normally distributed with mean zero and covariance matrix  $\Sigma = \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix}$ . The likelihood function can be written as:

$$\begin{aligned} L &= \prod_0 P(y_1^* \leq 0) \prod_{c_1} P(y_1^* > 0; y_2^* \leq c_1) \\ &\quad \prod_{y_2^*} f(y_2|y_1^* > 0; c_1 < y_2 \leq c_2) P(c_1 < y_2 \leq c_2|y_1^* > 0) P(y_1^* > 0) \\ &\quad \prod_{c_2} P(y_1^* > 0; y_2^* > c_2). \end{aligned} \quad (8)$$

The contribution of a non-innovator to the likelihood function (first expression) is written as  $\Phi_1(-\beta_1 x_1)$ , where  $\Phi_1$  is the univariate standard normal cumulative distribution function. The contribution of a process-only innovator (second expression) involves the calculation of a double integral, i.e.  $\int_{-\beta_1 x_1}^{\infty} \int_{-\infty}^{\frac{c_1 - \beta_2 x_2}{\sigma}} \phi_2(\varepsilon_1, \frac{\varepsilon_2}{\sigma}, \rho) d\varepsilon_1 d\frac{\varepsilon_2}{\sigma}$ , where  $\phi_2$  is the bivariate standard normal density function. This latter expression can be written as  $\Phi_1(\frac{c_1 - \beta_2 x_2}{\sigma}) - \Phi_2(-\beta_1 x_1, \frac{c_1 - \beta_2 x_2}{\sigma}, \rho)$ , where  $\Phi_2$  is the bivariate standard normal cumulative distribution function. The third expression, corresponding to a not too large product innovator, is calculated as  $\frac{f(y_2|y_1^* > 0)}{P(c_1 < y_2 \leq c_2|y_1^* > 0)} P(c_1 < y_2 \leq c_2|y_1^* > 0) P(y_1^* > 0)$  so that it becomes

$$L = \prod_{y_2^*} f(y_2|y_1^* > 0) P(y_1^* > 0). \quad (9)$$

From the definition of a truncated distribution, we derive  $L_y$  as  $\int_0^{\infty} f(y_2, y_1^*) dy_1^*$ , where  $f$  is the bivariate normal density function. In order to make the computation easier, we write the bivariate density as a conditional multiplied with a marginal density so that:

$$\int_0^{\infty} f(y_2, y_1^*) dy_1^* = f(y_2) \int_0^{\infty} f(y_1^*|y_2) dy_1^*. \quad (10)$$

**Proposition 1** *If  $y$  is normally distributed with mean  $\mu$  and variance  $\sigma^2$ , then  $\int_0^{\infty} f(y) dy = \Phi(\frac{\mu}{\sigma})$ .*

**Proof.** The density function of  $y$  is  $\frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right]$  so that:

$$\int_0^{\infty} f(y) dy = \frac{1}{\sigma} \int_0^{\infty} \phi_1\left(\frac{y-\mu}{\sigma}\right) dy = \frac{1}{\sigma\sqrt{2\pi}} \int_0^{\infty} \exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right] dy.$$

Let  $z = \frac{y-\mu}{\sigma}$  such that  $dy = \sigma dz$ , then

$$\int_0^{\infty} f(y) dy = \frac{1}{\sigma} \int_{-\frac{\mu}{\sigma}}^{\infty} \phi_1(z) \sigma dz. \quad (11)$$

Using the property of symmetry of the normal distribution, the integral (11) is finally written as:

$$\int_0^{\infty} f(y)dy = \int_{-\infty}^{\frac{\mu}{\sigma}} \phi_1(z) dz = \Phi_1\left(\frac{\mu}{\sigma}\right).$$

■

Applying the results of Proposition 1 to the conditional distribution of  $y_1^*|y_2$  which is normal with mean  $\mu_{y_1^*|y_2}$  and variance  $\sigma_{y_1^*|y_2}^2$ , we finally obtain the contribution to the likelihood function of one product innovator as:<sup>21</sup>

$$L_{y_2} = \frac{1}{\sigma_2} \phi_1\left(\frac{y_2 - \beta_2 x_2}{\sigma_2}\right) \Phi_1\left(\frac{\mu_{y_1^*|y_2}}{\sigma_{y_1^*|y_2}}\right).$$

Finally, the contribution of a large product innovator (fourth expression in (8)) also involves a double integral  $\int_{-\infty}^{\infty} \beta_1 x_1 \int_{\frac{c_2 - \beta_2 x_2}{\sigma}}^{\infty} \phi_2(\varepsilon_1, \frac{\varepsilon_2}{\sigma}, \rho) d\varepsilon_1 d\frac{\varepsilon_2}{\sigma}$  which equals  $1 - \Phi_1(-\beta_1 x_1) - \Phi_1(\frac{c_2 - \beta_2 x_2}{\sigma}) + \Phi_2(-\beta_1 x_1, \frac{c_2 - \beta_2 x_2}{\sigma}, \rho)$ .

## References

- [1] Amemiya, T. (1978), "The Estimation of a Simultaneous Equation Generalized Probit Model", *Econometrica* 46 (5), 1193-1205
- [2] Amemiya, T. (1985), *Advanced Econometrics*, Harvard University Press, Cambridge.
- [3] Arvanitis, S. and H. Hollenstein (1996), "Industrial Innovation in Switzerland: A Model-Based Analysis with Survey Data", in A. H. Kleinknecht (ed.) *Determinants of Innovation : The Message from New Indicators*, Mc Millan, London.
- [4] Baldwin, J. R. and G. Gellatly (2000), "A Firm-Based Approach to Industry Classification: Identifying the Knowledge-Based Economy" in L.-A. Lefebvre, E. Lefebvre and P. Mohnen (eds), *Doing Business in a Knowledge-Based Economy. Facts and Policy Challenges*. Boston, Mass.: Kluwer Academic Publishers.
- [5] Baldwin, J., P. Hanel and D. Sabourin (2002), "Determinants of Innovative Activity in Canadian Manufacturing Firms", in A. Kleinknecht and P. Mohnen (eds.), *Innovation and Firm Performance. Econometric Explorations of Survey Data*, Palgrave, London.

<sup>21</sup>If  $(y_1, y_2)'$  follows a bivariate normal distribution with mean  $(\mu_1, \mu_2)'$  and covariance matrix  $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$ , then the conditional distribution of  $y_1|y_2$  is normal the mean and variance of which are respectively

$$\mu_{y_1|y_2} = \mu_1 + \frac{\rho\sigma_1}{\sigma_2}(y_2 - \mu_2),$$

and

$$\sigma_{y_1|y_2}^2 = \sigma_1^2(1 - \rho^2).$$

- [6] Brouwer, E. and A. Kleinknecht (1996), “Determinants of Innovation: A Micro Econometric Analysis of Three Alternative Innovative Output Indicators”, in A. H. Kleinknecht (ed.), *Determinants of Innovation: The Message from New Indicators*, Mc Millan London.
- [7] Calvert, J., C. Ibarra, P. Patel and K. Pavitt (1996), “Innovation Outputs in European Industry (CIS)”, SPRU (UK).
- [8] Crépon, B., E. Duguet and I. Kabla (1996), “Schumpeterian Conjectures: A Moderate Support from Various Innovation Measures”, in A. H. Kleinknecht (ed.) *Determinants of Innovation : The Message from New Indicators*, Mc Millan, London.
- [9] Crépon, B., E. Duguet and J. Mairesse (1998), “Research and Development, Innovation and Productivity: An Econometric Analysis at the Firm Level”, *Economics of Innovation and New Technology* 7 (2): 115-158.
- [10] Diederer, B. (2001), MICRONOOM Micro-integratie van Economische Statistieken.
- [11] Eurostat (1999): “Annex II.4. Control and Logical Checks”, in *The Second Community Innovation Survey*, Statistical Office of the European Communities, Luxembourg.
- [12] Eurostat (1999): “Annex II.9. Micro-aggregation Process”, in *The Second Community Innovation Survey*, Statistical Office of the European Communities, Luxembourg.
- [13] Felder, J., G. Licht, E. Nerlinger and H. Stahl (1996), “Factors Determining R&D and Innovation Expenditure in German Manufacturing Industries” in A. H. Kleinknecht (ed.) *Determinants of Innovation : The Message from New Indicators*, Mc Millan, London.
- [14] Gouriéroux, C. (2000), *Econometrics of Qualitative Dependent Variables*, Cambridge University Press, Cambridge.
- [15] Greene, W. H. (2003), *Econometric Analysis*, 5<sup>th</sup> ed., Prentice Hall International.
- [16] Hatzichronoglou, T. (1997), “Revision of the High-technology Sector and Product Classification”, STI Working Paper 1997/2, OECD, Paris.
- [17] Janz, N., H. Lööf, and B. Peters (2003), “Innovation and Productivity: A Cross-country Comparison Between Germany and Sweden”, mimeo.
- [18] Janz, N. and B. Peters (2002), “Innovation and Innovation Success in the German Manufacturing Sector: Econometric Evidence at Firm Level”, Paper presented at EARIE 2002.
- [19] Kleinknecht A. (2000) “Indicators of Manufacturing and Service Innovation: Their Strengths and Weaknesses”, in J.S. Metcalf and I. Miles (eds.), *Innovation Systems in the Service Economy*, Kluwer Academic Publishers, Boston, Dordrecht, London.

- [20] Klomp, L. and G. Van Leeuwen (2001), "Linking Innovation and Firm Performance: A New Approach", *International Journal of the Economics of Business* 8 (3), 343-364.
- [21] Lööf, H. and A. Heshmati (2002), "On the Relationship Between Innovation and Performance: A Sensitivity Analysis", mimeo.
- [22] Maddala, G. S. (1983), *Limited-dependent and Qualitative Variables in Econometrics*, Econometric Society Monographs, 3, Cambridge University Press, Cambridge.
- [23] Mairesse, J. and P. Mohnen (2001), "To Be or Not To Be Innovative: An Exercise in Measurement", *STI Review Special Issue on New Science and Technology Indicators*, OECD, 27, 103-129.
- [24] Mohnen, P. and M. Dagenais (2001), "Towards an Innovation Intensity Index. The Case of CIS 1 in Denmark and Ireland", in A. Kleinknecht and P. Mohnen (eds.), *Innovation and Firm Performance. Econometric Explorations of Survey Data*, Palgrave, London.
- [25] OECD (1997), *OECD Proposed Guidelines for Collecting and Interpreting Technological Innovation Data -Oslo Manual*, second edition, OECD, Paris.
- [26] OECD (1998), *Technology, Productivity and Job Creation: Best Policy Practices*. OECD, Paris.
- [27] Organization for Economic Cooperation and Development (1999), *Science, Technology and Industry Scoreboard. Benchmarking Knowledge-Based Economies*, Paris.
- [28] Pavitt, K (1984), "Sectoral Patterns of Technical Change: Towards a Taxonomy and a Theory" *Research Policy* 13 (6): 343-373.
- [29] Rosett, R. N. (1959), "A Statistical Model of Friction in Economics", *Econometrica* 27 (2), 141-146.
- [30] Thomas, A. (2000), *Econométrie des Variables Qualitatives*, Dunod, Paris.
- [31] Van Leeuwen, G. (2002), "Linking Innovation to Productivity Growth Using Two Waves of the Community Innovation Survey", STI Working Paper 2002/8.