Helsinki Center of Economic Research

Discussion Papers

# Eligibility, awareness and the application decision: An empirical study of firm participation in an R&D subsidy program

Tanja Tanayama
Helsinki School of Economics and HECER

Discussion Paper No. 161
April 2007

ISSN 1795-0562

HECER – Helsinki Center of Economic Research, P.O. Box 17 (Arkadiankatu 7), FI-00014 University of Helsinki, FINLAND, Tel +358-9-191-28780, Fax +358-9-191-28781,
E-mail info-hecer@helsinki.fi, Internet www.hecer.fi

**HECER**
**Discussion Paper No. 161**

# Eligibility, awareness and the application decision: An empirical study of firm participation in an R&D subsidy program*

## Abstract

This paper analyzes the application for research and development (R&D) subsidies in Finland. Finnish firm-level data on applicants and potential applicants is used to characterize firms application behavior. In addition to analyzing the characteristics underlying application for R&D subsidies, this paper also examines the use of count data models in modeling the application for R&D subsidies. The findings of this paper suggest that firms that are the most likely to have eligible projects, are also aware of the R&D subsidy program. The results also suggest that the opportunity cost of applying is lower for firms quite at the beginning of their life cycle, and provide evidence that external knowledge is important in lowering the application cost. Industry level heterogeneity in application behavior seems to be related to the application activity of potential applicants rather than the awareness of the program. The model selection exercise indicates that in using a count data framework to model the application behavior it is important to take into account both unobserved heterogeneity and excess zeros.

**JEL Classification**: D21, O31, O38

**Keywords**: R&D, subsidies, application decision, eligibility, awareness, count data models

Tanja Tanayama

HECER
University of Helsinki
P.O. Box 17 (Arkadiankatu 7)
FI-00014 University of Helsinki
FINLAND

e-mail: tanja.tanayama@helsinki.fi

* I would like to thank especially Otto Toivanen and Tuomas Takalo for their advice and suggestions. In addition I would like to thank Pekka Ilmakunnas, Mariagrazia Squicciarini and participants of the FDPE workshop on Microeconomics and Industrial Organization for helpful comments. Financial support from Tekes is gratefully acknowledged. This research is part of the research project: the Allocation of Government R&D Subsidies: Selection, Screening and Strategic Firm Behavior.

# 1 Introduction

Direct research and development (R&D) subsidies to business sector are a widely used policy tool to encourage industrial R&D. They are the second largest and fastest growing form of industrial aid in developed countries (Nezu, 1998). In Finland where the data of this paper originates, R&D subsidies are the most important tool of innovation policy (Georghiu et al., 2003). Given the importance of R&D subsidies we know surprisingly little about the processes that allocate them. There is a widespread political urge to get plausible evidence about the effectiveness of this policy tool in terms of additionality, productivity and growth, but it seems that this pronounced focus on impact estimates has diverted attention from the issue of allocation. To get reliable evidence of the effectiveness and functioning of a policy tool like R&D subsidies, the participation process determining who is finally granted a subsidy has to be well understood (Heckman and Smith, 2004). The participation process consists of two decisions: application decision and granting decision. In addition to asking who is selected into R&D subsidy programs and how, it should be asked who applies for R&D subsidies and why. The objective of this paper is to provide a first explorative step toward understanding firms application behavior by analyzing the application for research and development subsidies in Finland.

Heckman and Smith (2004) provide three reasons why understanding the participation process is important:

1. Helps to identify the sources of inequalities in the receipt of government services.

2. Reveals information about the functioning of the program.

3. Provides information for more reliable econometric evaluation.

The first point stresses the need to have a thorough understanding of how different stages of the participation process shape the participation of different groups in a program. If there is unequal participation in a program it is important to know which stage of the process creates this inequality. For example is it the case that a specific group is less aware of a program than others, or is there unequal behavior at the application phase. The second point highlights that the participation process helps to understand how a program actually operates. Understanding the outcomes of choices made by potential participants on the one hand, and government bureaucrats on the other hand at different stages of the participation process helps identifying potential unexpected behavior not intended by the policy design.

The last point has to do with the selection bias related to microeconometric evaluation of different programs. There is a growing literature on quantitative evaluations of the effects of public R&D subsidies on private R&D activities, but results of the analyses are contradictory (see David, Hall, and Toole, 2000, and Klette, Møen and Griliches, 2000, for surveys). The confusing empirical findings have raised the question whether the econometric setups have been adequately specified (Klette, Møen and Griliches, 2000 and Jaffe, 2002). One of the major problems of these studies has been selection bias, which reflects the fact that outcomes of potential applicants who have not received a subsidy may differ systematically from what the outcomes of subsidized participants would have been in the absence of subsidies.[1] This selection bias makes it difficult to identify the effect of a public subsidy. An understanding of the participation process creating the selection problem provides basis for more reliable microeconometric evaluation studies.[2]

As mentioned above, the participation process consists of two decisions: an application decision by firms and a granting decision by the government. Potential participants decide whether to apply for a subsidy or not and the government bureaucrats administering R&D subsidies decide to which applicants to grant a subsidy. Often the latter is highlighted. The discussion about the allocation of R&D subsidies has centered on the question of whether the government can identify projects with high social returns that the private sector would not undertake on its own. Little attention is paid to the application behavior of firms.

In this paper I use Finnish firm-level data on applicants and potential applicants to characterize firms application behavior. To the best of my knowledge there are no previous studies focusing on the application phase of R&D subsidy programs. Blanes and Busom (2004) analyze the participation of firms in R&D subsidy programs, but their data does not allow distinguishing between the application and approval phases. Lichtenberg (1998) analyzes the determinants of allocation of public biomedical research expenditure. More specifically, he ana-

[1]Heckman, Ichimura, Smith and Todd (1998) provide an extensive treatment of the selection bias.

[2]There is a vast literature discussing the role of participation process in econometric evaluations of social programs. Heckman, LaLonde and Smith (1999) cover extensively issues related to econometric evaluation, and Heckman, Ichimura and Todd (1997, 1998) discusses participation process especially in relation to the method of matching.

lyzes how different characteristics of disease burden affect the amount of public research expenditure allocated on a disease. Feldman and Kelley (2001) in turn study how the attributes of a firm's R&D strategy affect the chances of winning an award from the Advanced Technology Program.

In addition to analyzing the characteristics underlying application for R&D subsidies, this paper also examines the use of count data models in modeling the application for R&D subsidies. The rich data at hand allows the identification of applicants and non-applicants, but it also contains information on the number of applications a firm has submitted during the observation period. This kind of data calls for a count data model. Given that there is little evidence on how to use count models in modeling application for R&D subsidies, it is not straightforward to decide what kind of a count data model should be used. As a result, various count models are estimated and compared.

The model selection exercise indicates that in using a count data framework to model the application behavior it is important to take into account both unobserved heterogeneity and excess zeros. Ignoring the issue that the sample consists of both non-applicants and potential applicants can distort the results. The interpretation of several regressors changes under the assumption that the sample is a mixture compared to an analysis conducted under the assumption that all the observations come from the same data generating process. Considering the sample as a mixture has also intuitive appeal. It provides a statistical method for identifying whether a firm is aware of the program or not.

The findings of this paper suggest that firms that are the most likely to have eligible projects, are also aware of the R&D subsidy program. In other words, the program seems to reach firms that are the most potential participants. In addition the results indicate that the opportunity cost of applying is lower for firms quite at the beginning of their life cycle. The results also provide evidence that external knowledge is important in lowering the application cost. Industry level heterogeneity in application behavior seems to be related to the application activity of potential applicants rather than the awareness of the program.

The structure of this paper is the following. Section 2 overviews the business funding activities of the Finnish Funding Agency for Technology and Innovation. Section 3 discusses the application decision in relation to the whole participation process that determines who participates in R&D subsidy programs. Section 4 introduces count data models, discusses features of the sample in question and presents model selection process. Section 5 describes the data and Section 6 provides the results obtained from different count data models. Section 7 presents some conclusions.

## 2 Overview of business R&D funding activities of Tekes[3]

The Finnish Funding Agency for Technology and Innovation (Tekes) is the principal public promoter of private R&D in Finland and also the most important public financier of business R&D. The primary objective of Tekes is to promote the competitiveness of Finnish industry and the service sector by technological means. This is done by providing funding and expert services to both business and public R&D. Public R&D consists of research conducted in universities, academic institutions and research institutes. According to the Tekes annual report 2004, Tekes funding amounted to 409 million euros in 2004, of which 237 million euros was allocated to the business sector. In terms of projects this translates into 2242 projects of which 1464 were business R&D projects. In this paper the focus is on the business R&D funding activities of Tekes.

Business R&D funding is meant for firms operating in Finland and striving to improve business operations by technological means (www.tekes.fi). However, one clear trend in the business funding of Tekes has been the increasing role of small and medium sized firms (SMEs).[4]Large firms are not excluded from Tekes funding, but requirements imposed on them are somewhat more stringent compared to SMEs. Large firms' projects should fulfil at least one of the following criteria: networking with SMEs and universities or research institutes, participation in a technology program (technology programs are explained in the next subsection), participation in an international R&D project and network, a project consisting mainly of industrial research, or research outcomes have to become public. Both in terms of applications received and amounts granted the relative share of SMEs grew steadily especially over the 90s. The same trend has continued after the 2000, but to a lesser extent. The share of applications by firms with less than 100 employees increased from 36 percent in 1990 to 69 percent in 2000 and the share of business funding allocated to SMEs rose from

[3]This section relies on publicly available material that consists of Tekes annual report 2004, "World of technology, Joy of innovation" brochure of Tekes and information from website www.tekes.fi concerning the business funding of Tekes.

[4]An enterprise is considered a SME if 1) it has less than 250 employees, 2) large firms ownership is under 25 % and 3) its yearly turnover is less than 40 million euros or its balance sheet total is not over 27 million euros.

22 percent in 1990 to 53 percent in 2000. In 2004, SMEs received 55 percent of the total business R&D funding of Tekes .

## 2.1 Funding instruments

Key funding instruments of Tekes are grants and low-interest loans. In 2004, 70 percent of the business R&D funding consisted of grants. In general the same criteria apply to both grants and subsidized loans. However, distance to market is a key element determining the suitable funding instrument: grants are directed to R&D work done at the early phases of the innovation process that involve greater uncertainty, namely generation of new knowledge and prototypes that provide a basis for new marketable applications. Subsidized loans and capital loans are aimed at R&D work in the later stages of the innovation process in which the focus is on developing a complete marketable product or service. In practice the distinction between different phases of the innovation process is not clear-cut and a project can incorporate both stages. As a result, Tekes funding can be a combination of several instruments.

Almost half of Tekes business R&D funding is steered through technology programs. Technology programs aim at strengthening key technologies and expertise from the perspective of Finland's future. In addition technology programs aim at promoting collaboration, networking, and the diffusion of technologies.

## 2.2 Funding criteria

Tekes uses a selective funding practice that follows specific predefined criteria to allocate the funding. The qualification criteria used in the project evaluation are related to:

a) *the business activity to be pursued* - The goal is to promote projects that generate profitable business opportunities for global markets.

b) *the technology, innovation or competence to be developed* - The technology, innovation or competence to be developed should be technologically new and challenging at least to the company itself. In addition, knowledge and know-how created within the project should generate long lasting competitive advantage to the company.

c) *the resources reserved for the proposed project* - To be realistic the project proposal should incorporate adequate human and economic resources and the overall economic stance of the company should be in order.

d) *co-operation within the project* - One central aim of Tekes funding is to promote both domestic and international networking with other companies, universities and research centers.

e) *societal benefits of the project* - Societal benefits that favor Tekes funding are: positive environmental effects, balanced regional development, amelioration of the Finnish working and living conditions, improvements to back up the development of social welfare,health-care and equality, and promotion of the national energy strategy.

f) *the effect of Tekes funding on the project* - The aim is that with Tekes funding the companies are willing to carry out more challenging R&D projects than they otherwise would and that by providing resources for efficient networking the funding enhances the widespread use of the benefits of the project in the Finnish economy.

Technical advisers evaluate each project proposal and compare it to other project proposals. Since the amount of funding is limited, it is not enough for a project proposal to fulfill the Tekes criteria in order to get funding, but it must also succeed in the competition against other proposals. The evaluation is done compared to the relevant domestic and international reference group. In addition to the project, also the company is evaluated.

# 3 Application as part of the participation process

Heckman and Smith (2004) decompose the participation process into five different stages: eligibility, awareness, application, acceptance and enrollment. Even though Heckman and Smith discuss participation in a social program such as a job training program, their framework can also be applied to R&D subsidies. The main scope of this paper is to analyze the application stage. In addition, eligibility and awareness will be discussed, even though the data at hand do not allow an empirical analysis of those phases. The last two stages acceptance and enrollment are outside the scope of this paper and are not discussed further. The third essay examines the acceptance stage.

## 3.1 Eligibility for R&D subsidies

Eligibility determines potential applicants - the target group for the policy. In relation to R&D subsidy programs, determining eligible applicants is often not a straightforward exercise. Project level eligibility is difficult to define in practice, and even if this could be done, there would hardly be appropriate project level data needed to construct a sample of potential applicants. As a result, it is common to define potential applicants based upon firm level eligibility. This is the case also in this study. All the manufacturing firms and firms belonging to the knowledge intensive service sector operating in Finland are regarded as potential applicants in this paper.

In the case of R&D subsidies, eligibility has a somewhat different connotation compared to social programs that often have explicit eligibility rules. R&D subsidies are allocated to specific innovation projects, so both the applicant (firm) and the innovation project have to be eligible for a subsidy. In Finland the basic eligibility criteria for firms is that that the firm operates in Finland and strives to improve business operations by technological means (www.tekes.fi). This means that basically any firm operating in Finland can apply for R&D subsidies. Eligibility at project level is far more complicated. There are no explicit eligibility rules for projects. The overall guideline is that in the long run, tax revenue from a project should outweigh the tax-paid public investment (www.tekes.fi). Publicly stated funding criteria basically determine eligibility, but they are very broad, abstract, numerous and rely on subjective evaluation (see previous section). This gives room for a variety of interpretations. Based on the official funding criteria, it is difficult to judge whether a project is eligible for Tekes funding or not.

## 3.2 Awareness of R&D subsidies

The difficulty in defining eligibility has implications for awareness. As Heckman and Smith (2004) argue, applicants have to be aware of the program and of their eligibility for it. A firm may be aware of R&D subsidies, but misinterprets eligibility. In the Finnish case I would argue that Tekes as such is well known in Finland. Tekes was established in 1983, so it has a well established position among the actors of Finnish technology policy. This is further supported by the role of Tekes as a centralized agency administering government R&D subsidies. In addition, Tekes has a strong regional representation through regional Employment and economic development centers (see www.te-keskus.fi). Tekes also has quite a good visibility in the Finnish media. In 2005, electronic media coverage of Tekes consisted of 2860 news.[5]

In terms of the funding Tekes provides the situation may be different. As noted above, the official funding criteria give room for a variety of interpretations and based on them, it is difficult to judge whether a project is eligible for Tekes funding or not. A potential applicant may be aware of Tekes, but incorrectly thinks that the project is not suitable for Tekes funding. This argument is supported by the fact that the majority of the applicants contact Tekes before submitting an application.[6] In fact, on their website, Tekes suggests potential applicants to contact Tekes to discuss the project idea before submitting an application (www.tekes.fi). Even though this kind of services provided by Tekes reduce the information barrier due to difficulties in determining whether a project is eligible for Tekes funding or not, they are unlikely to completely eliminate problems related to awareness.

## 3.3 The application decision

In the application stage potential applicants that are aware of Tekes funding decide whether to submit an application or not. In making this decision, a firm weights expected benefits against the costs of applying. The main benefit to the firm from R&D subsidies is that they reduce the cost of R&D. In relation to technology programs, Tekes also highlights the benefits from networking and information sharing between companies and research communities (http://www.tekes.fi/English/programmes/what/what.html).

There are also costs associated with applying. It takes time and effort to gather the information required in the application process and to fill in the application form. Moreover the opportunity costs of the effort of making and promoting an application are probably far greater than the direct monetary costs of filling in and filing it. There are also additional administrative procedures associated with R&D subsidy programs: firms have to organize a separate bookkeeping for the subsidized project, Tekes approval is needed if the content of the project changes once the project is launched, and firms have to report about the progress and outcomes of the project during, at the end and after

[5]Information obtained from Tekes in May 2006. Tekes uses News Now -service from M-Brain (www.m-brain.fi/English/newsnow.html) to get information about media coverage.

[6]This observation came up during the interviews and discussions I conducted while staying 11 months at Tekes in 2001.

the project.[7] In addition, publicity requirements, related especially to projects funded within technology programs, may prevent some firms from applying.[8]

The above discussion highlights that there are several reasons why a firm may not send an application. First, a firm simply is not aware of the program. Second, a firm may be aware of the program but misinterprets eligibility. Third, the activities of a firm may be outside the scope of the program. Fourth, application costs are so high that it is not profitable for a firm to apply.

# 4 The econometric setup

The data at hand does not only provide information on whether a firm submitted an application to Tekes, but also the number of applications submitted by a firm during the observation period is known. As a result I analyze the application for R&D subsidies by examining the number of applications a firm has submitted to Tekes during the observation period. This set up calls for a count data model. A standard candidate for a count data model would be the Poisson regression model (PRM) (see e.g. Greene, 1997 or Wooldridge, 2002 ). One characteristic of the PRM is that it assumes equidispersion. However, often the conditional variance exceeds the conditional mean, i.e. there is overdispersion. This overdispersion can be a consequence of unobserved heterogeneity, excess zeros, occurrence dependence between events or a combination of them. Especially in cross section data, it is difficult to identify the source of overdispersion.

## 4.1 Overdispersion due to unobserved heterogeneity

The solution to unobserved heterogeneity lies in more flexible modeling of the variance function.[9] This can be done in two ways: 1) moving away from the complete distributional specification to a specification of the first two moments, or 2) specifying a distribution that permits more flexible modeling of the variance than the Poisson. The first solution relies on the fact that the maximum likelihood (ML) estimator provides consistent estimates of PRM as long as the conditional mean function is correctly specified. The complete distributional assumption can thus be relaxed in favor of more general modeling of the variance function without loosing consistency of the estimates. This leads to what

[7] These problems related to administrative burden are not specific to Finland, but applies to R&D programs in general (Investing in Research and Innovation, 2004).
[8] Tekes publishes abstracts of projects funded within technology programs.
[9] Sections 4.1 and 4.2 rely on Cameron and Trivedi (1998).

Cameron and Trivedi (1998, 1986) call the Poisson pseudo maximum likelihood (PML) estimator. There are various standard error estimators depending on what functional form, if any, is assumed for the variance function.

One way to apply the second solution is to use continuous mixture models. In a continuous mixture model a stochastic error term is introduced into the conditional mean function reflecting the fact that the true mean is not fully observed. One common approach is to use a multiplicative stochastic error term. Various generalized count models can be generated by mixture models. One example is the widely used negative binomial model that can be represented as a Poisson-gamma mixture. The two common versions of the negative binomial model are what Cameron and Trivedi (1998) call the the NB2 and the NB1 models.[10]

## 4.2 Overdispersion due to excess zeros

In some cases, data display overdispersion through excess zeros. This means that the probability of obtaining a zero count is higher than what is consistent with the Poisson or some other specified distribution. The underlying reason for excess zeros is that zeros and the positives are generated by different data generating processes. Hurdle and zero-inflated models are the commonly used modified count models that deal with excess zeros.[11] These models alter both the conditional mean and the conditional variance functions relative to the PRM.

In the hurdle model the underlying idea is that a binomial probability model determines whether a zero or a positive realization is observed. If the hurdle is crossed, then a truncated-at-zero count model determines the conditional distribution of the positives. If $y_i$ is the observed count for observation $i$, then the probability mass function is of the form

$$\Pr(y_i \quad = \quad j) = \begin{cases} f_1(0), & j = 0 \\ \frac{1-f_1(0)}{1-f_2(0)} f_2(j), & j > 1, 2, .. \end{cases} \tag{1}$$

Where $f_1(.)$ and $f_2(.)$ are the probability mass functions related to the binomial

[10] Let $\mu_i$ denote the expected count for observation $i$. NB2 model yields a variance function $\mu_i(1+\alpha\mu_i)$ and NB1 model a variance function $(1+\alpha)\mu_i$. Both versions imply overdispersion as long as $\alpha$ is greater than zero. If $\alpha = 0$ we are back at the PRM. Estimation of PML with variance function$(1+\alpha)\mu_i$ yields the ML estimates of the NB1 model.

[11] Hurdle model dates back to Cragg (1971) and Mullahy (1986), whereas Lambert (1992) and Greene (1994) have introduced the zero-inflated model.

probability model and the count model respectively. Various probability mass functions can be specified. In this study the binomial probability model is a logit model with parameter vector $\gamma$, and the truncated at zero count process is specified to follow either a Poisson or a negative binomial distribution, with parameters $\beta$ related to covariate-vector $\boldsymbol{x}$.

In the zero-inflated count models it is in turn assumed that zeros can occur in two distinct states. There are two populations: one for which the event of interest is unlikely to occur and the other that experiences the event of interest according to a count data process. There are two data-generating processes at work: the first data-generating process determines whether an observation remains in a stage in which the event of interest does not occur or moves to a stage in which events occur at some rate. The second data-generating process generates the observed count that can also be zero.

Let $q_i$ denote the probability that observation $i$ stays at the state in which events do not occur. Correspondingly $(1 - q_i)$ denotes the probability that observation $i$ moves to the state in which the observed count is generated. The zero-inflated count data model specification for the probability of observing a count $j$ for observation $i$ is

$$\Pr(y_i \quad = \quad j) = \begin{cases} q_i + (1-q_i)f(j), & j = 0 \\ (1-q_i)f(j), & j > 1, 2, .. \end{cases} \tag{2}$$

where $f(.)$ is the probability mass function of the chosen probability distribution related to the count data process, usually a Poisson or a negative binomial distribution, with parameters $\beta$ related to covariate-vector $\boldsymbol{x}$.

$q_i$ is allowed to be determined by a binomial probability model with a set of covariates $\boldsymbol{w}$.Let $z$ denote a binary indicator variable that takes a value 1 if observation $i$ stays at state one, and a value 0 if observation $i$ moves to the second state. Then

$$q_i = \Pr(z_i = 1) = F(\mathbf{w}_i, \gamma). \tag{3}$$

$F(. \ , .)$ is the cumulative distribution function related to the chosen binomial probability model. Standard candidates for the distribution are the standard normal distribution (generating a probit model) and the logistic distribution (generating a logit model) with parameter vector $\gamma$ reflecting the impact of changes in $\boldsymbol{w}$ on the probability. In this study *F(.,.)* is the cumulative distribution function of a logistic distribution and *f(.)* is the probability mass function of either a Poisson or a negative binomial distribution.

## 4.3 Characteristics of the data and reflections on the application process

Returning to the current application, the data indicates there are signs of both overdispersion and excess zeros in the data. Table 1 below reveals that the variance is over six times larger than the mean, suggesting that overdispersion is left even once the effect of covariates is taken into account. Intuition suggests that unobserved heterogeneity may be present at least through occurrence dependence. Once a firm has applied for R&D subsidies, it is likely that filling the second application requires less effort. In other words, it is likely that previous applications increase the probability of subsequent applications. This would favor a negative binomial distribution over a Poisson distribution in modeling the occurrence of applications. In addition, Table 1 reveals that the data contain significantly more zeros than would be predicted by a Poisson distribution with a mean of 0.1497. This suggests the presence of excess zeros. Is it reasonable to suppose that excess zeros are the result of the underlying data generating process?

Table 1: *Summary statistics and the frequency distributions of the number of applications per firm.*

| | Mean | Median | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| # applications per firm | 0,1497 | 0 | 0,9693 | 0 | 66 |

| Frequency | 0 | 1 | 2 | 3 | 4+ |
|---|---|---|---|---|---|
| Actual (# firms) | 11242 | 709 | 189 | 67 | 68 |
| Predicted by Poisson with mean 0,1497 | 10569 | 1582 | 118 | 6 | 0 |

There are two circumstances under which a firm does not send any applications during the observation period. First, a firm may send an application at some other time, or may have sent an application in the past, but during the observation period the firm does not submit any applications. This can happen for example because the firm does not launch any suitable new projects during the observation period or it does not see it profitable to send an application for the kind of projects launched during the period in question due to e.g. variation in opportunity costs. In the following, I call this kind of firms interim non-applicants. Second, a firm may never even consider submitting an application. This can happen because the firm is not aware of the program, or because the scope of the firms activities in general is not suitable for the program.[12] In other words, there are firms among the potential applicants that do not consider submitting an application under any circumstances. These firms are called real non-applicants in the following. Potential applicants consist of both applicants and interim non-applicants.

The main underlying difference between the hurdle model and the zero-inflated model is that in the hurdle model only positives are allowed in the count data process part of the model, whereas the zero-inflated model allows some zeros to arise also from the count data process. This difference could be interpreted so that the hurdle model makes a distinction between those firms that apply, and those that do not apply. The zero-inflated model, in turn, separates between firms that are likely to apply and firms that do not consider applying. When considering the application process generating the observed count of applications, both setups could be plausible. If the data at hand consist of a well defined sample of potential applicants, i.e. applicants and interim non-applicants, then intuition supports the hurdle model. A zero observation is generated when a potential applicant decides not to send an application and the hurdle model separates between interim non-applicants and applicants. An alternative is the data may consist of a more general sample of firms: applicants, interim non-applicants and also real non-applicants. In this case, intuition favors the zero-inflated model.

As explained in section 3.1, it is in general not straightforward to define the eligibility for R&D subsidies. This being the case also here, the data at hand consist of a relatively broadly defined sample of firms that is likely to cover both real and interim non-applicants - in addition to applicants. Therefore the intuition would favor the zero-inflated model over the hurdle model.

### 4.4 The modeling approach

Given that the true cause of overdispersion is difficult to identify, the modeling approach chosen here is to start from the standard Poisson model and then test and evaluate more general models. First, models that treat unobserved hetero-

[12] Arundel and Hollanders (2005) report that 55 percent of Finnish firms do not innovate.

geneity as the cause of overdispersion and allow for more flexible modeling of the variance than the PRM are estimated and tested. More specifically, the Poisson PML estimator is used with different variance function specifications and then the negative binomial model is estimated. Second, models that consider overdispersion as a consequence of excess zeros generated by true underlying behavior, namely the hurdle Poisson and zero-inflated Poisson models, are estimated. These models alter the conditional mean function with respect to the PRM. Finally hurdle and zero-inflated versions of the negative binomial model (NB2) are estimated. These models allow for both sources of overdispersion - unobserved heterogeneity and true underlying behavior generating excess zeros. Estimations are based on the method of maximum likelihood and model comparison and testing will be based on information criteria, overdispersion tests, comparison of average predicted probabilities of counts with empirical relative frequencies and chi-square goodness of fit test.

# 5 Data

The firm data I use, covering originally 14 657 Finnish firms, come from Asiakastieto Ltd. Asiakastieto is a for-profit company collecting, standardizing, and selling firm specific quantitative information. The sample is drawn according to the following criteria: the most recent financial statement of the firm in the register is for either 2001 or 2000, the firm is a corporation, and the industrial classification of the firm belongs to the manufacturing, computer and related activities, research and development, architectural and engineering activities and related technical consultancy, technical testing and analysis. The data are based on firms' official profit sheet and balance sheet statements, plus other information disclosed by the firms to public registries like the industrial classification, geographical location, number of employees, whether a firm is an exporter or an importer, and information related to the ownership of the firm and the board composition. After cleaning the data of firms with missing values, we are left with 12 275 firms.

These 12 275 firms were matched with application data from Tekes that covers business sector applications Tekes received during the period January 1st 2000 to June 30th 2002. In total there were 2170 enterprises that applied for product development or industrial research funding from Tekes during the period. The matching was done using the business codes of firms. There were 31 firms in the Tekes application data for which no business code was available so they could not be matched with the Asiakastieto data. In total 1030 applicants were found in the Asiakastieto data. In addition, Tekes provided information on the number of applications the 12 275 firms had submitted to Tekes before January, 2000. There are 1232 firms that have submitted applications to Tekes before January, 2000, but not in the sample period.

There are three principal reasons why some 1000 Tekes-applicants could not be identified in our sample of potential applicants: 1) the firm did not operate in the industries from which the sample was formed; 2) the firm was so small that it was not obliged by law to send its balance and profit sheets to the official registry[13]; and 3) the firm did not have an official financial statement either in 2001 or in 2000, because it was so recently established or had not yet been officially established. In addition 109 Tekes-applicants drop when the original data with 14 657 firms is cleaned of firms with missing values.

## 5.1 Quality of the data

As explained in the previous section, all manufacturing firms and firms belonging to knowledge intensive business sectors are regarded as constituting the population of interest in this study. It is important to assess how well the data at hand describes the overall population of chosen industries in Finland and the population of Tekes applicants. This was done by comparing the distributions of firms in the Asiakastieto data to the overall distributions of Finnish firms in manufacturing and knowledge intensive business sectors, and by comparing all the Tekes applicants to those identified in the Asiakastieto data. The overall distribution of Finnish firms in the relevant industries is provided by Statistics Finland.

The comparisons according to the industry classification reveal that the available data constitute a relatively good representation of the actual populations, both in general and in terms of Tekes applicants. In other words, the distribution of firms belonging to the Asiakastieto data is well in line with the overall distribution of Finnish firms across industries, and likewise the distribution of Tekes applicants identified in the Asiakastieto data fits well to the industry distribution of all Tekes applicants. However, in terms of size, the quality could be better. Very small firms are clearly underrepresented among firms in the Asiakastieto data as well as among Tekes applicants identified in

[13] Asiakastieto claims to cover well also these smaller firms, but apparently not all of them.

the Asiakastieto data. Looking at the granted funding further highlights this issue. Applicants identified in the Asiakastieto data cover 70 percent of granted Tekes funding compared to their 53 percent share of applicants. Taking into account the increased emphasis of Tekes on SMEs, the under-representation of micro firms is certainly something that has to be kept in mind when interpreting results.

## 5.2 Determinants of application

Table 2 displays summary statistics of explanatory variables for the whole sample, and Table 3 conditions the statistics on the application decision.[14] As Table 2 shows, firms in the sample are heterogeneous. They are on average 12 years old with 36 employees. A very high proportion, 97 percent, are SMEs according to the official EU standard (see footnote 4). As explained in section 2, the funding criteria of Tekes favor SMEs. Sales per employee (SALES_EMPL), a measure of efficiency or a crude indication of value added, is 115 thousand euros, and some 22 per cent have exports (EXPORTS). We also have information on two corporate governance variables. In some 14 percent of potential applicants, the CEO is also the chairman of the board (CEO_CHAIR). Such an arrangement can, on the one hand, improve the information flow between the board and the executive but, on the other hand, weakens the board's independence. The board of an average potential applicant has four to five members (BOARD). A larger board is more likely to include members with outside knowledge that may be useful either in conducting R&D (choosing among competing projects, organizing management of current projects, monitoring), or in the application process itself. APPLICATIONS indicates the number of application a firm has submitted to Tekes before the year 2000. PARENT is an indicator variable getting a value one if the firm is a parent company. R&D_INV is the capitalized R&D investment in the balance sheet divided by firm's book value of total assets at the end of the year and R&D is an indicator variable taking value 1 if the firm has reported R&D investments in the balance sheet. I am well aware of the problems related to the balance sheet measure of R&D investment. For many firms,

[14]Sales figures and the R&D investment figures used are for the year preceding the first application or the nearest available of the preceding years back to 1999. There were 12 firms that applied in 2000 with 1999 figures missing. However 10 of them applied also in 2001 and the remaining two in 2002 so the figures for 2000 were used for these firms. 10 691 firms have figures for 1999, 1528 have figures for 2000 and 56 firms figures for 2001. R&D investment is measured as the share of total assets in the balance sheet. Other variables represent the state at the time of retrieval of the data, mainly the year 2000.

especially SMEs, it is difficult to separate R&D expenses from other activities of the firm. In addition, even larger firms with more established R&D departments do not necessarily want to announce figures related to their R&D expenditures. Unfortunately it is the only available measure of R&D investment.

Table 2: Descriptive statistics for the whole sample.

| | Mean | Median | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| AGE | 12 | 9 | 12.33 | 0 | 104 |
| # EMPLOYEES | 36 | 5 | 248.61 | 1 | 13541 |
| SALES_EMPL | 115 | 79 | 311 | -1.000 | 26100 |
| BOARD | 4.38 | 4 | 2.04 | 1 | 10 |
| APPLICATIONS | 0.6 | 0 | 4.67 | 0 | 287 |
| R&D_INV | 0.004 | 0 | 0.038 | 0 | 0.76 |
| R&D | 0.04 | 0 | 0.18 | 0 | 1 |
| PARENT | 0.11 | 0 | 0.32 | 0 | 1 |
| EXPORTER | 0.22 | 0 | 0.42 | 0 | 1 |
| SME | 0.97 | 1 | 0.16 | 0 | 1 |
| CEO_CHAIR | 0.14 | 0 | 0.35 | 0 | 1 |

NOTES: There are 12275 observations. Data sources: Asiakastieto Ltd. otherwise; for data on applications, Tekes.

From Table 3 we see that applicants are larger than non-applicants. The median number of employees for non-applicants is 5, for applicants 27. Also the sales per employee is somewhat larger for the applicants. Despite the problems related to the used R&D measures, applicants stand out as more engaged in R&D activities. However, only 13 percent of applicants have reported R&D investment in the balance sheet, which clearly indicates the flaws related to the balance sheet figure of R&D investment. The applicants also tend to have larger boards. Quite naturally, applicants have more previous applications on average than non-applicants. The difference in both means and medians is 4. Export orientation stands clearly out among the applicants. 57 percent of applicants have exports compared to 19 percent of non-applicants. In addition the share of parent companies is substantially higher among the applicants.

Table 3: Descriptive statistics for applicants and non-applicants

| | Non-applicants (11244) | | | Applicants (1033) | | |
|---|---|---|---|---|---|---|
| | Mean | Std. Dev | Median | Mean | Std. Dev | Median |
| AGE | 12 | 12.07 | 9 | 13 | 14.85 | 9 |
| # EMPLOYEES | 22 | 118.28 | 5 | 193 | 745.43 | 27 |
| SALES_EMPL | 113 | 320 | 78 | 138 | 186 | 100 |
| BOARD | 4.20 | 1.90 | 4 | 6.31 | 2.44 | 6 |
| APPLICATIONS | 0.2 | 1.23 | 0 | 4.5 | 15.1 | 1 |
| R&D_INV | 0.003 | 0.03 | 0 | 0.02 | 0.09 | 0 |
| R&D | 0.03 | 0.16 | 0 | 0.13 | 0.34 | 0 |
| PARENT | 0.09 | 0.29 | 0 | 0.36 | 0.48 | 0 |
| EXPORTER | 0.19 | 0.39 | 0 | 0.57 | 0.5 | 1 |
| SME | 0.99 | 0.12 | 1 | 0.84 | 0.37 | 1 |
| CEO_CHAIR | 0.14 | 0.35 | 0 | 0.15 | 0.35 | 0 |

NOTES: There are 12275 observations. Data sources: Asiakastieto Ltd. otherwise; for data on applications, Tekes.

# 6 Estimation results

## 6.1 Model evaluation and selection

Comparison and testing of models is based on Akaike and Bayesian information criteria (AIC and BIC respectively), a chi-square goodness of fit test (GoF-test), the likelihood ratio test (LR-test) and the Vuong-test. When comparing the standard Poisson regression model (PRM) to the pseudo maximum likelihood (PML) estimates and to the negative binomial model (NB2), estimated standard errors and overdispersion coefficient are also analyzed (Table 4). Table 5 summarizes the results of model comparison and testing.[15] In Table 6 the average predicted probabilities of counts generated by each model are compared with the empirical relative frequencies.

[15] Tables 8 and 9 in Appendix 1 present the results of model comparison and testing in more detail, and Table 10 in Appendix 2 shows full estimation results.

Table 4: Estimation results of Poisson PML and negative binomial model.

| Variable | Poisson ML and PML | | | | | | | NB2 | |
|---|---|---|---|---|---|---|---|---|---|
| | Coefficient | ML standard errors | | PML standard errors | | | \| t \|RS | Coefficient | \| t \|* |
| | | MLH | MLOP | NB1 | RS | Boot | | | |
| *ONE* | **-4.8674** | 0.1771 | 0.1454 | 0.2095 | 0.2406 | 0.2584 | 20.69 | **-5.0091** | 20.15 |
| *AGE* | **-0.0442** | 0.0043 | 0.0032 | 0.0051 | 0.0065 | 0.0083 | 6.81 | **-0.0421** | 6.45 |
| *AGE^2* | **0.0004** | 0.0000 | 0.0000 | 0.0001 | 0.0001 | 0.0001 | 6.29 | **0.0004** | 4.73 |
| *LNEMPL* | **0.6111** | 0.0469 | 0.0372 | 0.0560 | 0.0681 | 0.0838 | 8.97 | **0.7389** | 10.04 |
| *LNEMPL^2* | **-0.0453** | 0.0056 | 0.0039 | 0.0069 | 0.0095 | 0.0121 | 4.76 | **-0.0734** | 6.13 |
| *SALES_EMPL* | **0.0005** | 0.0001 | 0.0001 | 0.0002 | 0.0003 | 0.0003 | 2.02 | -0.0003 | 1.52 |
| *SALES_EMPL^2* | **0.0000** | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 1.98 | 0.0000 | 0.75 |
| *APPLICATIONS* | **0.0504** | 0.0023 | 0.0017 | 0.0028 | 0.0033 | 0.0123 | 15.52 | **0.1285** | 11.80 |
| *APPLICATIONS^2* | **-0.0001** | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0001 | 11.54 | **-0.0004** | 10.92 |
| *R&D_INV* | **1.8859** | 1.2937 | 1.1832 | 1.4724 | 1.5567 | 1.7700 | 1.21 | 1.1064 | 0.56 |
| *R&D_INV^2* | -1.1389 | 2.0895 | 2.0083 | 2.3520 | 2.2790 | 3.2153 | 0.50 | -0.4366 | 0.13 |
| *R&D* | **0.4428** | 0.1064 | 0.0894 | 0.1209 | 0.1336 | 0.1588 | 3.31 | **0.6916** | 4.26 |
| *SME* | -0.1487 | 0.9890 | 0.7220 | 0.1201 | 0.1459 | 0.1687 | 1.02 | -0.2389 | 1.28 |
| *PARENT* | **0.3871** | 0.0599 | 0.0443 | 0.0713 | 0.0850 | 0.0866 | 1.55 | **0.3881** | 4.67 |
| *EXPORTER* | **0.8800** | 0.0694 | 0.0567 | 0.0806 | 0.0877 | 0.0872 | 10.04 | **0.9506** | 10.41 |
| *BOARD* | **0.1819** | 0.0115 | 0.0085 | 0.0138 | 0.0176 | 0.0191 | 10.34 | **0.1784** | 11.56 |
| *CEO_CHAIR* | -0.0360 | 0.0763 | 0.0631 | 0.0866 | 0.0947 | 0.0939 | 0.38 | -0.0131 | 0.14 |
| $\alpha$ | | | | | | | | **1.3355** | 11.23 |
| regional dummies | YES | | | | | | | YES | |
| industry dummies | YES | | | | | | | YES | |
| Pseudo-$R^2$ | 0.42 | | | | | | | 0.26 | |

*Note1: The variance specifications underlying different standard error estimates are the following, MLH $\omega=\mu$, Hessian estimate; MLOP $\omega=\mu$, BHHH/outer product estimate; NB1 $\omega=\phi\mu=(1+\alpha)\mu$ where $\alpha=0.5640$; RS, unspecified $\omega$, robust sandwich estimate: Boot, unspecified $\omega$, bootstrap estimate with 200 iterations.*

*Note2: Industry and regional dummies were also included in the estimations and several of them were statistically significant. Appendix 2 presents full estimation results.*

*Note3: Figures in bold indicate that the coefficient estimate is statistically significant at least at the 10 % significance level.*

** Based on MLH standard error estimates. RS standard error estimates were also calculated, but they were in line with the MLH estimates.*

Comparison of the ML estimates of PRM to the PML estimates and to the NB2 model indicates that unobserved heterogeneity may be present. As Table 4 shows, ML based standard error estimates[16] are substantially lower than the Poisson PML estimates.[17] This is an indication of overdispersion and suggests

[16]Hessian (MLH) and outer product (MLOP) estimates.

[17]The NB1 estimates based on the NB1 variance function $(1+\alpha)\mu_i$, and robust sandwich (RS) and bootstrap (Boot) estimates based on an unspecified variance function.

that the ML standard errors should not be used. A comparison of the PRM and the NB2 model also conveys that PRM is not adequate for the data, because the overdispersion parameter $\alpha$ gets a highly significant value in the NB2 model as Table 4 reveals. Table 5 also shows that both the AIC and BIC favor the NB2 model over PRM. Moreover, the LR-test rejects the PRM over the NB2 model. However, the GoF-test based on actual and predicted frequencies rejects both the PRM and the NB2 model.

Table 5: Summary of model selection results

| | PRM | NB2 | ZIP | Hurdle-Poisson | ZINB | Hurdle-NegBin |
|---|---|---|---|---|---|---|
| AIC | VI | V | II | IV | I | III |
| BIC | VI | IV | II | V | I | III |
| GoF-test | rejected | rejected | rejected | rejected | not rejected | not rejected |
| LR-test | rejected vs. NB2<br>rejected vs. H-P | rejected vs. H-NB | rejected vs. ZINB | rejected vs. H-NB | | |
| Vuong-test | rejected vs. ZIP | rejected vs. ZINB | | | | |

Estimation results of zero-inflated (ZIP) and hurdle-Poisson models suggest that overdispersion through excess zeros is also something that should be taken into account.[18] Table 5 shows that based on AIC and BIC the ZIP model is preferred over the PRM and the NB2 model. The hurdle-Poisson model is also preferred over PRM. However, the information criteria favor NB2 over the hurdle-Poisson. In addition the Vuong test rejects PRM in favor of ZIP, and the LR-test rejects PRM against the hurdle-Poisson model. However, based on the results it is difficult to conclude whether the zero-inflated or the hurdle specification should be used. AIC and BIC favor the zero-inflated specification, but in terms of actual and predicted frequencies shown in Table 6 it is difficult to discriminate between the two. The chi-square goodness-of-fit test rejects both specifications.

The above analyses provide evidence in favor of both unobserved heterogeneity and excess-zeros. Estimation results related to negative binomial versions of the zero-inflated (ZINB) and the hurdle (Hurdle-NegBin) specifications provide further evidence in this respect. Both the information criteria and goodness-of-fit test favor the negative binomial specifications over the Poisson specifications. In addition the LR-tests reject ZIP in favor of ZINB and hurdle-Poisson in favor of hurdle-NegBin. This suggests that unobserved heterogeneity is present. A

[18] A Logit model is used in the binomial part in both models and the count processes follow a Poisson distribution.

Table 6: Actual and predicted cell frequencies for different models.

| Frequency | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7+ |
|---|---|---|---|---|---|---|---|---|
| Actual | 11242 | 709 | 189 | 67 | 24 | 13 | 5 | 26 |
| Poisson | 11080 | 938 | 160 | 47 | 19 | 9 | 5 | 16 |
| NegBin | 11279 | 710 | 141 | 51 | 25 | 14 | 9 | 46 |
| ZIP | 11217 | 686 | 238 | 77 | 27 | 11 | 5 | 13 |
| Hurdle-Poisson | 11242 | 661 | 229 | 79 | 30 | 13 | 6 | 13 |
| ZINB | 11268 | 654 | 211 | 73 | 29 | 13 | 7 | 20 |
| Hurdle-Negbin | 11242 | 709 | 192 | 65 | 27 | 13 | 7 | 20 |

Vuong test of NB2 versus ZINB and a LR-test of NB2 versus Hurdle-NegBin in turn provide evidence that overdispersion through excess zeros is also present. Both ZINB and Hurdle-NegBin are chosen over NB2.

The results suggest that taking into account both unobserved heterogeneity and excess zeros could be an improvement. However, the choice between ZINB and Hurdle-NegBin is less clearcut. AIC and BIC favor the zero-inflated specification. Chi-square goodness-of-fit tests do not reject either of the models, but Hurdle-NegBin seems to provide a better fit to the data when comparing the actual and predicted cell frequencies presented in Table 6. Based on these model comparisons it is difficult to say whether the zero-inflated or hurdle specification should be used. Given that the intuition presented in section 4.3 supports the zero-inflated specification, ZINB is selected as the final model.

## 6.2 Estimation results

Before analyzing the estimation results of ZINB in detail, it is interesting to have a look at the estimated coefficients of the NB2 versus the ZINB specifications presented in Appendix 2. The results show how the explanatory power of regressors in the ZINB specification is divided between the two processes compared to the NB2 model. For example the regressor EXPORTER has a highly significant coefficient in the NB2 model, but the two part model indicate that EXPORTER determines whether an observation belongs to potential applicants or not rather than the number of events. The same happens with the indicator variable R&D. In addition most of the statistically significant regressors get a smaller coefficient estimate once overdispersion through excess zeros is taken into account. This suggests that in order to get a more solid interpretation of the results, it is important to take into account excess zeros.

Table 7 presents the full estimation results of the ZINB model including industry and regional dummies and also marginal effects for the regressors. Marginal effects are presented for both parts of the model; i.e. with respect the unconditional expected number of applications, and the probability of belonging to the group of potential applicants. Moreover the marginal effects with respect to the unconditional mean of applications are divided between the effect due to a change in the probability of belonging to the group of potential applicants and due to change in the count process. Calculated marginal effects represent average response over all observations.[19]

### 6.2.1 Binary process

Coefficient estimates for the binary process reported in Table 7 indicate whether a regressor has a positive or negative influence on the probability of being a potential applicant. Given that eligibility and awareness determine whether a firm is a potential applicant, interpretation of the estimation results reflects upon the effect the regressors might have on these two components. Size of the firm is positively related to the probability of being a potential applicant, i.e. the larger the firm the likelier it is that the firm is a potential applicant. This effect may be due to eligibility and awareness. Larger firms are likelier to conduct innovative activities that are eligible for R&D subsidies, but also larger firms are likelier to be better informed about the subsidy program.

The positive coefficient of sales per employee reflects the activities of innovative firms that on average generate higher sales per employee compared to non-innovative firms. As a result firms with higher sales per employee are likelier to launch projects that are eligible for R&D subsidies. The number of members in the board of the firm increases the probability of being a potential applicant. This indicates that a larger board is likelier to provide the firm with knowledge that increases the likelihood of the firm to be aware of the R&D subsidy program.

[19]For continuous regressors AGE, SALES_EMPL, APPLIC, and BOARD, marginal effect gives the change in the expected value of the dependent variable due to a one-unit change in a regressor. In the case of the logarithmic regressor LNEMPL, marginal effect gives the change due to one-percent change in the number of employees. The marginal effects of R&D_INV are calculated in terms of R&D_INV*100, giving the change in the dependent variable due to one-percent change in the share. Both the value of the variable and it's square are taken into account in the calculations. For dummy variables (R&D, SME, PARENT, EXPORTER, CEO_CHAIR, industry dummies and regional dummies) the calculated effect is the difference between the expected value of the dependent variable when the dummy variable gets a value one and the expected value of the dependent variable when the dummy variable gets a value zero.

Table 7: Estimation results of the zero-inflated negative binomial model.

| Variable | Count process | Binary process | Marginal effects* | | | |
|---|---|---|---|---|---|---|
| | Coefficient | Coefficient | unconditional expected no of applications** | | | probability of being a potential applicant |
| | | | total | count | binary | |
| *ONE* | **-1.7273** | **-4.8967** | | | | |
| *AGE* | **-0.0375** | -0.0159 | -0.0040 | -0.0035 | -0.0004 | -0.0011 |
| *AGE^2* | **0.0004** | -0 | | | | |
| *LNEMPL* | **0.2148** | **0.5421** | 0.0210 | 0.0147 | 0.0064 | 0.0199 |
| *LNEMPL^2* | -0.0139 | -0.0578 | | | | |
| *SALES_EMPL* | **-0.0015** | **0.0024** | -0.0002 | -0.0002 | 0.0001 | 0.0002 |
| *SALES_EMPL^2* | **0.0000** | **0.0000** | | | | |
| *APPLICATIONS* | **0.0549** | **2.9102** | 0.0865 | 0.0072 | 0.0792 | 0.2069 |
| *APPLICATIONS^2* | **-0.0002** | **-0.0100** | | | | |
| *R&D_INV* | -2.0647 | 17.2572 | 0.0015 | -0.3165 | 0.4664 | 0.0122 |
| *R&D_INV^2* | 3.6626 | -24.2221 | | | | |
| *BOARD* | **0.1068** | **0.1530** | 0.0215 | 0.0173 | 0.0042 | 0.0109 |
| *Dummy variables* | | | | | | |
| *R&D* | 0.1004 | **1.6663** | 0.0815 | 0.0198 | 0.0617 | 0.1660 |
| *SME* | -0.1783 | 0.4665 | -0.0158 | -0.0286 | 0.0128 | 0.0301 |
| *PARENT* | **0.2700** | 0.3329 | 0.0553 | 0.0449 | 0.0104 | 0.0252 |
| *EXPORTER* | 0.0457 | **1.6703** | 0.0642 | 0.0080 | 0.0562 | 0.1539 |
| *CEO_CHAIR* | -0.0624 | 0.0714 | -0.0080 | -0.0099 | 0.0019 | 0.0051 |
| *Industry dummies* | | | | | | |
| *FOOD* | -0.2616 | 0.6485 | -0.0222 | -0.0396 | 0.0174 | 0.0520 |
| *PAPER* | 0.0135 | -0.6556 | -0.0142 | 0.0021 | -0.0163 | -0.0415 |
| *CHEMI* | **0.5386** | -0.6031 | 0.0866 | 0.1067 | -0.0200 | -0.0380 |
| *RUBBER* | **0.3578** | -0.7429 | 0.0435 | 0.0651 | -0.0216 | -0.0461 |
| *MACHINE* | **0.2858** | 0.4039 | 0.0669 | 0.0537 | 0.0133 | 0.0308 |
| *ELECTRIC* | **0.6824** | 0.4120 | 0.1736 | 0.1562 | 0.0173 | 0.0318 |
| *RADIOTV* | **0.4322** | 0.0455 | 0.0877 | 0.0861 | 0.0016 | 0.0033 |
| *OTHERMAN* | **0.3618** | **-0.8757** | 0.0402 | 0.0654 | -0.0252 | -0.0558 |
| *TELECOM* | 0.5436 | -0.4484 | 0.0969 | 0.1123 | -0.0153 | -0.0291 |
| *DATAPRO* | **0.6080** | **1.9025** | 0.2287 | 0.1444 | 0.0843 | 0.1821 |
| *R_D* | **0.7310** | -0.1104 | 0.1528 | 0.1570 | -0.0042 | -0.0077 |
| *Regional dummies* | | | | | | |
| *REGION2* | -0.0055 | -0.2002 | -0.0063 | -0.0009 | -0.0054 | -0.0141 |
| *REGION3* | 0.0622 | **0.8630** | 0.0382 | 0.0111 | 0.0272 | 0.0715 |
| *REGION4* | 0.1537 | 0.4726 | 0.0424 | 0.0276 | 0.0147 | 0.0366 |
| *REGION56* | **0.4491** | **-1.5054** | 0.0413 | 0.0825 | -0.0411 | -0.0799 |
| $\alpha$ | **0.3760** | | | | | |

* For dummy variables the difference between value with dummy = 1 and value with dummy = 0 are reported instead of marginal effects, i.e. E(y|.,d=1)-E(y|.,d=0) averaged over all observations.

** The total marginal effects with respect to unconditional mean are divided between the effect due to change in pr(potential applicant) (binary) and the effect due to change in E(y|x, potential applicant) (count).

***Marginal effects of R&D% are calculated in terms of R&D%*100

*Note: Figures in bold indicate that the coefficient estimate is statistically significant at least at the 10 % significance level.*

The three most important factors affecting the probability of being a potential applicant are the number of previous applications, the R&D dummy variable and whether the firm has exports or not, all of which have a strong positive effect on the probability. Together these three factors capture firms that most evidently belong to the group of potential applicants. When looking at the marginal effects in the last column of Table 7, the magnitude of the effect these factors have on the probability stands out clearly. The pronounced effect of these factors suggests that the R&D subsidy program reaches the most evident potential applicants - i.e. R&D oriented firms operating in international markets. However, at the same time it raises the question of whether R&D subsidies are capable of encouraging established non R&D oriented firms to engage in R&D activities.

Across industries there are small differences in the probability of being a non-applicant. Firms belonging to other manufacturing industries have a lower probability of being a potential applicant, whereas firms belonging to the data-processing industry have a higher probability of belonging to potential applicants compared to the base category of metal and metal products industry. The marginal effect reveals that belonging to the data-processing industry increases the probability of being a potential applicant by 18 percent. This result may be due to the sample period in question that covers the years of the IT-boom.

### 6.2.2 Count process

Coefficient estimates of the count process in Table 7 indicate the effect of the regressors on the expected number of applications for potential applicants. Whether a potential applicant sends an application depends on the costs of applying vis a vis expected benefits. On average, younger firms tend to send more applications. This could indicate that younger firms, with less internal funding and possibly facing financing constraints, are more in need of R&D subsidies. In other words the opportunity cost of applying is lower for smaller firms. Firm size, measured as the number of employees, is positively related to the number of application. This result seems rather obvious. Larger firms are likelier to have several R&D projects underway simultaneously. Sales per employee is negatively related to the expected number of applications. This may reflect that innovative firms quite at the beginning of their life cycle are more in need of R&D subsidies. Those firms are at the stage in which the main focus is in developing something that is expected to generate revenues and higher sales per employee in the future. Together with the age of the firm this suggests that the opportunity cost of applying is lower for firms quite at the beginning of their life cycle.

The number of previous applications is positively related to the number of applications sent in the sample period. This result is intuitive. Numerous previous applications indicate that the firm is actively engaged in R&D. Moreover there may be learning going on. Through numerous applications the firm may have learnt a great deal about how to fill in the application and what kind of activities Tekes favours. Both of these factors support the observation that a firm with numerous previous applications is likelier to continue sending several applications also in the future compared to firms with no or just one application. It may even be the case, that for small firms with little R&D activities, a subsidy received just before the sample period might be negatively related to applications in the sample period. The intuition is that these small firms are busy with the ongoing project and do not launch new R&D projects right after starting the previous one.

The number of members in the board of the firm is positively related to the expected number of applications. This may suggest that a larger board is likelier to provide the firm with knowledge that lowers the application costs. In addition firms that are parent companies have on average higher expected number of application than other firms. This may reflect the tendency of concerns to establish centralized research-oriented R&D laboratories within parent companies. Another explanation might be that the parent company administers applications originating from various companies of the consolidated corporation. In terms of application costs this could be interpreted so that, given the experience of parent companies in filling applications, the application cost for them is lower.

Industry dummies indicate that belonging to almost any industry other than the metal industry increases the expected number of applications - only exceptions are the food -, paper - and telecommunications industry. At the regional level the only difference is the region 5, which stands for Northern Finland and Lapland. It is quite surprising to notice that firms in Northern Finland or Lapland are on average more active applicants than firms in Southern Finland. This result is driven by the fact that there are a couple of active applicants among the few applicants from Northern Finland.

Marginal effects with respect to the unconditional expected number of applications are divided between the effect due to change in the probability of being a potential applicant and the effect due to change in the expected number of applications conditional on being a potential applicant. Marginal effects indicate the magnitude of the effect the regressors have. The first observation is that in general, effects seem relatively small. However, one should bear in mind that 92 percent of the firms have zero applications in the sample period. As a result the average expected number of applications is only 0.1497. This means that although the marginal effects may seem small in absolute terms, the effect of regressors cannot be interpreted to be negligible.

Marginal effects reveal that two thirds of the effect that the size of the firm has on the expected number of applications is generated by the effect size has on the expected number of applications conditional on being a potential applicant. In other words, in relation to the unconditional expected number of applications, size is more important in determining the number of applications than the probability of being a potential applicant. The opposite is true for the number of previous applications. 80 percent of the effect previous applications has on the expected number of applications is generated by the change in the probability of being a potential applicant. The marginal effect with respect to the number of members in the board indicate that the change in the expected number of applications is mainly generated by the effect BOARD has on the conditional count process. 80 percent of the total marginal effect is generated through the change in the conditional expected number of applications. Marginal effects with respect to the industry dummies further strengthen the observation that in general, industry level differences in the expected number of applications are mainly generated by different application activity of the potential applicants in different industries.

# 7 Conclusions

This paper examined the application for R&D subsidies using count data models. Given the importance of R&D subsidy programs as one of the main innovation policy tools, we know surprisingly little about the process that allocates them - i.e. the participation process. Heckman and Smith (2004) define the participation in a program as consisting of five different stages: awareness, eligibility, application, acceptance and enrollment. There are two important decision problems underlying this participation process: application decision of firms and acceptance decision of government bureaucrats. By focusing on the application stage, this paper provides the first explorative step toward understanding firms' application behavior, which, to the best of my knowledge, has not been investigated before.

Applications for R&D subsidies are investigated by analyzing the number of applications a firm sent to the Finnish Funding Agency for Technology and Innovation (Tekes) during the period of January, 2000 - June, 2002. Tekes is the principal public promoter of private R&D in Finland and also the most important public financier of business R&D. I analyzed various count data models and found that it is important to take into account both unobserved heterogeneity and excess zeros when modeling the number of applications.

The zero-inflated negative binomial model provided an intuitive framework for the analysis and was chosen as the final model. The zero-inflated specification models the probability of being a potential applicant and the expected number of applications conditional on being a potential applicant. Estimation results indicated that omitting especially the problem of excess zeros may distort the results. The explanatory power of regressors in the zero-inflated specification is divided between the two processes compared to the corresponding non zero-inflated count model.

Estimation results yield several findings:

- The number of previous applications, reported R&D-investments prior to the sample period and export activities have a pronounced positive effect on the probability of being a potential applicant. This indicates that the R&D subsidy program seems to reach the most evident potential applicants - firms engaged in R&D activities and operating in international markets.

- Age of the firm and sales per employee have a negative effect on the application activity suggesting that the opportunity cost of applying is higher for well established, older firms with higher sales per employee.

- The number of members in the board of the firm is positively related to the number of applications. This suggests that external knowledge is valuable in lowering the application costs.

- Industry level heterogeneity is related to application activity of potential applicants rather than to the probability of being a potential applicant.

In terms of eligibility and awareness, the results of the binary process modeling the probability of being a potential applicant could be interpreted so that those firms that are the most likely to have eligible projects, are also aware of the R&D subsidy program. In other words, the program seems to reach firms that are the most potential participants, and in that sense the program could be considered to work well. However, if the aim of the program is also to encourage established firms to engage in R&D activities, the conclusion is less clear cut. The way firms engaged in R&D activities and operating in international markets stand out suggests that there may be problems related to the awareness of firms that are not "by definition" among the potential applicants.

In relation to the application decision, the results suggest that firms quite at the beginning of the life cycle are more in need of R&D subsidies and therefore have stronger incentives to apply. This result supports the policy argument related to R&D subsidy programs that due to market failures especially young innovative firms need public R&D support. In addition, this result suggests that an important target group of the policy finds the program attractive.

The finding that supports the usefulness of external knowledge in lowering application costs indicates that trying the reduce the applications costs firms face, could be important in increasing application activity. Industry level conclusions are that there are no considerable differences in the awareness of the program across industries, but the application behavior is somewhat heterogeneous. This may be due to both different industry characteristics and policy emphasis that favor specific industries over others.

## References


[1] Arundel, Anthony and Hugo Holanders (2005): Innovation Strengths and Weaknesses, 2005 European Scoreboard on Innovation Paper.

[2] Blanes, Vicente J. and Isabel, Busom (2004): Who participates in R&D subsidy programs? The case of Spanish manufacturing firms, *Research Policy*, 33, 1459-1476.

[3] Cameron and Trivedi (1998): Regression Analysis of Count Data, Cambridge University Press.

[4] Cameron and Trivedi (1986): Econometric Models Based on Count Data: Comparison and Applications of Some Estimators, *Journal of Applied Econometrics,* 1, 29-53.

[5] Cragg, J. G. (1971): Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods, *Econometrica*, 39, 829-844.

[6] David, Hall and Toole (2000): Is public R&D a complement or substitute for private R&D? A review of the econometric evidence, *Research Policy*, 29, 497-529.

[7] Feldman, M. P. and M. R. Kelley (2001): Winning an award from the advanced technology program: pursuing R&D strategies in the public interest and benefiting from a Halo effect, NISTIR 6577, US Department of Commerce.

[8] Georghiou, Luke, Keith Smith, Otto Toivanen and Pekka Ylä-Anttila (2003): Evaluation of the Finnish Innovation Support System, Ministry of Trade and Industry Publications, 5/2003.

[9] Greene, W. (1994): Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models, Working Paper No. EC-94-10, Department of Economics, Stern School of Business, New York University.

[10] Greene, William H. (1998): Econometric Analysis, third edition, Prentice-Hall Inc., US.

[11] Heckman, James (1992): Randomization and Social Program Evaluation, , in Manski, Charles and Irwin Garfinkel (eds.), Evaluating Welfare and Training Programs, Cambridge, Mass: Harvard University Press, 231-252.

[12] Heckman, James J., Hidehiko Ichimura, Jeffrey Smith and Petra Todd (1998): Characterizing Selection Bias Using Experimental Data, *Econometrica*, 66, 1017-1098.

[13] Heckman, James J., Hidehiko Ichimura and Petra Todd (1997): Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme, *Review of Economic Studies*, 64, 605-654.

[14] - (1998): Matching as an Econometric Evaluation Estimator, *Review of Economic Studies*, 65, 261-294.

[15] Heckman James J., Robert, J. LaLond and Jeffrey A. Smith (1999): The Economics and Econometrics of Active Labor Market Programs, in Handbook of Labor Economics, Volume 3A, 1865-2097, Amsterdam: North-Holland.

[16] Heckman, James J. and Jeffrey A. Smith (1995): Assessing the case for Social Experiments, *The Journal of Economic Perspectives*, 9, 85-110.

[17] Heckman, James J. and Jeffrey A. Smith (2004): The Determinants of Participation in a Social Program: Evidence from a Prototypical Job Training Program, *Journal of Labor Economics*, 22, 243-298.

[18] Investing in Research and Innovation: Realising the Potential of Public - Private Interaction, Report summarising the results of the conference, held in Noordwijk, the Netherlands, October 12 and 13 2004, as part of the Dutch EU Presidency.

[19] Jaffe , Adam B. (2002): Building Programme Evaluation into the Design of Public Research-Support Programmes, *Oxford Review of Economic Policy*, 18, 1, 22-34.

[20] Klette, Moen and Griliches (2000): Do subsidies to commercial R&D reduce market failures? Microeconometric evaluation studies, *Research Policy*, 29, 471-495.

[21] Lambert , D. (1992): Zero-Inflated Poisson Regression with an Application to Defects in Manufacturing, *Technometrics*, 34, 1-14.

[22] Lichtenberg, Frank R. (1998): The Allocation of Publicly-Funded Biomedical Research, NBER Working Paper 6601.

[23] Moffit, Robert (1992): Evaluation Methods for Program Entry Effects, in Manski, Charles and Irwin Garfinkel (eds.), Evaluating Welfare and Training Programs, Cambridge, Mass: Harvard University Press, 231-252.

[24] Mullahy, J. (1986): Specification and Testing of Some Modified Count Data Models, *Journal of Econometrics*, 33, 341-365.

[25] Nezu, Risaburo (1998): Trends and Patterns of Public Support to Industry in the OECD Area, STI Review, 21, 12-24.

# Appendix 1

Table 8: Model selection criteria for the estimated count data models.

| | PRM | NB2 | ZIP | Hurdle-Poisson | ZINB | Hurdle-NegBin |
|---|---|---|---|---|---|---|
| -lnL | 3742 | 3510 | 3209 | 3379 | 3161 | 3341 |
| AIC | 7516 | 7052 | 6450 | 6822 | 6354 | 6747 |
| BIC | 7785 | 7322 | 6719 | 7361 | 6623 | 7294 |
| $T_{GoF}$ (df=7)** | 81 | 32 | 13 | 27 | 12 | 3 |
| α* | | 0.336 (11.23) | | | 0.376 (6.39) | 0.719 (3.24) |

* t-value in parenthesis

** Goodness-of-fit test statistic. $\chi^2(7)$ critical value at the 5 percent level is 12.02.

Table 9: Tests of different models.

| LR-test | | | Vuong-test | | |
|---|---|---|---|---|---|
| PRM vs. NB2 | $T_{LR}$ = 464 df=1 | PRM rejected | PRM vs. ZIP | $T_V$ = 12.6 | PRM rejected |
| PRM vs. H-P | $T_{LR}$ = 724 df=32 | PRM rejected | NB2 vs. ZINB | $T_V$ = 12.3 | NB2 rejected |
| NB2 vs. H-NB | $T_{LR}$ = 338 df=32 | NB2 rejected | | | |
| ZIP vs. ZINB | $T_{LR}$ = 96 df=1 | ZIP rejected | | | |
| H-P vs. H-NB | $T_{LR}$ = 76 df=1 | H-P rejected | | | |

# Appendix 2

Table 10: Estimation results

| | Poisson PMLE | | NB2 | | ZIP | | | | Hurdle-Poisson | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | count process | | binary process* | | positives | | zeros** | |
| Variable | Coefficient | $\lvert t \rvert^{RS}$ | Coefficient | $\lvert t \rvert$ | Coefficient | $\lvert t \rvert$ | Coefficient | $\lvert t \rvert$ | Coefficient | $\lvert t \rvert$ | Coefficient | $\lvert t \rvert$ |
| *ONE* | **-4.8674** | 20.69 | **-5.0091** | 20.15 | **-1.5394** | 6.17 | **-4.4389** | 6.70 | -1.8013 | 5.64 | **-5.1050** | 16.85 |
| *AGE* | **-0.0442** | 6.81 | **-0.0421** | 6.45 | **-0.0378** | 7.62 | -0.0174 | 1.17 | **-0.0367** | 5.74 | **-0.0475** | 5.91 |
| *AGE^2* | **0.0004** | 6.29 | **0.0004** | 4.73 | **0.0004** | 6.99 | 0.0001 | 0.33 | **0.0003** | 5.07 | **0.0004** | 4.19 |
| *LNEMPL* | **0.6111** | 8.97 | **0.7389** | 10.04 | **0.1156** | 1.90 | **0.6282** | 3.76 | **0.2036** | 2.6 | **0.7037** | 7.41 |
| *LNEMPL^2* | **-0.0453** | 4.76 | **-0.0734** | 6.13 | 0.0028 | 0.41 | **-0.0830** | 2.69 | -0.0011 | 0.13 | **-0.0787** | 4.81 |
| *SALES EMPL**** | **0.0519** | 2.02 | -0.0365 | 1.52 | **-0.1690** | 5.33 | **0.2500** | 2.67 | **-0.1880** | 4.66 | -0.0211 | 0.67 |
| *SALES EMPL^2* | **0.0002** | 1.98 | 0.0001 | 0.75 | **0.0057** | 3.94 | **0.0078** | 2.66 | **0.0067** | 3.2 | 0.0000 | 0.01 |
| *APPLICATIONS* | **0.0504** | 15.52 | **0.1285** | 11.80 | **0.0363** | 14.33 | **2.1933** | 10.12 | **0.0325** | 10.61 | **0.3376** | 14.75 |
| *APPLICATIONS^2* | **-0.0001** | 11.54 | **-0.0004** | 10.92 | **-0.0001** | 9.09 | **-0.0075** | 5.24 | **-0.0001** | 6.55 | **-0.0012** | 13.09 |
| *R&D INV* | **1.8859** | 1.21 | 1.1064 | 0.56 | -1.7848 | 1.09 | 4.7436 | 0.45 | -1.5002 | 0.72 | 0.5355 | 0.23 |
| *R&D INV^2* | -1.1389 | 0.50 | -0.4366 | 0.13 | 3.3175 | 1.34 | -6.7845 | 0.44 | 2.0005 | 0.62 | 2.2089 | 0.56 |
| *R&D* | **0.4428** | 3.31 | **0.6916** | 4.26 | 0.0285 | 0.21 | **1.6714** | 2.64 | 0.0940 | 0.53 | **0.7669** | 4.01 |
| *SME* | -0.1487 | 1.02 | -0.2389 | 1.28 | -0.0432 | 0.38 | -0.0226 | 0.04 | -0.0315 | 0.25 | -0.2147 | 0.86 |
| *PARENT* | **0.3871** | 1.55 | **0.3881** | 4.67 | **0.2595** | 3.75 | 0.2545 | 1.21 | **0.1999** | 2.28 | **0.4829** | 4.80 |
| *EXPORTER* | **0.8800** | 10.04 | **0.9506** | 10.41 | 0.0382 | 0.41 | **1.4999** | 6.84 | 0.1041 | 0.85 | **1.0843** | 10.28 |
| *BOARD* | **0.1819** | 10.34 | **0.1784** | 11.56 | **0.1049** | 7.74 | **0.1436** | 4.03 | **0.1038** | 5.91 | **0.1840** | 9.63 |
| *CEO CHAIR* | -0.0360 | 0.38 | -0.0131 | 0.14 | -0.0828 | 0.87 | 0.1168 | 0.53 | -0.2095 | 1.57 | 0.0760 | 0.70 |
| *Industry dummies* | | | | | | | | | | | | |
| *FOOD* | -0.0058 | 0.03 | -0.0113 | 0.05 | -0.1907 | 0.82 | 0.4034 | 0.76 | -0.3328 | 0.88 | -0.0122 | 0.05 |
| *PAPER* | 0.0306 | 0.17 | -0.1585 | 0.85 | 0.1542 | 0.86 | **-0.8587** | 1.96 | 0.1121 | 0.47 | -0.2034 | 0.99 |
| *CHEMI* | **0.5432** | 2.34 | **0.5559** | 2.37 | **0.5690** | 3.78 | -0.4931 | 0.86 | **0.7864** | 4.25 | 0.1918 | 0.64 |
| *RUBBER* | **0.4927** | 2.69 | **0.3354** | 1.92 | **0.4644** | 2.78 | **-0.9245** | 2.13 | **0.6130** | 2.71 | 0.1109 | 0.55 |
| *MACHINE* | **0.6350** | 3.81 | **0.4701** | 3.13 | **0.4045** | 3.03 | 0.1474 | 0.44 | **0.5050** | 2.75 | **0.4094** | 2.45 |
| *ELECTRIC* | **1.0969** | 6.46 | **0.9664** | 5.88 | **0.7388** | 5.37 | 0.2346 | 0.60 | **0.8212** | 4.4 | **0.8703** | 4.52 |
| *RADIOTV* | **0.8125** | 3.85 | **0.7332** | 3.29 | **0.4195** | 2.47 | 0.1706 | 0.28 | 0.4602 | 2.02 | **0.5169** | 1.88 |
| *OTHERMAN* | 0.0217 | 0.13 | -0.0498 | 0.33 | **0.4140** | 2.79 | **-0.8412** | 2.63 | **0.4594** | 2.28 | -0.1921 | 1.13 |
| *TELECOM* | 0.3488 | 1.11 | 0.1747 | 0.55 | **0.5963** | 2.21 | -0.5336 | 0.86 | **0.6643** | 2.37 | -0.2210 | 0.51 |
| *DATAPRO* | **1.5997** | 10.39 | **1.7015** | 11.60 | **0.6170** | 4.21 | **1.6238** | 5.09 | **0.8338** | 4.16 | **1.6236** | 9.86 |
| *R D* | **1.1863** | 7.34 | **0.9805** | 6.64 | **0.8220** | 5.86 | -0.2598 | 0.82 | **1.1165** | 5.86 | **-0.6538** | 3.84 |
| *Regional dummies* | | | | | | | | | | | | |
| *REGION2* | -0.0633 | 0.80 | -0.0103 | 0.13 | -0.0395 | 0.56 | -0.1389 | 0.78 | -0.1144 | 1.23 | 0.0293 | 0.32 |
| *REGION3* | **0.3797** | 3.38 | **0.5206** | 4.41 | 0.0079 | 0.11 | **0.8449** | 3.01 | -0.2130 | 1.24 | **0.6556** | 4.87 |
| *REGION4* | **0.2553** | 1.87 | **0.3250** | 2.53 | 0.1501 | 1.15 | 0.4108 | 1.38 | 0.0570 | 0.39 | **0.3996** | 2.58 |
| *REGION56* | 0.2452 | 0.78 | 0.0800 | 0.37 | **0.4785** | 2.29 | **-1.1701** | 2.12 | 0.2614 | 1.09 | 0.0903 | 0.34 |
| $\alpha$ | | | **0.3355** | 2.82 | | | | | | | | |
| Pseudo-$R^2$ | 0.42 | | 0.26 | | 0.39 | | | | 0.43 | | 0.32 | |
| -lnL | 3742 | | 3510 | | 3209 | | | | 3379 | | | |
| AIC | 7516 | | 7052 | | 6450 | | | | 6822 | | | |
| BIC | 7785 | | 7322 | | 6719 | | | | 7361 | | | |
| $T_{GoF}$(df=7) | 81 | | 32 | | 12.55 | | | | 27 | | | |

** Pr(potential applicant)*

*** Pr(at least one application)*

**** In 100 000 euros*

*Note: Figures in bold indicate that the coefficient estimate is statistically significant at least at the 10 % significance level.*

Table 11: Estimation results (continued)

| | ZINB | | | | Hurdle-NegBin | | | |
|---|---|---|---|---|---|---|---|---|
| | count process | | binary process* | | positives | | zeros** | |
| Variable | Coefficient | \| t \| | Coefficient | \| t \| | Coefficient | \| t \| | Coefficient | \| t \| |
| *ONE* | **-1.7273** | 6.13 | **-4.8967** | 6.20 | **-2.4359** | 5.15 | **-5.1050** | 16.85 |
| *AGE* | **-0.0375** | 5.75 | -0.0159 | 0.91 | **-0.0375** | 3.63 | **-0.0475** | 5.91 |
| *AGE^2* | **0.0004** | 4.91 | 0.0000 | 0.00 | **0.0004** | 3.03 | **0.0004** | 4.19 |
| *LNEMPL* | **0.2148** | 2.82 | **0.5421** | 2.81 | **0.2940** | 2.42 | **0.7037** | 7.41 |
| *LNEMPL^2* | -0.0139 | 1.35 | -0.0578 | 1.57 | -0.0104 | 0.66 | **-0.0787** | 4.81 |
| *SALES EMPL**** | **-0.0015** | 3.58 | **0.0024** | 2.23 | **-0.0021** | 3.18 | -0.0002 | 0.67 |
| *SALES EMPL^2* | **0.0000** | 2.74 | **0.0000** | 2.38 | **0.0000** | 2.24 | 0.0000 | 0.01 |
| *APPLICATIONS* | **0.0549** | 9.80 | **2.9102** | 8.27 | **0.0567** | 6.56 | **0.3376** | 14.75 |
| *APPLICATIONS^2* | **-0.0002** | 7.60 | **-0.0100** | 2.83 | **-0.0002** | 5.15 | **-0.0012** | 13.09 |
| *R&D INV* | -2.0647 | 1.27 | 17.2572 | 1.05 | -0.7750 | 0.27 | 0.5355 | 0.23 |
| *R&D INV^2* | 3.6626 | 1.41 | -24.2221 | 1.05 | 0.2022 | 0.04 | 2.2089 | 0.56 |
| *R&D* | 0.1004 | 0.70 | **1.6663** | 2.18 | 0.0767 | 0.31 | **0.7669** | 4.01 |
| *SME* | -0.1783 | 1.14 | 0.4665 | 0.68 | -0.0565 | 0.24 | -0.2147 | 0.86 |
| *PARENT* | **0.2700** | 3.33 | 0.3329 | 1.32 | 0.1964 | 1.53 | **0.4829** | 4.80 |
| *EXPORTER* | 0.0457 | 0.45 | **1.6703** | 6.64 | -0.0435 | 0.26 | **1.0843** | 10.28 |
| *BOARD* | **0.1068** | 6.85 | **0.1530** | 3.71 | **0.1072** | 4.38 | **0.1840** | 9.63 |
| *CEO CHAIR* | -0.0624 | 0.60 | 0.0714 | 0.29 | -0.2012 | 1.16 | 0.0760 | 0.70 |
| *Industry dummies* | | | | | | | | |
| *FOOD* | -0.2616 | 1.05 | 0.6485 | 1.14 | -0.3369 | 0.70 | -0.0122 | 0.05 |
| *PAPER* | 0.0135 | 0.06 | -0.6556 | 1.28 | -0.0554 | 0.15 | -0.2034 | 0.99 |
| *CHEMI* | **0.5386** | 2.45 | -0.6031 | 0.88 | **0.9655** | 2.71 | 0.1918 | 0.64 |
| *RUBBER* | **0.3578** | 1.88 | -0.7429 | 1.51 | 0.6142 | 1.87 | 0.1109 | 0.55 |
| *MACHINE* | **0.2858** | 1.81 | 0.4039 | 1.06 | 0.3944 | 1.37 | **0.4094** | 2.45 |
| *ELECTRIC* | **0.6824** | 4.15 | 0.4120 | 0.92 | **0.8769** | 3.05 | **0.8703** | 4.52 |
| *RADIOTV* | **0.4322** | 2.10 | 0.0455 | 0.06 | 0.5962 | 1.69 | **0.5169** | 1.88 |
| *OTHERMAN* | **0.3618** | 2.04 | **-0.8757** | 2.39 | 0.4356 | 1.42 | -0.1921 | 1.13 |
| *TELECOM* | 0.5436 | 1.47 | -0.4484 | 0.65 | 0.9663 | 1.84 | -0.2210 | 0.51 |
| *DATAPRO* | **0.6080** | 3.61 | **1.9025** | 5.14 | **0.8866** | 3.00 | **1.6236** | 9.86 |
| *R D* | **0.7310** | 4.40 | -0.1104 | 0.31 | **1.1152** | 3.82 | **-0.6538** | 3.84 |
| *Regional dummies* | | | | | | | | |
| *REGION2* | -0.0055 | 0.07 | -0.2002 | 1.00 | -0.1126 | 0.85 | 0.0293 | 0.32 |
| *REGION3* | 0.0622 | 0.50 | **0.8630** | 2.62 | -0.1665 | 0.77 | **0.6556** | 4.87 |
| *REGION4* | 0.1537 | 1.18 | 0.4726 | 1.44 | 0.1268 | 0.63 | **0.3996** | 2.58 |
| *REGION56* | **0.4491** | 1.81 | **-1.5054** | 2.24 | 0.2256 | 0.62 | 0.0903 | 0.34 |
| $\alpha$ | **0.3760** | 6.39 | | | **0.7192** | 3.24 | | |
| Pseudo-$R^2$ | 0.33 | | | | 0.19 | | 0.32 | |
| -lnL | 3161 | | | | 3341 | | | |
| AIC | 6354 | | | | 6747 | | | |
| BIC | 6623 | | | | 7294 | | | |
| $T_{GoF}$(df=7) | 12.29 | | | | 2.8 | | | |

** Pr(potential applicant)*

*** Pr(at least one application)*

**** In 100 000 euros*

*Note: Figures in bold indicate that the coefficient estimate is statistically significant at least at the 10 % significance level.*