Czado, Claudia; Erhardt, Vinzenz; Min, Aleksey

Working Paper

Zero-inflated generalized Poisson models with regression effects on the mean, dispersion and zero-inflation level applied to patent outsourcing rates

Discussion Paper, No. 482

Provided in Cooperation with:
Collaborative Research Center (SFB) 386: Statistical Analysis of discrete structures - Applications in Biometrics and Econometrics, University of Munich (LMU)

Suggested Citation: Czado, Claudia; Erhardt, Vinzenz; Min, Aleksey (2006) : Zero-inflated generalized Poisson models with regression effects on the mean, dispersion and zero-inflation level applied to patent outsourcing rates, Discussion Paper, No. 482, Ludwig-Maximilians-Universität München, Sonderforschungsbereich 386 - Statistische Analyse diskreter Strukturen, München, https://doi.org/10.5282/ubm/epub.1850

This Version is available at:
https://hdl.handle.net/10419/30994

Zero-inflated generalized Poisson models with regression effects on the mean, dispersion and zero-inflation level applied to patent outsourcing rates

Claudia CZADO, **Vinzenz ERHARDT**[1] and **Aleksey MIN**

Center for Mathematical Sciences
Munich University of Technology
Boltzmannstr. 3
D-85747 Garching, Germany

Abstract

This paper focuses on an extension of zero-inflated generalized Poisson (ZIGP) regression models for count data. We discuss generalized Poisson (GP) models where dispersion is modelled by an additional model parameter. Moreover, zero-inflated models in which overdispersion is assumed to be caused by an excessive number of zeros are discussed. In addition to ZIGP regression introduced by Famoye and Singh (2003), we now allow for regression on the overdispersion and zero-inflation parameters. Consequently, we propose tools for an exploratory data analysis on the dispersion and zero-inflation level. An application dealing with outsourcing of patent filing processes will be used to compare these nonnested models. The model parameters are fitted by maximum likelihood. Asymptotic normality of the ML estimates in this non-exponential setting is proven. Standard errors are estimated using the asymptotic normality of the estimates. Appropriate exploratory data analysis tools are developed. Also, a model comparison using AIC statistics and Vuong tests (see Vuong (1989)) is carried out. For the given data, our extended ZIGP regression model will prove to be superior over GP and ZIP models and even ZIGP models with constant overall dispersion and zero-inflation parameters demonstrating the usefulness of our proposed extensions.

Keywords: maximum likelihood estimator; overdispersion; patent outsourcing; Vuong test; zero-inflated generalized Poisson regression; zero-inflation

[1]Corresponding author.
Tel: +49/8122/86691
E-mail address: erhardt@ma.tum.de

1 Introduction

This paper considers zero-inflated generalized Poisson (ZIGP) regression models. The generalized Poisson distribution has first been introduced by Consul and Jain (1970). ZIGP models have recently been found useful for the analysis of count data with a large amount of zeros (see e.g. Famoye and Singh (2003), Gupta et al. (2004), Joe and Zhu (2005), Bae et al. (2005) and Famoye and Singh (2006)). It is a large class of regression models which contains zero-inflated Poisson (ZIP), generalized Poisson (GP) and Poisson regression (for ZIP, see e.g. Lambert (1992), for GP see e.g. Consul and Famoye (1992) and Famoye (1993)). The interest in this class of regression models is driven by the fact that it can handle overdispersion and/or zero-inflation, which count data very often exhibit. In this paper we now allow for regression not only on the mean but on the overdispersion and zero-inflation parameters. The aim is to improve model fit in those cases in which overall dispersion or zero-inflation parameters are insufficient. At the same time, we are interested in keeping the model complexity in terms of additional parameters low.

Since the ZIGP distribution does not belong to the exponential family, the regression model is no generalized linear model (GLM). Thus, we cannot rely on asymptotic properties known for GLM's. We develop the appropriate asymptotic theory and investigate the small sample properties of the maximum likelihood estimates.

Also, a comparison of nine models extending the regular Poisson GLM by dispersion and zero-inflation parameters will be facilitated. We will illustrate that these models need not have nested design matrices and therefore cannot be compared on the basis of partial deviance or likelihood ratio tests. Therefore, we use the Akaike Information Criterion and Vuong tests (see Vuong (1989)).

The usefulness of our model extensions will be demonstrated in an application dealing with patent outsourcing. We investigate make-or-buy decision drivers for the patent filing process. This dataset has already been examined by Wagner (2005) who used a negative binomial regression approach. Currently, there are only basic studies on the general determinants of outsourcing available. Sako (2005) states that offshoring even of services such as medical diagnosis, patent filing, payroll and benefits administration has become easy with the growth of information technology. Abraham and Taylor (1996) name possible reasons for outsourcing behaviour, such as wage and benefit savings or availability of specialized skills. We will focus on firm specific attributes such as the R&D spending per employee or patent. On the other hand, as suggested by Grossman and Hart (1986), an incitement for vertical integration might be the control over actions of the executing patent attorney. This will be corroborated by our results. Also, Amit and Schoemaker (1993) recommend that a company's decision should depend on the value of the corporate tasks and hence the resources necessary for the provision of these tasks. To analyse this complex dataset, we develop tools for exploratory data analysis for overdispersion and zero-inflation. We will see that heterogeneity is high and strongly depends on a company's industry. We will illustrate step-by-step that all model enhancements are useful for the analysis of this dataset. A graphical interpretation of the model will be performed comparing our results with those obtained by Wagner (2005).

This paper is organized as follows: Section 2 introduces our regression model. The necessary asymptotic theory is discussed in Section 3. Section 4 gives an overview of possible model extensions of the Poisson GLM and summarizes AIC and the Vuong test for model selection. Tools for an exploratory data analysis will be proposed in Section 5.1 and applied to our data

afterwards. Section 6 investigates covariate effects on the mean response and the coefficient of variation as well as interprets the results. We conclude with a summary and discussion.

2 $ZIGP(\mu_i, \varphi_i, \omega_i)$ regression

Famoye and Singh (2003) introduced a zero-inflated generalized Poisson $ZIGP(\mu_i, \varphi, \omega)$ regression model. The generalized Poisson $GP(\mu, \varphi)$ distribution was first introduced by Consul and Jain (1970) and subsequently studied in detail by Consul (1989). We refer to its mean parameterisation. One particular property of the GP distribution is that the variance is greater than, equal to or less than the mean according to whether the second parameter φ is greater than, equal to or less than 1. A ZIGP distribution is defined analogously to a zero-inflated Poisson (ZIP) distribution (see Mullahy (1986)) with an additional zero-inflation parameter ω. Thus, this distribution has three parameters μ, φ and ω and will be denoted by $ZIGP(\mu_i, \varphi, \omega)$. Mean and variance of the ith observation Y_i of the ZIGP distribution are given by

$$E(Y_i|\boldsymbol{X} = \boldsymbol{x_i}) = (1-\omega)\mu_i \tag{2.1}$$

$$\text{and } \sigma_i^2 := Var(Y_i|\boldsymbol{X} = \boldsymbol{x_i}) = E(Y_i|\boldsymbol{X} = \boldsymbol{x_i})\left(\varphi^2 + \mu_i\omega\right). \tag{2.2}$$

One of the main benefits of considering a regression model based on the ZIGP distribution is that it is allowing in two ways for overdispersion by using an additional overdispersion parameter φ and a zero-inflation parameter ω. In particular, it reduces to Poisson regression when $\varphi = 1$ and $\omega = 0$, to GP regression when $\omega = 0$ and to ZIP regression when $\varphi = 1$.

In some data sets a constant overdispersion and/or constant zero-inflation parameter might be too restrictive. Therefore, we now extend the regression model of Famoye and Singh (2003) by allowing for regression on φ and ω. We denote this model as a $ZIGP(\mu_i, \varphi_i, \omega_i)$ regression model with response Y_i and (known) explanatory variables $\boldsymbol{x}_i = (1, x_{i1}, \ldots, x_{ip})^t$ for the mean, $\boldsymbol{w}_i = (1, w_{i1}, \ldots, w_{ir})^t$ for overdispersion and $\boldsymbol{z}_i = (1, z_{i1}, \ldots, z_{iq})^t$ for zero-inflation, $i = 1, \ldots, n$. For individual observation periods, we allow exposure variables E_i, which satisfy $E_i > 0 \; \forall \; i$.

1. *Random components:*
   $\{Y_i,\ 1 \leq i \leq n\}$ are independent with $Y_i \sim ZIGP(\mu_i, \varphi_i, \omega_i)$.

2. *Systematic components:*
   Three linear predictors $\eta_i^{\mu}(\boldsymbol{\beta}) = \boldsymbol{x}_i^t\boldsymbol{\beta}$, $\eta_i^{\varphi}(\boldsymbol{\alpha}) = \boldsymbol{w}_i^t\boldsymbol{\alpha}$ and $\eta_i^{\omega}(\boldsymbol{\gamma}) = \boldsymbol{z}_i^t\boldsymbol{\gamma}$, $i = 1, \ldots, n$ influence the response Y_i. Here, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^t$, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_r)^t$ and $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_q)$ are unknown regression parameters. The matrices $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^t$, $\boldsymbol{W} = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n)^t$ and $\boldsymbol{Z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n)^t$ are called design matrices.

3. *Parametric link components:*
   The linear predictors $\eta_i^{\mu}(\boldsymbol{\beta})$, $\eta_i^{\varphi}(\boldsymbol{\alpha})$ and $\eta_i^{\omega}(\boldsymbol{\gamma})$ are related to the parameters $\mu_i(\boldsymbol{\beta})$, $\varphi_i(\boldsymbol{\alpha})$ and $\omega_i(\boldsymbol{\gamma})$, $i = 1, \ldots, n$ as follows:

   (i) *Mean level*

$$\begin{aligned} E(Y_i \mid \boldsymbol{\beta}) = \mu_i(\boldsymbol{\beta}) &:= E_i e^{\boldsymbol{x}_i^t\boldsymbol{\beta}} = e^{\boldsymbol{x}_i^t\boldsymbol{\beta} + log(E_i)} > 0 \\ \Leftrightarrow \eta_i^{\mu}(\boldsymbol{\beta}) &= \log(\mu_i(\boldsymbol{\beta})) - \log(E_i) \text{ (log link)}, \end{aligned} \tag{2.3}$$

   (ii) *Overdispersion level*

$$\varphi_i(\boldsymbol{\alpha}) := 1 + e^{\boldsymbol{w}_i^t\boldsymbol{\alpha}} > 1$$

$$\Leftrightarrow \eta_i^{\varphi}(\boldsymbol{\alpha}) = \log(\varphi_i(\boldsymbol{\alpha})-1)) \text{ (modified log link)}, \tag{2.4}$$

(iii) *Zero-inflation level*

$$\begin{aligned}\omega_i(\boldsymbol{\gamma}) &:= \frac{e^{\boldsymbol{z}_i^t\boldsymbol{\gamma}}}{1+e^{\boldsymbol{z}_i^t\boldsymbol{\gamma}}} \in (0,1)\\ \Leftrightarrow \eta_i^{\omega}(\boldsymbol{\gamma}) &= \log\left(\frac{\omega_i(\boldsymbol{\gamma})}{1-\omega_i(\boldsymbol{\gamma})}\right) \text{ (logit link)}.\end{aligned} \tag{2.5}$$

We would like to note that a ZIGP regression model is not a generalized linear model (GLM). We denote the joint vector of the regression parameters $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ by $\boldsymbol{\delta}$, i.e. $\boldsymbol{\delta} := (\boldsymbol{\beta}^t, \boldsymbol{\alpha}^t, \boldsymbol{\gamma}^t)^t$, and its maximum likelihood (ML) estimate by $\hat{\boldsymbol{\delta}}$. Additionally, the following abbreviations for $i = 1, \ldots, n$ will be used throughout the paper:

$$\begin{aligned}\mu_i(\boldsymbol{\beta}) &:= e^{\boldsymbol{x}_i^t\boldsymbol{\beta}+log(E_i)}, & & \\ b_i(\boldsymbol{\alpha}) &:= e^{\boldsymbol{w}_i^t\boldsymbol{\alpha}}, & \varphi_i(\boldsymbol{\alpha}) &:= 1+b_i(\boldsymbol{\alpha}),\\ k_i(\boldsymbol{\gamma}) &:= e^{\boldsymbol{z}_i^t\boldsymbol{\gamma}}, & \omega_i(\boldsymbol{\gamma}) &:= \tfrac{k_i(\boldsymbol{\gamma})}{1+k_i(\boldsymbol{\gamma})}\\ \text{and } P_i^0(\boldsymbol{\delta}) &:= \exp\left(-\tfrac{E_i\cdot e^{\boldsymbol{x}_i^t\boldsymbol{\beta}}}{1+e^{\boldsymbol{w}_i^t\boldsymbol{\alpha}}}\right).\end{aligned}$$

For observations $y_1, \ldots, y_n$, the log-likelihood $l(\boldsymbol{\delta})$ of a $ZIGP(\mu_i, \varphi_i, \omega_i)$ regression can be written as

$$\begin{aligned}l(\boldsymbol{\delta}) &= \sum_{i=1}^{n} \mathbb{1}_{\{y_i=0\}}\left[\log\left(e^{\boldsymbol{z}_i^t\boldsymbol{\gamma}}+\exp\left(-\frac{E_i\cdot e^{\boldsymbol{x}_i^t\boldsymbol{\beta}}}{1+e^{\boldsymbol{w}_i^t\boldsymbol{\alpha}}}\right)\right)-\log(1+e^{\boldsymbol{z}_i^t\boldsymbol{\gamma}})\right]\\ &\quad+\mathbb{1}_{\{y_i>0\}}\Big[-\log(1+e^{\boldsymbol{z}_i^t\boldsymbol{\gamma}})+\log(E_i)+\boldsymbol{x}_i^t\boldsymbol{\beta}+(y_i-1)\\ &\quad\times\log(E_i\ e^{\boldsymbol{x}_i^t\boldsymbol{\beta}}+e^{\boldsymbol{w}_i^t\boldsymbol{\alpha}}y_i)-\log(y_i!)-y_i\log(1+e^{\boldsymbol{w}_i^t\boldsymbol{\alpha}})\\ &\quad-\frac{E_i\ e^{\boldsymbol{x}_i^t\boldsymbol{\beta}}+e^{\boldsymbol{w}_i^t\boldsymbol{\alpha}}y_i}{1+e^{\boldsymbol{w}_i^t\boldsymbol{\alpha}}}\Big].\end{aligned} \tag{2.6}$$

Second, the score vector, i.e. the vector of the first derivatives of $l(\boldsymbol{\delta})$, has the following representation:

$$\mathbf{s}_n(\boldsymbol{\delta}) = (s_0(\boldsymbol{\delta}), \ldots, s_p(\boldsymbol{\delta}), \ldots, s_{p+r+1}(\boldsymbol{\delta}), \ldots, s_{p+r+q+2}(\boldsymbol{\delta}))^t, \tag{2.7}$$

where

$$\begin{aligned}s_m(\boldsymbol{\delta}) := \frac{\partial}{\partial\beta_m}l(\boldsymbol{\delta}) &= \sum_{i=1}^{n} x_{im}\Bigg(\mathbb{1}_{\{y_i=0\}}\left[\frac{-P_i^0(\boldsymbol{\delta})\mu_i(\boldsymbol{\beta})/\varphi_i(\boldsymbol{\alpha})}{k_i(\boldsymbol{\gamma})+P_i^0(\boldsymbol{\delta})}\right]\\ &\quad+\mathbb{1}_{\{y_i>0\}}\left[1+\frac{(y_i-1)\mu_i(\boldsymbol{\beta})}{\mu_i(\boldsymbol{\beta})+b_i(\boldsymbol{\alpha})y_i}\right.\\ &\quad\left.-\frac{1}{\varphi_i(\boldsymbol{\alpha})}\mu_i(\boldsymbol{\beta})\right]\Bigg), \text{ for } m=0,\ldots,p,\\ s_{p+1+m}(\boldsymbol{\delta}) := \frac{\partial}{\partial\alpha_m}l(\boldsymbol{\delta}) &= \sum_{i=1}^{n} w_{im}b_i(\boldsymbol{\alpha})\Bigg(\mathbb{1}_{\{y_i=0\}}\left[\frac{P_i^0(\boldsymbol{\delta})\mu_i(\boldsymbol{\beta})/\varphi_i(\boldsymbol{\alpha})^2}{k_i(\boldsymbol{\gamma})+P_i^0(\boldsymbol{\delta})}\right]\\ &\quad+\mathbb{1}_{\{y_i>0\}}\left[\frac{(y_i-1)y_i}{\mu_i(\boldsymbol{\beta})+b_i(\boldsymbol{\alpha})y_i}-\frac{y_i}{\varphi_i(\boldsymbol{\alpha})}\right.\end{aligned} \tag{2.8}$$

$$+\frac{\mu_i(\boldsymbol{\beta})-y_i}{\varphi_i(\boldsymbol{\alpha})^2}\Bigg]\Bigg),\ \text{for } m=0,\dots,r, \tag{2.9}$$

$$s_{p+r+2+m}(\boldsymbol{\delta}) := \frac{\partial}{\partial\gamma_m}l(\boldsymbol{\delta}) = \sum_{i=1}^{n} z_{im}k_i(\boldsymbol{\gamma})\Bigg(\mathbb{1}_{\{y_i=0\}}\left[\frac{1}{k_i(\boldsymbol{\gamma})+P_i^0(\boldsymbol{\delta})}\right] - \frac{1}{1+k_i(\boldsymbol{\gamma})}\Bigg),\ \text{for } m=0,\dots,q. \tag{2.10}$$

To compute the ML estimate $\hat{\boldsymbol{\delta}}$, we simultaneously solve the equations obtained by equating the score vector (2.7) to zero. The Fisher information is needed for the variance estimation of the ML estimates. It is calculated in the appendix.

3 Asymptotic Theory and small sample properties for $ZIGP(\mu_i,\varphi_i,\omega_i)$ regression models

In this section, we will prove consistency and asymptotic normality of the ML estimates. In analogy to Fahrmeir and Kaufmann (1985) we will use the Cholesky square root of the Fisher information matrix to norm the ML estimator. The left Cholesky square root $\boldsymbol{A}^{1/2}$ of a positive definite matrix $\boldsymbol{A}$ is given by the unique lower triangular matrix $\boldsymbol{A}^{1/2}(\boldsymbol{A}^{1/2})^t=\boldsymbol{A}$, which has positive diagonal elements. We write $\boldsymbol{A}^{t/2}:=(\boldsymbol{A}^{1/2})^t$. In addition to that, let $\lambda_{max}(\boldsymbol{A})$ and $\lambda_{min}(\boldsymbol{A})$ be the largest and smallest eigenvalues of $\boldsymbol{A}$, respectively. For vectors we use the L^2 norm $\|\cdot\|_2$, for matrices the spectral norm $\|\boldsymbol{A}\|_2=\lambda_{max}(\boldsymbol{A}^t\boldsymbol{A})^{1/2}=\sup_{\|u\|_2=1}\|\boldsymbol{A}\boldsymbol{u}\|_2$. The vector $\boldsymbol{\delta}_0:=\left(\boldsymbol{\beta}_0^t,\boldsymbol{\alpha}_0^t,\boldsymbol{\gamma}_0^t\right)^t$ consists of the true - yet unknown - model parameters. In addition, we define a neighborhood of the true parameter vector $\boldsymbol{\delta}_0$ by $N_n(\varepsilon):=\{\boldsymbol{\delta}:\ \|\boldsymbol{F}_n^{t/2}(\boldsymbol{\delta}_0)(\boldsymbol{\delta}-\boldsymbol{\delta}_0)\|\le\varepsilon\}$ for $\varepsilon>0$. Also, let $\partial N_n(\varepsilon)=\{\boldsymbol{\delta}:\ \|\boldsymbol{F}_n^{t/2}(\boldsymbol{\delta}_0)(\boldsymbol{\delta}-\boldsymbol{\delta}_0)\|=\varepsilon\}$. For simplicity, we omit the arguments $\boldsymbol{\delta}_0$, $\boldsymbol{\beta}_0$, $\boldsymbol{\alpha}_0$ and $\boldsymbol{\gamma}_0$. Then, we write μ_i instead of $\mu_i(\boldsymbol{\beta}_0)$, φ_i instead of $\varphi_i(\boldsymbol{\alpha}_0)$, k_i for $k_i(\boldsymbol{\gamma}_0)$ and P_i^0 for $P_i^0(\boldsymbol{\delta}_0)$. Admissible sets for $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ are B, A and G. We assume deterministic compact regressors. Further assumptions for the theorem are:

(A1) (Divergence) Let $\frac{n}{\lambda_{min}(\boldsymbol{F}_n)}\le C_1\ \forall\ n\ge 1$, where C_1 is a positive constant,

(A2) (Compact regressors)
$\{\boldsymbol{x}_n,\ n\ge 1\}\subset K_x$, where $K_x\subset\mathbb{R}^{p+1}$ is a compact set,
$\{\boldsymbol{w}_n,\ n\ge 1\}\subset K_w$, where $K_w\subset\mathbb{R}^{r+1}$ is a compact set and
$\{\boldsymbol{z}_n,\ n\ge 1\}\subset K_z$, where $K_z\subset\mathbb{R}^{q+1}$ is a compact set as well.

(A3) ($ZIGP(\mu_i,\varphi_i,\omega_i)$ regression)
Link functions are used as introduced for $ZIGP(\mu_i,\varphi_i,\omega_i)$ regression. Moreover, let $\boldsymbol{\delta}_0$ be an interior point of $B\times A\times G$, where $B\subset\mathbb{R}^{p+1}$, $A\subset\mathbb{R}^{r+1}$ and $G\subset\mathbb{R}^{q+1}$ are open sets.

THEOREM 1 (Consistency and Asymptotic Normality of the ML estimates). Given (A1) - (A3), there is a sequence of random variables $\hat{\boldsymbol{\delta}}_n$ such that

(i) $P(s_n(\hat{\boldsymbol{\delta}}_n)=0)\to 1$, for $n\to\infty$ (asymptotic existence (**AE**)),

(ii) $\hat{\boldsymbol{\delta}}_n\xrightarrow{P}\boldsymbol{\delta}_0$, for $n\to\infty$ (weak consistency (**WC**)),

(iii) $\boldsymbol{F}_n^{t/2}(\hat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_0) \stackrel{D}{\Rightarrow} N_p(\mathbf{0}, \boldsymbol{I}_{p+r+q+3})$, for $n \to \infty$ (asymptotic normality (**AN**)).

A sketch of the proof can be found in the appendix.

Further, we investigated the convergence speed of the ML estimates by generating sets of ZIGP random vectors in different settings (i.e. different designs for μ_i, φ_i and ω_i) and adapting ZIGP regression models on them afterwards. We generated sets of 100 random vectors of length n for each setting and assessed their estimated biases and mean squared errors. For each of the three parameters we used linear predictors consisting of an intercept and an additional equidistant regressor. Increasing sample size n or increasing the μ_i range both improved convergence speed. The results for the sample size are as expected. For the mean we saw that for smaller means the zero-probability grows, which, however, might also arise from high overdispersion or zero-inflation. Therefore, we needed larger sample sizes n to achieve similar results as for smaller means. In our setting, the detection of overdispersion required medium sample sizes of $n \geq 200$. Another result was that increasing dispersion or zero-inflation lowers convergence speed. This is due to higher heterogeneity in the generated data. Details can be found in Erhardt (2006, Sec. 2.4, Fig. 2.1 - 2.4).

4 Model Comparison

4.1 Comparison of several zero-inflated and overdispersed Poisson regression models

As illustrated above, we enhance the Poisson GLM by an overdispersion parameter and a zero-inflation parameter. Models which allow regression on a parameter are denoted by parameter index i. The tree in Fig. 1 sketches an evolution of nine models starting from Poisson regression to (9) $ZIGP(\mu_i, \varphi_i, \omega_i)$ regression described in Section 2.

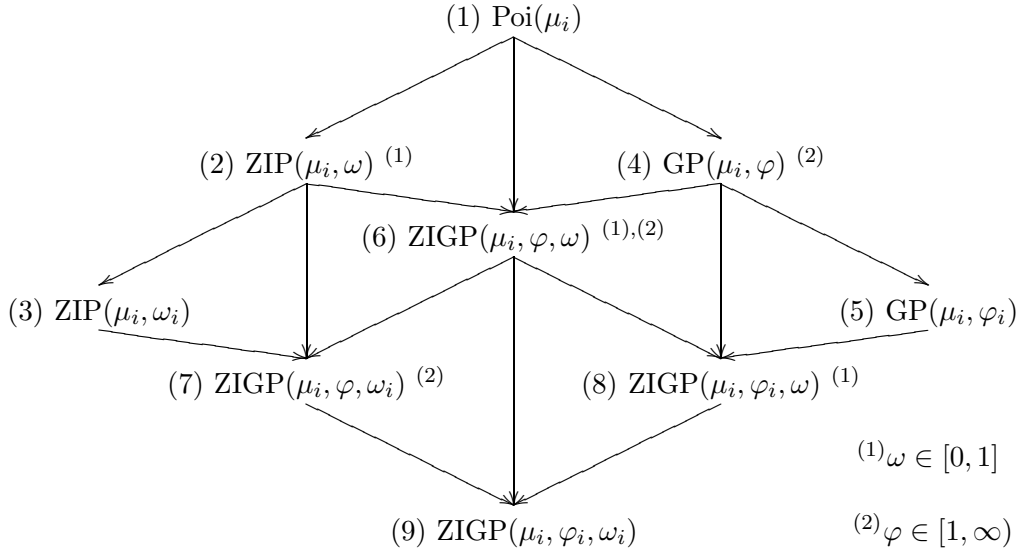

Figure 1: Overview of model enhancements of the Poisson GLM

A covariate being significant in terms of the Wald test (e.g. in the mean design of $Poi(\mu_i)$) can be insignificant in another model (say $ZIGP(\mu_i, \varphi, \omega)$). The same holds for dispersion and

zero-inflation designs. Therefore, all variables chosen in an exploratory data analysis have to be checked for significance and need to be eliminated sequentially. As their design matrices may thus be different, these models need not be nested. With nonnested models, we cannot make use of partial deviance or likelihood ratio tests for model comparison. Therefore, we use the Akaike Information Criterion (AIC) (see e.g. Heiberger and Holland (2004, p. 572)) and the Vuong test for nonnested models (see Vuong (1989)).

The AIC is given by $AIC := -2\ l(\hat{\boldsymbol{\delta}}) + 2k$, where $\hat{\boldsymbol{\delta}}$ is the vector of parameter estimates and k is the number of its components. The smaller the AIC, the better the model is assumed to fit the data. It balances model fit as measured by the log likelihood with model complexity measured by the number of parameters.

Vuong tests compare two regression models which need not to be nested (see Vuong (1989) and Clarke (2003, pp. 3f)). The Kullback–Leibler information criterion $KLIC$ (Kullback and Leibler (1951)) is a measure for the 'distance' between two statistical models. We have

$$KLIC := E_0[\log h_0(Y_i|\boldsymbol{x}_i)] - E_0[\log f(Y_i|\boldsymbol{v}_i, \hat{\boldsymbol{\delta}})], \tag{4.1}$$

where $h_0(\cdot|\cdot)$ is the true conditional density of Y_i given $\boldsymbol{x}_i$ (i.e., the true but unknown model), E_0 is the expectation given the true model, and $\hat{\boldsymbol{\delta}}$ is an estimate of $\boldsymbol{\delta}$ in model with $f(Y_i|\boldsymbol{v}_i, \hat{\boldsymbol{\delta}})$ (which is not the true model). Generally, the better of two models is the one with smaller $KLIC$, for it is closer to the true, but unknown, specification. If model 1 is closer to the true specification, we have

$$\begin{aligned} E_0[\log h_0(Y_i|\boldsymbol{x}_i)] - E_0[\log f_1(Y_i|\boldsymbol{v}_i, \hat{\boldsymbol{\delta}}^1)] &< E_0[\log h_0(Y_i|\boldsymbol{x}_i)] - E_0[\log f_2(Y_i|\boldsymbol{w}_i, \hat{\boldsymbol{\delta}}^2)] \\ \Leftrightarrow E_0 \log \frac{f_1(Y_i|\boldsymbol{v}_i, \hat{\boldsymbol{\delta}}^1)}{f_2(Y_i|\boldsymbol{w}_i, \hat{\boldsymbol{\delta}}^2)} &> 0. \end{aligned} \tag{4.2}$$

Vuong defines statistics $m_i := \log\left(\frac{f_1(y_i|\boldsymbol{v}_i, \hat{\boldsymbol{\delta}}^1)}{f_2(y_i|\boldsymbol{w}_i, \hat{\boldsymbol{\delta}}^2)}\right),\ i = 1, \ldots, n$. Then $\boldsymbol{m} = (m_1, \ldots, m_n)^t$ is a random vector with mean $\boldsymbol{\mu}_0^m = (\mu_1^m, \ldots, \mu_n^m)^t := E_0(\boldsymbol{m})$, if h_0 is the true probability mass function. Hence, we can test H_0: $\boldsymbol{\mu}_0^m = \mathbf{0}$ against $H_1 : \boldsymbol{\mu}_0^m \neq \mathbf{0}$. In other words: 'both models are equally close to the true specification.' Mean $\boldsymbol{\mu}_0^m$, however, is unknown. Further, Vuong defines a test statistic ν and shows that under H_0

$$\nu := \frac{\sqrt{n}[\frac{1}{n}\sum_{i=1}^n m_i]}{\sqrt{\frac{1}{n}\sum_{i=1}^n (m_i - \bar{m})^2}} \xrightarrow{D} N(0,1), \tag{4.3}$$

where $\bar{m} := 1/n \sum_{i=1}^n m_i$. This allows to construct an asymptotic α-Level test of H_0: $\boldsymbol{\mu}_0^m = \mathbf{0}$ versus H_1: not H_0. It rejects H_0 if and only if $|\nu| \geq z_{1-\frac{\alpha}{2}}$, where $z_{1-\frac{\alpha}{2}}$ is the $(1 - \frac{\alpha}{2})$–quantile of the standard normal distribution. The test chooses model 1 over 2, if $\nu \geq z_{1-\frac{\alpha}{2}}$. This is reasonable since significantly high values of ν indicate a higher $KLIC$ of model 1 compared to model 2 according to formula (4.2). Analogously, model 2 is chosen, if $\nu \leq -z_{1-\frac{\alpha}{2}}$.

5 Application: Outsourcing of patent applications

5.1 Data description and exploratory data analysis

The dataset consists of patent information of the European Patent Office. It has been examined and amended with corporate information by Wagner (2005) of *INNO-tec* (Institut für Innovationsforschung, Technologiemanagement und Entrepreneurship). Wagner (2005) points out that

there are two ways of filing a patent application: a company's internal patent department can undergo the application process itself or the company may delegate it to an external patent attorney. Wagner (2005) examines decision drivers using negative binomial panel regression. The survey considers 107 European companies ($i = 1, \ldots, 107$) over a period of eight years from 1993 to 2000. Since data for each company is aggregated over one year, we expect the a company's correlation over years to be small in comparison to a trend effect. Table 1 gives an overview of all influential variables used. A more detailed description can be found in Wagner (2005). Moreover, we used standard exploratory data analysis tools to investigate main effects and two-dimensional interactions on the mean level. In particular, we grouped the data and calculated group means of log observations standardized by their exposure ignoring possible zero-inflation. For details, see Erhardt (2006, Fig. 4.2 - 4.6). The four strongest two-dimensional interactions were **LN.COV** $*$ **BREADTH**, **CHEM.PHA** $*$ **LN.COV**, **CHEM.PHA** $*$ **SQRT.EMP** and **RDmiss** $*$ **CHEM.PHA** (see Erhardt (2006, Fig. 4.11)). We assume independent observations and therefore write Y_i instead of Y_{it}, E_i for E_{it}. We collect all observations in a data vector $\boldsymbol{Y} := (Y_1, \ldots, Y_{856}) = (Y_{1,1}, Y_{1,2}, \ldots, Y_{1,8}, Y_{2,1}, \ldots, Y_{2,8}, Y_{3,1}, \ldots, Y_{107,8})^t$. The observation year will be taken into account by using **YEAR** as a covariate.

There is no established method for finding covariates that have significant influence on the overdispersion parameter. We now propose such a method based on the modified log link (2.4) and ignore zero-inflation. We assume $Y_i \sim GP(\mu_i, \varphi_i) = ZIGP(\mu_i, \varphi_i, \omega_i = 0)$. Then for a class $j, j = 1, \ldots, J$ with n_j members, we have $\sigma_i^2 = \mu_i \cdot \varphi_i^2 \Leftrightarrow \varphi_i = \underset{(-)}{+}\sqrt{\sigma_i^2/\mu_i}$ for $i \in j$. With $\varphi_i = 1 + e^{\boldsymbol{w}_i^t \boldsymbol{\alpha}}$ we get

$$\boldsymbol{w}_i^t \boldsymbol{\alpha} = \log\left(\sqrt{\frac{\sigma_i^2}{\mu_i}} - 1\right) =: f_i(\mu_i, \sigma_i^2). \tag{5.1}$$

We estimate $f_i(\mu_i, \sigma_i^2)$ for observations $i \in$ class j by $f_j(\hat{\mu}_j, \hat{\sigma}_j^2)$, where $\hat{\mu}_j := \frac{1}{n_j} \sum_{i \in j} Y_i$ and $\hat{\sigma}_j := \frac{1}{n_j - 1} \sum_{i \in j} (Y_i - \hat{\mu}_j)^2$. Now we calculate these values for each category of categorical covariates and for each scoring class of metric covariates. If there were no overdispersion in class j, mean and variance would be identical and the fraction σ_j^2/μ_j would be 1. Before taking the logarithm in (5.1), all values would be close to 0. In fact, if we look at the fractions separated by year and industry, the fractions range from 4.8 to 443. Thus, we can use $f_j(\hat{\mu}_j, \hat{\sigma}_j^2)$ as an overdispersion indicator.

In order to determine appropriate covariates for zero-inflation modelling, we calculate empirical logits. This approach arises from binary regression: the event 'observation Y_i is zero' is a binary random variable. Then

$$\hat{\omega}_j := \frac{\#\{Y_i = 0, \delta_{ij} = 1\}}{\#\{Y_i = 0\}}, \text{ where } \delta_{ij} := \begin{cases} 1 & i \in \text{class } j \\ 0 & \text{else} \end{cases}, \tag{5.2}$$

$$\text{and } \hat{logit}(\hat{\omega}_j) := \log\left(\frac{\hat{\omega}_j + \frac{1}{2}}{1 - \hat{\omega}_j + \frac{1}{2}}\right), \; i = 1, \ldots, 856. \tag{5.3}$$

A shift of 1/2 is used in (5.3) to assure calculation also for cases in which a class has not a single observation $Y_i = 0$. If scoring classes are determined by quantiles, the numbers of observations n_j in each class are expected to be roughly equal. Hence, a covariate $\boldsymbol{X}$ with J scoring classes having no influence on the number of zeros in $\boldsymbol{Y}$ is expected to have around $1/J$ of $\boldsymbol{Y}$-zeros in every class.

Variable	Description
$\boldsymbol{Y}$	Response $\boldsymbol{Y}$ represents the number of patents being filed by a company in one year *by an external patent attorney.* The values lie in $[0, 953]$, where zero occurs 129 times.
$\boldsymbol{E}$	This exposure is the yearly total of a company's applications regardless of the application procedure. We have $E_i > 0 \; \forall \; i$.
COV	This metric covariate is the coefficient of variation of a company's number of applications referring to the past five years. For mean modelling we use a log transformation **LN.COV**.
BREADTH	Metric covariate **BREADTH** is a measure for the number of scientific fields a company has handed patent applications in for. High values correspond to broad areas of research. On dispersion level, dummy **BREADTH.49.72** indicates if **BREADTH** is in $(0.490, 0.721]$ (0) or not (1), where 0.490 is the 20%, 0.721 the 60% quantile of the observations. For zero-inflation modelling dummy **BREADTH.06** indicates if a company has a higher (0) or lower (1) **BREADTH** than 0.642 (40% quantile).
EMP	A company's number of employees. For the mean level we use the square root of EMP denoted by **SQRT.EMP**. On dispersion level, dummy **EMP.11291** indicates if a company has more (0) or less (1) than 11 291 employees (40% quantile). On zero-inflation level, dummy **EMP.2023.11291** indicates if **EMP** lies in $(2023, 11291]$ (1) or not (0), which are the $(20\%, 40\%]$ quantiles.
RDP	The average amount spent for a patent in MN Euros is given by **RDP**. It describes the average research and development (R&D) cost per patent. On the mean level, we transform RDP by using its inverse **INV.RDP**. For zero-inflation modelling, dummy **RDP.34** indicates if a company has a higher (0) or lower (1) **RDP** than 3.353 (67% quantile).
RDE	Covariate **RDE** is the average R&D cost per employee in 1000 Euros. Hence, it is a measure for the research intensity. For mean modelling we try the linear, quadratic and cubical transformations, i.e. **RDE1**, **RDE2** and **RDE3**. On dispersion level, dummy **RDE.63** indicates if $RDE \geq 6.3$ (0) or $RDE < 6.3$ (1) (67% quantile).
RDmiss	Dummy variable **RDmiss** indicates if R&D data is missing (1) or not (0).
CHEM.PHA **ELEC.TEL** **ENGINEER** **CAR.SUPP** **MED.BIOT** **OTHER**	These are six industry dummies: Chemical / Pharma, Electro / Telecommunication, Engineering, Cars and Suppliers, Medtech / Biotech and other industries. We also use industry *group* dummies **ELEC.TEL.OTHER**, **CAR.SUPP.OTHER** and **CHEM.PHA.ENGIN**.
YEAR	This is the observation year with values from 1993 to 2000.

Table 1: Description of variables considered in the regression models for the patent data

So, high deviation of the empirical logit from $\log\left(\frac{1/J+1/2}{1-1/J+1/2}\right)$ indicates high influence of $\boldsymbol{X}$ on zero-inflation. Table 2 shows empirical logits of **EMP**, where class 2 has the highest deviation of the reference value $\hat{logit}(0.2) = \log\left(\frac{0.2+0.5}{1-0.2+0.5}\right) = -0.62$.

	Class 1	Class 2	Class 3	Class 4	Class 5
Interval	$[0, 2020]$	$(2020, 11320]$	$(11320, 30249]$	$(30249, 75322]$	$(75322, 466938]$
n_j	172	171	171	171	171
$\hat{logit}(\hat{\omega}_j)$	-0.48	-0.21	-0.72	-0.83	-0.9

Table 2: Empirical logits of five scoring classes of **EMP** calculated according to (5.3)

For mean, dispersion and zero-inflation regression we select the following covariates according to the above strategies. For the mean level, they are **INTERCEPT**, **LN.COV**, **BREADTH**, **SQRT.EMP**, **INV.RDP**, **RDE1**, **RDE2**, **RDE3**, **RDmiss**, **CHEM.PHA**, **ELEC.TEL.OTHER**, **YEAR**, **LN.COV** $*$ **BREADTH**, **CHEM.PHA** $*$ **LN.COV**, **CHEM.PHA** $*$ **SQRT.EMP** and **RDmiss** $*$ **CHEM.PHA**. For overdispersion we select **INTERCEPT**, **ENGINEER**, **CAR.SUPP.OTHER**, **MED.BIOT**, **YEAR**, **BREADTH.49.72**, **EMP.11291** and **RDE.63**. For regression on zero-inflation the chosen covariates are **INTERCEPT**, **EMP.2023.11291**, **BREADTH.06**, **RDP.34** and **CHEM.PHA.ENGIN**. By sequential elimination on an α-level of 5%, however, we get the following regression equations given in Table 3. All covariates have been centered and standardized for numerical stability.

Model comparison for our data is carried out by using the methods discussed in Section 4. Models (1) thru (9) can also be found in Table 4 listed in rows (I) and columns (II). Entries of this lower triangular matrix show the results of Vuong tests for every model combination of a (I) with a (II) model. We choose a significance level of $\alpha = 5\%$, so $z_{1-\frac{\alpha}{2}} = 1.96$. In the first line of each cell, the Vuong statistic ν is given. In the second row the decision of the Vuong test is shown, i.e. if model (I) or (II) is better. Next to that we see the p-values of ν. For example, the most upper left cell refers to model (I) = (2) $ZIP(\mu_i, \omega)$ compared to (II) = (1) $Poi(\mu_i)$. The Vuong statistic is $\nu = 4.2$, which implies that Vuong prefers model (I) $ZIP(\mu_i, \omega)$ (see line 2). The p-value of ν is $< 10^{-4}$.

We now discuss the consequences of the Poisson GLM enhancements.

Adding a zero-inflation parameter: Taking zero-inflation into a model, the AIC decreases. Comparing (1) $Poi(\mu_i)$ with model (2) $ZIP(\mu_i, \omega)$, the AIC decreases from 23 896 to 19 183 (see Table 3). The Vuong test also prefers the ZIP model with a test statistic of $\nu = 4.2$ (see Table 4).

Adding a dispersion parameter: Adding a dispersion parameter has the strongest positive impact on model quality. The AIC for (1) $Poi(\mu_i)$ is 23 896, whereas for (4) $GP(\mu_i, \varphi)$ it is already 6 852, a drop of 71%. The Vuong statistic is very high, it is $\nu = 10.8$. This is a strong indication that our data is in fact overdispersed.

Regression on the zero-inflation parameter: If we allow regression on the zero-inflation parameter, the AIC decreases again. For instance, comparing model (2) $ZIP(\mu_i, \omega)$ with (3) $ZIP(\mu_i, \omega_i)$, the AIC falls from 19 183 to 19 110. Further, $\nu = 4.32$, so Vuong prefers model (2) $ZIP(\mu_i, \omega)$, too. The p-value is < 0.01.

Model	Model Equation μ	Model Equation φ	Model Equation ω	$l(\hat{\boldsymbol{\delta}})$	$p + r+q$	AIC
(1) $Poi(\mu_i)$	offset(**E**) + 1 + **LN.COV** + **BREADTH** + **SQRT.EMP** + **INV.RDP** + **RDE1** + **RDE2** + **RDE3** + **RDmiss** + **CHEM.PHA** + **ELEC.TEL.OTHER** + **YEAR** + **LN.COV** * **BREADTH** + **CHEM.PHA** * **LN.COV** + **CHEM.PHA** * **SQRT.EMP** + **RDmiss** * **CHEM.PHA**	$\varphi_i = 1\ \forall i$	$\omega_i = 0\ \forall i$	-11 931.9	16	23 896
(2) $ZIP(\mu_i, \omega)$	offset(**E**) + 1 + **LN.COV** + **BREADTH** + **SQRT.EMP** + **INV.RDP** + **RDE1** + **RDE2** + **RDE3** + **RDmiss** + **CHEM.PHA** + **ELEC.TEL.OTHER** + **YEAR** + **LN.COV** * **BREADTH** + **CHEM.PHA** * **LN.COV** + **CHEM.PHA** * **SQRT.EMP** + **RDmiss** * **CHEM.PHA**	$\varphi_i = 1\ \forall i$	$\omega_i = \omega\ \forall i$	-9 574.6	17	19 183
(3) $ZIP(\mu_i, \omega_i)$	offset(**E**) + 1 + **LN.COV** + **BREADTH** + **SQRT.EMP** + **INV.RDP** + **RDE1** + **RDE2** + **RDE3** + **RDmiss** + **CHEM.PHA** + **ELEC.TEL.OTHER** + **YEAR** + **LN.COV** * **BREADTH** + **CHEM.PHA** * **LN.COV** + **CHEM.PHA** * **SQRT.EMP** + **RDmiss** * **CHEM.PHA**	$\varphi_i = 1\ \forall i$	1 + **BREADTH.06** + **EMP.2023.11291** + **RDP.34** + **CHEM.PHA.ENGIN**	-9 533.8	21	19 110
(4) $GP(\mu_i, \varphi)$	offset(**E**) + 1 + **BREADTH** + **SQRT.EMP** + **INV.RDP** + **RDE1** + **RDE2** + **RDE3** + **CHEM.PHA** + **ELEC.TEL.OTHER**	$\varphi_i = \varphi\ \forall i$	$\omega_i = 0\ \forall i$	-3 416.1	10	6 852
(5) $GP(\mu_i, \varphi_i)$	offset(**E**) + 1 + **BREADTH** + **SQRT.EMP** + **INV.RDP** + **RDE1** + **RDE2** + **RDE3** + **RDmiss** + **CHEM.PHA** + **ELEC.TEL.OTHER** + **RDmiss** * **CHEM.PHA**	1 + **CAR.SUPP.OTHER** + **MED.BIOT** + **YEAR** + **BREADTH.49.72** + **EMP.11291** + **RDE.63**	$\omega_i = 0\ \forall i$	-3 356.8	18	6 750
(6) $ZIGP(\mu_i, \varphi, \omega)$	offset(**E**) + 1 + **LN.COV** + **BREADTH** + **SQRT.EMP** + **INV.RDP** + **RDE1** + **RDmiss** + **CHEM.PHA** + **ELEC.TEL.OTHER** + **YEAR** + **LN.COV** * **BREADTH** + **RDmiss** * **CHEM.PHA**	$\varphi_i = \varphi\ \forall i$	$\omega_i = \omega\ \forall i$	-3 308.6	14	6 645
(7) $ZIGP(\mu_i, \varphi, \omega_i)$	offset(**E**) + 1 + **LN.COV** + **BREADTH** + **SQRT.EMP** + **INV.RDP** + **RDE1** + **CHEM.PHA** + **ELEC.TEL.OTHER** + **LN.COV** * **BREADTH**	$\varphi_i = \varphi\ \forall i$	1 + **BREADTH.06**	-3 298.5	12	6 621
(8) $ZIGP(\mu_i, \varphi_i, \omega)$	offset(**E**) + 1 + **LN.COV** + **BREADTH** + **SQRT.EMP** + **RDE1** + **RDmiss** + **CHEM.PHA** + **ELEC.TEL.OTHER** + **LN.COV** * **BREADTH**	1 + **CAR.SUPP.OTHER** + **YEAR** + **EMP.11291** + **RDE.63**	$\omega_i = \omega\ \forall i$	-3 269	15	6 568
(9) $ZIGP(\mu_i, \varphi_i, \omega_i)$	offset(**E**) + 1 + **LN.COV** + **BREADTH** + **SQRT.EMP** + **INV.RDP** + **RDE1** + **CHEM.PHA** + **ELEC.TEL.OTHER** + **LN.COV** * **BREADTH**	1 + **ENGINEER** + **CAR.SUPP.OTHER** + **YEAR** + **EMP.11291** + **RDE.63**	1 + **BREADTH.06** + **RDP.34** + **CHEM.PHA.ENGIN**	-3 245.2	19	6 528

Table 3: Model equations and AIC for each of the nine models after sequential elimination of non-significant covariates

(II) / **(I)**	(1) $Poi(\mu_i)$	(2) ZIP (μ_i, ω)	(3) ZIP (μ_i, ω_i)	(4) GP (μ_i, φ)	(5) GP (μ_i, φ_i)	(6) $ZIGP$ (μ_i, φ, ω)	(7) $ZIGP$ $(\mu_i, \varphi, \omega_i)$	(8) $ZIGP$ $(\mu_i, \varphi_i, \omega)$
(2) ZIP (μ_i, ω)	$\nu = 4.2$ V: (I) $< 10^{-4}$							
(3) ZIP (μ_i, ω_i)	$\nu = 4.27$ V: (I) $< 10^{-4}$	$\nu = 4.32$ V: (I) $< 10^{-4}$						
(4) GP (μ_i, φ)	$\nu = 10.8$ V: (I) $< 10^{-22}$	$\nu = 9.94$ V: (I) $< 10^{-22}$	$\nu = 9.88$ V: (I) $< 10^{-22}$					
(5) GP (μ_i, φ_i)	$\nu = 10.8$ V: (I) $< 10^{-22}$	$\nu = 10.1$ V: (I) $< 10^{-22}$	$\nu = 9.99$ V: (I) $< 10^{-22}$	$\nu = 3.94$ V: (I) $< 10^{-4}$				
(6) $ZIGP$ (μ_i, φ, ω)	$\nu = 10.8$ V: (I) $< 10^{-22}$	$\nu = 10.3$ V: (I) $< 10^{-22}$	$\nu = 10.2$ V: (I) $< 10^{-22}$	$\nu = 4.21$ V: (I) $< 10^{-4}$	$\nu = 2.08$ V: (I) 0.04			
(7) $ZIGP$ $(\mu_i, \varphi, \omega_i)$	$\nu = 10.8$ V: (I) $< 10^{-22}$	$\nu = 10.3$ V: (I) $< 10^{-22}$	$\nu = 10.2$ V: (I) $< 10^{-22}$	$\nu = 4.29$ V: (I) $< 10^{-4}$	$\nu = 2.32$ V: (I) 0.02	$\nu = 1.88$ V: none 0.06		
(8) $ZIGP$ $(\mu_i, \varphi_i, \omega)$	$\nu = 10.8$ V: (I) $< 10^{-22}$	$\nu = 10.3$ V: (I) $< 10^{-22}$	$\nu = 10.2$ V: (I) $< 10^{-22}$	$\nu = 5.15$ V: (I) $< 10^{-6}$	$\nu = 4.01$ V: (I) $< 10^{-4}$	$\nu = 3.1$ V: (I) < 0.002	$\nu = 2.16$ V: (I) 0.03	
(9) $ZIGP$ $(\mu_i, \varphi_i, \omega_i)$	$\nu = 10.8$ V: (I) $< 10^{-22}$	$\nu = 10.3$ V: (I) $< 10^{-22}$	$\nu = 10.3$ V: (I) $< 10^{-22}$	$\nu = 5.69$ V: (I) $< 10^{-6}$	$\nu = 4.73$ V: (I) $< 10^{-4}$	$\nu = 4.28$ V: (I) $< 10^{-4}$	$\nu = 3.91$ V: (I) $< 10^{-4}$	$\nu = 2.9$ V: (I) < 0.004

Table 4: Model comparison using the Vuong test

Regression on the dispersion parameter: Comparing model (4) $GP(\mu_i, \varphi)$ with model (5) $GP(\mu_i, \varphi_i)$, we see a drop in AIC from 6 852 to 6 750. The Vuong statistic is $\nu = 3.94$.

All in all, model (9) $ZIGP(\mu_i, \varphi_i, \omega_i)$ seem to be the 'best' for our data in terms of AIC and Vuong. By eliminating non-significant covariates in $Poi(\mu_i)$ and adding overdispersion and zero-inflation regression, the AIC falls from 23 896 to 6 528. That is a AIC drop of 73%. Vuong applied to these two models has the highest test significance of all models compared, its statistic is $\nu = 10.8$.

6 Model interpretation

We will now interpret model (9) $ZIGP(\mu_i, \varphi_i, \omega_i)$. Parameter estimates and their estimated standard errors together with the p-values of the corresponding Wald tests can be found in Table 5.

	Estimate	Std. Error	p-value
MEAN REGRESSION			
INTERCEPT	-0.951	0.059	0.000
log(**COV**)	0.031	0.036	0.384
BREADTH	0.041	0.032	0.195
$\mathbf{EMP}^{1/2}$	-0.394	0.027	0.000
$\mathbf{RDP}^{-1}$	-0.123	0.033	0.000
RDE	0.124	0.033	0.000
Chemical / Pharma	-0.272	0.099	0.006
Electro / Telecommunication / Other	0.306	0.066	0.000
log(**COV**) * **BREADTH**	-0.104	0.034	0.003
OVERDISPERSION REGRESSION			
INTERCEPT	1.968	0.086	0.000
Engineering	-0.426	0.171	0.013
Cars / Suppliers / Other	-0.488	0.096	0.000
YEAR	0.161	0.042	0.000
$\mathbb{1}\{\mathbf{EMP} < 11\ 291\}$	-0.587	0.103	0.000
$\mathbb{1}\{\mathbf{RDE} < 6.3\}$	-0.305	0.096	0.002
ZERO-INFLATION REGRESSION			
INTERCEPT	-4.282	0.586	0.000
$\mathbb{1}\{\mathbf{BREADTH} < 0.642\}$	2.271	0.578	0.000
$\mathbb{1}\{\mathbf{RDP} < 3.353\}$	-1.241	0.522	0.017
Chemical / Pharma / Engineering	1.085	0.413	0.009
Mean range $\hat{\mu}$	[0.18, 446.3]		
Overdispersion parameter range $\hat{\varphi}$	[2.41, 10.15]		
Zero-inflation parameter range $\hat{\omega}$	[0.00, 0.28]		

Table 5: Summary of model (9) $ZIGP(\mu_i, \varphi_i, \omega_i)$ using centered and standardized covariates

First of all, we see that **RDmiss** has been dropped during the sequential elimination of non-significant variables, which is in line with Wagner (2005, p. 27), whose model has insignificant

RDmiss as well. This is comforting since **RDmiss** indicates missing R&D data. Otherwise, we would have a significant systematic error. We now want to perform an analysis of the impact of mean, overdispersion and zero-inflation regressors. Thus, we calculate mean, overdispersion and zero-inflation functions of these covariates and fix all remaining covariates. For metric covariates we use their mode or mean, for categorical covariates we compare their categories. Exposure **E** will also be replaced by its mode.

6.1 Outsourcing rate as a function of covariates influencing the mean level

We are interested in outsourcing rates $E(Y_i)/E_i$ rather than absolute 'outsourced' patent numbers. In order to find modes of metric variables, we used empirical density estimates. The mode in terms of transformed but unnormed values for **SQRT.EMP** is 58.35, for **INV.RDP** it is 2.824 and **RDE1** has -0.0153. The common mode of the interacting covariates (**LN.COV**, **BREADTH**) is $(-1.000, 0.7769)$. The exposure mode will be denoted as E^M and is calculated to be 13.36. Since we are interested in seeing the effects of the covariates on the original scale we have to reverse any centering and scaling and any covariate transformation used. For details, again see Erhardt (2006, p. 129).

For all outsourcing functions we have to separate between the three industry groups Chemical / Pharma, Electro / Telecommunication / Others and all remaining industries. Fig. 2 shows the resulting outsourcing rates as functions of covariates **EMP**, **RDP** and **RDE**.

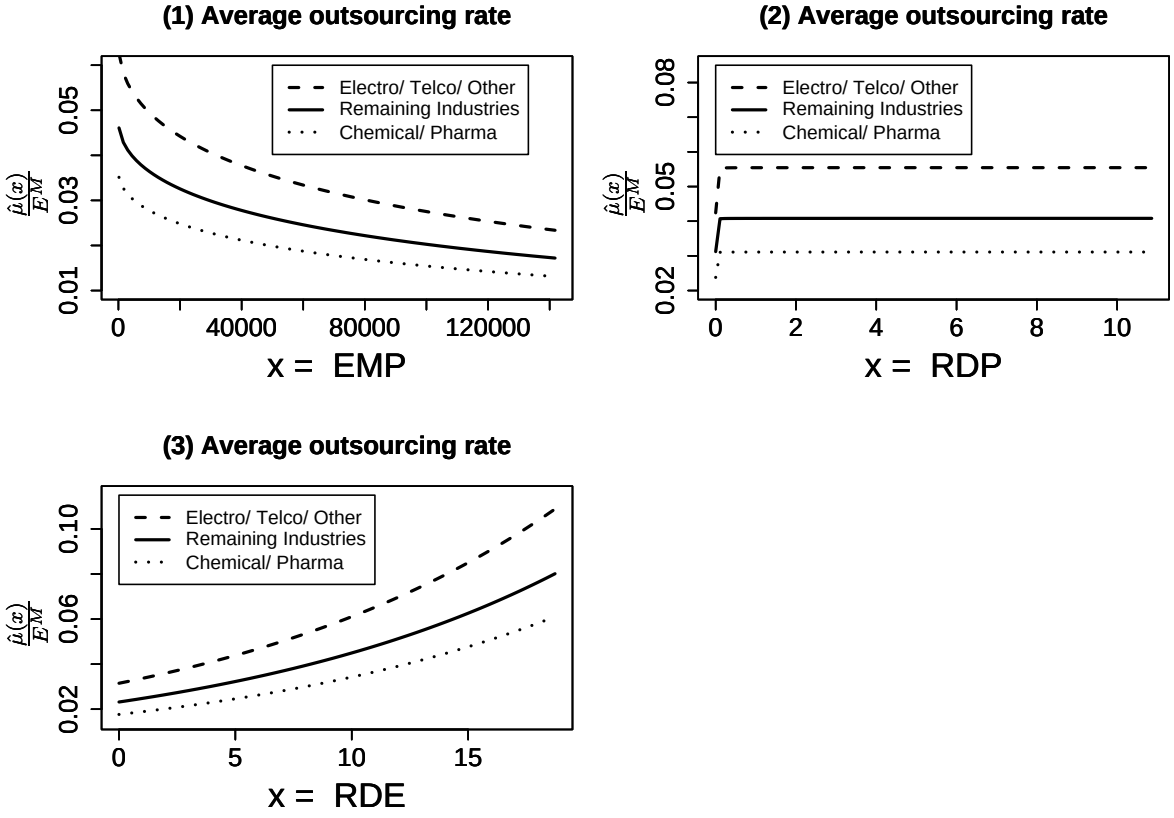

Figure 2: Influence of **EMP**, **RDP** and **RDE** on the outsourcing rate while fixing other covariates by their modes

The resulting outsourcing rates range between 1.5% and 12% depending on the setting.

Again, these figures are related to a fixed exposure. Hence, we can interpret their relation among each other rather than the absolute rates. Increasing firm size in terms of employees reduces the share of outsourced applications. As Wagner (2005, p. 28) explains, 'larger firms are more likely to have their own IP-department and hence more likely to process a higher share of the workload internally'. For **RDP**, we get a fairly small coefficient $\hat{\beta}_4 = -0.123$. Also, for very small values **RDP**< 10 the outsourcing share is low. For larger values, it is high and quite constant. A reason for that is that 262 observations have no R&D information and hence have **RDP**$= 0$. These companies, however, have an average of only 20 000 employees, whereas the overall average is 50 000. It seems like smaller companies (which we know show higher tendency to contract out) are likely to be unable to provide R&D data. Whether or not expensive R&D preceded a patent (high **RDP**) has minor impact on the outsourcing rate. This is in line with Wagner (2005, p. 28). Our model predicts that for higher R&D intensity, companies are likely to file their patents themselves. Wagner (2005, p. 28), who obtains a positive influence of **RDE** as well, stresses that a negative development would have been more plausible: the higher spending on R&D per employee, the greater a company's focus on research. These companies are more likely to have their own patent departments.

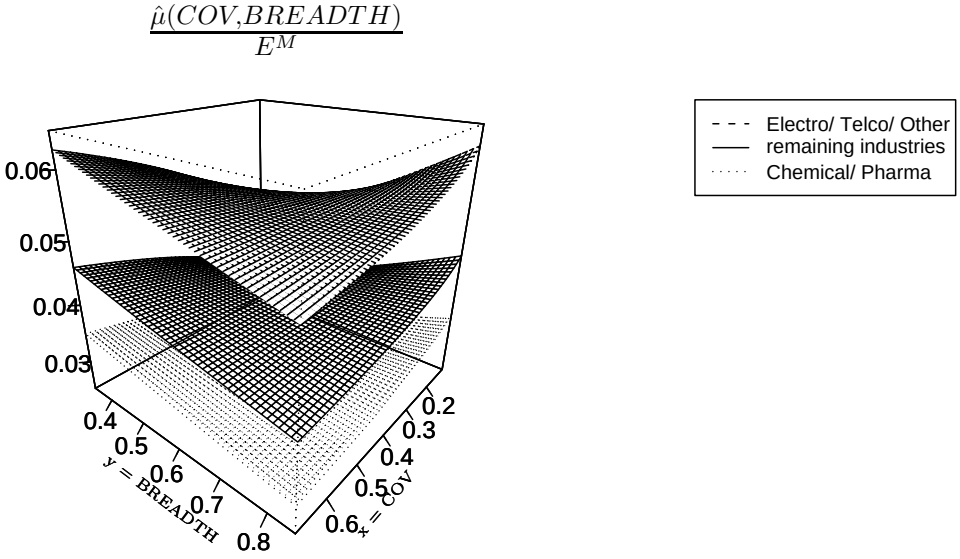

Figure 3: Common influence of COV and BREADTH on the estimated outsourcing rate $\hat{\mu}(COV, BREADTH)/E^M$ per industries

Fig. 3 shows projected outsourcing rates affected by both **COV** and **BREADTH**. The deflection arises from the interaction of these two effects. The higher **BREADTH** or **COV** as singular effects, the less patents are outsourced. Interaction effects strongly decrease our outsourcing projection if both variables have low values. There are only few companies having low **BREADTH** and **COV** at the same time. Their low outsourcing rates seem to arise from their unique situation. Schneider Electronics for instance had to face severe losses throughout the nineties and therefore had to cut their R&D activities way below the industry average. Filing patents themselves might arise from expenditure reasons. Heidelberger Druckmaschinen, however, is the world market leader and went public in 1997. Market leaders are known to trust their own patent departments more than external attorneys.

Finally, we look at industry differences. The only industry dummies left are those for Chem-

ical / Pharma and for Electronics / Telecommunication / Other. Chemical / Pharma companies have the lowest outsourcing rates (3.11%). Especially for pharmaceutical companies, one single and very complex patent protects one product, e.g. a drug. An important part of each patent application is thorough research if a similar patent has already been filed. Chemical / Pharma companies were among the first ones to establish internal patent databases to ensure quick and reliable research for equal or similar patents. Thus, it makes sense to have an internal department undergo the whole application process. Electro / Telco & Other have a predicted outsourcing share of 5.54%. The reason is the different role patents play for them. Applications in the Electronics industry are mainly driven by semiconductor manufacturers, which file a lot of 'small' patents. In many cases, companies confederate and create a common unique product standard (crossover licenses). Therefore, an existing patent may be replaced by a common patent. This is often done by common (external) patent attorneys. For all remaining industries we predict an outsourcing rate of 4.08%.

6.2 Overdispersion and zero-probability as functions of regression covariates

According to (2.1) and (2.2) the coefficient of variation of a random $Y_i \sim ZIGP(\mu_i,\, \varphi_i, \omega_i)$ is $\frac{Var(Y_i)}{E(Y_i)} = \varphi_i^2 + \mu_i\omega_i$. There are only categorical covariates for overdispersion:
$\boldsymbol{w} := (\mathbf{1},\ \textbf{ENGINEER},\ \textbf{CAR.SUPP.OTHER},\ \textbf{YEAR}^s,\ \textbf{EMP.11291},\ \textbf{RDE.63})$, whereas $\textbf{YEAR}^s$ stands for centered and standardized **YEAR** values. With the modified log link we get an overdispersion function $\hat{\varphi}(\boldsymbol{w})$, which is defined as

$$\hat{\varphi}(\boldsymbol{w}) \;\; := \;\; 1 + \exp\Big(\hat{\alpha}_0 + w_1 \cdot \hat{\alpha}_1 + \ldots + w_5 \cdot \hat{\alpha}_5\Big). \tag{6.1}$$

We can use this to estimate $\frac{Var(Y)}{E(Y)} = \varphi^2 + \mu \cdot \omega$ by

$$\frac{\hat{Var}(Y|\boldsymbol{X}=\boldsymbol{x}, \boldsymbol{W}=\boldsymbol{w}, \boldsymbol{Z}=\boldsymbol{z})}{\hat{E}(Y|\boldsymbol{X}=\boldsymbol{x}, \boldsymbol{Z}=\boldsymbol{z})} := \hat{\varphi}(\boldsymbol{w})^2 + \hat{\mu}(\boldsymbol{x}) \cdot \hat{\omega}(\boldsymbol{z}). \tag{6.2}$$

Companies are interested in a prediction of their zero-probability $P(Y=0)$ rather than their zero inflation $\hat{\omega}$. This is the probability that a company *has filed every patent in a certain year itself*. A zero can arise from the binary zero-inflation process or from the GP variable, thus we predict

$$\hat{P}(Y=0|\boldsymbol{X}=\boldsymbol{x}, \boldsymbol{W}=\boldsymbol{w}, \boldsymbol{Z}=\boldsymbol{z}) := \hat{\omega}(\boldsymbol{z}) + (1-\hat{\omega}(\boldsymbol{z})) \cdot \exp\Big(-\frac{\hat{\mu}(\boldsymbol{x})}{\hat{\varphi}(\boldsymbol{w})}\Big). \tag{6.3}$$

Zero-inflation will be regarded as a function $\hat{\omega}(\boldsymbol{z})$ of covariate $\boldsymbol{z}$, where
$\boldsymbol{z} := (\mathbf{1},\ \textbf{BREADTH.06},\ \textbf{RDP.34},\ \textbf{CHEM.PHA.ENGIN})$. Then we get

$$\hat{\omega}(\boldsymbol{z}) = \frac{\exp(\hat{\gamma}_0 + z_1 \cdot \hat{\gamma}_1 + \ldots + z_q \cdot \hat{\gamma}_q)}{1 + \exp(\hat{\gamma}_0 + z_1 \cdot \hat{\gamma}_1 + \ldots + z_q \cdot \hat{\gamma}_q)} \tag{6.4}$$

We see that for estimating $\frac{Var(Y)}{E(Y)}$ and $P(Y=0)$, we need both $\hat{\varphi}(\boldsymbol{w})$ and $\hat{\omega}(\boldsymbol{z})$ (see (6.2) and (6.3)). Consequently, we have to define common parameters, i.e. a union of the categorical settings for overdispersion and zero-inflation regression. As we are only looking at grouped industries, we can define four new common groups as 'Cars / Supplier / Other', 'Medtech / Biotech / Electro / Telco', 'Engineering' and 'Chemical / Pharma'. Remaining dummies are

EMP.11291 and **RDE.63** for overdispersion and **RDP.34** and **BREADTH.06** for zero-inflation. Therefore, we have to consider sixteen settings for each of these four industry groups, i.e. 64 different classes. We investigate, however, only the 14 cases with the most observations, which account for 63 of the 107 companies. These classes can be found in Table 6 denoted by j. Columns $\bar{\mu}_j$ thru $\hat{\omega}_j$ are estimates of μ_j, φ_j and ω_j, where $\hat{\varphi}_j$ is calculated as in (6.1), $\hat{\omega}_j$ as in (6.4). We will look at year 2 000 since it should be most interesting. For both overdispersion and zero-inflation, we need an estimate of the mean $\hat{\mu}$. In order to get appropriate values for class j, we use means $\bar{\mu}_j$ of the fitted values for μ for those companies in class j. They are given by $\bar{\mu}_j := 1/n_j \sum_{i \in I_j} \hat{\mu}_i$, where $\hat{\mu}_i = \hat{\mu}(\boldsymbol{x}_i, \boldsymbol{w}_i, \boldsymbol{z}_i)$. Column $\frac{\hat{Var}(Y|\boldsymbol{x},\boldsymbol{w},\boldsymbol{z})}{\hat{E}(Y|\boldsymbol{x},\boldsymbol{z})}$ lists results from formula (6.2), $\hat{P}(Y = 0|\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{z})$ from (6.3).

The quotes $\frac{\hat{Var}(Y|\boldsymbol{x},\boldsymbol{w},\boldsymbol{z})}{\hat{E}(Y|\boldsymbol{x},\boldsymbol{z})}$ range between 11 and 136 and indicate high overdispersion. For large companies in terms of employees, overdispersion is especially high. For example, class 7 has high **EMP** while class 11 has low. All other parameters are equal. Here, $\frac{\hat{Var}(Y|\boldsymbol{x},\boldsymbol{w},\boldsymbol{z})}{\hat{E}(Y|\boldsymbol{x},\boldsymbol{z})}$ increases from 39.4 for small **EMP** to 135.9 for high **EMP**. This is mainly driven by the higher mean $\bar{\mu}_7 = 277.8$ compared to $\bar{\mu}_{11} = 19.5$. Additionally, we have a higher range in employee numbers and therefore higher heterogeneity for large **EMP**.

The same holds for **RDE**: overdispersion rises as R&D intensity increases. An example are classes 1 and 2, where the coefficient of variation increases from 26.7 to 44.3. While R&D intensity with **RDE** $\in [0; 6.344)$ is quite compact, **RDE** $\in [6.344; 215.17]$ is a broad range.

Industry Engineering has the lowest overdispersion, Cars / Suppliers and Other are second. All remaining industries show high overdispersion. For Engineering we can state that it is typical for this industry that a large number of patents are developed. The number of patents filed is often regarded as a mean to boost the company's competitive position. Thus, the management works with patent number objectives the R&D departments have to fulfil. Accomplishing these aims is easier for an industry which needs many patents as they might just file minor inventions as patents. Often, this results in a 'precision landing' as far as patent numbers are concerned. Also, the number of patents filed the year before is often regarded as a minimum goal for the current year. These effects decrease the application variance and hence overdispersion.

Fig. 4 shows the influence of the observation year on the coefficient of variation per industry. The legend lists the classes in descending order. We predict a positive super-linear development. Again, Engineering shows lowest overdispersion with $\frac{\hat{Var}(Y|\boldsymbol{x},\boldsymbol{w},\boldsymbol{z})}{\hat{E}(Y|\boldsymbol{x},\boldsymbol{z})}$ in $[14.3, 14.5]$. Highest overdispersion occurs in 'Medtech / Biotech / Electro / Telco' which have values in $[37.1, 135.9]$.

The zero-probabilities range between 0.4% and 34%. A small **BREADTH** has a strong positive impact on $\hat{P}(Y = 0|\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{z})$. See for example classes 5 and 6, where $\hat{P}(Y = 0|\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{z})$ rises from 1.37% to 18.77%, as **BREADTH** decreases. This is evident since small research areas make it easier for companies to have their patent activity covered by internal patent attorneys.

Engineering is likely to have high zero-probability, see for instance class 13, where $\hat{P}(Y = 0|\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{z}) = 33.68\%$. Chemical / Pharma show very low zero-probability. Especially pharmaceutical companies have developed own patent databases and therefore are likely to file all patents themselves.

If we compare high **RDP** with low ones, we predict a higher zero-probability for the first setting (compare for instance classes 1 and 4). Here, $\hat{P}(Y = 0|\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{z})$ rises from 0.85% to 4.43%. Expensive patents (high **RDP**) are likely to be filed exclusively by internal departments. It seems like in crucial situations, companies trust their own patent departments more than

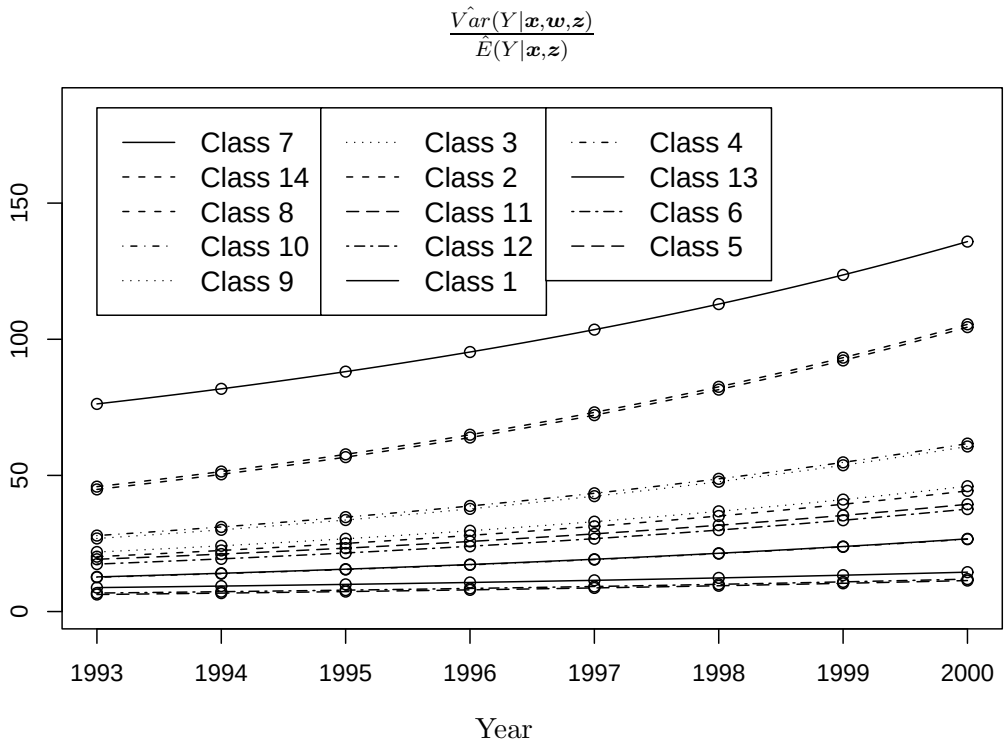

Figure 4: Influence of YEAR on the coefficient of variation while fixing other covariates by their modes; legend in descending order

external patent attorneys.

7 Conclusions and Discussions

We introduced a $ZIGP(\mu_i, \varphi_i, \omega_i)$ regression model, which not only extends the known Poisson GLM by overdispersion and zero-inflation parameters but also allows for regression on these parameters. Further, we developed the necessary asymptotic theory for these models, thus filling a theoretical and practical gap in this research area. In particular, we showed that there exists a sequence of random variables, whose representations are solutions of the score equations for each n. These solutions converge in probability to the true model parameter vector. In addition to that, we showed the asymptotic normality of these score solutions. The detection of overdispersion requires medium sample sizes of $n \geq 200$.

Moreover, we carried out a comparison of different models based on the Poisson model using a dataset investigating the determinants of patent outsourcing. We illustrated that every extension of our $ZIGP(\mu_i, \varphi_i, \omega_i)$ regression model improved model fit in terms of the AIC statistic for the given application. Vuong tests have been used to compare these nonnested models. Both AIC and Vuong tests chose the introduced $ZIGP(\mu_i, \varphi_i, \omega_i)$ regression model as the one fitting our data best. All in all, the AIC decreased by 73% as compared to the Poisson GLM.

		Class defining covariates								
Class			dispersion covariates		zero-inflation covariates					
j	n_j	Industry	**EMP**	**RDE**	**RDP**	**BREADTH**	$\bar{\mu}_j$	$\hat{\varphi}_j$	$\hat{\omega}_j$	$\frac{\hat{Var}(Y\|\boldsymbol{x},\boldsymbol{w},\boldsymbol{z})}{\hat{E}(Y\|\boldsymbol{x},\boldsymbol{z})}$	$\hat{P}(Y=0\|\boldsymbol{x},\boldsymbol{w},\boldsymbol{z})$
1.	6	Cars / Suppl. / Other	$\geq 11\,291$	< 6.3	≥ 3.353	≥ 0.642	18.0	5.14	1.36%	26.7	4.34%
2.	5	Cars / Suppl. / Other	$\geq 11\,291$	≥ 6.3	≥ 3.353	≥ 0.642	36.4	6.62	1.36%	44.3	1.77%
3.	5	Cars / Suppl. / Other	$\geq 11\,291$	≥ 6.3	≥ 3.353	< 0.642	18.6	6.62	11.81%	46.0	17.08%
4.	5	Cars / Suppl. / Other	$\geq 11\,291$	< 6.3	< 3.353	≥ 0.642	27.8	5.14	0.40%	26.6	0.85%
5.	5	Cars / Suppl. / Other	$< 11\,291$	< 6.3	≥ 3.353	≥ 0.642	33.8	3.30	1.36%	11.4	1.37%
6.	5	Cars / Suppl. / Other	$< 11\,291$	< 6.3	≥ 3.353	< 0.642	8.4	3.30	11.81%	11.9	18.77%
7.	4	Medt. / Biot. / Elec. / Telc.	$\geq 11\,291$	≥ 6.3	≥ 3.353	< 0.642	277.8	10.15	11.81%	135.9	11.81%
8.	4	Medt. / Biot. / Elec. / Telc.	$\geq 11\,291$	≥ 6.3	< 3.353	≥ 0.642	350.3	10.15	0.40%	104.5	0.40%
9.	4	Medt. / Biot. / Elec. / Telc.	$\geq 11\,291$	< 6.3	≥ 3.353	≥ 0.642	43.8	7.75	1.36%	60.6	1.71%
10.	4	Medt. / Biot. / Elec. / Telc.	$\geq 11\,291$	< 6.3	≥ 3.353	< 0.642	14.1	7.75	11.81%	61.7	26.18%
11.	4	Medt. / Biot. / Elec. / Telc.	$< 11\,291$	≥ 6.3	≥ 3.353	< 0.642	19.5	6.09	11.81%	39.4	15.39%
12.	4	Medt. / Biot. / Elec. / Telc.	$< 11\,291$	≥ 6.3	< 3.353	< 0.642	14.4	6.09	3.73%	37.6	12.76%
13.	4	Engineering	$< 11\,291$	< 6.3	≥ 3.353	< 0.642	9.0	3.45	28.38%	14.5	33.68%
14.	4	Chemical / Pharma	$\geq 11\,291$	≥ 6.3	≥ 3.353	≥ 0.642	61.6	10.15	3.93%	105.5	4.15%

Table 6: Estimated coefficient of variation and probability of no outsourced patent application for 14 classes which have the largest numbers of observations in year 2000

A model interpretation confirmed insights of former work on the given dataset from an economic point of view. We added an analytical and economic interpretation for overdispersion and zero-inflation drivers. The expected outsourcing rate is driven by the industry a company belongs to. Electronic and Telecommunication companies show particularly high, Chemical / Pharma companies low outsourcing shares. The number of employees has a strong negative, R&D costs per employee a positive influence. Overdispersion, in terms of the predicted coefficient of variation of the outsourcing shares, strongly depends on the industry as well. Engineering companies are likely to have low overdispersion. Large companies with high R&D spending per employee are predicted to have high overdispersion. Zero-probability (i.e. the probability of no outsourcing of patent applications whatsoever) grows with the observation year. Low R&D breadth and high R&D expenditures per patent have a positive influence on zero-probability.

Although correlation should be low, for a more complex model including a parameter for time correlation see for instance Hausmann et al. (1984). Time dependency may also be modelled through Generalized Estimating Equations (GEE), see e.g. Hardin and Hilbe (2003) and will be the subject of further research.

Appendix

Hessian matrix and Fisher information

The Hessian matrix $\mathcal{H}_n(\boldsymbol{\delta})$ in the ZIGP regression may be partitioned as

$$\mathcal{H}_n(\boldsymbol{\delta}) = \begin{pmatrix} \frac{\partial l_n(\boldsymbol{\delta})}{\partial \boldsymbol{\beta}\boldsymbol{\beta}^t} & \frac{\partial l_n(\boldsymbol{\delta})}{\partial \boldsymbol{\beta}\boldsymbol{\alpha}^t} & \frac{\partial l_n(\boldsymbol{\delta})}{\partial \boldsymbol{\beta}\boldsymbol{\gamma}^t} \\ \frac{\partial l_n(\boldsymbol{\delta})}{\partial \boldsymbol{\alpha}\boldsymbol{\beta}^t} & \frac{\partial l_n(\boldsymbol{\delta})}{\partial \boldsymbol{\alpha}\boldsymbol{\alpha}^t} & \frac{\partial l_n(\boldsymbol{\delta})}{\partial \boldsymbol{\alpha}\boldsymbol{\gamma}^t} \\ \frac{\partial l_n(\boldsymbol{\delta})}{\partial \boldsymbol{\gamma}\boldsymbol{\beta}^t} & \frac{\partial l_n(\boldsymbol{\delta})}{\partial \boldsymbol{\gamma}\boldsymbol{\alpha}^t} & \frac{\partial l_n(\boldsymbol{\delta})}{\partial \boldsymbol{\gamma}\boldsymbol{\gamma}^t} \end{pmatrix}, \tag{7.1}$$

where $\frac{\partial l_n(\boldsymbol{\delta})}{\partial \boldsymbol{\beta}\boldsymbol{\beta}^t} \in \mathbb{R}^{p\times p}$, $\frac{\partial l_n(\boldsymbol{\delta})}{\partial \boldsymbol{\beta}\boldsymbol{\alpha}^t} \in \mathbb{R}^{p\times r}$, $\frac{\partial l_n(\boldsymbol{\delta})}{\partial \boldsymbol{\beta}\boldsymbol{\gamma}^t} \in \mathbb{R}^{p\times q}$, $\frac{\partial l_n(\boldsymbol{\delta})}{\partial \boldsymbol{\alpha}\boldsymbol{\alpha}^t} \in \mathbb{R}^{r\times r}$, $\frac{\partial l_n(\boldsymbol{\delta})}{\partial \boldsymbol{\alpha}\boldsymbol{\gamma}^t} \in \mathbb{R}^{r\times q}$ and $\frac{\partial l_n(\boldsymbol{\delta})}{\partial \boldsymbol{\gamma}\boldsymbol{\gamma}^t} \in \mathbb{R}^{q\times q}$. Entries $h_{lm}(\boldsymbol{\delta})$'s of $\mathcal{H}_n(\boldsymbol{\delta})$ can be straightforwardly computed. For instance, entries of the matrix $\frac{\partial l_n(\boldsymbol{\delta})}{\partial \boldsymbol{\beta}\boldsymbol{\beta}^t}$ are given by

$$\begin{aligned} h_{lm}(\boldsymbol{\delta}) \;:=\; & \frac{\partial l_n(\boldsymbol{\delta})}{\partial \boldsymbol{\beta}_l \boldsymbol{\beta}_m} = -\sum_{i=1}^{n} \mathbb{1}_{\{y_i=0\}} x_{il} x_{im} \mu_i(\boldsymbol{\beta}) \\ & \times \frac{-P_i^0(\boldsymbol{\delta})^2/\varphi_i(\boldsymbol{\alpha}) \;+ \big(\mu_i(\boldsymbol{\beta}) - \varphi_i(\boldsymbol{\alpha})\big)/\varphi_i(\boldsymbol{\alpha})^2 P_i^0(\boldsymbol{\delta}) k_i(\boldsymbol{\gamma})}{(k_i(\boldsymbol{\gamma}) + P_i^0(\boldsymbol{\delta}))^2} \\ & - \sum_{i=1}^{n} \mathbb{1}_{\{y_i>0\}} x_{il} x_{im} \mu_i(\boldsymbol{\beta}) \left[\frac{(\varphi_i(\boldsymbol{\alpha})-1)(y_i-1)y_i}{(\mu_i(\boldsymbol{\beta}) + (\varphi_i(\boldsymbol{\alpha})-1)y_i)^2} - \frac{1}{\varphi_i(\boldsymbol{\alpha})} \right] \end{aligned} \tag{7.2}$$

for $l, m = 0, \ldots, p$.

Now set $\mathbf{H}_n(\boldsymbol{\delta}) := -\mathcal{H}_n(\boldsymbol{\delta})$. It is well known (see for example Mardia et al. (1979), p.98) that under mild general regularity assumptions, which are satisfied here, the Fisher information

matrix $\mathbf{F}_n(\boldsymbol{\delta})$ is equal to $E_\delta \mathbf{H}_n(\boldsymbol{\delta})$. Thus, entries of $\mathbf{F}_n(\boldsymbol{\delta})$ are given by

$$
\begin{aligned}
f_{l,m}(\boldsymbol{\delta}) &= f_{m,l}(\boldsymbol{\delta}) = -\sum_{i=1}^{n} x_{il}x_{im}\mu_i(\boldsymbol{\beta}) \\
&\quad \times\Bigg(\frac{-P_i^0(\boldsymbol{\delta})^2/\varphi_i(\boldsymbol{\alpha}) \ + \big(\mu_i(\boldsymbol{\beta})-\varphi_i(\boldsymbol{\alpha})\big)/\varphi_i(\boldsymbol{\alpha})^2 P_i^0(\boldsymbol{\delta})k_i(\boldsymbol{\gamma})}{(k_i(\boldsymbol{\gamma})+P_i^0(\boldsymbol{\delta}))(1+k_i(\boldsymbol{\gamma}))} \\
&\quad + \frac{b_i(\boldsymbol{\alpha})\mu_i(\boldsymbol{\beta})}{\varphi_i(\boldsymbol{\alpha})^2(\mu_i(\boldsymbol{\beta})-2+2\varphi_i(\boldsymbol{\alpha}))(1+k_i(\boldsymbol{\gamma}))} - \frac{1-P_i^0(\boldsymbol{\delta})}{\varphi_i(\boldsymbol{\alpha})(1+k_i(\boldsymbol{\gamma}))}\Bigg), \\
&\quad \text{for } l,m = 0,\ldots,p; \qquad (7.3)
\end{aligned}
$$

$$
\begin{aligned}
f_{l,p+1+m}(\boldsymbol{\delta}) &= f_{p+1+m,l}(\boldsymbol{\delta}) = -\sum_{i=1}^{n} x_{il}w_{im}\mu_i(\boldsymbol{\beta})b_i(\boldsymbol{\alpha}) \\
&\quad \times \Bigg(\frac{-P_i^0(\boldsymbol{\delta})/\varphi_i(\boldsymbol{\alpha})^3\mu_i(\boldsymbol{\beta})k_i(\boldsymbol{\gamma}) + P_i^0(\boldsymbol{\delta})/\varphi_i(\boldsymbol{\alpha})^2(k_i(\boldsymbol{\gamma})+P_i^0(\boldsymbol{\delta}))}{(k_i(\boldsymbol{\gamma})+P_i^0(\boldsymbol{\delta}))(1+k_i(\boldsymbol{\gamma}))} \\
&\quad - \frac{\mu_i(\boldsymbol{\beta})}{\varphi_i(\boldsymbol{\alpha})^2(\mu_i(\boldsymbol{\beta})-2+2\varphi_i(\boldsymbol{\alpha}))(1+k_i(\boldsymbol{\gamma}))} + \frac{1-P_i^0(\boldsymbol{\delta})}{\varphi_i(\boldsymbol{\alpha})^2(1+k_i(\boldsymbol{\gamma}))}\Bigg), \\
&\quad \text{for } l = 0,\ldots,p,\ m = 0,\ldots,r; \qquad (7.4)
\end{aligned}
$$

$$
\begin{aligned}
f_{l,p+2+m}(\boldsymbol{\delta}) &= f_{p+2+m,l}(\boldsymbol{\delta}) = -\sum_{i=1}^{n} x_{il}z_{im}\frac{P_i^0(\boldsymbol{\delta})/\varphi_i(\boldsymbol{\alpha})\mu_i(\boldsymbol{\beta})k_i(\boldsymbol{\gamma})}{(k_i(\boldsymbol{\gamma})+P_i^0(\boldsymbol{\delta}))(1+k_i(\boldsymbol{\gamma}))}, \\
&\quad \text{for } l = 0,\ldots,p,\ m = 0,\ldots,q; \qquad (7.5)
\end{aligned}
$$

$$
\begin{aligned}
f_{p+1+l,p+1+m}(\boldsymbol{\delta}) &= f_{p+1+m,p+1+l}(\boldsymbol{\delta}) = -\sum_{i=1}^{n} b_i(\boldsymbol{\alpha})\ w_{il}\ w_{im} \\
&\quad \times\Bigg(\frac{P_i^0(\boldsymbol{\delta})\mu_i(\boldsymbol{\beta})}{(k_i(\boldsymbol{\gamma})+P_i^0(\boldsymbol{\delta}))(1+k_i(\boldsymbol{\gamma}))}\left[\frac{\mu_i(\boldsymbol{\beta})b_i(\boldsymbol{\alpha})k_i(\boldsymbol{\gamma})}{\varphi_i(\boldsymbol{\alpha})^4}\right. \\
&\quad \left.+\left(\frac{1}{\varphi_i(\boldsymbol{\alpha})^2}-2\frac{b_i(\boldsymbol{\alpha})}{\varphi_i(\boldsymbol{\alpha})^3}\right)\big(k_i(\boldsymbol{\gamma})+P_i^0(\boldsymbol{\delta})\big)\right] \\
&\quad + \frac{\mu_i(\boldsymbol{\beta})^2}{\varphi_i(\boldsymbol{\alpha})^2(\mu_i(\boldsymbol{\beta})-2+2\varphi_i(\boldsymbol{\alpha}))(1+k_i(\boldsymbol{\gamma}))} + \frac{\mu_i(\boldsymbol{\beta})}{1+k_i(\boldsymbol{\gamma})} \\
&\quad \times\left[\frac{-2}{\varphi_i(\boldsymbol{\alpha})^2} + \frac{\varphi_i(\boldsymbol{\alpha})-P_i^0(\boldsymbol{\delta})(1-b_i(\boldsymbol{\alpha}))}{\varphi_i(\boldsymbol{\alpha})^3}\right]\Bigg), \\
&\quad \text{for } l,m = 0,\ldots,r; \qquad (7.6)
\end{aligned}
$$

$$
\begin{aligned}
f_{p+1+l,p+2+m}(\boldsymbol{\delta}) &= f_{p+2+m,p+1+l}(\boldsymbol{\delta}) \\
&= -\sum_{i=1}^{n} w_{il}\ z_{im}\ b_i(\boldsymbol{\alpha})\frac{-P_i^0(\boldsymbol{\delta})/\varphi_i(\boldsymbol{\alpha})^2\mu_i(\boldsymbol{\beta})k_i(\boldsymbol{\gamma})}{(k_i(\boldsymbol{\gamma})+P_i^0(\boldsymbol{\delta}))(1+k_i(\boldsymbol{\gamma}))}, \\
&\quad \text{for } l = 0,\ldots,r,\ m = 0,\ldots,q, \text{and} \qquad (7.7)
\end{aligned}
$$

$$
\begin{aligned}
f_{p+2+l,p+2+m}(\boldsymbol{\delta}) &= f_{p+2+m,p+2+l}(\boldsymbol{\delta}) = -\sum_{i=1}^{n} z_{il}z_{im}k_i(\boldsymbol{\gamma}) \\
&\quad \times\Bigg(\frac{P_i^0(\boldsymbol{\delta})}{(k_i(\boldsymbol{\gamma})+P_i^0(\boldsymbol{\delta}))(1+k_i(\boldsymbol{\gamma}))} - \frac{1}{(1+k_i(\boldsymbol{\gamma}))^2}\Bigg), \\
&\quad \text{for } l,m = 0,\ldots,q. \qquad (7.8)
\end{aligned}
$$

Proof of Theorem 1

The proof follows Fahrmeir and Kaufmann (1985) and Czado and Min (2005). A detailed German version can be found in Erhardt (2006, Sec. 2.3). First, we will derive a useful property.

LEMMA 1 (Convergence and Continuity). Given (A1)-(A3), for all $\varepsilon > 0$,

$$\max_{\boldsymbol{\delta} \in N_n(\varepsilon)} \| \boldsymbol{V}_n(\boldsymbol{\delta}) - \boldsymbol{I}_{p+r+q+3} \| \xrightarrow{P} 0, \tag{7.9}$$

where $\boldsymbol{V}_n(\boldsymbol{\delta}) := \boldsymbol{F}_n^{-1/2} \, \boldsymbol{H}_n(\boldsymbol{\delta}) \, \boldsymbol{F}_n^{-t/2}$ is the normed information matrix.

PROOF. [Sketch] The proof can be found in Czado and Min (2005, Lemma 6), where entry $h_{p,p}(\boldsymbol{\delta})$ of the Hessian matrix has to be replaced according to (7.2). In addition, we use the compactness not only of $\boldsymbol{x}_n$ but also of $\boldsymbol{w}_n$ and $\boldsymbol{z}_n$ when finding upper boundaries. □

For proving Theorem 1 (i), we first derive an equivalent representation of the asymptotic existence. Fahrmeir and Kaufmann (1985) state that (**AE**) is equivalent to (**AE***): for every $\tilde{\eta} > 0$ there is a $\varepsilon > 0$ and n_1 so that

$$(\mathbf{AE}^*) \; P\left\{ l_n(\boldsymbol{\delta}) - l_n(\boldsymbol{\delta}_0) < 0 \text{ for all } \boldsymbol{\delta} \in \partial N_n(\varepsilon) \right\} \geq 1 - \tilde{\eta}. \tag{7.10}$$

Instead of proving (**AE**), we can now show (**AE***).

PROOF. [(**AE**) $\Leftrightarrow$ (**AE***) *(Sketch)*] Let $\boldsymbol{\lambda} = \frac{1}{\varepsilon} \mathbf{F}_n^{t/2} (\boldsymbol{\delta} - \boldsymbol{\delta}_0)$, then the Taylor expansion of the log likelihood becomes

$$l_n(\boldsymbol{\delta}) - l_n(\boldsymbol{\delta}_0) = \varepsilon \boldsymbol{\lambda}^t \mathbf{F}_n^{-1/2} \mathbf{s}_n - \frac{\varepsilon^2}{2} \boldsymbol{\lambda}^t \mathbf{F}_n^{-1/2} \mathbf{H}_n(\tilde{\boldsymbol{\delta}}) \mathbf{F}_n^{-t/2} \boldsymbol{\lambda}, \tag{7.11}$$

where $\tilde{\boldsymbol{\delta}}$ lies between $\boldsymbol{\delta}$ and $\boldsymbol{\delta}_0$. Note that for $\boldsymbol{\delta} \in \partial N_n(\varepsilon)$ we have $\boldsymbol{\lambda}^t \boldsymbol{\lambda} = \frac{\| \mathbf{F}_n^{t/2} (\hat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_0) \|^2}{\varepsilon^2} = 1$. Using $\| \mathbf{F}_n^{-1/2} \mathbf{s}_n \| := \sup_{\|\lambda\|=1} \boldsymbol{\lambda}^t \mathbf{F}_n^{-1/2} \mathbf{s}_n$, we recognize that it is sufficient to show that for any $\tilde{\eta} = \eta_1 + \eta_2 > 0$:

$$P\left\{ \bigcap_{\lambda : \|\lambda\|=1} \left\{ \omega : \varepsilon \| \mathbf{F}_n^{-1/2} \mathbf{s}_n \| < \frac{\varepsilon^2}{2} \boldsymbol{\lambda}^t \mathbf{F}_n^{-1/2} \mathbf{H}_n(\tilde{\boldsymbol{\delta}}) \mathbf{F}_n^{-t/2} \boldsymbol{\lambda} \right\} \right\} \geq 1 - \tilde{\eta} \tag{7.12}$$

or equivalently

$$\Delta_1(n) := P\left\{ \bigcup_{\lambda : \|\lambda\|=1} \left\{ \omega : \varepsilon \| \mathbf{F}_n^{-1/2} \mathbf{s}_n \| \geq \frac{\varepsilon^2}{2} \boldsymbol{\lambda}^t \mathbf{F}_n^{-1/2} \mathbf{H}_n(\tilde{\boldsymbol{\delta}}) \mathbf{F}_n^{-t/2} \boldsymbol{\lambda} \right\} \right\} < \tilde{\eta}. \tag{7.13}$$

For further details, see Erhardt (2006, p. 50). □

In order to prove weak consistency in Theorem 1 (ii), we need **(A1)** and (**AE***).

PROOF. [Theorem 1 (ii)] Assumption **(A1)** implies that the neighborhoods $N_n(\varepsilon)$ shrink to $\boldsymbol{\delta}_0$ and $\hat{\boldsymbol{\delta}} \in N_n(\varepsilon)$. Hence, (**AE***) implies (ii). In particular, $N_n(\varepsilon)$ shrinks to $\boldsymbol{\delta}_0$ because with $n \to \infty$, $\lambda_{\min}(\mathbf{F}_n)$ is growing according to **(A1)** ($\frac{n}{c_1} \leq \lambda_{\min}(\mathbf{F}_n)$). Since $\lambda_{\min}(\mathbf{F}_n) \cdot \| \hat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_0 \|^2 \leq (\hat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_0)^t \mathbf{F}_n (\hat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_0) = \| \mathbf{F}_n^{t/2} (\hat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_0) \|^2 < \varepsilon^2$ a.s., we get $\hat{\boldsymbol{\delta}} \xrightarrow{p} \boldsymbol{\delta}_0$ a.s. from the last inequation. □

For the proof of the asymptotic normality of the solutions of the score equations we need the asymptotic normality of the score equations themselves.

LEMMA 2 (Asymptotic normality of the score equations). Given $(A1)$ - $(A3)$, $\boldsymbol{F}_n^{-1/2}\boldsymbol{s}_n \stackrel{D}{\Rightarrow} N_{p+r+q+3}(\mathbf{0}, \boldsymbol{I}_{p+r+q+3})$, for $n \to \infty$. Here, $N_{p+r+q+3}(\mathbf{0}, \boldsymbol{I}_{p+r+q+3})$ is the $(p+r+q+3)$ - dimensional normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{I}_{p+r+q+3}$.

PROOF. [Sketch] The proof can be found in Czado and Min (2005, Lemma 5), where the dimension of the identity matrix has to be replaced by $(p+r+q+3)$. Further, when showing the boundedness of $\max_{i=1,\ldots,n} E|s_{r,i}|^3$ for the Lyapunov condition, we now have individual values φ_i and k_i for each observation. Here, we use the compactness of the regressors $\boldsymbol{w}_n$ and $\boldsymbol{z}_n$ to find an upper boundary. □

The proof of the asymptotic normality of $\hat{\boldsymbol{\delta}}_n$ in Theorem 1 (iii) follows the line of arguments in Fahrmeir and Kaufmann (1985, Theorem 3) and uses Lemmas 1 and 2. For details, see Erhardt (2006, pp. 56f).

Acknowledgement

This work was supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 386 *Statistical Analysis of Discrete Structures*. We would like to thank *INNO-tec* for providing the data and especially Stefan Wagner for the helpful discussion of our results.

References

Abraham, K. G. and S. K. Taylor (1996). Firms' Use of Outside Contractors: Theory and Evidence. *Journal of Labor Economics 14*(3), 394–424.

Amit, R. and P. J. H. Schoemaker (1993). Strategic Assets and Organizational Rent. *Strategic Management Journal 14*(1), 33–46.

Bae, S., F. Famoye, J. T. Wulu, A. A. Bartolucci, and K. P. Singh (2005). A rich family of generalized Poisson regression models. *Math. Comput. Simulation 69*(1-2), 4–11.

Clarke, K. A. (2003). A Simple Distribution-Free Test for Nonnested Hypotheses. Department of Political Science, University of Rochester, Harkness Hall.

Consul, P. C. (1989). *Generalized Poisson distributions*, Volume 99 of *Statistics: Textbooks and Monographs*. New York: Marcel Dekker Inc. Properties and applications.

Consul, P. C. and F. Famoye (1992). Generalized Poisson regression model. *Comm. Statist. Theory Methods 21*(1), 89–109.

Consul, P. C. and G. C. Jain (1970). On the generalization of Poisson distribution. *Ann. Math. Statist. 41*, 1387.

Czado, C. and A. Min (2005). Consistency and asymptotic normality of the maximum likelihood estimator in a zero-inflated generalized Poisson regression. (*http://www.stat.uni-muenchen.de/sfb386/papers/dsp/paper423.ps*).

Erhardt, V. (2006). Verallgemeinerte Poisson und Nullenüberschuss- Regressionsmodelle mit regressiertem Erwartungswert, Dispersions- und Nullenüberschuss-Parameter und eine Anwendung zur Patentmodellierung. Diplomarbeit, Centre of Mathematical Sciences, Munich University of Technology, *http://www-m4.ma.tum.de/Diplarb/*, Garching bei München.

Fahrmeir, L. and H. Kaufmann (1985). "Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in Generalized Linear Models". *Ann. Statist. 13*(1), 342–368.

Famoye, F. (1993). Restricted generalized Poisson regression model. *Comm. Statist. Theory Methods 22*(5), 1335–1354.

Famoye, F. and K. P. Singh (2003). On inflated generalized Poisson regression models. *Adv. Appl. Stat. 3*(2), 145–158.

Famoye, F. and K. P. Singh (2006). Zero-inflated generalized Poisson model with an application to domestic violence data. *Journal of Data Science 4*(1), 117–130.

Grossman, S. J. and O. D. Hart (1986). The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration. *The Journal of Political Economy 94*(4), 691–719.

Gupta, P. L., R. C. Gupta, and R. C. Tripathi (2004). Score test for zero inflated generalized Poisson regression model. *Comm. Statist. Theory Methods 33*(1), 47–64.

Hardin, J. W. and J. M. Hilbe (2003). *Generalized estimating equations.* Boca Raton, FL: Chapman and Hall/CRC. xiii, 222 p.

Hausmann, J., B. H. Hall, and Z. Griliches (1984). Econometric Models for Count Data with an Application to the Patents - R&D Relationsship. *Econometrica Vol. 52, No. 4*, 909–938.

Heiberger, R. M. and B. Holland (2004). *Statistical analysis and data display. An intermediate course with examples in S-Plus, R, and SAS.* Springer Texts in Statistics. New York, NY: Springer. xxiv, 729 p. EUR 79.95/net; sFr 135.50.

Joe, H. and R. Zhu (2005). Generalized Poisson distribution: the property of mixture of Poisson and comparison with negative binomial distribution. *Biom. J. 47*(2), 219–229.

Kullback, S. and R. Leibler (1951). On Information and Sufficiency. *Annals of Mathematical Statistics 2*, 79–86.

Lambert, D. (1992). Zero-inflated poisson regression, with an application to defects in manufacturing. *Technometrics 34*(1), 1–14.

Mardia, K. V., J. T. Kent, and J. M. Bibby (1979). *Multivariate analysis.* London: Academic Press [Harcourt Brace Jovanovich Publishers]. Probability and Mathematical Statistics: A Series of Monographs and Textbooks.

Mullahy, J. (1986). Specification and testing of some modified count data models. *J. Econometrics 33*(3), 341–365.

Sako, M. (2005). Outsourcing and Offshoring: Key Trends and Issues. *Said Business School, Oxford* (*www.sbs.ox.ac.uk/NR/rdonlyres/FC74DDA9-8480-4EE1-A324-2176269F1753/1308/EMFOutsourcingNov06.doc*).

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and nonnested hypotheses. *Econometrica 57*(2), 307–333.

Wagner, S. (2005). An Empirical Analysis of Make-or-Buy Decisions in Patenting. *Munich School of Management, LMU München.*