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], **Aleksey MIN** and **Stefan WAGNER**

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 models considered by several authors, 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 using our *R* package "ZIGP" available on CRAN. 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 is carried out. For the given data, our extended ZIGP regression model will prove to be superior over GP and ZIP models and even over 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: +49-89-289-17437
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 (for score tests, see e.g. Famoye and Singh (2003), Gupta et al. (2004), Bae et al. (2005) and Famoye and Singh (2006), for mixing property see Joe and Zhu (2005) and for a Baeysian approach see Gschlößl and Czado (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. Here we allow for regression not only on the mean but on the overdispersion and zero-inflation parameters. This allows us to model individual overdispersion and zero-inflation by groups (e.g. by gender of an insured person). 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. Here we would like to mention a paper by Ghosh et al. (2006) which considers an alternative flexible class of zero-inflated regression models associated with power series distributions. The authors present a Bayesian approach for the above class of regressions models and also allow for regression on zero-inflation level.

Since the ZIGP distribution does not belong to the exponential family, the regression model is no generalized linear model (GLM). Further the existing asymptotic theory for GLM and its extensions considered so far is not applicable for these models. Therefore we develop the appropriate asymptotic theory and investigate the small sample properties of the maximum likelihood estimates.

A comparison of nine models extending the regular Poisson GLM by dispersion and zero-inflation parameters will be facilitated. Since these models might be nonnested, partial deviance, likelihood ratio tests or the AIC criterion are not applicable. Instead we use a test proposed by Vuong (1989) for nonnested models.

The usefulness of our extensions will be demonstrated in an application dealing with patent outsourcing. We investigate make-or-buy decision drivers for the patent filing process. This data has already been examined by Wagner (2006), 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 focus on firm specific attributes such as the R&D spending per employee or patent. 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 data, we develop exploratory data analysis tools for overdispersion and zero-inflation. We see that heterogeneity is high and strongly depends on a company's industry. We illustrate step-by-step that all enhancements are useful for this data. A graphical model interpretation will be performed comparing our results with those obtained by Wagner (2006).

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 and applied to our data afterwards. Section 6 investigates covariate effects on the mean response and the overdispersion factor as well as interprets the results. We conclude with a summary and discussion. The Fisher information matrix and a sketch of the proof of Theorem 1 are given in the Appendix.

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 parametrization (see e.g. Consul and Famoye (1992)). For $X \sim GP(\mu, \varphi)$ we have $Var(X) > (=, <)\ E(X) \Leftrightarrow \varphi > (=, <)\ 1$. This allows for modelling over- and underdispersion. However in the case of underdispersion ($\varphi < 1$), the support of the distribution depends on μ and φ, which is difficult to enforce when μ and φ need to be estimated. Further, in the regression context this fact implies that the support of a link function for φ depends on μ. Therefore, we restrict to the case of overdispersion. 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 $X \sim ZIGP(\mu, \varphi, \omega)$. Its probability density function (pdf) is given by

$$\begin{aligned} P(Y = y|\ \mu, \varphi, \omega) &= \mathbb{1}_{\{y=0\}} \left[\omega + (1-\omega)e^{-\frac{\mu}{\varphi}}\right] \\ &\quad + \mathbb{1}_{\{y>0\}} \left[(1-\omega)\frac{\mu(\mu+(\varphi-1)y)^{y-1}}{y!}\varphi^{-y}e^{-\frac{1}{\varphi}(\mu+(\varphi-1)y)}\right], \end{aligned} \tag{2.1}$$

where $\varphi > \max(\frac{1}{2}, 1 - \frac{\mu}{m})$ and m is the largest natural number with $\mu + m(\varphi - 1) > 0$, if $\varphi < 1$. Mean and variance of the ith observation $Y_i \sim ZIGP(\mu_i, \varphi, \omega)$ are given by

$$E(Y_i|\boldsymbol{X} = \boldsymbol{x_i}) = (1-\omega)\mu_i \tag{2.2}$$

$$\text{and } \sigma_i^2 := Var(Y_i|\boldsymbol{X} = \boldsymbol{x_i}) = E(Y_i|\boldsymbol{X} = \boldsymbol{x_i})\left(\varphi^2 + \mu_i\omega\right). \tag{2.3}$$

One of the main benefits of considering a regression model based on the ZIGP distribution is that it allows for two sources of overdispersion, one by mixing (see Joe and Zhu (2005)) and one by zero-inflation. 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 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:*
   $\{Y_i,\ 1 \le i \le n\}$ 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:

$$\begin{aligned}
&\text{(i) } \textit{Mean level} \\
E(Y_i \mid \boldsymbol{\beta}) = \mu_i(\boldsymbol{\beta}) &:= E_i e^{\boldsymbol{x}_i^t \boldsymbol{\beta}} = e^{\boldsymbol{x}_i^t \boldsymbol{\beta} + log(E_i)} > 0 \\
\Leftrightarrow \eta_i^{\mu}(\boldsymbol{\beta}) &= \log(\mu_i(\boldsymbol{\beta})) - \log(E_i) \text{ (log link)}, \qquad (2.4) \\
&\text{(ii) } \textit{Overdispersion level} \\
\varphi_i(\boldsymbol{\alpha}) &:= 1 + e^{\boldsymbol{w}_i^t \boldsymbol{\alpha}} > 1 \\
\Leftrightarrow \eta_i^{\varphi}(\boldsymbol{\alpha}) &= \log(\varphi_i(\boldsymbol{\alpha}) - 1)) \text{ (modified log link)}, \qquad (2.5) \\
&\text{(iii) } \textit{Zero-inflation level} \\
\omega_i(\boldsymbol{\gamma}) &:= \frac{e^{\boldsymbol{z}_i^t \boldsymbol{\gamma}}}{1 + e^{\boldsymbol{z}_i^t \boldsymbol{\gamma}}} \in (0, 1) \\
\Leftrightarrow \eta_i^{\omega}(\boldsymbol{\gamma}) &= \log\left(\frac{\omega_i(\boldsymbol{\gamma})}{1 - \omega_i(\boldsymbol{\gamma})}\right) \text{ (logit link)}. \qquad (2.6)
\end{aligned}$$

It should be noted that the link function for the zero-inflation level does not allow for no zero-inflation. A modification which would allow this, would require complex restrictions on the range of the mean values, which is not desirable. Since we will conduct nonnested model comparisons, this model restriction is not severe. The unknown parameters are collected in $\boldsymbol{\delta}$, i.e. $\boldsymbol{\delta} := (\boldsymbol{\beta}^t, \boldsymbol{\alpha}^t, \boldsymbol{\gamma}^t)^t$, and its maximum likelihood (ML) estimate will be denoted by $\hat{\boldsymbol{\delta}}$. The following abbreviations for $i = 1, \ldots, n$ will be used:

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

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

$$\begin{aligned}
l(\boldsymbol{\delta}) = &\sum_{i=1}^{n} \mathbb{1}_{\{y_i = 0\}} \left[\log\left(e^{\boldsymbol{z}_i^t \boldsymbol{\gamma}} + \exp\left(-\frac{E_i \cdot e^{\boldsymbol{x}_i^t \boldsymbol{\beta}}}{1 + e^{\boldsymbol{w}_i^t \boldsymbol{\alpha}}}\right)\right) - \log(1 + e^{\boldsymbol{z}_i^t \boldsymbol{\gamma}})\right] \\
&+ \mathbb{1}_{\{y_i > 0\}} \Big[-\log(1 + e^{\boldsymbol{z}_i^t \boldsymbol{\gamma}}) + \log(E_i) + \boldsymbol{x}_i^t \boldsymbol{\beta} - \log(y_i!) - y_i \log(1 + e^{\boldsymbol{w}_i^t \boldsymbol{\alpha}}) \\
&+ (y_i - 1) \log(E_i \; e^{\boldsymbol{x}_i^t \boldsymbol{\beta}} + e^{\boldsymbol{w}_i^t \boldsymbol{\alpha}} y_i) - \frac{E_i \; e^{\boldsymbol{x}_i^t \boldsymbol{\beta}} + e^{\boldsymbol{w}_i^t \boldsymbol{\alpha}} y_i}{1 + e^{\boldsymbol{w}_i^t \boldsymbol{\alpha}}}\Big]. \qquad (2.7)
\end{aligned}$$

Second, the score vector has the following representation:

$$\boldsymbol{s}_n(\boldsymbol{\delta}) = \left(s_0(\boldsymbol{\delta}), \ldots, s_p(\boldsymbol{\delta}), \ldots, s_{p+r+1}(\boldsymbol{\delta}), \ldots, s_{p+r+q+2}(\boldsymbol{\delta})\right)^t, \qquad (2.8)$$

where

$$
\begin{aligned}
s_m(\boldsymbol{\delta}) &:= \frac{\partial}{\partial \beta_m} l(\boldsymbol{\delta}) = \sum_{i=1}^{n} x_{im} \Bigg(\mathbb{1}_{\{y_i=0\}} \left[\frac{-P_i^0(\boldsymbol{\delta})\mu_i(\boldsymbol{\beta})/\varphi_i(\boldsymbol{\alpha})}{k_i(\boldsymbol{\gamma}) + P_i^0(\boldsymbol{\delta})} \right] \\
&\quad + \mathbb{1}_{\{y_i>0\}} \left[1 + \frac{(y_i-1)\mu_i(\boldsymbol{\beta})}{\mu_i(\boldsymbol{\beta}) + b_i(\boldsymbol{\alpha})y_i} - \frac{1}{\varphi_i(\boldsymbol{\alpha})}\mu_i(\boldsymbol{\beta}) \right] \Bigg), m = 0, \ldots, p, \qquad (2.9) \\
s_{p+1+m}(\boldsymbol{\delta}) &:= \frac{\partial}{\partial \alpha_m} l(\boldsymbol{\delta}) = \sum_{i=1}^{n} w_{im} b_i(\boldsymbol{\alpha}) \Bigg(\mathbb{1}_{\{y_i=0\}} \left[\frac{P_i^0(\boldsymbol{\delta})\mu_i(\boldsymbol{\beta})/\varphi_i(\boldsymbol{\alpha})^2}{k_i(\boldsymbol{\gamma}) + P_i^0(\boldsymbol{\delta})} \right] \\
&\quad + \mathbb{1}_{\{y_i>0\}} \left[\frac{(y_i-1)y_i}{\mu_i(\boldsymbol{\beta}) + b_i(\boldsymbol{\alpha})y_i} - \frac{y_i}{\varphi_i(\boldsymbol{\alpha})} + \frac{\mu_i(\boldsymbol{\beta}) - y_i}{\varphi_i(\boldsymbol{\alpha})^2} \right] \Bigg), \\
&\quad m = 0, \ldots, r, \qquad (2.10) \\
s_{p+r+2+m}(\boldsymbol{\delta}) &:= \frac{\partial}{\partial \gamma_m} l(\boldsymbol{\delta}) = \sum_{i=1}^{n} z_{im} k_i(\boldsymbol{\gamma}) \Bigg(\mathbb{1}_{\{y_i=0\}} \left[\frac{1}{k_i(\boldsymbol{\gamma}) + P_i^0(\boldsymbol{\delta})} \right] - \frac{1}{1 + k_i(\boldsymbol{\gamma})} \Bigg), \\
&\quad m = 0, \ldots, q. \qquad (2.11)
\end{aligned}
$$

To compute $\hat{\boldsymbol{\delta}}$, we simultaneously solve the equations obtained by equating the score vector (2.8) to zero. The Fisher information $\boldsymbol{F}_n(\boldsymbol{\delta})$ is needed for the variance estimation of the ML estimates. It is calculated in the appendix. The R package "ZIGP" available on CRAN facilitates ML estimation of the $ZIGP(\mu_i, \varphi_i, \omega_i)$ models.

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

Since the ZIGP regression model is not a GLM, we need to prove existence, consistency and asymptotic normality of the ML estimate $\hat{\boldsymbol{\delta}}$. 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 $N_n(\varepsilon) := \{\boldsymbol{\delta} : \|\boldsymbol{F}_n^{t/2}(\boldsymbol{\delta}_0)(\boldsymbol{\delta} - \boldsymbol{\delta}_0)\| \leq \varepsilon\}$ 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 $\frac{n}{\lambda_{min}(\boldsymbol{F}_n)} \leq C_1 \ \forall\ n \geq 1$, where C_1 is a positive constant,

(A2) (Compact regressors) The sets $\{\boldsymbol{x}_n,\ n \geq 1\} \subset K_x \subset \mathbb{R}^{p+1}$, $\{\boldsymbol{w}_n,\ n \geq 1\} \subset K_w \subset \mathbb{R}^{r+1}$ and $\{\boldsymbol{z}_n,\ n \geq 1\} \subset K_z \subset \mathbb{R}^{q+1}$ are compact sets,

(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) $\boldsymbol{F}_n^{t/2}(\hat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_0) \overset{D}{\Rightarrow} N_p(\mathbf{0}, \boldsymbol{I}_{p+r+q+3})$, 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 distribution of the ML estimates to their normal approximation in a simulation study.

We generated sets of 1000 random vectors of length n. For each of the three parameters we used linear predictors consisting of an intercept and an additional regressor with equidistant values. Corresponding QQ plots of the standardized ML estimates are given in Figure 1. We see that the performance for the mean ($\boldsymbol{\beta}$) and zero-inflation ($\boldsymbol{\gamma}$) regression parameters is satisfactory even for a smaller sample size, while large sample sizes are needed for the overdispersion ($\boldsymbol{\alpha}$) regression parameters. Further parameter settings were investigated. In particular increasing the μ_i range results in a lower convergence speed, while increasing dispersion or zero-inflation resulted in lower convergence speed (for details see Erhardt (2006)).

4 Model Comparison

The tree in Figure 2 sketches an evolution of nine models starting from (1) Poisson regression to (9) $ZIGP(\mu_i, \varphi_i, \omega_i)$ regression described in Section 2. 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, the full pool of covariates chosen in an exploratory data analysis will be checked for significance in each model individually. We used sequential elimination of insignificant effects. As design matrices may thus be different, these models need not be nested. For nested model comparisons one can use the AIC (see e.g. Heiberger and Holland (2004, p. 572)), while for nonnested models we use a test proposed by Vuong (1989).

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, the full pool of covariates chosen in an exploratory data analysis will be checked for significance in each model individually. We used sequential elimination of insignificant effects. As design matrices may thus be different, these models need not be nested. For nested model comparisons one can use the AIC (see e.g. Heiberger and Holland (2004, p. 572)), while for nonnested models we use a test proposed by Vuong (1989).

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

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

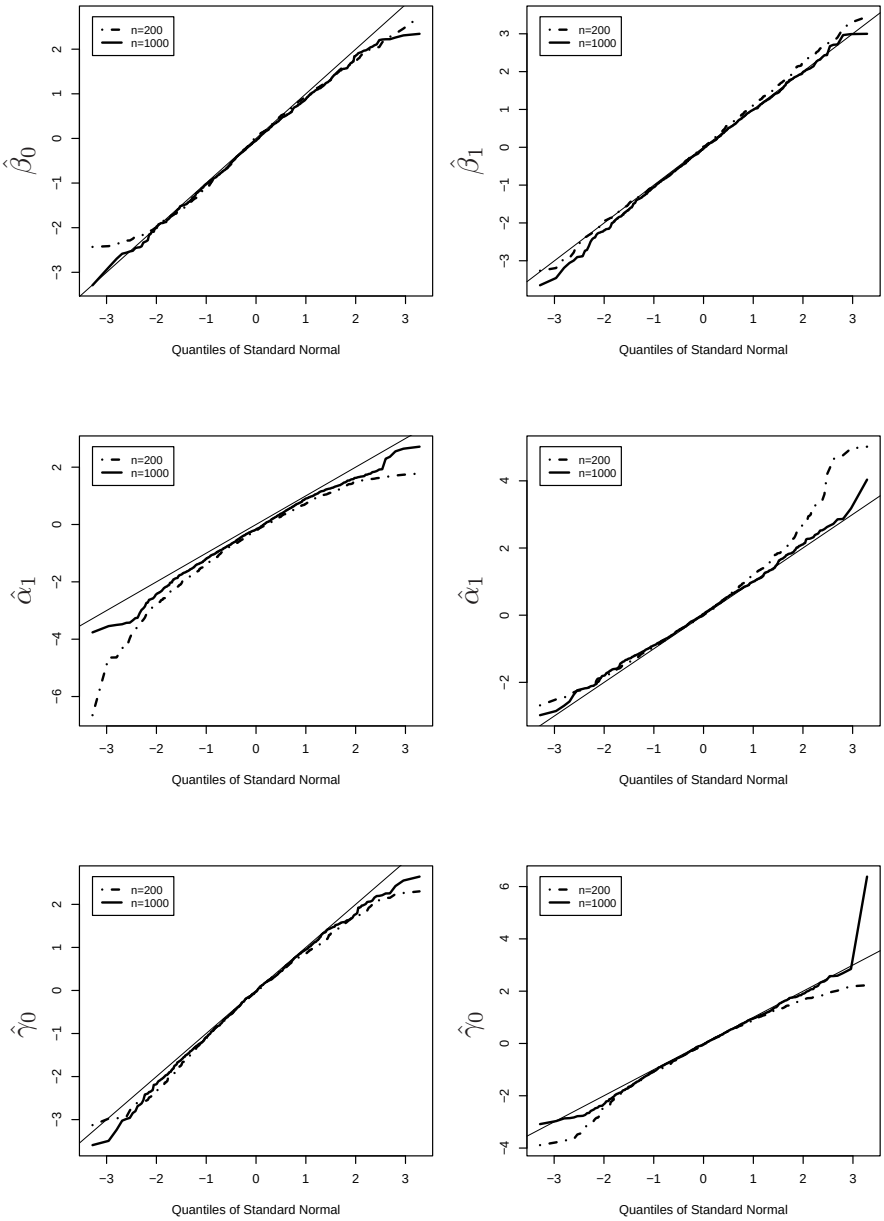

Figure 1: QQ-Plots of centered and normed ML estimates based on $N = 1000$ data sets

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

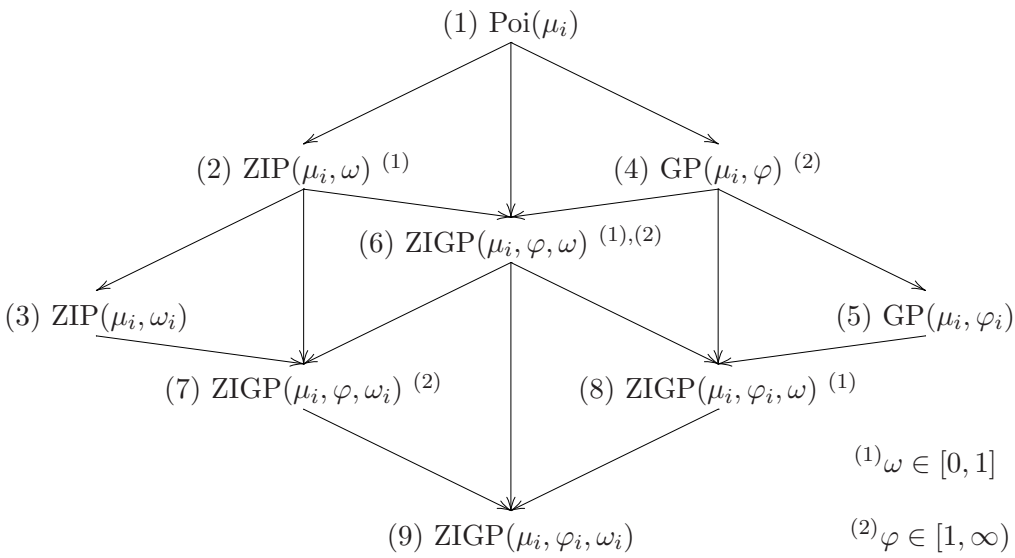

Figure 2: Overview of model enhancements of the Poisson GLM

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. Therefore, Vuong considers the test statistic ν defined below 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

The data consists of patent information of the European Patent Office. It has been examined and amended with corporate information by Wagner (2006). 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 (2006) examines decision drivers using negative binomial panel regression. The survey considers 107 European companies ($i = 1, \ldots, 107$) over a eight years (1993 to 2000). Since data for each company is aggregated over one year, we expect a company's correlation over years to be small in comparison to a trend effect. Figure 3 shows boxplots of the estimated residual autocorrelations for time lags of 1 and 2 years. We see evidence that the autocorrelation is indeed low. Residuals have been calculated according to our best-fit model (9) $ZIGP(\mu_i, \varphi_i, \omega_i)$, which will be introduced in Table 3.

Table 1 gives an overview of all influential variables. For more details see Wagner (2006, pp. 119-121). 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

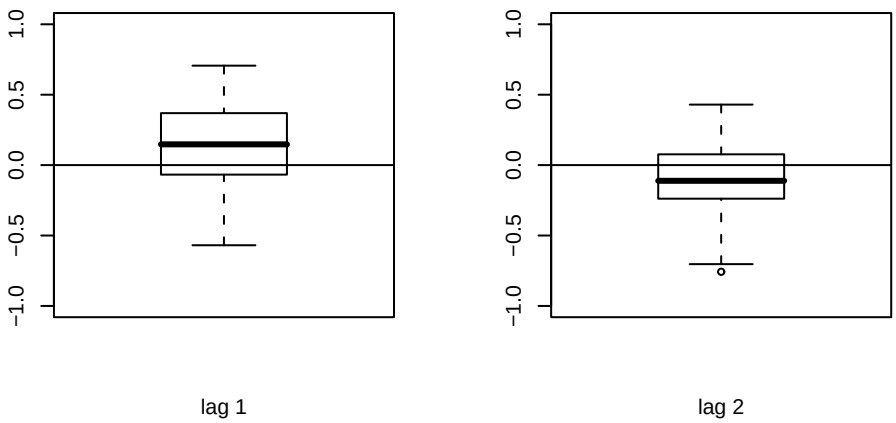

Figure 3: Boxplots of estimated residual autocorrelations for time lags of 1 and 2 years based on Model (9) given in Table 3

group means of log observations standardized by their exposure ignoring possible zero-inflation. The four strongest two-dimensional interactions were **LN.COV** $*$ **BREADTH**, **CHEM.PHA** $*$ **LN.COV**, **CHEM.PHA** $*$ **SQRT.EMP** and **RDmiss** $*$ **CHEM.PHA**. 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 accounted for 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 (2.5) and ignore zero-inflation, i.e. we assume $Y_i \sim GP(\mu_i, \varphi_i)$. For each potential covariate, we consider classes consisting of categories for categorical covariates or scoring classes for metric covariates. For metric covariates we select five scoring classes formed by the 0%, 20%, 40%, ..., 100% quantiles. We have $\sigma_i^2 = \mu_i \cdot \varphi_i^2 \Leftrightarrow \varphi_i = \pm\sqrt{\sigma_i^2/\mu_i}$. 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, $j = 1, \ldots, J$ with n_j observations 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 $\hat{\sigma}_j := \frac{1}{n_j - 1}\sum_{i \in j}(Y_i - \hat{\mu}_j)^2$. 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. The logarithm in (5.1) is taken to detect classes with large overdispersion. We will select a covariate for the overdispersion level if the corresponding fractions $f_j(\hat{\mu}_j, \hat{\sigma}_j^2) \geq 2$ for some class j of this covariate. For example, the fractions separated by year and industry range from 4.8 to 443.

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

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

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

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

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. So, large 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 $\hat{logit}(0.2) = \log\left(\frac{0.2+0.5}{1-0.2+0.5}\right) = -0.62$. For

	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)

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 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, in Table 3 we get the following regression equations for each model class considered in Figure 2. Fixed parameters are counted in the AIC statistic if they are estimated (such as $\omega_i = \omega$ in (2) $ZIP(\mu_i, \omega)$) and are not counted if they are not estimated (such as $\varphi_i = 1$ in (2) $ZIP(\mu_i, \omega)$.) All covariates have been centered and standardized for numerical stability. From Table 3 we see that only three nested model comparisons can be conducted using the AIC; namely (2) vs. (3), (4) vs. (5) and (7) vs. (9). In all these comparisons the more complex model is the better fitting model. All other test decisions have to be based on the Vuong test. Another approach to select models within each model class is 'backward selection', i.e. to sequentially eliminate the covariate from the full model which minimizes the AIC the most (as long as the AIC shrinks). In case of (9) $ZIGP(\mu_i, \varphi_i, \omega_i)$ this leads to almost the same model as in Table 3. It's AIC of 6 526 is only two points below the suggested model whereas there are two additional effects **RDE.2** and **EMP.2023.11291** which are insignificant at the 5% level. Model comparison is carried out using the methods discussed in Section 4. Table 4 again lists models (1) through (9) in rows (I) and columns (II). The entries show Vuong test results for every combination of an (I) and a (II) model. We choose an α-level of 5%, i.e. $z_{1-\frac{\alpha}{2}} = 1.96$. 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: Comparing (1) $Poi(\mu_i)$ with model (2) $ZIP(\mu_i, \omega)$, the Vuong test prefers the latter model with a test statistic of $\nu = 4.2$ (see Table 4).

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**	$\varphi_i = 1\ \forall i$ (not estimated)	$\omega_i = 0\ \forall i$ (not estimated)	-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**	$\varphi_i = 1\ \forall i$ (not estimated)	$\omega_i = \omega\ \forall i$	-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**	$\varphi_i = 1\ \forall i$ (not estimated)	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**	$\varphi_i = \varphi\ \forall i$	$\omega_i = 0\ \forall i$ (not estimated)	-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**	$\omega_i = 0\ \forall i$ (not estimated)	-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 insignificant 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

Adding a dispersion parameter: Adding a dispersion parameter has a strong positive impact on model quality. Comparing (1) $Poi(\mu_i)$ and (4) $GP(\mu_i, \varphi)$, the Vuong statistic is $\nu = 10.8$. This again indicates 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 the nested models (2) $ZIP(\mu_i, \omega)$ and (3) $ZIP(\mu_i, \omega_i)$, the AIC falls from 19 183 to 19 110 (see Table 3). Further, $\nu = 4.32$, so Vuong prefers model (3) $ZIP(\mu_i, \omega_i)$, too. The p-value is < 0.01.

Regression on the dispersion parameter: Comparing model (4) $GP(\mu_i, \varphi)$ with (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) $ZIGP(\mu_i, \varphi_i, \omega_i)$ seems fit our data in terms of the Vuong test best. This model is preferred over all other models discussed (see the last row in Table 4). Applied to (1) $Poi(\mu_i)$ vs. (9) $ZIGP(\mu_i, \varphi_i, \omega_i)$, the Vuong test has the largest significance of all comparisons, 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.

First of all, we see that **RDmiss** has been dropped during the sequential elimination of insignificant variables, which is in line with Wagner (2006, p. 133). 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 empirical mode or mean, for categorical covariates we compare their categories. Exposure $\boldsymbol{E}$ will also be replaced by its empirical mode E^M.

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. For all outsourcing functions we have to separate between the three industry groups Chemical / Pharma, Electro / Telecommunication / Others and all remaining industries. Figure 4 shows the resulting outsourcing rates as functions of covariates **EMP**, **RDP** and **RDE**. Increasing firm size in terms of employees reduces the share of outsourced applications. As Wagner (2006, p. 133) 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 $\mathbf{RDP}^{-1}$, we get a fairly small coefficient $\hat{\beta}_4 = -0.123$. Also, for very small values **RDP**< 0.1 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. German accounting standards did not require firms to publish their R&D expenses in their balance sheets and it is likely that only large corporations voluntarily did this in the explanatory notes of their balance sheets. Whether or not expensive R&D preceded a patent (high **RDP**) has minor impact on the outsourcing rate. This is in line with Wagner (2006, p. 133). Our model predicts that for higher R&D intensity, companies are likely to file their patents themselves. Wagner (2006, p. 133),

	Estimate	Std. Error	p-value	
MEAN REGRESSION				
INTERCEPT	-0.951	0.059	0.000	
log(**COV**)	0.031	0.036	0.384	[a]
BREADTH	0.041	0.032	0.195	[a]
$\mathbf{EMP}^{1/2}$	-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]			

[a] Although insignificant, this covariate remains in the model because of a significant interaction.

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

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 5 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. Chemical / Pharma companies have the lowest outsourcing rates (3.11%). Especially for pharmaceutical companies, one single and very complex

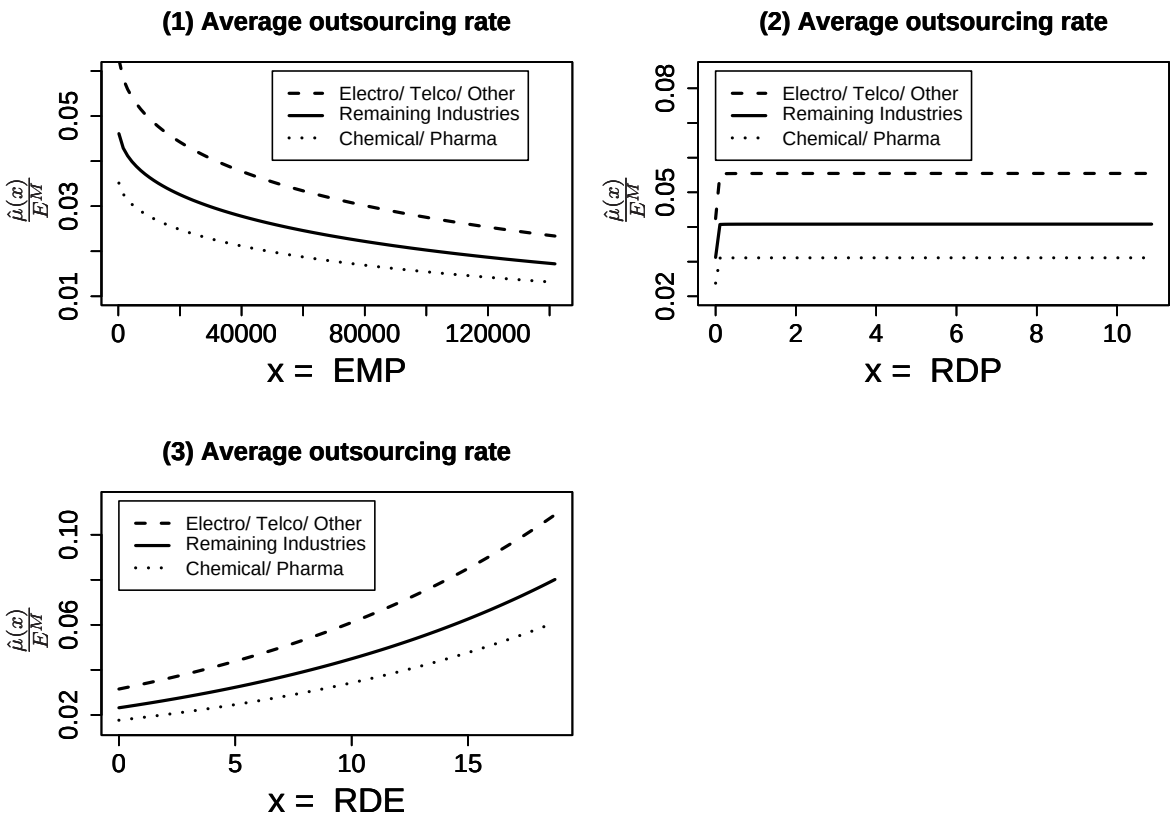

Figure 4: Influence of **EMP**, **RDP** and **RDE** on the outsourcing rate while fixing other covariates by their empirical modes

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. It is well known that patenting is largely driven by portfolio strategies in the Electronics industry in general and particularly in the semi-conductor sector (Hall and Ham (2001)). Due to overlapping technologies in these areas, firms often need to gain access to competitors' technologies and therefore engage in vast cross-licensing agreements. In the negotiations for cross-licenses the total size of the patent portfolio of the licensees has higher relevance than the characteristics of individual patents. Therefore the strategic value of individual patent is comparably low to electronic firms which might lead to a higher willingness to delegate the application procedure to external attorneys. For all remaining industries we predict an outsourcing rate of 4.08%.

6.2 Overdispersion and zero-probability as functions of regression covariates

We define the overdispersion factor of a random $Y_i \sim ZIGP(\mu_i,\ \varphi_i, \omega_i)$ as $V_i := \frac{Var(Y_i)}{E(Y_i)} = \varphi_i^2 + \mu_i\omega_i$. There are only categorical covariates for overdispersion: $\boldsymbol{w} := (\mathbf{1}$, **ENGINEER**, **CAR.SUPP.OTHER**, **YEAR**, **EMP.11291**, **RDE.63**), Using (2.5), we get

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

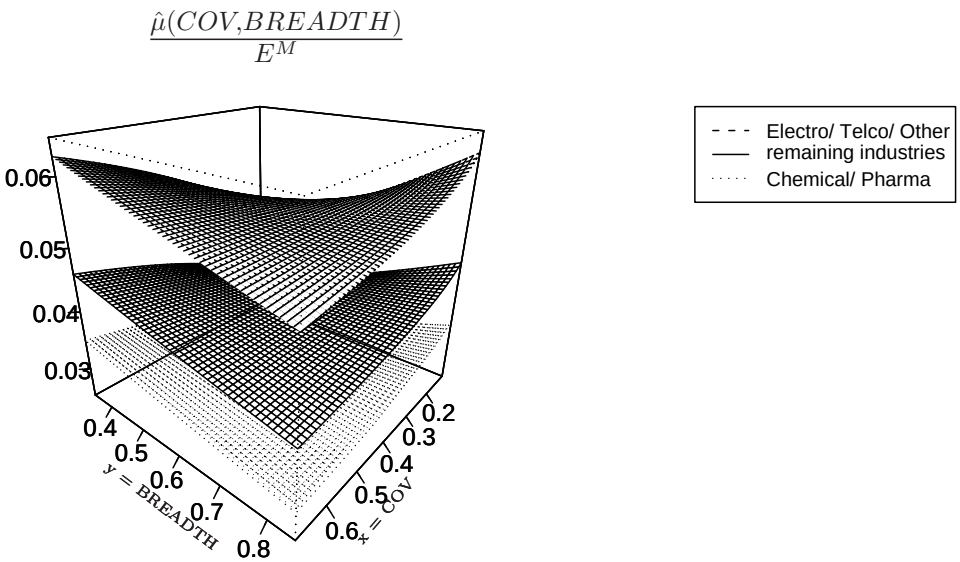

Figure 5: Common influence of COV and BREADTH on the estimated outsourcing rate $\hat{\mu}(COV, BREADTH)/E^M$ per industry

We can use this overdispersion function to estimate $V(\boldsymbol{X} = \boldsymbol{x}, \boldsymbol{W} = \boldsymbol{w}, \boldsymbol{Z} = \boldsymbol{z}) = \varphi^2 + \mu \cdot \omega$ by

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

Companies are interested in a prediction of their zero-probability rather than their zero inflation. 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

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

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

where $\boldsymbol{z} := (\mathbf{1}, \textbf{BREADTH.06}, \textbf{RDP.34}, \textbf{CHEM.PHA.ENGIN})$.

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 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. 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 $\hat{\mu}_i = \hat{\mu}(\boldsymbol{x}_i, \boldsymbol{w}_i, \boldsymbol{z}_i)$.

The quotes $\hat{V}(\boldsymbol{X}, \boldsymbol{W}, \boldsymbol{Z})$ range between 11 and 136 and indicate high overdispersion. For large companies in terms of employees, overdispersion is especially high (see for example class

7 vs. class 11). The same holds for **RDE**: overdispersion rises as R&D intensity increases (see for instance class 1 vs. 2). 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.

Figure 6 shows the influence of the observation year on the estimated overdispersion factor per industry. The legend lists the classes in descending order. We predict a positive super-linear development. Again, Engineering shows lowest overdispersion with $\hat{V}(\boldsymbol{X}, \boldsymbol{W}, \boldsymbol{Z})$ in $[14.3, 14.5]$. Largest overdispersion occurs in 'Medtech / Biotech / Electro / Telco' which have values in $[37.1, 135.9]$.

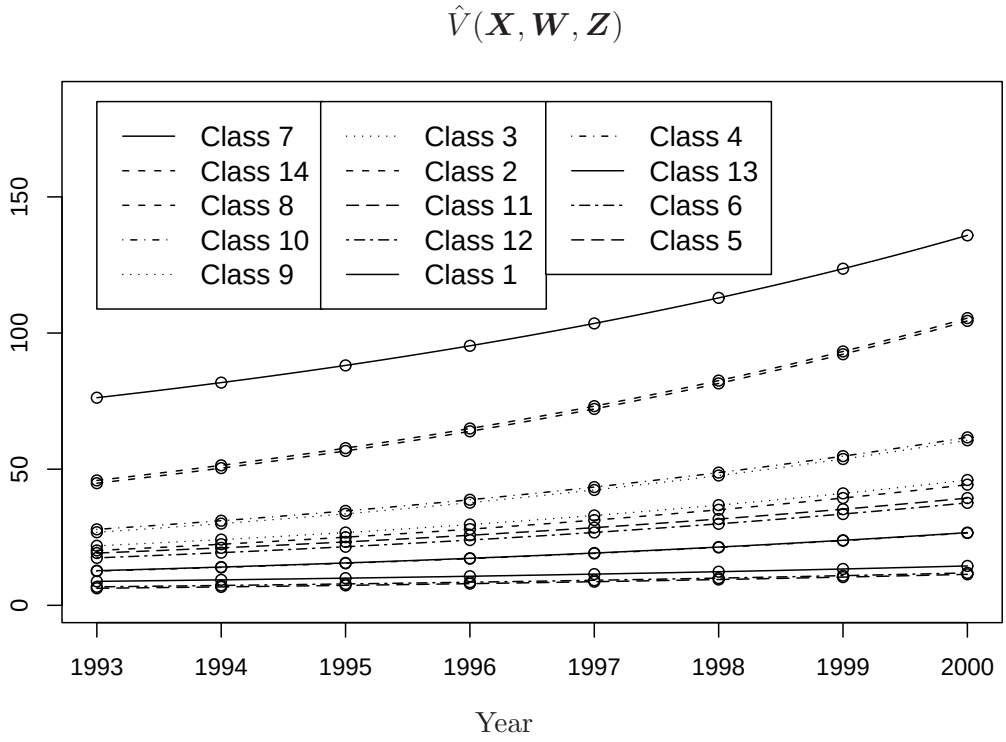

Figure 6: Influence of YEAR on the estimated overdispersion factor while fixing other covariates by their modes; legend in descending order

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 class 5 vs. 6). 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 e.g. class 13, where $\hat{P}(Y = 0|\boldsymbol{x}, \boldsymbol{w}, \boldsymbol{z}) = 33.68\%$. Chemical / Pharma show low zero-probability. Especially pharmaceutical companies have developed own patent databases and therefore are likely to file all

		Class defining covariates									
Class			dispersion covariates		zero-inflation covariates						
j	n_j	Industry	**EMP**	**RDE**	**RDP**	**BREADTH**	$\bar{\mu}_j$	$\hat{\varphi}_j$	$\hat{\omega}_j$	$\hat{V}(\boldsymbol{X}, \boldsymbol{W}, \boldsymbol{Z})$	$\hat{P}(Y = 0 \vert \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{z})$
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 overdispersion factor and probability of no outsourced patent application for 14 classes which have the largest numbers of observations in year 2000

patents themselves. We predict higher zero-probability for high-RDP companies (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 by internal departments exclusively. It seems like in crucial situations, companies trust their own patent departments more than external 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. Also, we developed the necessary asymptotic theory for these models, thus filling a theoretical and practical gap in this research area. From a simulation study, we saw that medium sample sizes of $n \geq 200$ are necessary to estimate the regression parameters on the mean and zero-inflation level, while one needs larger sample sizes to estimate the ones on the overdispersion level well.

Moreover, we carried out a comparison of different models based on the Poisson model using data 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 nested comparisons and Vuong statistics for nested and nonnested comparisons. 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.

A model interpretation confirmed insights of former work on the given data 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 overdispersion factor 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 correlations are low (see Figure 3), 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)) or a Bayesian approach involving Markov Chain Monte Carlo (MCMC) based inference. Czado and Song (2006) developed such a MCMC based inference for state space mixed models for binomial observations without zero-inflation. These possible extensions will be the subject of further research.

As an Associate Editor pointed out it is possible to apply a zero-inflated binomial mixed model to the patent data (see e.g. Hall and Berenhaut (2002)). In this case the likelihood does not have a closed form and involves multidimensional integration. Further Laplace-like approximations to the log-likelihood or MCMC methods should here be utilized. These model approaches are computationally demanding and therefore we prefer our ZIGP regression approach for the analysis of the patent data.

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 $\frac{\partial l_n(\boldsymbol{\delta})}{\partial \boldsymbol{\beta}\boldsymbol{\beta}^t} \in \mathbb{R}^{(p+1)\times(p+1)}$, $\frac{\partial l_n(\boldsymbol{\delta})}{\partial \boldsymbol{\beta}\boldsymbol{\alpha}^t} \in \mathbb{R}^{(p+1)\times(r+1)}$, $\frac{\partial l_n(\boldsymbol{\delta})}{\partial \boldsymbol{\beta}\boldsymbol{\gamma}^t} \in \mathbb{R}^{(p+1)\times(q+1)}$, $\frac{\partial l_n(\boldsymbol{\delta})}{\partial \boldsymbol{\alpha}\boldsymbol{\alpha}^t} \in \mathbb{R}^{(r+1)\times(r+1)}$, $\frac{\partial l_n(\boldsymbol{\delta})}{\partial \boldsymbol{\alpha}\boldsymbol{\gamma}^t} \in \mathbb{R}^{(r+1)\times(q+1)}$ and $\frac{\partial l_n(\boldsymbol{\delta})}{\partial \boldsymbol{\gamma}\boldsymbol{\gamma}^t} \in \mathbb{R}^{(q+1)\times(q+1)}$. Entries $h_{lm}(\boldsymbol{\delta})$'s can be computed easily. For instance, entries of $\frac{\partial l_n(\boldsymbol{\delta})}{\partial \boldsymbol{\beta}\boldsymbol{\beta}^t}$ 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 $\boldsymbol{H}_n(\boldsymbol{\delta}) := -\mathcal{H}_n(\boldsymbol{\delta})$. 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 $\boldsymbol{F}_n(\boldsymbol{\delta})$ is equal to $E_\delta \boldsymbol{H}_n(\boldsymbol{\delta})$. Thus, entries of $\boldsymbol{F}_n(\boldsymbol{\delta})$ are given by

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

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

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

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

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

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

Proof of Theorem 1

The proof of the Theorem follows Fahrmeir and Kaufmann (1985) and Czado and Min (2005). For the reader familiar with German, a detailed version can be found in Erhardt (2006, Sec. 2.3). First we would like to recall that the GP distribution with parameters μ and φ does not belong to the exponential family even if parameter φ is known. Therefore the score vector, the Hessian matrix and the Fisher information matrix have a more complex structure compared to the GLM case. Further a verification of the condition

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

under Assumptions (A1)-(A3) and for all $\varepsilon > 0$ requires much more effort than in Fahrmeir and Kaufmann (1985). Here $\boldsymbol{V}_n(\boldsymbol{\delta}) := \boldsymbol{F}_n^{-1/2}\ \boldsymbol{H}_n(\boldsymbol{\delta})\ \boldsymbol{F}_n^{-t/2}$ denotes the normed information matrix. To prove (7.9) we need to consider moments of $\boldsymbol{V}_n(\boldsymbol{\delta})$, i.e. moments of $\boldsymbol{H}_n(\boldsymbol{\delta})$. However this would involve moments of the form $E(\mu_i(\boldsymbol{\beta}) + (\varphi_i(\boldsymbol{\alpha}) - 1)y_i)^{-k}$ which are only well defined for $\varphi_i(\boldsymbol{\alpha}) > 1$. This is another reason, why we restrict to the case of overdispersion.

Let us sketch the verification of (7.9). It is not difficult to see that (7.9) follows now from

$$\max_{\delta \in N_n(\varepsilon)} \frac{|h_{lm}(\boldsymbol{\delta}) - Eh_{lm}(\boldsymbol{\delta})|}{n} \xrightarrow{P} 0 \tag{7.10}$$

and

$$\max_{\delta \in N_n(\varepsilon)} \frac{|Eh_{lm}(\boldsymbol{\delta}) - f_{lm}(\boldsymbol{\delta})|}{n} \xrightarrow{P} 0 \tag{7.11}$$

for all $l, m = 0, \ldots, p+r+q+2$. Note that the Hessian matrix (7.1) and the Fisher information matrix have 6 different entry types. We shall only illustrate (7.10) for $h_{lm}(\boldsymbol{\delta})$'s defined in (7.2). The convergence result (7.11) can be shown analogously. Letting $Z_i := \mathbb{1}_{\{y_i>0\}} Y_i(Y_i - 1)$, $U_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) := \mu_i(\boldsymbol{\beta}) + (\varphi_i(\boldsymbol{\alpha}) - 1)Y_i$, $q_{i,p}(\boldsymbol{\delta}) := x_{ip}^2 \mu_i(\boldsymbol{\beta})(\varphi_i(\boldsymbol{\alpha}) - 1)$ and

$$v_{i,p}(\boldsymbol{\delta}) := x_{ip}^2 \; \mu_i(\boldsymbol{\beta}) \frac{-\frac{1}{\varphi_i(\boldsymbol{\alpha})} \; P_i^0(\boldsymbol{\delta})^2 + \frac{\mu_i(\boldsymbol{\delta}) - \varphi_i(\boldsymbol{\alpha})}{\varphi_i(\boldsymbol{\alpha})^2} P_i^0(\boldsymbol{\delta}) k_i(\boldsymbol{\gamma})}{(k_i(\boldsymbol{\gamma}) + P_i^0(\boldsymbol{\delta}))^2}$$

it easy to see that (7.10) follows from

$$\begin{aligned}
\max_{\boldsymbol{\delta} \in N_n(\varepsilon)} \left| \frac{1}{n} \sum_{i=1}^n v_{i,p}(\boldsymbol{\delta}) \left(\mathbb{1}_{\{y_i=0\}} - E(\mathbb{1}_{\{y_i=0\}}) \right) \right| &\xrightarrow{P} 0, \\
\max_{\boldsymbol{\delta} \in N_n(\varepsilon)} \left| \frac{1}{n} \sum_{i=1}^n \frac{q_{i,p}(\boldsymbol{\delta})}{\varphi_i(\boldsymbol{\alpha})(\varphi_i(\boldsymbol{\alpha}) - 1)} \left(\mathbb{1}_{\{y_i>0\}} - E(\mathbb{1}_{\{y_i>0\}}) \right) \right| &\xrightarrow{P} 0, \\
\max_{\boldsymbol{\delta} \in N_n(\varepsilon)} \left| \frac{1}{n} \sum_{i=1}^n q_{i,p}(\boldsymbol{\delta}) \left[\frac{Z_i}{[U_i(\boldsymbol{\beta}, \boldsymbol{\alpha})]^2} - E\left(\frac{Z_i}{[U_i(\boldsymbol{\beta}, \boldsymbol{\alpha})]^2} \right) \right] \right| &\xrightarrow{P} 0.
\end{aligned} \tag{7.12}$$

Now we proceed as for the proof of Lemma 4 in Czado and Min (2005) to establish convergence result (7.12).

Acknowledgement

This work was supported by the Deutsche Forschungsgemeinschaft through grant Cz 86/1-1 and by Sonderforschungsbereich 386 *Statistical Analysis of Discrete Structures*. We are also grateful to the Editor, an Associate Editor and two anonymous referees for detailed suggestions that greatly improved this paper.

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.

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*).

Czado, C. and P. X.-K. Song (2006). State space mixed models for longitudinal observations with binary and binomial responses. *To appear in Statistical Papers (current version: Discussion paper 503 of SFB 386, http://www.stat.uni-muenchen.de/sfb386/).*

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.

Ghosh, S. K., P. Mukhopadhyay, and J.-C. Lu (2006). Bayesian analysis of zero-inflated regression models. *J. Stat. Plann. Inference 136*(4), 1360–1375.

Gschlößl, S. and C. Czado (2006). Modelling count data with overdispersion and spatial effects. *To appear in Statistical Papers (current version: Discussion paper 412 of SFB 386, http://www.stat.uni-muenchen.de/sfb386/).*

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.

Hall, B. H. and R. M. Ham (2001). The Determinants of Patenting in the US Semiconductor Industry, 1980-1994. *Rand Journal of Economics 32*, 101–128.

Hall, D. B. and K. S. Berenhaut (2002). Score tests for heterogeneity and overdispersion in zero-inflated Poisson and binomial regression models. *Canad. J. Statist. 30*(3), 415–430.

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. M. (2006). *Economic Analyses of the European Patent System*, Volume 1. Deutscher Universitätsverlag, 139 p.