

Discussion Papers No. 611, March 2010  
Statistics Norway, Research Department

*Torbjørn Skardhamar, Tore Schweder and  
Simen Gan Schweder*

## **Modelling ‘crime-proneness’** A comparison of models for repeated count outcomes

**Abstract:**

In the criminal career literature, the individual-level age-crime relationship is commonly modelled using generalized linear mixed models, where between-individual heterogeneity is then handled through specifying random effect(s) with some distribution. It is common to specify either a normal or discrete distribution for the random effects. However, there are also other options, and the choice of specification might have substantial effect on the results. In this article, we compare how various methods perform on Norwegian longitudinal data on registered crimes. We also present an approach that might be new to criminologists: the Poisson-gamma regression model. This model is interpretable, parsimonious, and quick to compute. For our data, the distributional assumptions have not dramatic effect on substantive interpretation. In criminology, the mixture distribution is also of theoretical interest by its own right, and we conclude that a gamma distribution is reasonable. We emphasize the importance of comparing multiple methods in any setting where the distributional assumptions are uncertain.

**Keywords:** criminal careers, repeated count data, random effects, Poisson-gamma regression, comparing methods

**JEL classification:** C02, C23, K40

**Acknowledgement:** The authors would like to thank Arvid Raknerud for useful comments on the manuscript. Financial support from the Norwegian Research Council and Statistics Norway is gratefully acknowledged.

**Address:** Torbjørn Skardhamar, Statistics Norway, Research Department. E-mail:

[torbjorn.skardhamar@ssb.no](mailto:torbjorn.skardhamar@ssb.no)

Tore Schweder, University of Oslo, Department of Economics, E-mail:

[tore.schweder@econ.uio.no](mailto:tore.schweder@econ.uio.no)

Simen Gan Schweder, University of Oslo, Department of Mathematics, E-mail:

[simen.schweder@gmail.com](mailto:simen.schweder@gmail.com)

---

**Discussion Papers**

comprise research papers intended for international journals or books. A preprint of a Discussion Paper may be longer and more elaborate than a standard journal article, as it may include intermediate calculations and background material etc.

Abstracts with downloadable Discussion Papers  
in PDF are available on the Internet:

<http://www.ssb.no>

<http://ideas.repec.org/s/ssb/dispap.html>

For printed Discussion Papers contact:

Statistics Norway  
Sales- and subscription service  
NO-2225 Kongsvinger

Telephone: +47 62 88 55 00

Telefax: +47 62 88 55 95

E-mail: [Salg-abonnement@ssb.no](mailto:Salg-abonnement@ssb.no)

## Introduction

The field of criminal career research is the study of individual's crimes over a longer period of time, and is therefore also often referred to as developmental or life-course criminology. Data will often be a panel with repeated measurements of offending for each individual in the sample, say for each year. Individuals vary with respect to crime proneness, also when observable determinants have been controlled for, and the correlation between observations on each individual must be accounted for. There are various models and techniques for dealing with this, and several of them are easily available in standard statistical software packages. The methods applied in the criminological literature are to a large extent based on generalized linear mixed models (GLMM), and as offending is often measured as counts, models tend to be based on the Poisson distribution. Due to software limitations, the standard approach assumes that the random effect(s) have a normal distribution (normal-GLMM). But the normality assumption might not hold, and it is uncertain whether this assumption has substantial consequences for the results (Lindsey & Lindsey, 2000; Litière, Alonso, & Molenberghs, 2007). An alternative is to relax this assumption by estimating the random effect nonparametrically (NP-GLMM). When lacking information about the true distribution, it is recommended to compare results from parametric and nonparametric models (Agresti, Caffo, & Ohman-Strickland, 2004, Skrondal & Rabe-Hesketh, 2004).

The conventional wisdom seems to be that wrongly specified mixture distribution does not have dramatic consequences, but this is not always the case (Agresti et al., 2004). The literature diverges somewhat on this point, but the issue cannot be set aside, particularly in the case of nonlinear models such as in Poisson regression models and logit models (Lindsey & Lindsey, 2000; Litière et al., 2007; Verbeke & Lesaffre, 1997).

It has been argued that non-parametric mixture model is preferable for criminal career data as the true distribution is unknown, and misspecification of the random effects might be problematic (Nagin & Tremblay 2005: 85). We think that it is not strictly correct to say that criminological theory does not suggest anything about the correct form of the latent variable, although we agree that it is quite vague. It is actually of considerable theoretical interest to study the distribution of unobserved heterogeneity in criminal proneness, although it is not given that the "true" latent distribution is identifiable. Some have argued that the offender population is best understood as discrete groups (Blumstein et al 1986; Lykken, 1995; Moffitt, 2006). Against this view is the majority of sociological theories that explicitly suggest that the distribution is continuous (e.g. Gottfredson & Hirschi, 1990; Laub & Sampson, 2003),

but without trying to specify its functional form. However, it is reasonable to suggest that the latent propensity to commit crimes is skewed with a long tail to the right. That is: most people are law-abiding to a high degree, and the number of people decreases with increasing criminal propensity. It should also be noticed that Nagin and Land (1993: 344), commented that a panel model with a discrete random intercept with five support points give a “gamma-like distribution”. Similar arguments have also been put forward by others (Greenberg, 1991; Rowe, Osgood, & Nicewander, 1990). We therefore find it reasonable to suggest a model with a gamma distributed heterogeneity.

In this paper, we compare alternative model specifications for criminal career data with count outcomes. The purpose is to see to what extent alternative specifications of the random variable has an impact on the substantive results. This exercise also warrants a discussion of what is the true distribution of “crime-proneness”. Our approach is to fit alternative models, and comparing their performance, to a longitudinal dataset following a Norwegian birth cohort from age 10 to 22 years old. One of the models, which is denoted a Poisson-gamma regression model, might be new to criminologists, but is known from other areas of research (see e.g. Rolph, Adams, and McGuigan, 2007).

It might be reasonable to suggest that a fraction of the population is law abiding to a high extent and commit no crimes. The Poisson-gamma model can handle this situation by including a discrete mass for this group, and let the latent “crime-proneness” vary in degree only for the remaining sample. We refer to this modification of the Poisson-gamma model as a zero-inflated Poisson-gamma regression model, which is also discussed and fitted.

Residuals are useful to assess model fit, and to diagnose potential sources of misfit. In the present context, residuals might be defined as predicted values of individual latent crime-proneness. A simple formula is developed for such residuals in the Poisson-gamma model. The residuals are used to reveal the distributional form of the latent crime-proneness.

An advantage for the applied researcher is that all methods considered here can be estimated using easily available software packages. The standard models are computationally intensive, and with large datasets, the computation time may be substantial. An advantage of our new proposed model is that it is estimated analytically and is thus very quick to compute. We restrict the discussion to count data and Poisson models with a random intercept. We also present some additional elaborations upon the

suggested Poisson-gamma model that might be useful for criminologists. In conclusion, we discuss the appropriateness of the models, and the consequences of misspecification.

## Methods for repeated count data

### Standard random intercept models

Criminal career data are usually structured as panels with repeated measurements of offending within subjects. In surveys, the respondents might have filled out a questionnaire each year, and with register data one might have the total number of arrests (etc.) for each person each year. When the outcome is counts, it is often assumed that the outcome variable is Poisson distributed, and Poisson GLMMs can be estimated with standard software. The distribution of unobserved heterogeneity between individuals is of primary interest, we will restrict attention to models with two levels.<sup>1</sup> The basic unit of observation is, say the number of convictions an individual has during a period of observation, say a given year. The individual is then represented by its time-series of basic observation. The periods are at level 1 and the individual at level 2. There might be further levels, say family, neighbourhood, period etc, but it is beneficial for our discussions to concentrate on the two-level models.

The observations consist of a response  $Y$  and a covariate vector  $x$ , possibly with individual specific components such as sex. If the outcome is a count variable (eg number of crimes), the preferred method is Poisson regression, where we model the log expected number of crimes,  $\log(\mu)$ .

Individuals are indexed by  $j$  and occasions by  $t$ , and a simple linear model with unobserved heterogeneity only in the level is then written as

$$(1) \quad \begin{aligned} \log(\mu_{jt}) &= \eta_j + \beta'x_{jt} \\ \eta_j &= \gamma + \zeta_j \end{aligned}$$

where  $\beta$  is a regression coefficient, and  $\eta_j$  is individual-specific random intercept with mean  $\gamma$  and variation  $\zeta_j$ . Thus, each individual has its own intercept drawn from some distribution and we have to make some assumptions about the distribution of these between-individual differences. The most common application is to assume that  $\eta_j$  is normally distributed, but other options are available. So instead of estimating one intercept for each individual, it is sufficient to estimate the parameters for

---

<sup>1</sup> The use of the term “level” is used as in multilevel modelling even though it is a traditional panel design, which alludes to that there might be more levels of nesting, say, family and neighbourhoods (see, Skrondal and Rabe-Hesketh, 2004).

this distribution (e.g. the mean  $\gamma$  and variance  $\sigma^2$ ). In the following, we will refer to this model as Poisson-normal model. This modelling approach might be attractive since it allows structure in the random effects, and can be estimated with the most commonly used computer software.

Another possibility is to let  $\zeta_j$  have a discrete distribution. Then we have a group-based model with as many groups as there are support points for the intercept and a common effect of  $x$ . A discrete distribution would be appropriate if indeed the population is stratified into distinct statistically homogeneous groups. It is also a fact that any distribution can be approximated by a finite discrete distribution, and latent class models with only a few classes are often found to approximate models with continuously distributed random effects very well (Skrondal & Rabe-Hesketh, 2004). In this case, we need to search for the number of support points  $g$  that best captures the variation around  $\gamma$ . As this kind of model is a GLMM with non-parametric intercept, we refer to this model as Poisson-NP model. A disadvantage of this model is that it can lose some efficiency since there are many parameters to be estimated (Agresti et al., 2004). Both the Poisson-normal and -NP models can be computationally demanding, especially with large samples.

Importantly, in ordinary GLMMs with Poisson specification, the linear term, including the random term, is put on the logarithmic scale. The distribution of individuals' "crime-proneness" is then symmetric on the logarithmic scale, which is a skewed distribution on the natural scale. The points of support for the non-parametric distribution are of course also on the logarithmic scale, and in both cases, the random effects might be easier to interpret when taking the exponential to bring parameters onto the natural scale.

Although Poisson GLMMs are convenient, the Poisson distribution is restrictive in the sense that the variance is assumed to be equal to the mean, and this assumption might be violated. Booth et al. (2003) relax this assumption and assume that given the individual random effect for criminal proneness the counts  $Y_{j1}, \dots, Y_{jT}$  are independent and negatively binomially distributed. The negative binomial model should be well-known in criminology (see e.g. Land, McCall, Nagin, 1996), but the most commonly used softwares have not a procedure for estimating it with a random effect. The conditional distribution for  $Y_j$  is thus

$$(2) \quad P(Y_j = y_j | \zeta_j) = \frac{\Gamma(\alpha + y_j)}{\Gamma(\alpha) y_j!} \left( \frac{\alpha}{\alpha + \mu_j} \right)^\alpha \left( \frac{\mu_j}{\alpha + \mu_j} \right)^{y_j}.$$

Booth et al. (2003) suggest modelling the individual random effects by the normal distribution. The variance  $\sigma^2$  in this latent distribution reflects the degree of intra-individual correlation, while  $\alpha$  reflects over-dispersion relative to the Poisson distribution with more over-dispersion the smaller the parameter. Their model might be called the Negative binomial normal model. It must be fitted by numerical integration of the individual random effects.<sup>2</sup> The log likelihood for unit  $j$  is

$$l_j = \log \Gamma(y_{ij} + 1/\alpha) - \log \Gamma(y_{ij} + 1) - \log \Gamma(1/\alpha) + \\ (1/\alpha) \log(1 - \alpha\mu_{ij}) / (1 + \alpha\mu_{ij}) + y_{ij} \log(\alpha\mu_{ij} / (1 + \alpha\mu_{ij}))$$

### A Poisson-gamma regression model

In the negative binomial normal model, there is one parameter reflecting over-dispersion, and a separate parameter reflecting intra-individual correlation. The Poisson gamma model suggested in this section has a common parameter  $\alpha$  reflecting both these extensions from the simple Poisson regression model.

Others have hinted that a gamma distribution for unobserved heterogeneity in crime proneness might be reasonable (Greenberg, 1991; Rowe et al 1990). We follow this lead. In our model there is an individual-specific random effect, which scales the mean parameters in the Poisson distribution for the counts. Given the random effect, the counts are assumed (conditionally) independent given the covariate vectors for the individual counts. Since the gamma distribution is conjugate to the Poisson distribution we choose the multiplicative random effect  $\tau$  to be gamma distributed.

Individuals might not be exposed for criminal activity during the whole of a given period. The individual might actually be in jail, be living abroad, or even be dead. Let  $h_t$  be the fraction of time the individual was exposed for period  $t$ . For covariate vectors  $x_t$  the log-linear Poisson-gamma regression model is:

- The counts  $Y_t$   $t=1, \dots, T$  are, conditional on the random effect  $\tau$  (and possibly on previous counts) independent and Poisson distributed with mean value  $EY_t = \mu_t = \tau h_t e^{x_t \beta}$  where  $\beta$  is a vector of regression parameters;

---

<sup>2</sup> Booth et al provide code to fit this model using SAS NLMIXED at the journal home page: <http://stat.uibk.ac.at/SMIJ/>

- $\tau$  is gamma distributed with both shape and rate parameter  $\alpha$  to make  $E\tau = 1$ . The random effects are independent across individuals.

An intercept term is usually included in the regression in the exponent. This parameter will then scale the mean count. For this reason the gamma distribution is required to have mean 1 by having the same shape and rate parameter. The exposure fraction might be lifted up in the exponent, making  $\log(h_t)$  an offset term in the conditional Poisson regression. To simplify notation we assume  $h_t = 1$  when not otherwise stated.

We assume the mixing distribution to be gamma, with shape parameter to be estimated. This distribution can be heavily skewed, with a long tail towards higher values. The smaller  $\alpha > 0$  is, the more dispersed and skewed is the distribution. Since  $\text{var}(\tau) = 1/\alpha$  the latent distribution might have been parametrized by  $\theta = 1/\alpha$  rather than by  $\alpha$ . The gamma distribution is, as mentioned, conjugate to the Poisson distribution, and integrating out the latent variable is done analytically, leaving us with a closed form likelihood.

Let  $Y_{jt}$  be the number of charges individual  $j$  experience in year  $t$ , which, given the individual's latent proneness  $\tau_j$ , is Poisson distributed with mean  $\tau_j \mu_{jt}$  and independent for  $t = 1, \dots, T$ . Unconditionally  $Y_{jt}$  has a negative binomial distribution when  $\tau$  has a gamma distribution. In order to identify the intercept term in the log-linear regression,

$$(3) \quad \log(\mu_{jt}) = \beta' x_{jt},$$

we require the gamma distribution to have a mean 1, and thus density

$$(4) \quad f(\tau) = \frac{\alpha^\alpha}{\Gamma(\alpha)} \tau^{\alpha-1} e^{-\alpha\tau}.$$

Unconditionally the expected number of charges  $Y_{jt}$  is thus  $\mu_{jt}$ , and the variance is  $\mu_{jt} + \mu_{jt}^2 / \alpha$ .

One of the covariates in the log-linear regression (3) is the constant 1 related to the intercept term. Usually age appears as a third order polynomial orthogonal to the constant, and the parameter vector for this is the career profile. There might also be cohort and period effects together with sex, education etc.



The joint distribution of an observed growth curve  $Y = (Y_1, \dots, Y_T)$  for a fully exposed individual in all the  $T$  periods is

$$\begin{aligned}
 f(y_{1j}, \dots, y_{Tj} | x, \alpha, \beta) &= \int_0^\infty \left[ \prod_{t=1}^T \left( \frac{(\tau \mu_{tj})^{y_{tj}}}{y_{tj}!} e^{-\tau \mu_{tj}} \right) \right] \frac{\alpha^\alpha}{\Gamma(\alpha)} \tau^{\alpha-1} e^{-\alpha \tau} d\tau \\
 (5) \quad &= \prod_{t=1}^T \left( \frac{\mu_{tj}^{y_{tj}}}{y_{tj}!} \right) \frac{\Gamma(\alpha + y_{+j})}{\Gamma(\alpha)} \frac{\alpha^\alpha}{(\alpha + \mu_{+j})^{y_{+j} + \alpha}} \\
 &= \frac{y_{+j}!}{\prod_{t=1}^T y_{tj}!} \prod_{t=1}^T \left( \frac{\mu_{tj}}{\mu_{+j}} \right)^{y_{tj}} \cdot \frac{\Gamma(\alpha + y_{+j})}{\Gamma(\alpha) y_{+j}!} \left( \frac{\alpha}{\alpha + \mu_{+j}} \right)^\alpha \left( \frac{\mu_{+j}}{\alpha + \mu_{+j}} \right)^{y_{+j}},
 \end{aligned}$$

where  $y_{+j} = \sum_{t=1}^T y_{tj}$  is the total count for the individual, and  $\mu_{+j} = \sum_{t=1}^T \mu_{tj}$  is the expected total count.

The bottom expression in (5) shows that the total counts  $y_{+j}$  are negatively binomially distributed, while given the total the individual counts are multinomially distributed within individual.

The marginal distribution of the individual counts is a negative binomial as in (2), but then without any conditioning. The unconditional count variance is thus  $\text{var}(y_{ij}) = \mu_{ij} + \frac{\mu_{ij}^2}{\alpha}$  while the

unconditional covariance of different counts within individual is  $\text{cov ar}(y_{ij}, y_{uj}) = \frac{\mu_{ij} \mu_{uj}}{\alpha}$   $t \neq u$  and

the unconditional correlation is  $\left[ \left( 1 + \frac{\alpha}{\mu_{ij}} \right) \left( 1 + \frac{\alpha}{\mu_{uj}} \right) \right]^{-1/2}$ . The correlation decrease from 1 at  $\alpha = 0$  to 0 at  $\alpha = \infty$ .

The log likelihood for unit  $j$  is from (5)

$$(6) \quad l_j = \sum_t (y_{tj} \log(\mu_{tj}) - \log(y_{tj}!)) + \log \Gamma(\alpha + y_{+j}) - \log \Gamma(\alpha) + \alpha \log(\alpha) - (\alpha + y_{+j}) \log(\alpha + \mu_{+j})$$

where  $\log(\mu_{ij}) = x'_{ij} \beta$ , and the total log likelihood is the sum of these individual log likelihoods.

Since the likelihood (5) splits in a product of conditional multinomial likelihood components only involving the regression parameters, and a product of negative binomial components, the conditional

maximum likelihood estimator of  $\beta$  given the total counts is obtained from a multivariate logistic regression. When this is in hand, an approximate maximum likelihood estimator of  $\alpha$  is obtained by maximizing the negative binomial part of the likelihood. This procedure will not yield the exact maximum likelihood estimator because there is some information on regression parameters in the total counts. If the exact estimator is required, the approximate estimator provides an excellent starting point for the hill climbing algorithm used.

If the individual is not fully exposed,  $\log(\mu_{ij}) = \log(h_i) + x'_{ij}\beta$  where the offset  $\log(h_i)$  is a constant that is not to be estimated.

Due to the analytic nature of our Poisson-gamma regression model, the conditional distribution of  $\tau$  given an observed growth curve  $y$  is found to be gamma with shape parameter  $\alpha + y_+$  and rate parameter  $\alpha + \mu_+$ . The point-prediction of the realized value of  $\tau$  for the individual is thus

$\hat{\tau} = \frac{\alpha + y_+}{\alpha + \mu_+}$ , and the predicted future counts are obtained accordingly.

### **Zero-inflated Poisson-gamma regression model**

The gamma distribution is in some cases inappropriate for unobserved crime proneness. If, say, a large fraction of the individuals are law abiding with no charges the value of  $\alpha$  should be small. If, on the other hand the total number of charges among those with charges show little variation, a large value of  $\alpha$  is required. In such cases a remedy for the model is to augment the latent distribution with some discrete mass, say  $1 - \pi$  at  $\tau = 0$ , while retaining the gamma distribution with mean 1 for individuals with positive crime proneness, and  $\pi$  is a parameter to be estimated. This modification of the model does not only affect the estimation of  $\alpha$ , but also the  $\beta$ -coefficients because the  $\beta$ s are only relevant given that one belongs to the crime-prone group,  $\tau > 0$ . The log likelihood for an individual with at least one charge,  $y_{+j} > 0$ , is still given by (6) above, while for individuals with no charges recorded over the study period the log likelihood is

$$l_j = \log \left( 1 - \pi + \pi \left( \frac{\alpha}{\alpha + \mu_{+j}} \right)^\alpha \right).$$

This zero-inflated model has one extra parameter, is nearly as easy to fit as the Poisson-gamma model. We denote this model zero-inflated Poisson-gamma regression.

Note that we place the discrete mass  $1 - \pi$  at the a priori support point  $\tau = 0$ , and not at a support point to be estimated as in the Poisson-NP presented above. It seems to us more natural to have a latent group of completely law-abiding individuals, and to allow those with a more relaxed attitude to the law to have a continuous proneness distribution. The zero group might simply be more easily identified and handled than the less homogeneous groups of crime prone individuals. For small  $\alpha$ 's the density of  $\tau$  will decline monotonously towards higher values. The hypothesis of a distinct law abiding group,  $\pi < 1$  can be tested by the likelihood ratio test.

### **Assessing the distributional assumption**

The mentioned, models differ primarily in the distributional assumption with regard to the random intercept, and we would like to assess the impact of these assumptions. It is hard to assess the distributional assumptions of the random effects because these are unobserved and cannot be checked directly. An obvious first step is to compare relative fit, and we should put greater trust in models that more closely reproduce the data. Comparing raw log likelihood ratios do not apply as the models are not nested. However, using BIC and AIC, the relative fit are penalized according to the number of parameters and thus we can also make direct comparisons of non-nested models (Lindsey & Lindsey, 2000), although this cannot be the final answer (Agresti et al., 2004). The relative fit criteria are calculated as  $BIC = \log(L) - 0.5m \log(N)$  and  $AIC = \log(L) - 0.5m$ , where  $L$  is the log likelihood,  $m$  is the number of parameters in the model, and  $N$  is the number of independent observations (individuals).

We should also compare the substantial results. It is often the case that even though some models might be statistically better, they might give essentially the same results. However, if there are diverging results, this gives rise to concern. This will give an indication on whether the assumptions underlying the various models are important for estimating the fixed effects. If eg a Poisson-NP and a Poisson-normal give different results, there are reasons to doubt the normality assumption.

In the Poisson gamma model there are really two aspects to investigate: are the counts within individual multinomially distributed given the total, and are the total counts negatively binomially distributed. The first question can be assessed by calculating the conditional multinomial likelihoods in (5) and summing their logarithms for  $y_{+j} > 0$ . The likelihood ratio test of the multivariate logistic regression will be indicative of any possible over-dispersion relative to the conditional Poisson distribution. A large log likelihood ratio could also be due to the regression model not fitting well. These two aspects might be untangled by looking at plots of the raw residuals

$r_{ij} = (y_{ij} - \hat{\mu}_{ij}) / \sqrt{\hat{\mu}_{ij} + \hat{\mu}_{ij}^2 / \hat{\alpha}}$  against covariates and other potential systematic factors. If no pattern is seen, but the likelihood ratio is high, it is natural to conclude that the counts are over-dispersed. In that case the more complicated model of Booth et al. (2003) might be considered.

We would also like to investigate the mixture distribution more closely. Given a model, one can make a point prediction of the latent variable for each individual. If the assumption about the mixture distribution is correct, then the point predictions should give a similar distribution. As mentioned,

$$(9) \quad \hat{\tau}_j = \frac{\hat{\alpha} + y_{+j}}{\hat{\alpha} + \hat{\mu}_{+j}}$$

is the predicted value of crime proneness, corrected for covariate information, for individual  $j$  in the gamma-Poisson model. If this model holds, the predicted values, which might be regarded as individual-specific residuals, should be nearly gamma distributed. Although these residuals will be slightly correlated through the common parameter estimates, a histogram or other distributional plots will be informative. The so-called QQ-plot against the gamma distribution with shape parameter  $\hat{\alpha}$  consists of plotting the ordered residuals against corresponding quantiles of the gamma distribution. A straight line indicates a good fit, while a staircase shape indicates that the latent distribution is discrete. We suggest to use the residuals given by (9) also for other models than the Poisson-gamma model. Then the expected counts are estimated from that other model, while  $\hat{\alpha}$  is taken from the Poisson-gamma model.

Note that the Poisson-NP is in itself a test of the mixture distribution, as it can be seen as a nonparametric approximation to an unknown continuous distribution (Skrondal & Rabe-Hesketh, 2004). If the supportpoints looks symmetric, then a normal distribution might be reasonable.

## Data

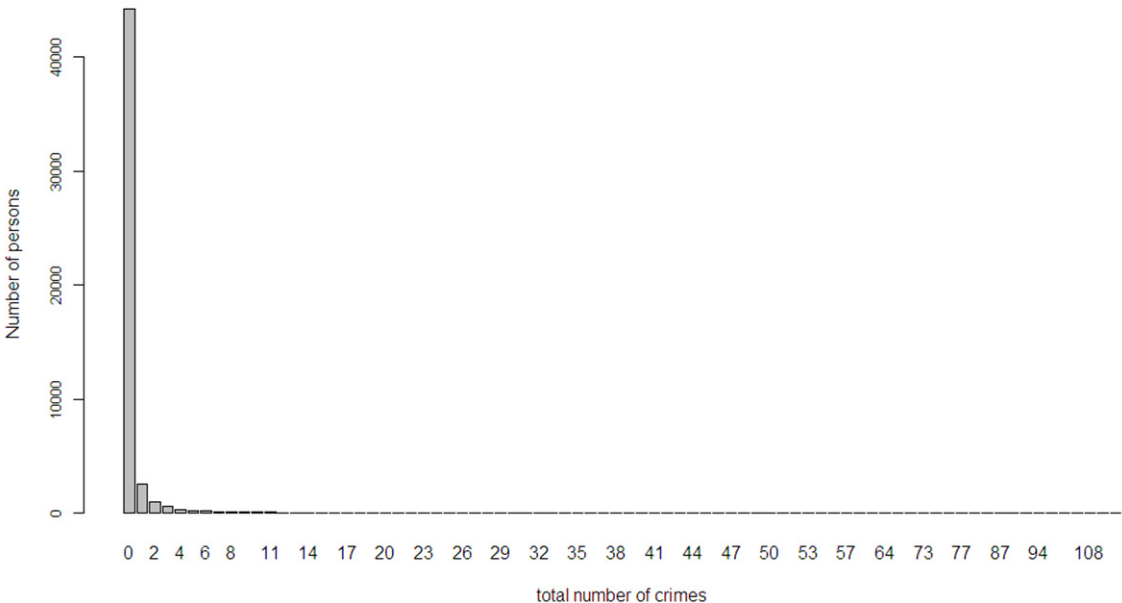
The data are gathered from the Norwegian administrative registers available for research purposes through Statistics Norway. The demographic databases contains the total population, and we have here chosen to look at one single birth cohort, born 1982, that is followed through the administrative registers. Information about charges for crimes is gathered from the official crime statistics where each individual is registered with their national identification number. This information is available from 1992 until 2004. Thus, we can follow these individuals through 13 years, from the year they turn 10 until 22 years old. We have included only persons that were alive and resident in Norway through

1992, and have also excluded all persons with immigrant background. The total sample is 49.975 persons. These data are analysed and presented in some more detail elsewhere (Skarðhamar, 2009).

We adhere to the standard definitions of offences as applied in the official statistics, where we distinguish between “serious crimes” and “misdemeanours”. The former is the more serious offences as defined in the Criminal Act, while misdemeanours is largely offences outside the Criminal Act, but for this age-group it is mainly shop-lifting and traffic offences. We chose to only look at the more serious offences, those defined as “serious crimes”. Each person might be charged with multiple crimes in a given year, or even the same day. Even though a person can commit multiple serious crimes simultaneously, we will here assume that crimes are committed one by one according to a Poisson process given the random effect and the covariates for the period.

Some descriptive statistics are given in figure 1. The majority (88%) had none charges for serious crimes during this time period. A total of 11 per cent have been charged for at least one serious crime, and this is much more common among boys than girls (not shown). Most of the offenders had been charged for only one (43%) or a couple (27%) of serious crimes, while very few have been charged with a larger number of serious crimes. The maximum count for a person was over 100 crimes.

**Figure 1 Barplot of total number of crimes per person for the period 10 to 23 years old**



## Results

We start by estimating an ordinary Poisson model ignoring the clustering of data within individuals, and then an ordinary negative binomial model recognizing over dispersion but still ignoring clustering. We proceed by fitting the four models for clustered counts discussed above; Poisson-normal, negative binomial-normal, and Poisson-NP, and Poisson-gamma models. The results are presented in Table 1. As expected, the ordinary Poisson model gives by far the worst relative fit to the data as it does not take neither clustering or dispersion into account. The ordinary negative binomial model gives much better relative fit, and the very large  $\alpha=50.2$  suggest that the data are highly dispersed. Moreover, both these two models give almost exactly the same parameter estimates for the fixed effects, except for the cubic term of age.

However, these two models do not take account for correlated observations within individuals, and are therefore inappropriate. The Poisson-normal model handles correlated observation by a random intercept, which gives greatly improved relative fit compared to the ordinary Poisson model, but worse than the negative binomial model. The Poisson-normal model only takes account for the clustering of observations, but not overdispersion which is clearly present in the data. Given the highly skewed characteristic of our data, it might be that normal distribution is inappropriate.

To take account of overdispersion, we fit the negative binomial-normal model which takes both clustering and dispersion into account. This improves the relative fit considerably relative to the Poisson-normal, giving evidence of additional over dispersion. The parameter estimates for the age effects are similar, while the effect for sex is similar as for the Poisson-normal model.

Our main interest is in the specification of the random effect, and we proceed by ignoring over dispersion and estimate the Poisson-NP model, which is more flexible considering the true distribution of heterogeneity. This model does not improve relative fit further, and the parameter estimates are almost identical to the estimates from the Poisson-normal model, except for the estimate for sex.

The Poisson-NP estimates can be used as a test for the distributional assumption, and suggests that the distribution is far too skewed to be approximated by a normal distribution. The mixture probabilities are plotted for each masspoint in figure 1. Seven masspoints capture the variance sufficiently in this case, and in this model, 86 per cent of the observations fall into the category with the lowest value on the intercept. The model is estimated including a mean intercept, so when the age variable (which is centred at age 18) is zero, the expected number of crimes for this group is  $\exp(-4.598 - .599) = .0055$ .

Thus, this group is “crime-prone” indistinguishable from zero crimes at all times. The other mass points decrease in sizes with increasing crime-proneness, which suggest a pattern that is very far from normal distributed. The pattern is plotted in figure 2, and we conclude that this give reason to be sceptical about the assumption under the Poisson-normal.

**Table 1 Results from alternative regression models. Standard errors and mixture probabilities reported in parenthesis**

	Poisson GLM (no random part)	Neg.Bin GLM (no random part)	Poisson normal- GLMM	Type II Neg.Bin. normal-GLMM	Poisson NP- GLMM	Poisson- gamma model
Constant	-2.471 (.010)	-2.464 (.021)	-5.805 (.046)	-5.484 (.047)	-4.598 (.032)	-2.471 (.027)
Age *	3.366 (.054)	3.535 (.097)	3.366 (.054)	3.663 (.097)	3.366 (.054)	3.362 (.054)
Age squared	-4.473 (.094)	-4.519 (.142)	-4.473 (.094)	-5.032 (.143)	-4.473 (.094)	-4.470 (.094)
Age cubic	-.696 (.268)	-1.393 (.429)	-.696 (.268)	-1.495 (.432)	-.696 (.268)	-.687 (.268)
Sex	-1.865 (.018)	-1.892 (.029)	-2.297 (.053)	-2.300 (.053)	-1.889 (.038)	-1.863 (.040)
$\alpha_1$		50.159		4.556		
$\alpha_2$						.062
$SD(\zeta_{\alpha_j})$			2.891	2.741		
MASS 1 ( $\pi_{\beta}$ )					-.599 (.856)	
MASS 2 ( $\pi_{\beta}$ )					2.951 (.097)	
MASS 3 ( $\pi_{\beta}$ )					4.249 (.028)	
MASS 4 ( $\pi_{\beta}$ )					5.119 (.011)	
MASS 5 ( $\pi_{\beta}$ )					5.861 (.005)	
MASS 6 ( $\pi_{\beta}$ )					6.458 (.002)	
MASS 7 ( $\pi_{\beta}$ )					7.065 (.001)	
No. parameters	5	6	6	7	19	6
Log likelihood	-119 298	-64 855	-69 278	-57 215	-69 327	-69 627
AIC	-119 300	-64 858	-69 281	-57 219	-69 336	-69 630
BIC	-119 309	-64 870	-69 292	-57 232	-69 371	-69 641
Time (hh:mm:ss)	00:00:13	00:01:27	07:10:16	06:38:16	08:36:19 **	00:02:34
Software / procedure	STATA / poisson	STATA / nbreg	STATA/ gllamm	SAS/ nlmixed	STATA / gllamm	R / nlm

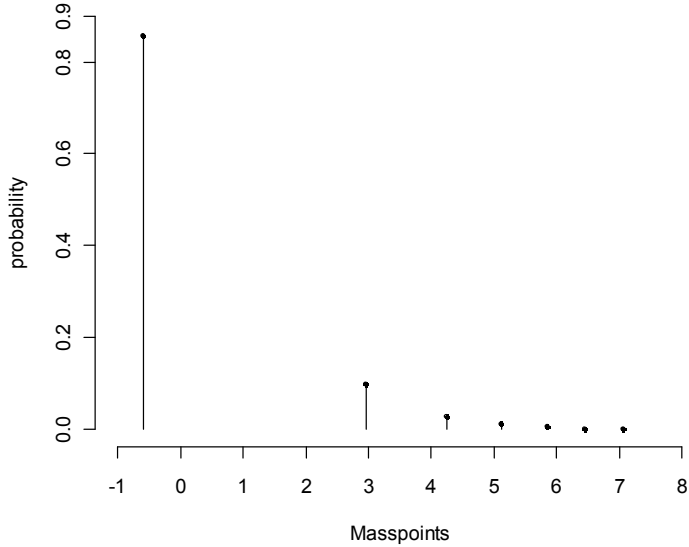
Note: The  $\alpha$  in the negative binomial models and the Poisson-gamma models do not have quite the same meaning and therefore subscripted in this table and placed on separate lines.

All estimations were done on a computer with Intel core duo E7300 processor and 2GB ram. All models were provided with good starting values which speed up the estimation. STATA and R read all the data into the memory, while SAS reads the observations one-by-one. This might affect the computation time so that SAS generally runs a little slower.

\* Age is rescaled to reduce computation time, so that age ranges from -0.7 to 0.5.

\*\* The computation time for Poisson-NP model is for the final model, but it requires running models from 2 through 6 masspoints first, giving a total computation time for the six models of more than 23 hours.

**Figure 2 Estimated masspoints from seven-group Poisson-NP model. Point estimates and mixture probabilities**



Finally, we estimate the Poisson-gamma model. Contrary to our expectations, this model does not improve the relative fit further, although the log likelihood is in about the same region as the Poisson-normal and Poisson-NP models. Thus, the negative binomial-normal model gives the best relative fit to the data.

Comparing parameter estimates, the main differences are in the estimated effect of sex. While the Poisson-normal and the negative binomial-normal model gives an estimate of about -2.3, while the estimate from the Poisson-NP model and the Poisson-gamma is -1.9. Although these estimates are perhaps not *dramatically* different, it is notable. Considering the age effects, these estimates are almost identical across models, except for the negative binomial normal model where the effect of age on order two and three. The differences in the combined effect of age parameters are hard to see from the parameters, and can be more easily inspected by plotting the expected trajectories. We plot the three directly comparable models that only differ in the specification of the random effect. The expected number of crimes at time  $t$  from the Poisson-normal model can be calculated as

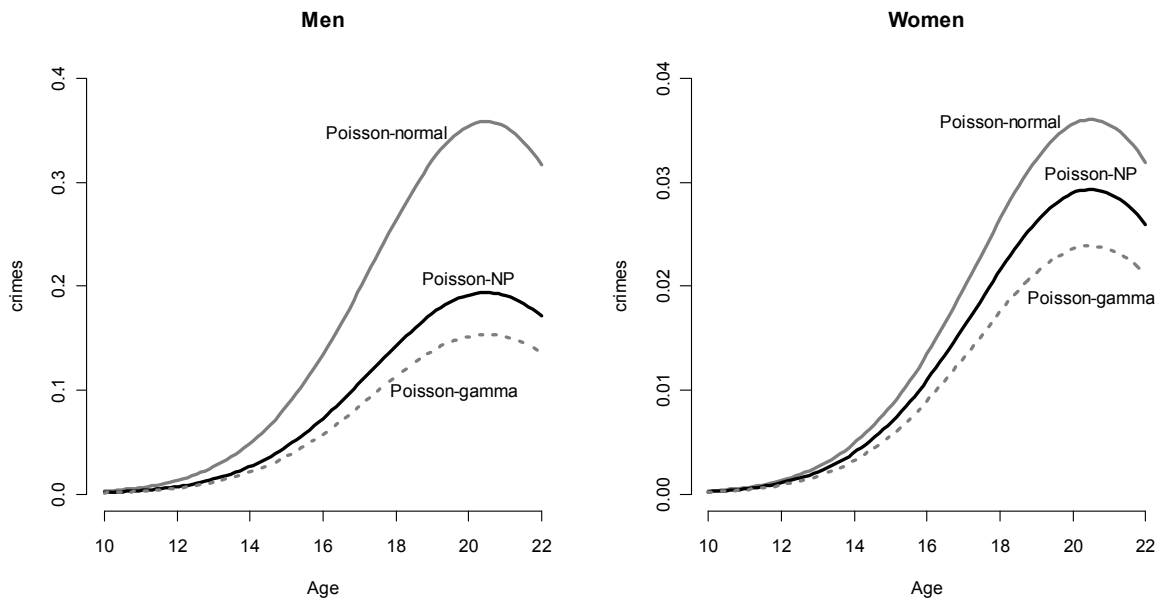
$E(\mu_t) = \exp\left(\frac{1}{2}\sigma^2 + \alpha + \beta X_t\right)$ , where  $\sigma^2$  is the variance of the random part  $\zeta$ . For the Poisson-NP model the expected count is  $E(\mu_t) = \sum_g \exp(\zeta_g + \alpha + \beta X_t) \pi_g$ , where  $\pi_g$  is the mixture probabilities

for masspoint  $g$ . For the Poisson-gamma model the expectation is  $E(\mu_t) = \exp(\alpha + \beta X_t) \tau_t$ , where



$E(\tau) = 1$ . The expected sequences of counts at times  $t = 1, 2, \dots, T$ , are plotted in figure 1. There are clearly differences in the *level* of the expected trajectories, which is due to the differences in the estimated intercepts which are much lower for the Poisson-normal model than for the Poisson-NP and the Poisson-gamma models. Note that the scale of the y-axis for men and women are different, reflecting the large coefficients for sex reported in table 1.

**Figure 3 Expected trajectories for men and women. Calculated from the alternative models**



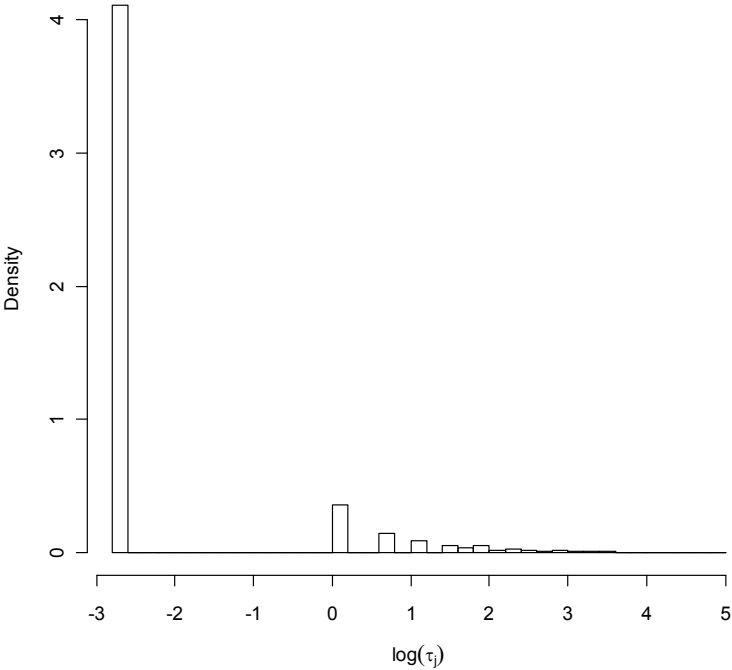
Although the models considered here rely on different assumptions, the consequences for the fixed effects are primarily related to the intercept, although there were also some differences in the effect of sex. The age parameters were largely unaffected by varying model specifications. The Poisson-NP model should be least affected by misspecification of the random effect as it should be better able approximate any distribution of the latent variable. It is then interesting that the estimate from the Poisson-gamma model is much closer to the NP-estimates than the other models. We also estimated a zero-inflated Poisson-gamma model, but without any improvement in relative fit and with estimated  $\pi \approx 1$ . Thus, the much skewed gamma random effect, effectively handles that most individuals are ‘crime-prone’ at a level indistinguishable from zero.

It is also worthy to note that the Poisson-normal, Poisson-NP and negative binomial-normal models are computational intensive, which may be time consuming if analyzing large samples. The two versions of the Poisson-gamma model are much quicker to estimate. If the substantive results do not

differ between models, then the Poisson-gamma models can be preferable from a pragmatic point of view.

We now study the distribution of residuals as specified in equation (9) above. We have plotted the log of the  $\tau_j$  residuals in figure 4. We use the log as the right tail would not show on the natural scale due to the high number of persons with a negative estimated  $\tau_j$ . The plot shows a clearly skewed distribution where the majority of individuals have a very low level of “crime-proneness”.

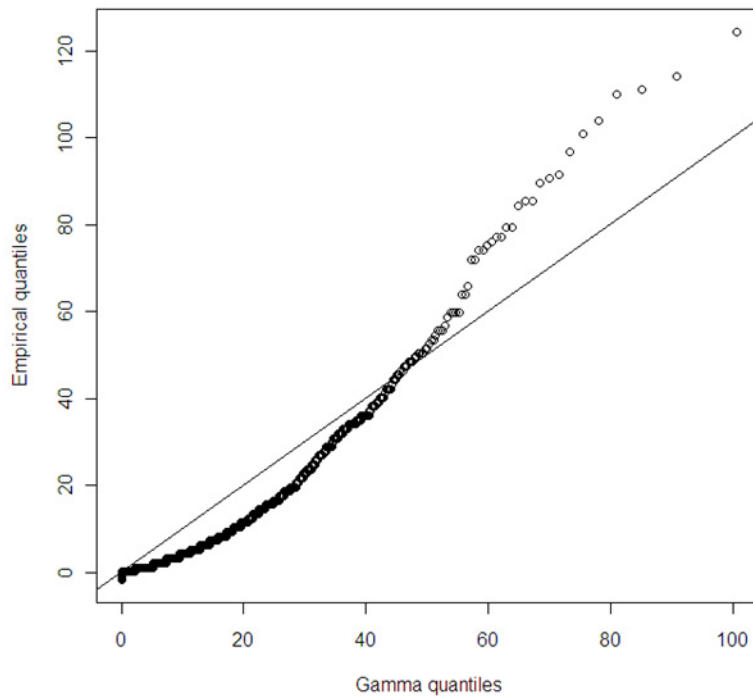
**Figure 4 Histogram of fitted  $\log(\tau_j)$**



To assess whether this distribution is reasonably well captured by a gamma distribution, we plot the empirical quartiles against the quartile of the theoretical gamma distribution as shown in Figure 5. The QQ-plot does not form quite a straight line, hence, the residuals are not quite well approximated by a gamma distribution. Most of the observations lie in the first quartiles, and the right tail of the distribution of  $\tau_j$  is too thick compared with the gamma distribution.

A second observation is that if the true distribution was approximately discrete, we would expect a staircase pattern, but this is not the case. We can therefore conclude that the true distribution is reasonably approximated by a skewed, continuous distribution – although not quite gamma distributed.

**Figure 5** QQ-plot of predicted versus the fitted gamma distribution



## Conclusion

Generalized linear mixed models (GLMM) is a popular technique of analysing clustered count data, and criminal career studies have often data with repeated measures of number of crimes, say each year. It is often assumed a normal distribution for the random effects as it can be easily estimated with standard software packages. However, other alternatives are almost equally easily available, and it might be important to consider what type of model would be appropriate for any particular study. Although the literature diverge on this point, it is possible that misspecification can have consequences for the estimation of the fixed effects (Lindsey & Lindsey, 2000; Litière et al., 2007).

When comparing models resting on different assumption, it is useful to apply models with varying parametric assumptions in addition to nonparametric models. We have presented a Poisson-gamma model as an additional tool for the criminologist's toolbox. This model has a number of additional

attractive properties, such as not being sensitive to start values, and being very quick to compute. This is beneficial because parametric models are more efficient if the mixture distribution is reasonable (Agresti et al., 2004). We also propose that the gamma distributed heterogeneity is a theoretically reasonable mixture distribution, and is therefore easier to give an explicit interpretation. Since the Poisson-gamma model is quick to compute, it can be useful to get start values for the more computational intensive nonparametric model.

We have compared the results from several models resting on different assumptions. Our data are highly overdispersed, thus the negative binomial normal model gave the best fit. Considering the Poisson models, specifying the random effect as normal-, gamma or nonparametric did not have much impact on the estimates for age, but did affect the estimates for the intercepts and sex. Given that a nonparametric model can approximate any true distribution, we put greater trust in the estimates from this model. It is then interesting that the Poisson-gamma model gave more similar result as the Poisson-NP model than the Poisson-normal model did. Misspecification of the random effect can then have substantive consequences, although perhaps not dramatic. Although the intercept is often not of substantive interest, any further calculations which include the intercept might be sensitive to specification of the random effect.

The distribution of heterogeneity might be of theoretical relevance in its own right. If there are meaningful clusters, as some suggest (Moffitt, 2006), a discrete distribution might be preferable to a continuous one. All models suggested a skewed distribution. As the random effects in the Poisson-normal are on the log scale, a large negative intercept suggests that a large majority of the individuals are crime-prone indistinguishable from zero (on the natural scale), but there is a small tail towards higher values of criminal dispositions. Our Poisson-gamma model estimated a gamma distribution with a very small shape-parameter, which would imply a similar distribution: most persons indistinguishable from zero but with a small tail to the right. Also the Poisson-NP model indicates that a skewed distribution is reasonable. The residual plots suggest that a highly skewed continuous distribution, such as the gamma distribution, is not unreasonable for our data.

The results show the importance of comparing alternative models for the same data, but also the difficulties of concluding on specification of the random effects. Improved relative fit can be achieved in several ways, and overdispersion appeared to be important in this respect for our data, so that better fit does not necessarily imply better estimates for the fixed effects. Accounting for over dispersion may improve the fit considerably without affecting the estimates much. Also; nonparametric

distribution should be more reliable when the true distribution is unknown, but is less efficient than correctly specified parametric distribution, and thus may give worse fit. Nevertheless, comparing alternative models and inspecting residuals is important in any empirical setting. If the models give different results, there are reasons for concern. In our study, the results did not diverge dramatically, and the largest difference was for the intercept, so the substantive conclusion would be similar even in presence of some bias.

## References

- Agresti, A., Caffo, B., & Ohman-Strickland, P. (2004). Examples in which misspecification of a random effects distribution reduced efficiency and possible remedies, *Computational statistics and data analysis*, 47: 639-653.
- Blumstein, A., Cohen, J., Roth, J., & Visher, C. (eds.). (1986). *Criminal careers and "career criminals"*. Washington D.C.: National Academy Press.
- Booth, J. G., Casella, G., Fridl, H., & Hobert, J. P. (2003). Negative Binomial Loglinear Mixed Models, *Statistical Modelling*, 3/3: 179-191.
- D'Unger, A., Land, K. C., McCall, P. L., & Nagin, D. S. (1998). How Many Latent Classes of Delinquent/Criminal Careers? Results from Mixed Poisson Regression Analyses, *American Journal of Sociology*, 103: 1593-1631.
- Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime*. Stanford, Calif., Stanford University Press.
- Greenberg, D. F. (1991). Modeling criminal careers, *Criminology*, 29: 17-46.
- Hipp, J. R., & Bauer, D. J. (2006). Local solutions in the estimation of growth mixture models, *Psychological Methods*, 11: 36-53.
- Jones, B. L., Nagin, D. S., & Roeder, K. (2001). A SAS Procedure Based on Mixture Models for Estimating Developmental Trajectories, *Sociological Methods and Research*, 29: 374-394.
- Land, K. C., McCall, P. L., & Nagin, D. S. (1996). A comparison of poisson, negative binomial and semiparametric mixed poisson regression models, *Sociological Methods and Research*, 24: 387-442.
- Laub, J. H., & Sampson, R. J. (2003). *Shared beginnings, divergent lives: delinquent boys to age 70*. Cambridge, Massachusetts, Harvard University Press.
- Lindsey, P. J., & Lindsey, J. K. (2000). Diagnostic tools for random effects in the repeated measures growth curve model, *Computational statistics and data analysis*, 33: 79-100.
- Litière, S., Alonso, A., & Molenberghs, G. (2007). Type I and type II error under random-effects misspecification in generalized linear mixed models, *Biometrics*, 63: 1038-1044.
- Lykken, D. T. (1995). *The antisocial personalities*. Hillsdale, New Jersey and Hove, UK, Lawrence Erlbaum Associates, Publishers.
- Moffitt, T. E. (2006). Life-course persistent versus adolescent-limited antisocial behavior. In D. Cicchetti & D. J. Cohen (eds) *Developmental psychopathology*. (Vol. 3). N.J., John Wiley, pp. 570-598
- Molenberghs, G., Verbeke, G., & Demétrio, C. G. B. (2007). An extended random-effects approach to modeling repeated, overdispersed count data, *Lifetime Data Anal*, 13.
- Nagin, D. S., & Land, K. C. (1993). Age, criminal careers, and population heterogeneity: specification and estimation of a nonparametric, mixed poisson model, *Criminology*, 31: 327-362.

- Nagin, D. S., & Tremblay, R. E. (2005). What has been learned from group-based trajectory modeling? Examples from physical aggression and other problem behaviors. In R. J. Sampson & J. H. Laub (eds) *Developmental criminology and its discontents: Trajectories of crime from childhood to old age*. (Vol. 602). Thousand Oaks-London-New Delhi, Sage publications, pp. 82-117
- Rolph, J. E., Adams, J. L., & McGuigan, K. A. (2007). Identifying Malpractice-Prone Physicians, *Journal of Empirical Legal Studies*, 4: 125-153.
- Rowe, D. C., Osgood, D. W., & Nicewander, W. A. (1990). A latent trait approach to unifying criminal careers, *Criminology*, 28: 237-270.
- Sampson, R. J., & Laub, J. H. (2005). Seductions of method: Rejoinder to Nagin and Tremblay's "Developmental trajectory groups: Fact or fiction?" *Criminology*, 43: 905-913.
- Skardhamar, T. (2009). Family dissolution and children's criminal careers, *European Journal of Criminology*, 6: 203–23
- Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling. Multilevel, longitudinal, and structural equation models*. Boca Raton, Chapman & Hall/CRC.
- Verbeke, G., & Lesaffre, E. (1997). The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data, *Computational statistics and data analysis*, 23: 541-556.

## Script to estimate the Poisson-gamma model using R

R is available for free at [www.r-project.org](http://www.r-project.org), and the following script use only the R-function nlm for optimization. The following contain an example run followed by the script.

```
# Written by Simen Gan Schweder, 18.09.2008
# Example run using publicly available example data available at
# http://www.andrew.cmu.edu/user/bjones/sas/zip.sas

data <- scan("C:/CambridgeData.txt", dec='.', skip=1);
data <- matrix(data, byrow=T, ncol=49)
fit <- pgfit("y ~ x + I(x^2) + I(x^3) + z1 + z2", beta=c(-1,-1,0.2,-0.9,0,0.8,0.7),
            y=data[,2:24], nofHomogenousTerms=2, z1=data[,48], z2=data[,49], x=data[,25:47]);
#Display summary info
pgfit.summary(fit)

# The null-inflated version of the same as the previous model
# (Just add one parameter to beta...)
fit <- pgfit("y ~ x + I(x^2) + I(x^3) + z1 + z2", beta=c(0.3, -1,-1,0.3,-0.8,0.1,0.8,0.7),
            y=data[,2:24], nofHomogenousTerms=2, z1=data[,48], z2=data[,49], x=data[,25:47]);
#Summary
pgfit.summary(fit)

#####
#-loglikelihood, with shape=rate for Poisson-gamma model
pgfit.ll <- function(start, y, ypluss, modelmatrix, n, pp, nofHetrogenousTerms,
                    nofHomogenousTerms ) {
  beta <- pgfit.buildCoefficientsFromBeta(
    beta=start,
    nofHetrogenous=nofHetrogenousTerms,
    nofHomogenous=nofHomogenousTerms,
    time=pp
  );
  s <- pgfit.buildsummationmatrix(beta);
  mu <- exp(modelmatrix %*% s);
  mupluss <- apply(mu, 1, sum);
  ymu <- apply(y*log(mu)-log(factorial(y)), 1, sum);
  -sum( ymu
    + lgamma(ypluss + beta$alpha)
    - lgamma(beta$alpha)
    + beta$alpha*(log(beta$alpha) -log(mupluss + beta$alpha))
    - ypluss*(log(mupluss+beta$alpha)));
}
```



```

#####
#loglikelihood, med zero-inflated model
pgfit.ll.nullinflated <- function(start, y, ypluss, modelmatrix, n, pp,
    nofHetrogenousTerms, nofHomogenousTerms ) {
    beta <- pgfit.buildCoefficientsFromBeta(
        start,
        nofHetrogenous=nofHetrogenousTerms,
        nofHomogenous=nofHomogenousTerms,
        time=pp
    );

    s <- pgfit.buildsummationmatrix(beta);
    mu <- exp(modelmatrix %*% s);
    mupluss <- apply(mu, 1, sum);
    ymu <- apply(y*log(mu)-log(factorial(y)),1,sum);
    pii <- beta$pii;

    -sum( (ypluss == 0)*log(1-pii+pii*(beta$alpha/(beta$alpha+mupluss))^beta$alpha)
        +(ypluss > 0)*(log(pii) + ymu
            + lgamma(ypluss + beta$alpha)
            - lgamma(beta$alpha)
            + beta$alpha*(log(beta$alpha) -log(mupluss + beta$alpha))
            - ypluss*(log(mupluss+beta$alpha)))
    );
}

#####
#Internal method, just assembles the coefficients and some extra info in a standard way.
pgfit.buildCoefficients <- function(alpha=NULL, pii=NULL, intercept=NULL, hetrogenous=NULL,
homogenous=NULL, nofHetrogenous=0, nofHomogenous=0,time=NULL) {
    list(
        pii=pii,
        alpha=alpha,
        intercept=intercept,
        hetrogenous=hetrogenous,
        homogenous=homogenous,
        nofHetrogenous=nofHetrogenous,
        nofHomogenous=nofHomogenous,
        time=time
    );
}

```

```

#####
#Internal method, builds coefficients from the result of an nlm-call + some extra info.
#Note, the alpha is exp(beta[1]), if pii is present it is assumed to be at the first position
#in beta and is inserted as logit(pii).
#Dette blir galt når det ikke er noen homogene variabler.
pgfit.buildCoefficientsFromBeta <- function(beta, nofHetrogenous=0, nofHomogenous=0,time=NULL)
{
  if(nofHomogenous > 0) {
    if(length(beta)==(2+nofHetrogenous+nofHomogenous)) {
      pgfit.buildCoefficients(
        alpha=exp(beta[1]),
        intercept=beta[2],
        hetrogenous=beta[3:(2+nofHetrogenous)],
        homogenous=beta[(3+nofHetrogenous):(nofHetrogenous + 2 +
          nofHomogenous)],
        nofHetrogenous=nofHetrogenous,
        nofHomogenous=nofHomogenous,
        time=time
      );
    } else {
      pgfit.buildCoefficients(
        pii=logit(beta[1]),
        alpha=exp(beta[2]),
        intercept=beta[3],
        hetrogenous=beta[4:(3+nofHetrogenous)],
        homogenous=beta[(4+nofHetrogenous):(nofHetrogenous + 3 +
          nofHomogenous)],
        nofHetrogenous=nofHetrogenous,
        nofHomogenous=nofHomogenous,
        time=time
      );
    }
  } else {
    if(length(beta)==(2+nofHetrogenous+nofHomogenous)) {
      pgfit.buildCoefficients(
        alpha=exp(beta[1]),
        intercept=beta[2],
        hetrogenous=beta[3:(2+nofHetrogenous)],
        homogenous="NONE",
        nofHetrogenous=nofHetrogenous,
        nofHomogenous=nofHomogenous,
        time=time
      );
    } else {
      pgfit.buildCoefficients(

```

```

        pii=logit(beta[1]),
        alpha=exp(beta[2]),
        intercept=beta[3],
        hetrogenous=beta[4:(3+nofHetrogenous)],
        homogenous="NONE",
        nofHetrogenous=nofHetrogenous,
        nofHomogenous=nofHomogenous,
        time=time
    );
};
}
}

#####
#Builds the summation matrix s, "modelmatrix %*% s" gives mu.
pgfit.buildsummationmatrix <- function(beta) {
    #print(beta);
    #Builds a temporary summation-matrix
    if(beta$nofHetrogenous > 0 || beta$nofHomogenous > 0) {
        s <- diag(rep(beta$hetrogenous[1],beta$time)); #First hetro
        if(beta$nofHetrogenous > 1) {
            for(j in 2:beta$nofHetrogenous) {#The rest of the hetrogenous variables
                s <- rbind(s, diag(rep(beta$hetrogenous[j],beta$time)));
            }
        }
        if(beta$nofHomogenous > 0) {
            #the homogenous variables
            s <- rbind(s, beta$homogenous %*% t(rep(1, beta$time)));
        }
        #the intercept
        s <- rbind(rep(beta$intercept,beta$time), s); #the intercept
    } else {
        s <- rep(beta$intercept,beta$time);
    }
    s
}

#####
#Internal method, returns the standard logit function.
logit <- function(x) {
    1/(1+exp(-x));
}
inverselogit <- function(y) {
    -log(1/y -1)
}

```

```

#Prints summary information on the fit-object returned by pgfit.
pgfit.summary <- function(fit) {
  print(list("method:"=fit$method, "minimum:"=fit$minimum, "gradient:"=fit$gradient,
"code:"=fit$code, "coefficients:"=fit$coefficients));
  print("The alfa value is log(true alfa), the pii value is logit(true pii)");
}

#####
#Fits the model.
#Uses the standardmodel if number of startparameters(beta) = 2+nofHetrogenous + nofHomogenous,
#and the nullinflated model if startparameters = 3+nofHetrogenous+nofHomogenous.
pgfit <- function(fun, beta=NULL, y, nofHomogenousTerms, ...) {
  #Build the dataframe
  variables <- list(...);
  df <- data.frame(y=I(y));
  for(i in 1:length(variables)) {
    df[[names(variables)[[i]]]] <- variables[[i]];
  }
  #Build the modelmatrix
  form <- as.formula(fun);
  modelmatrix <- model.matrix(form, data=df);

  p <- length(y[,1]);
  nofHetrogenousTerms = floor((length(modelmatrix[1,]) - 1 - nofHomogenousTerms)/p);
  n <- length(y[,1]);
  ypluss <- apply(y, 1, sum)

  #Test for correct start params and fit the appropriate model
  if(length(beta) == (2 + nofHetrogenousTerms + nofHomogenousTerms)) {
    method <- "standard poisson-gamma model"
    print("Running standard poisson-gamma model");
    temp <- nlm(pgfit.ll, p=beta, hessian=TRUE, iterlim=1000,
y=y, ypluss=ypluss, modelmatrix=modelmatrix, n=n, pp=p,
nofHetrogenousTerms=nofHetrogenousTerms,
nofHomogenousTerms=nofHomogenousTerms);
  } else if(length(beta) == (3 + nofHetrogenousTerms + nofHomogenousTerms)) {
    method <- "null-inflated poisson-gamma model"
    print("Running null-inflated poisson-gamma model");
    temp <- nlm(pgfit.ll.nullinflated, p=beta, hessian=TRUE, iterlim=1000,
y=y, ypluss=ypluss, modelmatrix=modelmatrix, n=n, pp=p,
nofHetrogenousTerms=nofHetrogenousTerms,
nofHomogenousTerms=nofHomogenousTerms);
  } else {
    stop("Wrong number of start-parameters, should be
(alpha,intercept,hetrogenous_1,...,hetrogenous_n,homogenous_1,...,homogenous_m)");
  }
  if(temp$code > 1) print("WARNING: Code returned from nlm > 1");
  list(method=method,

```

```
formula=fun,  
minimum=I(temp$minimum),  
estimate=I(temp$estimate),  
gradient=I(temp$gradient),  
hessian=I(temp$hessian),  
code=I(temp$code),  
iterations=I(temp$iterations),  
response=I(y),  
modelmatrix=I(modelmatrix),  
coefficients=I(pgfit.buildCoefficientsFromBeta(beta=temp$estimate,  
nofHetrogenous=nofHetrogenousTerms, nofHomogenous=nofHomogenousTerms,time=p))  
);  
}
```