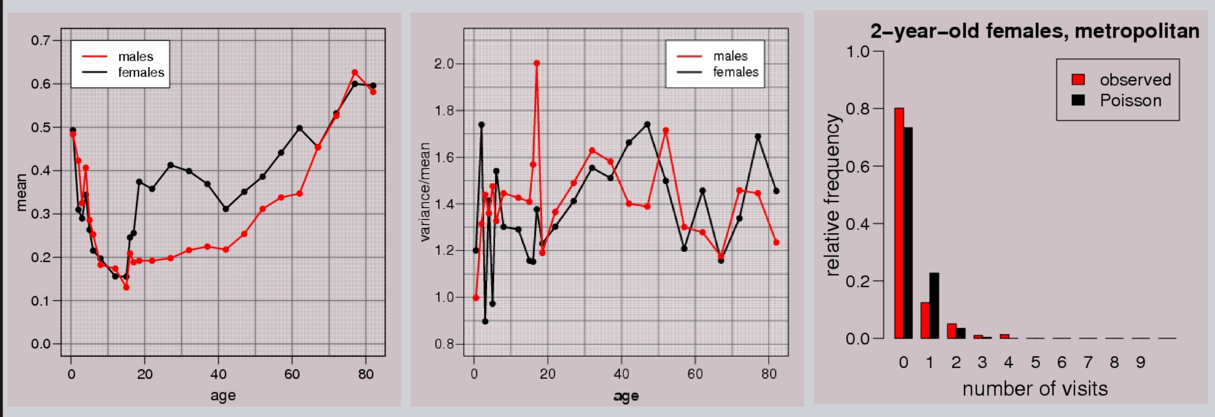
Comparison of Methods for Modelling Dispersion in Additive Negative Binomial Regression Using a Community Health Application

Andreas Berzel, Institute for Statistics and Econometrics, University of Goettingen

The Data

We consider the distribution for the number of visits to the doctor and model its dependence on a number of demographic factors.
The database, from the Australian National Health Survey 1995, covers 53828 individuals. The variable of interest is the number of visits to a doctor in the preceding two weeks (0,1,2,...,9,10+). The covariates considered are age (in 25 categories), gender and geographical area (in 3 categories). We focus here on the observations for metropolitan areas (in total 36207).

The data are characterized by a high proportion of zero counts, and marked overdispersion relative to the Poisson distribution.



Preliminary analysis and model selection revealed that the negative binomial is appropriate for these counts (Berzel, Heller and Zucchini, 2004).

The Model

The negative binomial distribution was parameterized as follows (Cameron and Trivedi, 1998):

$$p_X(x) = \begin{cases} \frac{\Gamma(x+1/\eta)}{\Gamma(1/\eta)\Gamma(x+1)} \left(\frac{1}{1+\eta\mu}\right)^{1/\eta} \left(\frac{\eta\mu}{1+\eta\mu}\right)^{x} & x = 0,1,2,\ldots \\ 0 & \text{otherwise} \end{cases}$$

$$E(X) = \mu \qquad Var(X) = \mu(1+\eta\mu)$$

We modelled the parameters of the negative binomial distribution in dependence on the covariates, i.e.

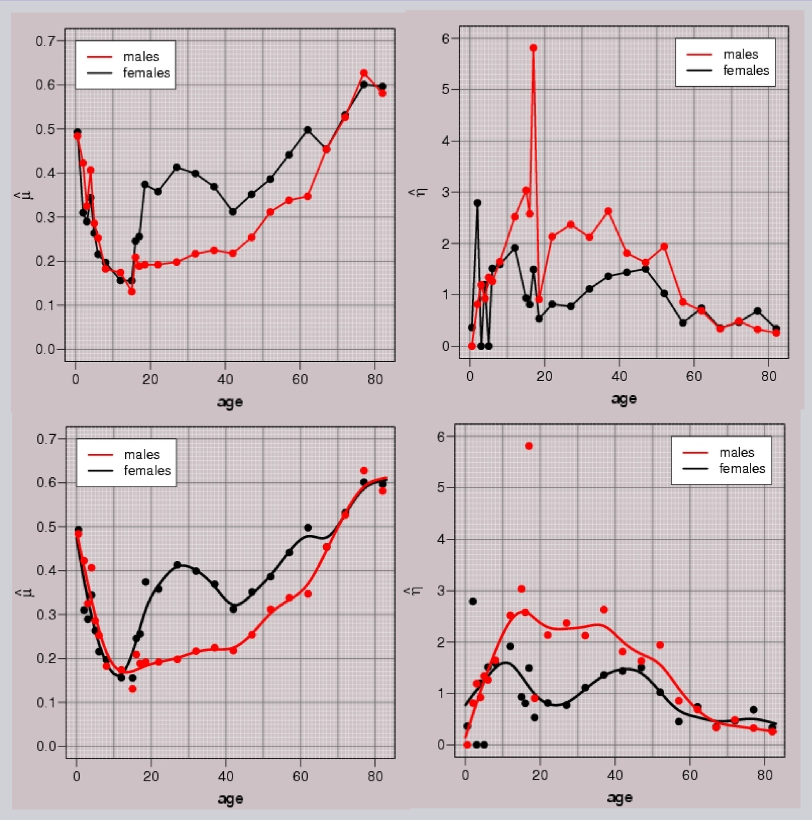
$$\mu = \mu(\mathbf{x}) \quad \text{and} \quad \eta = \eta(\mathbf{x})$$

The censoring of the response variable was taken into account in the simple modelling approach but was ignored in more complex models. (Only 9 out of 36,207 observations belong to the 10+ category.)

Simple Modelling Approach

The raw data were partitioned into 50 age-gender subgroups. We fitted a negative binomial distribution to each subgroup using the software R (Ihaka and Gentleman, 1996) to numerically maximize the likelihood function.

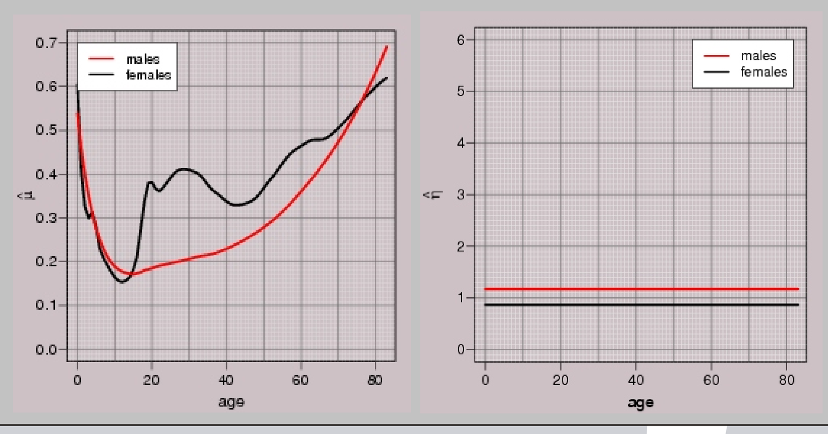
In a second step we smoothed the parameter estimates using weighted smoothing splines (using the R-function "smooth.spline()") with weights proportional to the number of observations in the respective subgroups (Green and Silverman, 1994). The penalty factor which minimized the AIC was selected for males; for females a local minimum of the AIC was chosen in order to obtain a smoother function.



Modelling Dispersion in Additive Negative Binomial Regression

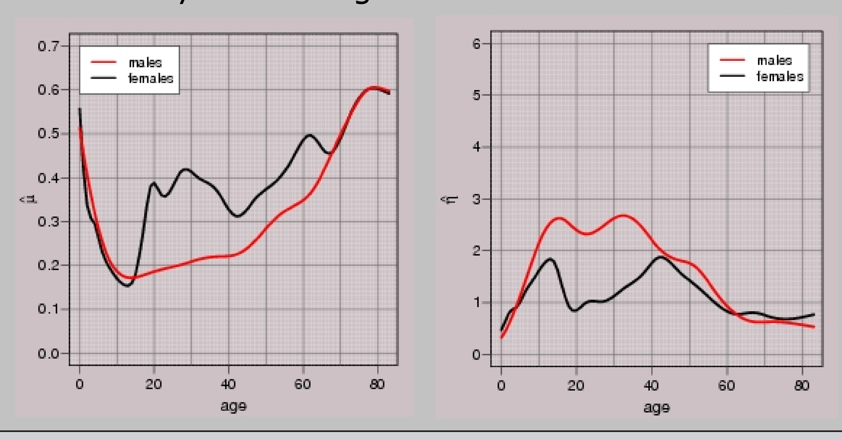
Negative Binomial Additive Model (NBAM)

Thurston, Wand and Wiencke (2000) proposed an extension of the generalized additive model to handle negative binomial responses. Parameter estimation is done using an alternating profile likelihood algorithm that switches between fitting a generalized additive model with negative binomial response and fixed shape parameter, and maximum likelihood estimation of the shape parameter given the current estimates of the mean. The former step uses a local scoring algorithm incorporating a backfitting algorithm for a weighted additive model and weighted local polynomial smoothing. As our application has just one additive term, we implemented a simplified version of that algorithm in R, and choose the degrees of freedom which minimized the AIC.



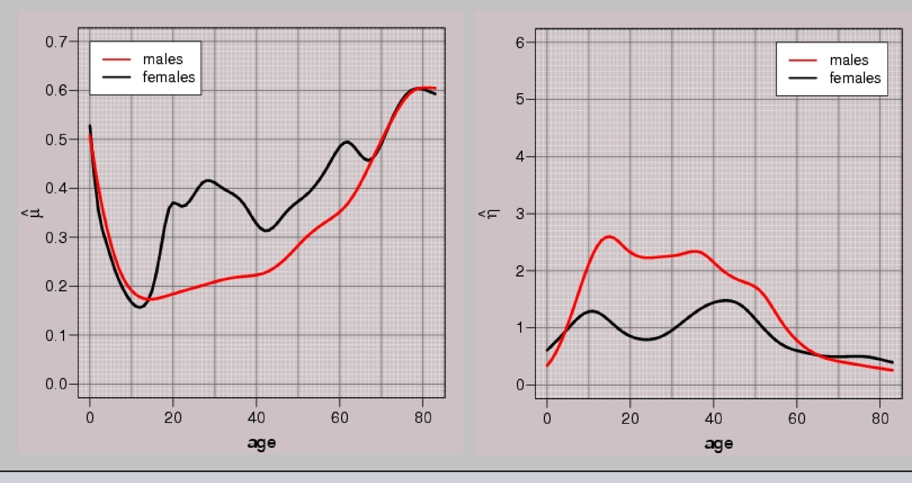
Mean And Dispersion Additive Model (MADAM)

Rigby and Stasinopoulos (1996) introduced a general flexible model for the mean and variance of a dependent variable. They model the variance as a product of a dispersion parameter and a known variance function of the mean, and use a semi-parametric additive model to describe the dependence of the parameters on explanatory variables. The model can be fitted by maximization of the penalized extended quasi-likelihood, or by pseudo-maximization of the penalized normal likelihood. This involves a successive relaxation algorithm that alternates between fitting the mean and the dispersion, together with a local scoring algorithm which combines a generalized linear model procedure and a backfitting algorithm incorporating a cubic spline smoother.
The algorithm is implemented in the GLIM4 macro MADAM, but here we applied a simplified version of that algorithm using R. We fitted a mean and dispersion additive model assuming an overdispersed Poisson distribution and, in a second step, computed the negative binomial parameters from the mean and dispersion estimates. The degrees of freedom were again estimated by minimizing the AIC.



GAM for Location, Scale and Shape (GAMLSS)

Generalized additive models for location, scale and shape, introduced by Rigby and Stasinopoulos (2001), are a generalization of the generalized additive model class, where the exponential family distribution assumption is relaxed and replaced by a general distribution family. The mean and the other parameters of the distribution are modelled as linear parametric, or additive non-parametric, functions of the covariates.
Maximum (penalized) likelihood estimation is used applying a Newton-Raphson/Fisher scoring algorithm with the backfitting algorithm for the additive components. (One of the two optional algorithms for model fitting is in fact based on MADAM.)
We used the R-package GAMLSS (kindly provided by Rigby and Stasinopoulos) to fit the model. Again the equivalent degrees of freedom were obtained by minimizing the AIC for males and by choosing a local minimum of the AIC for females.

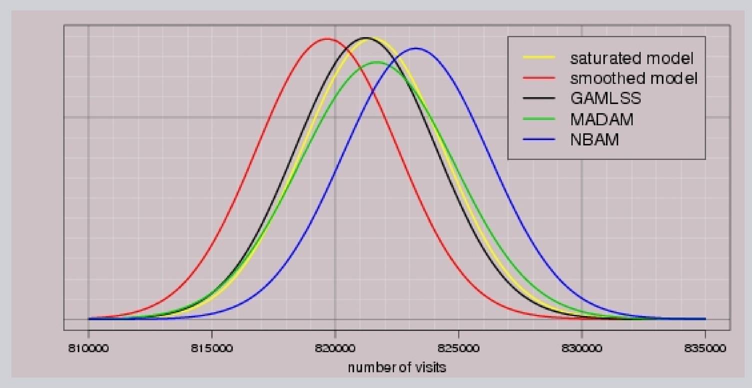


Note: A further possibility of modelling dispersion in parametric or nonparametric negative binomial regression is using Vector Generalized Additive Models that were introduced by Yee and Wild (1996). For fitting these models the R-package VGAM is provided.

Practical Implications

To illustrate one practical implication of the model selection, we applied each of the models to predict the distribution of the yearly number of consultations for the population of the Sydney suburb Ryde (total population in 2001: 94243) using the method outlined in Berzel, Heller and Zucchini (2004).

model	μ	σ
saturated	821,514	2,870
smooth	819,669	2,872
NBAM	823,267	2,975
MADAM	821,681	3,135
GAMLSS	821,235	2,864



Obviously, the model selection affects the results of practical applications.Above all, the different degrees of smoothness of the models lead to different predicted values.The high value of the predicted mean for the NBAM is due to deficiencies in extrapolating the model on the edges.

Results and Conclusions

The following table gives the AIC-selected model, separately for males and females, for each of the above model families, indicating the log-likelihood (logL), the equivalent degrees of freedom (edf) and the respective value of the AIC.

	males			females		
model	logL	edf	AIC	logL	edf	AIC
smooth	-11354	50	22808	-14351	50	28803
simple	-11370	19.6	22780	-14381	22.5	28808
NBAM	-11455	6.3	22923	-14405	15.9	28842
MADAM	-11376	19.2	22791	-14385	24.9	28820
GAMLSS	-11374	18.9	22785	-14377	27.2	28809

Clearly, the negative binomial additive model, which fits a constant scale parameter over the subgroups, is not flexible enough for these data.
The GAMLSS provides a better fit than does the MADAM. The latter can be regarded as a GAMLSS model in which the scale parameter of the NBD is being computed from a smooth of the dispersion parameter instead of being smoothed directly.
Surprisingly, the simple modelling approach leads to a lower AIC than the GAMLSS. This might be due to the high proportion of zero counts, or the censoring of the target variable not taken into account in the GAMLSS.
According to this table, the saturated model provides the best fit for females, but this is only due to the fact that for the smooth model and for the GAMLSS a local minimum of the AIC was selected in order to obtain smoother functions.

References
[1] Berzel A., Heller G. and Zucchini W. (2004). *Estimating the Number of Visits to the Doctor*. Submitted.
[2] Cameron A.C. and Trivedi P.K. (1998). *Regression Analysis of Count Data*. Cambridge et al.: Cambridge University Press.
[3] Green P.J. and Silverman B.W. (1994). *Nonparametric Regression and Generalized Linear Models: A roughness penalty approach.* London et al.: Chapman & Hall.
[4] Ihaka R. and Gentleman R. (1996). R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics* **5**(3), 299-314.
[5] Rigby R.A. and Stasinopoulos M.D. (1996). Mean and Dispersion Additive Models. In *Statistical Theory and Computational Aspects of Smoothing.* (ed: Härdle W. And Schimek M.G.), Physica-Verlag, 215-230.
[6] Rigby R.A. and Stasinopoulos M.D. (2001). *Generalised Additive Models for Location, Scale and Shape*. Submitted for publication.
[7] Thurston S.W., Wand M.P. and Wiencke J.K. (2000). Negative Binomial Additive Models. *Biometrics* **56**, 139-144.
[8] Yee T.W. and Wild C.J. (1996). Vector Generalized Additive Models. *Journal of the Royal Statistical Society, Series B* **58**, 481-493.