

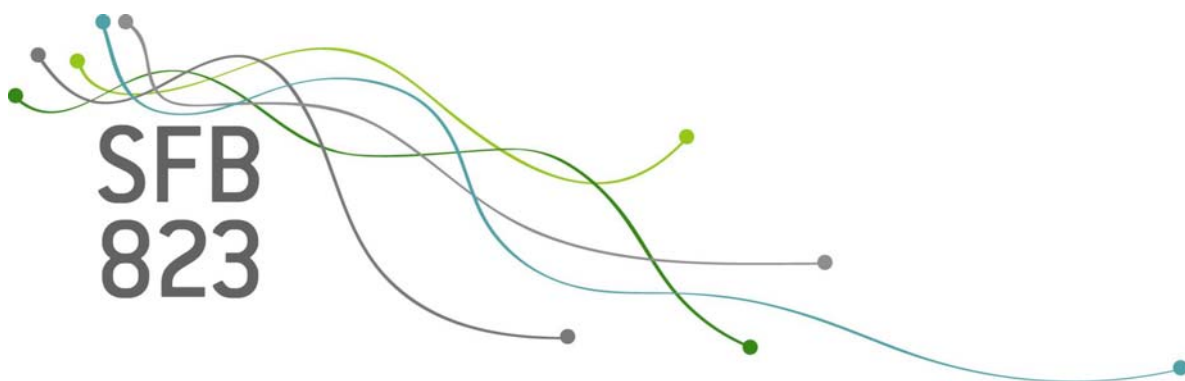
SFB
823

Consistency for the negative binomial regression with fixed covariate

Rafael Weißbach, Lucas Radloff

Nr. 26/2018

Discussion Paper



Consistency for the Negative Binomial Regression with Fixed Covariate

Rafael Weißbach and Lucas Radloff*

Chair in Statistics and Econometrics, University of Rostock, Germany

Abstract

We model an overdispersed count as a dependent measurement, by means of the Negative Binomial distribution. We consider quantitative regressors that are fixed by design. The expectation of the dependent variable is assumed to be a known function of a linear combination involving regressors and their coefficients. In the NB1-parametrization of the negative binomial distribution, the variance is a linear function of the expectation, inflated by the dispersion parameter, and not a generalized linear model. We apply a general result of Bradley and Gart (1962) to derive weak consistency and asymptotic normality of the maximum likelihood estimator for all parameters. To this end, we show (i) how to bound the logarithmic density by a function that is linear in the outcome of the dependent variable, independently of the parameter. Furthermore (ii) the positive definiteness of the matrix related to the Fisher information is shown with the Cauchy-Schwarz inequality.

Keywords: negative binomial regression, NB1-parametrization, fixed covariates, weak consistency, asymptotic normality

2000 MSC: 62F12, 62J02

*Corresponding author: Prof. Dr. Rafael Weißbach, Chair in Statistics and Econometrics, Faculty for Economic and Social Sciences, University of Rostock, 18051 Rostock, Germany, Phone: +49-381-4984428, Fax: +49-381-4984401.

Email address: rafael.weissbach@uni-rostock.de (Rafael Weißbach and Lucas Radloff)

URL: <http://www.wiwi.uni-rostock.de/vwl/statistik/> (Rafael Weißbach and Lucas Radloff)

1. Introduction

When modelling count data, overdispersion occasionally prevents applying a Poisson regression. Examples include numbers of sick pupils in schools (Weißbach et al., 2015), of insolvencies in credit portfolios (Weißbach et al., 2009) or of bookings on diverse occasions in a business context. Overdispersion is commonly modelled with a negative binomial distribution (see Cameron and Trivedi, 2008, Chap. 3). Different parametrizations of this distribution result from different relations between the expectation and the variance. In the NB1-model, the variance is linear in the expectation. As a consequence, it can not be represented as a member of the linear exponential family, and, hence, is not included in the class of generalized linear models (GLM) (see Cameron and Trivedi, 2008, Sec. 3.3). In contrast, in the NB2-model the relation is quadratic and it is a GLM, but only if one assumes the dispersion parameter to be known (see Hilbe, 2011, Chap. 8). Parametrizations with other polynomial relations are also reported (see Winkelmann, 2000, Sec. 2.3).

In this paper, we establish consistency of the maximum likelihood estimator (MLE) for the NB1-regression-model in the case of fixed covariates, i.e. when the regressors are deterministic vectors instead of random variables. A less efficient method, but more robust to model misspecification is the method of moments, with asymptotic properties discussed in Moore (1986). The central limit theorems for identically distributed random variables do not apply. If a Poisson regression were applicable, consistency in the case of fixed covariates could have been derived using results from Fahrmeir and Kaufmann (1985). They provide conditions for the consistency of the MLE, as the Poisson regression is a GLM. In order to prove consistency in such a case of independent, but not identically distributed observations, Kremer et al. (2014) exploit of the concavity of the criterion function. In contrast,

as classically done, we use properties of the densities' derivatives. For the NB1-model, we apply a theorem of Bradley and Gart (1962) and validate the provided conditions. It should be mentioned, that the method of proof also applies to the Poisson regression, for which the derivatives are fairly simple and the representations of Fisher matrices are explicit.

2. Model and Notation

2.1. Negative binomial distribution

The negative binomial distribution is a discrete distribution for which several parametrizations are possible (see Johnson et al., 2005, Chap. 5). Throughout this paper, we will work with the following definition, which is usually referred to as the NB1-model:

Definition 1. *We call a probability distribution with the probability mass function*

$$f(y; \lambda, \sigma) = \frac{\Gamma(\frac{\lambda}{\sigma} + y)}{y! \Gamma(\frac{\lambda}{\sigma})} \left(\frac{1}{1 + \sigma} \right)^{\frac{\lambda}{\sigma}} \left(\frac{\sigma}{1 + \sigma} \right)^y, \quad y \in \mathbb{N}_0 \quad (1)$$

a negative binomial distribution with parameters $\lambda > 0$ and $\sigma > 0$. It is denoted by $\mathcal{NB}(\lambda, \sigma)$.

For $X \sim \mathcal{NB}(\lambda, \sigma)$ we have (see Winkelmann, 2000, Sec. 2.3.1)

$$\mathbb{E}X = \lambda \quad \text{and} \quad \text{Var}(X) = \lambda(1 + \sigma).$$

The dispersion parameter σ scales the variance, linearly dependent on λ . Clearly, we have $\mathbb{E}X < \text{Var}(X)$, which constitutes modelling overdispersion. In contrast, the Poisson model with $\mathbb{E}X = \text{Var}(X)$ arises for $\sigma \rightarrow 0$.

It will occasionally prove helpful to rewrite

$$\frac{\Gamma(a + y)}{\Gamma(a)} = \prod_{j=0}^{y-1} (a + j) \quad \forall a \in \mathbb{R}_+, y \in \mathbb{N}. \quad (2)$$

2.2. Regression model and further assumptions

Let $Y_i, i \in \mathbb{N}$, be random variables defined on the probability space $(\Omega, \mathfrak{A}, P)$ which describe the count data. Further let $\mathbf{z}_i = (z_{i1}, \dots, z_{ip})^T \in \mathcal{Z} \subset \mathbb{R}^p, i \in \mathbb{N}$, denote the deterministic covariate vectors. The negative binomial regression model is defined by the following assumptions.

NBR1 The random variables $Y_i, i \in \mathbb{N}$, are stochastically independent.

NBR2 $Y_i \sim \mathcal{NB}(\lambda_i^0, \sigma_0), i \in \mathbb{N}$. (Hence, the dispersion parameter $\sigma_0 \in \mathcal{S} \subset \mathbb{R}_+$ is identical for all $i \in \mathbb{N}$.)

NBR3 The expected values are given by the equations

$$\lambda_i^0 = \mathbb{E}Y_i = g(\mathbf{z}_i, \boldsymbol{\beta}_0), i \in \mathbb{N}, \quad (3)$$

where g is a function to be specified and $\boldsymbol{\beta}_0 \in \mathcal{B} \subset \mathbb{R}^p$ is an unknown parameter vector.

We consider the infinite sequence of random variables $(Y_i)_{i \in \mathbb{N}}$, but for the estimation of parameters always only a finite sample Y_1, \dots, Y_n is available. We are interested in the asymptotic behaviour of the maximum likelihood estimator when increasing the sample size, i.e. the case $n \rightarrow \infty$. In order to establish weak consistency, and asymptotic normality later in this paper, we introduce some further assumptions:

CN1 The set \mathcal{Z} of all possible regressor outcomes is compact. The sets \mathcal{B} and $\mathcal{S} \subset (0, \infty)$ of all possible parameters are compact intervals of the corresponding dimensions.

CN2 The equation (3) has the special form

$$\lambda_i^0 = \mathbb{E}Y_i = g(\mathbf{z}_i^T \boldsymbol{\beta}_0), i \in \mathbb{N}, \quad (4)$$

where $g : \mathcal{G} \rightarrow \mathbb{R}_+$ with $\mathcal{G} := \{\mathbf{z}^T \boldsymbol{\beta} \in \mathbb{R} | \mathbf{z} \in \mathcal{Z}, \boldsymbol{\beta} \in \mathcal{B}\}$ is a three times continuously differentiable function of which the first derivative has no zeros. In particular, it is strictly monotonous.

CN3 Denote by $\lambda_{\min}[A]$ the smallest eigenvalue of the matrix A . For the sequence of regressors $(\mathbf{z}_i)_i$ it is valid that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \lambda_{\min} \left[\sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^T \right] > 0. \quad (5)$$

CN4 For all open sets $A \subset \mathcal{Z}$ there is a $c_A \in [0, 1]$, such that

$$\frac{1}{n} \sum_{i=1}^n \mathbf{1}_A(\mathbf{z}_i) \xrightarrow[n \rightarrow \infty]{} c_A.$$

The more specific model equation (4) contains the standard choice

$$g(\mathbf{z}, \boldsymbol{\beta}) = \exp(\mathbf{z}^T \boldsymbol{\beta}).$$

Assumption CN3 was inspired by Fahrmeir and Kaufmann (1985), who propose a similar condition in the context of a generalized linear model. Here, it ensures the asymptotic occurrence of the deterministic regressors $(\mathbf{z}_i)_i$. In order to see that such requirements are necessary, consider, for example, a 0-1-valued covariate. If this regressor, for $n \rightarrow \infty$, is zero or one only for a finite number of times, a parameter that describes the difference between occurrence and non-occurrence cannot be estimated consistently. To render the argument explicit, note that CN3 is actually not fulfilled in such a case. Let $p = 2$, with the first covariate being constantly one and a 0-1-valued second, i.e. $\mathbf{z}_i = (1, z_i)^T$ with $z_i \in \{0, 1\}$. Using the notation $k_n := \sum_{i=1}^n z_i$ for the number of ones occurring until n we obtain

$$\mathbf{z}_i \mathbf{z}_i^T = \begin{pmatrix} 1 & z_i \\ z_i & z_i^2 \end{pmatrix} = \begin{pmatrix} 1 & z_i \\ z_i & z_i \end{pmatrix} \text{ and thus } \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^T = \begin{pmatrix} n & k_n \\ k_n & k_n \end{pmatrix}.$$

Therefore, the smallest eigenvalue of $\sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^T$ can be calculated by:

$$\begin{aligned} \lambda_{\min} \left[\sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^T \right] &= \frac{n + k_n}{2} - \sqrt{\frac{(n + k_n)^2}{4} - (nk_n - k_n^2)} \\ &= \frac{1}{2} \left(n + k_n - \sqrt{(n - k_n)^2 + 4k_n^2} \right) \\ &\leq \begin{cases} \frac{1}{2} \left(n + k_n - \sqrt{(n - k_n)^2} \right) & = k_n \\ \frac{1}{2} \left(n + k_n - \sqrt{4k_n^2} \right) & = \frac{1}{2}(n - k_n) \end{cases} \end{aligned}$$

Thus the sequence of the smallest eigenvalues is actually bounded if k_n or $n - k_n$ is bounded and therefore, CN3 is not fulfilled. Condition CN4 is another claim to the asymptotic behaviour of the sequence of deterministic covariables. It ensures a property that would have been guaranteed for stochastic regressors by the law of large numbers.

In the following analysis, we assume Conditions NB1-3 and CN1-4 to hold. Often, we will use the notation $\boldsymbol{\theta} := (\boldsymbol{\beta}, \sigma)$ for a complete parameter vector (respectively $\boldsymbol{\theta}_0 := (\boldsymbol{\beta}_0, \sigma_0)$ for the complete true parameter vector) and $\Theta := \mathcal{B} \times \mathcal{S}$ for the corresponding parameter space. $f_i(\cdot; \boldsymbol{\theta}), i \in \mathbb{N}$, will denote negative binomial densities with expectation parameters $\lambda_i = g(\mathbf{z}_i^T \boldsymbol{\beta})$ and dispersion parameter σ . Hence, $f_i(\cdot; \boldsymbol{\theta}_0)$ is the probability mass function of Y_i . The model is smooth in the following sence.

Lemma 1. *For all $i \in \mathbb{N}$, for all $y \in \mathbb{N}_0$, for all $\boldsymbol{\theta} \in \Theta$ and for all $j, k, l = 1, \dots, p + 1$ the following partial derivatives exist,*

$$\frac{\partial \ln f_i}{\partial \theta_j}, \quad \frac{\partial^2 \ln f_i}{\partial \theta_j \partial \theta_k}, \quad \text{and} \quad \frac{\partial^3 \ln f_i}{\partial \theta_j \partial \theta_k \partial \theta_l}.$$

Proof: Considering (1), (2) and condition CN2, one can easily see that $\ln f_i, i \in \mathbb{N}$, are compositions of sufficiently often differentiable functions of the parameters and thus sufficiently often differentiable themselves. \square

2.3. Maximum likelihood estimation

Using (1) and (2) we obtain the log-likelihood function:

$$\begin{aligned} L(\boldsymbol{\beta}, \sigma; \mathbf{y}) &= \sum_{i=1}^n \left[\sum_{j=0}^{y_i-1} \ln \left(\frac{\lambda_i}{\sigma} + j \right) - \ln(y_i!) - \left(y_i + \frac{\lambda_i}{\sigma} \right) \ln(1 + \sigma) + y_i \ln(\sigma) \right] \\ &= \sum_{i=1}^n \left[\sum_{j=0}^{y_i-1} \ln \left(\frac{g(\mathbf{z}_i^T \boldsymbol{\beta})}{\sigma} + j \right) - \ln(y_i!) \right. \\ &\quad \left. - \left(y_i + \frac{g(\mathbf{z}_i^T \boldsymbol{\beta})}{\sigma} \right) \ln(1 + \sigma) + y_i \ln(\sigma) \right] \end{aligned}$$

This will be maximized with respect to $\boldsymbol{\beta}$ and σ simultaneously, so that the maximum likelihood estimator is defined as

$$(\hat{\boldsymbol{\beta}}, \hat{\sigma}) := \arg \max_{\boldsymbol{\beta}, \sigma} L(\boldsymbol{\beta}, \sigma; \mathbf{y}).$$

It is necessary for the estimators to solve the likelihood equations (using the chain rule):

$$\begin{aligned} \frac{\partial L(\boldsymbol{\beta}, \sigma; \mathbf{y})}{\partial \sigma} &= \sum_{i=1}^n \left[\sum_{j=0}^{y_i-1} \frac{-\lambda_i}{\sigma^2} + \frac{y_i}{\sigma} - \frac{y_i}{1+\sigma} - \frac{\lambda_i}{\sigma(1+\sigma)} \right. \\ &\quad \left. + \frac{\lambda_i}{\sigma^2} \ln(1+\sigma) \right] \Bigg|_{\lambda_i=g(\mathbf{z}_i^T \boldsymbol{\beta})} \\ &= \sum_{i=1}^n \left[- \sum_{j=0}^{y_i-1} \frac{\lambda_i}{\lambda_i \sigma + j \sigma^2} + \frac{y_i - \lambda_i}{\sigma(1+\sigma)} + \frac{\lambda_i}{\sigma^2} \ln(1+\sigma) \right] \Bigg|_{\lambda_i=g(\mathbf{z}_i^T \boldsymbol{\beta})} \stackrel{!}{=} 0 \\ \frac{\partial L(\boldsymbol{\beta}, \sigma; \mathbf{y})}{\partial \beta_k} &= \frac{\partial L(\boldsymbol{\beta}, \sigma; \mathbf{y})}{\partial \lambda_i} \Bigg|_{\lambda_i=g(\mathbf{z}_i^T \boldsymbol{\beta})} \cdot \frac{\partial \lambda_i}{\partial \beta_k} \\ &= \sum_{i=1}^n \left[\sum_{j=0}^{y_i-1} \frac{1}{\lambda_i + j \sigma} - \frac{\ln(1+\sigma)}{\sigma} \right] \Bigg|_{\lambda_i=g(\mathbf{z}_i^T \boldsymbol{\beta})} \\ &\quad \cdot g'(\mathbf{z}_i^T \boldsymbol{\beta}) z_{ik} \stackrel{!}{=} 0, \quad k = 1, \dots, p \end{aligned} \tag{6}$$

They do not have closed-form solutions, but are accessible by numerical methods.

3. Weak Consistency

3.1. Smoothness properties

Lemma 2 will help to prove ones, and claims that the partial derivatives of the log-densities are bounded by a linear function in y . Broadly spoken it ensures condition (D) from Wald (1949) in the notation of (Gourieroux and Monfort, 1995, Sec. 24.2).

Lemma 2. *For $k = 1, 2, 3$ and $l_1, \dots, l_k \in \{1, \dots, p+1\}$ there are constants $C, D < \infty$ independent of $i \in \mathbb{N}$ and $\boldsymbol{\theta} \in \Theta$, such that*

$$\left| \frac{\partial^k \ln f_i(y; \boldsymbol{\theta})}{\partial \theta_{l_1} \dots \partial \theta_{l_k}} \right| \leq Cy + D, \quad y \in \mathbb{N}_0.$$

For $k > 3$, the statement remains valid as long as continuous derivatives of the respective order exist, which basically depends on the differentiability of g .

For this lemma and for all other lemmas in the sequel with no display of the proof in the body of the paper, proofs are provided in Appendix A.

Lemma 3. (1) For all $i \in \mathbb{N}$ and $j, k = 1, \dots, p + 1$ the function series

$$\sum_{y=0}^{\infty} \frac{\partial f_i(y, \boldsymbol{\theta})}{\partial \theta_j} \quad \text{and} \quad \sum_{y=0}^{\infty} \frac{\partial^2 f_i(y, \boldsymbol{\theta})}{\partial \theta_j \partial \theta_k}$$

converge uniformly in $\boldsymbol{\theta} \in \Theta$.

(2) For all $i \in \mathbb{N}$ and $j, k, l = 1, \dots, p + 1$ there are functions $H_{ijkl}(y)$, independent of $\boldsymbol{\theta}$, and constants M_i , only dependent on i , such that

$$\left| \frac{\partial^3 \ln f_i(y, \boldsymbol{\theta})}{\partial \theta_j \partial \theta_k \partial \theta_l} \right| < H_{ijkl}(y) \quad \text{with} \quad \sum_{y \in \mathbb{N}_0} H_{ijkl}(y) f_i(y) < M_i.$$

(3) Furthermore there is a finite number M , such that for all $n \in \mathbb{N}$

$$\frac{1}{n} \sum_{i=1}^n M_i < M.$$

Proof: (1): Using the chain rule, the generalized product rule and formula (2) as the first partial derivative of the density with respect to σ there arises

$$\begin{aligned} \frac{\partial f_i(y; \boldsymbol{\beta}, \sigma)}{\partial \sigma} = & \frac{1}{y!} \left[\left(\frac{1}{1 + \sigma} \right)^{\frac{\lambda_i}{\sigma}} \left(\frac{\sigma}{1 + \sigma} \right)^y \sum_{j=0}^{y-1} - \frac{\lambda_i}{\sigma^2} \prod_{k \neq j} \left(\frac{\lambda_i}{\sigma} + k \right) \right. \\ & + \frac{\Gamma(\frac{\lambda_i}{\sigma} + y)}{\Gamma(\frac{\lambda_i}{\sigma})} \left(\frac{1}{1 + \sigma} \right)^{\frac{\lambda_i}{\sigma}} y \left(\frac{1}{(1 + \sigma)^2} \right)^{y-1} \\ & \left. + \frac{\Gamma(\frac{\lambda_i}{\sigma} + y)}{\Gamma(\frac{\lambda_i}{\sigma})} \left(\frac{\sigma}{1 + \sigma} \right)^y \left(\ln(1 + \sigma) \frac{\lambda_i}{\sigma^2} - \frac{1}{1 + \sigma} \frac{\lambda_i}{\sigma} \right) \left(\frac{1}{1 + \sigma} \right)^{\frac{\lambda_i}{\sigma}} \right]. \end{aligned}$$

As finite sums of uniformly converging sequences of functions are uniformly converging, we can examine the summands separately. Indeed, we will only investigate the first summand in detail here, because the arguments are very similar for the others. The same applies to the derivatives with respect to

$\beta_j, j = 1, \dots, p$ and the partial derivatives of second order.

According to the Weierstrass criterion (e.g. Amann and Escher (2005), Theorem V.1.6) it suffices to proof the convergence of the supremum norms (on Θ). Denoting the first summand with (I) we obtain:

$$\begin{aligned} |(I)| &\stackrel{(i)}{=} \frac{1}{y!} \left(\frac{1}{1+\sigma} \right)^{\frac{\lambda_i}{\sigma}} \left(\frac{\sigma}{1+\sigma} \right)^y \sum_{j=0}^{y-1} \frac{\lambda_i}{\sigma^2} \frac{\Gamma(\frac{\lambda_i}{\sigma} + y)}{\Gamma(\frac{\lambda_i}{\sigma})} \frac{1}{\frac{\lambda_i}{\sigma} + j} \\ &\stackrel{(ii)}{\leq} \frac{1}{y!} \left(\frac{1}{1+\sigma} \right)^{\frac{\lambda_i}{\sigma}} \left(\frac{\sigma}{1+\sigma} \right)^y \left(y - 1 + \frac{\sigma}{\lambda_i} \right) \frac{\lambda_i}{\sigma^2} \frac{\Gamma(\frac{\lambda_i}{\sigma} + y)}{\Gamma(\frac{\lambda_i}{\sigma})} \\ &\stackrel{(iii)}{\leq} \frac{1}{y!} \cdot c \cdot q^y (y - 1 + d) \frac{\Gamma(m + y)}{\Gamma(m)} \end{aligned}$$

In equality (i) we obtain a representation, for which we can apply formula (2), by multiplying each summand with $1 = \frac{\lambda_i/\sigma+j}{\lambda_i/\sigma+j}$. Inequality (ii) is due to $1/(\frac{\lambda_i}{\sigma} + j) \leq 1, j \geq 1$. Using the compactness of \mathcal{B} and \mathcal{S} in combination with the continuity of all involved functions, we can find constants $0 < c, m, d < \infty$ and $0 < q < 1$, such that inequality (iii) is valid.

Thus, we have $\|(I)\|_\infty \leq \frac{1}{y!} \cdot c \cdot q^y (y - 1 + d) \frac{\Gamma(m+y)}{\Gamma(m)}$ and the corresponding series converges according to the ratio test (e.g. Amann and Escher (2005), Theorem II.8.6) since

$$\frac{\frac{1}{(y+1)!} \cdot c \cdot q^{y+1} (y + d) \frac{\Gamma(m+y+1)}{\Gamma(m)}}{\frac{1}{y!} \cdot c \cdot q^y (y - 1 + d) \frac{\Gamma(m+y)}{\Gamma(m)}} = \frac{y + d}{y - 1 + d} \cdot q \cdot \frac{y + m}{y + 1} \rightarrow q < 1, y \rightarrow \infty.$$

Here $x\Gamma(x) = \Gamma(x + 1)$ is used.

(2): According to Lemma 2, there are constants C_{jkl}, D_{jkl} independent of i such that

$$\left| \frac{\partial^3 \ln f_i(y, \boldsymbol{\theta})}{\partial \theta_j \partial \theta_k \partial \theta_l} \right| \leq y \cdot C_{jkl} + D_{jkl} =: H_{ijkl}(y).$$

Since there is only a finite number of combinations of j, k, l we also find constants C, D such that

$$\sum_{y=0}^{\infty} H_{ijkl}(y) f_i(y; \boldsymbol{\theta}) \leq \sum_{y=0}^{\infty} (yC + D) f_i(y; \boldsymbol{\theta}) = \lambda_i C + D \leq M =: M_i.$$

We can find the constant M , since λ_i is a continuous function of \mathbf{z}_i and β and both are from compact sets. Thus $(\lambda_i)_i$ is a bounded sequence.

(3): Obviously we have $\frac{1}{n} \sum_{i=1}^n M_i = \frac{1}{n} \sum_{i=1}^n M = M$. \square

3.2. D -sets

The next lemma shows how we can find an upper bound for certain sets that will occur afterwards.

Lemma 4. *Let $h_i : \mathbb{N}_0 \rightarrow \mathbb{R}$, $i \in \mathbb{N}$, be a sequence of sequences for which $|h_i(y)| \leq Cy + D$, $y \in \mathbb{N}_0, i \in \mathbb{N}$, with $C, D \in \mathbb{R}$. Let further $D_i^n := \{y \in \mathbb{N}_0 : |h_i(y)| > n\}$. Then there exists a $d > 0$ and an $n_0 \in \mathbb{N}$, such that for all $n \geq n_0$ and for all $i \in \mathbb{N}$:*

$$D_i^n \subset \{y \in \mathbb{N}_0 : y > d \cdot n\}$$

For the proof of Lemma 6, we need the following lemma.

Lemma 5. *Let $(a_y)_{y \in \mathbb{N}} \subset (0, \infty)$ be a sequence for which there exists an $y_0 \in \mathbb{N}$ and $0 < q < 1$ such that for all $y \geq y_0$, $a_{y+1}/a_y \leq q$, i.e. the corresponding series converges according to the ratio test. Then, for any $d > 0$,*

$$n \cdot \sum_{y=\lceil dn \rceil}^{\infty} a_y \xrightarrow{n \rightarrow \infty} 0.$$

Lemma 6. *For $k = 1, 2, 3$ and any $l_1, \dots, l_k \in \{1, \dots, p + 1\}$ let $D_{i_1}^{(n)} := \{y \in \mathbb{N}_0 : \left| \frac{\partial^k \ln f_i(y; \boldsymbol{\theta})}{\partial \theta_{l_1} \dots \partial \theta_{l_k}} \right| > n\}$. Then, for all $\boldsymbol{\theta} \in \Theta$,*

$$\sum_{i=1}^n \sum_{y \in D_{i_1}^{(n)}} f_i(y, \boldsymbol{\theta}) = o(1), \quad n \rightarrow \infty.$$

Here, for real valued sequences $(a_n)_n$ and $(b_n)_n$, the o -notation is defined by

$$a_n = o(b_n), n \rightarrow \infty \quad :\Leftrightarrow \quad \frac{a_n}{b_n} \xrightarrow{n \rightarrow \infty} 0.$$

Proof: In each of the cases $k = 1, 2, 3$ by combination of Lemma 2 and Lemma 4, we obtain a $d > 0$, such that for sufficiently large n and all $i \in \mathbb{N}$, we have $D_{i1}^{(n)} \subset \{y \in \mathbb{N}_0 : y > d \cdot n\}$. Thus, for large n :

$$\begin{aligned} \sum_{i=1}^n \sum_{y \in D_{i1}^{(n)}} f_i(y; \boldsymbol{\beta}, \sigma) &\stackrel{(i)}{\leq} \sum_{i=1}^n \sum_{y=\lceil dn \rceil}^{\infty} \frac{\Gamma(\frac{\lambda_i}{\sigma} + y)}{y! \Gamma(\frac{\lambda_i}{\sigma})} \left(\frac{1}{1+\sigma}\right)^{\frac{\lambda_i}{\sigma}} \left(\frac{\sigma}{1+\sigma}\right)^y \\ &\stackrel{(ii)}{\leq} \sum_{i=1}^n \sum_{y=\lceil dn \rceil}^{\infty} \frac{\Gamma(m+y)}{y!} D q^y \\ &\stackrel{(iii)}{=} Dn \sum_{y=\lceil dn \rceil}^{\infty} \frac{\Gamma(m+y)}{y!} q^y \xrightarrow[n \rightarrow \infty]{(iv)} 0 \end{aligned}$$

First, we use the upper bound of the set $D_{i1}^{(n)}$, derived above, to find inequality (i). Because of the continuity of λ_i in \mathbf{z}_i , the compactness of \mathcal{Z} and the continuity of the summands in λ_i , constants m, D exist, both independent of i , such that - with $q := \sigma/(1+\sigma) < 1$ - inequality (ii) is valid. As the summands are now independent of i , we can replace the sum by the factor n (equality (iii)). The sequence $\left(\frac{\Gamma(m+y)}{y!} q^y\right)_y$ fulfills the prerequisites of Lemma 5, because

$$\frac{\frac{\Gamma(m+y+1)}{(y+1)!} q^{y+1}}{\frac{\Gamma(m+y)}{y!} q^y} = \frac{y+m}{y+1} \cdot q \xrightarrow[y \rightarrow \infty]{} q < 1,$$

completing the proof of convergence (iv). \square

Lemma 7. For $k = 1, 2, 3$ and any $l_1, \dots, l_k \in \{1, \dots, p+1\}$ let $D_{i2}^{(n)} := \{y \in \mathbb{N}_0 : \left|\frac{\partial^k \ln f_i(y; \boldsymbol{\theta})}{\partial \theta_{l_1} \dots \partial \theta_{l_k}}\right| < n\}$. Then for all $\boldsymbol{\theta} \in \Theta$

$$\sum_{i=1}^n \sum_{y \in D_{i2}^{(n)}} \left(\frac{\partial \ln f_i(y, \boldsymbol{\theta})}{\partial \theta_{l_1} \dots \partial \theta_{l_k}}\right)^2 f_i(y, \boldsymbol{\theta}) = o(n^2), \quad n \rightarrow \infty.$$

Proof: In each of the cases $k = 1, 2, 3$ we obtain:

$$\begin{aligned}
\frac{1}{n^2} \sum_{i=1}^n \sum_{y \in D_{i2}^{(n)}} \left(\frac{\partial^k \ln f_i(y; \boldsymbol{\theta})}{\partial \theta_{l_1} \dots \partial \theta_{l_k}} \right)^2 f_i(y; \boldsymbol{\beta}, \sigma) &\leq \frac{1}{n^2} \sum_{i=1}^n \sum_{y=0}^{\infty} \left(\frac{\partial^k \ln f_i(y; \boldsymbol{\theta})}{\partial \theta_{l_1} \dots \partial \theta_{l_k}} \right)^2 f_i(y; \boldsymbol{\beta}, \sigma) \\
&\stackrel{(i)}{\leq} \frac{1}{n^2} \sum_{i=1}^n \sum_{y=0}^{\infty} (yC + D)^2 f_i(y; \boldsymbol{\beta}, \sigma) \\
&\stackrel{(ii)}{=} \frac{1}{n^2} \sum_{i=1}^n (C^2(\lambda_i(1 + \sigma) + \lambda_i^2) + 2CD\lambda_i + D^2) \stackrel{(iii)}{\leq} \frac{nE}{n^2} \xrightarrow{n \rightarrow \infty} 0.
\end{aligned}$$

The constants C and D , that explain inequality (i), exist according to Lemma 2. Using the known formulas for the expected value and the variance of the negative binomial distribution mentioned in chapter 2.1, establishes equality (ii). According to the assumptions, the λ_i are bounded across all i , so that the summands are bounded from above by a constant E , which is independent of i . Such constant leads to inequality (iii). \square

3.3. Positive definiteness of F

One can form an arithmetic mean from the information matrices of the single observations. That the limit of the mean, as a sequence in n , exists for our assumptions, is a result to which the following lemma contributes.

Lemma 8. *Let $(a_i)_{i \in \mathbb{N}}$ be a bounded sequence of real numbers, such that for any open set $A \subset \mathbb{R}$ a $c_A \in \mathbb{R}$ exists, such that*

$$\frac{1}{n} \sum_{i=1}^n \mathbf{1}_A(a_i) \xrightarrow{n \rightarrow \infty} c_A. \tag{7}$$

Then, for $n \rightarrow \infty$, a $c \in \mathbb{R}$ exists with $(\sum_{i=1}^n a_i)/n \rightarrow c$.

For the notation in the proof of Lemma 10, we want to introduce two real-valued random variables:

$$\begin{aligned}
A_i &:= A(Y_i, \mathbf{z}_i, \boldsymbol{\theta}) := g'(\mathbf{z}_i^T \boldsymbol{\beta}) \frac{\partial \ln f_i(Y_i, \boldsymbol{\theta})}{\partial \lambda_i} \\
&= g'(\mathbf{z}_i^T \boldsymbol{\beta}) \left[\sum_{j=0}^{Y_i-1} \frac{\sigma^{-1}}{\lambda_i \sigma^{-1} + j} - \frac{1}{\sigma} \ln(1 + \sigma) \right]
\end{aligned}$$

and

$$\begin{aligned} C_i := C(Y_i, \mathbf{z}_i, \boldsymbol{\theta}) &:= \frac{\partial \ln f_i(Y_i, \boldsymbol{\theta})}{\partial \sigma} \\ &= \sum_{j=0}^{Y_i-1} \frac{-\lambda_i \sigma^{-2}}{\lambda_i \sigma^{-1} + j} - \frac{1}{1 + \sigma} \left(\frac{\lambda_i}{\sigma} + Y_i \right) + \frac{\lambda_i}{\sigma^2} \ln(1 + \sigma) + \frac{Y_i}{\sigma} \end{aligned}$$

We summarize some of their properties.

Lemma 9. *For any $\boldsymbol{\theta} \in \Theta$, the following statements hold:*

- (a) $\mathbb{E}_{\boldsymbol{\theta}} A_i^2$ and $\mathbb{E}_{\boldsymbol{\theta}} C_i^2$ exist.
- (b) $\mathbb{E}_{\boldsymbol{\theta}} A_i^2$, $\mathbb{E}_{\boldsymbol{\theta}} C_i^2$ and $\mathbb{E}_{\boldsymbol{\theta}} A_i C_i$ are continuous functions of \mathbf{z}_i .
- (c) $(\mathbb{E}_{\boldsymbol{\theta}} A_i^2)_i$ and $(\mathbb{E}_{\boldsymbol{\theta}} C_i^2)_i$ are sequences both bounded from above and away from zero.
- (d) An $\epsilon > 0$ exists, such that

$$|\mathbb{E}_{\boldsymbol{\theta}} A_i C_i| \leq (\mathbb{E}_{\boldsymbol{\theta}} A_i^2 \mathbb{E}_{\boldsymbol{\theta}} C_i^2)^{\frac{1}{2}} - \epsilon$$

for all $i \in \mathbb{N}$.

Lemma 10. *For $j, k = 1, \dots, p + 1$ the following limits exist:*

$$\begin{aligned} \mathbf{F}_{jk}(\boldsymbol{\theta}) &:= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sum_{y \in \mathbb{N}_0} -\frac{\partial^2 \ln f_i(y, \boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} f_i(y, \boldsymbol{\theta}) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sum_{y \in \mathbb{N}_0} \frac{\partial \ln f_i(y, \boldsymbol{\theta})}{\partial \theta_j} \frac{\partial \ln f_i(y, \boldsymbol{\theta})}{\partial \theta_k} f_i(y, \boldsymbol{\theta}) \end{aligned}$$

The resulting matrix $\mathbf{F}(\boldsymbol{\theta}) := [\mathbf{F}_{jk}(\boldsymbol{\theta})]_{j,k}$ is positive definite and has a finite determinant.

Proof:

Step 1: Existence of limit: First we note that both versions of the inner series (which are expected values with respect to the parameter $\boldsymbol{\theta}$) of the sequences considered in the lemma coincide, which follows from Lemma 3, but will not be discussed in further detail (see e.g. Bickel and Doksum, 2007, Proposition 3.4.4, for details).

Using the denotations introduced in front of this lemma, we can describe the matrix \mathbf{F} as follows:

$$\mathbf{F}(\boldsymbol{\theta}) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} \mathbf{z}_i \mathbf{z}_i^T \mathbb{E}_{\boldsymbol{\theta}} A_i^2 & \mathbf{z}_i \mathbb{E}_{\boldsymbol{\theta}} A_i C_i \\ \mathbf{z}_i^T \mathbb{E}_{\boldsymbol{\theta}} A_i C_i & \mathbb{E}_{\boldsymbol{\theta}} C_i^2 \end{pmatrix} \in \mathbb{R}^{(p+1) \times (p+1)}, \quad (8)$$

for which the existence of the limits is still to be proven. Thus, each of the sequences of interest has the form

$$\frac{1}{n} \sum_{i=1}^n a_i = \frac{1}{n} \sum_{i=1}^n a(\mathbf{z}_i),$$

with a function $a : \mathcal{Z} \rightarrow \mathbb{R}$, which is continuous according to Lemma 9 (b). According to Lemma 8, this sequence converges if, for all open sets $A \subset \mathbb{R}$, there is a $c_A \in [0, 1]$ such that $\frac{1}{n} \sum_{i=1}^n \mathbf{1}_A(a_i) \rightarrow c_A$. Since pre-images of open sets under continuous functions are open and $\mathbf{1}_A(a_i) = \mathbf{1}_{a^{-1}(A)}(\mathbf{z}_i)$, this is the case according to assumption CN4.

Step 2: Positive definiteness: In order to prove the positive definiteness, we consider an arbitrary vector $(\mathbf{x}_1^T, x_2)^T \in \mathbb{R}^{p+1} \setminus \{\mathbf{0}\}$ and denote $c_i := \mathbf{z}_i^T \mathbf{x}_1$. We note that in case of $\mathbf{x}_1 \neq \mathbf{0} = (0, \dots, 0)^T$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n c_i^2 = \mathbf{x}_1^T \left[\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^T \right] \mathbf{x}_1 > 0, \quad (9)$$

because, according to assumption CN3, $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^T$ is positive definite (the existence of the limit, which is not required in CN3, emerges from identical arguments to those regarding the existence of \mathbf{F}).

We now obtain

$$\begin{aligned}
(\mathbf{x}_1^T, x_2)\mathbf{F}(\boldsymbol{\theta})(\mathbf{x}_1^T, x_2)^T &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_1^T, x_2) \begin{pmatrix} \mathbf{z}_i \mathbf{z}_i^T \mathbb{E}_{\boldsymbol{\theta}} A_i^2 & \mathbf{z}_i \mathbb{E}_{\boldsymbol{\theta}} A_i C_i \\ \mathbf{z}_i^T \mathbb{E}_{\boldsymbol{\theta}} A_i C_i & \mathbb{E}_{\boldsymbol{\theta}} C_i^2 \end{pmatrix} \begin{pmatrix} \mathbf{x}_1 \\ x_2 \end{pmatrix} \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n [(\mathbf{x}_1^T \mathbf{z}_i)(\mathbf{z}_i^T \mathbf{x}_1) \mathbb{E}_{\boldsymbol{\theta}} A_i^2 + (\mathbf{x}_1^T \mathbf{z}_i)x_2 \mathbb{E}_{\boldsymbol{\theta}} A_i C_i \\
&\quad + (\mathbf{z}_i^T \mathbf{x}_1)x_2 \mathbb{E}_{\boldsymbol{\theta}} A_i C_i + x_2^2 \mathbb{E}_{\boldsymbol{\theta}} C_i^2] \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n [c_i^2 \mathbb{E}_{\boldsymbol{\theta}} A_i^2 + 2c_i x_2 \mathbb{E}_{\boldsymbol{\theta}} A_i C_i + x_2^2 \mathbb{E}_{\boldsymbol{\theta}} C_i^2] \\
&\stackrel{(i)}{\geq} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n [c_i^2 \mathbb{E}_{\boldsymbol{\theta}} A_i^2 - 2|c_i x_2| |\mathbb{E}_{\boldsymbol{\theta}} A_i C_i| + x_2^2 \mathbb{E}_{\boldsymbol{\theta}} C_i^2] \\
&\stackrel{(ii)}{\geq} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \left[c_i^2 \mathbb{E}_{\boldsymbol{\theta}} A_i^2 - 2|c_i x_2| \left(\sqrt{\mathbb{E}_{\boldsymbol{\theta}} A_i^2 \mathbb{E}_{\boldsymbol{\theta}} C_i^2} - \epsilon \right) + x_2^2 \mathbb{E}_{\boldsymbol{\theta}} C_i^2 \right] \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \left[\left(c_i \sqrt{\mathbb{E}_{\boldsymbol{\theta}} A_i^2} - x_2 \sqrt{\mathbb{E}_{\boldsymbol{\theta}} C_i^2} \right)^2 + 2|c_i x_2| \epsilon \right] \stackrel{(iii)}{>} 0.
\end{aligned}$$

Here, inequality (i) is due to the rule $x \geq -|x|$, $x \in \mathbb{R}$, and inequality (ii) is due to Lemma 9 (d). In order to establish the strict inequality (iii), we distinguish between several cases. In case $x_2 = 0$, we have $\mathbf{x}_1 \neq \mathbf{0}$ and hence (9). Together with the boundedness away from zero of $(\mathbb{E}_{\boldsymbol{\theta}} A_i^2)_i$, according to Lemma 9 (c), this provides the inequality. In case $\mathbf{x}_1 = \mathbf{0}$, we have $x_2 \neq 0$, and thus the inequality is guaranteed by the boundedness away from zero of $(\mathbb{E}_{\boldsymbol{\theta}} C_i^2)_i$ according to Lemma 9 (c). In case of $\mathbf{x}_1 \neq \mathbf{0}$ and $x_2 \neq 0$, we use (9) and the boundedness of $(c_i)_i$ (by a constant $d > 0$) to see that

$$0 < \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n c_i^2 \leq d \cdot \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n |c_i|,$$

which implies

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n 2|c_i x_2| \epsilon = 2|x_2| \epsilon \cdot \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n |c_i| > 0.$$

Thus, the positive definiteness is shown. The finiteness of the determinant is clear, since $(\mathbb{E} A_i^2)_i$, $(\mathbb{E} C_i^2)_i$ and all entries of $(\mathbf{z}_i)_i$ are bounded. \square

3.4. Result

We are now able to establish the weak consistency of the maximum likelihood estimator.

Theorem 1. *Let, for a negative binomial regression, conditions NBR1-3 and CN1-4 be fulfilled. Further let $\hat{\boldsymbol{\theta}} := (\hat{\boldsymbol{\beta}}, \hat{\sigma})$ denote a solution of the likelihood equations (6). Then*

$$\hat{\boldsymbol{\theta}} \xrightarrow[n \rightarrow \infty]{P} \boldsymbol{\theta}_0.$$

Proof: Since the distributions of the different observations differ, dependent on the fixed covariate, we are in the situation of independent, but not identically distributed observations. Bradley and Gart (1962) provide conditions for consistency in such a case. Lemma 1 states that assumption I(i) from the mentioned paper is fulfilled, and Lemma 3 states the validity of assumption I(ii). Assumption II(i) is proven by Lemmas 6 and 7, Assumption II(ii) by Lemmas 6, 7 and 10. The validity of assumption II(iii) is guaranteed by Lemmas 6, 7 and 3. Thus, we can apply Theorem 2(i) from Bradley and Gart (1962), which proves our theorem. \square

4. Asymptotic Normality

In order to establish the asymptotic normality of the maximum likelihood estimator, we need to prove one further lemma:

Lemma 11. *For all $\boldsymbol{\theta} \in \Theta$ and all $\epsilon > 0$ let*

$$D_{i3}^{(n,\epsilon)} := \left\{ y \in \mathbb{N}_0 : \left[\sum_{j=1}^{p+1} \left(\frac{\partial \ln f_i(y, \boldsymbol{\theta})}{\partial \theta_j} \right)^2 \right]^{1/2} > \epsilon \sqrt{n} \right\}. \text{ Then for all } \boldsymbol{\theta} \in \Theta$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sum_{y \in D_{i3}^{(n,\epsilon)}} \sum_{j=1}^p \left(\frac{\partial \ln f_i(y, \boldsymbol{\theta})}{\partial \theta_j} \right)^2 f_i(y, \boldsymbol{\theta}) = 0.$$

Proof: First, we find an upper bound for large n for $D_{i3}^{(n,\epsilon)}$:

$$\begin{aligned} D_{i3}^{(n,\epsilon)} &\stackrel{(i)}{\subset} \left\{ y : \left(\sum_{j=1}^{p+1} (C_j y + D_j)^2 \right)^{1/2} > \epsilon \sqrt{n} \right\} \\ &\stackrel{(ii)}{=} \{ y : (C y^2 + D y + E) > \epsilon^2 n \} \\ &\stackrel{(iii)}{\subset} \{ y : F y^2 > \epsilon^2 n \} \stackrel{(iv)}{=} \{ y : y > d \sqrt{n} \} \end{aligned}$$

The positive constants C_j and D_j in (i), which do not depend on i , exist according to Lemma 2. Squaring both sides, expanding the left-hand side and summarizing the constants leads to equality (ii). Defining $F := 2C$, (iii) is valid for sufficiently large n . Defining $d := \frac{\epsilon}{\sqrt{F}}$ we obtain identity (iv).

Thus, for large n , we obtain

$$\begin{aligned} 0 &\leq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sum_{y \in D_{i3}^{(n,\epsilon)}} \sum_{j=1}^{p+1} \left(\frac{\partial \ln f_i(y, \boldsymbol{\theta})}{\partial \theta_j} \right)^2 f_i(y; \boldsymbol{\theta}) \\ &\stackrel{(i)}{\leq} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sum_{y=\lceil d\sqrt{n} \rceil}^{\infty} (C y^2 + D y + E) f_i(y; \boldsymbol{\theta}) \\ &\stackrel{(ii)}{\leq} \lim_{n \rightarrow \infty} \frac{n}{n} F \sum_{y=\lceil d\sqrt{n} \rceil}^{\infty} (C y^2 + D y + E) \frac{\Gamma(m+y)}{\Gamma(m)y!} q^y \\ &= 0. \end{aligned}$$

Applying the bound of $D_{i3}^{(n,\epsilon)}$ found before, as well as Lemma 2, leads to inequality (i). In inequality (ii), we bound the density uniformly across all i and can thus substitute the sum by the factor n . Again, we have $q < 1$, and n/n cancels out. For a fixed starting index, the series converges according to the ratio test. Thus, the term approaches zero when shifting the starting index to infinity. \square

Theorem 2. *We consider the negative binomial regression model including the conditions NBR1-3 and CN1-4.*

Let $\hat{\boldsymbol{\theta}} := (\hat{\boldsymbol{\beta}}, \hat{\sigma})$ denote a solution of the likelihood equations (6). Then

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathbf{Z} \quad \text{with} \quad \mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{F}(\boldsymbol{\theta}_0)^{-1}),$$

where $\mathbf{F}(\boldsymbol{\theta})$ is defined in Lemma 10.

Proof: Lemma 11 states that assumption III from Bradley and Gart (1962) is also fulfilled in our scenario. Thus, the application of their Theorem 2(iv) proves asymptotic normality for our model. \square

5. Discussion and Conclusion

Practically, in some applications, the counts are events for (latent) subunits being exposed (see Winkelmann (2000), Sec. 2.1.3, and Dobson (2002), Sec. 9.1). Different kinds of exposed subunits are possible. For instance, a time period can be of a certain length when counting the number of events, or a number of people or objects may cause or suffer from certain events. For instance, a number of people on a ship cruise may - or may not - book one or more shore leaves. A couple of loans in a bank portfolio or a number of policies in an insurance portfolio may cause losses. If the amount of exposure is the same for all observed units, it may be normalized to one, as was implicitly the case in the present model. But if exposure varies between units, it should be involved in the model as a factor in (3). Under justifiable assumptions, our proofs are not affected by such factors.

Theoretically, there is an advantage of the NB1-parametrization over the NB2-one. It is reproducible, i.e. the convolution of two NB1-distributions – with coinciding dispersion parameters – is a NB1-distribution (see Winkelmann, 2000, Proposition 2.3.2). Such a property may facilitate analyzing data at different levels of aggregation. One should finally note the theoretical disadvantage that the Poisson regression, i.e. a dispersion parameter equal to zero, is not nested in our model. Hence, for example a likelihood ratio test for overdispersion cannot be derived directly from the asymptotic parameter normality (as in Weißbach and Walter, 2010). However, reasons for overdispersion such as correlation between subunits or unobserved heterogeneity between units are often present beyond doubt, at least for observational data.

Acknowledgment

The financial support from the Deutsche Forschungsgemeinschaft (DFG) is gratefully acknowledged (Grant WE3573/3-1 “Multi-state, multi-time, multi-level analysis of health-related demographic events: Statistical aspects and applications” and CRC 823 “Statistical modelling of nonlinear dynamic processes”, Project A1: Dynamic Dependence Structures in Risky Asset Returns).

6. References

- Amann, H., Escher, J., 2005. Analysis I, Springer, Basel.
- Bauer, H., 2001. Measure and integration theory. De Gruyter, Berlin.
- Bickel, P. J., Doksum, K. A., 2007. Mathematical statistics: Basic ideas and selected topics, 2nd Edition. Pearson Prentice Hall, Upper Saddle River.
- Bradley, R. A., Gart, J. J., 1962. The asymptotic properties of ML estimators when sampling from associated populations. *Biometrika* 49, 205-214.
- Cameron, A. C., Trivedi, P. K., 2008. Regression analysis of count data, 7th Edition. Cambridge University Press, Cambridge.
- Dobson, A. J., 2002. An introduction to generalized linear models, 2nd Edition. Chapman & Hall/CRC, Boca Raton.
- Fahrmeir, L., Kaufmann, H., 1985. Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Annals of Statistics* 13, 342–368.
- Feller, W., 1971. An introduction to probability theory and its applications Vol. 2, 2nd Edition. Wiley, New York.

- Gourieroux, C., Monfort, A., 1995. *Statistics and econometric models*, Vol. 2. Cambridge University Press, Cambridge.
- Hilbe, J., 2011. *Negative binomial regression*, 2nd Edition. Cambridge University Press, Cambridge.
- Johnson, N. L., Kemp, A. W., Kotz, S., 2005. *Univariate discrete distributions*, 3rd Edition. Wiley, Hoboken.
- Kremer, A., Weißbach, R., Liese, F., 2014. Maximum likelihood estimation for left-censored survival times in an additive hazard model. *Journal of Statistical Planning and Inference* 149, 33–45.
- Moore, D.F., 1986. Asymptotic properties of moment estimators for overdispersed counts and proportions. *Biometrika* 73, 583–588.
- Wald, A., 1949. Note on the consistency of the maximum likelihood estimate. *Annals of Mathematical Statistics* 20, 595–601.
- Weißbach, R., Herzog, M., Menzel, G., 2015. Regionaler Anteil kariesfreier Vorschulkinder – eine cluster-randomisierte Studie in Südhessen. *AStA Wirtschafts- und Sozialstatistisches Archiv* 9, 27–39.
- Weißbach, R., Tschiersch, P., Lawrenz, C., 2009. Testing time-homogeneity of rating transitions after origination of debt. *Empirical Economics* 36, 575–596.
- Weißbach, R., Walter, R., 2010. A likelihood ratio test for stationarity of rating transitions. *Journal of Econometrics* 155, 188–194.
- Winkelmann, R., 2000. *Econometric analysis of count data*, 3rd Edition. Springer, Berlin.

Appendix A. Proof of Lemmas

Proof of Lemma 2:

Statement 1: The partial derivatives of order $k = 1, 2, 3$ for $\ln f_i$ have the form

$$\frac{\partial^k \ln f_i(y; \boldsymbol{\theta})}{\partial \theta_{l_1} \dots \partial \theta_{l_k}} = \sum_{j=0}^{y-1} a(j, \boldsymbol{\theta}, \mathbf{z}_i) + y \cdot b(\boldsymbol{\theta}, \mathbf{z}_i) + c(\boldsymbol{\theta}, \mathbf{z}_i),$$

where a is a rational function in j , the degree of the denominator of which is greater than the degree of the numerator. Let $\mathcal{M} := \Theta \times \mathcal{Z}$. The coefficient functions of a are continuous on \mathcal{M} . In the denominator, the absolute term is not zero and all non-zero coefficient functions are strictly positive on \mathcal{M} . b and c are continuous functions on \mathcal{M} .

Proof of statement 1: For $k = 1$, the partial derivatives are:

$$\begin{aligned} \frac{\partial \ln f_i(y; \boldsymbol{\theta})}{\partial \sigma} &= \sum_{j=0}^{y-1} \frac{-\lambda_i \sigma^{-2}}{\lambda_i \sigma^{-1} + j} - \frac{1}{1 + \sigma} \left(\frac{\lambda}{\sigma} + y \right) + \frac{\lambda_i}{\sigma^2} \ln(1 + \sigma) + \frac{y}{\sigma} \\ \frac{\partial \ln f_i(y; \boldsymbol{\theta})}{\partial \beta_l} &= \left[\sum_{j=0}^{y-1} \frac{\sigma^{-1}}{\lambda_i \sigma^{-1} + j} - \frac{1}{\sigma} \ln(1 + \sigma) \right] z_{il} g'(\mathbf{z}_i^T \boldsymbol{\beta}) \end{aligned}$$

Since λ_i is continuous in $\boldsymbol{\beta}$ and \mathbf{z}_i , the statement is obvious. The linear structure in y obviously remains valid after further differentiation. The continuity of the respective versions of b and c is ensured by assumption CN2. Using the quotient rule, it is easy to see that the mentioned properties of a also remain valid after further differentiating.

Statement 2: Let $a(j, \mathbf{x}), \mathbf{x} \in \mathcal{X} \subset \mathbb{R}^k$ be a rational function in j , whose coefficient functions are continuous on the compact set \mathcal{X} . In addition, let the absolute term function of the denominator be strictly positive and all coefficient functions of the denominator non-negative on \mathcal{X} . Furthermore, let the degree of the denominator μ be greater than the degree of the numerator λ . Then, there is a constant M , independent of j and \mathbf{x} such that $|a(j, \mathbf{x})| \leq M$ for all $j \geq 0, \mathbf{x} \in \mathcal{X}$.

Proof of statement 2: Using the continuity of the coefficient functions (see Lemma 1) and the compactness of \mathcal{X} , we obtain

$$\begin{aligned} \max_{\mathbf{x} \in \mathcal{X}} |a(j, \mathbf{x})| &= \max_{\mathbf{x} \in \mathcal{X}} \left| \frac{\sum_{l=0}^{\nu} a_l(\mathbf{x}) j^l}{\sum_{k=0}^{\mu} b_k(\mathbf{x}) j^k} \right| \leq \max_{\mathbf{x} \in \mathcal{X}} \frac{\sum_{l=0}^{\nu} |a_l(\mathbf{x})| j^l}{\sum_{k=0}^{\mu} b_k(\mathbf{x}) j^k} \\ &\leq \frac{\sum_{l=0}^{\nu} \max_{\mathbf{x}} |a_l(\mathbf{x})| j^l}{\sum_{k=0}^{\mu} \min_{\mathbf{x}} b_k(\mathbf{x}) j^k} =: \frac{\sum_{l=0}^{\nu} a_l j^l}{\sum_{k=0}^{\mu} b_k j^k} =: a_{max}(j). \end{aligned}$$

$a_{max}(j)$ is a rational function in j independent of \mathbf{x} . According to the assumptions, the denominator has no zeros for $j \geq 0$ and thus, $a_{max}(j)$ is defined and continuous in j on $[0, \infty)$. Since $\mu > \lambda$, we have $\lim_{j \rightarrow \infty} a_{max}(j) = 0$. Therefore, $M := \max_{j \geq 0} a_{max}(j)$, which obviously fulfills the condition claimed in the statement.

For $k \in \{1, 2, 3\}$ we combine both statements. \mathcal{M} from the first statement corresponds to \mathcal{X} from the second. Let M be the constant from statement 2 and $b_{max} := \max_{(\boldsymbol{\theta}, \mathbf{z}_i) \in \mathcal{M}} b(\boldsymbol{\theta}, \mathbf{z}_i) < \infty$, $c_{max} := \max_{(\boldsymbol{\theta}, \mathbf{z}_i) \in \mathcal{M}} c(\boldsymbol{\theta}, \mathbf{z}_i) < \infty$, which exist because of the compactness of \mathcal{M} and the continuity of b, c . Then

$$\begin{aligned} \left| \frac{\partial^k \ln f_i(y; \boldsymbol{\theta})}{\partial \theta_{l_1} \dots \partial \theta_{l_k}} \right| &\leq \sum_{j=0}^{y-1} |a(j, \boldsymbol{\theta}, \mathbf{z}_i)| + y \cdot |b(\boldsymbol{\theta}, \mathbf{z}_i)| + |c(\boldsymbol{\theta}, \mathbf{z}_i)| \\ &\leq y \cdot M + y \cdot b_{max} + c_{max} =: Cy + D \end{aligned}$$

□

Proof of Lemma 4: W.l.o.g. let $C > 0$ and define $c := 2C$. Then, there is a $\tilde{y} \in \mathbb{N}$, such that $cy \geq Cy + D$ for all $y \geq \tilde{y}$. Define $n_0 := \lceil \tilde{y}c \rceil$. Then, for all $n \geq n_0$ and all $y \leq \tilde{y}$

$$Cy + D \leq C\tilde{y} + D \leq c\tilde{y} \leq n_0 \leq n. \quad (\text{A.1})$$

Thus, for all $n \geq n_0$ and all i follows

$$\begin{aligned} D_i^n &\subset \{y \in \mathbb{N}_0 : Cy + D > n\} \stackrel{(i)}{=} \{y \geq \tilde{y} : Cy + D > n\} \\ &\stackrel{(ii)}{\subset} \{y \geq \tilde{y} : c \cdot y > n\} \subset \{y \in \mathbb{N}_0 : c \cdot y > n\} \stackrel{(iii)}{=} \{y \in \mathbb{N}_0 : y > d \cdot n\} \end{aligned}$$

Here, equality (i) is due to $n \geq n_0$ in connection with (A.1). Since we have $c \cdot y \geq Cy + D$ for all $y \geq \tilde{y}$, the subset relation (ii) follows. With the definition $d := \frac{1}{c}$, identity (iii) is clear. \square

Proof of Lemma 5: First, we observe that for all $y \geq y_0$, $a_y \leq a_{y_0} \cdot q^{y-y_0} =: c \cdot q^y$. This is easy to prove inductively, because for $y = y_0$, the statement is clear and for $y \rightarrow y + 1$ we obtain

$$a_{y+1} = \frac{a_{y+1}}{a_y} a_y \stackrel{(i)}{\leq} q \cdot a_y \stackrel{(ii)}{\leq} q \cdot q^{y-y_0} a_{y_0} = q^{(y+1)-y_0} a_{y_0} = c \cdot q^{y+1},$$

where (i) is due to the prerequisites of the lemma and (ii) due to the induction hypothesis. Thus, for $n \geq \frac{y_0}{d}$ we have

$$n \cdot \sum_{y=\lceil dn \rceil}^{\infty} a_y \leq c \cdot n \cdot \sum_{y=\lceil dn \rceil}^{\infty} q^y = c \cdot n \cdot \frac{q^{\lceil dn \rceil + 1}}{1 - q} \leq \frac{c}{1 - q} \cdot n q^{dn}.$$

The convergence to zero now follows according to L'Hospital's rule, since

$$\lim_{x \rightarrow \infty} x q^{dx} = \lim_{x \rightarrow \infty} \frac{x}{q^{-dx}} = \lim_{x \rightarrow \infty} \frac{1}{-d \ln(q) q^{-dx}} = \frac{1}{-d \ln(q)} \lim_{x \rightarrow \infty} q^{dx} = 0.$$

\square

Proof of Lemma 8: First, we observe that condition (7) is automatically fulfilled for closed sets B . Additionally, because the compliment B^c is open, $c_B = 1 - c_{B^c}$ easily arises. It follows that the condition is also fulfilled for all intervals $I \subset \mathbb{R}$, because each interval is the disjoint union of an open set A_I and a closed set B_I , so that we obtain $c_I = c_{A_I} + c_{B_I}$.

Because of the boundedness, a $K \in \mathbb{R}_+$ exists, such that $(a_n)_{n \in \mathbb{N}} \subset [-K, K]$. We define the intervals $I_j^{(m)} := [-K + \frac{j-1}{m} 2K, -K + \frac{j}{m} 2K)$, $j = 1, \dots, m$, $m \in \mathbb{N}$, and denote the corresponding constants from the assumptions of the lemma with $c_{jm} := c_{I_j^{(m)}}$. Since, for all $m \in \mathbb{N}$ $\{I_j^{(m)}\}_{j=1, \dots, m}$ is a partition of $[-K, K)$, we have

$$\sum_{j=1}^m c_{jm} = 1, m \in \mathbb{N}. \tag{A.2}$$

Note that the finiteness of m is important here. Otherwise this 'additive property' would not be guaranteed. Using the definition of the intervals $I_j^{(m)}$, we find the inequality

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n a_i &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m \mathbf{1}_{I_j^{(m)}}(a_i) \cdot a_i \leq \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m \mathbf{1}_{I_j^{(m)}}(a_i) \left(-K + \frac{j}{m} 2K \right) \\ &= \sum_{j=1}^m \left(-K + \frac{j}{m} 2K \right) \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{I_j^{(m)}}(a_i) \end{aligned}$$

and obtain, by taking the limit superior on both sides (which coincides with the limit on the right-hand side),

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n a_i \leq \sum_{j=1}^m \left(-K + \frac{j}{m} 2K \right) c_{jm} =: U^{(m)}, m \in \mathbb{N}.$$

Analogously, we obtain

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n a_i \geq \sum_{j=1}^m \left(-K + \frac{j-1}{m} 2K \right) c_{jm} =: L^{(m)}, m \in \mathbb{N}.$$

Altogether, we therefore obtain

$$L^{(m)} \leq \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n a_i \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n a_i \leq U^{(m)}, m \in \mathbb{N}.$$

Together with

$$U^{(m)} - L^{(m)} = \sum_{j=1}^m \frac{1}{m} 2K c_{jm} = \frac{1}{m} 2K \sum_{j=1}^m c_{jm} \stackrel{(A.2)}{=} \frac{1}{m} 2K \xrightarrow{m \rightarrow \infty} 0$$

follows the convergence of $(\sum_{i=1}^n a_i)/n$. \square

Proof of Lemma 9: (a): First, we note that there are finite constants M and N , such that for all i and θ

$$|A_i| \leq M Y_i + N \quad \text{and} \quad |C_i| \leq M Y_i + N. \quad (\text{A.3})$$

For C_i this is a statement of Lemma 2, for A_i Lemma 2 does not apply directly, but the proof works identically in this case. Thus, the existence of moments is ensured.

(b): With (A.3) and the usual arguments of compactness and continuity, we find constants independent of \mathbf{z}_i such that

$$\mathbb{E}_{\boldsymbol{\theta}} A_i^2 = \sum_{y=0}^{\infty} A(y, \mathbf{z}_i, \boldsymbol{\theta})^2 f_i(y; \boldsymbol{\theta}) \leq \sum_{y=0}^{\infty} (My + N)^2 \cdot E \cdot \frac{\Gamma(m+y)}{\Gamma(m)} q^y < \infty.$$

This series still converges according to the ratio test (where $q < 1$ is important). The continuity of the expected values now follows from a well-known lemma about the continuity of parameter integrals (e.g. Bauer (2001), Lemma 16.1). For $\mathbb{E}_{\boldsymbol{\theta}} C_i^2$ and $\mathbb{E}_{\boldsymbol{\theta}} A_i C_i$, the proofs are identical.

(c): Obviously, neither A_i nor C_i are constant zero, so that $\mathbb{E}_{\boldsymbol{\theta}} A_i^2$ and $\mathbb{E}_{\boldsymbol{\theta}} C_i^2$ are positive for all i (remember $g \neq 0$ according to CN2). Since the dependence on i is only via \mathbf{z}_i , the dependence on \mathbf{z}_i is continuous (see (b)) and all \mathbf{z}_i are from the compact set \mathcal{Z} , the statement is clear.

(d): First, we note that for any $\boldsymbol{\theta} \in \Theta$ and $i \in \mathbb{N}$ A_i and C_i are linear independent random variables in the space $\mathcal{L}^2(\Omega, \mathfrak{A}, P_{\boldsymbol{\theta}})$. To prove this, it is sufficient to show that there is no $d = d(\boldsymbol{\theta}) > 0$ in existence, such that

$$A_i(y) := A(y, \mathbf{z}_i, \boldsymbol{\theta}) = d \cdot C(y, \mathbf{z}_i, \boldsymbol{\theta}) =: d \cdot C_i(y)$$

for all $y \in \mathbb{N}_0$. If such a d existed, we would also have

$$\frac{A_i(y+1) - A_i(y)}{C_i(y+1) - C_i(y)} = \frac{dC_i(y+1) - dC_i(y)}{C_i(y+1) - C_i(y)} = d$$

for all $y \in \mathbb{N}_0$. But since

$$\begin{aligned} \frac{A_i(y+1) - A_i(y)}{C_i(y+1) - C_i(y)} &= \frac{\frac{\sigma^{-1}}{\lambda_i \sigma^{-1+y}}}{\frac{-\lambda_i \sigma^{-2}}{\lambda_i \sigma^{-1+y}} - \frac{1}{1+\sigma} + \frac{1}{\sigma}} \\ &= \left[-\frac{\lambda_i}{\sigma} + \left(\frac{1}{\sigma} - \frac{1}{1+\sigma} \right) \frac{\lambda_i}{\sigma} + \left(\frac{1}{\sigma} - \frac{1}{1+\sigma} \right) y \right]^{-1}, \end{aligned}$$

which is constant in y if and only if $\frac{1}{\sigma} - \frac{1}{1+\sigma} = 0$, being impossible, we can exclude the existence of such a d , and the linear independence is proven.

According to the strict version of the Cauchy-Schwarz inequality (e.g. Feller, 1971, Chap. V, Sec. 8(a)), we thus have

$$|\mathbb{E}_{\boldsymbol{\theta}} A_i C_i| < (\mathbb{E}_{\boldsymbol{\theta}} A_i^2 \mathbb{E}_{\boldsymbol{\theta}} C_i^2)^{\frac{1}{2}} \Leftrightarrow |\mathbb{E}_{\boldsymbol{\theta}} A_i C_i| - (\mathbb{E}_{\boldsymbol{\theta}} A_i^2 \mathbb{E}_{\boldsymbol{\theta}} C_i^2)^{\frac{1}{2}} < 0$$

for all $i \in \mathbb{N}$. Because, on the left-hand side of the right inequality the dependence on i is only via \mathbf{z}_i , the dependence on \mathbf{z}_i is continuous according to (b). As all \mathbf{z}_i come from the compact set \mathcal{Z} , there even an $\epsilon > 0$ exists, such that

$$|\mathbb{E}_{\boldsymbol{\theta}} A_i C_i| - (\mathbb{E}_{\boldsymbol{\theta}} A_i^2 \mathbb{E}_{\boldsymbol{\theta}} C_i^2)^{\frac{1}{2}} \leq -\epsilon,$$

which proves (d). □

