Lecture 7: Overdispersion in Poisson regression

Claudia Czado

TU München

Overview

- Introduction
- Modeling overdispersion through mixing
- Score test for detecting overdispersion

Introduction

McCullagh and Nelder (1989) p.198-200

Overdispersion is present if

$$Var(Y_i) > E(Y_i)$$

Overdispersion can be modeled using a mixing approach:

$$Y_i|Z_i \sim Poisson(Z_i) \quad Z_i \geq 0 \quad \text{ind. rv's}$$

with $E(Z_i) = \mu_i$

Distributional assumption for Z_i

1) $Z_i \sim$ Gamma with mean μ_i and index $\phi\mu_i$

Recall $X \sim Gamma(\mu, \nu) \Rightarrow f_X(x) = \frac{1}{\Gamma(\nu)} \left(\frac{\nu x}{\mu}\right)^{\nu} \exp\left(-\frac{\nu x}{\mu}\right) \frac{1}{x}$

μ mean ν index

$\Rightarrow E(X) = \mu \qquad Var(X) = \mu^2/\nu$

Here $f_{Z_i}(z_i) = \frac{1}{\Gamma(\phi\mu_i)} \left(\frac{\phi\mu_i z_i}{\mu_i}\right)^{\phi\mu_i} \exp\left(-\frac{\phi\mu_i z_i}{\mu_i}\right) \frac{1}{z_i}$

$\Rightarrow E(Z_i) = \mu_i \qquad Var(Z_i) = \frac{\mu_i^2}{\phi\mu_i} = \mu_i/\phi$

$$
\begin{aligned}
P(Y_i = y_i) &= \int_0^\infty \frac{e^{-z_i}(z_i)^{y_i}}{y_i!} \frac{1}{\Gamma(\phi\mu_i)} \left(\frac{\phi\mu_i z_i}{\mu_i}\right)^{\phi\mu_i} \exp\{-\phi z_i\} \frac{1}{z_i} dz_i \\
&= \frac{\phi^{\phi\mu_i}}{\Gamma(\phi\mu_i) y_i!} \int_0^\infty e^{-z_i(1+\phi)} z_i^{(y_i+\phi\mu_i-1)} dz_i \\
&\overset{s_i=z_i(1+\phi)}{=} \frac{\phi^{\phi\mu_i}}{\Gamma(\phi\mu_i) y_i!} \int_0^\infty e^{-s_i} \frac{1}{(1+\phi)^{(y_i+\phi\mu_i-1)}} s_i^{(y_i+\phi\mu_i-1)} \frac{1}{\phi+1} ds_i \\
&= \frac{\phi^{\phi\mu_i}}{\Gamma(\phi\mu_i) y_i! (1+\phi)^{(y_i+\phi\mu_i)}} \int_0^\infty e^{-s_i} s_i^{(y_i+\phi\mu_i-1)} ds_i \\
&= \frac{\phi^{\phi\mu_i} \Gamma(y_i+\phi\mu_i)}{\Gamma(\phi\mu_i) y_i! (1+\phi)^{(y_i+\phi\mu_i)}} = \frac{\Gamma(y_i+\phi\mu_i)}{\Gamma(\phi\mu_i)\Gamma(y_i+1)} \cdot \underbrace{\left(\frac{\phi}{1+\phi}\right)}_{\frac{1}{1/\phi+1}}^{\phi\mu_i} \underbrace{\left(\frac{1}{1+\phi}\right)}_{\frac{1/\phi}{1/\phi+1}}^{y_i}
\end{aligned}
$$

Negative binomial distribution

$X \sim negbin(a, b)$

$$P(X = k) = \frac{\Gamma(a+k)}{\Gamma(a)\Gamma(k+1)} \left(\frac{1}{1+b}\right)^a \left(\frac{b}{1+b}\right)^k \quad k = 0, 1, \ldots \quad a, b \geq 0$$

$$E(X) = ab \qquad Var(X) = ab(1 + b)$$

Marginal distribution of Y_i

is given by

$$Y_i \sim negbin(a_i, b_i) \quad \text{with} \quad a_i = \phi\mu_i \quad b_i = 1/\phi$$
$$\Rightarrow \quad E(Y_i) = \frac{\phi\mu_i}{\phi} = \mu_i \quad \text{und} \quad Var(Y_i) = \frac{\phi\mu_i}{\phi}(1 + 1/\phi) = \mu_i(1 + 1/\phi)$$

Remarks:

- $Var(Y_i) > E(Y_i)$ if $1/\phi > 0$, i.e. overdispersion

- $\phi = \infty \Rightarrow$ no overdispersion

- To estimate β and ϕ jointly one needs to maximize the negative binomial likelihood. No standard software available but one can use the Splus library $negbin()$ from $http://lib.stat.cmu.edu/S/$

- $Var(Y_i) = \mu_i(1 + 1/\phi)$ is linear in μ_i

2) $Z_i \sim$ Gamma with mean μ_i and index ν

$\overset{\text{(exercise)}}{\Rightarrow}$ $Y_i \sim negbin(\nu, \mu_i/\nu)$ such that

$$E(Y_i) = \nu \cdot \mu_i/\nu = \mu_i$$

$$Var(Y_i) = \mu_i + \mu_i^2/\nu$$

$\Rightarrow$ quadratic variance function in μ_i

Multiplicative heterogeneity in Poisson regression

Another approach for modeling overdispersion is to use

$$Y_i|Z_i \sim Poisson(\mu_i Z_i)$$

with $E(Z_i) = 1$ and $Var(Z_i) = \sigma_Z^2$, i.e. Z_i i.i.d., Z_i is called multiplicative random effect

$$\overset{\text{(exercise)}}{\Rightarrow} \quad \begin{array}{l} E(Y_i) = \mu_i \\ Var(Y_i) = \mu_i + \sigma_Z^2 \mu_i^2 \end{array}$$

If $Z_i \sim Gamma$ with expectation 1 and index ν

$$\Rightarrow Y_i \text{ is } negbin(a_i, b_i) \quad a_i = \nu, \quad b_i = \frac{\mu_i}{\nu}$$

Summary

1) $Y_i|Z_i \sim Poisson(Z_i) \qquad E(Z_i) = \mu_i$

a)

$$
\begin{aligned}
& Z_i \sim Gamma(\mu_i, \phi\mu_i) \\
\Rightarrow \quad & Y_i \sim Negbin(\phi\mu_i, 1/\phi) \\
& E(Y_i) = \mu_i \quad Var(Y_i) = \mu_i(1 + 1/\phi) \quad \text{linear}
\end{aligned}
$$

b)

$$
\begin{aligned}
& Z_i \sim Gamma(\mu_i, \nu) \\
\Rightarrow \quad & Y_i \sim Negbin(\nu, \mu_i/\nu) \\
& E(Y_i) = \mu_i \quad Var(Y_i) = \mu_i + \mu_i^2/\nu \quad \text{quadratic}
\end{aligned}
$$

2) $Y_i|Z_i \sim Poisson(\mu_i Z_i) \qquad E(Z_i) = 1$

$$
\begin{aligned}
& Z_i \sim Gamma(1, \nu) \\
\Rightarrow \quad & Y_i \sim Negbin(\nu, \mu_i/\nu) \\
& E(Y_i) = \mu_i \quad Var(Y_i) = \mu_i + \mu_i^2/\nu \quad \text{quadratic}
\end{aligned}
$$

Test for overdispersion

Dean (1992)

Assume

$$Y_i \sim Poisson(\mu_i) \quad \text{with} \quad \mu_i = e^{\mathbf{x_i}^t \beta}$$

$$\Rightarrow \quad \theta_i = \ln(\mu_i) = \boldsymbol{x_i^t \beta}$$

To model overdispersion we assume that the canonical parameters θ_i are not fixed but random quantities θ_i^* with

$$E(\theta_i^*) = \theta_i$$

$$Var(\theta_i^*) = \tau k_i(\theta_i) > 0 \quad \text{for} \quad \tau \geq 0 \quad \text{and} \quad k_i(\theta_i) \quad \text{differentiable}$$

To test for overdispersion, want to test

$$H_0 : \tau = 0 \qquad \text{versus} \qquad H_1 : \tau > 0$$

Example:

$$\theta_i^* = \boldsymbol{x_i^t}\boldsymbol{\beta} + Z_i \quad \text{with} \quad Z_i \; i.i.d. \quad E(Z_i) = 0 \quad Var(Z_i) = \tau < \infty$$

$$\Rightarrow \quad \begin{array}{l} E(\theta_i^*) = \theta_i \\ Var(\theta_i^*) = \tau \qquad \text{i.e.} \quad k_i(\theta_i) = 1 \end{array}$$

Compare to:

$$\begin{array}{rl} & Y_i|Z_i \sim Poisson(\mu_i Z_i) \quad E(Z_i) = 1 \\ \Rightarrow & \theta_i^* = \ln(\mu_i Z_i) = \ln(\mu_i) + \ln(Z_i) \quad E(\ln(Z_i)) \approx 0 \end{array}$$

A score test will now be developed for $H_0 : \tau = 0$ versus $H_1 : \tau > 0$.

General score test

Let $Y_1, \ldots, Y_n$ ind. r.v. with densities $f_i(y, \boldsymbol{\theta})$

$$\boldsymbol{\theta} \in \Omega \qquad \text{and} \qquad H_0 : \boldsymbol{\theta} \in \Omega_0 \quad H_1 : \boldsymbol{\theta} \in \Omega_1$$

$$\Omega_0 + \Omega_1 = \Omega \subset \mathbb{R}^d \quad \Omega_0 \bigcap \Omega_1 = \emptyset$$

Loglikelihood: $$l(\boldsymbol{\theta}) = \sum_{i=1}^{n} l_i(\boldsymbol{\theta}) \qquad l_i(\boldsymbol{\theta}) = \ln f_i(y_i; \boldsymbol{\theta})$$

$$I(\boldsymbol{\theta}) := E\left(\left[\frac{\partial l(\theta)}{\partial \theta}\right] \left[\frac{\partial l(\theta)}{\partial \theta}\right]^t \right) \in \mathbb{R}^{d \times d} \quad \text{Fisher information}$$

$$= -E\left(\frac{\partial l^2(\theta)}{\partial \theta \partial \theta^t}\right) \quad \text{under regularity conditions}$$

We are interested in the following composite hypothesis for $\boldsymbol{\theta} = (\boldsymbol{\beta}^t, \tau)^t$:

$$H_0 : \tau = \tau_0 \qquad \text{versus} \qquad H_1 : \tau \neq \tau_0$$

Consider the score statistic

$$T_s = \left[\frac{\partial}{\partial \theta} l(\boldsymbol{\theta}) \Big|_{\theta=\hat{\theta}_0}\right]^t I^{-1}(\hat{\boldsymbol{\theta}}_0) \left[\frac{\partial}{\partial \theta} l(\boldsymbol{\theta}) \Big|_{\theta=\hat{\theta}_0}\right]$$

where $\hat{\boldsymbol{\theta}}_0$ is the MLE of $\boldsymbol{\theta}$ in model $\boldsymbol{\theta} = (\boldsymbol{\beta}^t, \tau_0)^t$, i.e.

$$\hat{\boldsymbol{\theta}}_0 = (\hat{\boldsymbol{\beta}}^t, \hat{\tau}_0)^t \quad \text{satisfies} \quad \frac{\partial}{\partial \theta} l(\boldsymbol{\theta}) = \left(\frac{\partial l(\beta, \tau)^t}{\partial \beta}, \frac{\partial l(\beta, \tau)}{\partial \tau}\right)^t \Big|_{\theta=\hat{\theta}_0} = \mathbf{0}$$

Under regularity conditions we have $T_S \approx \chi^2_{\text{dim}\Omega - \text{dim}\Omega_0}$ under H_0.

Remark:

There are regularity condition under which the standard asymptotic remains valid if τ_0 is at the boundary of Ω, as for example at $H_0 : \tau = 0$ versus $H_1 : \tau > 0$.

Here:

$\boldsymbol{\theta} = (\boldsymbol{\beta}^t, \tau)^t$ For $\tau = 0$ we have a Poisson GLM and one can show that

$\hat{\boldsymbol{\theta}}^0 = (\hat{\boldsymbol{\beta}}^{0t}, 0)^t$ where $\hat{\boldsymbol{\beta}}^0$ is the MLE of the Poisson GLM.

Score test for overdispersion in count regression

$$Y_i|\theta_i^* \sim Poisson(e^{\theta_i^*})$$

$$\text{with} \quad E(\theta_i^*) = \boldsymbol{x_i^t \beta}$$

$$\text{and} \quad Var(\theta_i^*) = \tau k_i(\theta_i)$$

Let $f(y_i, \theta_i^*)$ be the conditional density of Y_i given θ_i^*, then with Taylor we have

$$\begin{aligned} f(Y_i, \theta_i^*) &= f(Y_i, \theta_i) + \frac{\partial}{\partial \theta_i^*} f(Y_i, \theta_i^*)|_{\theta_i^* = \theta_i}(\theta_i^* - \theta_i) \\ &+ \frac{1}{2} \frac{\partial^2}{\partial \theta_i^{*2}} f(Y_i, \theta_i^*)|_{\theta_i^* = \theta_i}(\theta_i^* - \theta_i)^2 + R \\ R &= \text{remainder term} \end{aligned}$$

$$
\begin{aligned}
\Rightarrow \; f_M(Y_i, \theta_i) &= \text{marginal likelihood of } Y_i \\
&= \int f(Y_i, \theta_i^*) f(\theta_i^*) d\theta_i^* \\
&= E_{\theta_i^*}(f(Y_i, \theta_i^*)) \\
&= f(Y_i, \theta_i) + 0 + \frac{1}{2}\frac{\partial^2}{\partial \theta_i^{*2}} f(Y_i, \theta_i^*)|_{\theta_i^*=\theta_i} \underbrace{E[(\theta_i^* - \theta_i)^2]}_{Var(\theta_i^*)=\tau k_i(\theta_i)} + R \\
&= f(Y_i, \theta_i)\Big[1 + \frac{1}{2}\tau k_i(\theta_i) f^{-1}(Y_i, \theta_i) \frac{\partial^2}{\partial \theta_i^{*2}} f(Y_i, \theta_i^*)|_{\theta_i^*=\theta_i} + R f^{-1}(Y_i, \theta_i)\Big]
\end{aligned}
$$

Marginal log likelihood

$$
\begin{aligned}
l(\boldsymbol{\theta}) &= \sum_{i=1}^{n} l_i(\theta_i) = \sum_{i=1}^{n} \ln f_M(Y_i, \theta_i) \\
&= \sum_{i=1}^{n} [\ln f(Y_i, \theta_i) \\
&+ \ln \left\{ 1 \; + \; \tfrac{1}{2}\tau k_i(\theta_i) \; f^{-1}(Y_i, \theta_i) \; \tfrac{\partial^2}{\partial \theta_i^{*2}} f(Y_i, \theta_i^*)|_{\theta_i^* = \theta_i} \; + \; R f^{-1}(Y_i, \theta_i) \right\}]
\end{aligned}
$$

$$
\Rightarrow \quad \frac{\partial l(\theta)}{\partial \theta}\bigg|_{\theta = (\hat{\beta}^{0t}, 0)^t} \quad = \quad \begin{pmatrix} \frac{\partial l(\beta,\tau)}{\partial \beta}\Big|_{\beta=\hat{\beta}^0, \tau=0} \\ \\ \frac{\partial l(\beta,\tau)}{\partial \tau}\Big|_{\beta=\hat{\beta}^0, \tau=0} \end{pmatrix} \quad = \quad \begin{pmatrix} \mathbf{0} \\ \\ \frac{\partial l(\beta,\tau)}{\partial \tau}\Big|_{\beta=\hat{\beta}^0, \tau=0} \end{pmatrix}
$$

$$
\theta_i = \boldsymbol{x}_i^t \boldsymbol{\beta} \Rightarrow \hat{\theta}_i^0 := \mathbf{x_i}^t \hat{\boldsymbol{\beta}}^0
$$

We assume

$$E((\theta_i^* - \theta_i)^r) \quad =: \quad d_r \quad \text{with} \quad d_r = o(\tau) \text{ for } r \geq 3$$

$$\Rightarrow \quad \frac{\partial l(\beta,\tau)}{\partial \tau}\bigg|_{\beta=\hat{\beta}^0,\tau=0} \quad \approx \quad \sum_{i=1}^{n} \frac{\frac{1}{2}k_i(\theta_i)\ f^{-1}(Y_i,\theta_i)\ \frac{\partial^2}{\partial \theta_i^{*2}} f(Y_i,\theta_i^*)|_{\theta_i^*=\theta_i}}{1 + \frac{1}{2}\ \tau\ k_i(\theta_i)\ f^{-1}(Y_i,\theta_i)\ \frac{\partial^2}{\partial \theta_i^{*2}} f(Y_i,\theta_i^*)|_{\theta_i^*=\theta_i}}\Bigg|_{\beta=\hat{\beta}^0,\tau=0}$$

$$= \quad \sum_{i=1}^{n} \underbrace{\frac{1}{2}k_i(\hat{\theta}_i^0)\ f^{-1}(Y_i,\hat{\theta}_i^0)\ \frac{\partial^2}{\partial \theta_i^{*2}} f(Y_i,\theta_i^*)|_{\theta_i^*=\hat{\theta}_i^0}}_{=:T_i(\hat{\theta}_i^0)}$$

(denominator $= 1$, since $\tau{=}0$)

Additionally

$$I(\hat{\boldsymbol{\theta}}^0) = \begin{pmatrix} I_{\beta\beta} & I_{\beta\tau} \\ I_{\beta\tau} & I_{\tau\tau} \end{pmatrix} \qquad \hat{\boldsymbol{\theta}}^0 = (\hat{\boldsymbol{\beta}}^{0t}, 0)^t$$

$$T_s = \left[\frac{\partial}{\partial\theta} l(\boldsymbol{\theta})\Big|_{\theta=(\hat{\beta}^0,0)} \right]^t I^{-1}(\hat{\boldsymbol{\beta}}^0, 0) \left[\frac{\partial}{\partial\theta} l(\boldsymbol{\theta})\Big|_{\theta=(\hat{\beta}^0,0)} \right]$$

$$= \left(\frac{\partial l(\beta,\tau)}{\partial\tau} \Big|_{\beta=\hat{\beta}^0,\tau=0} \right)^2 \left(I^{-1}(\hat{\boldsymbol{\beta}}^0, 0) \right)_{\tau\tau} = \left[\sum_{i=1}^n T_i(\hat{\theta}_i^0) \right]^2 \left(I^{-1}(\hat{\boldsymbol{\beta}}^0, 0) \right)_{\tau\tau}$$

since components corresponding to $\boldsymbol{\beta}$ are zero.
Further

$$I^{-1}(\hat{\boldsymbol{\beta}}^0, 0) = (I_{\tau\tau} - I_{\beta\tau}^t I_{\beta\beta}^{-1} I_{\beta\tau})^{-1} \qquad \text{(partioned matrixes)}$$

$$\Rightarrow \quad T_S = \left[\sum_{i=1}^n T_i(\hat{\theta}_i^0) \right]^2 / V^2, \qquad V^2 = (I_{\tau\tau} - I_{\beta\tau}^t I_{\beta\beta}^{-1} I_{\beta\tau})$$

Overdispersion test

Reject $H_0 : \tau = 0$ versus $H_1 : \tau > 0$ at level α

$\Leftrightarrow \quad T_S > \chi^2_{1,1-\alpha}$

Ex.: Poisson model with random effects

$$\theta_i^* = \boldsymbol{x_i^t}\boldsymbol{\beta} + Z_i \qquad E(Z_i) = 0 \qquad Var(Z_i) = \tau$$

$$Y_i|\theta_i^* \sim Poisson(e^{\theta_i^*}) \qquad \theta_i = \ln \mu_i = \boldsymbol{x_i^t}\boldsymbol{\beta}$$

$$\begin{aligned}
\Rightarrow E(Y_i) &= E_{\theta_i^*}(E(Y_i|\theta_i^*)) = E_{\theta_i^*}(e^{\theta_i^*}) \approx E_{\theta_i^*}(1+\theta_i^*) = 1 + \theta_i \\
&= 1 + \ln \mu_i \approx 1 + \mu_i - 1 = \mu_i \\
Var(Y_i) &= E_{\theta_i^*}\underbrace{(Var(Y_i|\theta_i^*))}_{E(Y_i|\theta_i^*)} + Var_{\theta_i^*}\underbrace{(E(Y_i|\theta_i^*))}_{e^{\theta_i^*}} \\
&= E(Y_i) + (e^{\theta_i})^2 \underbrace{Var(e^{Z_i})}_{\approx Var(1+Z_i) = Var(Z_i) = \tau} \\
&\approx \mu_i + \mu_i^2 \tau
\end{aligned}$$

$\Rightarrow$ can model overdispersion

Overdispersion test in Poisson model with random effect

Model: $Y_i|\theta_i^* \sim Poisson(e^{\theta_i^*})$ $\theta_i^* = \boldsymbol{x_i^t}\boldsymbol{\beta} + Z_i$ $E(Z_i) = 0$ $Var(Z_i) = \sigma^2$

Score test: Reject $H_0 : \tau = 0$ versus $H_1 : \tau > 0$ $\Leftrightarrow$

$$\frac{\left|\sum_{i=1}^{n} T_i(\hat{\theta}_i^0)\right|}{V} = \frac{\frac{1}{2}\sum_{i=1}^{n}\left[(Y_i - \hat{\mu}_i^0)^2 - \hat{\mu}_i^0\right]}{\sqrt{\frac{1}{2}\sum_{i=1}^{n}\hat{\mu}_i^{0^2}}} > Z_{1-\alpha/2}$$

References

Dean, C. (1992). Testing for overdispersion in Poisson and binomial regression models. *JASA 87*(418), 451–457.

McCullagh, P. and J. Nelder (1989). *Generalized linear models*. Chapman & Hall.