

Ein stochastisches Modell für multivariate Surveillance von Infektionskrankheiten

Mathias Hofmann, Volker Schmid, Michael Höhle,
Leonhard Held

SFB 386, Institut für Statistik, LMU München

Remagen, 08.10.2004



Ziel

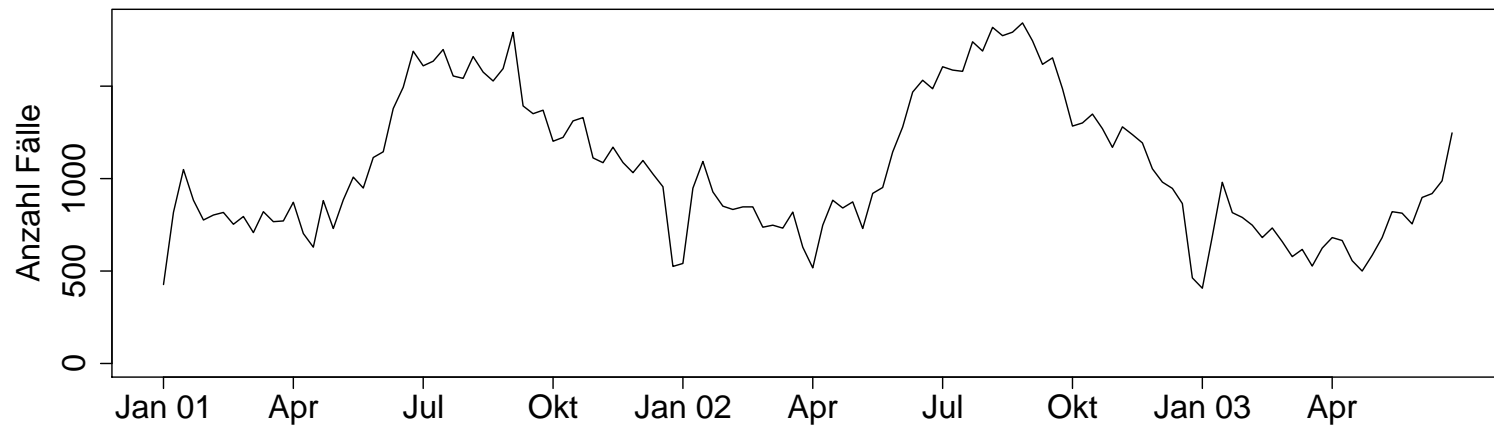
- Entwicklung von realistischen Raum-Zeit-Modellen für Surveillance-Daten

Idee

- Explizite Dekomposition der Gesamtinzidenz in endemische und epidemische Komponente.
- Modellierung durch Verzweigungsprozess mit Immigration.

Campylobacter-Daten vom RKI

Gesamtinzidenz seit 2001 in der BRD



Modell

- Gesamtinzidenz Z_{it} , $t = 1, \dots, n$, $i = 1, \dots, I$, setzt sich zusammen aus endemischen Fällen X_{it} und epidemischen Fällen Y_{it} :

$$Z_{it} = X_{it} + Y_{it}$$

$$X_{it} \sim \text{Poisson}(\nu \xi_i)$$

$$Y_{it} \sim \begin{cases} \text{Poisson}(\omega_{s,i} \cdot \frac{\lambda \nu \xi_i}{1-\lambda}) & t = 1, \\ \text{Poisson}(\lambda(Y_{i,t-1} + X_{i,t-1})) & t \geq 2. \end{cases}$$

- Parameter:
 - ν : Rate der endemischen Fälle
 - λ : Reproduktionsrate
 - $\omega_{s,i} = 1$ bzw. $\omega_{s,i} \sim \text{Ga}(\nu(1 + \lambda), \nu(1 + \lambda))$
- ξ_i Anteil der Population der Region i an der Gesamtpopulation

Überdispersion

- Die beobachtete Zeiteinheit (Wochen) entspricht typischerweise nicht einer “epidemischen” Zeiteinheit
- Simulationsstudien zeigen: Überdispersion umso größer, je stärker die Aggregation der epidemischen Zeiteinheiten
- Integration im Modell durch:

$$Y_{it} | \omega_{it} \sim \text{Poisson}(\lambda \omega_{it} (Y_{i,t-1} + X_{i,t-1}))$$
$$\omega_{it} \sim \text{Gamma}(\psi, \psi)$$

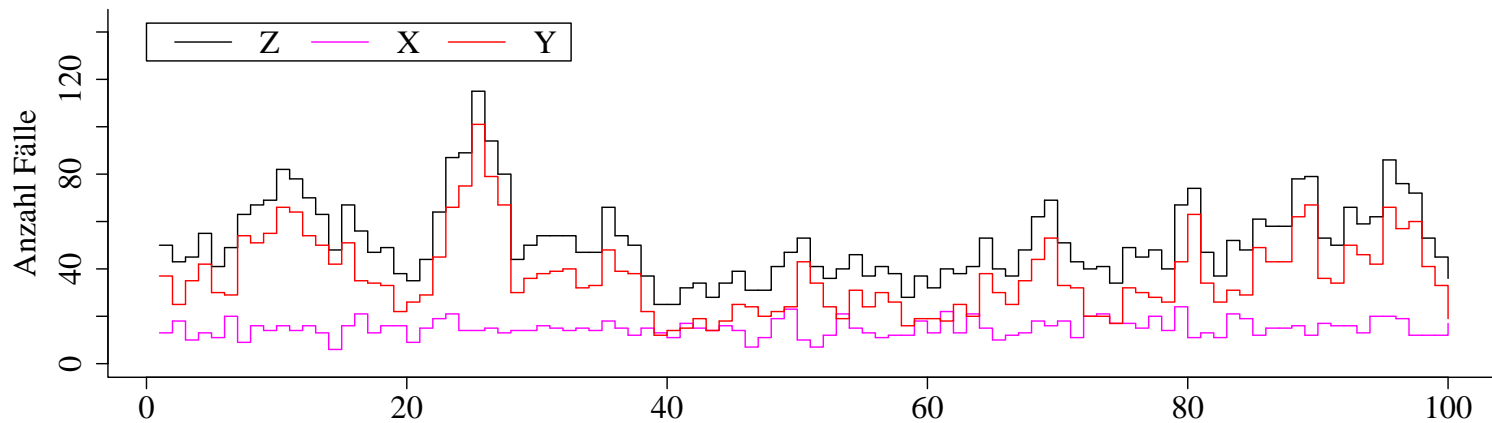
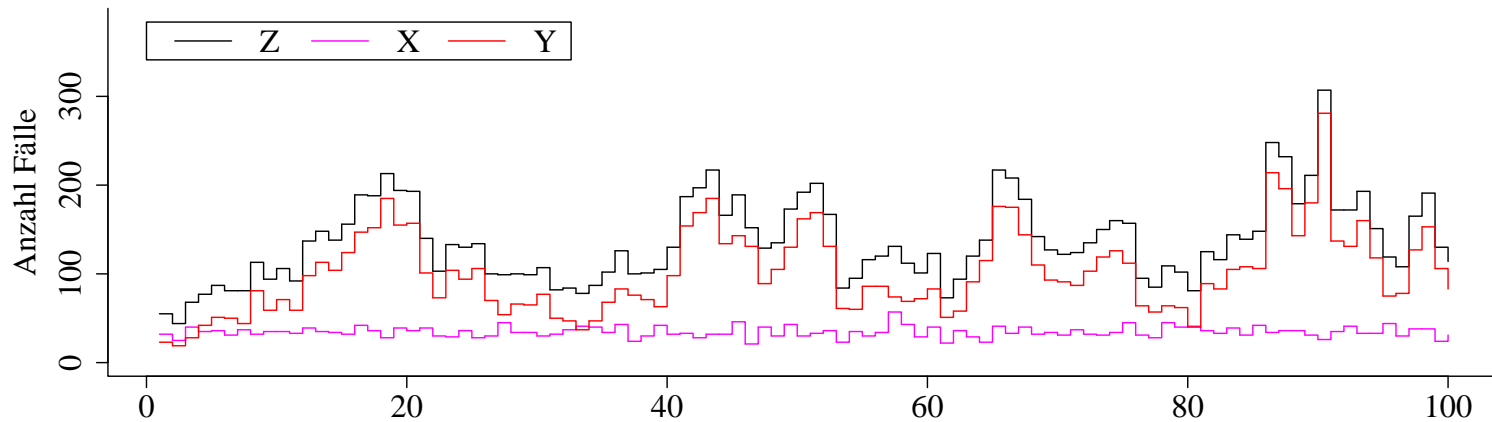
- Y_{it} negativ-binomial verteilt, $\text{Var}(Y_{it}) = E(Y_{it})(1 + E(Y_{it})/\psi)$
- weiterer Modellparameter ψ , mit Priori $\psi \sim \text{Gamma}(\alpha_\psi, \beta_\psi)$

Bayesianische Inferenz

- Schätzung der Modellparameter ν , λ und ψ mit MCMC
- Die latenten Größen X_{it} und Y_{it} werden als Parameter mitgeschätzt
- Beurteilung der Modellanpassung durch Berechnung der Posteriori-Devianz D
- Priori Verteilungen: $\nu \sim \text{Gamma}(\alpha_\nu, \beta_\nu)$, $\lambda \sim \text{Beta}(\alpha_\lambda, \beta_\lambda)$ und $\psi \sim \text{Gamma}(\alpha_\psi, \beta_\psi)$

Simulation

$I = 2, n = 100, \nu = 50, \lambda = 0.7, \psi = 20$ und $\xi = (0.7, 0.3)$



Raum-Zeit-Modellierung für ν

- Surveillance Daten haben häufig eine saisonale Struktur
- Modellierung der Saison durch raum-zeitlich variierendes ν mit saisonaler Komponente $\zeta(t)$:

$$\log(\nu_{it}) = \alpha_i + \beta_t + \zeta(t)$$

Räumliche Effekte

Die räumlichen Effekte α_i werden a priori unabhängig normalverteilt angenommen

$$\begin{aligned}\alpha_i &\sim \text{N}(\log(\xi_i), \sigma_\alpha^2) \\ \sigma_\alpha^2 &\sim \text{IG}(1, 0.001)\end{aligned}$$

Zeitlicher Trend

Für den Zeittrend β wird ein Random Walk zweiter Ordnung als Priori angesetzt

$$\begin{aligned}p(\beta_1) = p(\beta_2) &\propto \text{const.} \\ \beta_t &\sim \text{N}(2\beta_{t-1} - \beta_{t-2}, \sigma_\beta^2) \quad \text{für alle } t \geq 3 \\ \sigma_\beta^2 &\sim \text{IG}(1, 10^{-5})\end{aligned}$$

Saison-Komponente

Der saisonale Trend $\zeta(t)$ wird durch die Überlagerung von L Sinus-Schwingungen modelliert

$$\zeta(t) = \gamma_0 + \sum_{l=1}^L (A_l \sin(\rho j t + \phi_l)).$$

Dies entspricht der Überlagerung von von L Sinus- und L Cosinus-Schwingungen

$$\zeta(t) = \gamma_0 + \sum_{j=1}^J (\gamma_{2j-1} \sin(\rho j t) + \gamma_{2j} \cos(\rho j t))$$

mit $J = 2L$.

Saison-Komponente

Mit

$$\begin{aligned} s_j(t) &= \sin\left(\frac{\rho t j}{2}\right) && \text{für } j \text{ gerade,} \\ s_j(t) &= \cos\left(\frac{\rho t (j+1)}{2}\right) && \text{für } j \text{ ungerade} \\ s_0(t) &= 1. \end{aligned}$$

läßt sich die Modellierung der Saison-Komponente $\zeta(t)$ auf ein einfaches Regressionsproblem zurückführen

$$\zeta(t) = \sum_{j=0}^J (\gamma_j s_j(t)).$$

Als Priori für γ wird $\gamma \sim N(0, \sigma_\gamma^2 \mathbf{I})$ mit $\sigma_\gamma^2 = 10$ angenommen.

Ergebnisse für die raum-zeitlichen *Campylobacter* Daten beobachtet in den 16 Bundesländer

Vergleich der vier Modelle, jeweils mit fixem und variierendem ν und ohne und mit Überdispersion.

Überdisp.	ν fest		ν variierend	
	ohne	mit	ohne	mit
λ	0.948	0.956	0.424	0.594
ν	36.8	32.2	749.9	533.4
ψ	-	11.83	-	7.51
Devianz	8228.7	2439.7	5287.4	2399.9
p_D	12.9	1418.5	3.6	928.4
DIC	8215.7	3858.2	5283.8	3328.3

Tabelle 1: Posteriori-Mittelwerte der geschätzten Parameter.

Räumliche Effekte bei variierendem ν

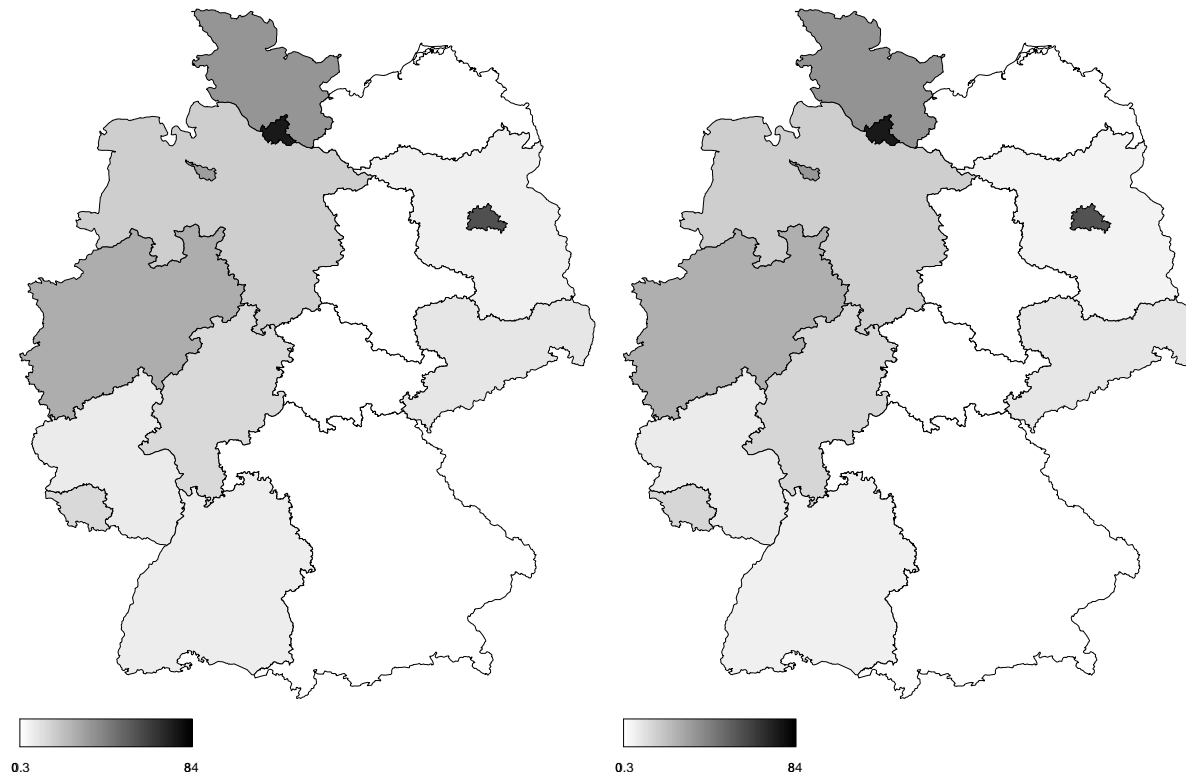


Abbildung 1: Räumlicher Effekt des variierenden ν nach Berücksichtigung der Population bei den Modellen ohne (links) und mit (rechts) Überdispersion.

Zeitlicher Trend bei variierendem ν

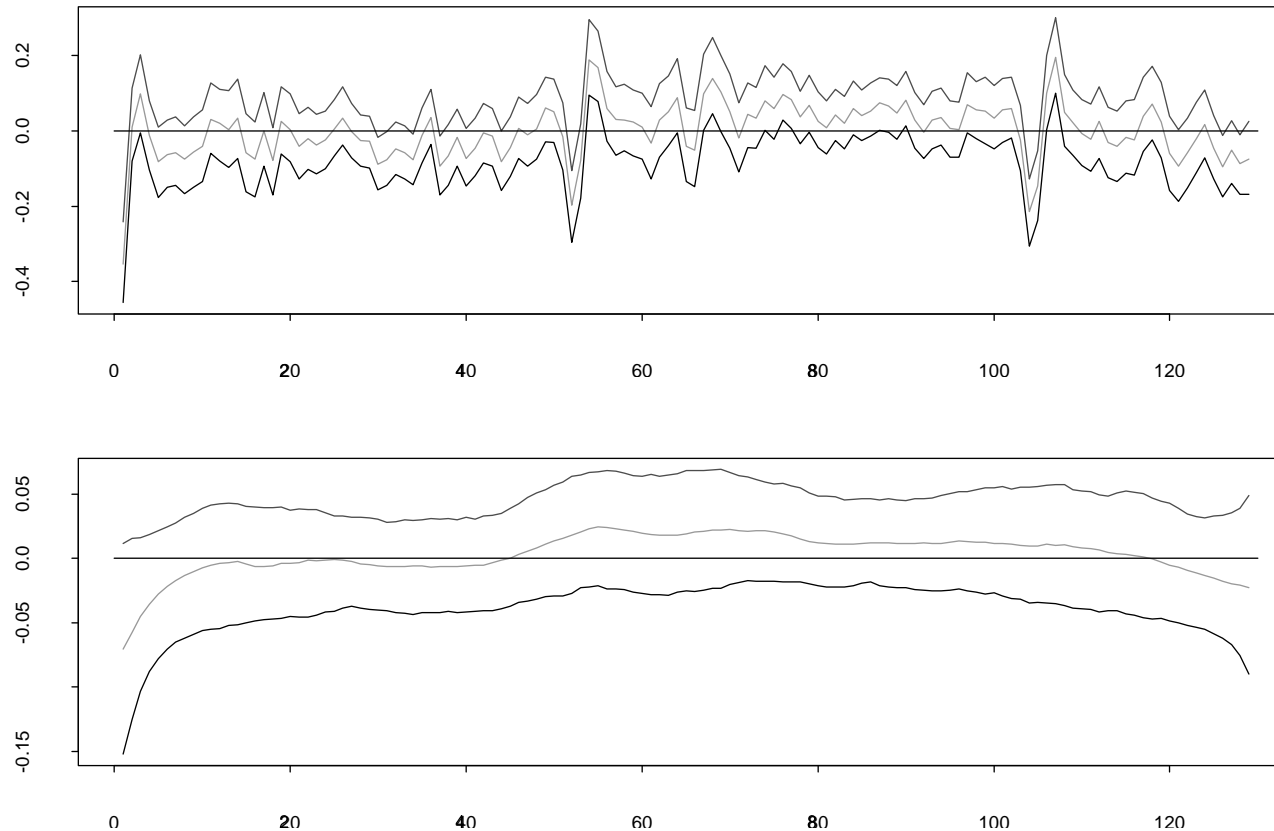


Abbildung 2: Zeittrend des variierenden ν beim Modell ohne (oben) und mit (unten) Überdispersion. Für den Zeittrend sind Median und 90%-Intervall abgebildet.

Saison-Komponente bei variierendem ν

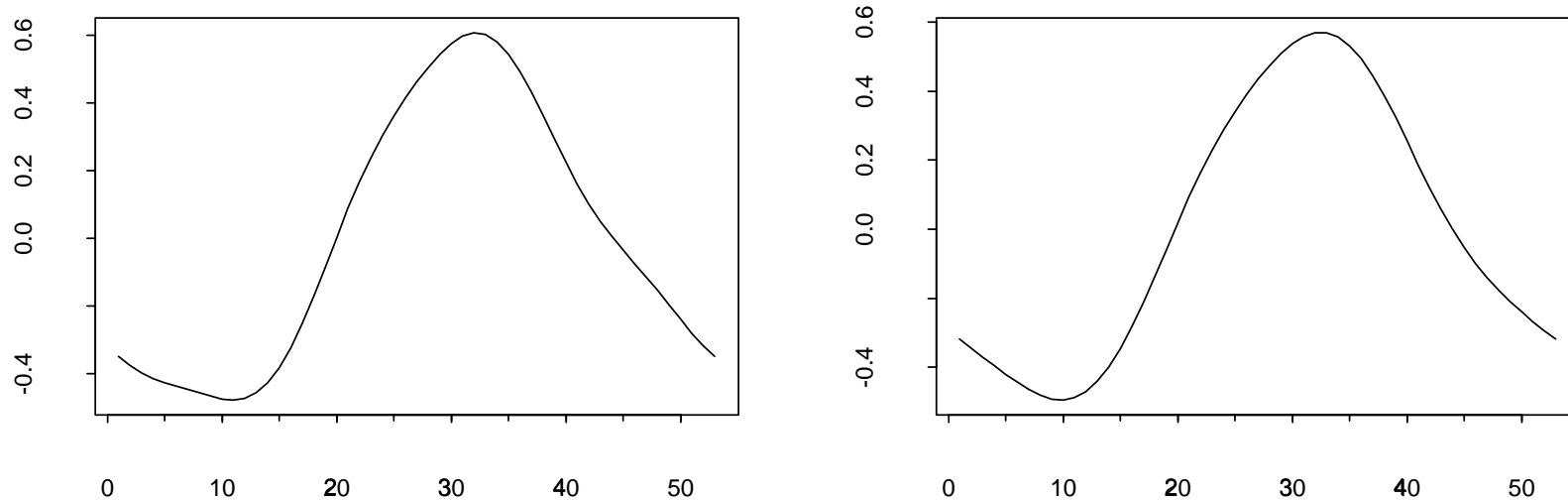
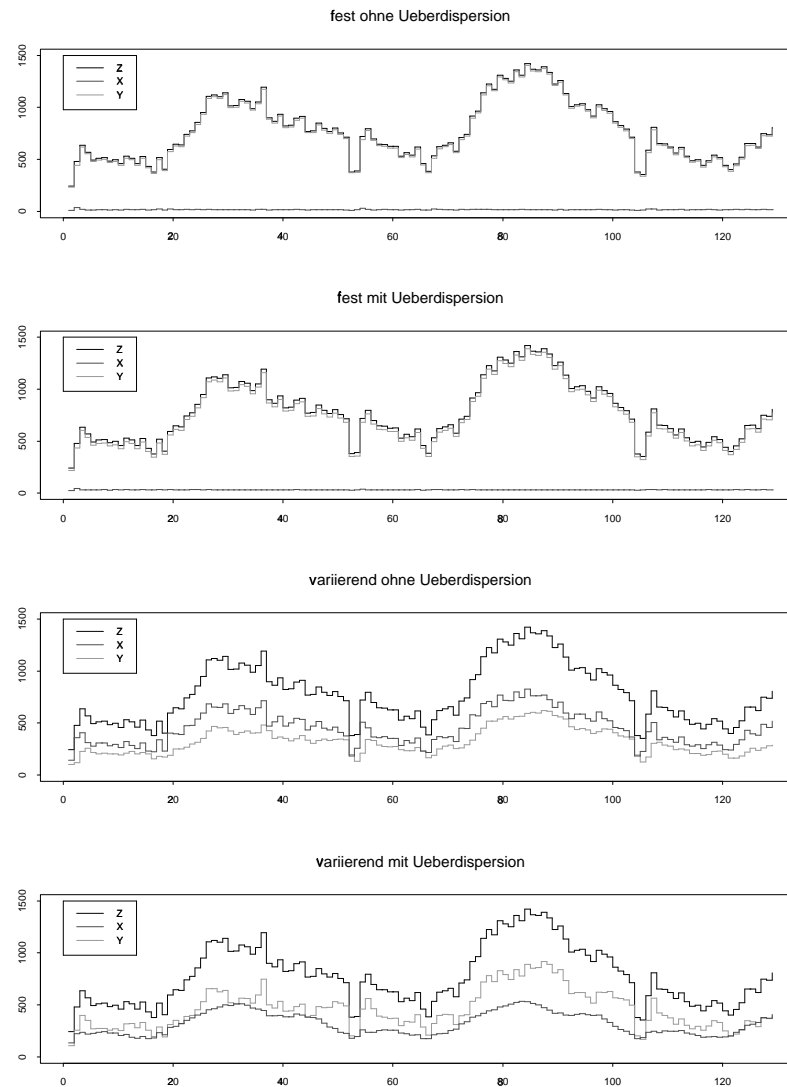
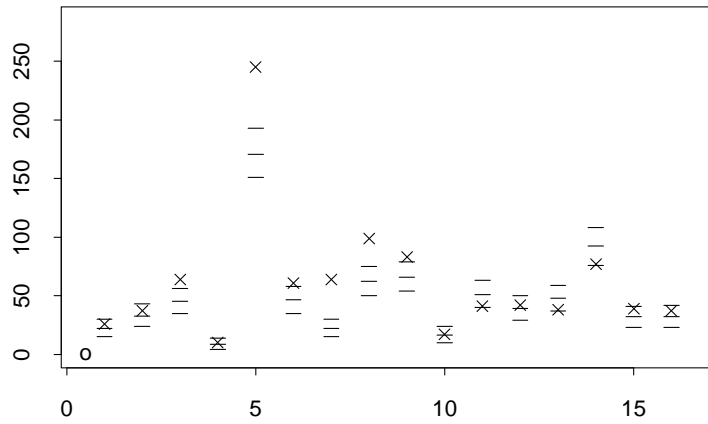


Abbildung 3: Saisonkomponente des variierenden ν beim Modell ohne (oben) und mit (unten) Überdispersion. Die Saisonkomponente wurde aus den Medianen der Regressionsparameter γ berechnet.

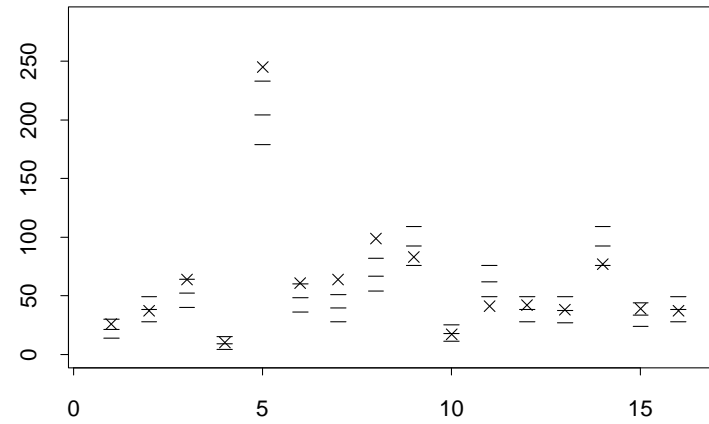
Endemische und epidemische Komponente



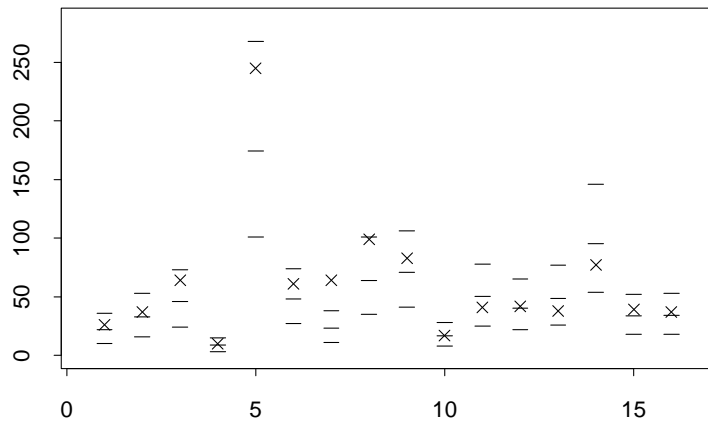
Prädiktive Verteilung



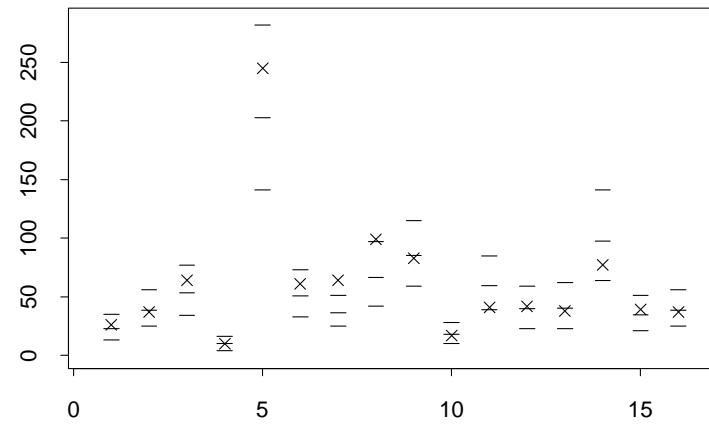
fix ohne Ueberdispersion



variierend ohne Ueberdispersion



fix mit Ueberdispersion



variierend mit Ueberdispersion

Erweiterungen

- Raum-zeitlich variierendes λ_{it} mit saisonaler Komponente und Aufhebung der Annahme $\lambda_{it} \leq 1$
- Berücksichtigung räumlicher Verbreitung der epidemischen Komponente zwischen Regionen durch multitype Verzweigungsprozess
- Einbeziehung von Kovariablen