State estimation of Covid-19 disease in Denmark

Kasper Schou Telkamp

2022-11-24

Contents

Introduction

Early detection of outbreaks with communicable diseases are of great importance in order to initiate timely interventions and help prevent disease spread. In this report non-normal mixed effects models will be evaluated on their ability to identify outbreaks of Covid-19 disease using data over new hospital admissions with Covid-19 in Denmark.

Materials and method

Data

In this project, the daily record of new hospital admissions with Covid-19 in Denmark grouped by region of residence and totals are used. The data is publicly available and were obtained from Statens Serum Institut (SSI) website[1]. SSI collects the data from the National Patient Registry (NPR), which contains information about outpatient contacts from Danish public as well as private hospitals. The data from NPR has some delay. Therefore, the inventory is updated daily with real-time data from the regions. The regions provide snapshot-data twice to SSI daily at 7am and 3pm. A hospital admission related to Covid-19 is defined as an admission, where a patient is admitted within 14 days after a positive SARS-CoV-2 test. Patients that are tested positive for SARS-CoV-2 during an admission is also registered as a Covid-19 related admission. Furthermore, admissions with Covid-19 are only registered for patients that are present in at least one snapshot, or if the patient have been admitted for more than 12 hours according to NPR.

In Figure 1 the total number of new admissions to the hospital are visualized.

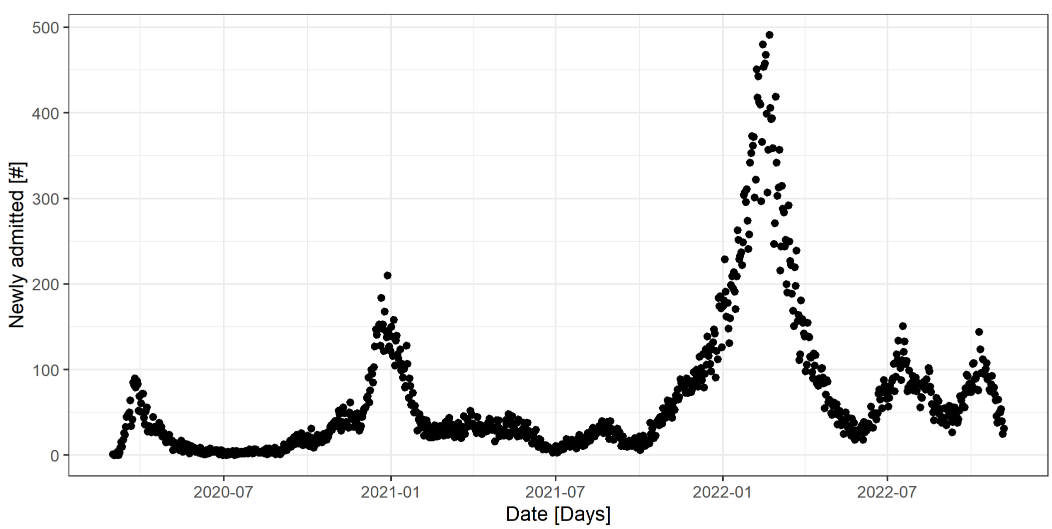

Figure 1: Daily number of new hospital admissions in Denmark.

In Figure 2 the daily number of new hospital admissions grouped by region of residence

[1] https://covid19.ssi.dk/overvagningsdata/download-fil-med-overvaagningdata

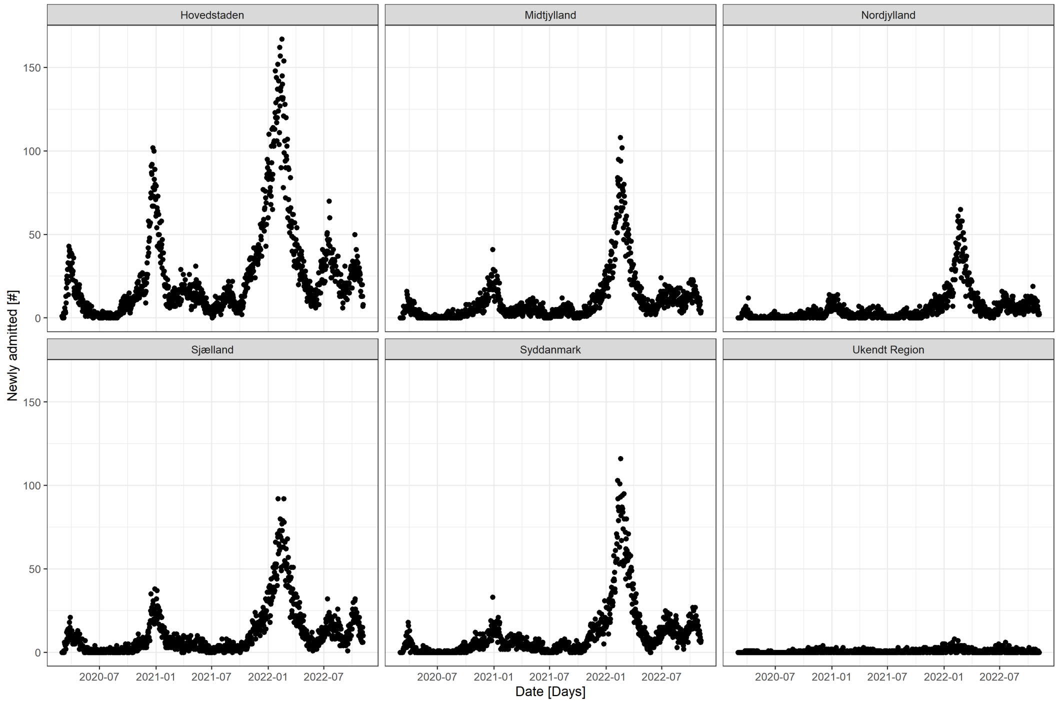

Figure 2: Total number of new hospital admissions in Denmark.

Generalized linear mixed effects model

Consider a hierarchical model for Y specified by

$$Y \sim \text{Pois}(\lambda) \tag{1a}$$
$$\log(\lambda) = \mathbf{X}\beta + u \tag{1b}$$

where $u \sim \text{N}\big(0, \Sigma(\psi)\big)$.

Seen in this book Madsen & Thyregod (2011)

Model

In order to analyze the data a state space model first proposed by Zeger (1988) and extended by Chan & Ledolter (1995), where the count observation y_i are assumed to follow a Poisson distribution with intensities λ_i is given by

$$\log(\lambda_i) = \mathbf{X}_i^T\beta + u_i$$

Here $\mathbf{X}_i$ is a ?-dimensional vector of covariates and β is the corresponding fixed effect model parameters. The random effects u_i are assumed to follow a first order autoregressive process

$$u_i = au_{i-1} + \epsilon_i$$

where $\epsilon \sim N(0, \sigma^2), i > 1$ is a white noise process, and a and σ are model parameters. Using the results from Madsen (2007), it is assumed that the first random effect follow the stationary distribution of the first order autoregressive process $u_1 \sim N(0, \sigma^2/(1-a^2))$.

The joint likelihood becomes

$$L(a, \beta, \sigma; \mathbf{y}) = \phi_{0, \frac{\sigma^2}{1-\sigma^2}}(u_1) \prod_{i=2}^{n} \left(\phi_{0,\sigma^2}(u_i - a u_{i-1})\right) \prod_{i=1}^{n} \left(p_{\lambda_i}(y_i)\right) \tag{2}$$

where ϕ_{μ,σ^2} is the probability density function (pdf) of the normal distribution with mean μ and variance σ_2, and p_λ is the pdf of the Poisson distribution with mean λ. In order to make computation of the joint likelihood function in (2) feasible, the estimation is carried out using the multivariate Laplace approximation.

Laplace approximation

The joint log-likelihood $l(a, \beta, \sigma; \mathbf{y}) = \log(L(a, \beta, \sigma; \mathbf{y})$ is approximated by a second order Taylor approximation around the optimum $\tilde{u} = \hat{u}_\theta$ of the log-likelihood function w.r.t. the unobserved random variables u, i.e.,

$$l(a, \beta, \sigma; \mathbf{y}) \approx l()$$

glmmTMB

This section describes the R package `glmmTMB` by Brooks et al. (2017) for linear and generalized linear mixed models (GLMMs) using Template Model Builder (TMB). The models are estimated using maximum likelihood estimation via TMB. Random effects are assumed to be Gaussian on the scale of the linear predictor and are integrated out using Laplace approximation. Additionally, gradients are calculated using automatic differentiation.

KFAS

This section goes into detail with the R package `KFAS` by Helske (2017) for state space modelling with observations from the exponential family.

Results

Estimating the total number of new hospital admissions in Denmark

In Figure 3 the total number of new hospitals admissions in Denmark is visualized together with the smoothed estimates from `glmmTMB` and `KFAS`.

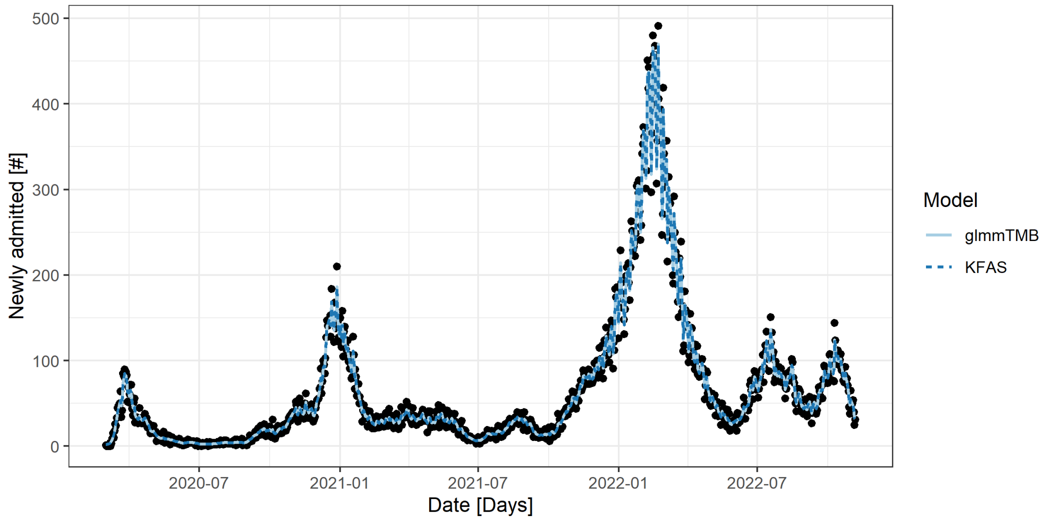

Figure 3: Total number of new hospital admissions in Denmark with smoothed estimates from glmmTMB and KFAS.

In Figure 4 the estimated latent intensity from `glmmTMB` and `KFAS` is visualized.

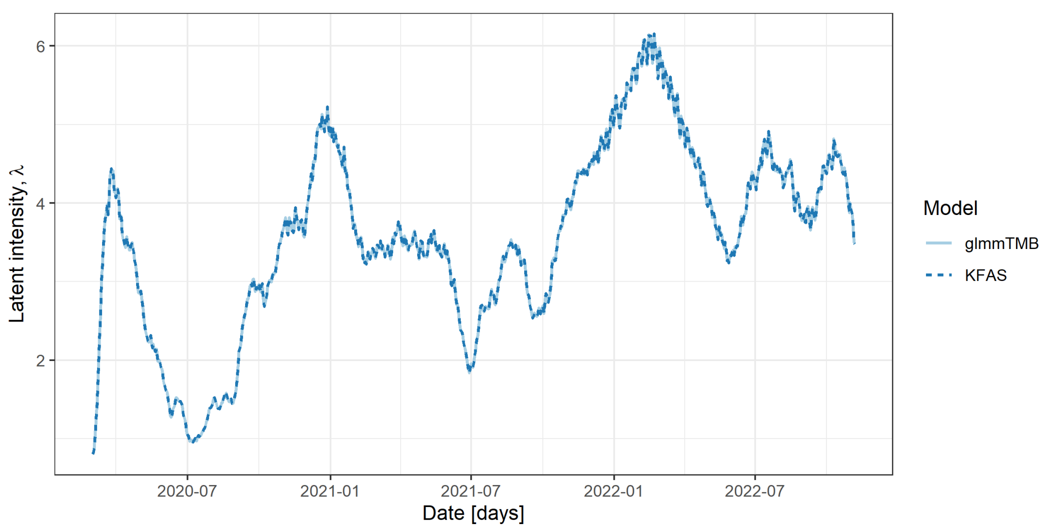

Figure 4: Estimated latent intensity from glmmTMB and KFAS.

Estimating the number of new hospital admissions in Denmark grouped by region

In Figure 5 the total number of new hospitals admissions in Denmark grouped by region is visualized together with the smoothed estimates from `glmmTMB` and `KFAS`.

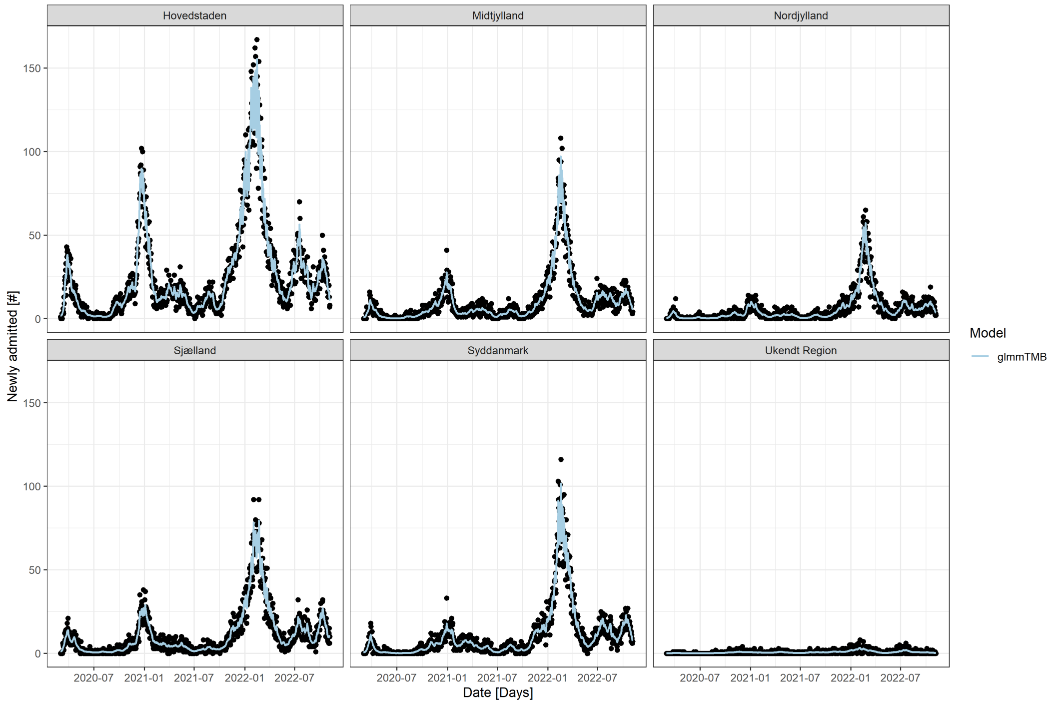

Figure 5: Total number of new hospital admissions in Denmark grouped by region with smoothed estimates from glmmTMB and KFAS.

Discussion

Write it later

References

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Maechler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, *9*(2), 378–400. https://doi.org/10.32614/RJ-2017-066

Chan, K. S., & Ledolter, J. (1995). Monte carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association*, *90*(429), 242–252. http://www.jstor.org/stable/2291149

Helske, J. (2017). KFAS: Exponential family state space models in R. *Journal of Statistical Software*, *78*(10), 1–39. https://doi.org/10.18637/jss.v078.i10

Madsen, H. (2007). *Time series analysis*. Chapman & Hall. https://doi.org/10.1201/9781420059687

Madsen, H., & Thyregod, P. (2011). *Introduction to general and generalized linear models*. CRC Press.

Zeger, S. L. (1988). A regression model for time series of counts. *Biometrika*, *75*(4), 621–629. http://www.jstor.org/stable/2336303