Comparing Prequential Selection Criteria for CIR and Continuous-type Spatio-temporal Models for Infectious Disease Data

Malesios C[1]., Demiris N[2]. and Kalogeropoulos K[3].
[1]Department of Rural Development, Democritus University of Thrace, Orestiada, Greece
[2]Department of Statistics, Athens University of Economics and Business, Athens, Greece
[3]Department of Statistics, London School of Economics, London, UK

Introduction

- In finance, the Cox-Ingersoll-Ross (CIR) model describes the evolution of interest rates.
- We model foot-and-mouth epidemic data, comparing a spatio-temporal regression-type model with a newly developed CIR model.
- We investigate the advantages of one-step-ahead predictions for assessing model fit in a Bayesian framework (Dawid, 1984).

Prequential methodology

- Examines each model's prediction for the next data point x_{n+1} after fitting the model to data $x_1, x_2, \ldots, x_n$.
- Criterion for model selection: distances, known as scoring rules, measuring the discrepancy between the forecasts and the data.
- The best model is the one with the smallest average score among the entertained models. We utilize four scoring rules for the evaluations:

Single-valued predictions:

$$AES = \left| x_t - \mu_{p_t} \right|$$

$$SES = \left(x_t - \mu_{p_t} \right)^2$$

Probabilistic predictions:

$$\log S = S\left(x_t, F_t\right) = -\log q(x_t)$$

$$RPS = \sum_{K=0}^{\infty} \left(P\left(X \le k\right) - 1\left(x \le k\right) \right)^2$$

Spatio-temporal models

The models are described by: Infections in $[t_i, t_{i+1}] \sim \text{Poisson}(\Lambda_i)$

$$\Lambda_i = \int_{t_i}^{t_{i+1}} \lambda_s ds$$

Model 1:

$$\lambda_i = \mu_i = \mathbf{X}_{(i)} \cdot \mathbf{\Theta}_{\beta} + K\left(d_i, \mathbf{\Theta}_{\mathbf{K}}\right) = \mathbf{X}_{(i)} \cdot \mathbf{\Theta}_{\beta} + \left(1 + \frac{d_i}{\alpha}\right)^{-\gamma}$$

Model 2: *CIR*

$$d\lambda_i = \phi(\mu_i - \lambda_i)dt + \sigma\sqrt{\lambda_i}\, dB_i$$

$$\mathbf{\Theta} = \left(\mathbf{\Theta}_{\beta}, \mathbf{\Theta}_{\mathbf{K}}\right)^t$$

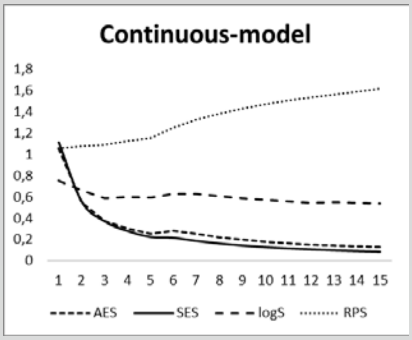
with model parameters including environmental factors, transmission kernel.

Results

Inference and sampling from the predictive distributions was performed in WinBUGS. Prequential analysis was done in R2winBUGS, a tool developed for combining WinBUGS and R.

Continuous type-model:

Model	Distribution	$\overline{SES}$	$\overline{AES}$	$\overline{\log S}$	$\overline{RPS}$	$\overline{D}$
Continuous-type spatio-temporal model	Poisson	2.933	1.025	0.887	10.436	4569

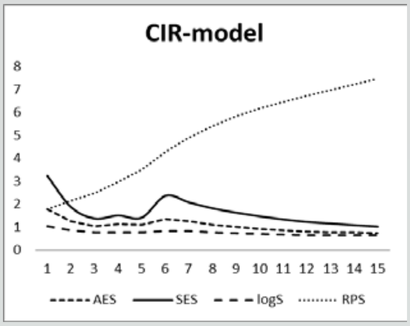


Series of one-step-ahead average prediction error estimates per mean prediction score for each data point.

The analysis confirmed theoretical claim of high sensitivity of RPS to distance (*Czado et al., 2009*).

CIR-model:

Model	Distribution	$\overline{SES}$	$\overline{AES}$	$\overline{\log S}$	$\overline{RPS}$	$\overline{D}$
CIR spatio-temporal model	Poisson	3.219	1.152	0.965	11.612	4972



Results did not confirmed the previous claim of sensitivity of logS to extreme cases (*Gneiting and Raftery, 2007*)

Parameter estimates indicated the relative importance of epidemic covariates in comparison to the endemic covariates for both models.

Covariate	Parameter Estimate: Continuous-type	Parameter Estimate: CIR
avg temp (β_1)	-0.000 (-0.003,0.002)	-0.004 (-0.07,0.096)
max temp (β_2)	0.000 (-0.012,0.009)	-0.058 (-0.128,0.029)
min temp (β_3)	-0.000 (-0.008,0.01)	0.074 (-0.007,0.141)
rel. humidity (β_4)	-0.000 (-0.004,0.006)	-0.008 (-0.129,0.14)
soil temp (β_5)	0.000 (-0.003,0.006)	0.019 (-0.063,0.097)
wind speed (β_6)	-0.000 (-0.004,0.004)	0.018 (-0.103,0.216)
α	**6.858 (0.353,22.2)**	**8.951 (0.346,16.41)**
γ	**6.759 (0.34,22.37)**	**1.527 (0.417,2.728)**

Discussion

- CIR and continuous type models showed that mortality during the epidemic is significantly associated with the spatial spread of the disease.
- Principled advantage of the prequential approach is that it respects the time ordering of the data and allows for the temporal assessment of each model's performance.
- We prefer the probabilistic one-step-ahead predictions based on logS and RPS since they properly account for uncertainty.