Spatial LC

Luiz Fernando Figueiredo

1 Introduction

Measuring mortality across populations to construct life tables is common practice in a variety of fields, such as: Demography, to help proposing tailored public policies, medicine, to measure and predict disease spread and impact, and actuarial sciences, to aid actuaries in pension and portfolio calculations. Modelling these mortality rates gives further insight about the subject population through the inferential process, capturing death trends and allowing for future behavior predictions.

Although many advances were made in the mortality rate modelling field in recent years, ranging from better data collection and filtering to state of the art stochastic models, smaller grids, such as counties or states, still suffer from missing data points due to the smaller population size and/or lack of data report resources.

Enter the spatial stochastic models: Designed to incorporate correlation between neighbour spaces thus allowing more data to be incorporated in the modelling process. This approach is built on the idea that neighbouring counties/places/points of interest in a data grid can carry information about the area they are close to, so that missing data problems in a individual area can readily be remediated by looking at the collective information of the immediate neighbourhood.

In this context, a stochastic model often used to fit mortality data is the Lee-Carter model (Lee and Carter), which introduces age and time specific parameters to accommodate temporal information about a specific population. Some extensions of the Lee-Carter model can also be found in literature, such as a bayesian approach (such as refBesag, refKnorr-Held, refPedroza), Poisson response (instead of the usual log normal response Brouhns et al) or discussing/adding an extra pair of parameters to model the spatial response to the Lee-Carter equation (ref LiandLee 2020, ref gomez rubio et al, refLiu-Sun-Wang).

ref gomez rubio et al used the spatio-temporal Lee Carter adaptation for the joint analysis of multiple diseases through three different causes of death registered in Spain at the province level. refLiu-Sun-Wang fitted their spatial Lee-Carter model to Japanese counties data to forecast mortality trends. Both expanded (to a certain degree) on the Poisson response Lee-Carter.

In this paper, we aim to extend on the bayesian Lee-Carter approach found in redPedroza to incorporate spatial information (as seen in those papers) and compare with the independent singular model, fitting Brazillian mortality data for Rio de Janeiro state at a county level, available via ref IBGE/DataSUS. We introduced a pair of spatial-related parameters to capture the correlation between neighbouring areas via CAR specification (refBanerjee), while maintaining the log normal response originally found in refLeeand-Carter. Practically, this means that we can still take advantage of the Kalman Filter (refKalmanFilter (?) to specify the temporal parameter and all of the extensions proposed by redPedroza could be applied to the model with no hassle (such as missing data treatment).

The remainder of the paper is organized as follows, in section 2...

2 Lee-Carter

The lee-carter model is described in refleecarter as:

$$ln(\mathbf{m}_{x,t}) = \mathbf{a}_x + \mathbf{b}_x k_t + \epsilon_{x,t} \tag{1}$$

where $\mathbf{a}$ and $\mathbf{b}$ are age-specific vectors and $\mathbf{k}$ is a time-varying index of the level of mortality. $\mathbf{a}$ represents the general shape across age of the mortality, $\mathbf{b}$ tells us which rates decline rapidly and wich rates decline slowly in response to changes in time $\mathbf{k}$. The error term $\epsilon_{x,t}$, with mean 0 and variance σ^2_ϵ, reflects particular age-specific historical influences not captured by the model. To ensure model identification, some constraints have been applied: $\mathbf{b}_x$ to sum to unity and $\mathbf{k}_t$ to sum to zero, which implies that the $\mathbf{a}_x$ are simply the averages over time of the log-mortality rates. por que negrito nos parametros k? ele especificado assim me parece ser so uma constante

Then, first presented in refPedroza2002, the Lee-Carter model can be reformulated as a state-space model:

$$\begin{aligned} \boldsymbol{y}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}\kappa_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t &\overset{iid}{\sim} N_p(\mathbf{0}, \sigma^2_\epsilon \boldsymbol{I}) \\ \kappa_t = \kappa_{t-1} + \theta + \omega_t, \quad \omega_t &\overset{iid}{\sim} N(0, \sigma^2_\omega) \end{aligned} \tag{2}$$

3 Spatial Lee-Carter

The approach to spatial modelling is based on the work found in refartigoPBLC...

$$\begin{aligned} \boldsymbol{y}^{(k)}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}\kappa_t + \boldsymbol{\gamma}\theta_k + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t &\overset{iid}{\sim} N_p(\mathbf{0}, \sigma^2_\epsilon \boldsymbol{I}) \\ \kappa_t = \kappa_{t-1} + \eta + \omega_t, \quad \omega_t &\overset{iid}{\sim} N(0, \sigma^2_\omega) \end{aligned} \tag{3}$$

with θ_i representing the spatial effects with respect to the region and $\boldsymbol{\gamma}$ the age spatial effects.

4 Estimation

The Bayesian estimation takes the usual form described in refpedroza, with a few changes to the FFBS and gibbs algorithm due to the addition of spatial parameters. Consider data with $x = 1, \ldots, n$ ages, $t = 1, \ldots, T$ years and $s = 1, \ldots, m$ regions.

4.1 The FFBS

Following refpedroza steps, the Forward Filtering Backward Sampling (FFBS) algorithm is applied to the estimation of parameter κ_t.

To incorporate the spatial data to the filtering recursions, we can rewrite the model equation stacking the log mortality rates to each region as such:

$$\mathbf{E}[\boldsymbol{Y}_t] = \boldsymbol{A\alpha} + \boldsymbol{A\beta}\kappa_t + (\boldsymbol{\gamma} \otimes \boldsymbol{\theta}) \tag{4}$$

where $\boldsymbol{Y}_t$ is a stacked log-mortality matrix with dimensions $nm \times 1$, $\boldsymbol{A}$ a $nm \times n$ stacking matrix for age related parameters and $(\boldsymbol{\gamma} \otimes \boldsymbol{\theta})$ representing the $nm \times 1$ kronecker matrix product between the age-vector $\boldsymbol{\gamma}$ and spatial vector $\boldsymbol{\theta}$). The stacked quantities can be illustrated as:

$$\boldsymbol{Y}_t = \begin{bmatrix} y_{t,1}^{(1)} \\ \cdots \\ y_{t,1}^{(m)} \\ y_{t,2}^{(1)} \\ \cdots \\ y_{t,n}^{(m)} \end{bmatrix}, \quad \boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix}, \quad (\boldsymbol{\gamma} \otimes \boldsymbol{\theta}) = \begin{bmatrix} \gamma_1\theta_1 \\ \cdots \\ \gamma_1\theta_m \\ \gamma_2\theta_1 \\ \cdots \\ \gamma_n\theta_m \end{bmatrix}, \tag{5}$$

the latent equation describing κ_t remains the same.

With our new specification, assuming that $(\kappa_{t-1} \mid D_{t-1}) \sim N(m_{t-1},\ R_{t-1})$ for each $t = 1, \ldots, p$, we can recalculate the Kalman filter quantities:

1. The 1-step ahead prior density is $(\kappa_t \mid D_{t-1}) \sim N(a_t,\ R_t)$, where

$$\begin{aligned} a_t &= \mathbf{E}[\kappa_t \mid D_{t-1}] = \mathbf{E}[\kappa_{t-1} + \eta + \omega_t \mid D_{t-1}] = \eta + m_{t-1} \\ R_t &= \mathbf{V}[\kappa_t \mid D_{t-1}] = \mathbf{V}[\kappa_{t-1} + \eta + \omega_t \mid D_{t-1}] = C_{t-1} + \sigma_\omega^2 \end{aligned} \tag{6}$$

2. The 1-step ahead predictive density is $(y_t \mid D_{t-1}) \sim N(\boldsymbol{f}_t,\ \boldsymbol{Q}_t)$, where

$$\begin{aligned} \boldsymbol{f}_t &= \mathbf{E}[\boldsymbol{Y}_t \mid D_{t-1}] = \boldsymbol{A\alpha} + \boldsymbol{A\beta}a_t + (\boldsymbol{\gamma} \otimes \boldsymbol{\theta}) \\ \boldsymbol{Q}_t &= \mathbf{V}[\boldsymbol{Y}_t \mid D_{t-1}] = \boldsymbol{A\beta}(R_t)\boldsymbol{\beta'A'} + \mathbf{1}\sigma_\epsilon^2 \end{aligned} \tag{7}$$

3. Finally, the posterior density is $(\kappa_t \mid D_t) \sim N(m_t,\ C_t)$, where

$$\begin{aligned} m_t &= a_t + R_t\boldsymbol{\beta'A'Q}_t^{-1}(\boldsymbol{Y}_t - \boldsymbol{f}_t) \\ C_t &= R_t - R_t\boldsymbol{\beta'A'Q}_t^{-1}\boldsymbol{A\beta}(R_t)R_t \end{aligned} \tag{8}$$

   resulting in a sample from $\kappa_t \sim N(m_t,\ C_t)$.

The Backward sampling algorithm remains as usual. For $T = p-1, p-2, \ldots, 0$, we sample from the smooth density $\pi(\kappa_T \mid D_{T+1}) \sim N(h_T, H_T)$, where:

$$\begin{aligned} B_T &= C_T R_{T+1}^{-1} \\ h_T &= m_T + B_T(\kappa_{T+1} - a_{T+1}) \\ H_T &= C_T - B_T R_{T+1} B_T' \end{aligned} \tag{9}$$

4.2 CAR specification

mudar pra IAR

The spatial parameter $\boldsymbol{\theta}$ is estimated via Conditional AutoRegressive (CAR) structure. According to refbanerjee and refartigoPBLC, we can write $\boldsymbol{\theta}$ as a multivariate Normal distribution in terms of mean vector and covariance matrix:

$$\boldsymbol{\theta} \sim N_m(\mathbf{0}, \sigma_\theta^2(\boldsymbol{M_\theta} - \lambda \boldsymbol{W_\theta})^{-1}) \tag{10}$$

where $\boldsymbol{M_\theta}$ is a diagonal $m \times m$ matrix with diagonal entries equal to the number of neighbours and $\boldsymbol{W_\theta}$ is an "adjancency matrix"whose generic ij entry $\omega_{ij} = 1$ if i and j are neighbours and $\omega_{ij} = 0$ otherwise.

The λ parameter is added to the CAR structure to guarantee a proper posterior distribution, being defined as $\lambda \in (\lambda_{min}^{-1}, \lambda_{max}^{-1})$ where λ_{min} and λ_{max} are the smallest and largest eigenvalues in $\boldsymbol{W_\theta}$. refbanerjee discusses the real usability of the parameter λ... TBD...

Denoting $\boldsymbol{Q_\theta} = \boldsymbol{M_\theta} - \lambda \boldsymbol{W_\theta}$, we can write

$$\boldsymbol{\theta} \sim N_m(\mathbf{0}, \sigma_\theta^2 \boldsymbol{Q_\theta}^{-1}). \tag{11}$$

Then, assuming the following conditional distribution of θ_i:

$$\begin{aligned} &\theta_i | \theta_{-i} \sim N(\lambda \sum_{i \neq j} b_{ij}\theta_j, \sigma_i^2), \quad \textit{where} \;\; i = 1, 2, \ldots, m, \\ &b_{ij} = \frac{\omega_{ij}}{\omega_{i+}}, \quad \sigma_i^2 = \frac{\sigma_\theta^2}{\omega_{i+}} \quad \text{and} \quad \omega_{i+} = \sum_j \omega_{ij} \end{aligned} \tag{12}$$

we can estimate the spatial parameters θ_i through a Gibbs sampler.

4.3 Gibbs sampling

Bayes theorem and posterior conditional complete distributions for the parameters in order to sample with Gibbs.

4.3.1 Prior specification

For prior specifications, we mainly follow refpedroza with the noninformative and flat priors:

$$
\begin{aligned}
& p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \eta) \propto 1 \\
& p(\sigma_\epsilon^2) \propto 1/\sigma_\epsilon^2 \\
& p(\sigma_\omega^2) \propto 1/\sigma_\omega \\
& p(\sigma_\theta^2) \propto 1/\sigma_\theta \\
\\
& \kappa_0 \sim N(m_0, C_0) \\
& \theta_i|\theta_{-i} \sim N(\lambda \sum_{i \neq j} b_{ij}\theta_j, \sigma_i^2)
\end{aligned}
\tag{13}
$$

with m_0 and C_0 assumed to be known.

4.3.2 Posterior specification

The likelihood function can be written as:

$$
\begin{aligned}
L(.) &= \prod_{x=1}^{n}\prod_{t=1}^{p}\prod_{k=1}^{m} N(\alpha_x + \beta_x\kappa_t + \gamma_x\theta_k, \quad \sigma_\epsilon^2) \\
&\propto (1/\sigma_\epsilon^2)^{\frac{npm}{2}} exp\left(-\frac{1}{2\sigma_\epsilon^2}\sum_{x=1}^{n}\sum_{t=1}^{p}\sum_{k=1}^{m}(y_{x,t,k} - \alpha_x - \beta_x\kappa_t - \gamma_x\theta_k)^2\right)
\end{aligned}
\tag{14}
$$

Then, following the bayes theorem, we can obtain the complete conditional distributions for the parameters to be estimated.

The posterior distribution for the σ_ϵ^2 parameter can be obtained as:

$$
\begin{aligned}
\pi(\sigma_\epsilon^2 \mid .) &= L(.) \times p(\sigma_\epsilon^2) \\
&= \prod_{x=1}^{n}\prod_{t=1}^{p}\prod_{k=1}^{m} N(\alpha_x + \beta_x\kappa_t + \gamma_x\theta_k, \quad \sigma_\epsilon^2) \times 1/\sigma_\epsilon^2 \\
&\propto (1/\sigma_\epsilon^2)^{\frac{npm}{2}+1} exp\left(-\frac{1}{2\sigma_\epsilon^2}\sum_{x=1}^{n}\sum_{t=1}^{p}\sum_{k=1}^{m}(y_{x,t,k} - \alpha_x - \beta_x\kappa_t - \gamma_x\theta_k)^2\right)
\end{aligned}
\tag{15}
$$

$$
\pi(\sigma_\epsilon^2 \mid .) \sim \text{InvGamma}\left(\frac{npm}{2}, \ \frac{1}{2}\sum_{x=1}^{n}\sum_{t=1}^{p}\sum_{k=1}^{m}(y_{x,t,k} - \alpha_x - \beta_x\kappa_t - \gamma_x\theta_k)^2\right)
$$

The posterior distribution for the $(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma})$ parameters can be built upon the refpedroza approach, performing separate regressions for each age group of y_x on $\boldsymbol{\kappa}$ and $\boldsymbol{\theta}$ as:

$$
\pi(\alpha_x, \beta_x, \gamma_x \mid \boldsymbol{\kappa}, \boldsymbol{\theta}, \sigma_\epsilon^2) \sim N((\boldsymbol{X'X})^{-1}\boldsymbol{X'y_x}, \ \sigma_\epsilon^2(\boldsymbol{X'X})^{-1})
\tag{16}
$$

where $\boldsymbol{X} = (\mathbf{1},\ \boldsymbol{\kappa},\ \boldsymbol{\theta})$ is a $nm \times 3$ matrix, with $\boldsymbol{\kappa}$ and $\boldsymbol{\theta}$ vectors for the time and space parameters respectively, and $\boldsymbol{y_x}$ is the log-mortality vector with respect to age x. Both are stacked on the time and space parameters, such as:

$$\boldsymbol{y_x} = \begin{bmatrix} y_{1,x}^{(1)} \\ \cdots \\ y_{1,x}^{(m)} \\ y_{2,x}^{(1)} \\ \cdots \\ y_{p,x}^{(m)} \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} 1 & \kappa_1 & \theta_1 \\ \cdots & \cdots & \cdots \\ 1 & \kappa_1 & \theta_m \\ 1 & \kappa_2 & \theta_1 \\ \cdots & \cdots & \cdots \\ 1 & \kappa_p & \theta_m \end{bmatrix} \tag{17}$$

The posterior distributions for the drift parameter η and σ^2_ω are the same as seen in refpedroza, since they don't depend on the space parameter

$$\begin{aligned} &\pi(\eta \mid \boldsymbol{\kappa},\ \kappa_0,\ \sigma^2_\omega) \sim N\left(\frac{\kappa_p - \kappa_0}{p},\ \frac{\sigma^2_\omega}{p}\right) \\ &\pi(\sigma^2_\omega \mid \boldsymbol{\kappa},\ \kappa_0,\ \eta) \sim \text{InvGamma}\left(\frac{p-1}{2},\ \frac{1}{2}\sum_{t=1}^{p}(\kappa_t - \kappa_{t-1} - \eta)^2\right) \end{aligned} \tag{18}$$

The posterior distribution for the σ^2_θ parameter can be written as:

$$\begin{aligned} \pi(\sigma^2_\theta \mid \lambda,\ \boldsymbol{\theta}) &= \pi(\boldsymbol{\theta} \mid \lambda,\ \sigma^2_\theta) \times p(\sigma^2_\theta) \\ &= N_p(0,\ \sigma^2_\theta \boldsymbol{Q}_\theta^{-1}) \times 1/\sigma_\theta \\ &\propto det(\sigma^2_\theta \boldsymbol{Q}_\theta^{-1})^{-\frac{1}{2}} exp\left(-\frac{1}{2}\boldsymbol{\theta'}(\sigma^2_\theta \boldsymbol{Q}_\theta^{-1})^{-1}\right) \times 1/\sigma_\theta \\ &\propto (1/\sigma^2_\theta)^{\frac{n}{2}}(1/\sigma^2_\theta)^{\frac{1}{2}} exp\left(-(\frac{1}{2}\boldsymbol{\theta'}\boldsymbol{Q}_\theta\boldsymbol{\theta})/\sigma^2_\theta\right) \\ &\propto (1/\sigma^2_\theta)^{\frac{n-1}{2}+1} exp\left(-(\frac{1}{2}\boldsymbol{\theta'}\boldsymbol{Q}_\theta\boldsymbol{\theta})/\sigma^2_\theta\right) \end{aligned} \tag{19}$$

$$\pi(\sigma^2_\theta \mid \lambda,\ \boldsymbol{\theta}) \sim \text{InvGamma}\left(\frac{n-1}{2},\ \frac{1}{2}\boldsymbol{\theta'}\boldsymbol{Q}_\theta\boldsymbol{\theta}\right)$$

The posterior distribution for the θ parameter is obtained through iteration on the $k = 1, \ldots, m$ regions:

$$\begin{aligned} \pi(\boldsymbol{\theta_k} \mid .) &= L(.) \times p(\theta_k \mid \theta_{-k}) \\ &= \prod_{x=1}^{n}\prod_{t=1}^{p} N(\alpha_x + \beta_x\kappa_t + \gamma_x\theta_k,\ \ \sigma^2_\epsilon) \times N(\lambda\sum_{k\neq j} b_{kj}\theta_j, \sigma^2_k) \end{aligned} \tag{20}$$

FAZER PRO THETA ....

4.3.3 Constraints

To ensure model identification for the SBLC model, we apply the constraints seen in refLeeandCarter to the new spatial parameters: γ_x and θ_s following the same constraints as β_x and κ_t, respectively.

We apply these constraints as a "block-step"inside the Gibbs algorithm, transforming the parameters as follows:

$$\beta_x^* \leftarrow \frac{\beta_x}{\beta_+}, \quad \beta_+ = \sum_x \beta_x \qquad \kappa_t^* \leftarrow \kappa_t - \overline{\kappa}, \quad \overline{\kappa} = \frac{1}{T}\sum_t^T \kappa_t$$

$$\gamma_x^* \leftarrow \frac{\gamma_x}{\gamma_+}, \quad \gamma_+ = \sum_x \gamma_x \qquad \theta_s^* \leftarrow \theta_s - \overline{\theta}, \quad \overline{\theta} = \frac{1}{m}\sum_s^m \gamma_s$$

As seen in refCzado2005, these transformations reflect on other parameters:

$$\alpha_x^* \leftarrow \alpha_x + \beta_x\overline{\kappa} + \gamma_x\overline{\theta} \qquad \kappa_t^* \leftarrow \kappa_t^*\beta_+ \qquad \theta_s^* \leftarrow \theta_s^*\gamma_+$$

so that the model equation leading to the response parameter $\boldsymbol{y}$ remains unchanged:

$$\begin{aligned}
\boldsymbol{y}_t^{(s)} &= \alpha_x^* + \beta_x^*\kappa_t^* + \gamma_x^*\theta_s^* \\
&= (\alpha_x + \beta_x\overline{\kappa} + \gamma_x\overline{\theta}) + \frac{\beta_x}{\beta_+}(\kappa_t - \overline{\kappa})\beta_+ + \frac{\gamma_x}{\gamma_+}(\theta_s - \overline{\theta})\gamma_+ \\
&= \alpha_x + (\beta_x\overline{\kappa} - \beta_x\overline{\kappa}) + (\gamma_x\overline{\theta} - \gamma_x\overline{\theta}) + \beta_x\kappa_t + \gamma_x\theta_s \\
&= \alpha_x + \beta_x\kappa_t + \gamma_x\theta_s
\end{aligned}$$

4.4 Missing data

5 Application

6 Conclusion