A Class of Random Field Memory Models for Mortality Forecasting

Yahia Salhi, ISFA, Université Lyon 1 / BNP Paribas Cardif - DAMI Chair

joint work with: P. Doukhan (Université de Cergy), J. Rynkiewicz (Sorbonne University) and D. Pommeret (Aix-Marseille University)

About the speaker

- **Yahia Salhi**
- Assistant Professor
- Yahia holds a Phd in applied mathematics from the University of Luon, a MSc in actuarial science and finance from ISFA, and a engineering diploma from Ecole des Mines. He is research associate at the BNP Paribas Cardif chair 'DAMI'. Yahia's main research interests include detection of abrupt changes, mortality modelling, pricing and management as well as surrender risk.

CHAIR OF EXCELLENCE
Data Analytics & Models for Insurance
BNP PARIBAS CARDIF | LABORATOIRE SAF

- **ISFA, Université Lyon 1 / DAMI Chair**

Agenda

Mortality Stylized Facts

Random Fields Models

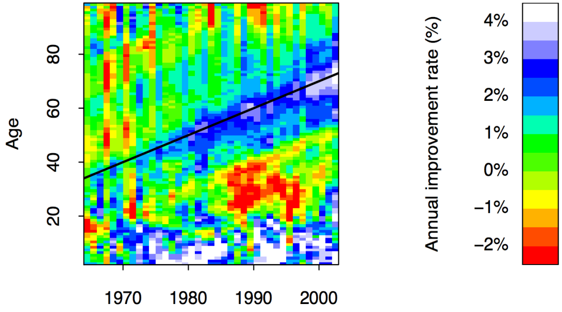
Illustrative Example

Some Statistical Inference

Empirical Analysis

Conclusion

Mortality Stylized Facts (1/4)



- Cohort effect (Willets [2004])
- Cross-cohort correlation (Loisel and Serant [2007], Jevtić et al. [2013] and Mavros et al. [2016])
- Heteroskedasticity (Giacometti et al. [2012], Chai et al. [2013] and Chen et al. [2015])

We propose to model the whole **surface of improvement rates**.

Mortality Stylized Facts (2/4)

Consider the process X_s parameterized by the lattice points $s = (a, t)$, with $a =$age and $t =$time, and defined as the centered mortality improvement rates, i.e., $\text{IR}_s = \log(q_{a,t}/q_{a,t-1})$

$$X_s = \text{IR}_s - \overline{\text{IR}},$$

such that the random field $(X_s)_{s \in I \times J}$ is:

- **Stationary**, in the sense of Doukhan and Truquet [2007]
- **Markov**, such that X_s depends on the information $\{X_{u,v}\,;\, u < a, v < t\}$, see Loisel and Serant [2007]

Mortality Stylized Facts (3/4)

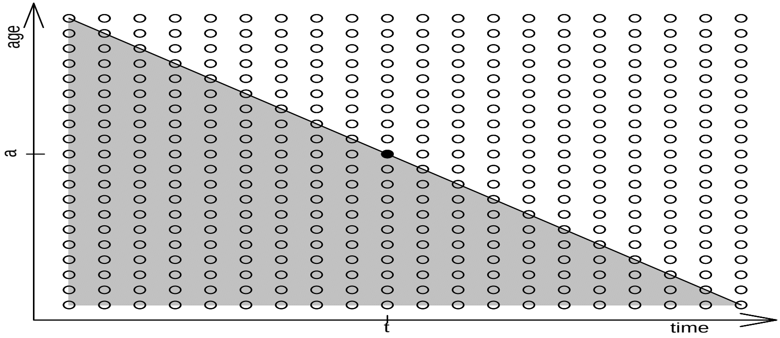


Figure: A bi-dimensional representation of the random field $X_s = \mathrm{IR}_s - \overline{\mathrm{IR}}$, with $s = (a, t)$.

Random Fields Models (1/5)

A candidate model can be specified in the following form:

$$X_s = F\Big((X_{s-v})_{v\in V}, \theta, \xi_s\Big), \quad \text{with } s \in \mathbb{N}^2,$$

where:

- F is a given parametric function taking values in $\mathbb{R}$, i.e. $F : \mathbb{R}^V \times \Theta \times \mathbb{R} \to \mathbb{R}$
- $V \subset \mathbb{N}^2 \setminus \{0\}$ is a neighborhood (characterized by the Markov property)
- $(\xi_s)_{s\in\mathbb{N}^2}$ is an independent identically distributed (i.i.d.) random field

Random Fields Models (2/5)

We consider the following assumptions (contraction principle):

A-1 $\|F(x_0, \theta, \xi)\|_p < \infty$ for some $x_0 \in \mathbb{R}^V$,

A-2 $\|F(x', \theta, \xi) - F(x, \theta, \xi)\|_p < \sum_{v \in V} \alpha_v \|x'_v - x_v\|$ for all $x = (x_v)_{v \in V}, x' = (x'_v)_{v \in V} \in \mathbb{R}^V$, where the coefficients α_v are such that $\sum_{v \in V} \alpha_v < 1$.

to ensure the existence and the uniqueness of a stationary solution.

Random Fields Models (3/5)

Based on this abstract formulation, we propose a specific form for the function F that is intended to capture the various stylized facts.

- We decompose the random field into a *conditional mean* m_s and *conditional variance* σ_s^2 in such a way that :

$$
\begin{aligned}
m_s &= \mathbb{E}\Big[X_s \Big| \{X_{u,v};\, u < a, v \le t\}\Big] = \sum_{v \in V_1} \beta_v X_{s-v} \\
\sigma_s^2 &= \mathrm{Var}\Big(X_s \Big| \{X_{u,v};\, u < a, v \le t\}\Big) = \alpha_0 + \sum_{v \in V_2} \alpha_v X_{s-v}^2,
\end{aligned}
$$

- V_1 and V_2 are two neighbourhoods characterizing the evolution of m_s and σ_s^2 of each X_s in terms of its own past values and the present and past values of the adjacent cohorts

Random Fields Models (4/5)

The combined model is referred to as the **AR-ARCH random field** and is given by

$$X_s = \xi_s \sqrt{\alpha_0 + \sum_{v \in V_2} \alpha_v X_{s-v}^2} + \sum_{v \in V_1} \beta_v X_{s-v}. \quad (1)$$

- The model (1) is a generalization of the now widely used AR-ARCH models for random processes.

Random Fields Models (5/5)

ICA CIA BERLIN 2018

- The function F is given by

$$F(x,\theta,z) = z\Big(\alpha_0 + \sum_{v\in V_2}\alpha_v x_v^2\Big)^{1/2} + \sum_{v\in V_1}\beta_v x_v,$$

for $x = (x_s)_{s\in\mathbb{N}^2}$, $\theta = ((\alpha_s)_{s\in V_2}, (\beta_s)_{s\in V_1})$, and $z\in\mathbb{R}$.

- The conditions **A-1** and **A-2** for the existence and uniqueness of stationary solution writes:

$$\|\xi_0\|_p < \infty, \quad \kappa_p \equiv \|\xi_0\|_p \textstyle\sum_{s\in V_2}\alpha_s + \sum_{s\in V_1}\beta_s < 1, \quad p\geq 1.$$

Illustrative Example (1/2)

Assume a Markov property for the random field such that the following property holds

$$\mathcal{L}(X_s | \{X_{u,v};\, u < a, v \leq t\}) = \mathcal{L}(X_s | X_{s^-}, X_{s^+}, X_{s^=}),$$

where

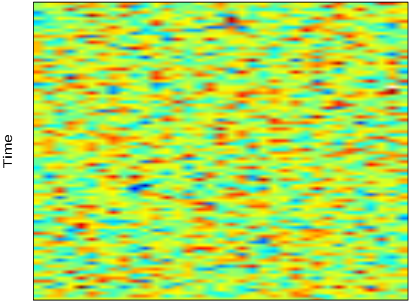
$$\underbrace{s^- = (a-1,t)}_{\text{young cohort}}, \quad \underbrace{s^+ = (a,t-1)}_{\text{old cohort}}, \quad \underbrace{s^= = (a-1,t-1)}_{\text{auto-regressive}}.$$

An example of potential models can be described using the causal neighborhoods $V_2 = \{(1,0),(0,1)\}$ and $V_1 = \{(1,1)\}$; so that the model in (1) can be simply rewritten as

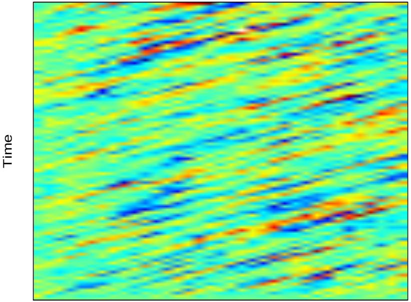
$$X_s = \xi_s \sqrt{\alpha_0 + \alpha^- X_{s^-}^2 + \alpha^+ X_{s^+}^2} + \underbrace{\beta}_{\text{cohort effect}} X_{s^=},$$

Illustrative Example (2/2)

$$X_s = \xi_s \sqrt{\alpha_0 + \alpha^- X_{s^-}^2 + \alpha^+ X_{s^+}^2} + \overbrace{\color{red}{\beta}}^{\text{cohort effect}} X_{s^=},$$



(a) $\beta \underset{\text{Age}}{=} 0.1$



(b) $\beta \underset{\text{Age}}{=} 0.5$

Some Statistical Inference (1/3)

ICA CIA BERLIN 2018

- We consider an approximation of the MLE called Quasi- Maximum Likelihood Estimator (QMLE):

$$L_T\left(x_s, s \in \mathcal{O}; \theta\right) = \frac{1}{T}\left(\sum_{s \in \mathcal{O}} -\frac{1}{2}\ln\left(\alpha_0 + \sum_{v \in V_2} \alpha_v x_{s-v}^2\right) - \frac{\left(x_s - \sum_{v \in V_1} \beta_v x_{s-v}\right)^2}{2\left(\alpha_0 + \sum_{v \in V_2} \alpha_v x_{s-v}^2\right)}\right).$$

- We will consider the estimator based on maximizing the above function (QMLE) over the set Θ, which will be denoted $\widehat{\theta}_T$, i.e.

$$\widehat{\theta}_T = \arg\max_{\theta \in \Theta} L_T\left(x_s, s \in \mathcal{O}; \theta\right), \qquad (2)$$

where Θ is the set of possible parameters.

Some Statistical Inference (2/3)

ICA CIA BERLIN 2018

We will need the following assumptions:

H-1 Finite second order moment, i.e. $\mathbb{E}(X_s^2) < \infty$.

H-2 The model is identifiable

H-3 The set of possible parameters Θ is compact and the true parameter θ^0 of the model (1) belongs to the interior of Θ.

Theorem (Consistency)

If the assumptions **H-1** *and* **H-2** *hold, then the QMLE estimator* $\hat{\theta}_T$ *is consistent:*

$$\hat{\theta}_T \xrightarrow{P} \theta^0.$$

Some Statistical Inference (3/3)

Theorem (Asymptotic Normality)

Under assumption **H-1**, **H-2**, **H-3** *and* **H-4**,

$$\sqrt{T}\left(\hat{\theta}_T - \theta^0\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, A_0^{-1} B_0 A_0^{-1}\right),$$

with A_0 and B_0 are some given matrices (see the paper).

- The parameter estimates using the QMLE most inevitably be subject to some degree of uncertainty as the amount of data is, generally, limited.
- This allows us to quantify the uncertainty on parameters based on their asymptotic distribution → ORSA (Model Risk)

Empirical Analysis (1/4)

- We focus on the illustrative model (called AR-ARCH three-level memory random field).
- We consider mortality records of E&W and US over the period 1960-2010 and age band 59-89.
- We check the robustness of the parameters estimation using two time-frame 1960-1999 and 1960-2010.

	1960-2010				1960-1999			
	β	α_0	α^+	α^-	β	α_0	α^+	α^-
US	-0.015	4.81E-4	0.221	0.195	-0.024	5.01E-4	0.208	0.172
E&W	0.28	6.79E-4	0.312	0.329	0.28	6.58E-4	0.334	0.341

ICA CIA
BERLIN 2018

Empirical Analysis (2/4)

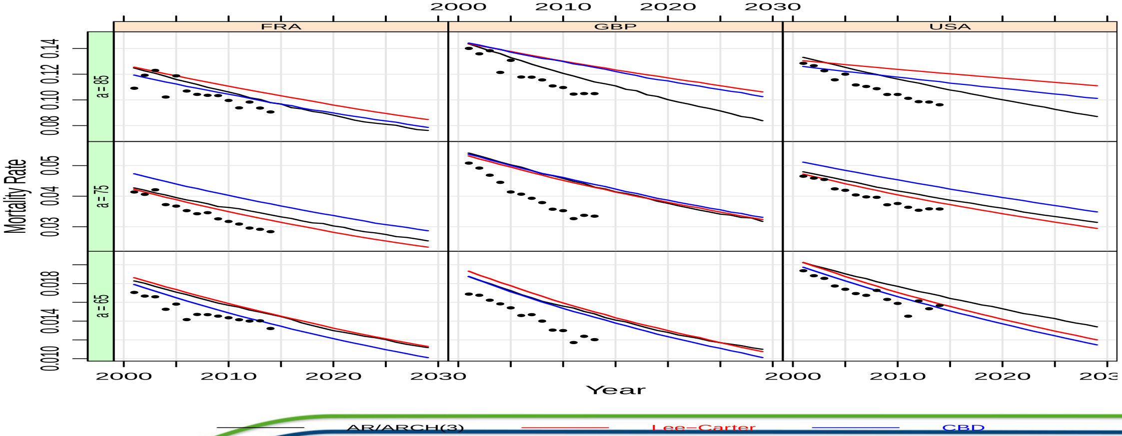
- We compare the model to the LC and CBD-M5 models
  - Goodness-of-fit: in terms of the root of the sum of squared residuals (for an estimation over the period 1960-2010)

	AR-ARCH	LC	CDB
US	4.71E-04	7.57E-04	1.32E-03
E&W	7.08E-04	8.61E-04	1.02E-03

  - Out-of-sample: Parameters estimation over the period 1960-1999 and comparison to the raw mortality over the period 2000-2010

Empirical Analysis (3/4)

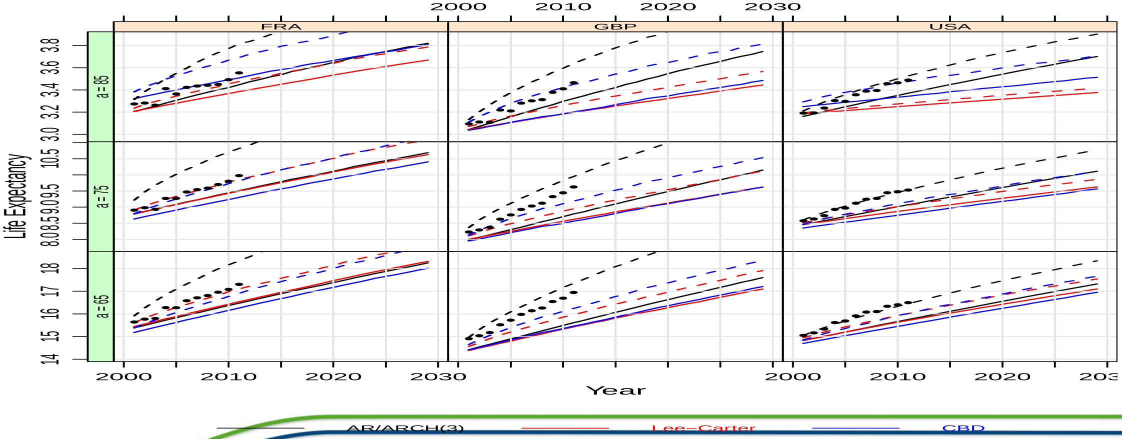
Mortality Rates



ICA CIA BERLIN 2018

Empirical Analysis (4/4)

Life Expectancy



Conclusion (1/2)

- In this paper we proposed a class of random field models with a given causal structure
- This class of model is a generalization of the AR-ARCH univariate process, capturing the cohort effect, the dependence between adjacent cohorts as well as the conditional heteroskedasticity.
- For such a class of models, we propose an estimation procedure for the parameters and exhibit their statistical inferences
- The illustrative model outperforms the Lee-Carter model, especially, for high ages but using few paramaters

Conclusion (2/2)

Next...

- A selection procedure is introduced in

  Doukhan, P., Rynkiewicz, J., & Salhi, Y. (2018). **Optimal neighborhoods selection for AR-ARCH random fields with application to mortality**. Working paper

- More sophisticated and fast selection procedures using machine learning (Bayesian)

- Develop the diagnostic checks (hypotheses testing, model validation... etc.)

- Take into account the non-stationarity of the improvement rates (locally stationary random fields)

Thank you very much for your attention!

Contact details:

Yahia Salhi

address:	50 avenue Tony Garnier 69007 Lyon France
phone:	+33 (0)665/2838 13
mail:	yahia.salhi@univ-lyon1.fr
web:	http://salhi.yahia.free.fr/

References

R.C. Willets. The cohort effect: insights and explanations. *British Actuarial Journal*, 10(04):833–877, 2004.

S. Loisel and D. Serant. In the core of longevity risk: hidden dependence in stochastic mortality models and cut-offs in prices of longevity swaps. 2007.

P. Jevtić, E. Luciano, and E. Vigna. Mortality surface by means of continuous time cohort models. *Insurance: Mathematics and Economics*, 53(1):122–133, 2013.

G. Mavros, A.J.G. Cairns, T. Kleinow, and G. Streftaris. Stochastic mortality modelling: Key drivers and dependent residuals. 2016.

R. Giacometti, M. Bertocchi, S.T. Rachev, and F.J. Fabozzi. A comparison of the lee–carter model and ar–arch model for forecasting mortality rates. *Insurance: Mathematics and Economics*, 50(1):85–93, 2012.

C.M.H. Chai, T.K. Siu, and X. Zhou. A double-exponential garch model for stochastic mortality. *European Actuarial Journal*, 3(2):385–406, 2013.

H. Chen, R. MacMinn, and T. Sun. Multi-population mortality models: A factor copula approach. *Insurance: Mathematics and Economics*, 63:135–146, 2015.

P. Doukhan and L. Truquet. A fixed point approach to model random fields. *ALEA: Latin American Journal of Probability and Mathematical Statistics*, 3:111–132, 2007.

Model Selection (1/2)

ICA CIA BERLIN 2018

We consider the penalized QMLE:

$$U_T(V) = T \times \sup_{\theta \in \Theta_V} L_T(x_s, s \in \mathcal{V}; \theta) - a_T(|V|) \tag{3}$$

where

- T is the number of observations in $\mathcal{O}$
- $|V|$ be the number of elements of the neighborhood associated to non-zero parameter α_v or β_v
- $a_T(|V|)$ is the penalty function of $|V|$ (model complexity), e.g. for the Bayesian Information Criterion (BIC) we have $a_T(|V|) = |V| \times \log(T)$

Model Selection (2/2)

The **best model** in the sense that the model is not only correct but also most economical among all the *correct* models

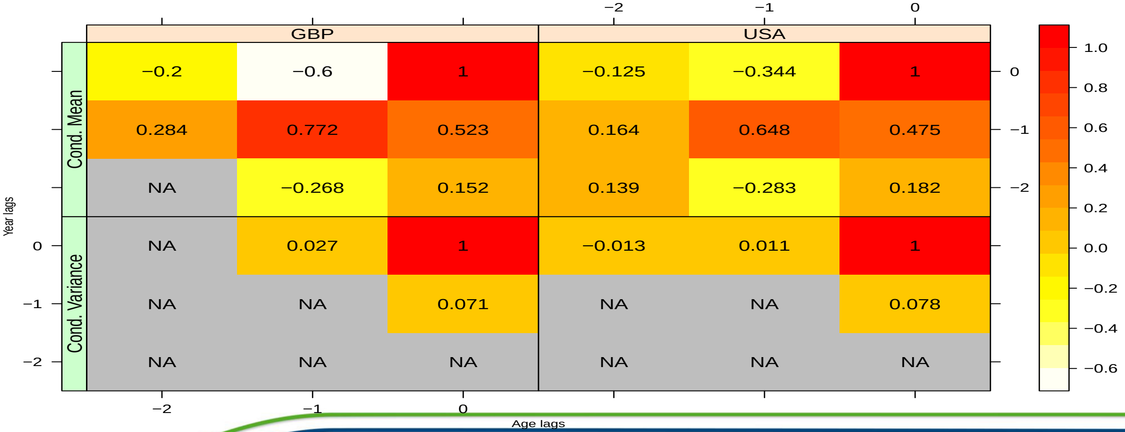
$$\hat{V} = \arg\min_{V \subset \mathcal{V}_1 \cup \mathcal{V}_2} U_T(V).$$

Theorem

Under some assumptions (see the paper) $\hat{V}$ converges in probability to the true neighborhood V^0.

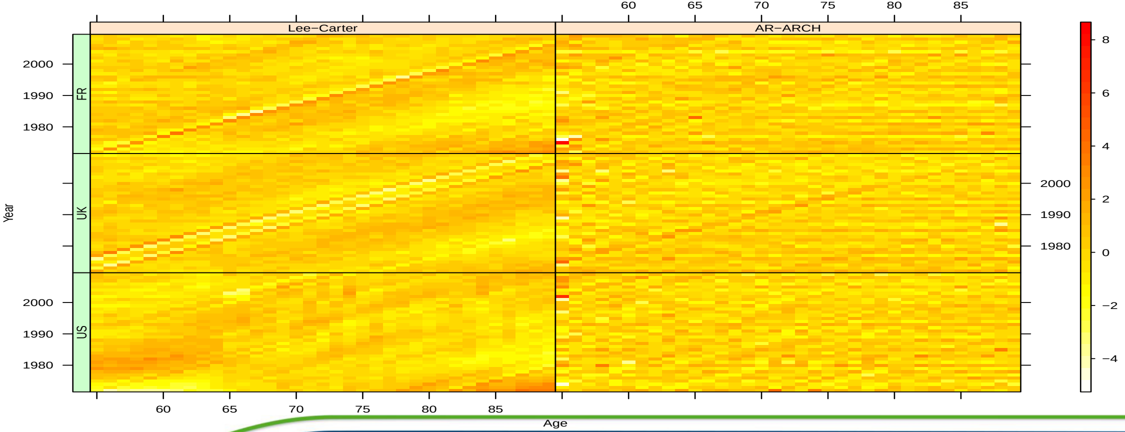
Application to Datasets (1/6)

Neighborhoods V_1 (top) and V_2 (bottom)



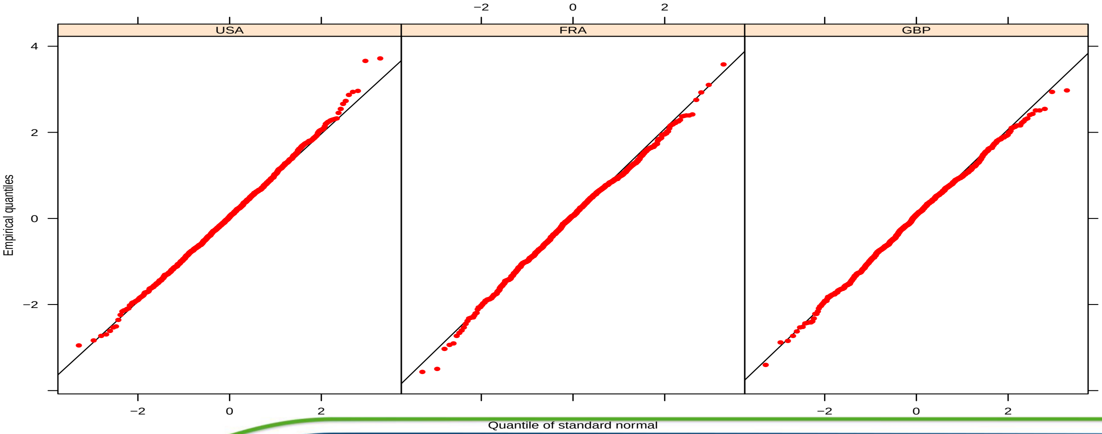
Application to Datasets (2/6)

Residuals



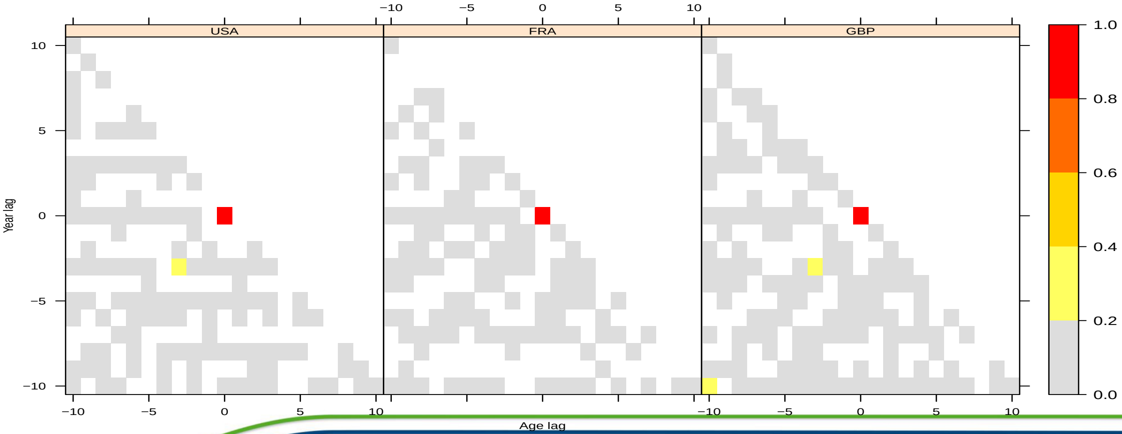
Application to Datasets (3/6)

Residuals



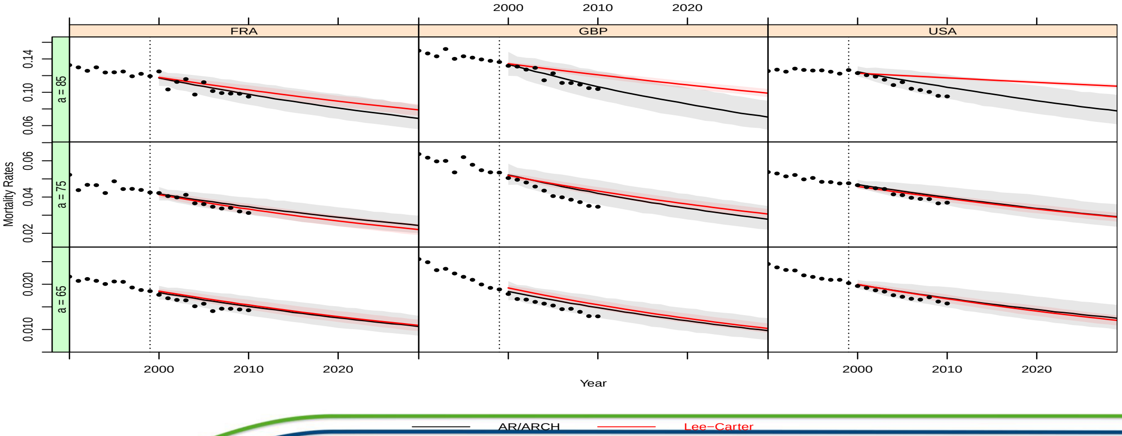
Application to Datasets (4/6)

Residuals' Spatial Autocorrelation



Application to Datasets (5/6)

Mortality Rates



Application to Datasets (6/6)

Life Expectancy