HMD Users Conference #2 – Paris, 8 mars, 2019

A Class of Random Field Memory Models for Mortality Forecasting

Yahia Salhi

yahia.salhi@gmail.com | salhi.yahia.free.fr

ISFA, Université Claude Bernard Lyon 1

&

Chair of excellence "DAMI"

Joint work with P. Doukhan, J. Rynkiewicz, D. Pommeret

LoLitA

ISFA

CHAIR OF EXCELLENCE

Data Analytics & Models for Insurance

This presentation is based on these papers

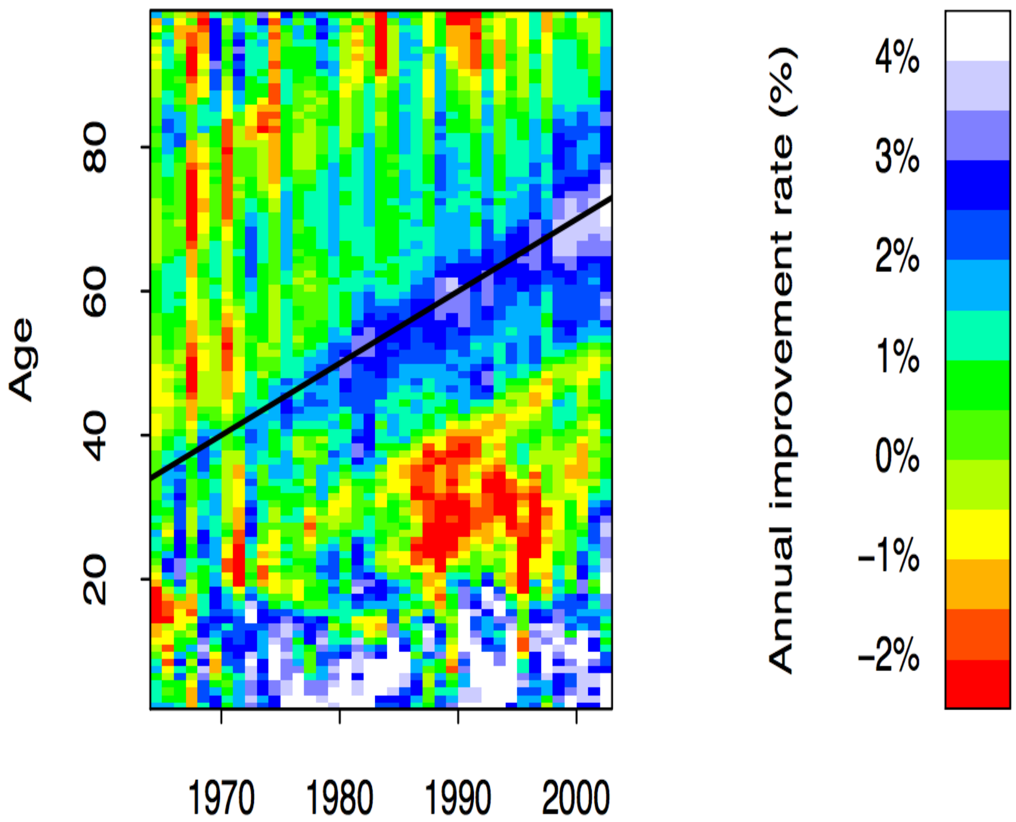
Doukhan, P., Pommeret, D., Rynkiewicz, J., & Salhi, Y. (2017). **A class of random field memory models for mortality forecasting**. *Insurance: Mathematics and Economics*, 77, 97-110.

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

SUMMARY

1. Mortality Stylized Facts
2. Random Fields Models
3. Illustrative Model
4. Some Statistical Inference
5. Model Selection

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)

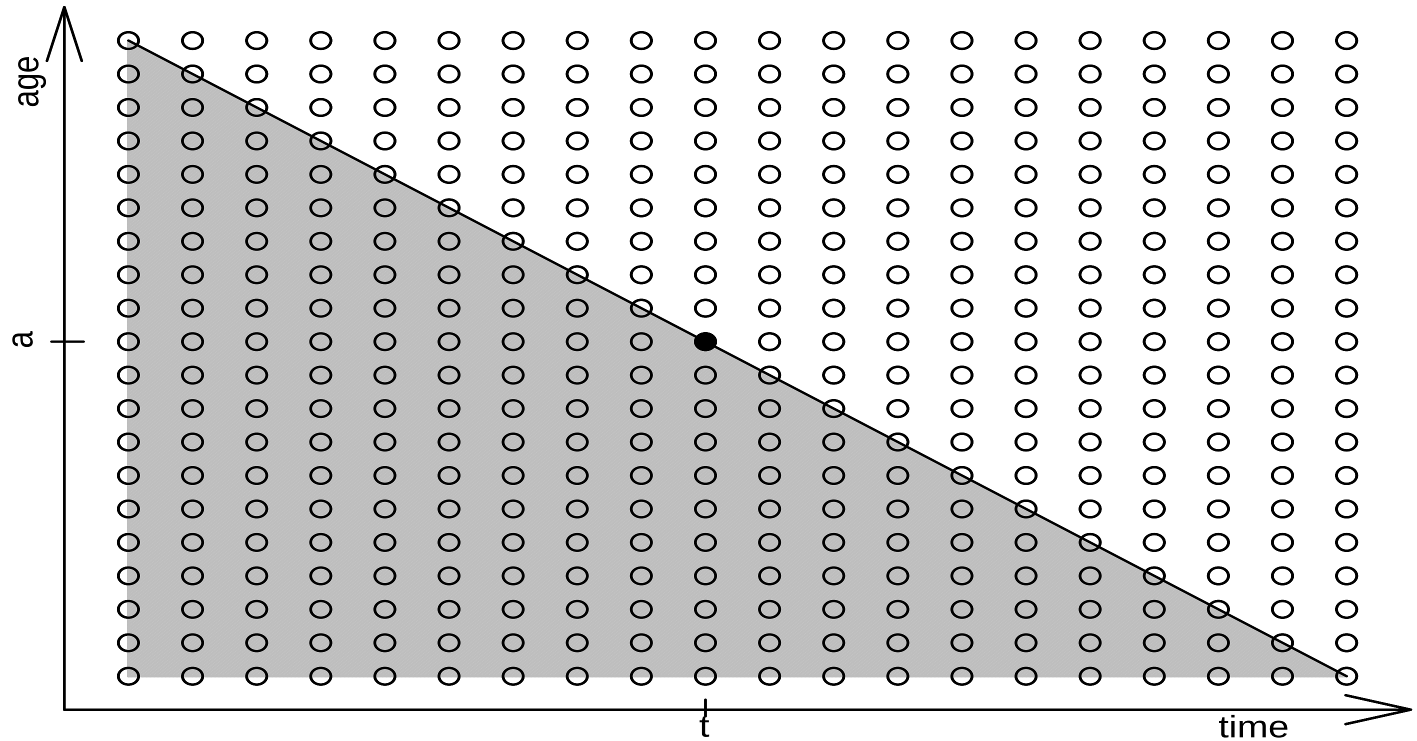


Figure 1: A bi-dimensional representation of the random field $X_s = \mathrm{IR}_s - \overline{\mathrm{IR}}$, with $s = (a, t)$.

Mortality Stylized Facts (4/4)

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

under the 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 $\sum_{v\in V} \alpha_v < 1$.

RANDOM FIELDS MODELS (1/3)

- Based on this abstract formulation, we aim, henceforth, at proposing 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 :

$$m_s = \mathbb{E}\Big[X_s \Big| \{X_{u,v};\, u < a, v \leq t\}\Big] = \sum_{v \in V_1} \beta_v X_{s-v}$$

$$\sigma_s^2 = \mathrm{Var}\Big(X_s \Big| \{X_{u,v};\, u < a, v \leq t\}\Big) = \alpha_0 + \sum_{v \in V_2} \alpha_v X_{s-v}^2,$$

  - ➤ V_1 and V_2 are two neighborhoods characterizing the evolution of conditional mean and variance of each X_s in terms of its own past values and the present and past values of the adjacent cohorts

Random Fields Models (2/3)

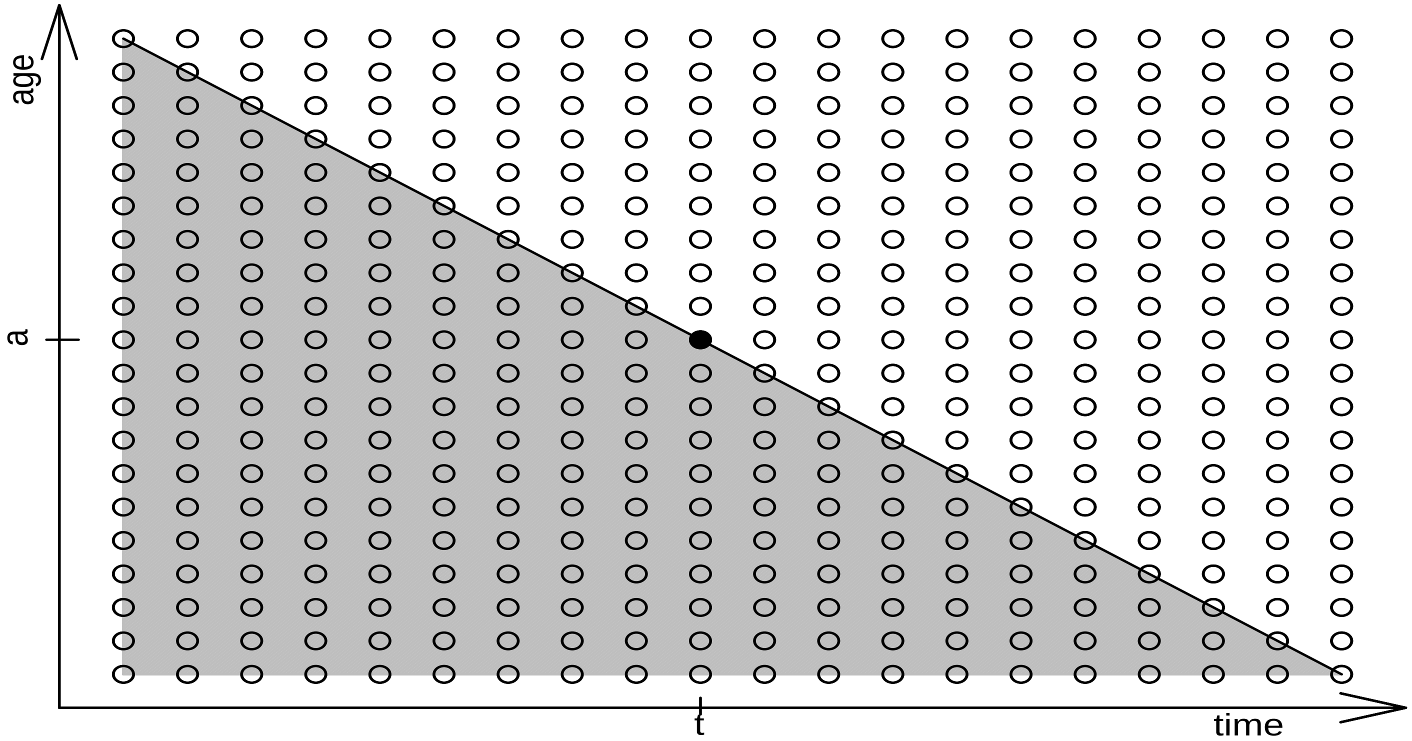


Figure 2: A bi-dimensional representation of the random field $X_s = \mathrm{IR}_s - \overline{\mathrm{IR}}$, with $s = (a, t)$. The grayed area repsent the causal neighbor V_1 and V_2 needed to characterize the evolution of X_s.

Random Fields Models (3/3)

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

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

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

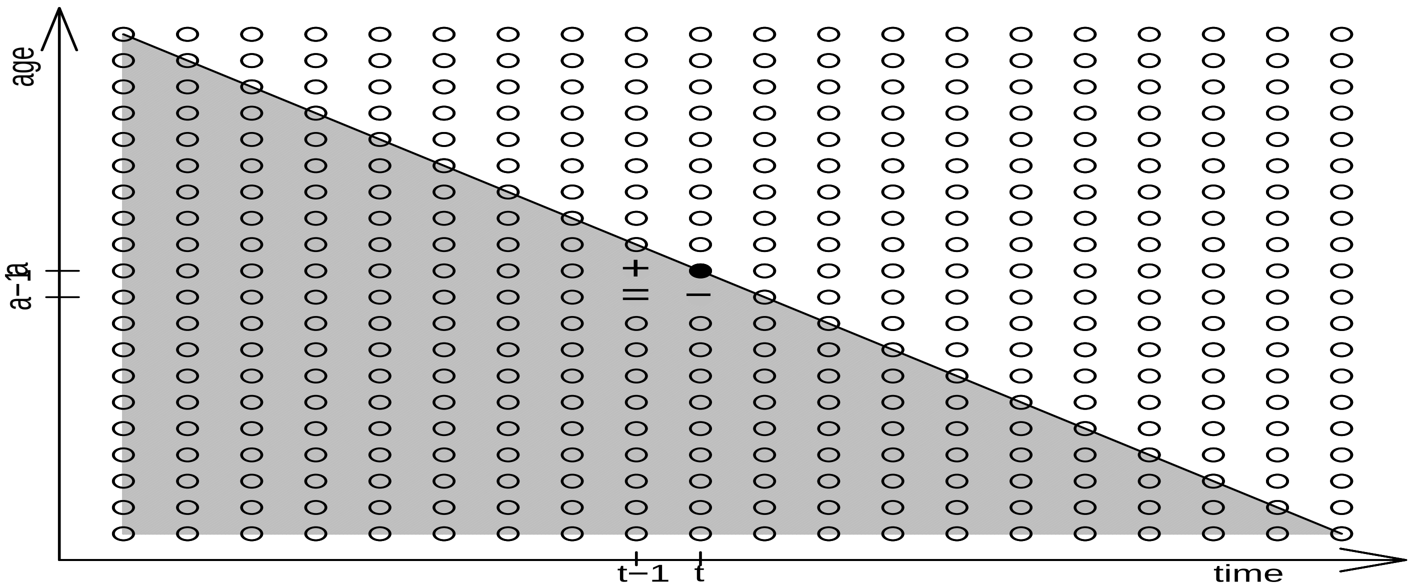
$$\|\xi_0\|_p < \infty, \quad \kappa_p \equiv \|\xi_0\|_p \sum_{s \in V_2} \alpha_s + \sum_{s \in V_1} \beta_s < 1, \quad p \geq 1.$$

Illustrative Example (1/2)

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

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

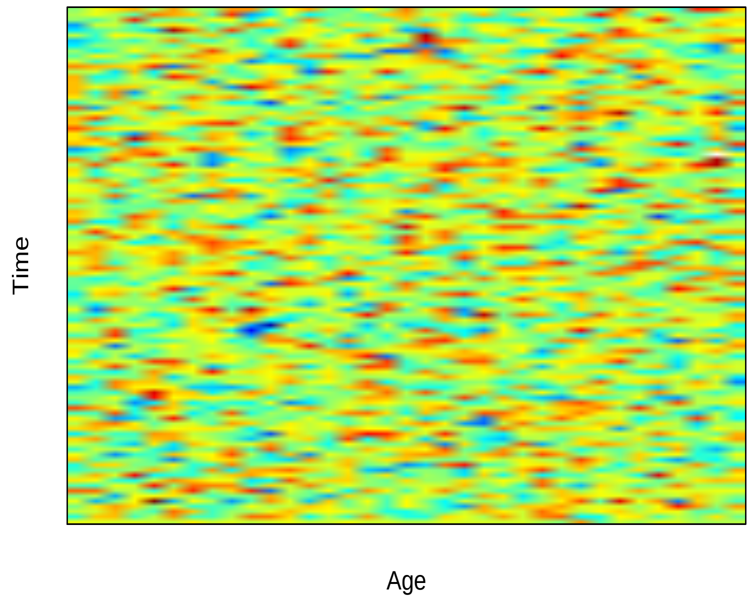
where $\underbrace{s^- = (a-1, t)}_{\text{young cohort}}, \quad \underbrace{s^+ = (a, t-1)}_{\text{old cohort}}, \quad \underbrace{s^= = (a-1, t-1)}_{\text{auto-regressive}}.$



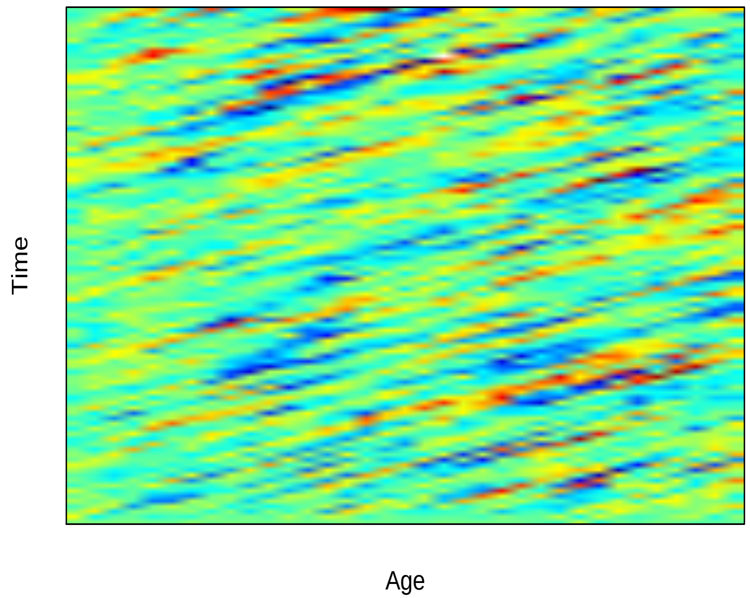
Illustrative Example (2/2)

- 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 (M) can be simply rewritten as

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



(a) $\beta \downarrow$



(b) $\beta \uparrow$

Some Statistical Inference (1/3)

- 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.

$$\widehat{\theta}_T = \arg\max_{\theta\in\Theta} L_T\left(x_s, s \in \mathcal{O}; \theta\right), \qquad (1)$$

where Θ is the set of possible parameters.

Some Statistical Inference (2/3)

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 (M) belongs to the interior of Θ.

Theorem 1 (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 2 (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.
- This is of paramount importance as the amount of data at our disposal is, generally, limited.
- This allows us to quantify the uncertainty on parameters based on their asymptotic distribution $\rightarrow$ 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

Empirical Analysis (2/4)

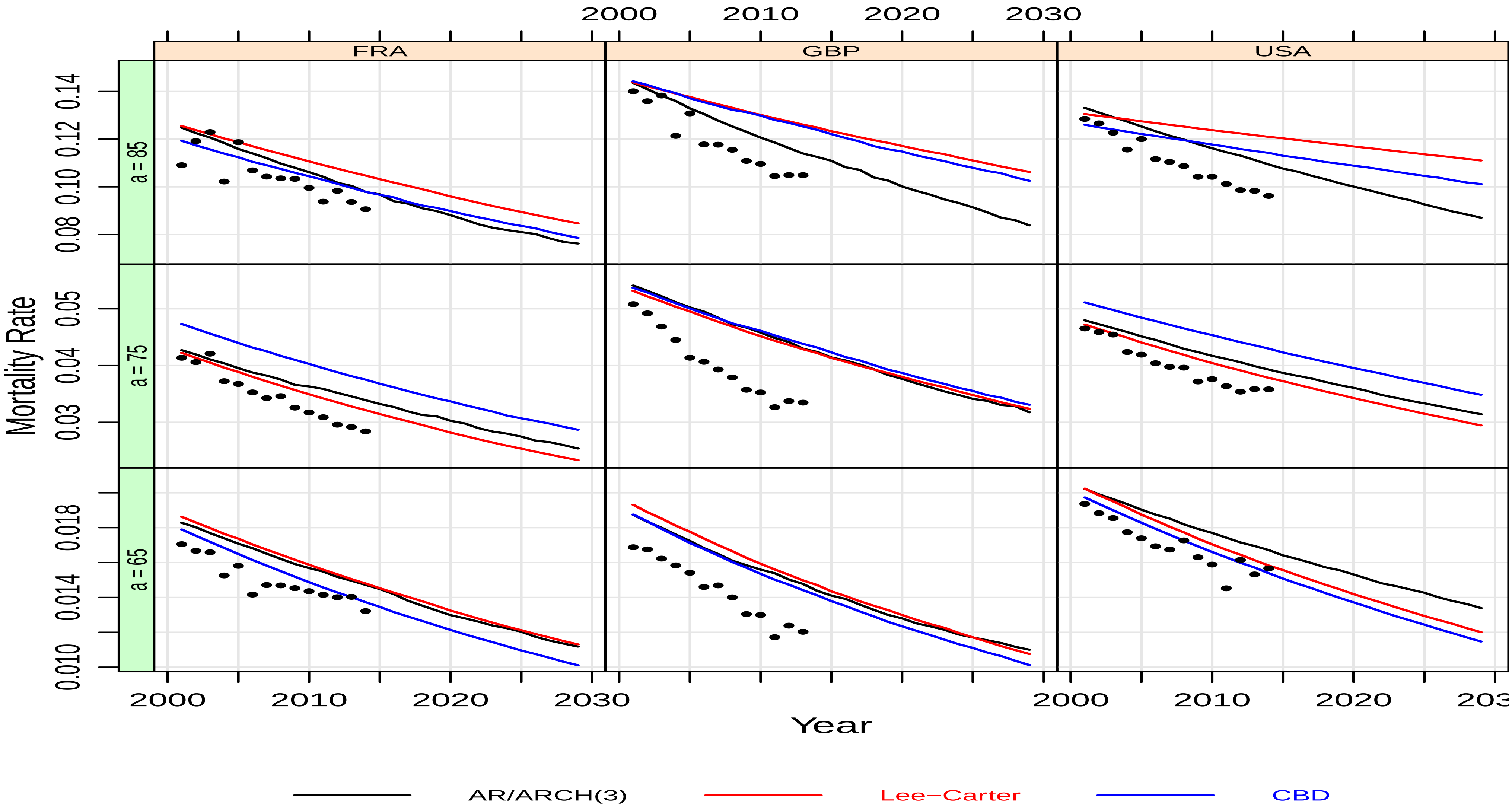
- We compare the model to the LC and CBD-M5 models
  - Goodness-of-fit: in terms of the root of the sum of squared residuals (for an estimation over the period 1960-2010)

	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

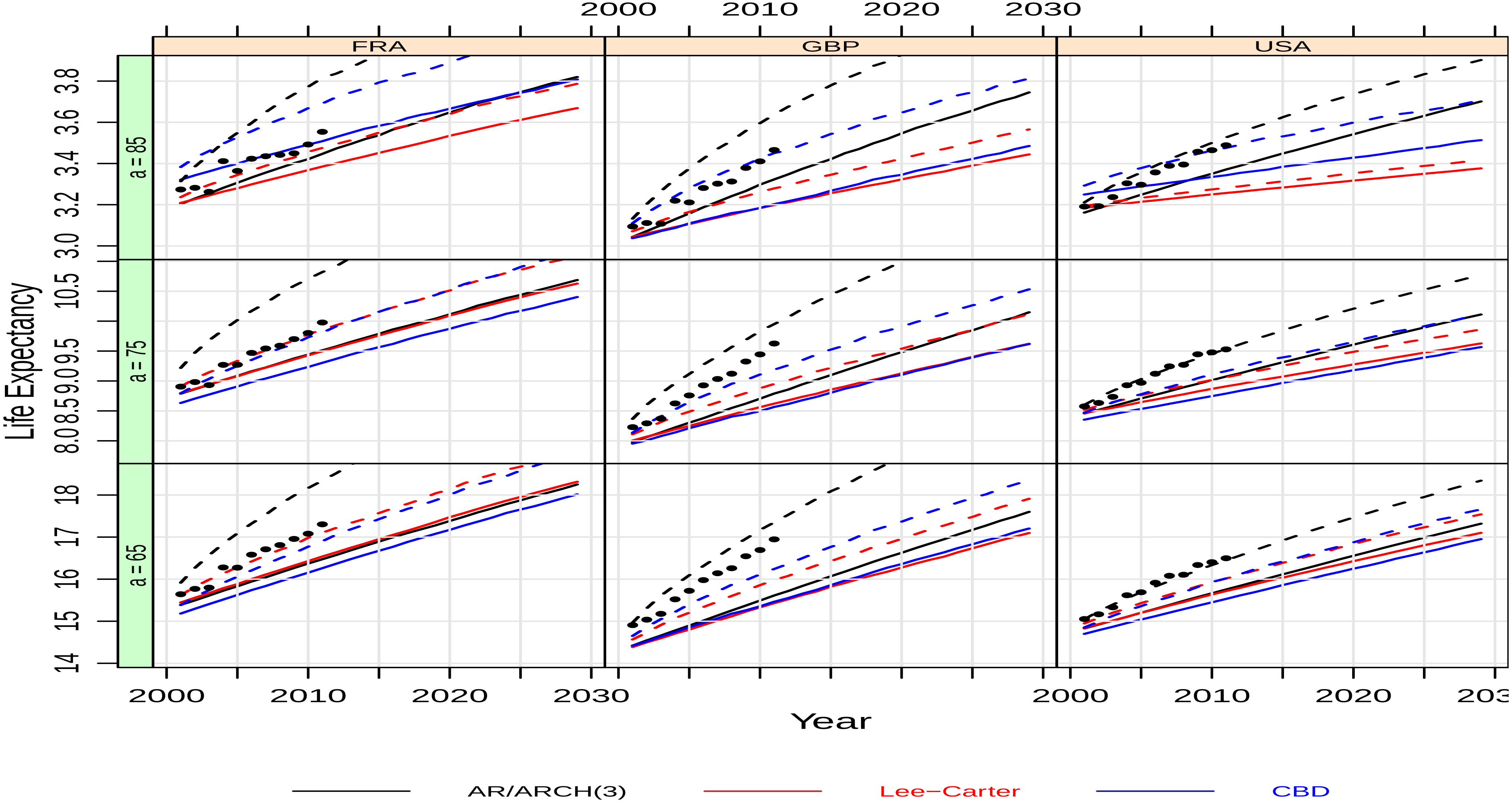
Empirical Analysis (3/4)

Mortality Rates



Empirical Analysis (4/4)

Life Expectancy



Model Selection (1/3)

We consider the penalized QMLE:

$$U_T(V) = T \times \sup_{\theta \in \Theta_V} L_T\left(x_s, s \in \mathcal{V}; \theta\right) - a_T(|V|) \tag{2}$$

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 (2/3)

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

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

Theorem 3. *Under some assumptions (see the paper) $\hat{V}$ converges in probability to the true neighborhood V^0.*

Model Selection (3/3)

The selection procedure requires:

1. The determination of *maximal* neighborhoods for the AR and ARCH components
2. Scanning (optimization of the QMLE) all the possible combinations
3. Ordering the models using the BIC

Application to Datasets (1/4)

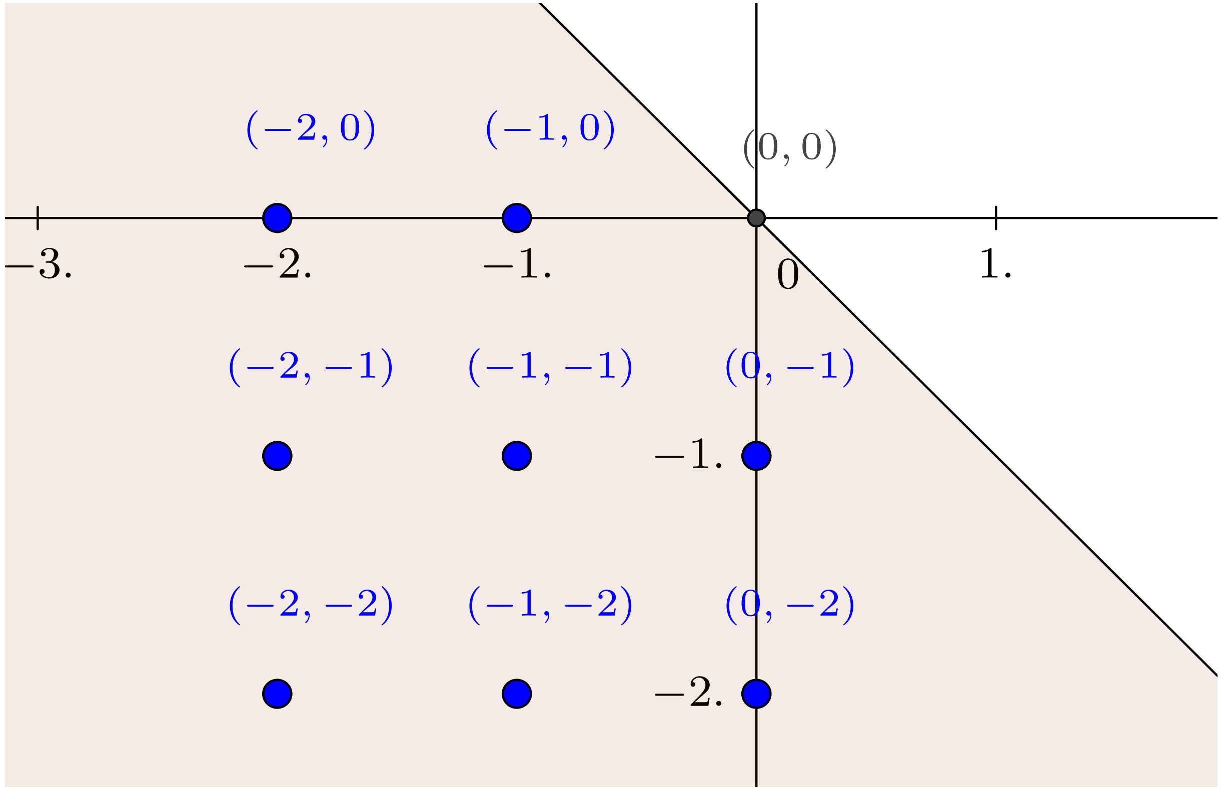


Figure 3: The initial (maximal) candidate neighborhoods $\mathcal{V}_1 \equiv \mathcal{V}_2$.

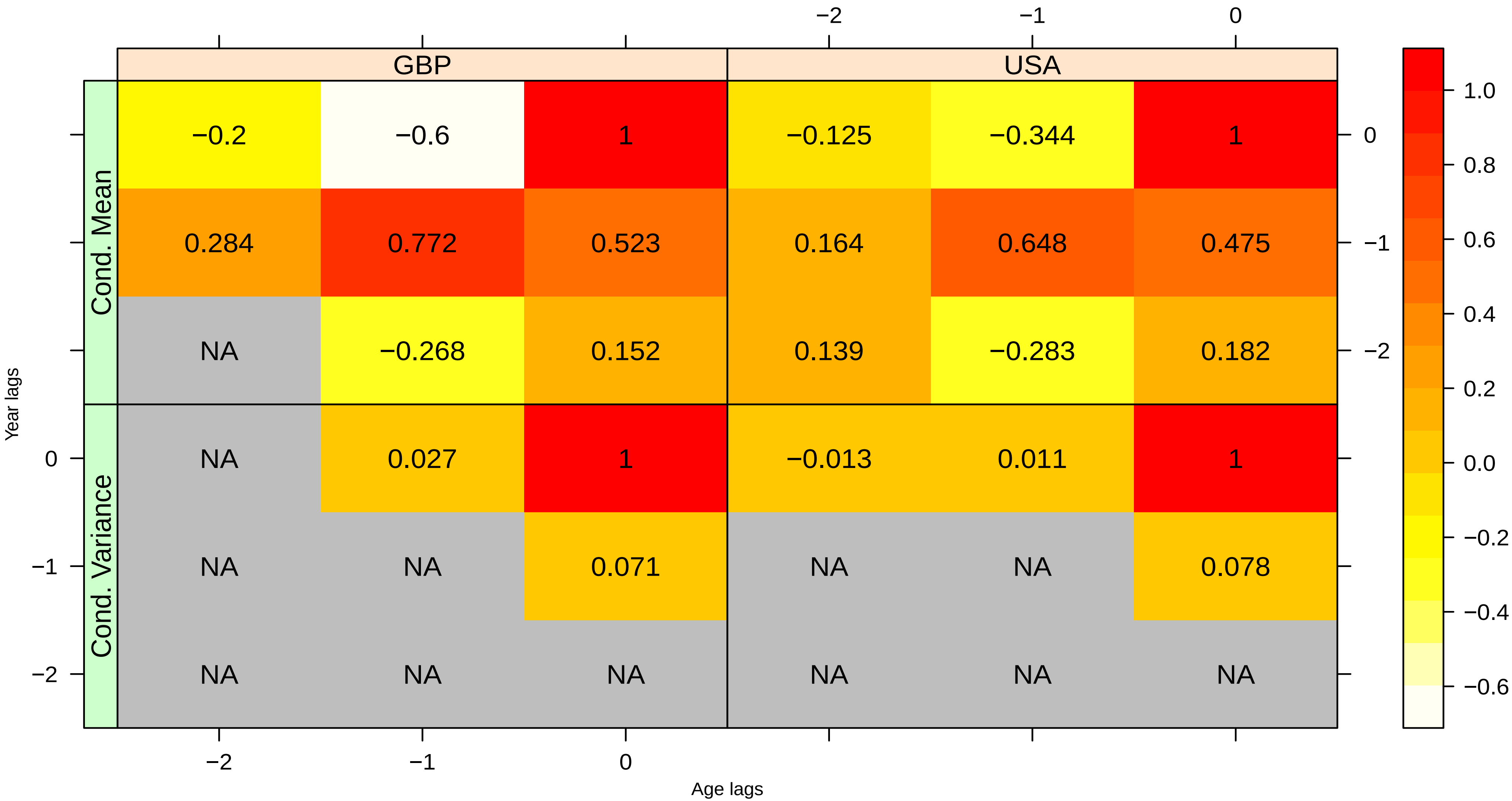
This gives rise to $2^{16} = 65536$ potential models (combinations)!

Application to Datasets (1/4)

- Data of Human Mortality Database, ages 59 – 89 and years 1960 – 2010 then 1960 – 1999 for model back-testing
- Codes in R with C++ component using parallel computation $\sim$ 8mn per population

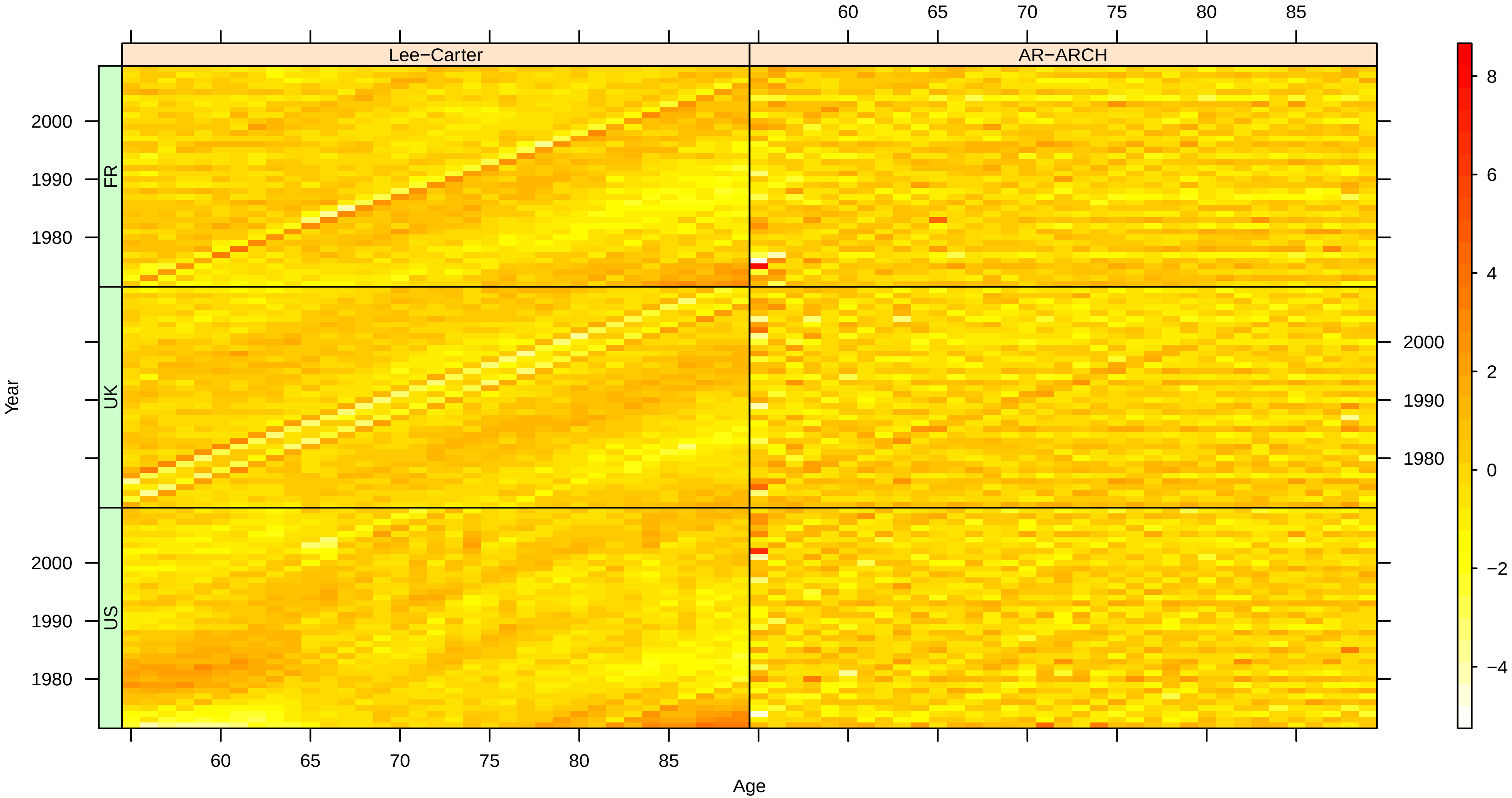
Application to Datasets (2/4)

Neighborhoods V_1 (top) and V_2 (bottom)



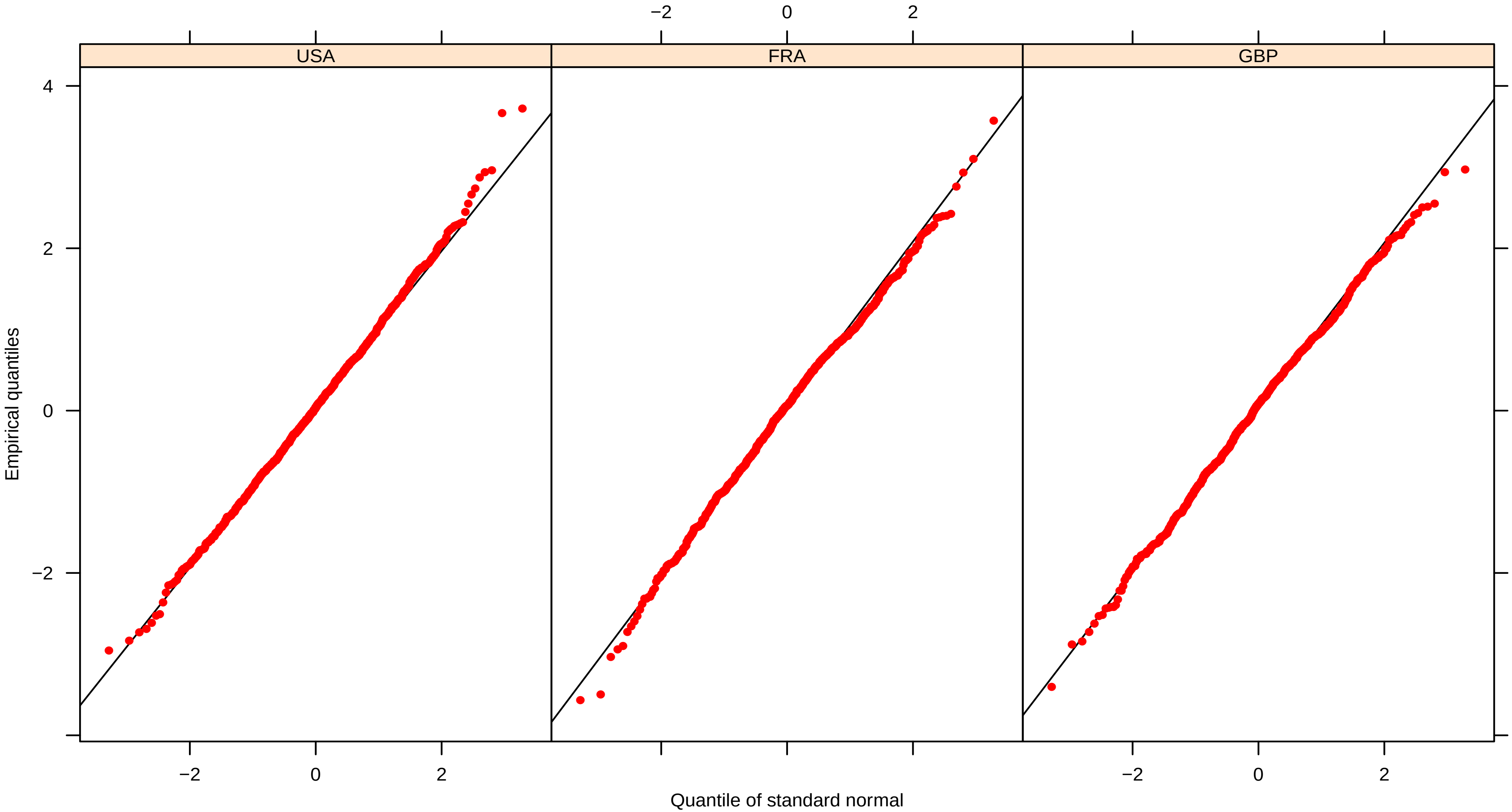
Application to Datasets (3/4)

Residuals



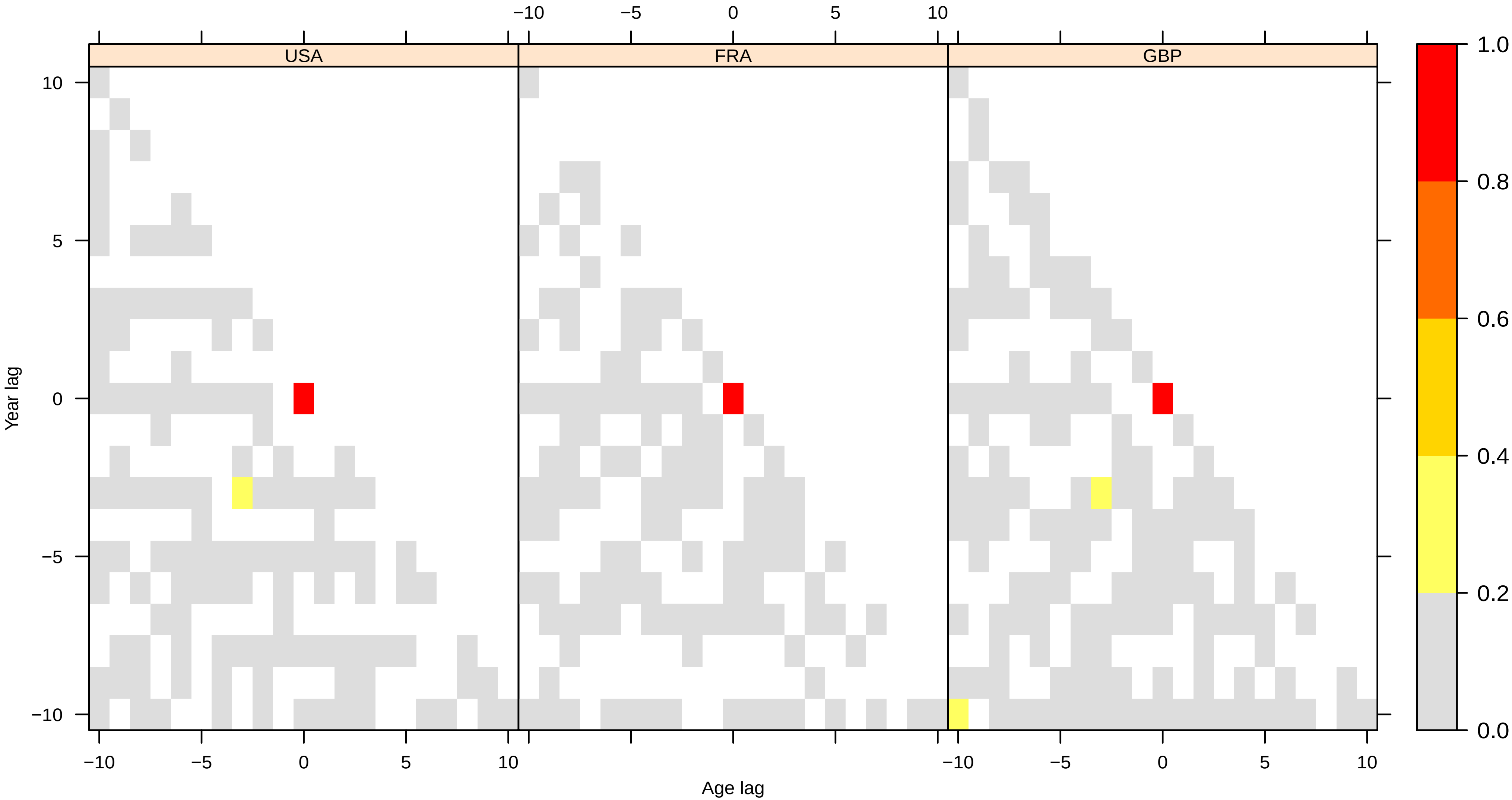
Application to Datasets (3/4)

Residuals



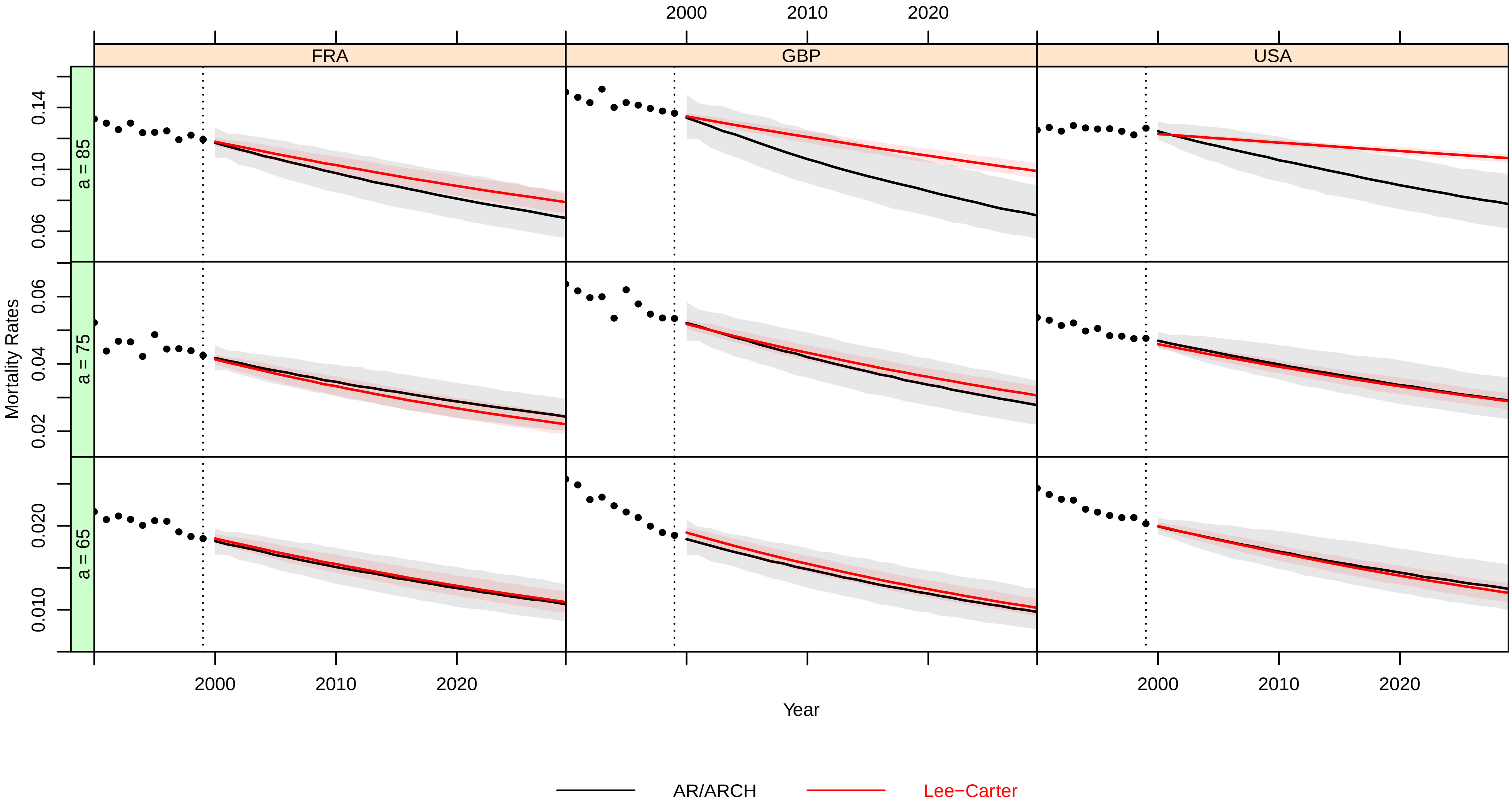
Application to Datasets (3/4)

Residuals' Spatial Autocorrelation



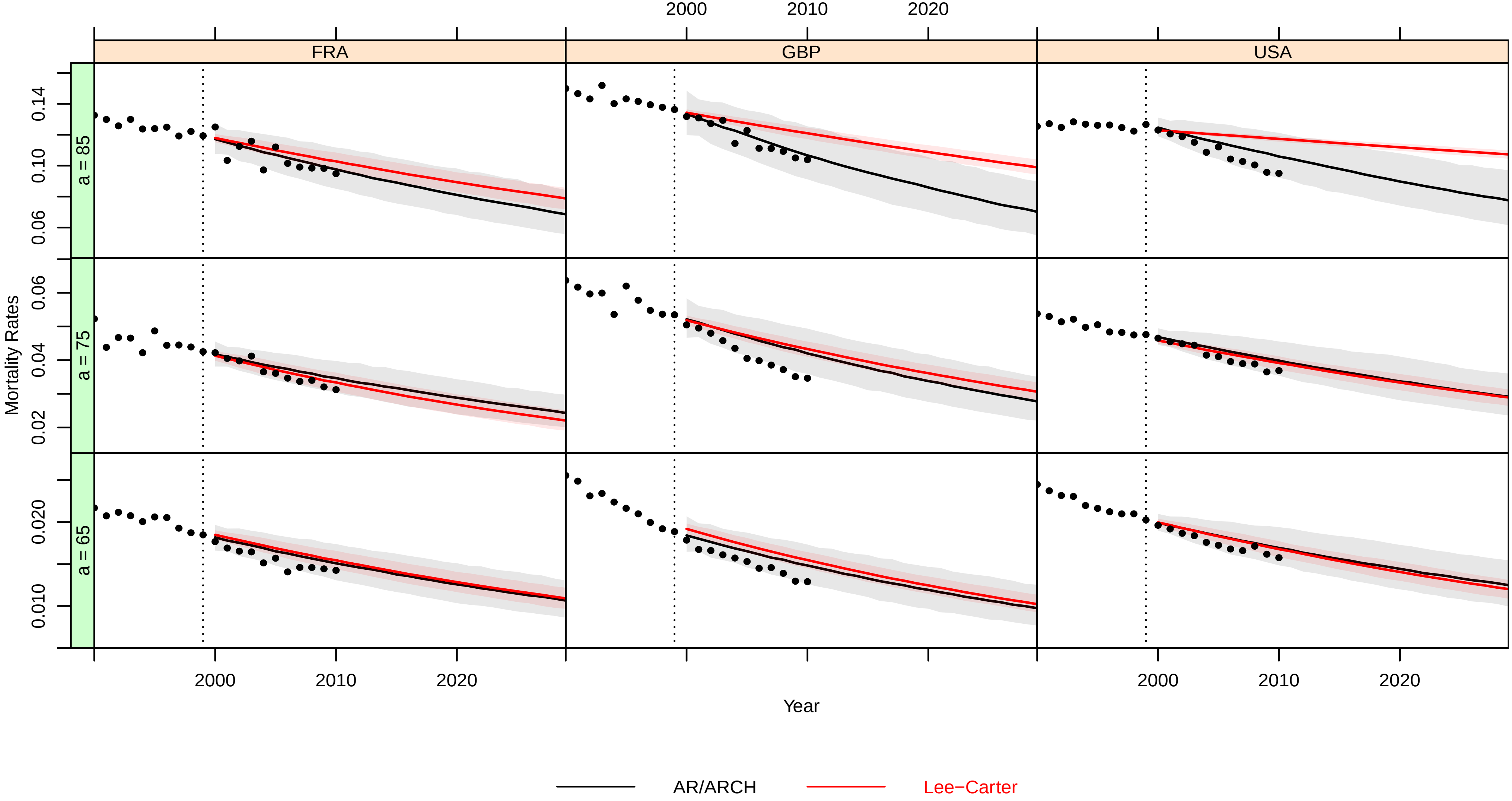
Application to Datasets (4/4)

Mortality Rates



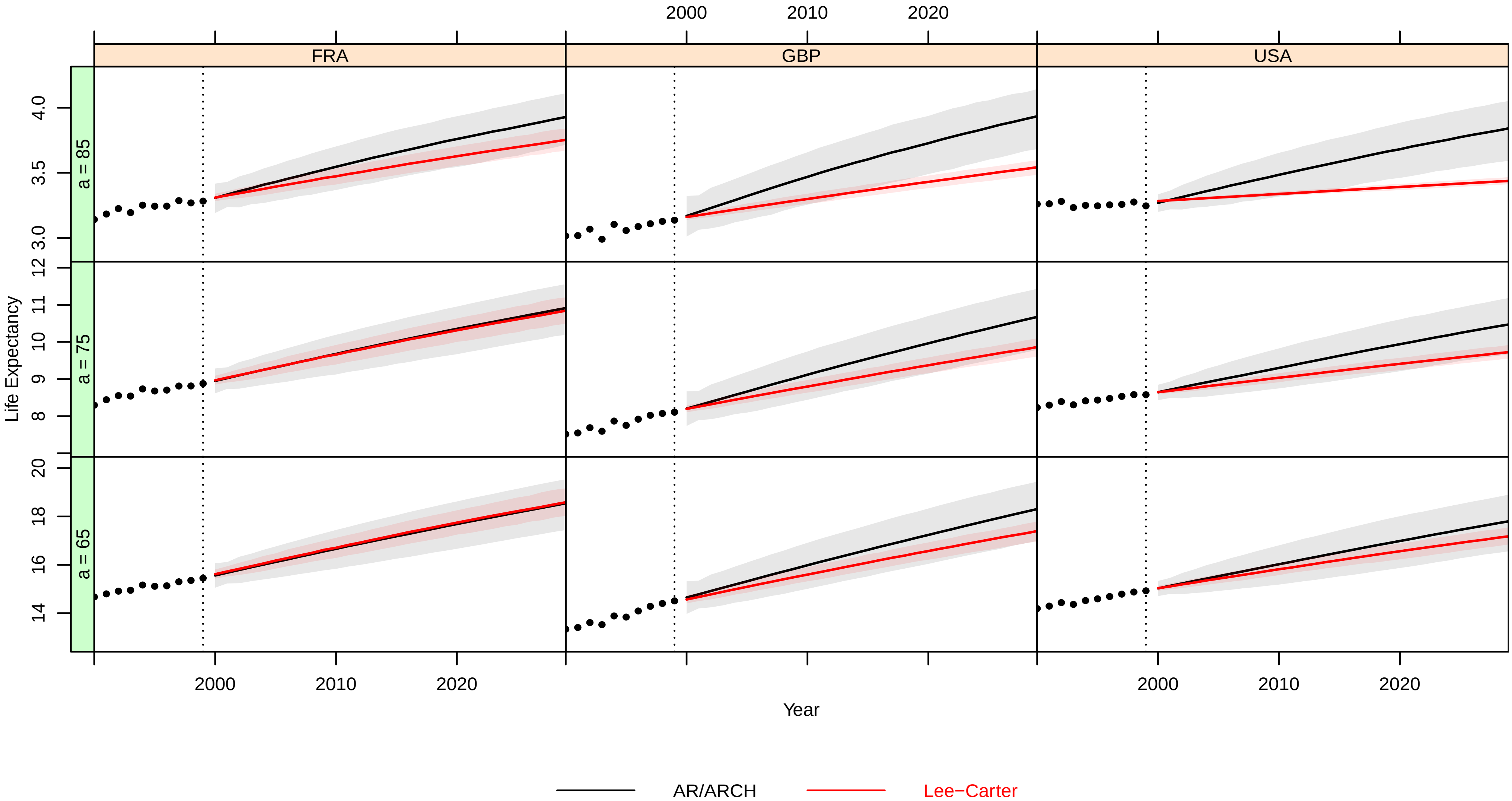
Application to Datasets (4/4)

Mortality Rates



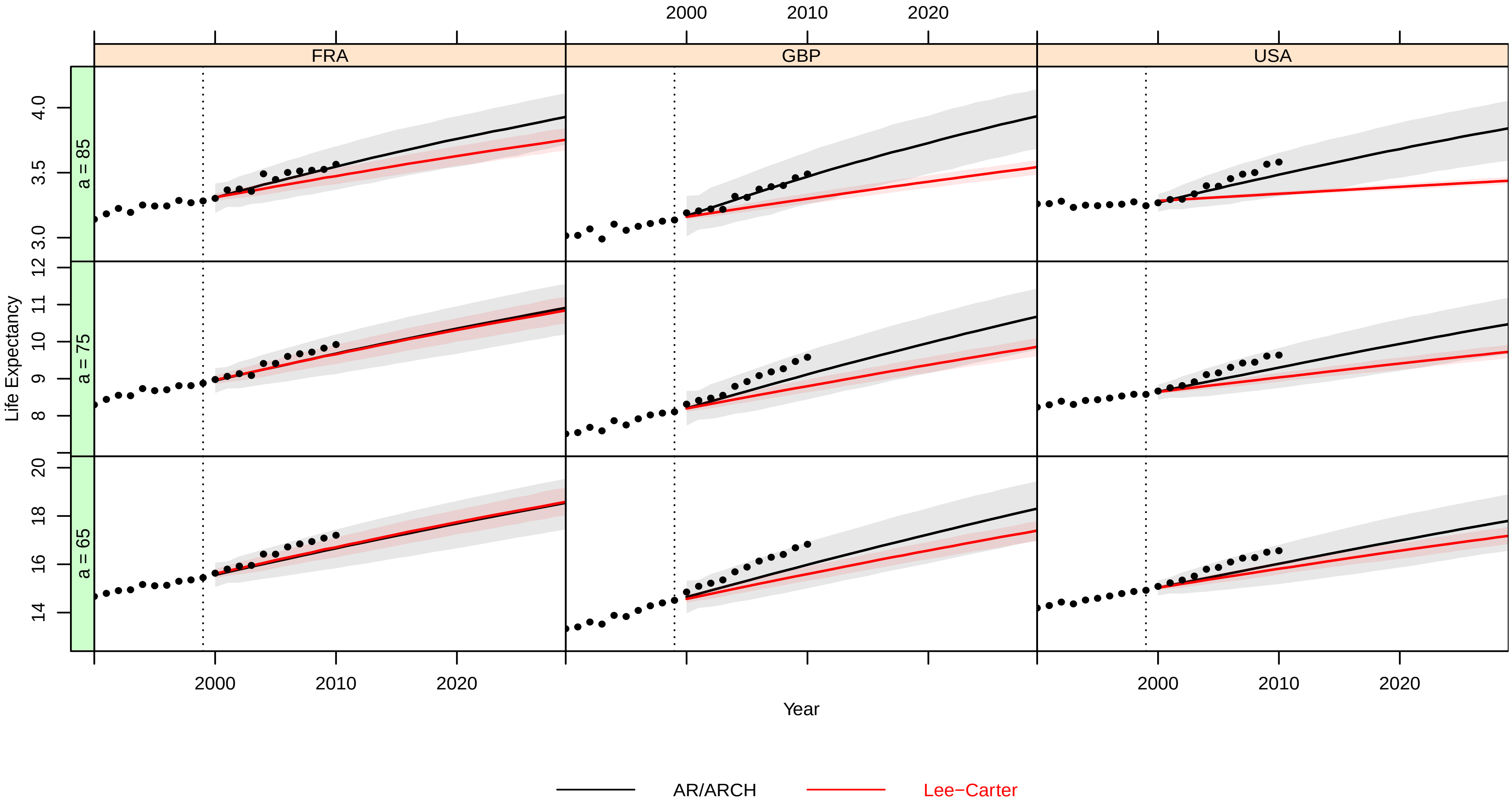
Application to Datasets (4/4)

Life Expectancy



Application to Datasets (4/4)

Life Expectancy



Conclusion

- 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 dependence between adjacent cohorts as well as the conditional heteroskedasticity.
- For such a class of models, we propose an estimation procedure for the parameters as well as a selection procedure
- The model outperforms the Lee-Carter model, especially, for high ages

Future work:

- In pursuit of swiftness, accelerating the optimal selection procedure using Bayesian *learning* algorithms
- Developing a selection procedure for the *maximal neighborhood* as for the time series (Box-Jenkins)
- Develop the diagnostic checks (hypotheses testing, model validation... etc.)
- Take into account the non-stationarity of the improvement rates (locally stationary random fields)

... but also,

- fit the model using the new updated HMD data

THANK YOU FOR YOUR ATTENTION

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.

Questions ?