

A Class of Random Field Memory Models for Mortality Forecasting*

P. Doukhan⁽¹⁾ D. Pommeret⁽²⁾ J. Rynkiewicz⁽³⁾ Y. Salhi⁽⁴⁾

⁽¹⁾Université de Cergy-Pontoise, UMR 8088 Analyse, Géométrie et Modélisation,
95302 Cergy-Pontoise cedex, France.

⁽²⁾Aix Marseille Univ, CNRS, Centrale Marseille, I2M,
13288 Marseille cedex 9, France

⁽³⁾Équipe SAMM, EA 4543 Université Paris I Panthéon-Sorbonne
90, rue de Tolbiac 75634 Paris cedex 13 France

⁽⁴⁾Univ Lyon, Université Lyon 1, ISFA, LSAF EA2429,
50 avenue Tony Garnier, F-69007 Lyon, France

March 19, 2017

Abstract

This article proposes a parsimonious alternative approach for modeling the stochastic dynamics of mortality rates. Instead of the commonly used factor-based decomposition framework, we consider modeling mortality improvements using a random field specification with a given causal structure. Such a class of models introduces dependencies among adjacent cohorts aiming at capturing, among others, the cohort effects and cross generations correlations. It also describes the conditional heteroskedasticity of mortality. The proposed model is a generalization of the now widely used AR-ARCH models for random processes. For such a class of models, we propose an estimation procedure for the parameters. Formally, we use the quasi-maximum likelihood estimator (QMLE) and show its statistical consistency and the asymptotic normality of the estimated parameters. The framework being general, we investigate and illustrate a simple variant, called the three-level memory model, in order to fully understand and assess the effectiveness of the approach for modeling mortality dynamics.

Keywords Mortality Rates, AR-ARCH Random Field, Estimation, QMLE, Inference

Mathematics Subject Classification (2000) Primary: 60G70 Secondary: 60G10, 60F99

*The work is supported by the ANR Research Project LoLitA (ANR-13-BS01-0011). P. Doukhan's work has been developed within the MME-DII center of excellence (ANR-11-LABEX-0023-01). Y. Salhi's work is supported by the BNP Paribas Cardif Chair "Data Analytics & Models for Insurance". The views expressed herein are the author's owns and do not reflect those endorsed by BNP Paribas. Contacts: P. Doukhan (paul.doukhan@u-cergy.fr), D. Pommeret (denys.pommeret@univ-amu.fr), J. Rynkiewicz (joseph.rynkiewicz@univ-paris1.fr) and Y. Salhi (yahia.salhi@univ-lyon1.fr – corresponding author). We thank the editor and the anonymous referees for their careful reading, and their useful comments and suggestions which made our earlier presentation better. We would also like to thank Quentin Guibert for his useful comments.

1 Introduction

The forecast of future mortality improvements pose a challenge not only for public retirement systems planning but also for the private life annuities business, due to the continuous longevity improvement. For public policy, as well as for the management of financial institutions, it is important to forecast future mortality rates in order to quantify the risk underlying their pension and annuities portfolios. To this end, a variety of models have been introduced in the literature during the last decades.

Most notably, there are the so-called factor-based models widely in use by practitioners, which know an increasing recognition from the actuarial community. These *traditional* mortality models rely on a factor-based decomposition of mortality surface. These factors are intended to capture the complex patterns of mortality evolution over time. Although these models are quite intuitive, their statistical properties are, however, not accurately understood. For instance, in their seminal and influential work, [Lee and Carter \(1992\)](#) have proposed a model that decomposes mortality surface into a latent trend, and two corresponding age-sensitive parameters, see also [Brouhns et al. \(2002\)](#). The other models that followed extend the idea underlying the [Lee and Carter \(1992\)](#)'s model by adding a mixture of additional components which capture age, period and, in some cases, cohort effects. As noted by [Mavros et al. \(2017\)](#), “*the number and form of these types of effects is usually what distinguishes one model from another*”. However, some recent works show their limits, e.g. [Giacometti et al. \(2012\)](#), [Chai et al. \(2013\)](#), [Hunt and Villegas \(2015\)](#) and [Mavros et al. \(2017\)](#) among others. In particular, one of the main drawbacks of these classical models relates to the assumption of the homoskedasticity of their residuals. In fact, the assumption of constant variance is always violated in practice as it is time varying, see e.g. [Lee and Miller \(2001\)](#) and [Gao and Hu \(2009\)](#). Furthermore, the mortality evolution is known to be related to the age of birth, see [Willets \(2004\)](#). This is generally referred to as the cohort effect and translates the persistent of some shocks on mortality among cohorts. It is observed when plotting the residuals of some models that rely on age and period factors as an apparent diagonal structure. These observations point to a need for additional univariate cohort-dependent process in some countries. Such a phenomenon has lead to various extensions, in the literature, of the initial Lee-Carter model, e.g. [Renshaw and Haberman \(2006\)](#) or [Cairns et al. \(2009\)](#). The incorporation of the cohort-specific process, for instance, has been suggested to overcome the so-called *non-stationary* effect, which corresponds to the diagonal structure observed in the plotting of the age-period models' residuals. Even if this undesired remaining diagonal effect is, generally, accommodated, it is still unclear how such a cohort-effect can be interpreted and identified, see [Hunt and Villegas \(2015\)](#). This is even more appealing in view of some recent empirical findings. Although, praise the goodness-of-fit performance of age-period-cohort models specification and meanwhile shed light on their instable forecasting performance. Furthermore, these mainstream models are over-parametrized and have tendency to over-fit and thus produce less reliable forecasts.

It is of course very important to tackle these limitations when considering a new modeling approach, but it is also essential to take into account the dependence structure between adjacent cohorts. Indeed, some recent works, and even common intuition, point out the importance of cross-cohorts correlation, see e.g. [Loisel and Serant \(2007\)](#) and [Jevtić et al. \(2013\)](#). In their empirical work, [Loisel and Serant \(2007\)](#) show that correlation among close generations is higher enough to be omitted. The same conclusions were drawn in the very recent work of [Mavros et al. \(2017\)](#).

In this paper, in contrast to this univariate factor-based framework, we approach the problem of

modeling mortality rates by considering the whole surface of mortality improvements as a sole random field without any further assumption on the particular dependence structure neither the factors driving its evolution. Thus, unlike to mainstream approach, our modeling framework is intended to accommodate cross-cohorts dependence as well as conditional heteroskedasticity. The starting point of our approach is a formulation of the mortality random field in the sense of [Doukhan and Truquet \(2007\)](#) with a given causal structure. Such a class of models introduces dependencies among adjacent cohorts aiming at capturing, among others, the cohort effects and cross-generations correlations. It also takes into account the conditional heteroskedasticity of mortality. The proposed model is a generalization of the now widely used AR-ARCH models for random processes. More formally, the conditional mean and variance of mortality rates are respectively described by linear combinations of the observed rates and their squared values on a given neighborhood. In [Section 2](#), we fully describe the model and give some intuitions on its construction. The specification of the causality structure is discussed and some first results on the stability as well as the identification of the model are introduced. For such a class of models, we also propose a robust estimation procedure for the parameters.

The rest of the paper is organized as follows. In [Section 3](#), we use the quasi-maximum likelihood estimator (QMLE) to estimate the parameters. Its statistical consistency and asymptotic normality are shown. The framework being general, we investigate and illustrate a simple variant, called the three-level memory model, in order to fully understand and assess the effectiveness of the approach for modeling mortality dynamics. This three-level memory level incorporates the correlations with the immediate cohorts and it is intended to capture the cohort effect in a natural manner. In [Section 4](#), the model is applied to the populations of US, France and England & Wales, and is compared to the benchmark models of [Lee and Carter \(1992\)](#) and [Cairns et al. \(2006\)](#) two-factor models.

2 Random Fields Memory Models

2.1 From Classical Mortality Models to a Random Field Memory Formulation

Denote by $m_{(a,t)}$ the crude death rate at age a and date t . Time is assumed to be measured in years, so that calendar year t has the meaning of the time interval $[t, t + 1)$. For expository purpose and since we will be working with only a subset of historical data, we will henceforth re-index the observable ages by $a = 0, 1, \dots, I - 1$ and the observable dates by $t = 0, 1, \dots, J - 1$; where I and J are, respectively the number of ages and years. Here, we introduce two benchmark models for mortality dynamics in order to motivate the development of the random field model discussed later on this section. We limit ourselves to these models for simplicity and other modeling frameworks are briefly discussed.

Classical Mortality Models. In their seminal paper, [Lee and Carter \(1992\)](#) postulated that the (log) mortality rates at different ages are captured by a common factor, and an age-specific coefficient with respect to this common trend. More precisely, we have for any a and t

$$\log m_{(a,t)} = \alpha_a + \beta_a \kappa_t + \epsilon_{(a,t)}, \quad \text{with } \epsilon_{(a,t)} \sim \mathcal{N}(0, \sigma) \quad (1)$$

where α_a is the time average level of $\log m_{(a,t)}$ at age a , κ_t is the common factor also known as the period mortality effect and β_a is the age-specific sensitivity coefficient with respect to κ_t . Another interesting

model was suggested by [Cairns et al. \(2006\)](#) and assumes that the one-year death mortality rates dynamics are given by the following modeling form:

$$\text{logit } q_{(a,t)} = \kappa_t^{(1)} + \kappa_t^{(2)}(a - \bar{a}) + \epsilon_{(a,t)}, \text{ with } \epsilon_{(a,t)} \sim \mathcal{N}(0, \sigma) \quad (2)$$

where $\kappa_t^{(1)}$ and $\kappa_t^{(2)}$ are two time varying stochastic period factors and \bar{a} is the mean of the ages in the data. The innovation $\epsilon_{(a,t)}$ is assumed to be drawn from an i.i.d. zero-mean Gaussian random variable with constant variance σ^2 . Such models describe the principal mortality dynamics in the sense that it includes the age related basis component and all of the non-stationary stochastic part of the mortality surface. The time-dependent parameters in both models are generally modeled using a simple ARIMA(0,1,0) model.

One of the drawbacks of these classical models relates to the assumption of homoskedastic error terms $\epsilon_{(a,t)}$. In fact, in practice, the assumption of constant variance is always violated: the observed logarithm of central death rates is much more variable and the volatility is time varying, see e.g. [Lee and Miller \(2001\)](#) and [Gao and Hu \(2009\)](#). Furthermore, the mortality evolution is known to be related to the age of birth, see [Willets \(2004\)](#). This is referred to as the cohort effect and translates the persistent of some shocks on mortality among cohorts. This is generally observed when plotting the residuals $\epsilon_{(a,t)}$ of models (1) and (2) as an apparent diagonal structure which requires additional univariate cohort processes in some countries. This phenomenon has lead to various extensions of the initial Lee-Carter model, e.g. [Renshaw and Haberman \(2006\)](#) or [Cairns et al. \(2009\)](#) and the reference therein. However, the inclusion of additional univariate processes enhances the goodness-of-fit of the model but *over-fit* the data and thus produces less reliable forecasts.

In contrast to this univariate factor-based framework, we approach the problem of modeling mortality rates by considering the whole surface as a sole random field without any further assumptions on the particular dependence structure neither the factors driving its evolution. Thus, unlike to mainstream approach, our modeling framework is intended to accommodate cross-cohorts dependence as well as conditional heteroskedasticity.

Through A Random Field Framework. Consider the process X_s parameterized by the lattice points $s = (a, t)$ and defined as the centered mortality improvement rates $\text{IR}_{(a,t)} = \log(m_{(a,t)}/m_{(a,t-1)})$. Formally, we let

$$X_s = \text{IR}_s - \overline{\text{IR}}, \quad (3)$$

where $\overline{\text{IR}}$ is the average improvement rate over $I \times J$. Empirical studies show that the differentiation of mortality rates (in the logarithmic scale) removes the non-stationarity, see for example [Chai et al. \(2013\)](#). This is also advocated by the models introduced above as the time-dependent factors are described by random walks with a constant drift. However, these models assume that IR_s only depends on the observed age. As noted before, the conditional average of the improvement rates should not only depend on the age but also on the cohort, i.e. $t - a$, as well as the rates experienced in adjacent generations. First, this is to allow for capturing the cohort effect, i.e. the persistent effect of mortality shocks in the same cohort. Second, including experience from neighboring generations allows improving the assessment of the global mortality. Formally, this will account for the diffusion phenomenon well known in demographic theory. Indeed, some changes in health risk behaviors are adopted first among some cohorts, and then

diffuse through the population. We can refer to this as learning or diffusive effects. Therefore, in order to account for correlations, in a natural way, across generations, a Markov property for the random field X_s is needed. Formally, we assume that X_s has interaction with a finite number of neighbors arranged in any manner around the point $s = (a, t)$. This neighborhood is denoted by $V \subset \mathbb{N}^2 \setminus \{0\}$. Its shape is of paramount importance as it directly conditions the causality of the random field. In [Figure 1](#), at the

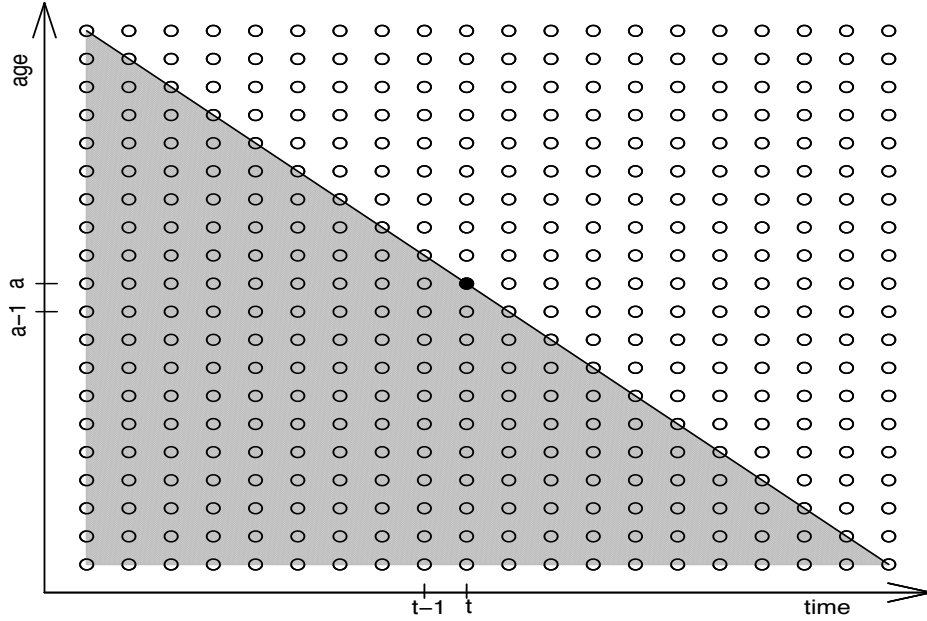


Figure 1: A bi-dimensional representation of the random field $X_s = \mathbb{I}R_s - \overline{\mathbb{I}R}$, with $s = (a, t)$. The grayed area represents the causal neighborhood V needed to characterize the evolution of X_s .

lattice points $s = (a, t)$, the neighborhood is depicted and corresponds to the grayed area, which obviously excludes the point s . This subset is causal in the sense of [Doukhan and Truquet \(2007\)](#), that is, stable by addition. Any, subset of this grayed area can be used to characterize the evolution of the process X_s .

Random Field Memory Models. The starting point of our approach is a formulation of the random field X in the sense of [Doukhan and Truquet \(2007\)](#). To this end, let $V \subset \mathbb{N}^2 \setminus \{0\}$ be a given (causal) neighborhood, let $\Theta \subset \mathbb{R}^d$ be a set of possible parameters, for d a positive integer, and consider $\xi = (\xi_s)_{s \in \mathbb{N}^2}$ an independent identically distributed (i.i.d.) random field, and F a given parametric function taking values in \mathbb{R} : $F : \mathbb{R}^V \times \Theta \times \mathbb{R} \rightarrow \mathbb{R}$ (endowed with a suitable norm $\|\cdot\|$), and consider solutions of the equation:

$$X_s = F\left((X_{s-v})_{v \in V}, \theta, \xi_s\right), \quad \text{with } s \in \mathbb{N}^2. \quad (4)$$

For $Z \in \mathbb{R}^k$ ($k \geq 1$) denote $\|Z\|_p = (\mathbb{E}\|Z\|^p)^{1/p}$ for an integer $p \geq 1$. In the above equation the set V , as discussed above, refers to a neighborhood of X_s used to characterize its evolution and $\theta \in \Theta$ are the parameters of the model. Recall that the set V does not contain the point $(0, 0)$, that is, $(s - v) \neq s$ for all $v \in V$. In this model, X_s is expressed in terms of its own past values and the present and past values of some adjacent cohorts but also a sequence of i.i.d. random variables. In the case where the set

V is causal, see [Doukhan and Truquet \(2007\)](#), the existence and uniqueness of a stationary solution of (4) in \mathbb{L}^p ($p \geq 1$) rely on the so-called contraction principle, which can be summarized in the following two properties:

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

The model in [Equation \(4\)](#) is a general formulation of random fields models with infinite interactions, which can be extended to the case where X takes values in \mathbb{R}^k and F takes values in \mathbb{R}^q , for integers $k, q > 1$. These models are not necessarily Markov, neither linear nor homoskedastic. Moreover the inputs do not need additional distributional assumptions. It is thus an extension of ARMA random fields which are special linear random fields, see e.g. [Loubaton \(1989\)](#) and [Guyon \(1995\)](#). Such an extension yields a novel random field which is capable of taking into account the nonlinearity and spatial dependency well adapted in our context. In other words, we can fit a particular spatial model of the form (4) to such data to give an appealing tool for investigating both spatiality and non-Gaussianity patterns in the evolution of mortality surface.

Based on this abstract formulation, we aim, henceforth, at proposing a specific form for the function F that is intended to tackle the various limitations encountered when using the factor-based mainstream approaches.

2.2 AR-ARCH-type random fields

We let $s = (a, t) \in \mathbb{N}^2$ and consider a first subset $V_1 \subset \mathbb{N}^2 \setminus \{0\}$ which characterizes the neighborhood associated to the autoregressive part of our model. That is, we consider a random field that satisfies the following autoregressive (AR) model

$$X_s = \xi_s \cdot \sigma_s + \sum_{v \in V_1} \beta_v X_{s-v}, \quad (5)$$

where the innovation $\sigma_s > 0$ is conditionally heteroscedastic in the sense that the conditional variance

$$\sigma_s^2 = \text{var}(X_s | \{X_{u,v}; u < a, v < t\})$$

is a non-deterministic random process. A class of conditionally heteroscedastic random field for the innovation is defined from a standardized random field ξ_s given as an autoregressive conditional heteroscedastic model (ARCH) as follows:

$$\sigma_s^2 = \alpha_0 + \sum_{v \in V_2} \alpha_v X_{s-v}^2,$$

with $V_2 \subset \mathbb{N}^2 \setminus \{0\}$, and where α_0 and $(\alpha_v)_{v \in V_2}$ are real positive coefficients. Such a specification is in accordance with some recent results obtained for mortality series that exhibit conditional heteroskedasticity and conditional non-normality features, which are significant empirical long-term mortality structures, see e.g. [Giacometti et al. \(2012\)](#), [Chai et al. \(2013\)](#) and [Chen et al. \(2015\)](#). The combined model is referred to as the AR-ARCH random field and is given by

$$X_s = \xi_s \sqrt{\alpha_0 + \sum_{v \in V_2} \alpha_v X_{s-v}^2} + \sum_{v \in V_1} \beta_v X_{s-v}. \quad (6)$$

The parameters of the model are then $\theta = (\alpha_0, (\alpha_v)_{v \in V_2}, (\beta_v)_{v \in V_1})$ where V_2 is the the neighborhood for the conditional variance and V_1 is the the neighborhood for the conditional mean. Note that the model in [Equation \(6\)](#) satisfies the assumptions **A-1** and **A-2** needed to ensure the existence and the uniqueness of a stationary solution. Indeed, note that the function F is given by

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

for $x = (x_v)_{v \in \mathbb{N}^2}$, $\theta = (\alpha_0, (\alpha_v)_{v \in V_2}, (\beta_v)_{v \in V_1})$, and $z \in \mathbb{R}$.

Then, letting μ be the law of the centered i.i.d. random field $(\xi_t)_{t \in \mathbb{N}^2}$, [Equation \(5\)](#) can be expressed in the following form

$$X_s = H((\xi_{s-t})_{t \in \mathbb{N}^2}), \quad H \in \mathbb{L}^p(\mathbb{R}^{\mathbb{N}^2}, \mu),$$

so that the conditions for the existence and uniqueness of stationary solution can be written as

$$\|\xi_0\|_p < \infty, \quad \kappa_p \equiv \|\xi_0\|_p \sum_{v \in V_1} |\alpha_v| + \sum_{v \in V_2} |\beta_v| < 1, \quad p \geq 1.$$

For $p = 2$, a weaker condition than the previous one is given by

$$\|\xi_0\|_2 < \infty, \quad \mathbb{E}\xi_0 = 0, \quad \kappa_2^2 \equiv \|\xi_0\|_2^2 (\sum_{v \in V_1} |\alpha_v|)^2 + (\sum_{v \in V_2} |\beta_v|)^2 < 1, \quad p = 2.$$

Example 1 (Three-Level Memory Model). We assume a Markov property for the random field $\{X_{(a,t)}; a = 1, \dots, I, t = 1, \dots, T\}$ in the sense that for all $s = (a, t)$ with $a, t > 1$, the following property holds

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

where for ease of notation we denoted

$$s^- = (a - 1, t), \quad s^+ = (a, t - 1), \quad s^- = (a - 1, t - 1).$$

In other words, the mortality improvement for any cohort s (born at time $t - a$) is solely related to the *immediately* adjacent cohorts. More precisely, s^- refers to the observation at time t stemming from the cohort born at date $t - a + 1$, whereas s^+ and s^- (respectively) refer to the observations at the last period $t - 1$ for the cohorts born at time $t - a - 1$ and $t - a$. These lattice points are represented in [Figure 2](#), with obvious notation. Next, in view of the Markov assumption [\(M\)](#), an example of a potential model can be described using the causal neighborhoods $V_2 = \{(1, 0), (0, 1)\}$ and $V_1 = \{(1, 1)\}$; so that the model in [Equation \(6\)](#) can be simply rewritten as

$$X_s = \xi_s \sqrt{\alpha_0 + \alpha^- X_{s^-}^2 + \alpha^+ X_{s^+}^2 + \beta X_{s^-}}, \quad (7)$$

where ξ_s is a standard i.i.d. random field, independent of the $X_{\tilde{s}}$ for any $\tilde{s} < s$ (in the sense that $\tilde{a} < a$ or $\tilde{t} < t$). Here, we make the convention $X_{s^-} = 0$ if $a = 0$, $X_{s^+} = 0$ if $t = 0$, $X_{s^-} = 0$ if $a = 0$ or $t = 0$.

The above model is referred to as the *three-level memory model*. The latter can be interpreted easily as follows. The autoregressive parameter β captures the cohort effect at the population level. Formally, it provides information on the persistence and inertia of the cohort effect. That is, the tendency of high (or low) improvement rates, for each cohort, to remain in the same level from one year to the next. In other words, the larger β the more persistent (strength) is the cohort effect. The remaining term of the

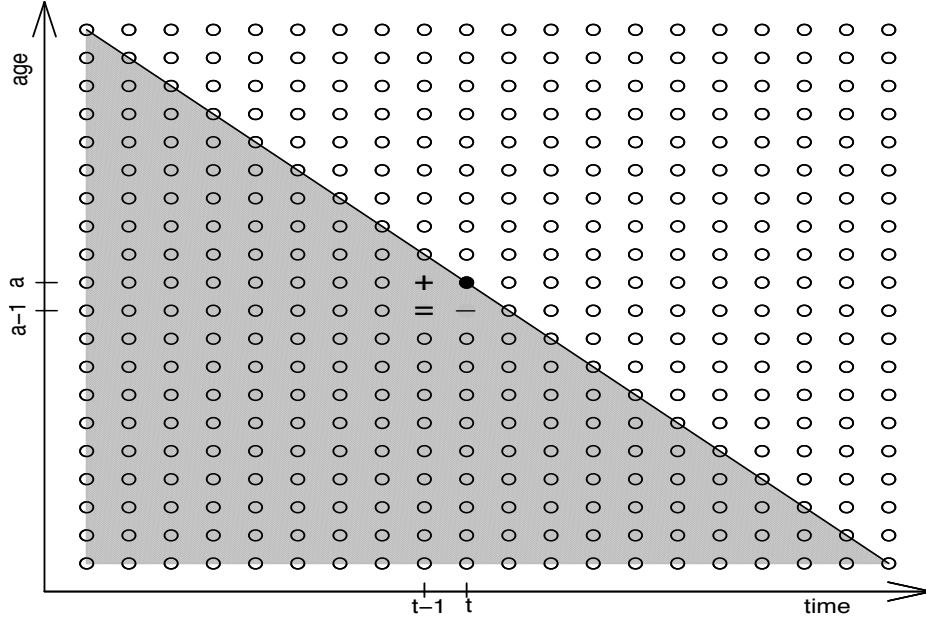


Figure 2: Evolution of mortality for three cohorts with respective year of birth: $t - a - 1$, $t - a$ and $t - a + 1$. The improvement rate at age a and calendar year t depends on the adjacent improvement rates (neighbor) (+) and (-) as well as the last observation on time $t - 1$, i.e. (=) .

right-hand side of Equation (7) allows to incorporate the conditional heteroskedasticity, in the sense that the conditional variance of X_s moves linearly in terms the adjacent cohorts (+) and (-) improvement rates. Accordingly, if the improvement rates for the neighboring cohort is large in magnitude, the forecast for next period's conditional variance, will be large.

It is important to recall that the three-level memory model is not necessary well adapted to all mortality surfaces. It is intended to capture the main behavioral aspects that may differ from one population to the other. For instance, the persistence of shocks over cohorts should need the inclusion of additional autoregressive factors at the cohort level but also at the adjacent neighbors. Thus, one has to enlarge the initial neighborhood $V_1 = \{(1, 1)\}$. The same remark applies to the choice of set V_2 . To do so, a rigorous inspection of the adequacy of the model (7) is needed. In other words, a model selection (neighborhoods selection) procedure is needed. However, this step is beyond the scope of this paper and will be considered in future research.

3 Estimation and Asymptotic Inference

In this section we study the maximum likelihood estimator (MLE) for the parameter θ in the AR-ARCH model (6). More precisely, we consider an approximation of the MLE called Quasi-Maximum Likelihood Estimator (QMLE) and prove its consistency, together with the asymptotic normality.

3.1 Quasi Maximum Likelihood Estimator (QMLE)

For the sake of simplicity we consider that the observation are real numbers, i.e. taking values in \mathbb{R} , the results we give can be extended to more general systems, but this remains a topic for future research. For a real function g and an integer p , let us write $\|g(X)\|_p := (\mathbb{E} |g(X)|^p)^{\frac{1}{p}}$, the \mathbb{L}^p norm.

Suppose that the random field is observed for $\mathcal{I} = \{s-v, v \in V_1 \cup V_2\}$ (the initial states) and $s \in \mathcal{O} \subset \mathbb{N}^2$. Let us write T the number of observations in \mathcal{O} , we assume that the following equation can always be written using $s \in \mathcal{O}$ and $s-v \in \mathcal{O} \cup \mathcal{I}$. Conditionally to \mathcal{I} the quasi log-likelihood of the model can be written as:

$$L_T(x_s, s \in \mathcal{O}; \theta) = \frac{1}{T} \left(\sum_{s \in \mathcal{O}} -\frac{1}{2} \ln \left(\alpha_0 + \sum_{v \in V_2} \alpha_v x_{s-v}^2 \right) - \frac{(x_s - \sum_{v \in V_1} \beta_v x_{s-v})^2}{2(\alpha_0 + \sum_{v \in V_2} \alpha_v x_{s-v}^2)} \right). \quad (8)$$

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

$$\hat{\theta}_T = \arg \max_{\theta \in \Theta} L_T(x_s, s \in \mathcal{O}; \theta). \quad (9)$$

We say that $\hat{\theta}_T$ is the QMLE estimator. In order to study the consistency of this estimator we assume the following properties:

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

H-2 The model is identifiable, in the sense that:

$$\begin{aligned} \left\| \sum_{v \in V_1} \beta_v X_{s-v} - \sum_{v \in V_1} \beta'_v X_{s-v} \right\|_2 = 0 &\Leftrightarrow (\beta_v)_{v \in V_1} = (\beta'_v)_{v \in V_1} \\ \frac{\alpha_0 + \sum_{v \in V_2} \alpha_v X_{s-v}^2}{\alpha'_0 + \sum_{v \in V_2} \alpha'_v X_{s-v}^2} \stackrel{a.s.}{=} 1 &\Leftrightarrow (\alpha_v)_{v \in V_2} = (\alpha'_v)_{v \in V_2}. \end{aligned}$$

H-3 The set of possible parameters Θ is compact and the true parameter θ^0 of the model (6) belongs to the interior of Θ .

Remark 1. These are classical assumptions required for consistency and asymptotic normality. The assumption **H-1** supposes that the variance of the random field X is finite, which is in line with the object of interest, i.e. mortality improvements. On the other hand, **H-2** ensures the identifiability of the model, which is a critical condition for consistency as we may see later on this section. This will impose that the quasi likelihood $L_T(x_s, s \in \mathcal{O}; \theta)$ has a unique maximum at the true parameter value θ^0 over the compact set Θ , from assumption **H-3**.

We can state the first result, where the proof is postponed to Appendix A.

Theorem 2. *For the model (6), under the assumptions **H-1**, **H-2** and **H-3***

$$L(\theta) = \lim_{T \rightarrow \infty} L_T(x_s, s \in \mathcal{O}; \theta)$$

exists a.s. for all $\theta \in \Theta$ and is uniquely maximized at θ^0 .

We can now proceed to show the consistency of the estimator $\widehat{\theta}_T$, i.e. the convergence (in probability) of the the estimated parameter $\widehat{\theta}_T$ to the true parameter θ . To this end, notice that the set of continuous functions \mathcal{G} defined as

$$\mathcal{G} = \left\{ g((x_s, (x_{s-v})_{v \in V_1}, (x_{s-v})_{v \in V_2}; \theta) = -\frac{1}{2} \ln \left(\alpha_0 + \sum_{v \in V_2} \alpha_v x_{s-v}^2 \right) - \frac{(x_s - \sum_{v \in V_1} \beta_v x_{s-v})^2}{2(\alpha_0 + \sum_{v \in V_2} \alpha_v x_{s-v}^2)}, \theta \in \Theta \right\},$$

is Glivenko-Cantelli for the \mathbb{L}_1 -norm since the set of possible parameters Θ is a compact set, i.e. assumption **H-3**. Now, applying Theorem 5.7 of [van der Vaart \(1998\)](#), we get:

Theorem 3 (Consistency). *If the assumptions in Theorem 2 hold, then the QMLE estimator $\widehat{\theta}_T$ in (9) is consistent:*

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

To prove the asymptotic normality, we need to reinforce the assumptions and add a finiteness condition on the fourth order moment of X_s , i.e.

$$\mathbf{H-4} \mathbb{E}(X_s^4) < \infty.$$

Under this additional assumption, we shall prove the asymptotic normality of the QMLE estimator. But before, let us write

$$A_0 = -\mathbb{E} \left(\frac{\partial^2 g(X_s, (X_{s-v})_{v \in V_1}, (X_{s-v})_{v \in V_2}; \theta^0)}{\partial \theta^2} \right), \quad (10)$$

and

$$B_0 = \mathbb{E} \left(\frac{\partial g(X_s, (X_{s-v})_{v \in V_1}, (X_{s-v})_{v \in V_2}; \theta^0)}{\partial \theta} \left(\frac{\partial g(X_s, (X_{s-v})_{v \in V_1}, (X_{s-v})_{v \in V_2}; \theta^0)}{\partial \theta} \right)^T \right), \quad (11)$$

then:

Theorem 4 (Asymptotic Normality). *Under assumption **H-1**, **H-2**, **H-3** and **H-4**,*

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

Remark 5. Let us give some interpretations of the above results.

(i) Asymptotic normal distribution result in [Theorem 4](#) allows us to approximate the distribution of the estimated parameters θ . This can be used, for instance, to obtain confidence intervals and to conduct hypothesis testing.

(ii) [Theorem 4](#) is also of paramount importance when it comes to mortality forecasting. As the amount of data at our disposal is, generally, limited, the parameter estimates using the QMLE in (9) must inevitably be subject to some degree of uncertainty. As demonstrated by [Cairns et al. \(2006\)](#), the uncertainty surrounding the parameters estimation forms a significant element of the uncertainty in forecasts of future mortality. Thus, [Theorem 4](#) allows us to quantify this uncertainty on parameters based on their asymptotic distribution.

Illustration with [Example 1](#). The parameters $(\alpha_0, \alpha^+, \alpha^-, \beta)$ of the three-level memory model in (7) can be estimated thanks to the QMLE in (8). The latter can be written, in this particular case, as follows:

$$L_T(x_s, s \in \mathcal{O}; \theta) = \frac{1}{T} \left(\sum_{s \in \mathcal{O}} -\frac{1}{2} \ln (\alpha_0 + \alpha^+ x_{s^+} + \alpha^- x_{s^-}) - \frac{(x_s - \beta x_{s^-})^2}{2(\alpha_0 + \alpha^+ x_{s^+} + \alpha^- x_{s^-})} \right),$$

with the notation introduced earlier: $s^+ = s - (1, 0)$, $s^- = s - (0, 1)$ and $s^\bar{=} = s - (1, 1)$. This (quasi) log-likelihood needs to be maximized numerically to get estimates for α_0 , α^+ , α^- and β . In the classical maximization approach, derivatives of the log-likelihood are required by numerical maximization algorithms. It is also of interest to derive the asymptotic distribution of the estimated parameters in [Theorem 4](#), which are used in [Equations \(10\) and \(11\)](#). The closed form formulas for the derivatives, in the general case, are given in [Appendix B](#).

3.2 Method of Moments

In this subsection, we briefly discuss an alternative method for estimating the model's parameters. We only focus on the illustrative example, but generalizing the approach to the class of models in [\(6\)](#) is straightforward. Formally, we introduce the method of moments, which is based on the derivation of some equations that relate the moments of the random field X_s to the parameters of the model. Besides the classical moments, numerous empirical studies have demonstrated the importance of the so-called leverages; that is, quantities of the form $\text{cov}(X_0, X_s^2) = \mathbb{E}(X_0 X_s^2)$. Their effects was first discovered by [Black \(1976\)](#) for some financial time series. In our case, we propose to adapt their expressions to obtain moment based estimators of the parameters $\theta = (\alpha_0, \alpha^+, \alpha^-, \beta)$, with the notation used in [Example 1](#). More precisely, we use second order moments, i.e. $\mathbb{E}(X_s^2)$, leverage moments, i.e. $\mathbb{E}(X_0 X_s^2)$, and extended leverage moments of the form $\mathbb{E}(X_0 X_s^p)$ for some integer p , to get estimators of α_0 , α^- and α^+ . Moreover, we use the relation

$$\mathbb{E}(X_s X_{\tilde{s}}^p) = \beta \mathbb{E}(X_s = X_{\tilde{s}}^p),$$

with $\tilde{s} < s$, to get an estimator of β as follows:

$$\hat{\beta} = \frac{1}{N} \sum_{(\tilde{s}, s) \in E} \frac{\widehat{\mathbb{E}}(X_s X_{\tilde{s}}^p)}{\widehat{\mathbb{E}}(X_s = X_{\tilde{s}}^p)}, \quad (12)$$

where $E = \{s^-, s^+, s^\bar{=}\} \times \{1, \dots, k\}$ for a given k , and $N = \text{card}(E)$, and where empirical moments can be estimated based on the raw data of the random field.

In order to estimate the remaining parameters α_0 , α^+ , and α^- , we use [Equation \(12\)](#) together with the expansions of $\mathbb{E}(X_0^2)$, $\mathbb{E}(X_0 X_{(1,0)}^2)$, $\mathbb{E}(X_0 X_{(0,1)}^2)$ and $\mathbb{E}(X_0 X_{(1,1)}^2)$ (which also may be written as $\mathbb{E}(X_s^2)$, $\mathbb{E}(X_s X_{(a+1,t)}^2)$, $\mathbb{E}(X_s X_{(a,t+1)}^2)$, $\mathbb{E}(X_s X_{(a+1,t+1)}^2)$) from stationarity), combined with the property that ξ_s and $X_{\tilde{s}}$ are independent for $\tilde{s} < s$ (in the sense that $\tilde{a} < a$ or $\tilde{t} < t$ for $s = (a, t)$ and $\tilde{s} = (\tilde{a}, \tilde{t})$), to get the following system:

$$\underbrace{\begin{pmatrix} 1 & M_2 & M_2 \\ 0 & M_3 & \mathbb{E}X_s X_{(a+1,t-1)}^2 \\ 0 & \mathbb{E}X_s X_{(a-1,t+1)}^2 & M_3 \end{pmatrix}}_A \underbrace{\begin{pmatrix} \alpha_0 \\ \alpha^- \\ \alpha^+ \end{pmatrix}}_\theta = \underbrace{\begin{pmatrix} M_2(1 - \hat{\beta}^2) \\ \mathbb{E}X_s X_{(a,t+1)}^2 - \mathbb{E}X_s X_{(a,t-1)}^2 \hat{\beta}^2 \\ \mathbb{E}X_s X_{(a+1,t)}^2 - \mathbb{E}X_s X_{(a-1,t)}^2 \hat{\beta}^2 \end{pmatrix}}_C,$$

of the form $A\theta = C$, with the notation $M_2 = \mathbb{E}(X_0^2)$ and $M_3 = \mathbb{E}(X_0^3)$.

Although this method is easy to implement, it is less attractive compared to the one based on the QMLE. Indeed, it is not an easy task to develop the asymptotic properties neither the consistency of the estimated parameters. In addition, the characterization of the asymptotic distribution made possible for

the QMLE, which we have derived in [Theorem 4](#), allows to quantify the parameters uncertainty and thus take into account the risk model. Therefore, in the following section, we mainly focus on the estimation procedure based on the QMLE method.

4 Numerical and Empirical Analyses

4.1 Simulation Study

Graphical Diagnosis. We start by discussing the specification of the model (6) in [Example 1](#). We illustrate [Example 1](#) with standard normal ξ_s and a parametrization such that assumptions **A-1** and **A-2** hold. The model parameter vector θ is chosen as follows.

$$\alpha^+ = \alpha^- = 0.1, \alpha_0 = 0.01 \text{ and } \beta = 0.7 \text{ or } 0.1.$$

We draw random samples of the underlying random field with these parameters values. We focus purely on the autoregressive part of the model to show how the illustrative model can reproduce the so-called cohort effect. As discussed earlier, the latter is mainly controlled by the autoregressive part of the model, i.e. the neighborhood V_1 in the general class and the parameter β in the illustrative model. In [Figure 3](#), we depicted two simulated random fields from the model in [Equation \(6\)](#) corresponding to $\beta = 0.1$ and $\beta = 0.7$, *ceteris paribus*. The right plot of [Figure 3](#) shows the improvement rates for a model with parameter $\beta = 0.1$. This would correspond to a population with weaker cohort effect. Indeed, the simulated improvement rates do not exhibit any *diagonal structure*, in extent that mortality rates, for a given cohort, do not persist from a year to another. This is a typical behavior of mortality patterns for some countries such as France, e.g. [Li et al. \(2016\)](#). The plot in the left-hand side of [Figure 3](#) depicts a sample with a higher value of β equal to 0.7. As we can see, for this simulation, a diagonal pattern appears due to the persistence of shocks. This effect is also observed for neighboring cohorts due to the dependency accounted in the ARCH part via the conditional variance. This particularity of the model is ignored in most mortality models. Therefore, a three-level memory model with higher values of β , tends

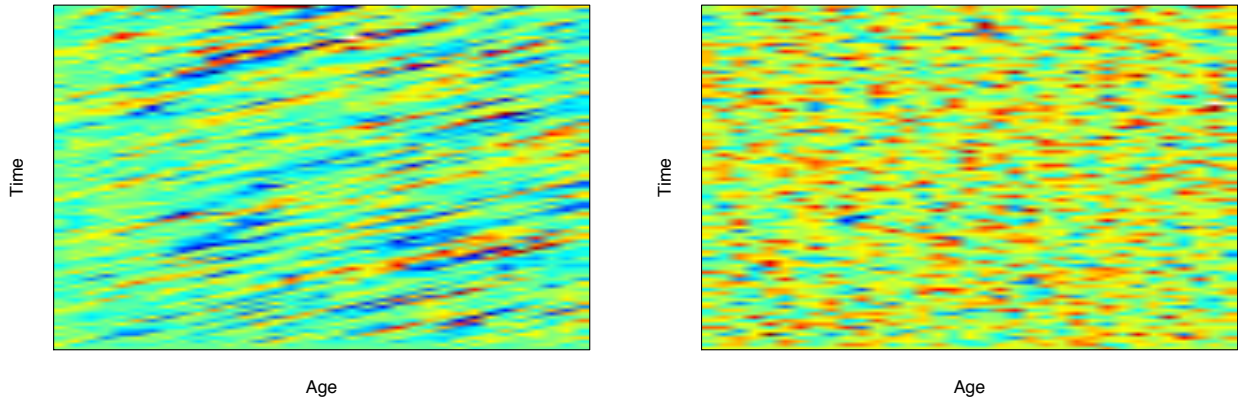


Figure 3: Simulations of the three-level memory random field with fixed conditional variance parameters: $\alpha^+ = \alpha^- = 0.1$. and $\alpha^0 = 0.01$. Two autoregressive parameters are considered: $\beta = 0.7$ (left) and $\beta = 0.1$ (right).

to respond to the criticisms which were addressed to the classical mortality models. Indeed, these have a tendency to omit, among others, some structural diagonal effects. As noted by [Hunt and Blake \(2014\)](#), the visual inspection of the benchmarking models, generally, exhibits obvious vertical and horizontal patterns but not diagonal banding evolution, which indicates the presence of further cohort effects. It is worth mentioning that models (1) and (2) produce improvement rates which are, visually, in accordance with the sample depicted in the right-hand side of [Figure 3](#).

It is important to point that another choice of the set V_1 , with a deeper dependency on the past as well as on the adjacent cohorts experiences, may produce a wider cohort effect. In other words, the diagonal patterns shown in the left-hand side of [Figure 3](#) will be larger, which is more in tune with the observed cohort effect, e.g., in England & Wales.

Monte-Carlo Analysis. This section examines the performance of the asymptotic estimation results in finite samples through Monte Carlo experiments. Data are generated using the three-level memory model in [Example 1](#), with parameters fulfilling the assumptions **A-1** and **A-2**. [Table 1](#) compares the

Table 1: Estimation of the parameters using the QMLE and comparison with theoretical values (between brackets)

		Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
α_0	0.06	0.001	0.034 (0.031)	0.060	0.061 (0.060)	0.078 (0.073)	0.231
α^-	0.01	0.006	0.009 (0.009)	0.010	0.010 (0.010)	0.011 (0.011)	0.013
α^+	0.20	0.082	0.166 (0.17)	0.193	0.194 (0.20)	0.220 (0.220)	0.332
β	0.50	0.331	0.453 (0.460)	0.496	0.498 0.500	0.544 (0.535)	0.708

distributions of the QMLE estimates of the parameters $\alpha_0, \alpha^+, \alpha^-$ and β over $N = 1000$ independent simulations of the illustrative model, for the sample size $I = 30$ and $J = 45$. These sizes are intended to replicate real-world cases where generally mortality is considered over an age band of 30 years, i.e. 60 to 89, and using the most recent available observations, i.e. approximatively 50 years. [Table 1](#) also reports the empirical estimates of the parameters values based on the Monte-Carlo experiment as well as the theoretical values given by [Theorem 4](#). We notice that the QMLE procedure succeed to give an accurate estimation of the initial values and the distributional values are also in accordance with the theoretical ones. However, we should note that there is a high uncertainty on the estimation of the parameter α_0 . This is not, nevertheless, of critical importance as the parameters does not greatly impact the mortality rates. Indeed, as we may see in the next experiment the values of such a parameter on real-world datasets is almost null.

4.2 Real-World Datasets

In this section we illustrate the three-level memory model introduced in [Example 1](#), with mortality data (death and exposure counts) obtained from the Human Mortality Database (HMD). We consider data from male population of United States (US), France and England & Wales (E&W). We use the logarithm

of the single-age crude death rates for ages 55 to 89, and from 1960 to 2012 for all the considered populations. First, we discuss the fit of the three-level memory model for the full data-set. Second, we conduct a comparative analysis with the LC and CBD models introduced, respectively, in Equations (1) and (2).

Parameters Estimates. The optimal values of α_0 , α^+ , α^- and β for the three populations are reported in Table 2. Here, we conduct a sensitivity analysis for the parameter estimates, initially, by fitting the model to the first 50 years of the data (right panel), from 1960 to 2009, and then, by fitting the model to the entire years of data from 1960 to 2012 (left panel). Note that, for each different estimation time-frame, the conditions for the existence and uniqueness of a stationary solution seems to be satisfied since $|\alpha_0| + |\alpha^+| + |\alpha^-| + |\beta| < 1$ holds for the three populations. Note also that the three estimations of α_0 , α^- , α^+ satisfy the condition of positivity. The sensitivity analysis aims at giving ground for robustness of the parameters with respect to the estimation time-frame. For the conditional variance (ARCH) as well as the conditional mean (AR) parameters, slight deviations occur with the exception of the AR parameter of the US population. Before trying to understand this difference for the US population, we shall, first, start by interpreting the parameters.

Table 2: The optimal values of three-level memory model estimated over two different periods for the three populations

	1960 – 2012				1960 – 2009			
	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
FR	-0.098	7.25E-4	0.315	0.445	-0.099	7.32E-4	0.328	0.463
E&W	0.028	6.79E-4	0.312	0.429	0.021	6.58E-4	0.334	0.441

(i) The AR parameter β governing the conditional mean is less than one (in absolute value) for the three populations. This is to say that the improvement rates for each cohort are stable over time, or the further back in time a given change or shock in mortality occurs, the less it will affect the present improvement. The given change dies out over time. The sign of coefficient as well as its absolute value provide information on the persistence or inertia of the cohort effect: the tendency of high (or low) improvement rates, for the same cohort, to remain in the same level from one year to the next. In other words, the larger β the more persistent (strength) is the cohort effect. In our case, E&W population has the most pronounced cohort effect with an AR coefficient of 0.028. This is consistent with the well-known E&W cohort effect which has been noted several times in literatures, see Willets (2004). Results for other populations (US and France) echo some recent works on cohort effects. Indeed, in Zhang and Zhao (2015), there are analyzes mortality cohort effect based on an approach that identifies and measures the strength of this effect in U.S. and E&W mortality datasets over time. They used a measure called *generation gap*, which describes how long the cohort effect maintains. Formally, it is the gap from the beginning to the end of a peak on the series of cohort effect, which can also be associated to the persistence mentioned above. Their finding shows that cohort effect in E&W tends to last, by far, longer than for U.S. population and thus for all generations. Similar conclusions were drawn by Li et al. (2016), using a different approach.

The latter observes a relatively strong cohort effect in the E&W, a medium level of cohort strength for the French males and finally a relatively weaker effect for the U.S. This is consistent with values reported in [Table 2](#).

(ii) The values of ARCH parameters α^+ and α^- , governing the conditional variance, are reported in [Table 2](#). As is the case of AR coefficient, the parameters values do not change substantially under the two estimation time-frames. Furthermore, note that these are not null, thus, conveying for a diffusion phenomena between adjacent generations, e.g. when some lifestyle factors highly affect the mortality pattern and thus diffuse through generations. As noted before, this may be interpreted as learning effect from the old and the young generations. Also, these parameters measure the extent to which shocks to cohorts (+) and (-) feed through into the cohort next-period improvement rate uncertainty. We should point out that the contribution of the young cohort (-) to the conditional variance is slightly lower than the old one, i.e. (+).

(iv) Coming back to the US males mortality, we observe that the persistence of the cohort effect, evidenced by the AR coefficient, recedes when we include the recent observation (ranging from 2010 to 2012). The parameter β goes from -0.024 to -0.013 . This deviation might not solely be explained by the uncertainty stemming from the estimation. A closer look into the recent advances in mortality patterns investigation may convey for a slowing cohort effect. This may be due, for example, to a changing regime leading to a weaker cohort effect for the very recent mortality experience.

(v) Here, we check the robustness of the parameters estimation a few steps further. Indeed, we consider the two time-frame where the academic literature has detected a shifting on the mortality behavior. Especially for E&W data, there is an obvious change in mortality dynamics (and hence, improvements as well) after 1980. This was evidenced, generally, by the change of the time-dependent component dynamics, i.e. κ , of the CBD as well as the LC. Therefore, we estimate parameters over the periods 1960 – 1980 and 1960 – 2012.

Table 3: The optimal values of three-level memory model estimated over two different periods, 1960 – 2012 and 1980 – 2012, for the E&W population

	1960 – 2012				1980 – 2012			
	β	α_0	α^+	α^-	β	α_0	α^+	α^-
E&W	0.028	6.79E-4	0.312	0.429	0.018	7.03E-04	0.321	0.4091

In [Table 3](#) we examine the impact of the two time-frame on the estimation of the parameters $\beta, \alpha_0, \alpha^+$ and α^- . The estimations are not so different for the two periods. We observe that the persistence parameter β is lower for the second period. This could be due to the uncertainty surrounding the estimation as the second time-frame has less observations, which makes the uncertainty increases compared to the first estimation period, see [Theorem 3](#). Another explanation for this slight difference on the parameter value may stem from a change on the cohort effect over the recent period. In other words, the records from the recent period 1980 – 2012 show that the mortality improvements at the cohort level is experiencing much lesser persistence than the full period 1960 – 2012 would suggest. Regardless of the reason, it must be noted that the AR-ARCH model’s estimation is not much dependent on the time-frames unless the dependence structure changes. Even if one could expect a more drastic change on these

estimates; as the academic literature has already mentioned a change on the drift of the mortality located in the early eighties; the dependence structure remains almost unchanged according to our analysis. To understand this, we recall that the quantities we are work with are the centered mortality rates. We can see that the average mortality, i.e. \bar{R} , over the two time periods has shifted significantly going from -1.837% to -2.486% , for the 1960 – 2012 and 1980 – 2012 estimations. This change on the average improvements is thus in accordance with the above mentioned observed behavior. This suggests that once the parameters are estimated one should take a careful look at the average mortality improvement incorporated in the model. However, this particular step is beyond the scope of this paper and should be considered in future works.

Forecasting Properties. Let us now evaluate the *out-of-sample* and *in-sample* forecasting performances of our model compared to the benchmarks, that are the [Lee and Carter](#)’s model in (1) and the [Cairns et al.](#)’s model in (2). Further, we calculate standard mortality risk metrics in order to quantify the differences between the forecasts of the mortality models.

(i) *In-Sample Analysis.* With the data described above we can now compare our model to the others using in-sample measures of goodness-of-fit. We first consider in-sample tests to stay consistent with existing literature. Here, we only rely on the RMSE. This measures the square root of the sum of squared differences between the actual crude mortality rates and the models described above compared to the three-level AR-ARCH memory model, i.e.

$$\text{RMSE} = \sqrt{\frac{1}{I \times J} \sum_{(a,t) \in \mathcal{I}} (m_{(a,t)} - \hat{m}_{(a,t)})^2},$$

where $\hat{m}_{(a,t)}$ denotes the observed crude mortality rate and $m_{(a,t)}$ is the output of the models.

In [Table 4](#) we report the results of all models in each country. For each population, we see that the

Table 4: Root mean squared of the in-sample errors

	AR-ARCH	LC	CBD
US	4.71E-04	7.57E-04	1.32E-03
FR	8.35E-04	1.07E-03	1.72E-03
UK	7.08E-04	8.61E-04	1.02E-03

AR-ARCH model has the least error and thus provides more accurate fit than the respective benchmarks. Therefore, our three-level memory model appears to better capture some of the complicated aspect of the mortality dynamics. Note also, that this performance is surprising in the extent that the model only works with four parameters conversely to the benchmarks, which tends generally to over-fit and better perform on the in-sample forecasting analysis. In order to understand this superiority, we investigate the forecasts of the mortality rates obtained form the model.

(ii) *Out-of-Sample Performance.* Even if the three-level memory model provides a good in-sample fit to historical data and produce forecasts that appear to be *biologically reasonable ex ante*, it is of interest to explore its *ex-post* forecasts. In other words, we carry out an out-of-sample analysis and we compare the forecast to the realized outcomes. To this end, we fit the models to data on the period ranging from

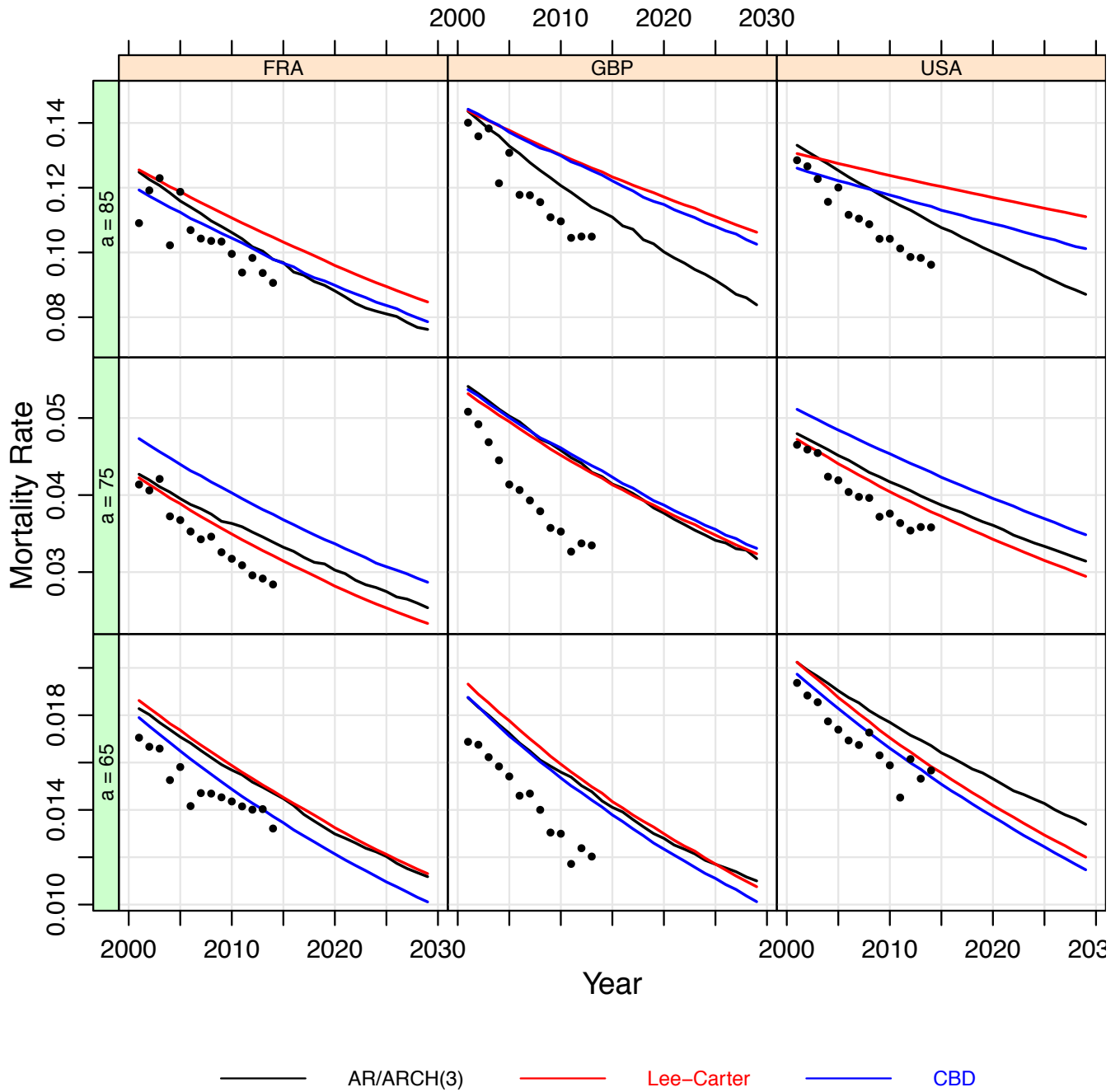


Figure 4: Projected median mortality rates for ages 65, 75 and 85. Models are fitted on the period 1960 – 1999 and compared to the raw rates (black circles). The graphs show the forecasts provided by the AR-ARCH model in [Example 1](#) (black line), Lee-Carter model (red line) and the CBD model (blue line) for the U.S. (right panel), E&W (middle) and France (left).

1960 to 1999 for each population, and forecast the future improvement rates from year 2000 to 2012. Next, we recover the corresponding mortality rates $m_{(a,t)}$. [Figure 4](#) displays the historical observations and the projected medians of the three models. At first sight, we see that the three-level memory model provides forecasts very close to the benchmarks. There is no tendency to a general out-performance of our model. Indeed, as we can see in [Figure 5](#), the plots of RMSE for different time horizons show that the tree-level memory model has comparable performance to the benchmarks. However, for high ages, our model seems to have more accurate performance. This confirms the graphical tendency observed in [Figure 4](#). Therefore, we can not draw a conclusion on the superiority or not of our model based on this comparative study. We shall examine some other mortality metrics in order to assess the behavior of the model and quantify the difference with the benchmarks.

(iii) *Life Expectancy*. Here, we use the forecasts provided by the out-of-sample analysis and derive the corresponding projected remaining period life expectancies. [Figure 6](#) shows 95% confidence intervals for life expectancies forecasts for ages 65, 75 and 85. We also depict with solid lines the median life expectancies forecasts. First, we can see that the three-level memory model produces forecasts at least equivalent to the LC or the CBD models. For some ages, our model gives more accurate predictions closer to the observed life expectancies. Unlike the forecasting performance of the crude mortality the illustrative model outperforms clearly the benchmarks when it comes to life expectancy prediction.

Regarding the confidence intervals, we can observe that those related to the benchmarks are quite narrow and underestimate the evolution of life expectancy, especially those related to the CBD model (red lines). This narrowness may result in underestimation of the risk of more extreme outcomes. Therefore, it should be point out that, for these datasets and the considered time-frame, the model in (1) and (2) underestimate largely the longevity risk. This may be explained by the rigid structure of the models as it was widely recognized in the literature. For instance, in [Lee \(2000\)](#), the LC model has been seen to generate overly narrow confidence intervals. However, the confidence intervals of the illustrative model appear more plausible and encompasses the observed life expectancies at the 95% level. This was made possible thanks to the inclusion of the heteroskedasticity of the conditional variance as well as the cohort effect allowing for a broader range of probable outcomes compared to the benchmarks.

(iv) *Comparison to the APC*. As the AR-ARCH random field is intended to capture the main stylized facts of mortality improvements and naturally addresses the cohort effect, we here compare the outputs of the illustrative model to a generalization of the models (1) and (2) with a cohort component. Here, we consider a modified version of the CBD model in [Equation \(2\)](#) that incorporates the cohort effect, see [Cairns et al. \(2009\)](#). Formally, we have the following formulation:

$$\text{logit } q_{(a,t)} = \kappa_t^{(1)} + \kappa_t^{(2)}(a - \bar{a}) + \gamma_{t-a} + \epsilon_{(a,t)}, \quad \epsilon_{(a,t)} \sim \mathcal{N}(0, \sigma), \quad (13)$$

where γ_{t-a} denote the stochastic cohort effect. Such a model is widely used in the literature, e.g. [Mavros et al. \(2017\)](#) and the reference therein. We fit this model on the same time-frame as well as the age groups used above. The forecasting performances of our three-level AR-ARCH model are compared to latter and are depicted in [Figures 7 and 8](#). We notice that the two models generate substantially similar outputs for the three ages, see [Figure 7](#). However, the forecast of our model are much smooth compared to the M6. This particular behavior of the age-period-cohort models is well known in the literature, see [Cairns et al. \(2009\)](#). Furthermore, when it comes to projection of life expectancy, we observe in [Figure 8](#) that the two models also behave in the same manner especially for the young ages. We notice that this

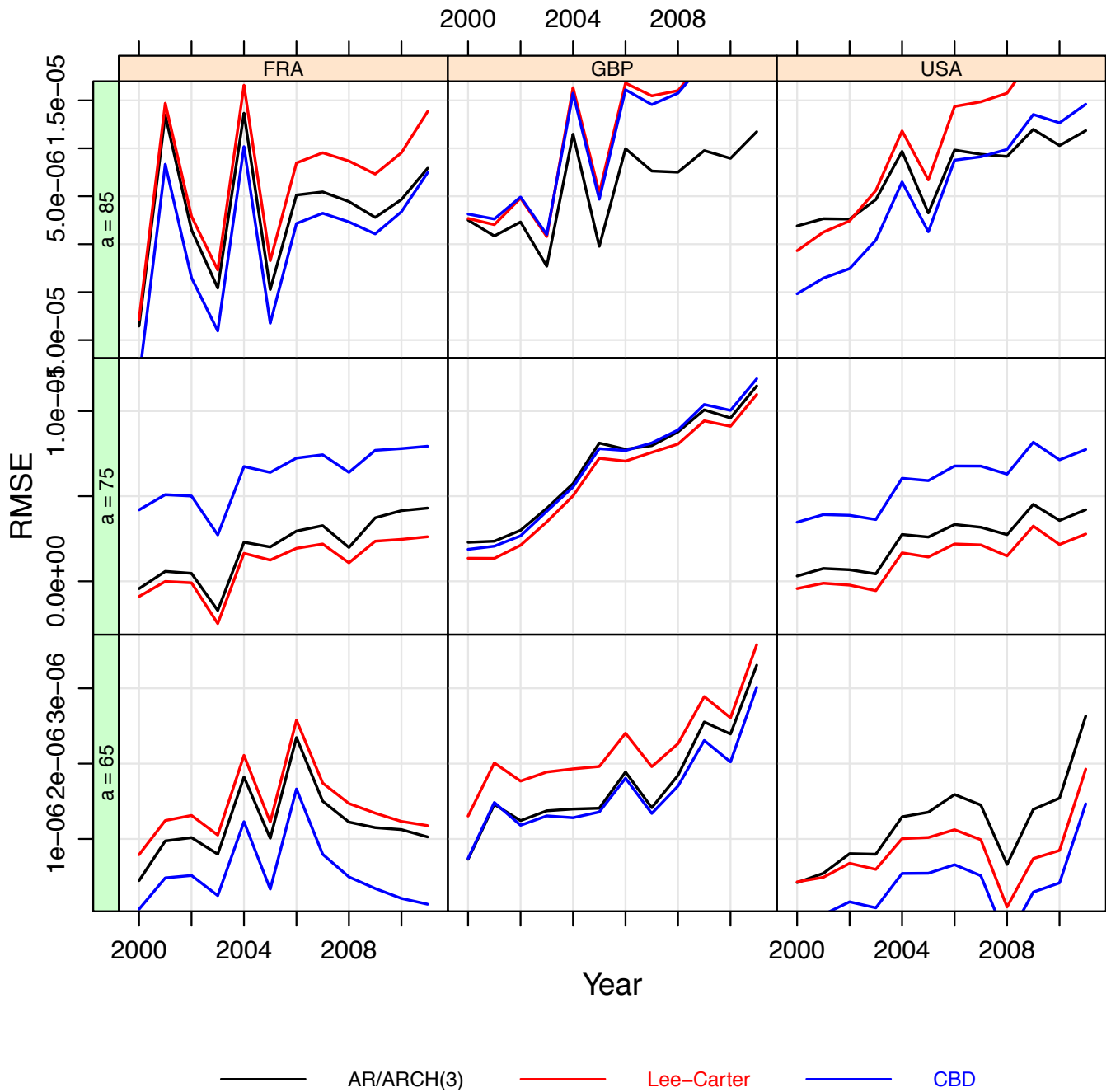


Figure 5: Evolution of mortality for three cohorts with respective year of birth: $t - a - 1$, $t - a$ and $t - a + 1$. The improvement rate at age a and calendar year t depends on the adjacent improvement rates (neighbor) as well as the last observation on time $t - 1$.

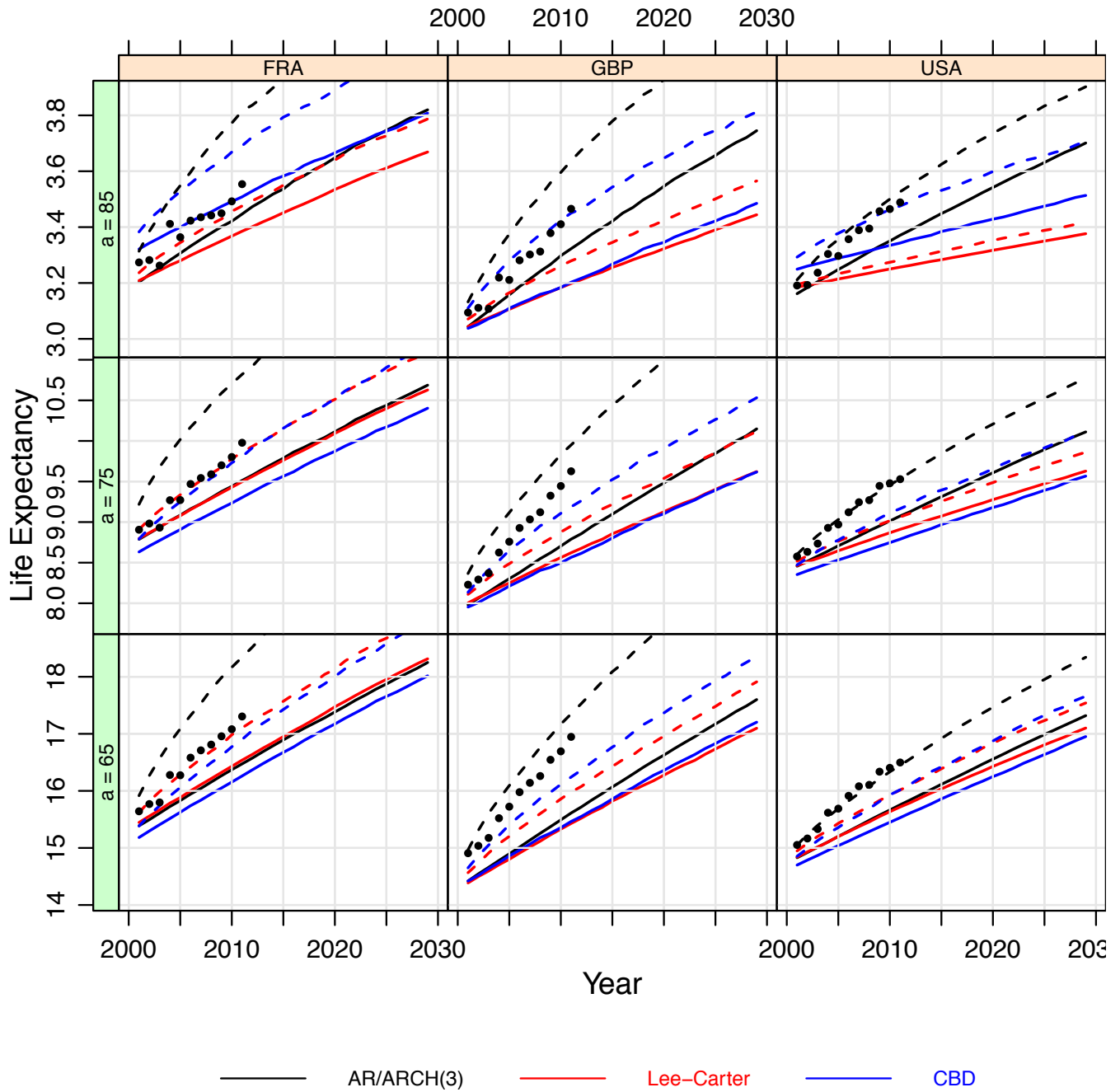


Figure 6: 90% extreme scenarii (dashed lines) and the projected median remaining period life expectancies (solid lines) for ages 65, 75 and 85. The graphs show the forecasts provided by the AR-ARCH model in [Example 1](#) (black line), Lee-Carter model (red line) and the CBD model (blue line) for the U.S. (right panel), E&W (middle) and France (left).

similarity tends to disappear as the time horizon increases. There is, in fact, a clear deviance of the M6 from the three-level model’s prediction and also from the actual life expectancies and the latter is systematically lower. In other words, the cohort-based model tends to underestimate the life expectancies in the long term. On the other hand, the inclusion of the cohort factor enhances slightly the confidence of the outputs compared to the models (1) and (2). In fact, the M6 model generates wider confidence intervals, compared to the LC and CBD models, but does not succeed to contain the actual values for the ages 65 and 75, which is the case of the AR-ARCH three-level memory model.

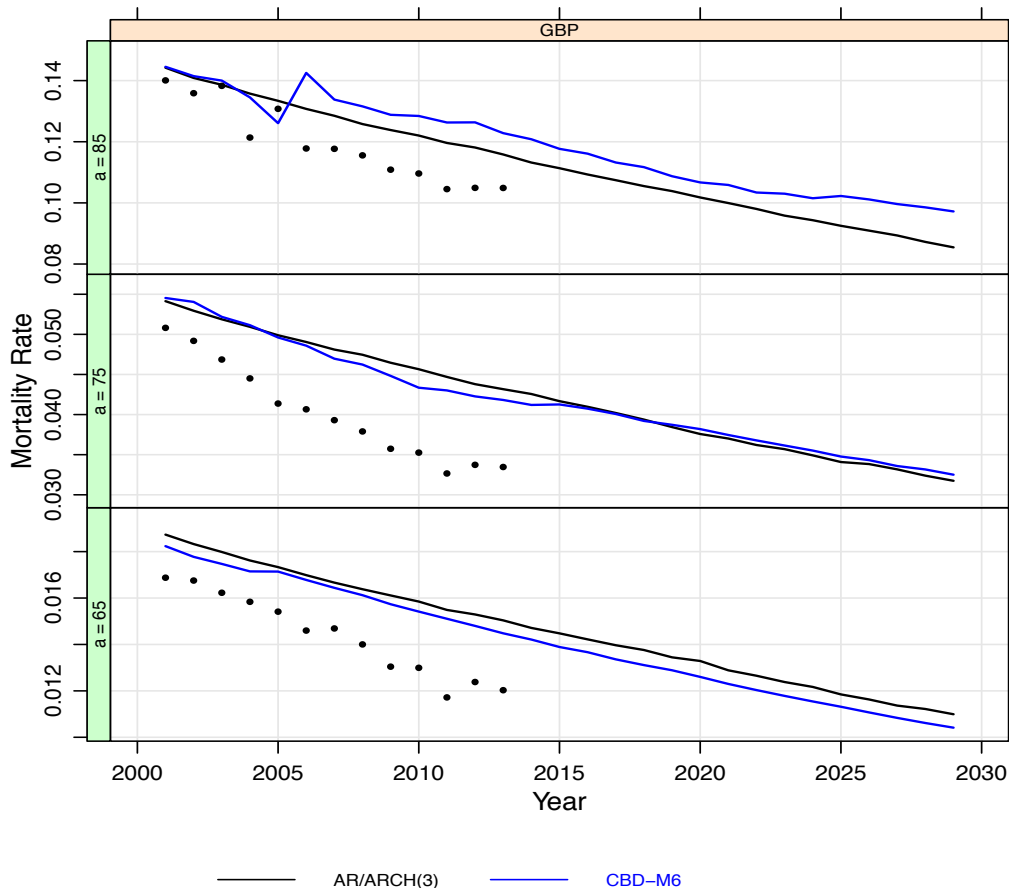


Figure 7: Projected median mortality rates for ages 65, 75 and 85. Models are fitted on the period 1960 – 1999 and compared to the raw rates (black circles). The graphs show the forecasts provided by the AR-ARCH model in [Example 1](#) (black line) and the CBD model with cohort effect in [Equation \(13\)](#) (blue line) for the E&W population.

5 Concluding Remarks

This paper proposes a class of random field models with a given causal structure. The underlying motivation is the desire to coherently model mortality improvement rates taking into account some stylized facts already mentioned in the academic actuarial and demographic literature. 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

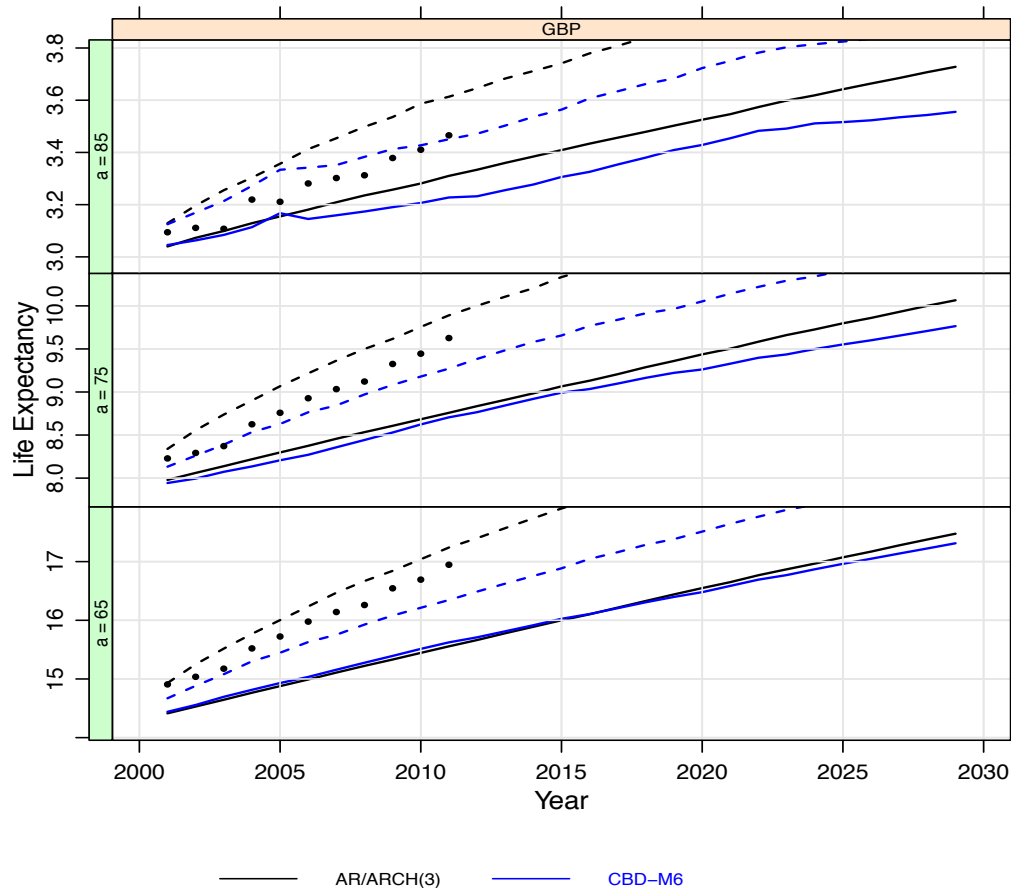


Figure 8: 90% extreme scenarii (dashed lines) and the projected median remaining period life expectancies (solid lines) for ages 65, 75 and 85. The graphs show the forecasts provided by the AR-ARCH in [Example 1](#) (black line) and the CBD model with cohort effect in [Equation \(13\)](#) (blue line) for the E&W male population.

for the parameters. Formally, we use the quasi-maximum likelihood estimator (QMLE), very popular for estimating univariate time-series. In our case, we show the statistical consistency of the QMLE and prove that the estimator is asymptotically normally distributed.

The framework being general, we investigate and illustrate a simple variant, called the three-level memory model, in order to fully understand and assess the effectiveness of the approach for modeling mortality dynamics. The predictive performance of such an approach, in comparison to two benchmark models, is studied, through the simple illustrative yet promising model. The empirical analysis uses data from male population of US, E&W and France. The results show that the three-level memory model provides, for high ages, an outstanding performance when it comes to the prediction of future life expectancy. However, the predictive ability of the model, for young ages, does not stand out from the references.

However, given the flexibility of the model, one can enhance the illustrative example. To do this, we should consider back the initial class and set an optimal characterization of the neighborhoods V_1 and V_2 . In fact, these should not be the same for all populations as the cohort effect and the inter-generational correlations are not behaving equally across populations. Although this model selection would be more

efficient, it goes beyond the scope of the present paper since it poses much more theoretical challenges. We leave the investigation of such interesting questions for future research.

On the other hand, for forecasting purposes, model bootstrapping is of paramount importance; after having estimated the parameters with any procedure developed in the current paper, it is also easy to fit residuals from the empirical characterization of the model in [Equation \(5\)](#).

$$X_s = \widehat{\xi}_s \cdot \widehat{\sigma}_s + \sum_{v \in V_2} \widehat{\beta}_v X_{s-v}, \quad \widehat{\sigma}_s^2 = \widehat{\alpha}_0 + \sum_{v \in V_1} \widehat{\alpha}_v X_{s-v}^2. \quad (14)$$

As this was also provided in both [Liebscher \(1999\)](#) for functional autoregressive models and in asymmetric ARCH models in [Doukhan and Mtibaa \(2016\)](#), we may proceed to the estimation of the innovation's density through such fitted innovations.

Also, next-period (one-step) prediction $X_s^{(1)}$ may also be considered if one observes all $X_{s'}$ with $s' < s$, this means $a' < a, t' \leq t$ or $a' \leq a, t' < t$ for $s = (a, t)$ and $s' = (a', t')$.

$$X_s^{(1)} = \sum_{v \in V_2} \beta_v X_{s-v}, \quad X_s^{(1)} - X_s = \sigma_s \xi_s. \quad (15)$$

Thus, in case $\mathbb{E}\xi_0 = 0$, $\mathbb{E}\xi_0^2 = 1$ and $\|\xi_0\|_m < \infty$ the prediction error may be controlled through the Markov inequality and $\mathbb{P}(|X_s^{(1)} - X_s| > x) \leq \mathbb{E}\sigma_s^m / x^m$. In case the innovations are known to be standard Gaussian, the values of Normal quantiles improve on this expression. Empirical versions of those relations yield predictions and confidence sets for those predictions.

Analogue 2-step predictors (this means that there exists some $s' < s$ such that one observes all the values of $X_{s''}$ for $s'' < s$) are deduced from iteration of the above relation,

$$X_s^{(2)} = \sum_{v \in V_2} \sum_{v' \in V_2} \beta_v \beta_{v'} X_{s-v-v'}.$$

The prediction error may now be written with the help of $\sigma_s^{(1)} \xi_s$, with a less tractable expression:

$$(\sigma_s^{(1)})^2 = \alpha_0 + \sum_{v \in V_1} \alpha_v (X_{s-v}^{(1)})^2.$$

Clearly further k -step predictions can be obtained with

$$X_s^{(k)} = \sum_{v_1 \in V_2} \cdots \sum_{v_k \in V_2} \beta_{v_1} \cdots \beta_{v_k} X_{s-v_1-v_2-\cdots-v_k}.$$

Anyway, the expression of the prediction error is then more cumbersome than for $k = 1$ and is left for further work since it writes recursively from $\sigma_s^{(k-1)} \xi_s$ with

$$(\sigma_s^{(k-1)})^2 = \alpha_0 + \sum_{v \in V_1} \alpha_v (X_{s-v}^{(k-1)})^2$$

The leverage properties of our model are investigated in [Subsection 3.2](#). Some easy improvement of the model would even increase this leverage effect as in [Doukhan et al. \(2016\)](#), which are exploited in the simplest possible AR(1)-model in [Doukhan and Mtibaa \(2016\)](#). For this, consider the model (5), with now:

$$\sigma_s^2 = \alpha_0 + \sum_{v \in V_1} (\alpha_v X_{s-v} + \gamma_v)^2.$$

Exploration of such new random fields models will be the aim of further studies.

References

- F. Black. Studies in stock price volatility changes. In *Proc. 1976 Business Meeting of the Business and Economics Statistics*, pages 177–181. Amer. Statist. Assoc., Chicago, 1976.
- N. Brouhns, M. Denuit, and J.K. Vermunt. A poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, 31(3):373–393, 2002.
- A.J.G. Cairns, D. Blake, and K. Dowd. A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance*, 73(4):687–718, 2006.
- A.J.G. Cairns, D. Blake, K. Dowd, G.D. Coughlan, D. Epstein, A. Ong, and I. Balevich. A quantitative comparison of stochastic mortality models using data from england and wales and the united states. *North American Actuarial Journal*, 13(1):1–35, 2009.
- 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 N. Mtibaa. Weak dependence: an introduction through asymmetric arch models. In F. Chaari, J. Leskow, A. Napolitano, and A. Sanchez-Ramirez, editors, *Cyclostationarity: Theory and Methods*, Lecture Notes in Mechanical Engineering. Springer, 2016.
- 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.
- P. Doukhan, I. Grublyte, and D. Surgailis. A nonlinear model for long memory conditional heteroscedasticity. *Lithuanian Journal of Mathematics*, 56(2):164–188, 2016.
- Q. Gao and C. Hu. Dynamic mortality factor model with conditional heteroskedasticity. *Insurance: Mathematics and Economics*, 45(3):410–423, 2009.
- 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.
- X. Guyon. *Random fields on a network: modeling, statistics, and applications*. Springer Science & Business Media, 1995.
- A. Hunt and D. Blake. A general procedure for constructing mortality models. *North American Actuarial Journal*, 18(1):116–138, 2014.
- A. Hunt and A.M. Villegas. Robustness and convergence in the lee–carter model with cohort effects. *Insurance: Mathematics and Economics*, 64:186–202, 2015.
- 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.
- R. Lee. The lee-carter method for forecasting mortality, with various extensions and applications. *North American actuarial journal*, 4(1):80–91, 2000.
- R. Lee and T. Miller. Evaluating the performance of the lee-carter method for forecasting mortality. *Demography*, 38(4):537–549, 2001.
- R.D. Lee and L.R. Carter. Modeling and forecasting US mortality. *Journal of the American statistical association*, 87(419):659–671, 1992.
- H. Li, C. O’hare, and F. Vahid. Two-dimensional kernel smoothing of mortality surface: An evaluation of cohort strength. *Journal of Forecasting*, 2016.
- E.A. Liebscher. Estimating the density of residuals in autoregressive models. *Stat. Inf. Stoch. Proc.*, 2:105–117, 1999.

- 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. Loubaton. *Champs stationnaires au sens large sur Z_2 : Propriétés structurelles et modèles paramétriques*. Ecole nationale supérieure des télécommunications, 1989.
- G. Mavros, A.J.G. Cairns, T. Kleinow, and G. Streftaris. Stochastic mortality modelling: Key drivers and dependent residuals. *to appear in North American Actuarial Journal*, 2017.
- A.E. Renshaw and S. Haberman. A cohort-based extension to the lee–carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, 38(3):556–570, 2006.
- A. van der Vaart. *Asymptotic statistics*. Cambridge University Press, 1998.
- R.C. Willets. The cohort effect: insights and explanations. *British Actuarial Journal*, 10(04):833–877, 2004.
- N. Zhang and L. Zhao. Mortality cohort effect detection and measurement based on differential geometry. *arXiv preprint arXiv:1504.00327*, 2015.

Appendix A Proof of Theorem 2.

Thanks to the law of large numbers applied to the stationary process $(X_s)_{s \in I \times J}$, we get

$$L(\theta) = \lim_{T \rightarrow \infty} L_T(X_s, s \in \mathcal{O}; \theta) = -\frac{1}{2} \mathbb{E} \left(\ln \left(\alpha_0 + \sum_{v \in V_2} \alpha_v X_{s-v}^2 \right) \right) - \mathbb{E} \left(\frac{(X_s - \sum_{v \in V_1} \beta_v X_{s-v})^2}{2(\alpha_0 + \sum_{v \in V_2} \alpha_v X_{s-v}^2)} \right).$$

This expectation is well defined since $\alpha_0 > 0$ and

$$\mathbb{E} \left(\ln \left(\alpha_0 + \sum_{v \in V_2} \alpha_v X_{s-v}^2 \right) \right) \leq \ln \left(\mathbb{E} \left(\alpha_0 + \sum_{v \in V_2} \alpha_v X_{s-v}^2 \right) \right) \leq \infty.$$

Now, for any $\alpha = (\alpha_v)_{v \in V_2}$, write $C(\alpha) = -\frac{1}{2} \mathbb{E} (\ln (\alpha_0 + \sum_{v \in V_2} \alpha_v X_{s-v}^2))$ then

$$\begin{aligned} L(\theta) &= C(\alpha) - \mathbb{E} \left(\frac{(X_s - \sum_{v \in V_1} \beta_v X_{s-v})^2}{2(\alpha_0 + \sum_{v \in V_2} \alpha_v X_{s-v}^2)} \right), \\ &= C(\alpha) - \mathbb{E} \left(\frac{(X_s - \sum_{v \in V_1} \beta_v^0 X_{s-v} + \sum_{v \in V_1} \beta_v^0 X_{s-v} - \sum_{v \in V_1} \beta_v X_{s-v})^2}{2(\alpha_0 + \sum_{v \in V_2} \alpha_v X_{s-v}^2)} \right), \end{aligned}$$

and $X_s - \sum_{v \in V_1} \beta_v^0 X_{s-v} = \xi_s \sqrt{\alpha_0^0 + \sum_{v \in V_2} \alpha_v^0 X_{s-v}^2}$, hence

$$L(\theta) = C(\alpha) - \mathbb{E} \left(\frac{\xi_s^2 (u_0^0 + \sum_{v \in V_2} u_v^0 X_{s-v}^2)}{2(\alpha_0 + \sum_{v \in V_2} \alpha_v X_{s-v}^2)} \right) + \mathbb{E} \left(\left(\frac{\sum_{v \in V_1} \beta_v^0 X_{s-v} - \sum_{v \in V_1} \beta_v X_{s-v}}{2(\alpha_0 + \sum_{v \in V_2} \alpha_v X_{s-v}^2)} \right)^2 \right).$$

Next, note that for any fixed $\alpha = (\alpha_v)_{v \in V_2}$, $L(\theta)$ is maximum for $(\beta_v)_{v \in V_1} = (\beta_v^0)_{v \in V_1}$ since the model is identifiable (assumption **H-2**).

Let $\theta(\alpha)$ be a parameter vector where $(\beta_v)_{v \in V_1} = (\beta_v^0)_{v \in V_1}$ and $\alpha = (\alpha_v)_{v \in V}$, then

$$\begin{aligned} &L(\theta(\alpha)) + \frac{1}{2} \mathbb{E} (\ln(\alpha_0^0 + \sum_{v \in V_2} u_v^0 X_{s-v}^2)) \\ &= -\frac{1}{2} \mathbb{E} \left(\ln \left(\frac{\alpha_0 + \sum_{v \in V_2} \alpha_v X_{s-v}^2}{u_0^0 + \sum_{v \in V_2} u_v^0 X_{s-v}^2} \right) + \frac{\xi_s^2 (\alpha_0^0 + \sum_{v \in V_2} \alpha_v^0 X_{s-v}^2)}{\alpha_0 + \sum_{v \in V_2} \alpha_v X_{s-v}^2} \right), \\ &= -\frac{1}{2} \mathbb{E} \left(\ln \left(\frac{\alpha_0 + \sum_{v \in V_2} \alpha_v X_{s-v}^2}{\alpha_0^0 + \sum_{v \in V_2} \alpha_v^0 X_{s-v}^2} \right) + \frac{\alpha_0^0 + \sum_{v \in V_2} \alpha_v^0 X_{s-v}^2}{\alpha_0 + \sum_{v \in V_2} \alpha_v X_{s-v}^2} \right), \end{aligned}$$

since ξ_s is independent from X_{s-v} and $\mathbb{E}(\xi_s^2) = 1$. Now, since $\ln(x) + 1/x \geq 1$ with equality if and only if $x = 1$ and thanks to assumption **H-2**, we conclude that $L(\theta(\alpha))$ is maximum for $\alpha = \alpha^0 = (\alpha_v^0)_{v \in V_2}$ and, finally, $L(\theta)$ is maximum for θ^0 . \square

Appendix B Proof of Theorem 4.

First, let us explicitly compute the derivatives appearing in Equations (10) and (11). After some algebra, we have

$$\frac{\partial g(X_s, (X_{s-v})_{v \in V_1}, (X_{s-v})_{v \in V_2}; \theta^0)}{\partial \alpha_0} = -\frac{1}{2} \left(\frac{1}{\alpha_0^0 + \sum_{v \in V_2} \alpha_v^0 X_{s-v}^2} - \frac{(X_s - \sum_{v \in V_1} \beta_v^0 X_{s-v})^2}{(\alpha_0^0 + \sum_{v \in V_2} \alpha_v^0 X_{s-v}^2)^2} \right),$$

- for v in V_2 :

$$\frac{\partial g(X_s, (X_{s-v})_{v \in V_1}, (X_{s-v})_{v \in V_2}; \theta^0)}{\partial \alpha_v} = -\frac{1}{2} \left(\frac{X_{s-v}^2}{\alpha_0^0 + \sum_{v \in V_2} \alpha_v^0 X_{s-v}^2} - \frac{(X_s - \sum_{v \in V_1} \beta_v^0 X_{s-v})^2 X_{s-v}^2}{(\alpha_0^0 + \sum_{v \in V_2} \alpha_v^0 X_{s-v}^2)^2} \right),$$

- for v in V_1 :

$$\frac{\partial g(X_s, (X_{s-v})_{v \in V_1}, (X_{s-v})_{v \in V_2}; \theta^0)}{\partial \beta_v} = \frac{(X_s - \sum_{v \in V_1} \beta_v^0 X_{s-v}) X_{s-v}}{\alpha_0^0 + \sum_{v \in V_2} \alpha_v^0 X_{s-v}^2}.$$

Since $(X_s - \sum_{v \in V_1} \beta_v^0 X_{s-v})^2 = \xi_s^2 (\alpha_0^0 + \sum_{v \in V_2} \alpha_v^0 X_{s-v}^2)$, under assumption **H-4**, B_0 is well defined. Moreover

$$\frac{\partial^2 g(X_s, (X_{s-v})_{v \in V_1}, (X_{s-v})_{v \in V_2}; \theta^0)}{\partial \alpha_0^2} = -\frac{1}{2} \left(-\frac{1}{(\alpha_0^0 + \sum_{v \in V_2} \alpha_v^0 X_{s-v}^2)^2} + 2 \frac{(X_s - \sum_{v \in V_1} \beta_v^0 X_{s-v})^2}{(\alpha_0^0 + \sum_{v \in V_2} \alpha_v^0 X_{s-v}^2)^3} \right),$$

- for v in V_2 :

$$\frac{\partial^2 g(X_s, (X_{s-v})_{v \in V_1}, (X_{s-v})_{v \in V_2}; \theta^0)}{\partial \alpha_0 \partial \alpha_v} = -\frac{1}{2} \left(-\frac{X_{s-v}^2}{(\alpha_0^0 + \sum_{v \in V_2} \alpha_v^0 X_{s-v}^2)^2} + 2 \frac{(X_s - \sum_{v \in V_1} \beta_v^0 X_{s-v})^2 X_{s-v}^2}{(\alpha_0^0 + \sum_{v \in V_2} \alpha_v^0 X_{s-v}^2)^3} \right),$$

- for v in V_1 :

$$\frac{\partial^2 g(X_s, (X_{s-v})_{v \in V_1}, (X_{s-v})_{v \in V_2}; \theta^0)}{\partial \alpha_0 \partial \beta_v} = \frac{-2 (X_s - \sum_{v \in V_1} \beta_v^0 X_{s-v}) X_{s-v}}{(\alpha_0^0 + \sum_{v \in V_2} \alpha_v^0 X_{s-v}^2)^2},$$

- for v and v' in V_1 :

$$\frac{\partial^2 g(X_s, (X_{s-v})_{v \in V_1}, (X_{s-v})_{v \in V_2}; \theta^0)}{\partial \alpha_v \partial \alpha_{v'}} = -\frac{1}{2} \left(-\frac{X_{s-v}^2 X_{s-v'}^2}{(\alpha_0^0 + \sum_{v \in V_2} \alpha_v^0 X_{s-v}^2)^2} + 2 \frac{(X_s - \sum_{v \in V_1} \beta_v^0 X_{s-v})^2 X_{s-v}^2 X_{s-v'}^2}{(\alpha_0^0 + \sum_{v \in V_2} \alpha_v^0 X_{s-v}^2)^3} \right),$$

- for v in V_2 and v' in V_1 :

$$\frac{\partial^2 g(X_s, (X_{s-v})_{v \in V_1}, (X_{s-v})_{v \in V_2}; \theta^0)}{\partial \alpha_v \partial \beta_{v'}} = -\frac{(X_s - \sum_{v \in V_1} \beta_v^0 X_{s-v}) X_{s-v}^2 X_{s-v'}}{(\alpha_0^0 + \sum_{v \in V_2} \alpha_v^0 X_{s-v}^2)^2},$$

- for v and v' in V_1 :

$$\frac{\partial^2 g(X_s, (X_{s-v})_{v \in V_1}, (X_{s-v})_{v \in V_2}; \theta^0)}{\partial \beta_v \partial \beta_{v'}} = -\frac{X_{s-v} X_{s-v'}}{\alpha_0^0 + \sum_{v \in V_2} \alpha_v^0 X_{s-v}^2}.$$

It thus follows that, under assumption **H-4**, A_0 defined in Equation (10) is well defined. By definition

$$\mathbb{E} \left(\frac{\partial g(X_s, (X_{s-v})_{v \in V_1}, (X_{s-v})_{v \in V_2}; \theta^0)}{\partial \theta} \Big|_{((x_{s-v})_{v \in V_1}, (x_{s-v})_{v \in V_2})} \right) = 0,$$

and

$$\frac{\partial L_T(X_s, s \in \mathcal{O}; \theta^0)}{\partial \theta} = \frac{1}{T} \sum_{s \in \mathcal{O}} \frac{\partial g(X_s, (X_{s-v})_{v \in V_1}, (X_{s-v})_{v \in V_2}; \theta^0)}{\partial \theta}.$$

Reorder \mathbb{N}^2 in the standard way:

$$s = (a, t) \mapsto C(s) = C(a, t) = 1/2(a + t)(a + t + 1) + t, \quad (16)$$

which is a bijection between \mathbb{N}^2 and \mathbb{N} with for any $v \in V_1 \cup V_2$, $C(s - v) < C(s)$. This leads us back to some process indexed by \mathbb{N} :

$$\frac{\partial L_T (X_s, s \in \mathcal{O}; \theta^0)}{\partial \theta} = \frac{1}{T} \sum_{s \in \mathcal{O}} \frac{\partial g(X_{C(s)}, (X_{C(s-v)})_{v \in V_1}, (X_{C(s-v)})_{v \in V_2}; \theta^0)}{\partial \theta}.$$

According to this new indexation and thanks to the property **(P)** in [Subsection 3.2](#), the above process is easily proved to be a martingale. So, the martingale central limit theorem and the Theorem 5.23 in [van der Vaart \(1998\)](#) conclude the proof. \square