

# Theoretical Application Analysis of Bayesian Inference in Time Series Forecasting of Mortality

Xuetong Wang

School of Finance, Actuarial Studies and Statistics, Australian National University, Canberra, ACT2601, Australia

---

**Abstract:** Population mortality rate prediction is a key issue in pension pricing and insurance product design and a fundamental problem in demography and actuarial science. The Lee-Carter model has been shown to fit well in the long run, however, the frequentist framework is not well adapted to fully account for parameter uncertainty and does not offer a natural way to incorporate prior information. The bias of estimation is particularly problematic when the sample size is small. Bayesian inference takes the prior distribution and the likelihood function and produces the posterior distribution, which is a full probabilistic statement of mortality rate prediction. This work presents a theoretical review on the internal logic and technological pathways of the Bayesian Lee-Carter model, hierarchical models and model averaging methods. It is observed that the advantages of Bayesian approaches are in combining many sources of data, measuring uncertainty and sparse data, especially ideal for long-term prediction scenarios. However, they still have substantial limitations in reliably recovering age-specific mortality statistics, in constructing suitable prior distributions, and in computational efficiency. The real methodological contribution of Bayesian inference is to encourage the move from point estimation to probability prediction in the mortality rate prediction. This change is of considerable practical importance for actuarial practice and for policy analysis.

**Keywords:** Bayesian inference; Mortality rate prediction; Lee-Carter model; Hierarchical model; Uncertainty quantification.

---

## 1. Introduction

The use of time series methods to forecast population mortality rates have a direct impact on pension actuarial calculations, pricing of insurance products, and the fiscal balance of the social security system. It is basic to demographic and actuarial practice. From deterministic extrapolation to stochastic modelling, from a single model to comparing many models, mortality rate prediction has progressed since the single-factor model provided by Lee and Carter (1992). But this paradigm shift has always been limited by a basic contradiction: census data is low frequency, sampling monitoring data is high frequency. The two measurements are not consistent in their standards, and it is difficult to fit the enormous amount of accessible information into the same analytical frame. Actuaries and policy makers need more than expected values, they need a whole probability distribution taking into account parameter uncertainty and model uncertainty.

This conflict has not been solved in the classic frequentist paradigm. The Lee-Carter model is popular for its straightforward form, however estimate of the parameter standard errors relies largely on large-sample asymptotic theory. We demonstrate that the robustness of inference is substantially weakened if the sample size is small or if there are structural breaks in the time dimension and the age dimension of the data. The frequentist school assumes parameters to be fixed constants and makes inferences based on the assumption of repeated sampling. This epistemic assumption has a tension with the fact that population mortality events are not repeatable and historical processes are unique.

The above situation offers a theoretical window of Bayesian inference. The Bayesian approach, in essence, reconstructs the logic of information use, where the prior distribution stands for the historical cognition and the likelihood function stands for the current facts in the same

probability framework. Then the full posterior distribution of the parameters is taken as the shared starting point for inference and prediction [1]. This paper discusses the theoretical logic and applicable conditions of Bayesian inference in the time series prediction of population mortality rates, investigates the technical presuppositions of its key models and their importance in the actuarial practice, and tries to clarify the role and limits of this methodology in the promotion of the transformation of the prediction paradigm.

## 2. The Epistemological Logic of Bayesian Inference and Its Applicability to Mortality Data Characteristics

The frequentist approach builds inference on the premise of fixed parameters and repeatable data sampling. Take China for example, the death records of a certain time and space is just a snapshot in the historical process, it can not be re-sampled. Bayesian inference does not assume "fixed parameters", but permits the researcher to treat parameters as random variables. The cognition outside the data can also be integrated into the prior distribution and the repeated sampling assumption is no longer an insurmountable obstacle. The great usefulness of the Bayesian technique is not in the modification of a set of calculating formulae. It is in another potential in epistemology.

The Bayesian framework is naturally suitable for mortality prediction, and it has to come back to the properties of the data itself. China is a classic example. The data in 1982, 1990, 2000 and 2010 are from the national census, while the data in other years are mostly from the 1% population sample survey or the yearly population change sampling survey. The census and the sampling surveys are not uniform in their coverage and sampling methodologies and merely attempting to create a time series would certainly result in systematic variations. In Bayesian inference, the variance penalty terms can be

assigned to various quality data, so that data of worse quality would be given less weight in the parameter estimation [2]. There is no natural channel in the frequentist school to reach equivalent flexible weighting.

Within the parameters of mortality itself (age and time), there are also clear structural limits. The mortality rates of neighbouring ages are usually gradually varying and do not fluctuate wildly, and the improvement of mortality rates also has continuity in time. These two properties can be integrated into the model through the prior distribution. For example, we can put a random walk prior on the age effect, which is effectively a penalty on the differences in mutations between ages. When original data are insufficient, information from nearby ages or adjacent periods are borrowed, making the estimation results more stable.

In the end, mortality prediction has practical implications, such as the capacity to pay pensions. Knowing only an expected value is not sufficient, one needs to know additionally how reliable this estimate is. Point estimation without uncertainty interval is almost useless in practice. The Bayesian posterior sampling gives the whole distribution of the projected value, and the uncertainty in the parameter estimate step is also completely passed to the prediction stage, which is far more reliable than providing a single number only [3].

### **3. Bayesian Lee-Carter Model: A Theoretical Leap from Point Estimation to Parameter Distributionalization**

The Lee-Carter model has a not too complicated mathematical structure but a quite solid empirical performance. Therefore, it is widely considered as the benchmark model in the area of mortality prediction. Its basic formula is relatively simple: the logarithm of the central mortality rate is split into three multiplied components: the average value  $\alpha_x$  of age-specific, the sensitivity  $\beta_x$  of age-specific and the temporal trend  $\kappa_t$ . However, the conventional estimation method can only give point estimates of the parameters, and it is hard to adequately assess the volatility of  $\beta_x$  and  $\kappa_t$ . The mortality prediction intervals obtained are frequently excessively narrow.

The Bayesian Lee-Carter model is not a shell-casting. It replaces the "deterministic parameter decomposition" by the "random parameter evolution" [4].  $\kappa_t$  is formulated as a drift-driven random walk, and prior distributions are specified for  $\alpha_x$ ,  $\beta_x$  and the observation error variance. The observation equation links data to potential states, while the state equation describes the stochastic evolution of the mortality improvement trend. Together they create a complete dynamic linear model system. Combining Gibbs sampling with forward filtering and backward sampling produces not only the posterior mean of the parameters but also the complete posterior distribution of the parameters conditional on the data.

There are two advantages to this theoretical jump. First, it resolves the long-standing problem of narrow prediction intervals in actuarial practice. In traditional approaches, the extrapolation of  $\kappa_t$  is treated as a deterministic number, and the estimation error of  $\kappa_t$  does not propagate to the prediction step. The Bayesian Lee-Carter model (BLC) is not a simple shell-casting, and the uncertainty of the parameter estimate stage is automatically propagated to the prediction stage,

resulting in more appropriate interval widths. The second is that it allows age specific differences to be diverse. The death rates of the young and old populations vary far more than the middle-aged and young people in real data. BLC allows the observation error variances of different ages to be specified independently. When fitting the fluctuating age groups, the fitting quality of other age groups is not compromised.

But the more rigorous the theory, the more specific the actual cost often becomes. If the age groups are further refined, for example in single-year age groups, the parameter space of BLC gets quite large and the convergence diagnostics for the MCMC chain becomes very difficult. BLC is always a single-population overall model and cannot address the concerns that policymakers are worried about, that is, how to jointly estimate the mortality improvement trends in different regions and different genders? The major technical difficulty is the generalisation of the mortality prediction to several regions or several groups from one country. This is a methodological problem that has to be tackled in modelling—how to preserve individual distinctions while without losing the general similarity [5].

### **4. Bayesian Hierarchical Model: The Theoretical Construction of Multisource Information Integration and Information Borrowing Mechanism**

The Bayesian hierarchical model provides the methodological basis for resolving this paradox. However, when independent models are trained for each region, the estimation of parameters in data-weak regions is not robust, and when all regions are pooled to process, the key distinctions between regions are likely to be eliminated.

The gist of this concept is in terms of "partial information borrowing". The mortality rates parameters of each region are not fully independent of each other, but are assumed to be drawn from a similar overall distribution and the parameters of this distribution are likewise incorporated in the estimating system. If the data stock of a given region is not enough, the posterior estimation will tend to converge toward the center of the overall distribution [6]. If the data in this region is enough, the posterior estimation mostly relies on its own observational knowledge. This approach provides a middle ground between extreme independence and extreme pooling, avoiding the problem of parameter explosion owing to sparse data and the suppression of underlying differences between regions due to forced pooling.

One such implementation of this approach is the U.N. probabilistic population projection. In many parts of the world the base of population data is somewhat inadequate. It is infeasible and difficult to guarantee the quality of inference if we totally depend on the mortality data of each country itself. The hierarchical model is able to compensate for the lack of local data by borrowing global overall trend information [7]. The mortality data of each province in China likewise has comparable imbalances, and some places have relatively complete data, and some have significantly lacking data, and there are both differences and common improving tendencies between provinces. The hierarchical model structure is precisely suitable for such multi-region collaborative inference.

But the theoretical advantages of the hierarchical approach still need to be validated at practical level. The model

normally predicts life expectancy as the highest level prediction variable and then transforms it to age-specific death rates using the model life table. Although this method can ensure the logical consistency of macro total changes, the original information at the age dimension is compressed in the conversion process, and the special nature of the mortality rate form of the high-age segment is also weakened by the smoothing setting of the life table itself. The use of the hierarchical model is a continual challenge to find a more precise link between macro resilience and micro age-specific precision. These restrictions at the model level suggest another, more fundamental problem: how can we address the uncertainty of the model structure given the right choice of the model structure?

## 5. Quantification of Model Uncertainty: Theoretical Logic and Practical Anchoring of Bayesian Model Averaging

As indicated before, although the BLC and the hierarchical models differ in their methodology, they have the same theoretical basis, i.e., there is an actual data creation mechanism and the model we choose is an efficient representation of this mechanism. Empirical investigations of the mortality rate prediction reveal that models with very distinct theoretical bases such as the Lee-Carter and the CBD models can give very varied long-term prediction outcomes for the same data. In the case of selecting a model alone and excluding the others, there is no basis for the robustness of the prediction result [8].

The purpose of proposing the Bayesian model average is exactly to avoid the subjectivity of model selection. The theoretical aim is not to establish the true model but to treat the model as an uncertain item for which we are able to make an estimate. The BMA puts the model level under the Bayesian update rule and calculates the posterior probability of each model by computing the marginal likelihood of each candidate model. The final predictive distribution is a weighted average of the individual model predictions, where the weights are the posterior model probabilities. This mechanism assigns weights according to the model's performance in predicting the data, so that an automatic trade-off between complexity and goodness of fit is obtained—the integration of the marginal likelihood includes the Occam factor, and complex models will be naturally penalised.

By including BMA into mortality rate prediction, not only do we improve accuracy, we also change the way researchers think about uncertainty. The prediction intervals obtained from previous approaches generally only represent the fluctuations of the parameters. The intervals provided by BMA, however, also take into account the uncertainty on the model level [9]. When the pension insurance solvency is evaluated, the intervals after adding model uncertainty are closer to the actual inferential state: prediction mistakes arise from both parameter variations and the limits of knowing the ageing process.

BMA is conceptually well founded but its actual implementation is hampered by computing challenges. The marginal likelihood estimation has to compute high-dimensional integration, which is difficult to compute exactly when there are numerous candidate models, therefore approximate techniques such as bridge sampling or harmonic means are commonly needed. The more fundamental problem

is that the definition of the model set includes the subjective judgement of the researcher. If all candidate models are inaccurate—and Box's famous quote, "All models are wrong", is especially pertinent in mortality rate prediction—then the theoretical justification for their weighted average must be re-examined. Such uncertainties do not diminish its practical significance but rather stimulate scholars to distinguish between statistical posterior probability and actual probability. In the end, theoretical conflicts must be tested in practice, and that is what the next chapter will do.

## 6. Empirical Application Review and Theoretical Reflection on Bayesian Mortality Prediction

Practice is the final guarantee of the method. The most important recent development of the Bayesian approach to mortality forecasting has been the official adoption of Bayesian probability forecasting by the United Nations Population Division in the World Population Prospects. This selection shows that this method has successfully moved from academic research to international statistical standards [10]. From an application point of view, the stratified model is more robust than the traditional deterministic extrapolation method in describing the trajectory of mortality improvement in different countries, especially for countries with weak data, and the information borrowing mechanism can effectively buffer the disturbance caused by local outliers.

And the actuarial industry is responding to this rationale, too. For longevity swaps, mortality linked securities etc we need mortality prediction for dynamic updates and even intraday computation capabilities. These constraints are equivalent to properties of Bayesian sequential update. When fresh mortality data are released, there is no need to conduct the whole MCMC sampling from the beginning; only utilising the posterior distribution of the last round as the prior distribution of the current round can swiftly update the prediction findings. Recently, some authors have embedded in the Bayesian framework ideas of "vanishing jumps" to turn exogenous shocks such as pandemics into part of the model. The conclusion is that uncommon but persistent structural fractures can occur from time to time while undergoing mortality improvement and should not be treated as normal anomalies.

Of course there are still theoretical problems with this approach. An example is the choice of the prior variance [11]. If the constraints are too strict, the model will overlook the future probable mutations. If they are too loose, the prediction interval may lose its reference value. There are also challenges at the computational level. MCMC sampling of high-dimensional stratified models is time expensive and the performance of convergence diagnostics like Gelman-Rubin is not always satisfactory in complex models. Even when they use the identical data, different researchers can nevertheless find slightly different posterior outcomes.

There's also a deeper issue worth talking about. Does the explanatory power of population transition theory on mortality changes get diluted when the Bayesian framework becomes increasingly dependent on exact prior structure and sophisticated calculation process? The key is about framing. Bayesian inference should not be seen as a black box which automatically produces predictions but as a tool to formalise theoretical assumptions. The prior selection should be made in accordance with the population theory for the balance of

the technical precision and the theoretical explanation.

## 7. Conclusion

The Bayesian inference introduces the prediction of the mortality rate and its value appears at two levels. On the technical side, it offers a number of operational frameworks that include uncertainty in statistical inference. On the epistemological side, it calls on academics to reconsider the nature of the prediction activity itself. Repeated trials cannot be used to observe the mortality rates; uncertainty arises from both sampling errors and structural blind spots in understanding complex population systems. The Bayesian framework, unlike the frequentist school, mathematizes these blind spots using posterior distributions.

The Bayesian Lee-Carter approach makes clear internally how the random fluctuations from the parameter estimate step are passed to the prediction step. The hierarchical approach incorporates multi-source data fusion and the borrowing of subgroup information into a unified framework, addressing the issue of data heterogeneity. Model averaging takes account of model structure uncertainty at a higher level. The three approaches differ in focus but are, in a sense, progressive in depth: from the random fluctuations in parameter estimates, to the structural variations in data heterogeneity, and to the uncertainty of the model structure itself.

This method also has limits. The choice of priors is hard to escape from subjective judgement based on historical trends, the computational burden of high dimensional MCMC sampling and the convergence diagnosis is not always satisfactory and the delineation of the candidate model set is not completely objective either. Recognising these constraints is to place it as an extension on the methodological range. Practically, it can be a more prudent choice to perform a cross-check of Bayesian inference with frequentist methods and to make a prior sensitivity analysis a fixed step.

Future advances may occur from the three-dimensional interplay of age-period-cohort combined with Bayesian deep probabilistic programming. But whatever the updating of tools, the most basic contribution of Bayesian inference is that it forces the researcher to confront his cognitive limits. China is undergoing rapid ageing and is improving its data gathering systems at the same time. Building a prior knowledge system

that represents the peculiarities of China's demographic transition is both a theoretical necessity and practical problem for China.

## References

- [1] Barigou, K., Goffard, P. O., Loisel, S., & Salhi, Y. (2023). Bayesian model averaging for mortality forecasting using leave-future-out validation. *International Journal of Forecasting*, 39(2), 674–690.
- [2] Li, J. S. H., Zhou, K. Q., Zhu, X., Chan, W. S., & Chan, F. W. H. (2019). A Bayesian approach to developing a stochastic mortality model for China. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 182(4), 1523–1560.
- [3] Fung, M. C., Peters, G. W., & Shevchenko, P. V. (2017). A unified approach to mortality modelling using state-space framework: characterisation, identification, estimation and forecasting. *Annals of Actuarial Science*, 11(2), 343–389.
- [4] Pedroza, C. (2006). A Bayesian forecasting model: predicting US male mortality. *Biostatistics*, 7(4), 530–550.
- [5] Antonio, K., Bardoutsos, A., & Ouburg, W. (2015). Bayesian Poisson log-bilinear models for mortality projections with multiple populations. *European Actuarial Journal*, 5(2), 245–281.
- [6] Alexander, M., Zagheni, E., & Barbieri, M. (2017). A flexible Bayesian model for estimating subnational mortality. *Demography*, 54(6), 2025–2041.
- [7] Ševčíková, H., & Raftery, A. E. (2016). bayesPop: probabilistic population projections. *Journal of Statistical Software*, 75, 1–29.
- [8] Bravo, J. M., & Ayuso, M. (2021). Linking pensions to life expectancy: tackling conceptual uncertainty through Bayesian model averaging. *Mathematics*, 9(24), 3307.
- [9] Wong, J. S., Forster, J. J., & Smith, P. W. (2018). Bayesian mortality forecasting with overdispersion. *Insurance: Mathematics and Economics*, 83, 206–221.
- [10] Raftery, A. E., Alkema, L., & Gerland, P. (2014). Bayesian population projections for the United Nations. *Statistical Science: A Review Journal of the Institute of Mathematical Statistics*, 29(1), 58.
- [11] Bohk-Ewald, C., & Rau, R. (2017). Probabilistic mortality forecasting with varying age-specific survival improvements. *Genus*, 73(1), 1.