A General Age-Specific Mortality Model with An Example Indexed by Child or Child/Adult Mortality

Yuxin Du

## Keywords

Mortality Model, Child Mortality, Singular Value Decomposition, Log-Quad Model

## Introduction

Many of the nations for the developing countries, especially in the continent of Africa, don’t have public registering and important statistical organizations which could be used functioning good enough to precisely account on either fertility data or mortality data. Most countries in Africa and almost one-third of countries all over the world need mortality models to infer a complete age table of mortality for population estimation and many other applications in demographic epidemiology. The child mortality estimation as well as the mortality in other age groups are all important, and because we have child mortality measures in nearly every country, not mortality in other age groups, therefore, it may be possible to estimate the mortality in other age groups with the information of child mortality. We defined mortality model for different age groups.

In this report, a mortality component model with singular value decomposition (sVD-Comp) was built and the relation between child mortality and mortality in other age groups from the Human Mortality Database were calibrated into the mortality component model. We validate the mortality component model through cross validation and also compare the results with the Log-Quad Model. The results in cross validation as well as in the comparison with Log-Quad Model showed the sVD-Comp mortality component model is useful to predict the mortality at other ages when we only have the child mortality information, also, the sVD-Comp mortality component model could be generally applied and used to develop other useful models.

## Methodology (Data and Model)

### Data

The data is from the Human Mortality Database (HMD). The project originated from the Department of demography at the University of California, Berkeley and the Max Planck Institute of population research in Rostock, Germany. The human mortality database provides detailed mortality and population data for 39 countries. The Human Mortality Database (HMD) (University of California, Berkeley and Max Planck Institute for Demographic Research, 2016) comprises thoroughly prepared, patterned and confirmed data on deaths and exposure from some developed countries. The statistics are combined and accessible in a varied format. The purpose of this report is to detention and describe as much inconsistency in age groups mortality as possible, and therefore we selected to practice the 1 × 1 HMD life tables for each gender. Those offer all columns of a typical lifetime table for single almanack years by every year of age from 0 to more than 110 years old. Every nation offers information for different past eras, and some nations are divided into more detailed sub groups. Then, a ‘nationwide population’ lifetime table is naturally provided which sums across the sub groups. Both the national and sub group people are involved in this report to make best use of the inconsistency in age groups mortality agendas in the total data set.

There are 4,486 lifetime tables for each gender, and there are 8,972 lifetime tables in total. The HMD data used in this report are downloaded from the file at <http://www.mortality.org/hmd/zip/all_hmd/hmd_statistics.zip>. Some of the 1 × 1 life tables were deleted due to no data (Belgium data between 1914 and 1918) or constant lifetime tables (males lifetime tables for Iceland) or similar implausible mortality rates (females lifetime tables for Iceland).

### Model

#### Singular Value Decomposition (SVD)

Singular value decomposition (SVD) is a widely used algorithm in the field of machine learning. It can be used not only in feature decomposition of dimension reduction algorithm, but also in recommendation system and natural language processing. It is the cornerstone of many machine learning algorithms.

A matrix could be decomposed to three components as follows:

where is the left singular vector matrix, is the right singular vector matrix, and the is the singular value diagonal matrix.

The left singular vector matrix and the right singular vector matrix are independent matrix and both of them have unit length. There are two useful formulas for the singular value decomposition that

and

where are the left singular vectors, are the right singular vectors, are the singular values, is the rankd of matrix , is the column of matrix .

#### SVD Component Model – ‘SVD-Comp’

SVD Component Mortality Model is a normal model framework for ASMR, and it is a one-year age model of lift tables for UNPD. The SVD-Comp Model combined unusual effects of HIV, battle and ordinary disasters through summary indicators and was rotated at older age groups and lower general mortality rate by smoothing, interpolation and/or extrapolation.

Suppose that a matrix is the mortality schedules for each gender, therefore, from the last part, we could decompose the matrix to three components that

Also, from equation , we have that

where is the set of , is the rank of matrix , and is the mortality schedules (Golub et al., 1987)

#### Calibration SVDs

The lifetime tables of the HMD are settled into two matrices of single-year, age-specific lifetime table probabilities of dying , one for each gender. is the number of age groups, which is 110, is the number of lifetime tables, which is 4,486, and . To make sure that all age clusters have roughly the similar effect when calculating the SVDs, every mortality schedule is offset from the by -10 and the offset is then added back to forecast mortality schedules.

#### Models for Predicting Weights

In regression analysis, if there are two or more independent variables, it is called multiple regression. In fact, a phenomenon is often associated with multiple factors. It is more effective and practical to predict or estimate dependent variables by the optimal combination of multiple independent variables. Therefore, multiple linear regression is more practical than univariate linear regression. Multiple linear regression model uses ordinary least squares (OLS) to estimate the parameters. The hypothesis of multiple linear regression model is: 1) the explanatory variable is not random or fixed, and there is no multicollinearity between X; 2) the random error has zero mean, homoscedasticity and nonsequential correlation; 3) the explanatory variable is not related to the random term; 4) the random term satisfies the normal distribution.

The final prediction models is

where and is the mortality schedules. Ordinary Least Square regression was used to estimate the coefficients for the above models.

#### Models for Adult Mortality

The models for the adult mortality are:

Again, the Ordinary Least Square regression was used to estimate the coefficients for the above models.

#### Models for Mortality in the First Year of Life

The models for mortality in the first year of life is expressed as follows:

Ordinary Least Square regression was used to estimate the coefficients for the above models.

#### Model Validation

Cross validation is a common method in machine learning to build models and verify model parameters. It is generally used to evaluate the performance of a machine learning model. More often, we also use cross validation for model selection. Cross validation, as the name implies, is the repeated use of data, the obtained sample data are segmented and combined into different training sets and test sets. The training set is used to train the model, and the test set is used to evaluate the quality of the model prediction. On this basis, different training sets and test sets can be obtained. A sample in a training set may become a sample in the test set next time, which is called “crossover”.

According to different segmentation methods, cross validation can be divided into the following three types: The first is simple cross validation, which is relatively simple compared with other cross validation methods. First of all, we randomly divide the sample data into two parts (for example, 70% of the training set and 30% of the test set), and then use the training set to train the model and verify the model and parameters on the test set. Then, we scramble the samples, re select the training set and test set, and continue to train the data and test the model. Finally, we choose the optimal model and parameters of the loss function.

The second is s-folder cross validation, which is often used. Different from the first method, s-fold cross validation first divides the data set D into s mutually exclusive subsets of the same size, and randomly selects S-1 as the training set and the remaining one as the test set. When this round is completed, the S-1 is randomly selected to train the data. After several rounds, the loss function is selected to evaluate the optimal model and parameters. Note that the stability and fidelity of cross validation evaluation results largely depend on s value.

The third is leave one out cross validation, which is a special case of the second case. In this case, s is equal to the number of samples n. thus, for n samples, n-1 samples are selected to train data, and one sample is reserved to verify the quality of model prediction. This method is mainly used when the sample size is very small, for example, for ordinary moderate problems, when n is less than 50, it is generally used to leave one cross validation.

Cross validation was used to test how the model performed and 25 random selected samples were used for prediction errors calculation for mortality schedules. In another method of cross validation, for sample fractions from 10%, 30%, 50%, 70% and 90%, 50 samples were randomly selected and used for modeling and the prediction errors were calculated and summarized.

#### Comparing with the Log-Quad Model

The Log-Quad model (Wilmoth et al., 2012) was compared with the SVD-comp mortality model. The prediction errors for both methods were calculated and compared.

## Results

### Data results

In Figure 1, we provided the mortality information in the HMD data as well as the prediction mortality from the SVD-Comp model. The Y axis is the logit for Sweden in the year of 1751 adn for France in the year in 1978, and the X axis is the as the independent variable only. The simple SVD-Comp model showed good prediction performance for older years and young people.

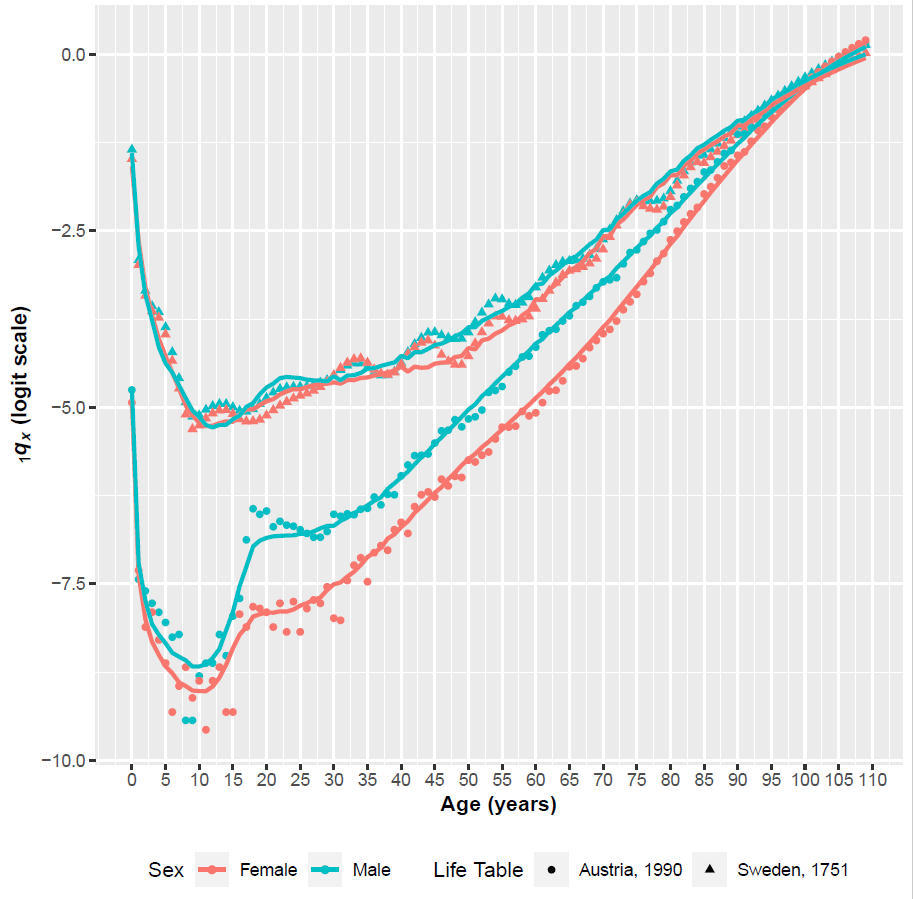


Figure 1: Example data set plotting and prediction

### SVD results

Figure 2 showed the sacled LSV from the SVD of the HMD mortality schedules.

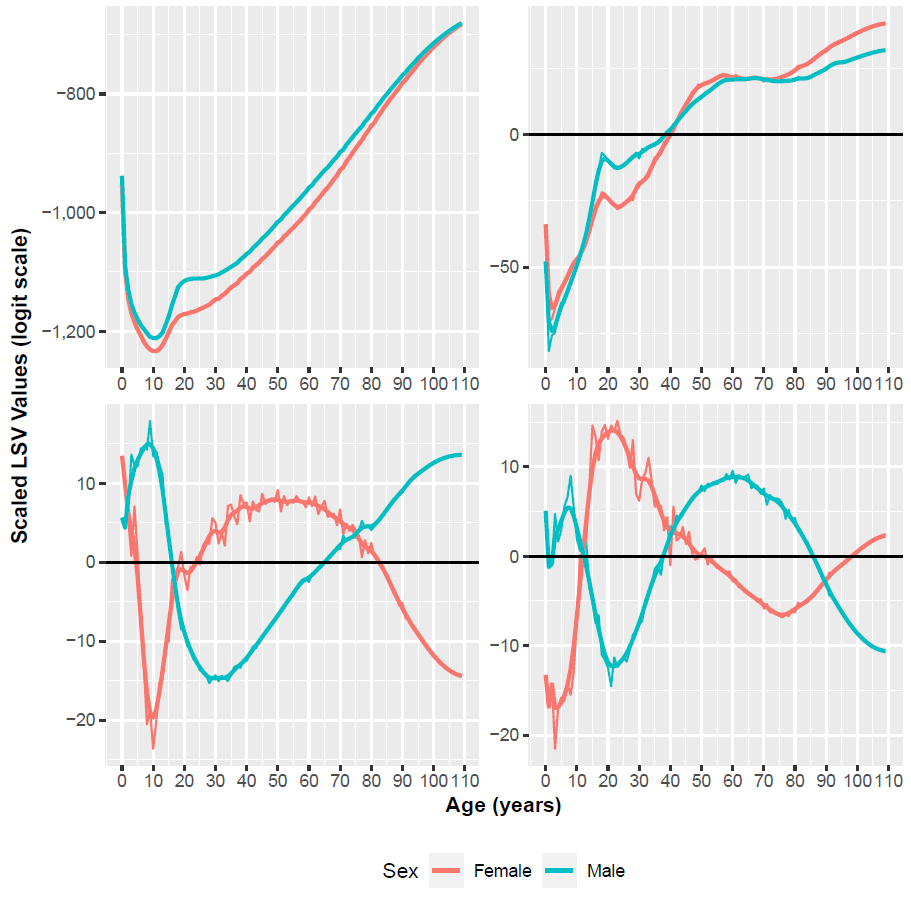


Figure 2: SVD Results

### Calibration relation results

Figure 3 to Figure 5 showed the true data and their predictions from the model in Equation (6), (7), and (8),as well as the estimted coefficients for the models. The plots indicated that there are quasilinear relationships between RSV Element Values and child mortality (), as in Figure 3 and Figure 4. Figure 5 is the prediction results for model Equation (8).

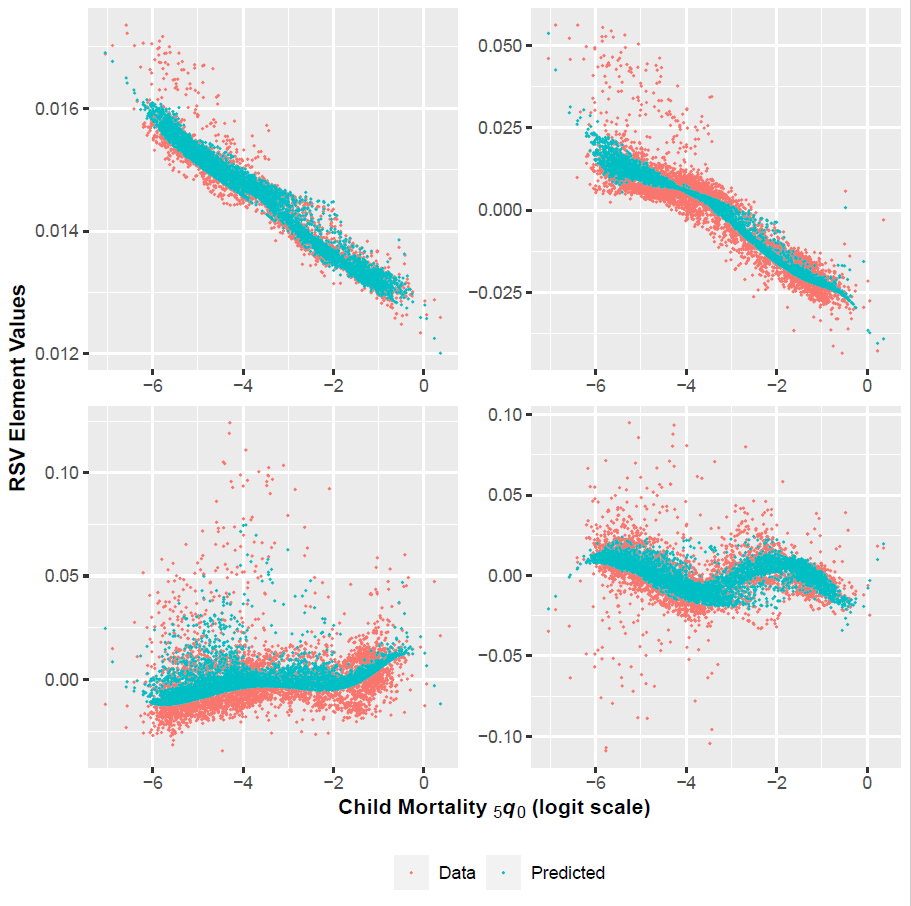


Figure 3: Right Singular Vector Element Values for Females

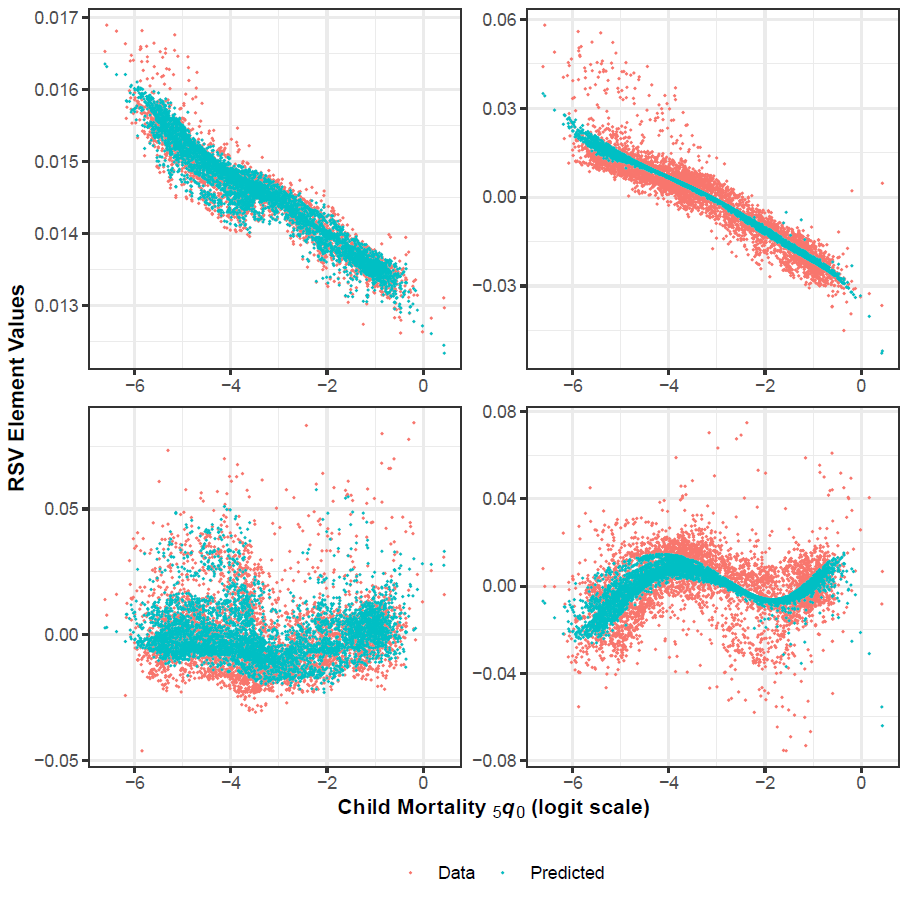


Figure 4: Right Singular Vector Element Values for Males

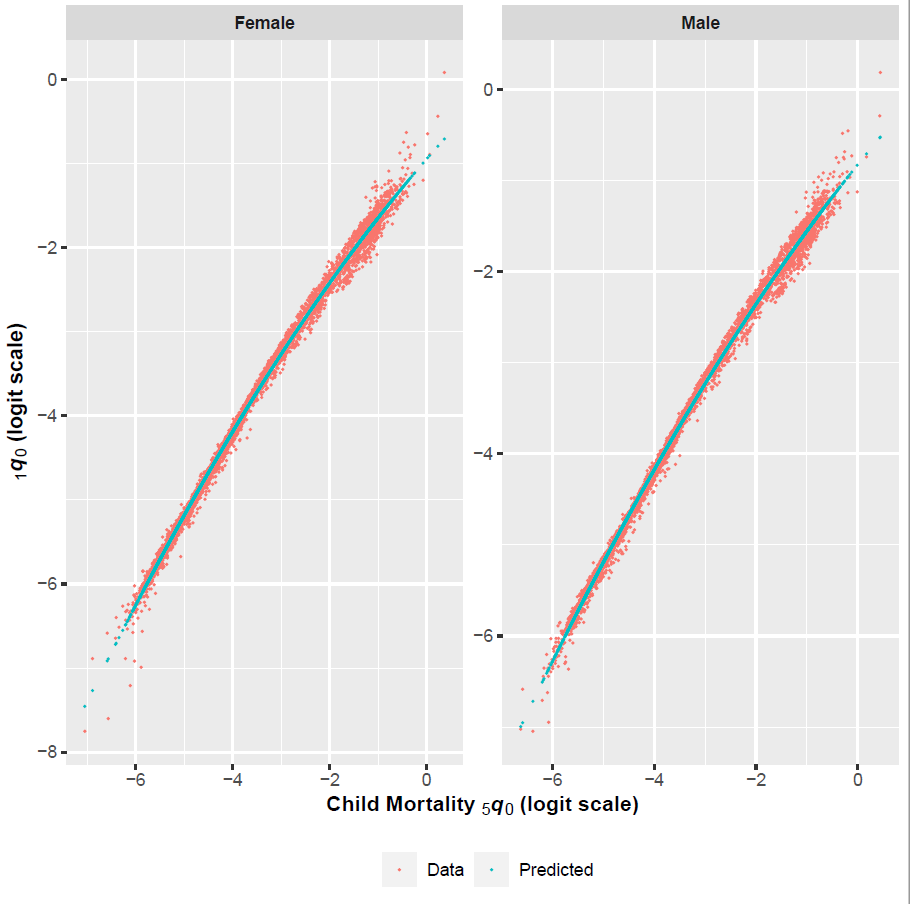


Figure 5: Age 0 Probability of Dying vs. Child Mortality.

### Cross Validation errors results

The cross-validation errors were plotted in Figure 6 and Figure 7. The median errors in Figure 6 for female and for male for each samples had the box plots with different sampel fraction from 10% to 90%. In Figure 7, it showed the interquartile range box plot of prediction errors by differrent sample fraction from 10% to 90%.

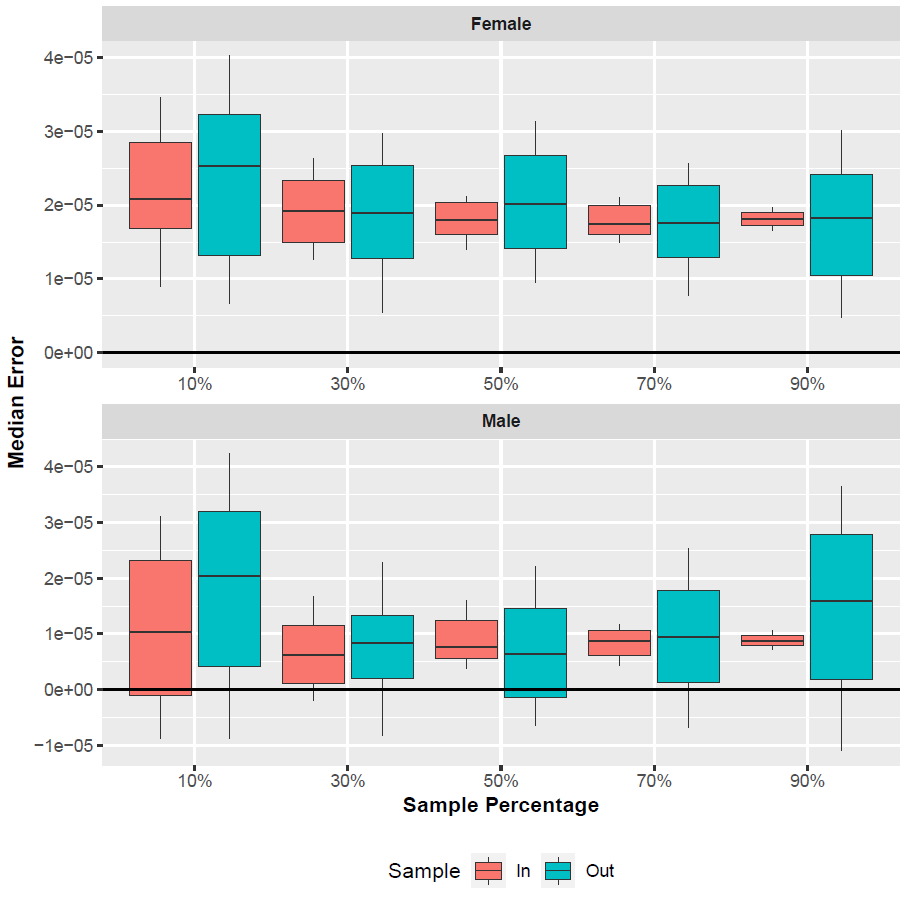


Figure 6: Median Prediction Error by Sample Fraction.

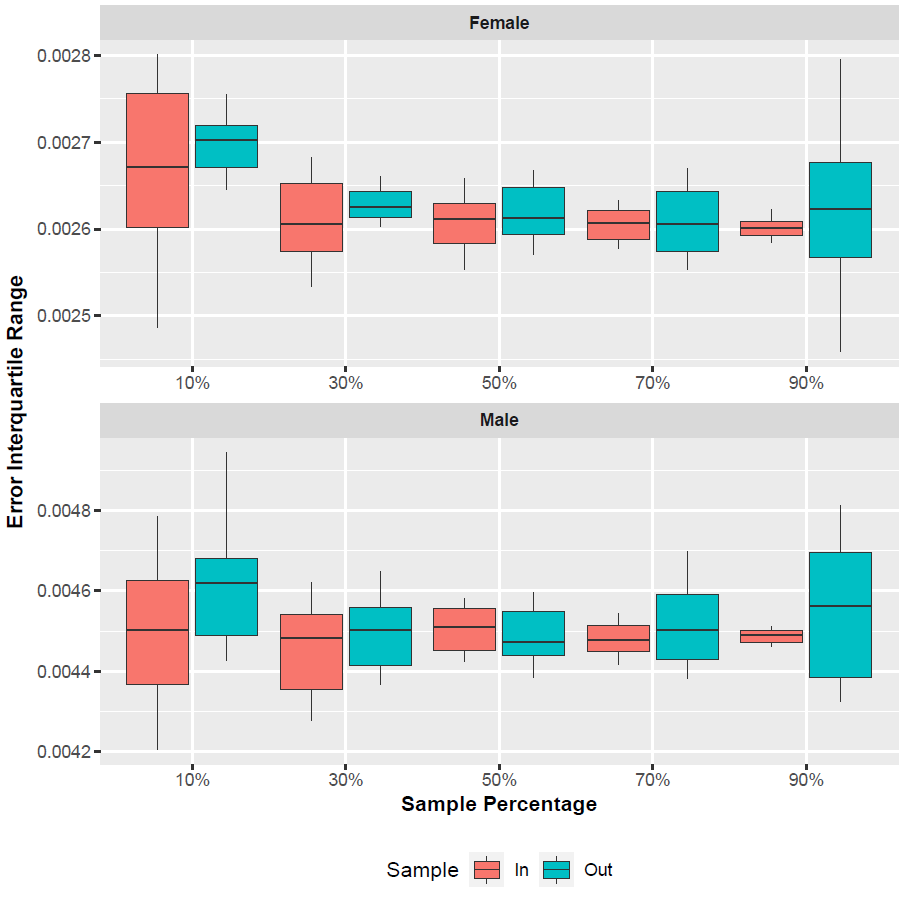


Figure 7: Interquartile Range of Prediction Error by Sample Fraction.

### Comparison results between Log-Quad model and SVD-comp mortality model

The comparison results between Log-Quad model and SVD-comp mortality model are summarized in Figure 8 and Figure 9. In Figure 8, the prediction errors for different age groups by SVD-Comp models and Log-Quad models were plotted and compared. In Figure 9, predicted at Three Levels of were plotted for different genders and different level of and it showed the level of increased the correlation between female mortality and male mortality differences.

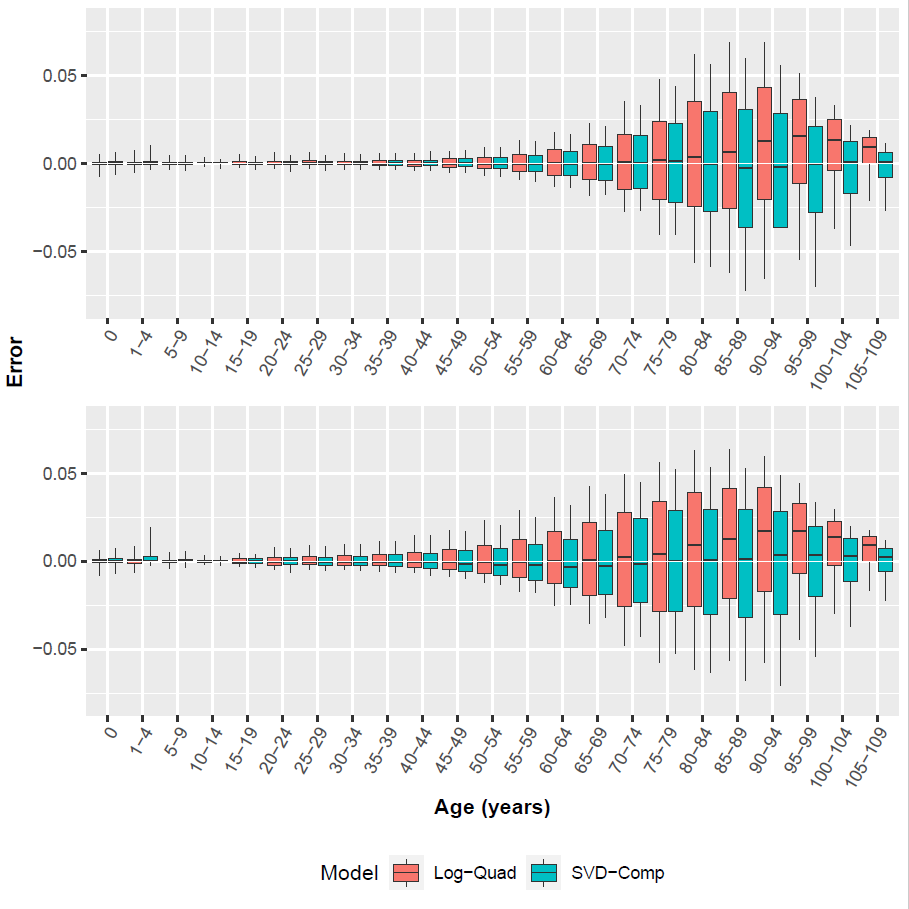


Figure 8: SVD-Comp and Log-Quad Prediction Errors.

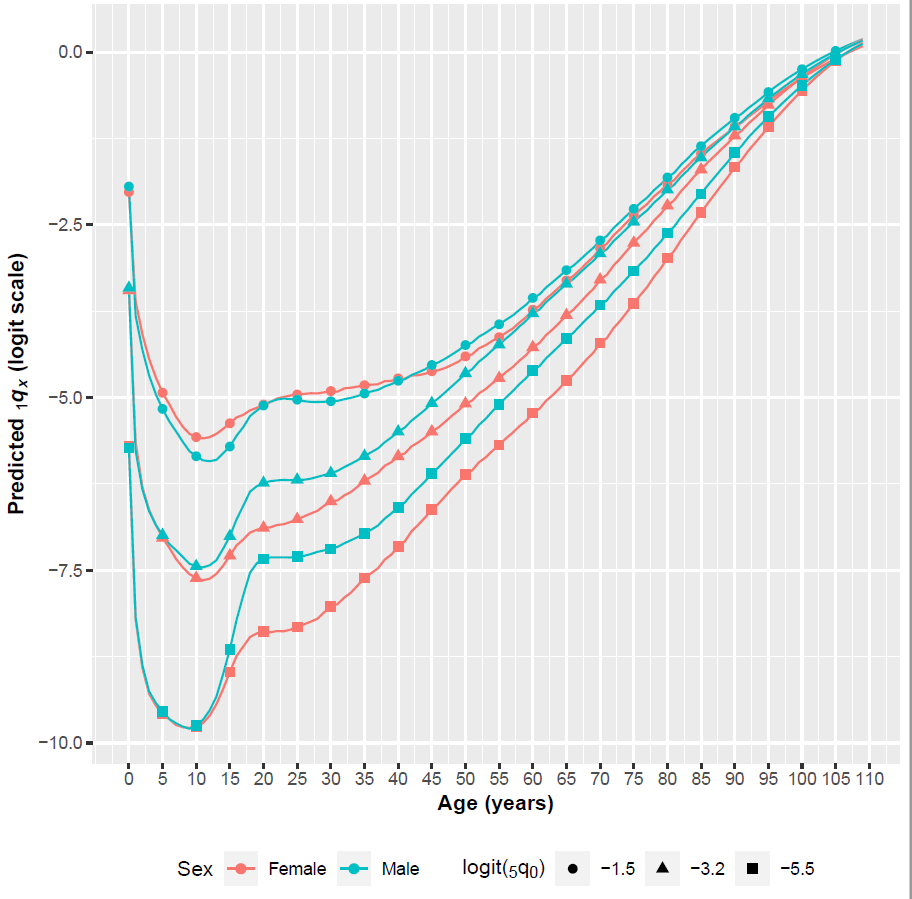


Figure 9: Predicted at Three Levels of .

## Discussion

Summary

In this analysis, a mortality component model based on singular value decomposition (sVD-Comp) was built and was compared with the Log-Quad Model. Also, the relationship between child mortality and mortality in other ages from the Human Mortality Database (HMD) were calibrated into the SVD-Comp model.

Conclusions

The results in cross validation as well as in the comparison with Log-Quad Model showed the sVD-Comp mortality component model is useful to predict the mortality at other ages when we only have the child mortality information.

Also, the sVD-Comp mortality component model could be generally applied and used to develop other useful models.

Weakness & Next Steps

In this analysis about SVD-Comp, we assume that the relationship of age groups had a constant components fromt he SVD, but in reality it may not be constant and it may vary. An empirical weights about the relationship could be applyed to get more accurate predictions about the mortality in other age groups.

Besides the model technical in this report, some other methods could also be used, such as the OLS regression as well as the Bayesian estimation could be used about the incomplete mortality with only child information and then predict the full mortality in all ages groups.

## References

Clark J. Samuel, 2019, ‘A General Age-Specific Mortality Model With an Example Indexed by Child Mortality or Both Child and Adult Mortality’, Demography, 56, p1131–1159

Golub, G. H., A. Hoffman, and G. W. Stewart (1987). A generalization of the eckart-young-mirsky matrix approximation theorem. Linear Algebra and Its Applications 88, 317–327.

University of California, Berkeley and Max Planck Institute for Demographic Research (Downloaded November 2016). Human Mortality Database. <http://www.mortality.org> or <http://www.humanmortality.de>.

Wilmoth, J., S. Zureick, V. Canudas-Romo, M. Inoue, and C. Sawyer (2012). A flexible twodimensional mortality model for use in indirect estimation. Population studies 66(1), 1–28.