RESEARCH REPORT SERIES
(Statistics #2014-03)

Forecasting Fertility and Mortality by Race/Ethnicity and Gender

Osbert Pang

Tucker McElroy

Center for Statistical Research & Methodology
Research and Methodology Directorate
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: July 8, 2014

Forecasting Fertility and Mortality by Race/Ethnicity and Gender

Osbert Pang and Tucker McElroy

Abstract

Forecasts of fertility and mortality rates are two of the components of population projections. Because the data consists of age-specific rates, the production of an age distribution curve leads to a high-dimensional problem, where there may be more ages in the data than there are observed years. This paper focuses on forecasting the total rates of fertility and mortality for a set of gender and race/ethnic subgroups, as well as the relative age-specific rates (or age distribution curves) for each. The forecasting of total rates is relatively straightforward, with comparisons done using time series models with and without a drift parameter. For age-specific rates, principal components analysis is used to reduce the dimensionality of the data. The procedure for carrying out the principal components as well as all time series modeling of the principal component series is discussed; included is some discussion of how to choose the number of principal components to retain. Results are produced using fertility and mortality data dating from 1989 to 2009. For total rates, it is determined that a model without drift produces more tenable forecasts in comparison to the occasionally implausible results from the model with drift. Some discussion of ways to improve the procedure for the future are provided.

1 Introduction

In order to obtain projections of future population, forecasts of fertility, mortality, and immigration are required, as those are the components of population change. These data are tabulated in the form of age-specific rates. This leads to a high-dimensional problem, because in producing forecasts of age-specific rates, there may be more ages than years of data. Principal components analysis proves to be a useful tool in reducing the dimensionality to a more manageable size. In this paper, we discuss the use of principal components analysis with respect to forecasting fertility and mortality rates. We treat the resulting principal components as time series in order to produce models and forecasts, and these forecasts are ultimately used to obtain forecasts of the corresponding rate data.

The use of principal components to forecast fertility and mortality rates is not a recent development. Bozik and Bell (1987) used principal components analysis on the covariance matrix of log-transformed relative fertility rates, where relative rates are just the age-specific rates of a particular year divided by the corresponding total fertility rate (TFR). Normalizing

the age-specific rates so that they sum to one lends itself to an interpretation of the relative fertility rate as an age distribution. In addition, they transformed the total fertility rate by simply taking the logarithm; this transformation eliminates the possibility of nonsensical negative TFR forecasts, since exponentiating any number will always result in a positive number. Ultimately, Bozik and Bell chose to fit a multivariate ARIMA model to the transformed TFR, together with the first four principal component series.

Similarly to Bozik and Bell (1987), Lee and Carter (1992) used principal components to help forecast U.S. mortality rates in 5-year age groups. Unlike Bozik and Bell (1987), however, Lee and Carter (1992) subtracted out age-specific means and used only a single principal component. They also constrained the number of deaths in their principal component approximation to be equal to the actual number of deaths seen in the data. Bell (1997) pointed out that this single principal component approach would have more approximation error compared to one that incorporated a greater number of principal components. To address this, he suggested that the single component used by Lee and Carter (1992) could be modified by forecasting all of the other principal components (excluding the first) as random walk models without drift, thereby keeping the forecasts constant at the last observed values. As the number of principal components used increases, however, this modification would have less of an impact.

As an alternative to principal components, curve fitting has also been used in an attempt to solve the issue of dimensionality. Bell (1997) compared curve-fitting approaches to a principal component approach for forecasting age-specific rates. He concluded that the curve-fitting approach produces a greater amount of approximation error than the principal component approach.

In the same vein as Lee and Carter (1992), McElroy and Bell (2004) used a single principal component approach in examining the age distribution curve for Hispanic employment immigrants data. Instead of using a log-transformed rate as was done previously, however, they proposed a generalized logistic transform, which has the benefit of producing forecasted rates that sum to one without requiring additional normalization at the end. They also included a step with a smoothing spline as well as a Bayesian modification to control the long-range movement of forecasts. With the Bayesian modification, upper and lower bounds can be set based on expert prior knowledge.

The procedure used here draws a bit from each of the aforementioned works. For age-specific rates, forecasts are computed for both the natural logarithm transform from Bozik and Bell (1987) and the generalized logistic transform proposed in McElroy and Bell (2004).

The principal components approach used in these earlier works is employed here, although we use univariate ARIMA (1, 1, 0) models instead of the multivariate model from Bozik and Bell (1987). This principal components approach does incorporate the use of the smoothing spline on age-specific means and the appropriate number of eigenvectors as was done in McElroy and Bell (2004). Modeling of the total fertility rate is done using the log of TFR minus one, which is coherent with work done for a previous set of Census Bureau fertility projections, as discussed by Thompson et al. (1989). For mortality rates, the sum of the age-specific rates is modeled using a standard log transform, and more than one principal component is retained. The forecasting of the remaining principal components proceeds as suggested by Bell (1997); that is, for all of the principal components that are not modeled using an ARIMA (1, 1, 0), the forecasts will remain constant at the last observed values. Lastly, there is no Bayesian attenuation as described in McElroy and Bell (2004), so there are limited means by which the procedure used here can control long-range movement.

Section 2 begins with discussion of the form of the data for both fertility and mortality. Then the segmentation of the data into disjoint sets is described, such that the fertility data is split by race/ethnic group and the mortality data is separated by the combination of race/ethnicity and gender. Section 3 elucidates how the total rates are handled in comparison to the age-specific rates. Section 4 describes the results for fertility rates, while Section 5 covers the same ground in regard to mortality rates. Section 6 considers the use of an age-adjusted approach to the mortality rates discussed in Section 5. Section 7 then summarizes the outcomes of the previous sections, as well as illuminating some possible areas that may warrant additional examination. All plots associated with Sections 4 through 6 can be found in Appendix A.

2 Data and Processing

The initial data set for fertility rates included variables for year, age, race/ethnic group, births, population of women, and age-specific fertility rate. The data set for mortality rates did not include similar variables for deaths or population, but did have a gender identifier. So, there were a total of five categories for age-specific fertility rates based solely on race/ethnicity (non-Hispanic whites, non-Hispanic blacks, American Indian & Alaska Natives, Asian & Pacific Islanders, and Hispanics), and ten for age-specific mortality rates using those same races/ethnicities for each gender.

The goal was to produce forecasts of two different pieces of information: the total rates, which were just the sum of the age-specific rates, and the relative age-specific rates, which

were the age-specific rates divided by their sum (thus guaranteeing that the relative rates sum to 1). For both items, the forecasts needed to be positive, so the use of a transformation like the natural logarithm or generalized logistic function was warranted. For the total rates, the transformations were not difficult to select. The sum of the age-specific mortality rates (denoted as the total mortality rate, even though such a quantity has a fairly nebulous interpretation) was subjected to a standard natural logarithm transformation. For the sum of age-specific fertility rates, or total fertility rate, the desire was to ensure forecasts greater than 1. This was achieved by the transformation $\log(\text{TFR} - 1)$.

While it was fairly straightforward to transform the sum of the age-specific rates, the relative rates proved to be more complicated. Both transforms considered involved logarithms, and logarithms do not accept zeroes. For the mortality rate, this was a minor concern because there were only isolated cases where the rate was zero. Hence, for mortality rate, any zeroes were simply replaced by a very small value, which was arbitrarily chosen as 10^{-8}. While the chosen number may have been smaller than necessary, the results suggested that the models were mostly unaffected by this particular selection.

On the other hand, the fertility rate had more occurrences of zero births (and thus more zeroes for age-specific fertility rates). An examination of the ages which experienced zero births for any race/ethnic group revealed that they were all concentrated in the lower and upper age ranges. That is, the only ages for which there were no births in one of the years fell into the ranges 10 through 13 and 47 through 54. We chose to omit the low end, but felt the high end should be retained. As a result, in order to eliminate the zeroes in the upper end, we performed a linear regression with no intercept for the number of births for ages 46 through 54 against the years remaining to age 55 and then replaced the number of births for ages 47 through 54 by the corresponding fitted values from the regression. For any race/ethnic group, if t denotes the year, and i denotes the age, let y_{it} represent the number of births in year t for women of age i; then the linear model takes the form $y_{it} = \alpha_t(55 - i) + \text{error}$, where the subscript t in α_t reflects how it varies depending on year (and race/ethnic group, although that is implicit in the formulation as described). Using the linear model, the estimate for α_t becomes

$$\widehat{\alpha}_t = \frac{\sum_{i=46}^{54} y_{it}(55 - i)}{\sum_{i=46}^{54}(55 - i)^2}.$$

This gives an interpolated value of births in year t of $\widehat{y}_{it} = \widehat{\alpha}_t(55-i)$, where $i \in \{47, 48, \ldots, 53, 54\}$. Even though the number of births for women of age 46 is used in estimating the linear model, its value is left unchanged. In part, this helps maintain a level of smoothness in moving from births at age 46 to births at age 47, but its inclusion also allows us to handle the situation

where, in a given year t, the number of births for women between the ages of 47 and 54 were all zero for some race/ethnic group (this would have been an issue with American Indian & Alaska Native). This choice ensures that the fitted values of the regression are always non-zero. These interpolated births are then used as the numerator with the original value of population as the denominator to obtain modified age-specific fertility rates at the upper range. The values for age-specific fertility rate between ages 14 and 46 are left intact by this procedure.

The procedure as described above is flawed in that it does not maintain total births; the use of fitted estimates that are always positive in an age range that has the possibility of zeroes forces this model-based value for total births to be larger than the actual observed value. However, since the denominator for the age-specific fertility rates is left unchanged, we felt that the increase in number of births would not amount to a substantive increase in the corresponding age-specific fertility rates. The total fertility rates incorporating the estimated births in the upper region experience an increase of no more than 0.002 in any instance, although we only use these inflated total fertility rates in converting the age-specific fertility rates from this procedure into relative rates.

3 Procedure for Total and Age-specific Rates

Section 2 includes a brief discussion about what transformation to use for the sum of the age-specific rates. For fertility, we argue that $\log(\text{TFR} - 1)$ is a good choice, in that it ensures forecasts that are strictly greater than 1. For mortality, we argue that $\log(\text{TMR})$ serves the purpose well enough without the need to subtract a lower limit beforehand. The transformed series are then fit to an ARIMA (1, 1, 0) time series model that includes parameters for both drift and the autoregressive coefficient. Forecasts can be produced using a variation of this model that excludes the drift parameter.

In contrast to the fairly straightforward approach to modeling total rates, the procedure for age-specific rates is more complicated. The transformations used for relative age-specific rates have been detailed in Section 2, and as such, need not be repeated here. Regardless of the transformation that is used for the relative rates, the method by which the forecasts are produced proceeds the same. Only the reversing of the transformation at the end requires different handling for the various transformations. The first step is to compute the age-specific mean curve (the average taken over time). Bell (1997) previously commented that subtracting out the age-specific means improves a low-dimensional principal component approximation, although it becomes less useful as the number of principal components used increases. Given

our preference for smooth forecasts of age distribution curves, combined with the fact that the age-specific means are added onto the forecasts produced using the mean-corrected data, we choose to run the age-specific means through a smoothing spline to help ensure that our ultimate forecasts are indeed smooth. Thus, this smoothed age-specific mean curve will be subtracted from the transformed data, and we perform a principal components analysis (PCA) on this mean-corrected data.

The goal behind PCA is dimension reduction: if there are K ages in the transformed data, one wishes to find J principal components that provide the best J-dimensional approximation of the (transformed) data, where ideally $J < K$. In general form, if X_t is a $K \times 1$ vector representing the transformed rates observed at time t, then X_t could be alternately expressed as

$$X_t = \Lambda\beta_t + \epsilon_t.$$

In the above expression, Λ would be a $K \times J$ matrix produced using PCA; the columns of Λ would be the eigenvectors corresponding to the J largest eigenvalues of the sum of squares and cross products matrix of the data. This expression looks remarkably similar to the matrix representation for ordinary least squares regression, in which case β_t would just be a $J \times 1$ vector of coefficients obtained by regressing X_t against Λ, but the β_t might also be viewed as loadings for the principal components. Also, ϵ_t has the usual regression interpretation of being the collection of errors.

It is possible to leave $J = K$, in which case the J principal components form an orthonormal basis that recreates the mean-corrected data with no error. However, one can typically obtain an approximation that accounts for at least 95% of the variation in the (mean-corrected) data with a far smaller number of principal components. In this specific application, where each race/ethnic group is handled separately, for a given year t with mean-corrected data $\gamma_t - \overline{\gamma}$, the approximation would be

$$\widehat{\gamma}_t = \overline{\gamma} + \Lambda\widehat{\beta}_t,$$

and hence

$$\widehat{\beta}_t = (\Lambda'\Lambda)^{-1}\Lambda'(\gamma_t - \overline{\gamma}).$$

Our procedure (see below) uses all K principal components, although the first J (chosen to explain 95% of the variation) of them have a different forecasting model from the rest. Thus, the Λ matrix is $K \times K$; let $\widehat{\beta} = (\widehat{\beta}_1, \widehat{\beta}_2, \ldots, \widehat{\beta}_n)$ be a $K \times n$ matrix where the columns are the $\widehat{\beta}_t$s mentioned earlier, and where n represents the number of years in the data (in this application, both mortality and fertility have $n = 21$). Each row of $\widehat{\beta}$ then corresponds

to a separate principal component series, so the first row of $\widehat{\beta}$ is the estimated time series associated with the first principal component, the second row is the estimated time series associated with the second principal component, and so on.

Using fewer principal components is generally desired, but the cost in doing so is a greater amount of approximation error. Bell (1997) proposed emulating the "bias adjustment" procedure of Thompson et al. (1989) as a means for reducing the approximation error. Thus, we retain all K principal components, but only the first J principal components will be modeled using time series methods; we forecast the remaining $K - J$ principal components using a simple random walk model, which amounts to using the last observed value for all forecasts. Hence, for each of the first J rows of $\widehat{\beta}$, a univariate ARIMA (1, 1, 0) model is estimated and forecasted over a horizon of h years, and the full forecast matrix $\widehat{\beta}^{(h)}$ is of dimension $K \times h$, where column l represents the forecast for the K principal component series at time $n + l$. Maintaining this notation, for a given year $n + l$, the forecast of γ_{n+l} ends up being

$$\widehat{\gamma}_{n+l} = \overline{\gamma} + \Lambda \widehat{\beta}_{n+l},$$

where $\overline{\gamma}$ is the smoothed mean curve of the transformed data, and Λ and $\widehat{\beta}_{n+l}$ are as defined previously.

It is preferable to obtain smooth forecasts for our age distribution curves. However, using Λ as is in forecasting $\widehat{\gamma}$ may not yield smoothness. This leads us to construct $\widetilde{\Lambda}$, which is a $K \times K$ matrix resulting from using a smoothing spline on each of the columns of Λ. We use the same smoothing parameter for each of the columns of Λ, although this smoothing parameter is not necessarily the same as the one used for the age-specific mean curve. Ultimately, we replace the Λ in constructing $\widehat{\gamma}$ with $\widetilde{\Lambda}$, and this helps produce smooth forecasts.

There is an apparent inconsistency in using the unsmoothed Λ to construct $\widehat{\beta}_t$, but the smoothed $\widetilde{\Lambda}$ to produce $\widehat{\gamma}$. It would seem logical to use one of Λ and $\widetilde{\Lambda}$ throughout the process, but not both. However, we found that using only Λ in the process of constructing $\widehat{\gamma}$ resulted in forecasted age distribution curves that became increasingly jagged as we moved into the future. To remedy this defect, we instead substituted $\widetilde{\Lambda}$ in for Λ. A few issues appeared in doing this; first, the value of the smoothing parameter had a profound impact on the scale of the values for individual rows. In particular, as more smoothing was performed, the values grew progressively larger in magnitude. Another problem with the sole use of $\widetilde{\Lambda}$ was that we saw frequent appearances of secondary humps and spikes in the forecasted age distribution curves, so that we might have a forecast 50 years ahead that had 3 distinct peaks. Overall, the results appeared smoothest when we used Λ to construct $\widehat{\beta}_t$s and $\widetilde{\Lambda}$ to construct $\widehat{\gamma}$.

Letting $\widehat{\gamma}^{(h)} = (\widehat{\gamma}_{n+1}, \widehat{\gamma}_{n+2}, \ldots, \widehat{\gamma}_{n+h})$ represent the forecast matrix, column l of this matrix is the forecast of the transformed relative rates for time $n+l$. Thus, reversing the original transformation and then re-normalizing each column would yield a forecasted age distribution curve for time $n+l$. For the natural logarithm transformation, reversing the transformation simply entails exponentiating the $\widehat{\gamma}^{(h)}$ and then dividing the forecasts in each year (each column) by the corresponding sum. For the generalized logistic transform, the process is described in McElroy and Bell (2004), but it shall be repeated here. For every age other than a reference age κ, the forecast of $\widehat{\gamma}_{n+l}$ is exponentiated and then each of the entries is divided by 1 plus their overall sum to obtain the relative age-specific rate $\widehat{r}_{i,n+l}$ (for age i):

$$\widehat{r}_{i,n+l} = \frac{e^{\widehat{\gamma}_{i,n+l}}}{1 + \sum_{k \neq \kappa} e^{\widehat{\gamma}_{k,n+l}}}, \qquad i \neq \kappa.$$

For the reference age κ at time $n+l$, the forecast is just 1 minus the sum of the other $\widehat{r}_{i,n+l}$, or alternatively, one could calculate it using

$$\widehat{r}_{\kappa,n+l} = \frac{1}{1 + \sum_{k \neq \kappa} e^{\widehat{\gamma}_{k,n+l}}},$$

where k in both equations given is allowed to range over the ages in the data. Note that for the generalized logistic transform, the transformed data actually only has $K-1$ ages, as compared to the K when using the natural logarithm. However, no additional re-normalizing is required because the forecasted relative rates under the generalized logistic transform automatically sum to 1.

To summarize, forecasting total rates is a straightforward procedure in which the only significant decision is whether to include a drift parameter in the designated time series model. A different story exists for forecasting relative age-specific rates (and thus, the age distribution curves). The procedure for producing forecasts of the age distribution curves can be laid out as follows:

1. Convert age-specific rates to relative rates after first eliminating all occurrences of zeroes in the rates.

2. Take a suitable transformation of the relative rates; denote the vector of transformed relative rates for year t as γ_t.

3. Calculate the age-specific mean curve for the transformed relative rates and then use a smoothing spline to obtain a smoothed version labeled $\overline{\gamma}$.

4. Subtract out the smoothed age-specific means from the transformed rates $(\gamma_t - \overline{\gamma})$ and determine the resulting sum of squares and cross-products matrix $S = \sum_t (\gamma_t - \overline{\gamma})(\gamma_t - \overline{\gamma})'$.

5. Determine the eigenvalues and eigenvectors of S. Let Λ be the $K \times K$ matrix of eigenvectors and keep all columns.

6. Calculate the $K \times 1$ vector β_t using $\widehat{\beta}_t = (\Lambda'\Lambda)^{-1}\Lambda'\gamma_t$, and collect the columns into the matrix $\widehat{\beta} = (\widehat{\beta}_1, \widehat{\beta}_2, \ldots, \widehat{\beta}_n)$, so that each row of $\widehat{\beta}$ corresponds to a time series for a principal component.

7. Choose J, which can be less than K. Fit a univariate ARIMA (1, 1, 0) with drift to each of the first J rows of $\widehat{\beta}$ and for a pre-specified value of h, use the estimated model to forecast all steps up to and including h. For the remaining $K - J$ rows of $\widehat{\beta}$, produce the forecasts by keeping the series constant at the observed value for time n. Label the matrix formed by collecting these forecasts of the $\widehat{\beta}$ matrix as $\widehat{\beta}^{(h)}$.

8. Use another smoothing spline on each of the columns of Λ and call the smoothed matrix $\widetilde{\Lambda}$. Then the matrix $\widehat{\gamma}^{(h)}$ containing all h-step ahead forecasts of the transformed relative rates is $\widehat{\gamma}^{(h)} = \overline{\gamma}\ \mathbf{1}_h' + \widetilde{\Lambda}\widehat{\beta}^{(h)}$, where $\mathbf{1}_h$ is a $h \times 1$ column vector containing all 1s.

9. Reverse the original transformation of $\widehat{\gamma}^{(h)}$ and re-normalize, if necessary, to obtain the matrix of relative rates $\widehat{r}^{(h)}$, whose columns represent each of the h-step ahead forecasted age distribution curves.

The procedure is fairly straightforward to encode, and our implementation in R runs quickly.

4 Results for Fertility Rates

4.1 Total Fertility Rate

We modeled log(TFR - 1) for each of the 5 race/ethnic groups using an ARIMA (1, 1, 0) model. We initially included a drift parameter in these models to help account for expected declines in total fertility rate over time. Subsequently, we produced alternate models that did not include a drift parameter. While the models without drift were somewhat unrealistic in their tendency to produce forecasts that leveled off very quickly, we considered this preferable to the substantial declines in total fertility rates that were predicted by the models with a drift parameter.

Figures 1 through 5 show the results for the 5 race/ethnic groups. In each plot, the solid black line represents the observed total fertility rates between the years 1989 and 2009, while the dashed red line indicates forecasts using the $(1, 1, 0)$ with drift and the solid blue line shows forecasts using the $(1, 1, 0)$ without drift; the forecasts are plotted for the years 2010 through 2069, so there are 60 years of forecasts provided. For non-Hispanic whites, as seen in Figure 1, the history of total fertility rate seems to fluctuate pretty wildly, but the scale on the vertical axis reveals that the range of movement has only been between about 1.77 and 1.91 during the 21 years of data, and going 60 years out results in a difference of less than 0.05 in the total fertility rate produced by the 2 competing forecasts. It would appear that the drift parameter for the model estimated for non-Hispanic white women does not yield substantial variation between the 2 different models. Also, the model without a drift parameter seems to hit a stable level within the first 2 or 3 years of forecasts, so the drift-less model levels off rapidly.

For non-Hispanic blacks (Figure 2), a different picture is obtained. The total fertility rate here is negatively trending, with a peak of about 2.5 occurring around 1990 and going down to approximately 2.0 by 2009. The range of movement is larger than it was for non-Hispanic whites, and the forecast for the model with drift reflects this. There is a noticeable curvature to the forecasts from the model with drift that was not apparent in the corresponding curve for non-Hispanic whites, and the difference over the 60 years of forecasts results in a difference exceeding 0.7 between the 2 different model specifications. The magnitude of the difference between the competing forecasts is not the only noticeable feature; looking closely at the curve for the model with drift, the implication is that in 60 years, the total fertility rate for non-Hispanic black women would be below 1.3, which would be roughly half of what their observed peak in 1990 was. It seems implausible that the total fertility rate for non-Hispanic blacks would fall so drastically over the forecast horizon. One last point to make is that the model without drift once again produces forecasts that level out within the first few years of the forecast horizon, which seems unlikely, but at least the plateau occurs at what seems to be a reasonable value for total fertility rate.

For American Indian/Alaska natives (abbreviated AIAN), the behavior of both history and forecasts in general is similar to that for non-Hispanic blacks, but the range of movement is more limited. The peak of total fertility rate for AIAN is around 2.4 in 1989, but at its lowest, it is still above 2.1. It is also apparent that the forecasts from the drift model for AIAN exhibit less curvature than the ones from the corresponding model for non-Hispanic blacks. This is a consequence of the drift parameter from the AIAN model being closer to 0

than the drift parameter of the ARIMA model for the non-Hispanic blacks. The difference between the model with drift and the model without over the 60 years of forecasts for AIAN seems to be no greater than 0.6 by the year 2069, as the forecast for the model with drift approaches 1.6 whereas the drift-less variant has a forecast around 2.2 for the total fertility rate. The model with drift continues to produce long-range forecasts that are not credible.

For Asian/Pacific islanders (abbreviated API) and Hispanics, the story is the same: they too have negatively-trending histories of total fertility rate, and they have a fairly pronounced curve in the long-range forecasts produced by the ARIMA (1, 1, 0) model with drift. While one might guess based on historical trends that the forecasts of total fertility rate among API would be consistently lower than the forecasts of total fertility rate among Hispanics, the forecasts for the model with drift for both API and Hispanics drop to around 1.3 by the end of the 60-year forecast horizon. This is rather surprising considering the peak for total fertility rate among API in our data is approximately 1 less than the peak value for total fertility rate among Hispanics. The more pronounced curvature for the forecasts of the model with drift are explained by the drift and autoregressive parameters for Hispanics being much larger in magnitude than those for API. One distinct feature of the plot for Hispanics is that the model without drift takes more time to level out compared to the rest of the race/ethnic groups, and it experiences a much steeper drop at the beginning of the forecast horizon for the model without drift.

Based on the visual evidence of Figures 1 through 5, if the same model must be chosen for all of the 5 race/ethnic groups, it would be hard to favor the model with drift over the one without drift. A drawback of the models without drift is that it takes very little time for the forecasts of total fertility rate produced under these models to stabilize. For long-range forecasts, it may prove beneficial that the models without drift give very stable results, however. On the other hand, while the models with drift allow for more movement over the course of the forecast horizon, there is no mechanism to control the amount of change in the total fertility rate, so the long-range forecasts of 3 of the race/ethnic groups all end up around 1.3, with 2 of those 3 being race/ethnic groups that have traditionally seen higher fertility rates. In addition, historical evidence would indicate the improbability of forecasts of total fertility rate for the non-white races/ethnicities falling below the level of whites, and yet the drift models indicate that the long-range forecasts of total fertility rate for whites would be the highest at the end of the 60-year forecast horizon. As a result, we found it difficult to accept the results of the drift models and thus chose to use the models without drift in the rest of this work involving fertility rates.

4.2 Relative Age-specific Fertility Rates

The general procedure for modeling relative age-specific fertility rates has been detailed in Section 3. There, the number of years of the data was generically referred to as n and the number of ages has been labeled as K. The data provided, as described above, contains the fertility rates for the years 1989 through 2009, thus giving $n = 21$. Similarly, although the original data covered fertility for females between the ages of 10 and 54, we decided to omit the ages 10 through 13. For practical reasons, we felt that the number of births that were likely to occur in that age range were not likely to form a substantial portion of fertility in the U.S. at any point in the near or distant future. As a result, there are a total of 41 ages to work with, so the K for the natural logarithm transform is $K = 41$, but for the generalized logistic transform, we only work with 40 ages.

The goal here is to produce an age distribution curve, so that one might answer: what percentage of the total fertility rate for a given race/ethnic group is accounted for by women of a particular age? So, the processing described earlier divides all age-specific fertility rate (or ASFR for short) values by their associated TFR, which yields series of relative age-specific fertility rates. These relative rates can be viewed as probabilities, and hence the curve of relative rates against age for a given year can be treated as a distribution curve of fertility rate. To ensure positivity in the forecasts of the relative rates, two transformations were considered: the first was the natural logarithm, which is only defined for positive values, and the second was the generalized logistic transform, which is a variation of the natural logarithm transform (it modifies the natural log transform by subtracting off the natural log of the rate for a reference age). One additional choice when using the generalized logistic transform came from the choice for reference age. The work here used the relative rates for the last age (54), even though those specific rates have been produced via the linear interpolation method described above. Originally, different ages were considered, as the initial belief was that an age with more stability in its ASFR would lead to better results. The idea was eventually abandoned when it became clear that forecasts for a reference age in the middle did not mesh with the forecasts for the neighboring ages.

For the principal components approach, there was some uncertainty about whether to use the same number of principal components for every race/ethnic group regardless of transformation or whether to use the bare minimum required to account for 95% of the total variation. Ultimately, for consistency, the number of principal components selected was equal to the maximum number of principal components required by any race/ethnic group (regardless of transformation) to achieve that 95% threshold. This number turned out to be $J = 8$. Each of

these 8 principal components is modeled univariately using an ARIMA (1, 1, 0) model with a drift parameter included. One could exclude the drift parameter, but the resulting forecasts for age distribution curves would result in very little deviation from the actual curve for the last year of the data. One could also choose to perform a multivariate time series model for the 8 principal components, but modeling 8 components simultaneously could prove awkward, and handling them univariately seems to work well enough. The remaining principal components that are not forecast using the ARIMA (1, 1, 0) model simply have forecasts that remain constant at the last observed value from 2009. This is another precautionary step to adjust for potential bias in the forecasts, when combined with the initial subtracting out of (smoothed) age-specific means for the transformed data.

Figures 6 through 10 show some results for the individual race/ethnic groups. The top plot in each figure displays the results using a natural logarithm transform, while the bottom plot shows the same results using a generalized logistic transform with age 54 fertility rates serving as the baseline comparison. The solid black line gives the observed age distribution curve from 2009, while the dashed red, green, blue, and gray lines give the projected age distribution curves for the years 2010, 2020, 2040, and 2060, respectively. These 4 years correspond to a 1-year forecast, an 11-year forecast, a 31-year forecast, and a 51-year forecast, and they may be viewed as forecasts for the very short term (2010), short-to-mid range (2010), mid-range (2030), and long-range (2060). Looking at the 5 plots for general behavior, it appears that there are only minor differences between the forecasts done using the natural logarithm compared to the ones done using the generalized logistic transform. For the most part, the differences that are visible seem to stem from the natural log shifting mass to the right at a slightly faster pace, but ultimately, either transform would be acceptable.

On the whole, the plots share many characteristics, although there are certainly features that appear to be specific to certain race/ethnic groups. One takeaway from the plots is that the 2010 forecast tends to mirror the actual 2009 curve. This is a positive sign, as fertility behavior seems unlikely to swing wildly from one year to the next, and it would be troublesome if the first forecast was vastly different from the last observation. The plots do appear to veer off fairly quickly from the 2009 curve, however; the forecast for 2020 shows a fairly noticeable gap between what was the forecast a decade earlier, and the shifts for 2040 and 2060 are even more pronounced. There is a common thread that all of the forecasts appear to be shifting toward a peak around age 34 or 35, which seems to be too large a change. The general uncertainty inherent in a really long-range forecast helps to explain this, but the result is still unsatisfying. One idea that could counter this problem with the

long-range forecasts is to stop after some number of forecasts and then extend that last accepted forecast to all remaining years of the forecast horizon. This possible solution would introduce some subjectivity with the choice for stopping time, however, and that may be just as unsettling as leaving the results as they are.

While all of the race/ethnic groups generally have plots that shift the mass and the peak to older ages, there are characteristics that are peculiar for certain races/ethnicities. The non-Hispanic white women have a larger occurrence of childbirths among older women, and the forecasts end up accentuating this feature by producing a curve over the mid-to-late 40s and early 50s that encompasses greater mass. Non-Hispanic blacks, American Indian & Alaska natives, and Hispanics all start off with peaks in their age distribution curves in the late teens or early 20s, so the general shift towards a peak in the mid 30s for these races/ethnicities has a strong ripple effect on the rest of the curve. For black females, the mass of the curve does shift to the right, but it also becomes less concentrated, and the peak does not concentrate in the long-run. On the other hand, the curves for Hispanic females also shift mass to the right, but there is a definite pronounced peak in that same long-run window. For the Asian & Pacific islanders, the age distribution curve begins with a peak around 30, and a more tightly packed mass than the other races/ethnicities. This feature becomes magnified over time, as the peak exceeds 0.1 in the mid 30s, compared to a peak in the 0.06 to 0.07 range for the other race/ethnic groups. This increased density around the peak may end up being hard to believe, however, and does lend some support to the proposal of truncating forecasts after some arbitrary time horizon.

5 Results for Mortality Rates

5.1 Total Mortality Rate

The total fertility rate is a fairly simple quantity to interpret: for a given race/ethnic group and year, the TFR represents the average number of children that would be born to a woman in the course of her life if she matches the ASFR for the specified year. The total mortality rate does not have a corresponding direct interpretation, although its role in projecting age-specific mortality rates is identical to that of TFR in projecting age-specific fertility rates. For total mortality rate, the natural logarithm transform is sufficient for the modeling done in this application. Staying consistent, the model of choice is again the ARIMA (1, 1, 0). Keeping in line with what was considered for TFR, forecasts were produced for all combinations of gender and race/ethnicity using a full model including drift and autoregressive parameters

as well as restricted models that omitted the drift parameter.

The results for the total mortality rate forecasts are presented in Figures 11 through 15. In each figure, the top plot gives the results for males, while the bottom shows the corresponding results for females. Looking at the history portion of the plots (the solid black lines stretching covering the range 1989 through 2009), given that males generally have shorter life expectancies than females, one would expect that the sum of age-specific mortality rates is greater for males than for females, and the plots do indeed support this belief. In addition, the sum of the mortality rates has generally been decreasing over the observed time period, with the major exception being American Indian & Alaska native females, whose TMR exhibits a positive trend; this seems to indicate an increase in the frequency of deaths, but it is unclear what the broader implications of this phenomenon are. A separate odd feature that is shared by multiple subgroups of gender and race/ethnicity is a sizable decrease in the sum going from the year 1999 to 2000. While the non-Hispanic whites and the Asian & Pacific islander females do not experience this, it is particularly pronounced in the case of males of the 4 non-white groups. While this may just be a consequence of the change in the 2000 Census for the question about race and ethnicity, it contributes to the difficulty in finding a reasonable interpretation for what the total mortality rate is.

For forecasts the dashed red line provides the outcomes associated with the ARIMA (1, 1, 0) that includes a drift parameter, and the solid blue line does the same for the model without a drift parameter. As was the case for TFR, from looking at the plots, the less objectionable long-range forecasts for TMR seem to have been produced using the models that excluded a drift parameter. One flaw with the model that includes a drift parameter appears to be its upward long-range forecasts for the sum of age-specific mortality rates among American Indian & Alaska native females and Hispanic females; it is somewhat counterintuitive that the forecasts would be increasing over time, especially considering how life expectancy has progressively increased over time. Another issue with the use of the full model is how much more rapidly the forecasts change for the race/ethnic groups that see more volatile movements in their history. For the non-Hispanic whites and Asian & Pacific islanders, the past movements of the sum of age-specific mortality rates covers a range of no greater than 0.8 units, but for the other races/ethnicities, the range must be close to double that number. This leads to some odd results: non-Hispanic black and Hispanic males have forecasts that drop well below the one for non-Hispanic white males, and males of some race/ethnic groups have lower forecasts for their sum than their female counterparts. Again, the major drawback to using a model that excludes a drift parameter is the somewhat rapid

convergence to a flat long-range behavior. There may be some early oscillation around the stable state for forecasts from these models, but there is essentially no movement after the first 5 years or so of forecasts. Ultimately, it is difficult to recommend selecting a model whose results are not trusted, so the models without drift, while boring, are preferred.

5.2 Relative Age-specific Mortality Rates

Modeling of age-specific mortality rates proceeds in much the same way as modeling of age-specific fertility rate does. As before, the goal is to produce forecasts of age distribution curves for mortality. If a zero rate is encountered, the situation was remedied by substituting a very small value for the mortality rate. In this application, the value was arbitrarily chosen as 10^{-8}. One other difference is that in addition to using the last age (100+) as a reference age, the first age (0) was also tried. This amounts to mortality rates compared against infant mortality rather than centenarian mortality. Because the observed age distribution curves experienced different movement depending on whether the age was above or below 20, it seemed that the behavior of mortality below age 20 may not have been adequately handled by a reference age at the upper extreme, so it was thought that using age 0 as an additional reference would help counteract this. Ultimately, it turned out that the gains resulting from the alternate baseline are minimal at best.

One distinction from the fertility rate section comes from the different value of the smoothing parameter on the spline smoother for the age-adjusted means in this particular application. In fertility rates, the smoothing parameter for R's *smooth.spline* function was defaulted to 0.5, and that number worked well enough that we did not see the need to adjust it. For mortality rates, a few options were considered, from very small values near 0, which would equate to almost no smoothing, to large values near 1, which signify heavy smoothing. Very large values for the smoothing parameters resulted in forecast curves that washed out features of the observed curves that should be retained, but minuscule values resulted in very jagged forecast curves, and given the preference for smoothness, this was not desirable either. The plots shown here will be correspond to analysis performed using a smoothing parameter of 0.25; this could be viewed as a low-to-medium amount of smoothing, and it represented a (subjective) compromise between handling the early features of the mortality rate curve correctly and achieving smoothness in the resulting forecast curve.

Similar to the earlier section on age-specific fertility rate, the principal components approach will be used again here. The age range going from 0 to 100+ means there are 101 ages to handle, so dimension reduction was required in this case. The previous threshold

rule of using as many principal components as will account for 95% of the total variation in the worst subgroup was employed here as well; in this case, the results will be displayed for each subgroup of gender and race/ethnicity using the analysis performed using 16 principal components. For some of these subgroups, this many principal components was required to achieve the 95% threshold rule, and for the ones that required less, there is no real drawback to using more principal components than the bare minimum. Once again, each of the 16 principal components will be modeled univariately using an ARIMA (1, 1, 0) with drift. A multivariate approach may actually make sense here, but it might be that any benefit to such an approach would be offset by the increased complexity in the calculations required to fit the models.

Because the mortality rates below age 40 (with the exception of infant mortality) are so small relative to the mortality rates post 40, it would be very difficult, if not impossible, to notice any movement in mortality rates below 40 if the whole age range was displayed on a regular scale. Hence, the plots all use a natural log scale for the relative mortality rates, which allows for better resolution in the lower ages. Thus, better insight is available into the behavior of the forecasts relative to the last observed curve from 2009. There is a flaw to using a logarithmic scale, though: the vertical difference on a logarithmic scale does not correspond to vertical distance on the regular scale. This helps explain some potential graphical oddities.

Only the plots for forecasts done using the natural logarithm will be displayed here because what differences do exist between the forecasts produced using the natural logarithm versus the logistic transforms are largely confined to the highest ages. For the rest of the age range, there is just some (expected) oscillation around 0. In addition, the natural logarithm and the logistic with baseline 0 track closely, as their maximum difference is only around magnitude 0.003; on the other hand, the logistic with baseline 100 tends to explode at 100, so the magnitude of its difference against one of the other transforms ends up topping out around 0.13, which is too high.

Figures 16 through 20 display the natural logarithms of the forecasted age distribution curves for age-specific mortality rates among the various race/ethnic groups, separated by gender. The plots again share some common characteristics, such as their shape. All start off high due to infant mortality, then drop to a minimum around age 10, swing back up to around age 20, and then increase steadily up to age 100. Another feature is the close tracking of the 1-year ahead forecast for 2010 with the actual 2009 curve. There is some concern that forecasts for the drop described between the ages 0 and 20 may be too large, although that

may be a consequence of using logarithms on very small numbers. Also, we see a similar magnifying effect that a long-term forecast produces on minor features of the curve. This is especially noticeable with black and Hispanic males, and white males to a lesser extent, as their curves reveal the presence of an increasingly pronounced dip in the curve going from around age 20 to age 40.

As was previously pointed out, there are some peculiarities apparent in the curves as well. For some groups, like black males, American Indian / Alaska native males, and Hispanic males, at the highest ages, there is a decrease in the age-specific mortality rates. Also, in some of the plots, the long-range forecast of the age distribution curve seems to be consistently less than the observed age distribution curve except at a few points, and yet the curves, once exponentiated, will all sum to 1. This is the aforementioned flaw of the logarithmic scale, in that the visual difference does not equal the numerical difference on a non-logarithmic scale. That is, a difference of 1 on a logarithmic scale when the values are around -10 is almost minimal on a non-logarithmic scale when compared to a difference of 1 when the values are around -4 (i.e., on a logarithmic scale, there is the same movement when going from -9 to -10 as there is going from -3 to -4, but eliminating the logarithmic scale gives $e^{-9} - e^{-10} \approx 8 \times 10^{-5}$, whereas $e^{-3} - e^{-4} \approx 0.03$).

6 An Alternative Approach for Mortality Rates

6.1 Age-adjusted Mortality Rate

The total mortality rate produced by summing up the age-specific mortality rates yields some implausible results. On the one hand, we believe that age-specific mortality rates should be trending downward in the long-run, but some of our drift models feature an upward trend. One drawback to a straight sum of the age-specific mortality rates is that it may be unduly influenced by the older ages, which tend to have the highest mortality rates. At the same time, that age range represents a relatively small proportion of the population, and as such, our total mortality rates may not be accurately depicting overall movements in mortality.

We consider here the possibility of using an age-adjusted mortality rate in lieu of the total mortality rate. The age-adjusted mortality rate takes the definition

$$(\text{age-adjusted mortality rate})_t = \sum_{k=0}^{K} w_k q_{kt},$$

where q_{kt} is the mortality rate at age k in year t, w_k is the proportion of the population aged k in some reference population, and K should be the upper limit of the mortality table (i.e.,

the age K for which $q_{Kt} = 1$). For our purposes, we use K as the upper limit of the age range that we have been provided. Because the w_k will be small at the upper limits of the age range, the age-adjusted mortality rate (hereafter abbreviated as AAMR) will place far less weight on the older ages than the total mortality rates from the previous section. Forecasting the AAMR will be done in exactly the same fashion as forecasting the TMR from before.

In order to produce the weights w_k, we examined a few options. The first possibility is to use the age distribution of the full U.S. population from the 2010 Census (data available on American Factfinder). This option would use the same weights for all of the gender by race/ethnic groups. Other weighting schemes would find the population proportions for the U.S. population stratified by gender and/or race/ethnic group. Separating solely by gender would create a distinct set of weights for all males and another for all females. Using weights based solely on race/ethnic group would yield 5 sets of weights and ignore gender differences. A stratification using both factors would end up creating 10 different sets of weights, one for each combination of gender and race/ethnic group.

Figure 21 displays the weights that would be used under two of the proposed schemes. The top plot shows the population proportions associated with the full U.S. population in 2010. We see that there is a steep decline in the age weights that starts in the early 60s. The curves in the bottom plot give the population proportions associated with each of the 5 major race/ethnic groups from the 2010 U.S. population. One feature to note from this lower plot is how similar the curve for non-Hispanic white proportions is to the one for the overall population from the top plot. This is attributable to the comparatively larger size of the non-Hispanic white group relative to the other 4 race/ethnic groups. We also observe that the non-Hispanic white curve has a much larger proportion of people over age 50. The non-Hispanic black and American Indian Alaska native populations are very close in overall shape, with the AIAN population having a slightly greater proportion below age 20. Lastly, the Hispanic population has a very large proportion of young people, but sees a very dramatic fall in the population proportion above age 40. This may indicate that Hispanic immigrants tend to be relatively young and are also more likely to return to their country of origin as they grow older.

We did not include a plot that separated by gender. There is not much to distinguish a curve of male population proportions and a curve of female population proportions. The shapes are quite similar, although there is a greater proportion for males below age 50 and a greater proportion for females above age 50. That is, there are a greater number of young males than young females, but conversely, there are a greater number of old females than old

males. This observed gender disparity holds for all of the race/ethnic groups, so we felt that using weights stratifying on both gender and race/ethnicity was not necessary.

We proceed to plot the age-adjusted mortality rate for each gender by race/ethnic group combination using weights derived from both the full U.S. population and the respective race/ethnic group. Figures 22 through 26 display the forecasts produced using ARIMA (1, 1, 0) models with drift parameters included for these two weighting schemes. For whites, we see that the use of the full population weights yields lower age-adjusted mortality rates and forecasts than the use of just the white population. The reverse statement holds for the other race/ethnic groups: using the full population weights pushes the age-adjusted mortality rates and forecasts higher. However, the forecast of age-adjusted mortality rate for AIAN females (see lower plot of Figure 24) is still showing a positive trend.

6.2 Age-specific Mortality Rates

In order to meld the age-adjusted mortality rate with the rest of the procedure described previously, we will transform from q_{kt} to $r_{kt} = w_k q_{kt} / \sum_{k=0}^{K} w_k q_{kt}$, which yields the individual age relative contributions to the AAMR, and these contributions will sum to 1. Taking the logarithm of the resulting r_{kt} allows us to use the procedure previously described in Section 3 for relative age-specific rates to obtain forecasts of $\widehat{r}_{kt}$. To transform the resulting forecasts $\widehat{r}_{kt}$ back to age-specific mortality rate forecasts $\widehat{q}_{kt}$, we would need to multiply the forecasts of AAMR, $\widehat{AAMR}_t$, by the forecasts $\widehat{r}_{kt}$ and then divide by the weights w_k; that is, we would get the forecasts as

$$\widehat{q}_{kt} = \frac{\widehat{r}_{kt}\widehat{AAMR}_t}{w_k}.$$

We display the results for log age-specific mortality rates from using the individual race/ethnic group weights in Figures 27 through 31. While we could have used the full population weights instead, the use of the proportions specific to each race/ethnic group would seem to produce a more accurate depiction of mortality trends. We chose not to convert the age-specific rate to its relative counterpart, as we were primarily interested in examining the shape of the log curve, and rescaling only stretches the general shape. For whites (see Figure 27), it appears that adjusting for age produces some kinks in the age-specific mortality forecasts. In particular, there is an obvious one that occurs right near the upper edge of the age range. The forecasts would thus suggest that the mortality rate is actually highest not at the 100+ age, but just below it, which is implausible. Similarly, we see that many of the features that were present in the log curves of the relative mortality rates that were produced

using the total mortality rates persist here. We conclude that an age-adjusted scheme does not eliminate the issues we encountered in using a straight sum for a mortality index.

7 Discussion and Summary

Before providing a summary of what was done in this application, a few other thoughts about changes that could be incorporated into future methodology will be briefly discussed. Some interesting features were observed in the course of producing the age-specific fertility and mortality rate forecasts. Regarding age-specific fertility rates, the mode of the age distribution curves drifted toward the mid 30s as the forecast horizon increased out to 60 steps ahead. This phenomenon occurred for all five race/ethnic groups, and it suggests that the age of peak fertility increases by at least 5 years over the course of 60 years. The rightward drift is not limited to just the mode, however; over the forecast horizon, regardless of transformation used, the mass of the age distribution curve also moves toward the higher ages. While such shifts are plausible, the magnitude appears to be somewhat exaggerated, so a good modification may restrain the extent of the rightward drift in the future.

Another concern manifests when considering the age distribution curves for mortality. Because the mortality rates in the ages below 40 are so small relative to those past age 40, a log transformation proves useful for revealing details. When plotting the log-transformed age distribution curves for mortality, one observes that the curve over the first 20 years behaves similarly for any combination of gender and race/ethnic origin. Specifically, there is a high infant mortality, which leads into a fall until about the age of 10. The (smoothed) log-transformed curve over this interval has somewhat of a "U" shape in this interval. Forecasting this movement accurately presents a challenge, however; over the course of a 60-step forecast horizon, the curve becomes steeper going down and coming back up, and the bottom around age 10 steadily decreases. It would be preferable to see less severe change in the forecasts over time in this area of the curve. On the other hand, the relative mortality rates in this region are very small as previously noted, so the movements would be difficult to see on a normal scale.

Because the vast majority of all deaths occur after age 40, it was thought that one could ignore ages below 40 for forecasting mortality rates. This would restrict the forecasts to a range that matters more, and censoring the data in this fashion would have mirrored the work of Lundström and Qvist (2004), who were forecasting Swedish mortality in order to estimate life expectancy. They chose to disregard mortality below the age of 40 in their work, as they found that over 98% of birth cohorts (for a Swedish life table) survived to age 40. This idea

was given a trial in the hope that it would smooth out some of the long-range oscillations in the ages above 40 (in particular the swing under the last observed curve), but the features in that interval that were originally observed in the forecasted age distribution curves produced using the full range of ages persisted even when the first 40 ages were censored. Thus, this idea was discarded, as it did not produce a measurable improvement.

Due to a preference for a simple model, an ARIMA (1, 1, 0) model was selected for forecasting the principal component series that go into producing the forecasts of age-specific rates. The same model was also used for forecasting the total fertility and mortality rates. A higher-order model may prove superior if a larger amount of data is available, as the movement of the forecasts under the $(1, 1, 0)$ may be too heavily influenced by the last observed difference. This influence could prove particularly problematic when the last observed difference coincides with a potential shock effect; the data used for this iteration extended through 2009, meaning the last observed difference would have lined up with the recent U.S. recession. Although mortality rates appear to be largely immune to economic stimuli, the total fertility rate did drop noticeably between 2008 and 2009, and this leads to overly aggressive downward forecasts for total fertility rate for the more volatile race/ethnic groups (non-Hispanic whites were not affected as much).

Error estimation is an additional issue that has not been discussed in the main body of this report. We briefly sketch a scheme based on the bootstrap procedure of Efron (1979) as a method for quantifying the amount of error in some of our estimates. Recall that our form of the principal component approximation of γ_t gives

$$\gamma_t = \overline{\gamma} + \widetilde{\Lambda}\widehat{\beta}_t + \widehat{\epsilon}_t,$$

where the first two terms on the right-hand side of this equation are treated as $\widehat{\gamma}_t$. Thus, resampling the residuals obtained by subtracting $\widehat{\gamma}_t$ from γ_t would allow us to obtain bootstrap sets γ_t^*, which we could feed into our procedure in order to check the variability of β_t. This in turn would lead to error estimates on both the h-step ahead forecasts of $\widehat{\gamma}^{(h)}$, and after reversing the original transformation, the h-step ahead forecasts of $\widehat{r}^{(h)}$. Some of these could ultimately take the form of a pointwise interval.

There are a couple of caveats attached to the outline of the bootstrap procedure above. The first stems from our initial decision to retain all of the principal components, but only forecast the first J of them. Using the full $K \times K$ matrix $\widetilde{\Lambda}$ would yield a $\widehat{\gamma}_t$ that is identical to γ_t, so we would not have any residuals to resample. Thus, in order for this bootstrap scheme to work, we need to use a lower-dimensional approximation, which would ideally arise from J

being strictly less than K. Given that that is not necessarily going to be the case, we choose some value M that is strictly less than K (the simplest choice being $M = 1$) and obtain $\widehat{\epsilon}_t$ as the residuals formed when approximating γ_t with just M principal components. Resampling these residuals then allows for the construction of new bootstrap set γ_t^*. The second point stems from the possibility that a bootstrapped set γ_t^*, when reverse transformed, yields a bootstrap sample of relative rates r_t^* that does not sum to 1 at every time point. Relative to the first issue, this only needs a fairly straightforward correction: once we construct a bootstrapped set γ_t^* as outlined above, we reverse the transformation, rescale so that the sum at each time point does sum to 1, and then transform again.

In summary, a description of the original decomposition of the full fertility and mortality data into disjoint subgroups according to race/ethnicity or the combination of gender and race/ethnicity has been provided by our work. The method used to forecast the sum of the age-specific rates included a logarithmic transformation in order to ensure positivity in the forecasts. For total fertility rate, an additional modification was included with the subtraction of 1 prior to taking the logarithm, as this change guaranteed that the forecasts would be greater than 1. A minor point about the interpretation of the sum of age-specific mortality rates (alternatively referred to as total mortality rate) was raised. For age-specific rates, the transforms that were used included the natural logarithm, but the generalized logistic transform as well, which had the extra benefit of producing forecasts that were already age distribution curves without any further normalization.

The more complicated principal components approach to producing age distribution curves for fertility and mortality rates was explained, with some discussion of the choice of smoothing parameter for the stages that called for a smoothing spline. The default smoothing parameter was 0.5, and that worked well enough for fertility rates. For mortality rates, however, 0.5 seemed to be too large, so smaller values were tried, with 0.25 being the eventual choice. The number of principal components to use was selected to satisfy a worst-case 95% scenario, in which the resulting dimension would ensure that at least 95% of the total variation would be accounted for by every subgroup.

The difference between an ARIMA (1, 1, 0) model with a drift parameter versus one without a drift parameter was discussed. For the total rates, it seemed that the models without drift yielded more sensible forecasts, whereas for age-specific rates, the drift parameter was included. Exclusion of a drift parameter in fitting time series models to the principal components found for the transformed relative age-specific rates would have resulted in forecasted curves that do a little oscillating around the last observed curve. Inclusion of drift, however,

resulted in curve features that became magnified over the course of the forecast horizon, so any small bumps end up turning into big ones. One remedy for this issue is to determine a truncation point beyond which all forecasts are held constant at the forecast of that time. This stopping rule would have a high degree of subjectivity, however.

References

[1] Bell, W.R. (1997), "Comparing and Assessing Time Series Methods for Forecasting Age-Specific Fertility and Mortality Rates," *Journal of Official Statistics* **13**, 279-303.

[2] Bozik, J.E. and Bell, W.R. (1987), "Forecasting Age Specific Fertility Using Principal Components," SRD Research Report No. RR-87/19, U.S. Census Bureau, available at http://www.census.gov/srd/papers/pdf/rr87-19.pdf.

[3] Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife," *Annals of Statistics* **7**, 1-26.

[4] Lee, R.D. and Carter, L.R. (1992), "Modeling and Forecasting the Time Series of U.S. Mortality," *Journal of the American Statistical Association* **87**, 659-671.

[5] Lundström, H. and Qvist, J. (2004), "Mortality Forecasting and Trend Shifts: An Application of the Lee-Carter Model to Swedish Mortality Data," *International Statistical Review* **72**, 37-50.

[6] McElroy, T.S. and Bell, W.R. (2004), "Forecasting Age Distribution Curves," *Proceedings of the 2005 Federal Forecasters Conference.*

[7] R Core Team (2012), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org.

[8] Thompson, P.A., Bell, W.R., Long, J.F., and Miller, R.B. (1989), "Multivariate Time Series Projections of Parameterized Age-Specific Fertility Rates," *Journal of the American Statistical Association* **84**, 689-699.

A All Plots

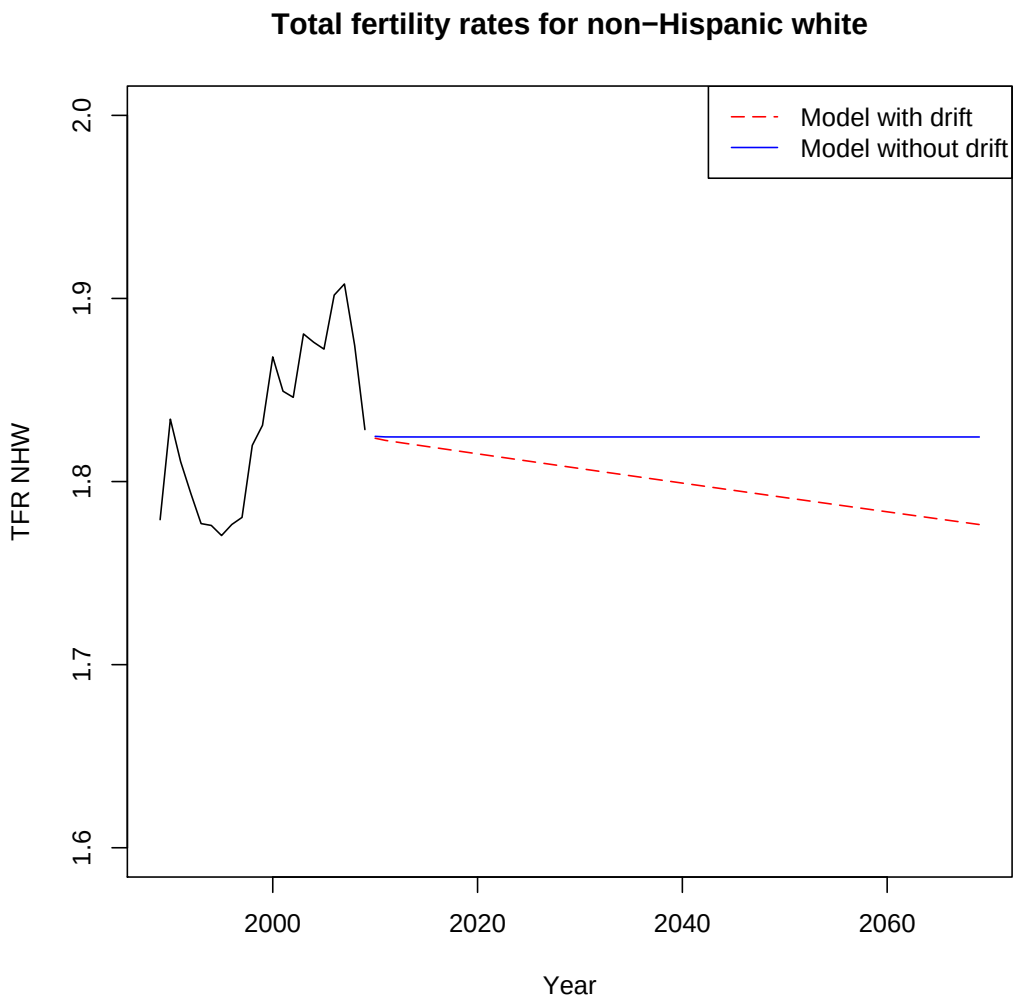


Figure 1: *History of TFR for non-Hispanic white women from 1989 - 2009 with ARIMA (1, 1, 0) forecasts for 2010 - 2069.*

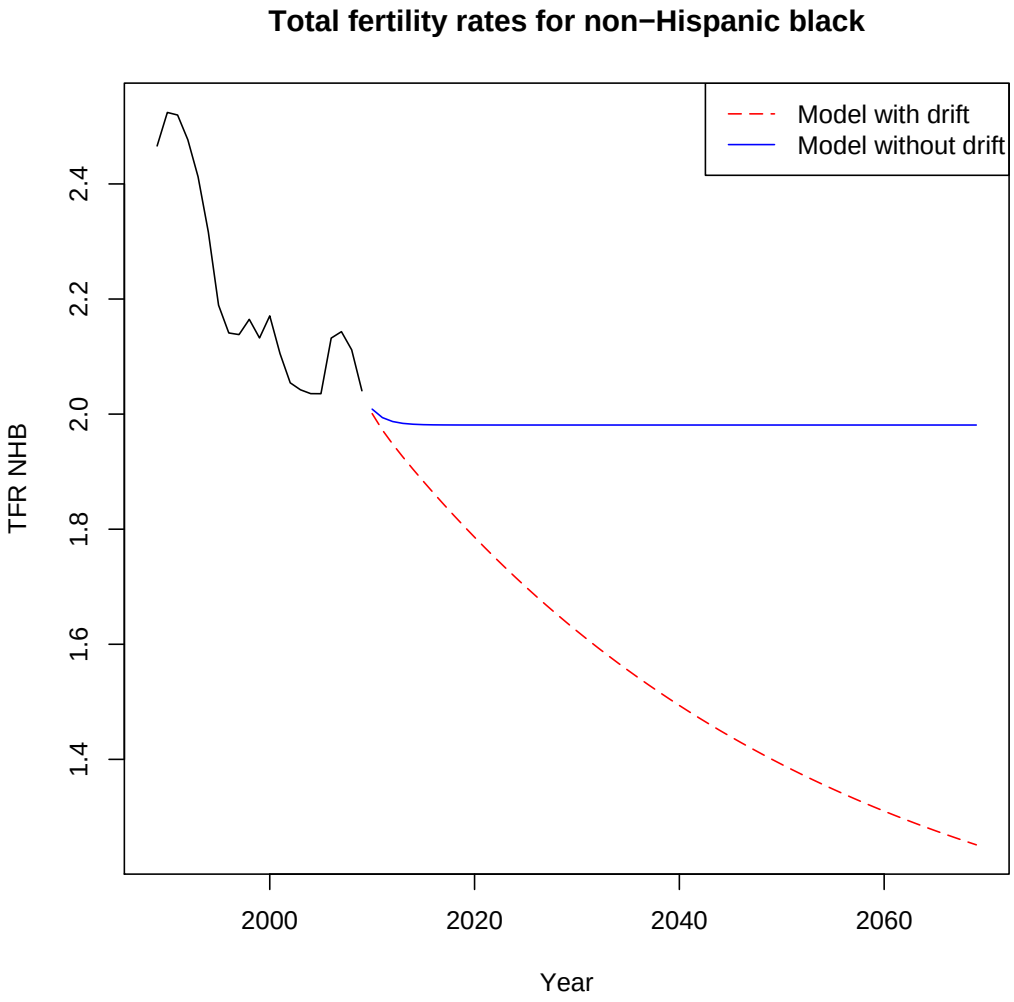


Figure 2: *History of TFR for non-Hispanic black women from 1989 - 2009 with ARIMA (1, 1, 0) forecasts for 2010 - 2069.*

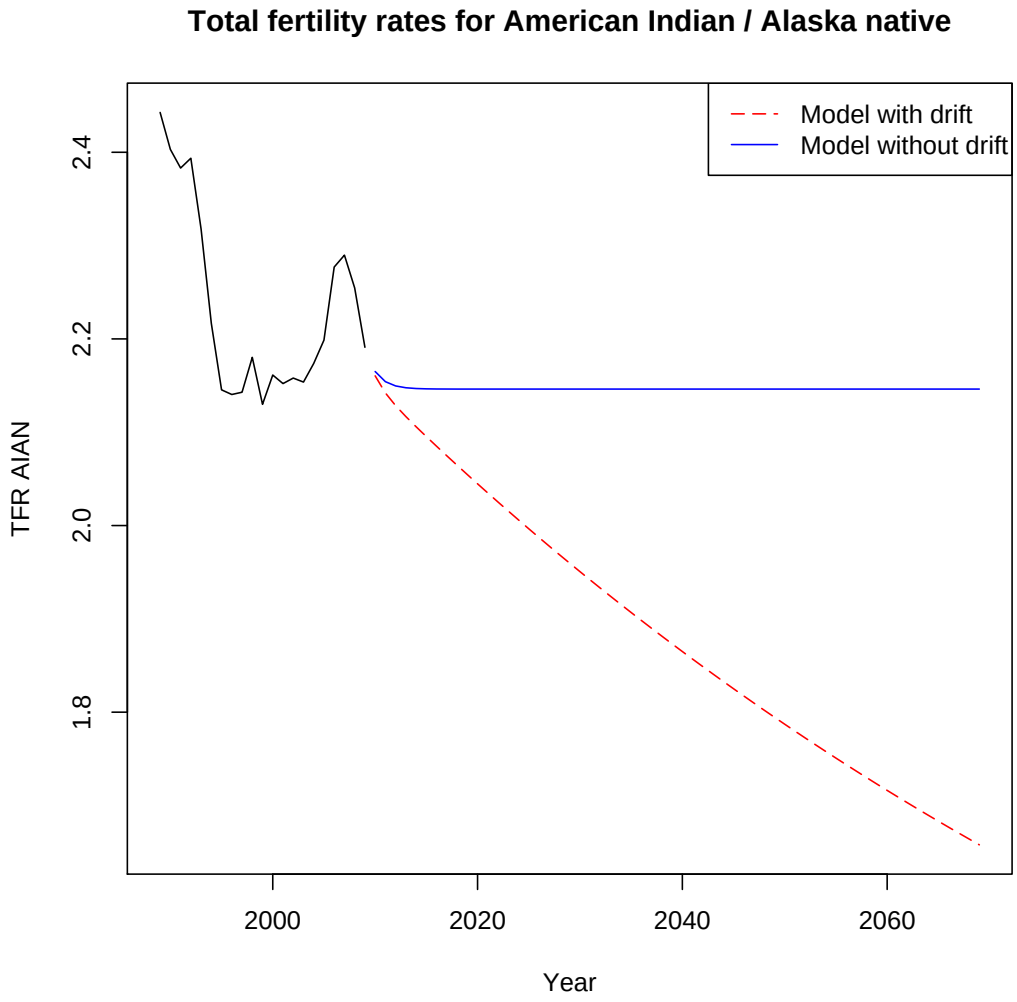


Figure 3: *History of TFR for American Indian & Alaska Native women from 1989 - 2009 with ARIMA (1, 1, 0) forecasts for 2010 - 2069.*

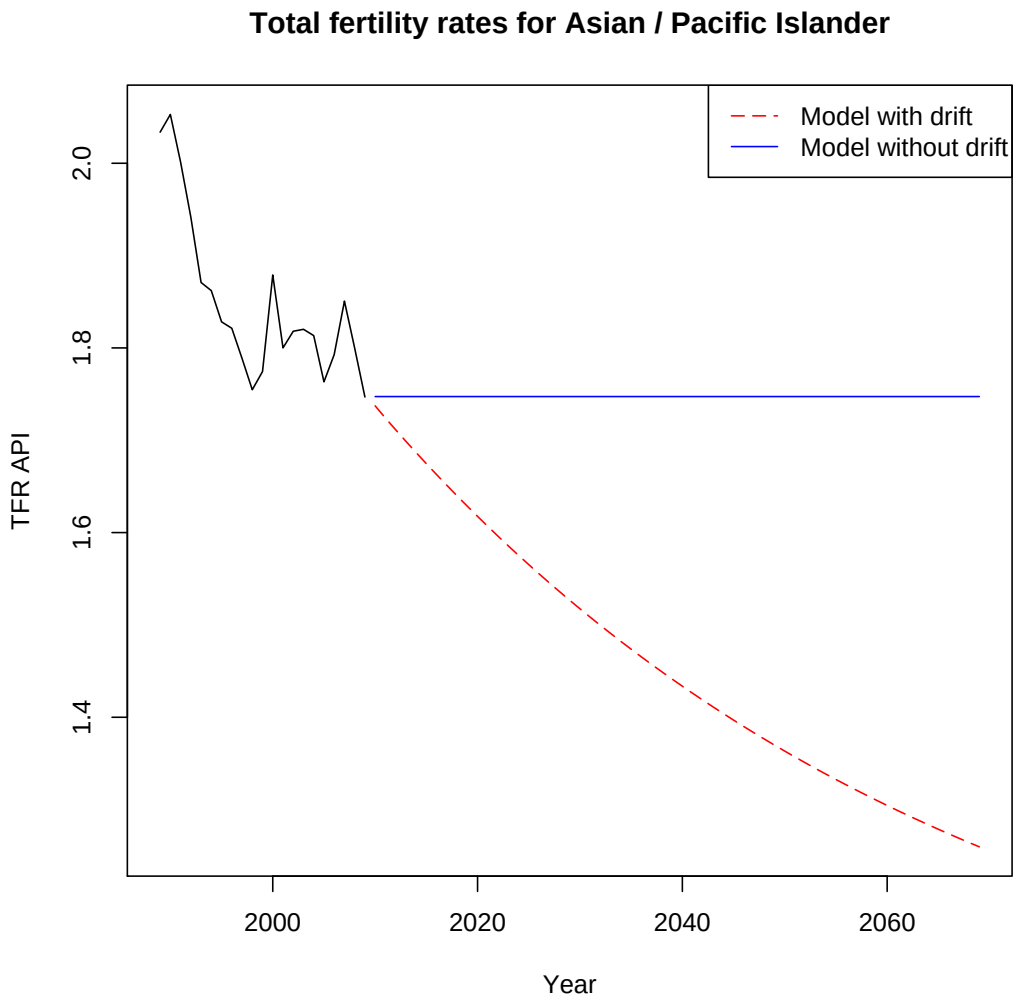


Figure 4: *History of TFR for Asian & Pacific Islander women from 1989 - 2009 with ARIMA (1, 1, 0) forecasts for 2010 - 2069.*

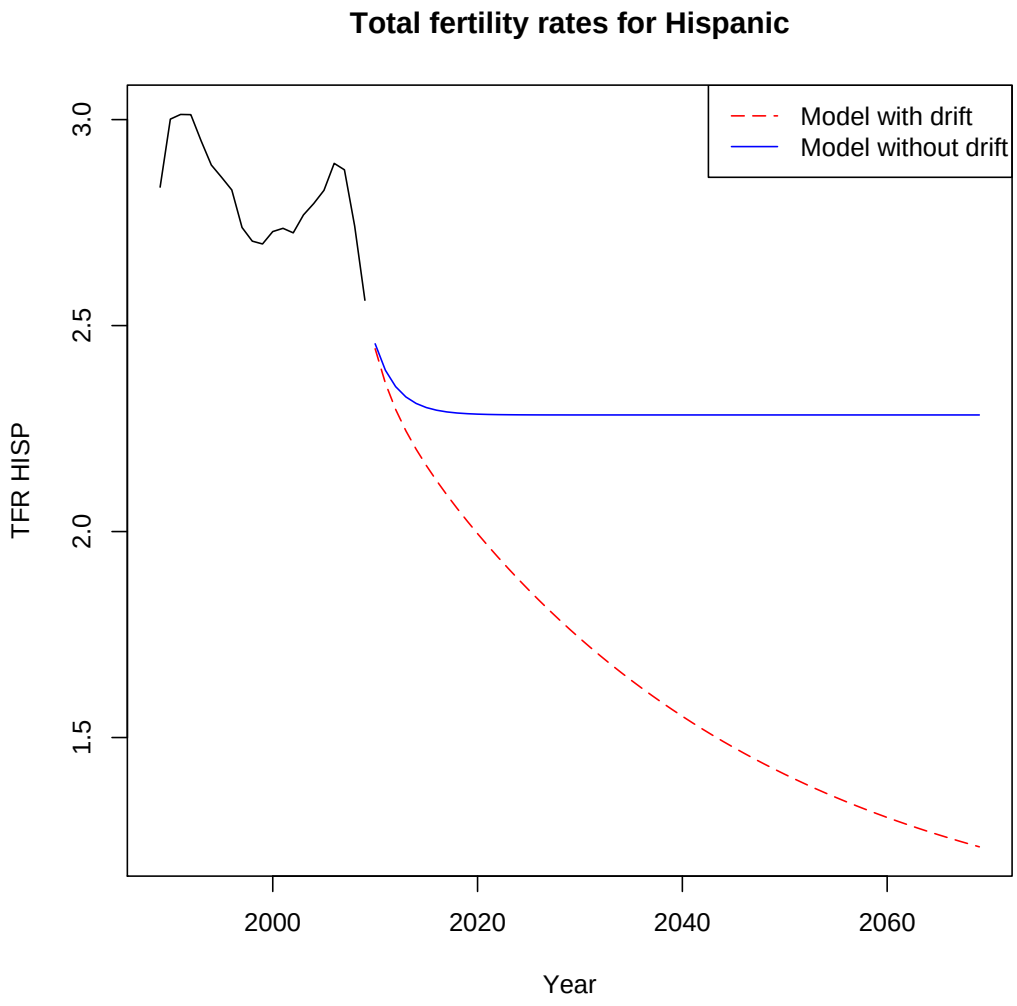
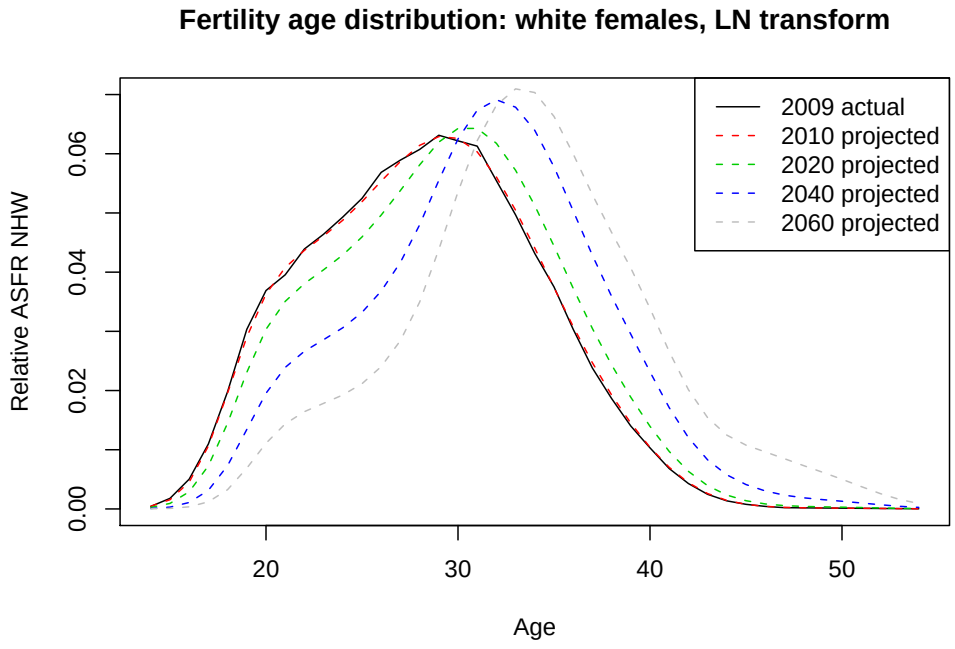


Figure 5: *History of TFR for Hispanic women from 1989 - 2009 with ARIMA (1, 1, 0) forecasts for 2010 - 2069.*



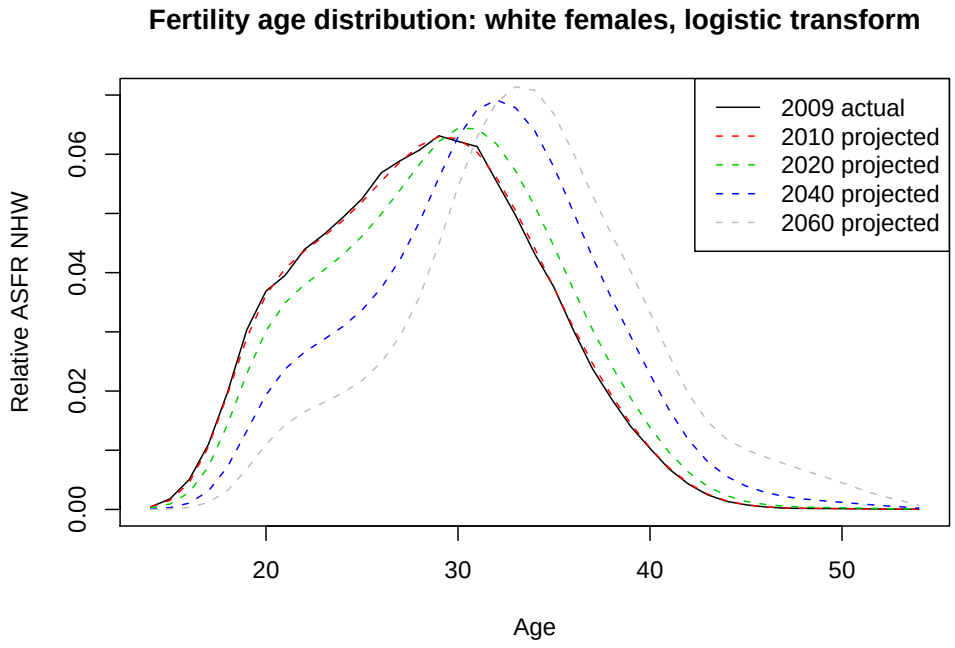


Figure 6: *Actual 2009 age distribution curve for non-Hispanic white females along with projected age distribution curves for 2010, 2020, 2040, and 2060. Top shows projected curves using a natural logarithm transform on relative asfr, while bottom shows same projections using generalized logistic transform.*

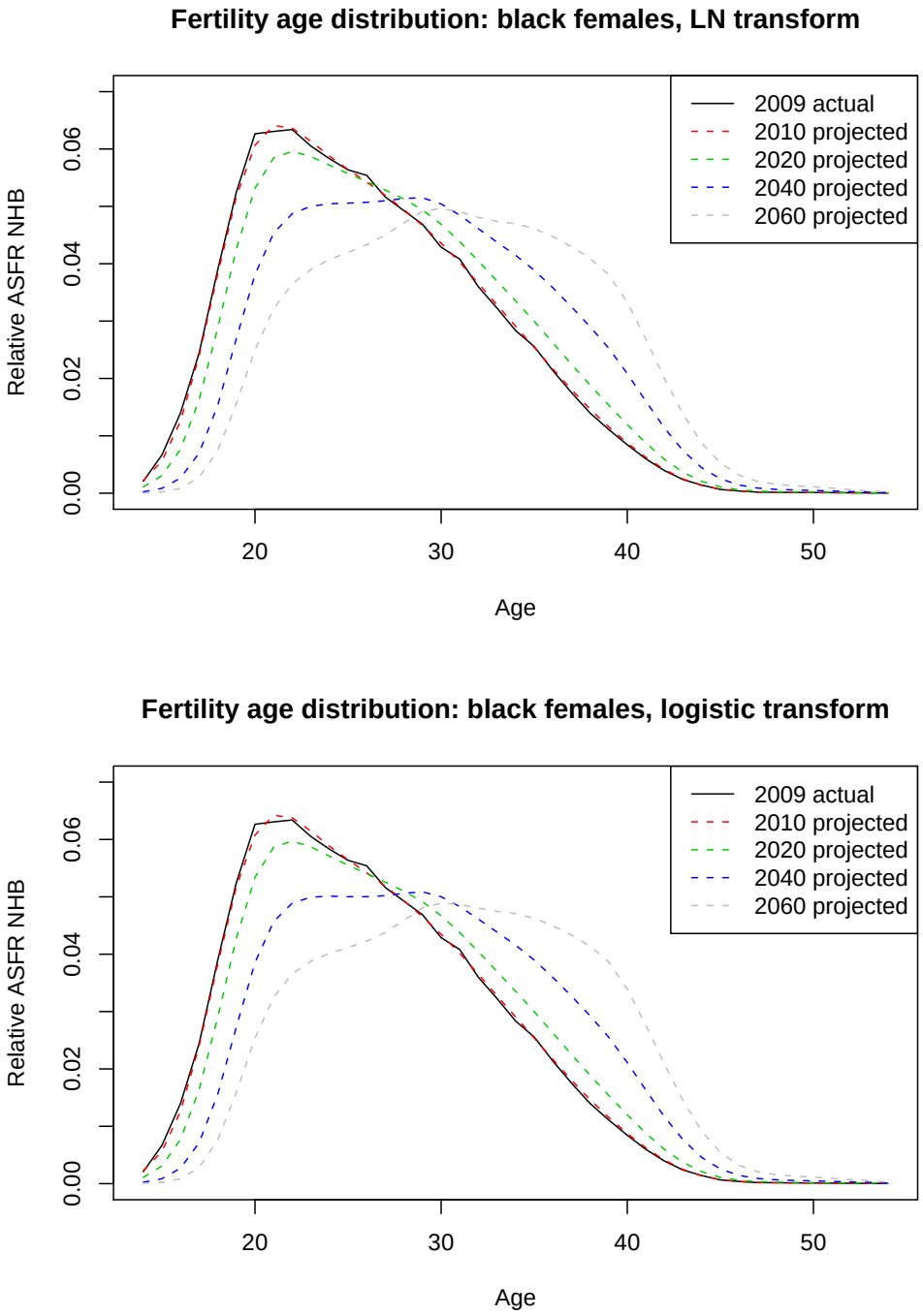


Figure 7: *Actual 2009 age distribution curve for non-Hispanic black females along with projected age distribution curves for 2010, 2020, 2040, and 2060. Top shows projected curves using natural logarithm transform on relative asfr, while bottom shows same projections using generalized logistic transform.*

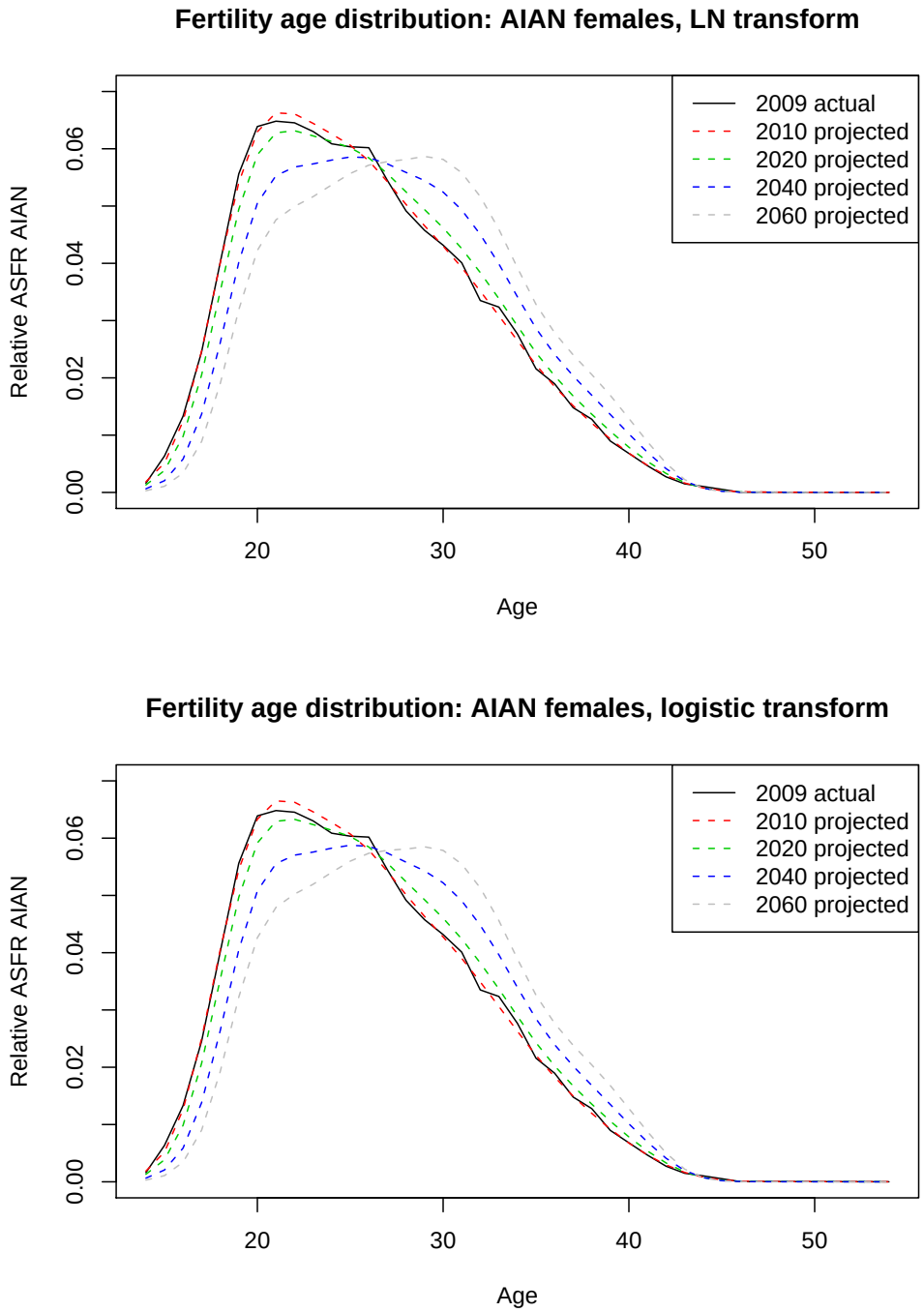


Figure 8: *Actual 2009 age distribution curve for American Indian & Alaska native females along with projected age distribution curves for 2010, 2020, 2040, and 2060. Top shows projected curves using natural logarithm transform on relative asfr, while bottom shows same projections using generalized logistic transform.*

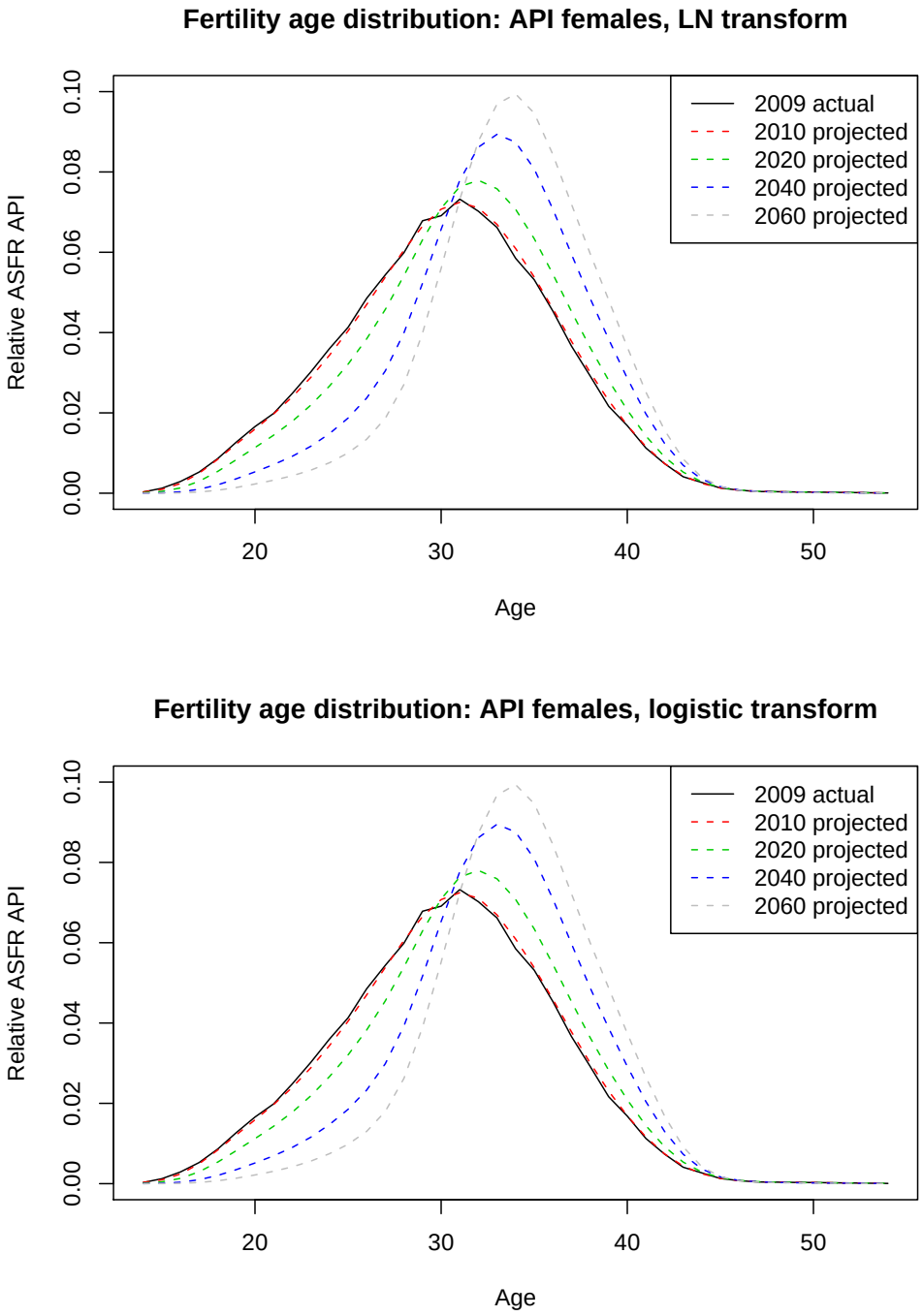
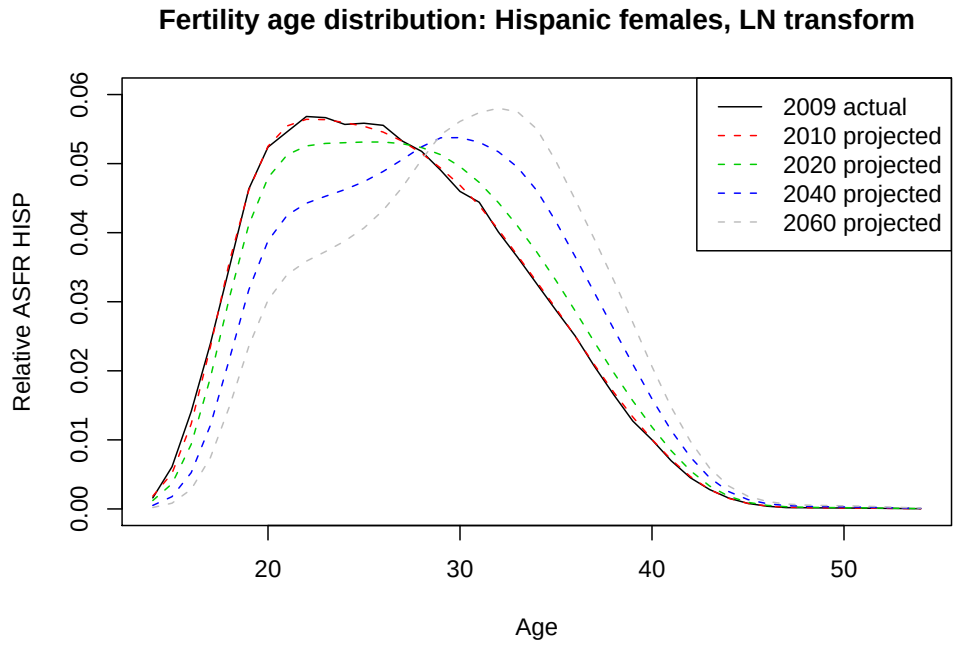


Figure 9: *Actual 2009 age distribution curve for Asian & Pacific Islander females along with projected age distribution curves for 2010, 2020, 2040, and 2060. Top shows projected curves using natural logarithm transform on relative asfr, while bottom shows same projections using generalized logistic transform.*



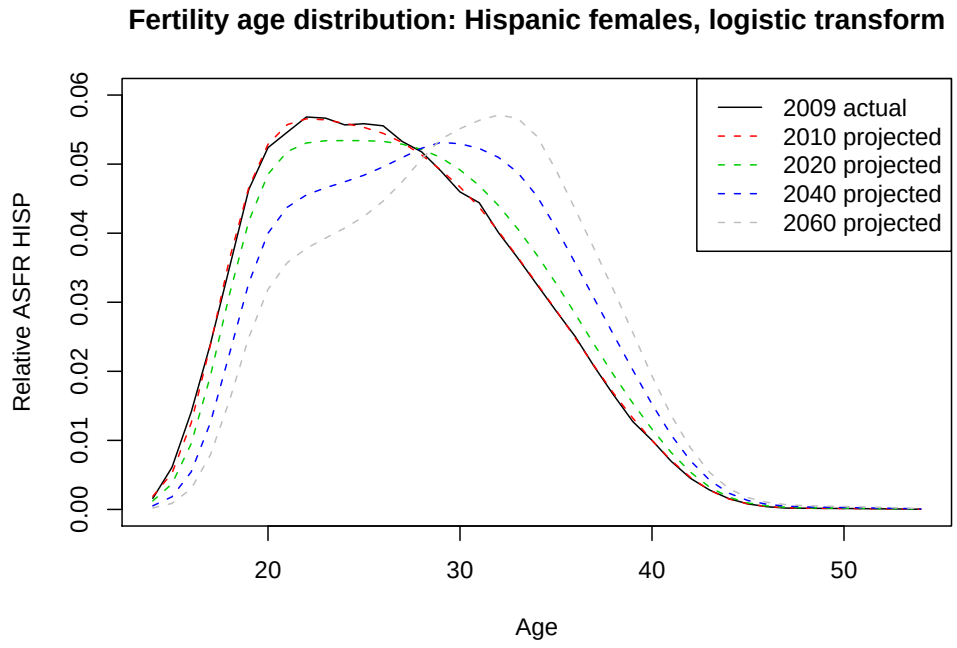


Figure 10: *Actual 2009 age distribution curve for Hispanic females along with projected age distribution curves for 2010, 2020, 2040, and 2060. Top shows projected curves using natural logarithm transform on relative asfr, while bottom shows same projections using generalized logistic transform.*

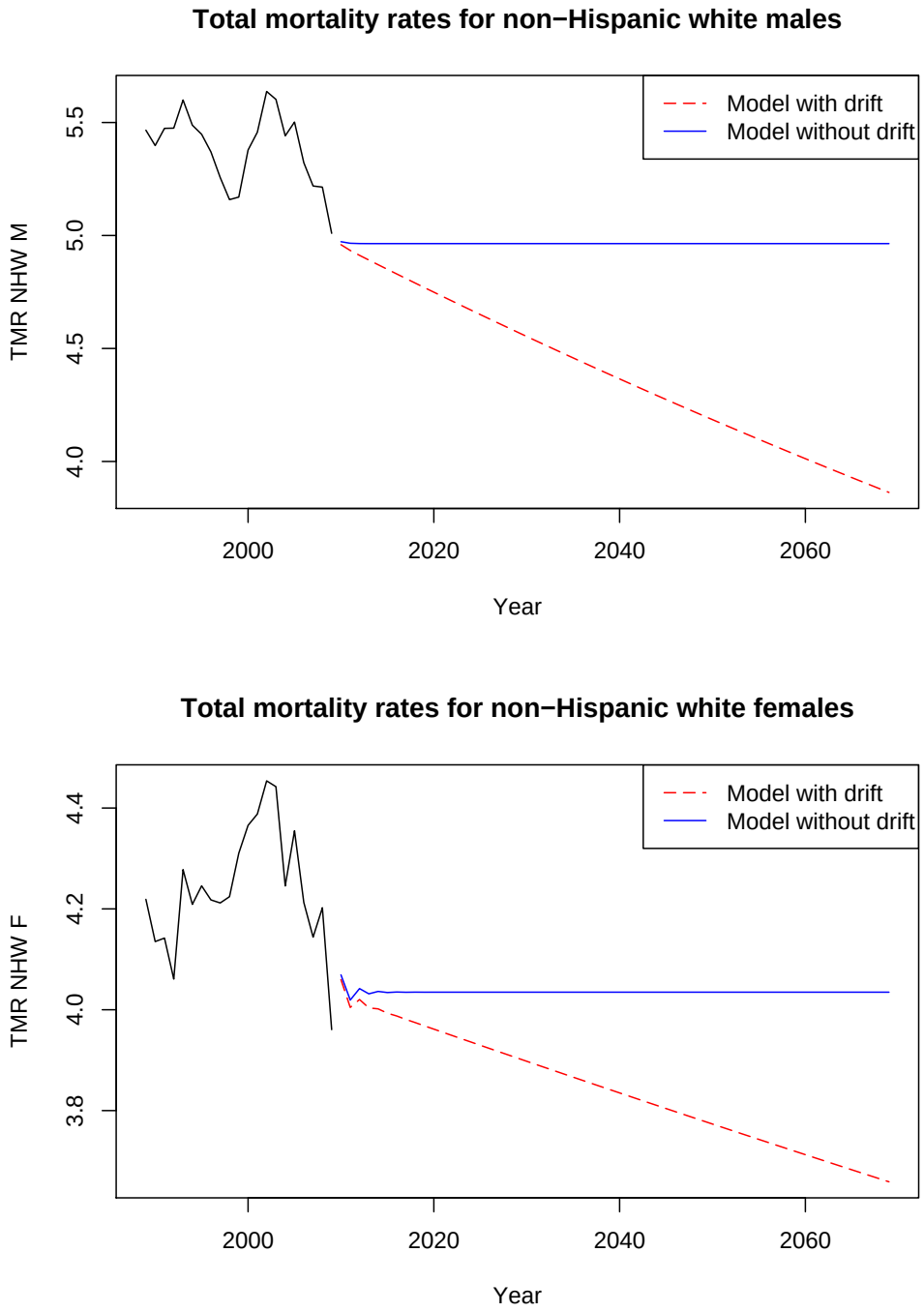


Figure 11: *History of total mortality rate for non-Hispanic white males (top) and females (bottom) from 1989 - 2009 with ARIMA (1, 1, 0) forecasts for 2010 - 2069.*

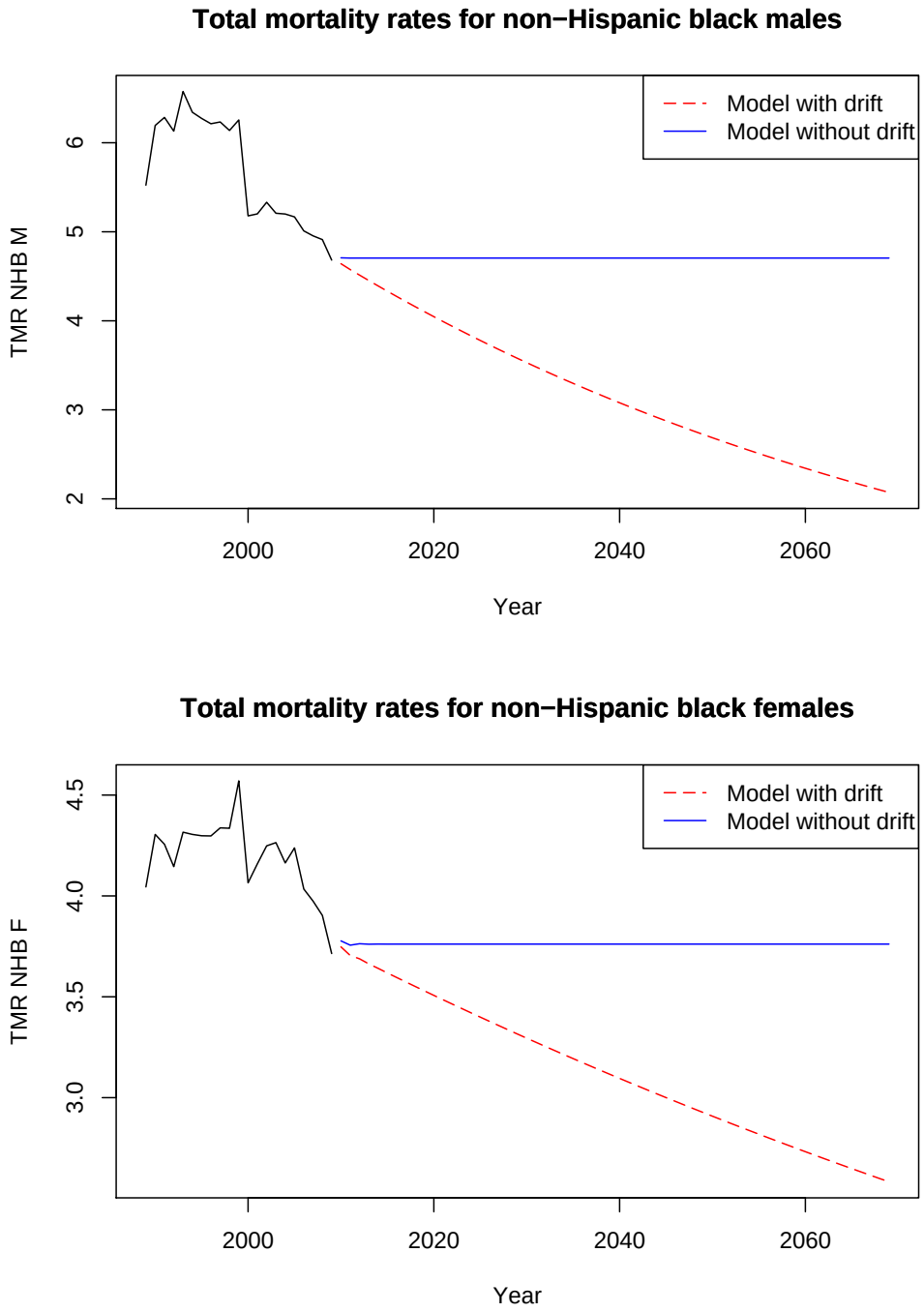


Figure 12: *History of total mortality rate for non-Hispanic black males (top) and females (bottom) from 1989 - 2009 with ARIMA (1, 1, 0) forecasts for 2010 - 2069.*

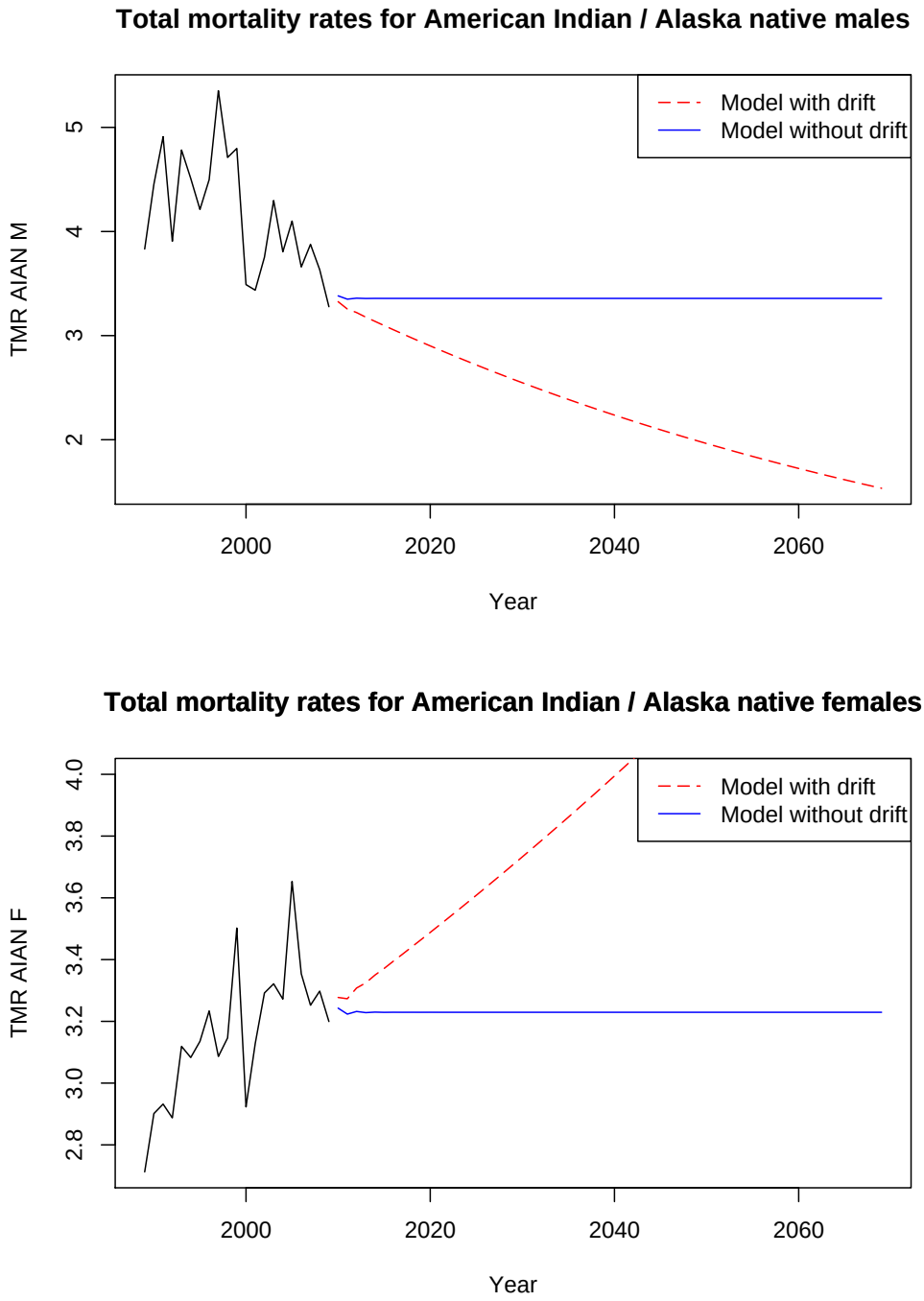


Figure 13: *History of total mortality rate for American Indian & Alaska native males (top) and females (bottom) from 1989 - 2009 with ARIMA (1, 1, 0) forecasts for 2010 - 2069.*

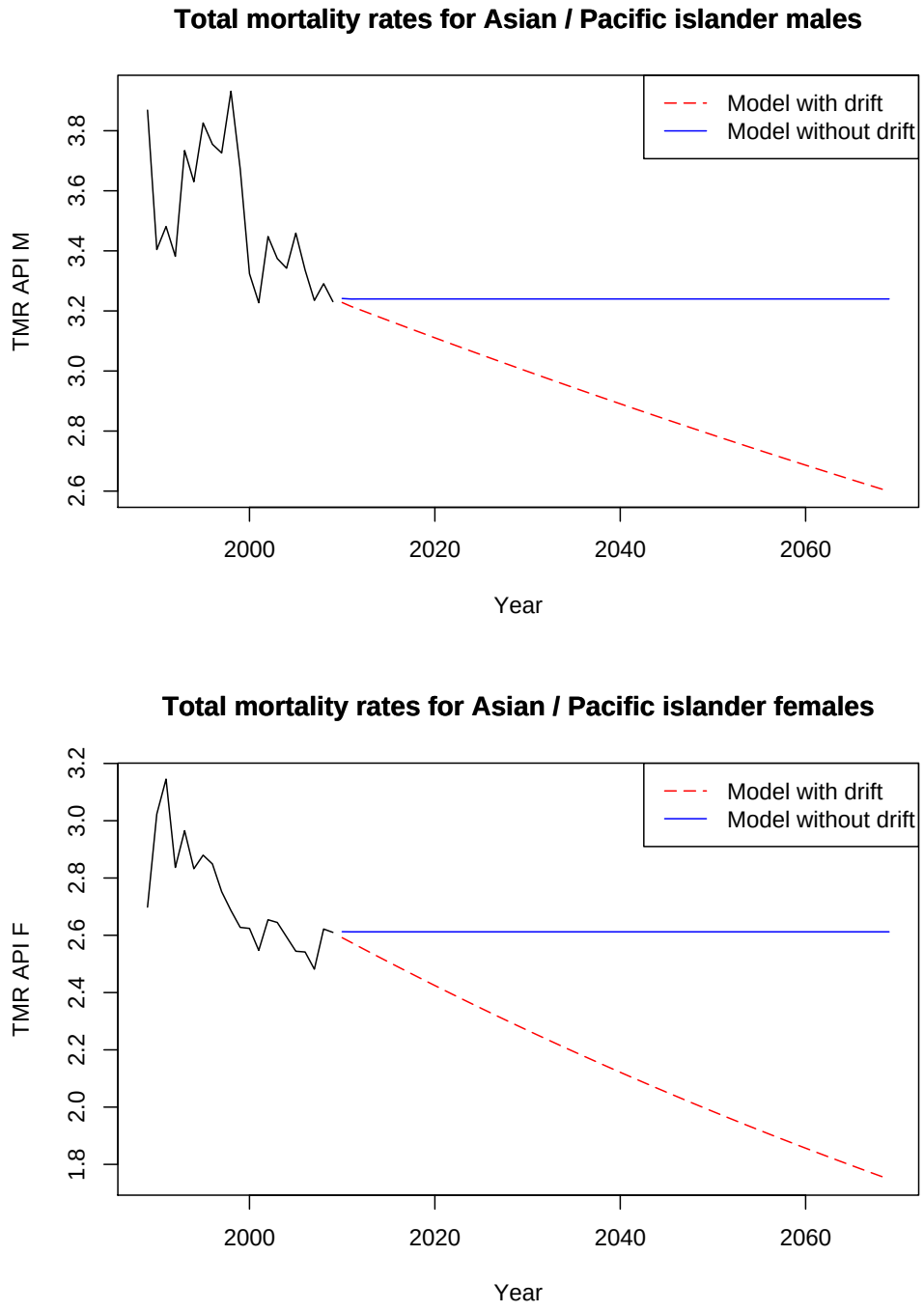


Figure 14: *History of total mortality rate for Asian & Pacific islander males (top) and females (bottom) from 1989 - 2009 with ARIMA (1, 1, 0) forecasts for 2010 - 2069.*

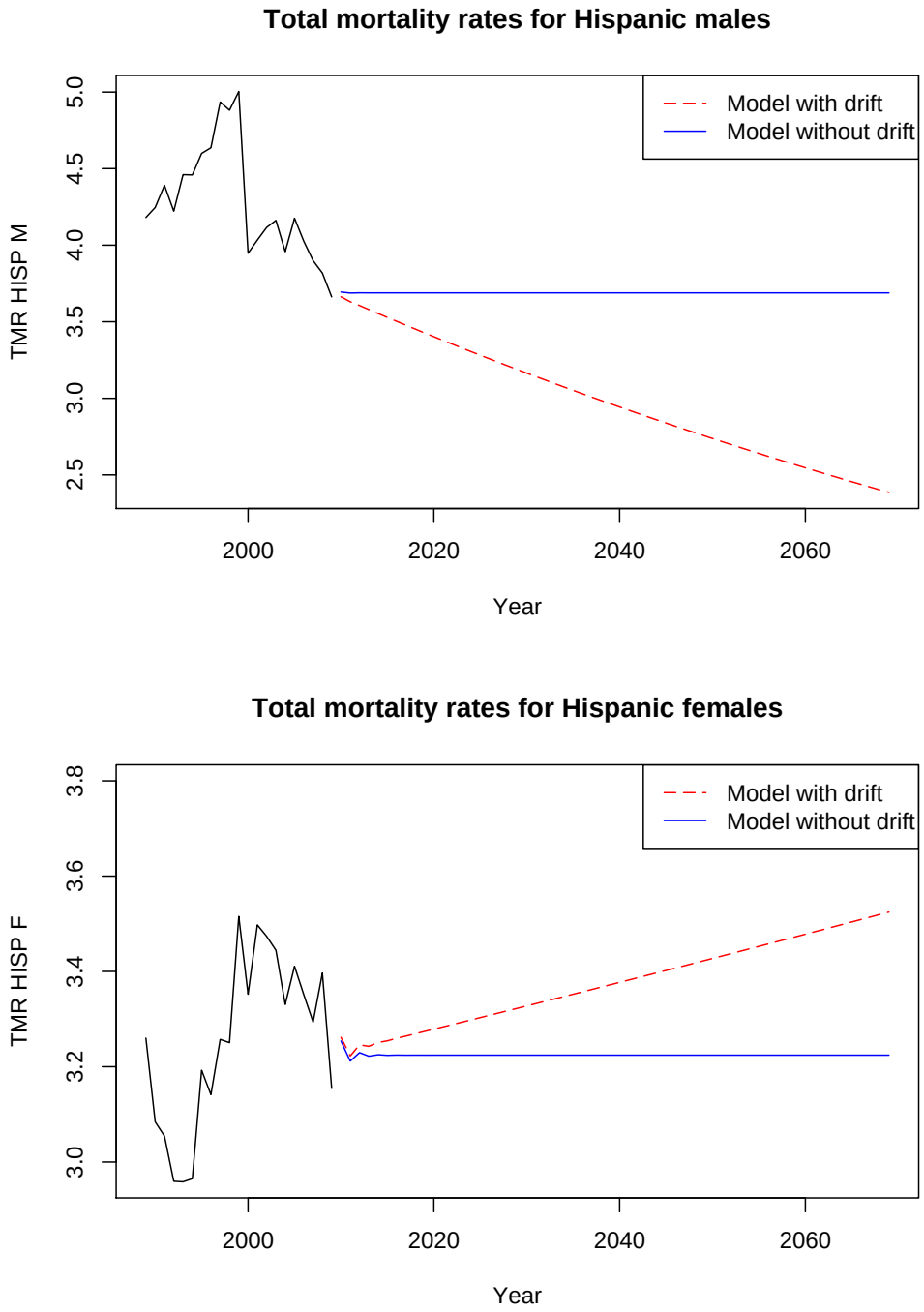


Figure 15: *History of total mortality rate for Hispanic males (top) and females (bottom) from 1989 - 2009 with ARIMA (1, 1, 0) forecasts for 2010 - 2069.*

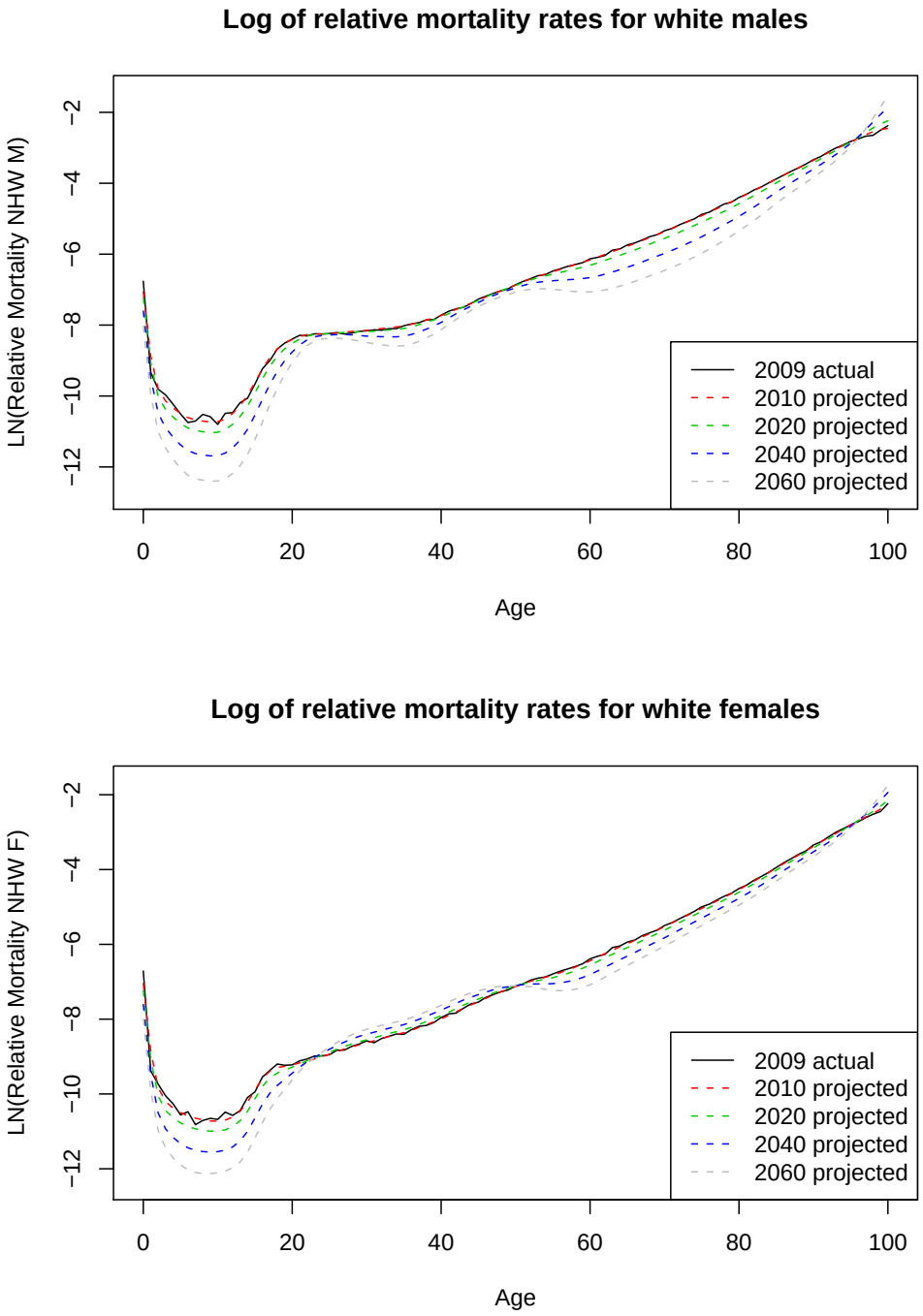


Figure 16: *Log of actual 2009 distribution curve of age-specific mortality rate for non-Hispanic whites along with projections for years 2010, 2020, 2040, and 2060. Top shows male rates, while bottom shows female rates.*

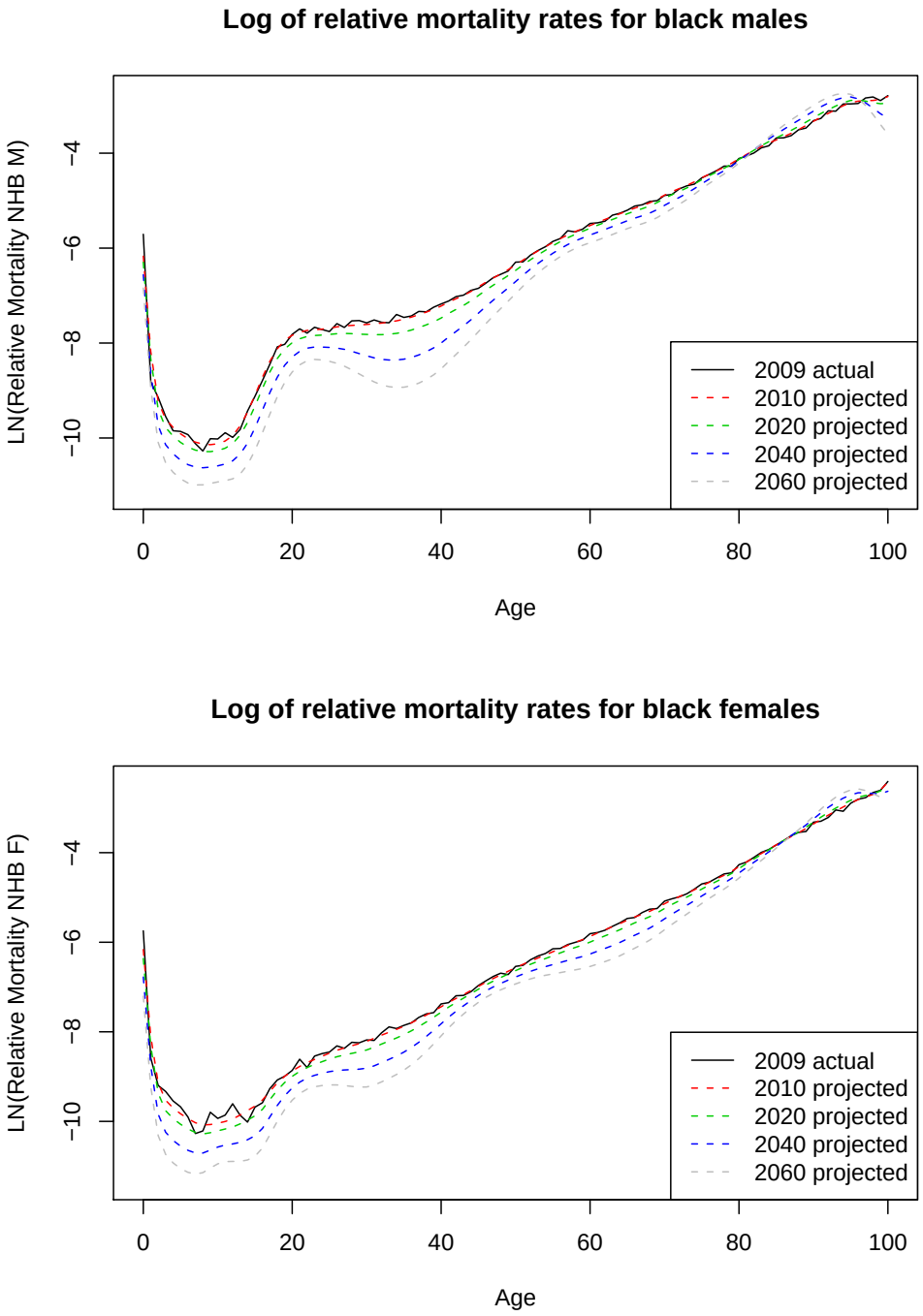


Figure 17: *Log of actual 2009 distribution curve of age-specific mortality rate for non-Hispanic blacks along with projections for years 2010, 2020, 2040, and 2060. Top shows male rates, while bottom shows female rates.*

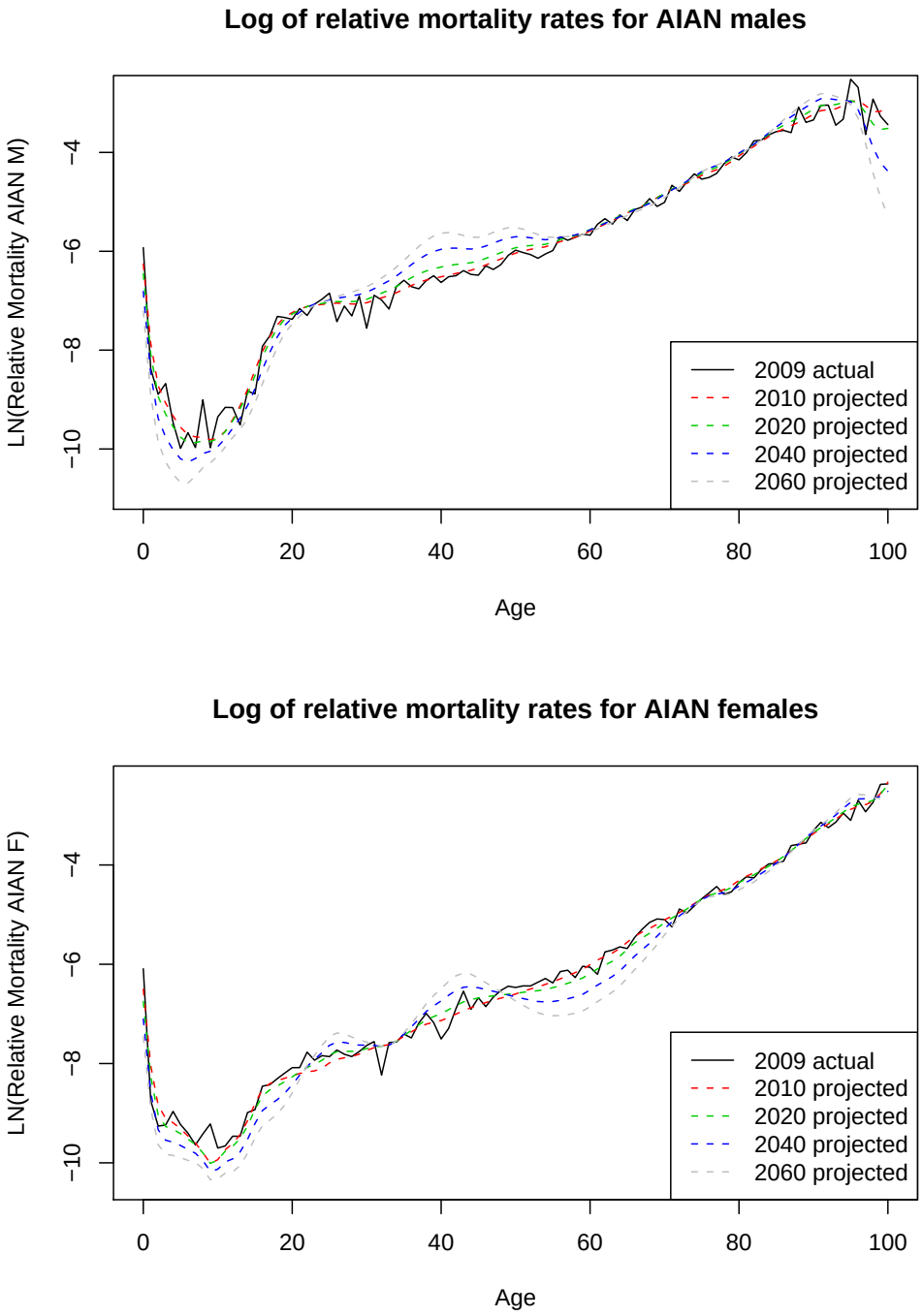


Figure 18: *Log of actual 2009 distribution curve of age-specific mortality rate for American Indian & Alaska native along with projections for years 2010, 2020, 2040, and 2060. Top shows male rates, while bottom shows female rates.*

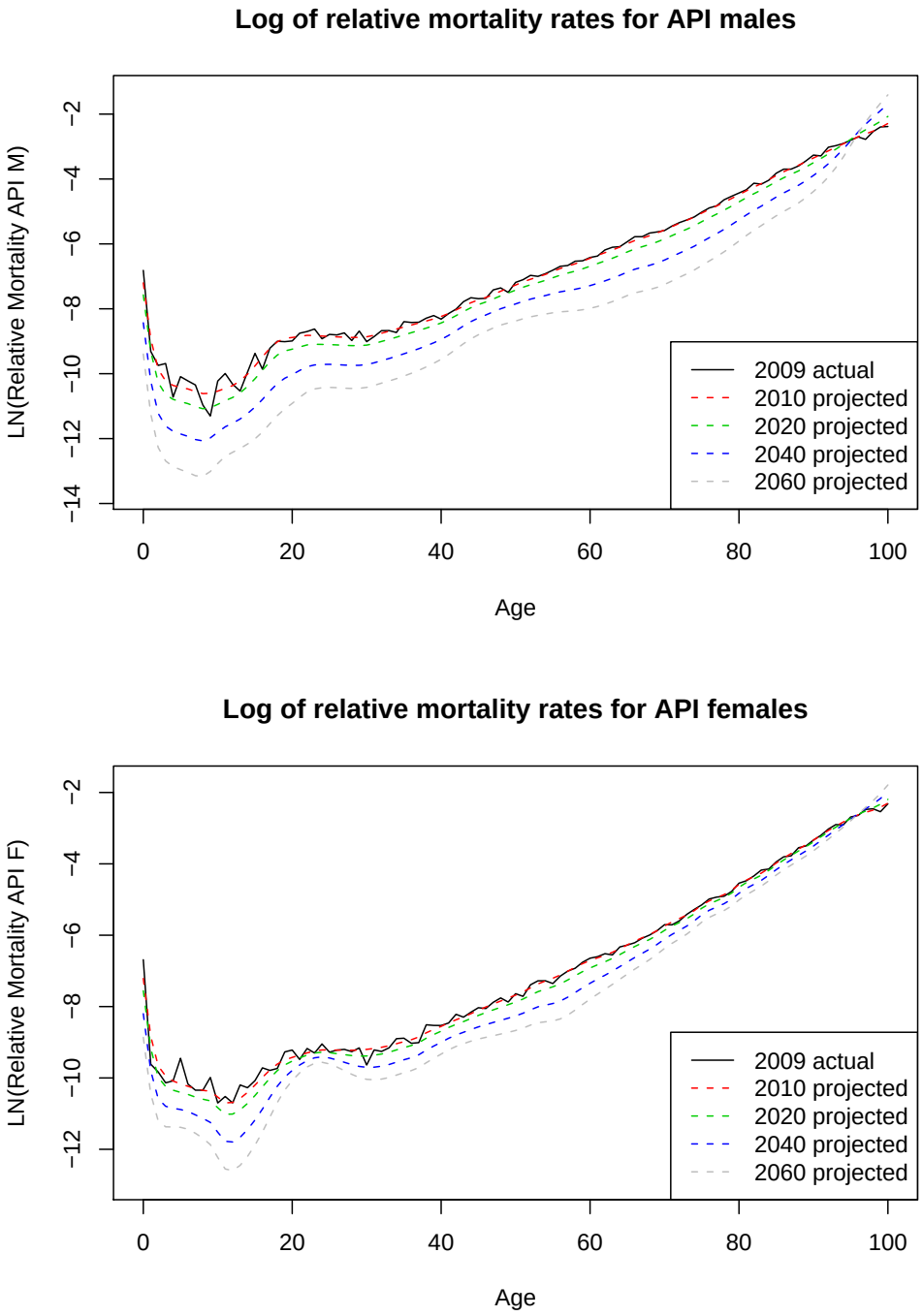


Figure 19: *Log of actual 2009 distribution curve of age-specific mortality rate for Asian & Pacific islanders along with projections for years 2010, 2020, 2040, and 2060. Top shows male rates, while bottom shows female rates.*

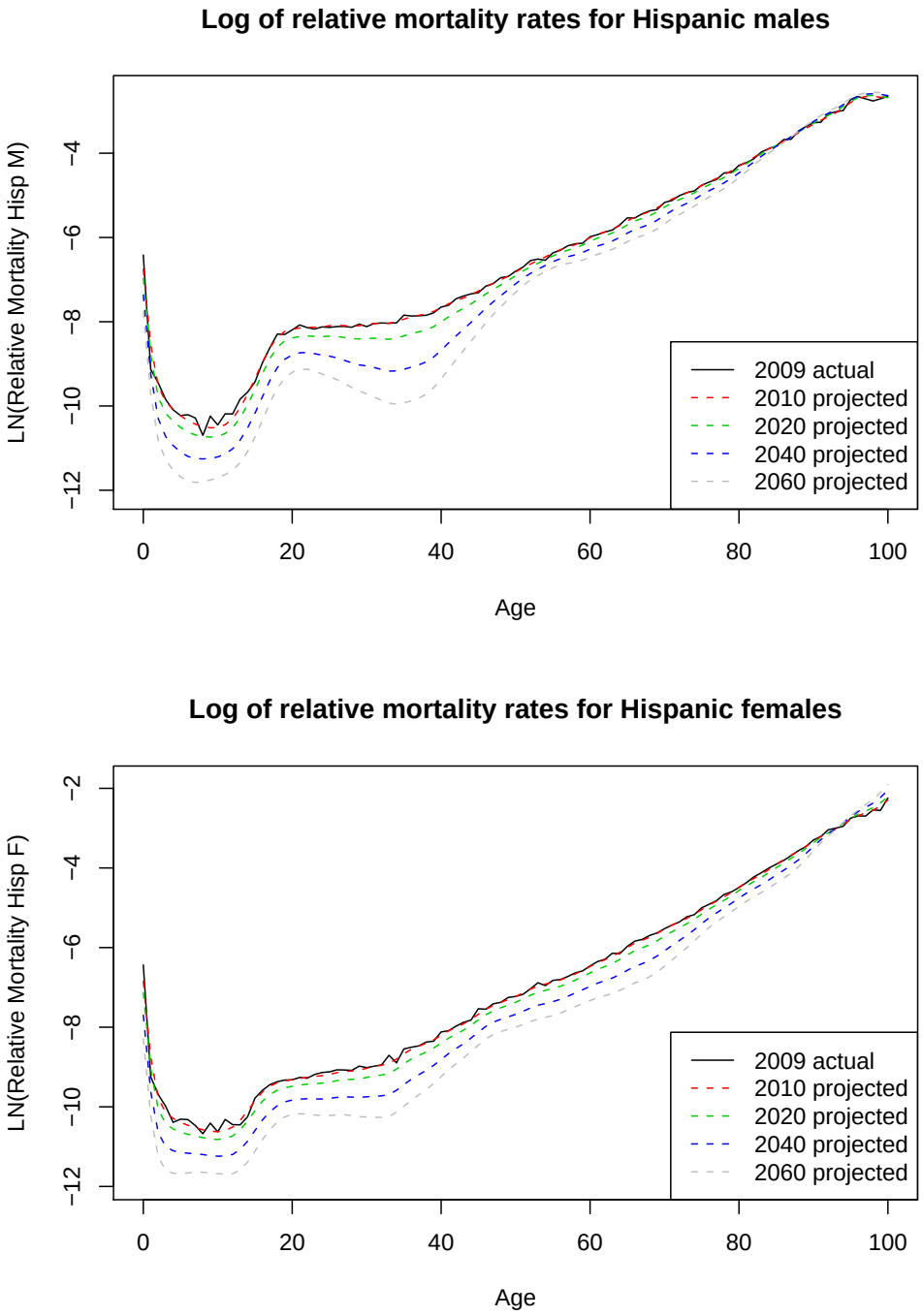


Figure 20: *Log of actual 2009 distribution curve of age-specific mortality rate for Hispanics along with projections for years 2010, 2020, 2040, and 2060. Top shows male rates, while bottom shows female rates.*

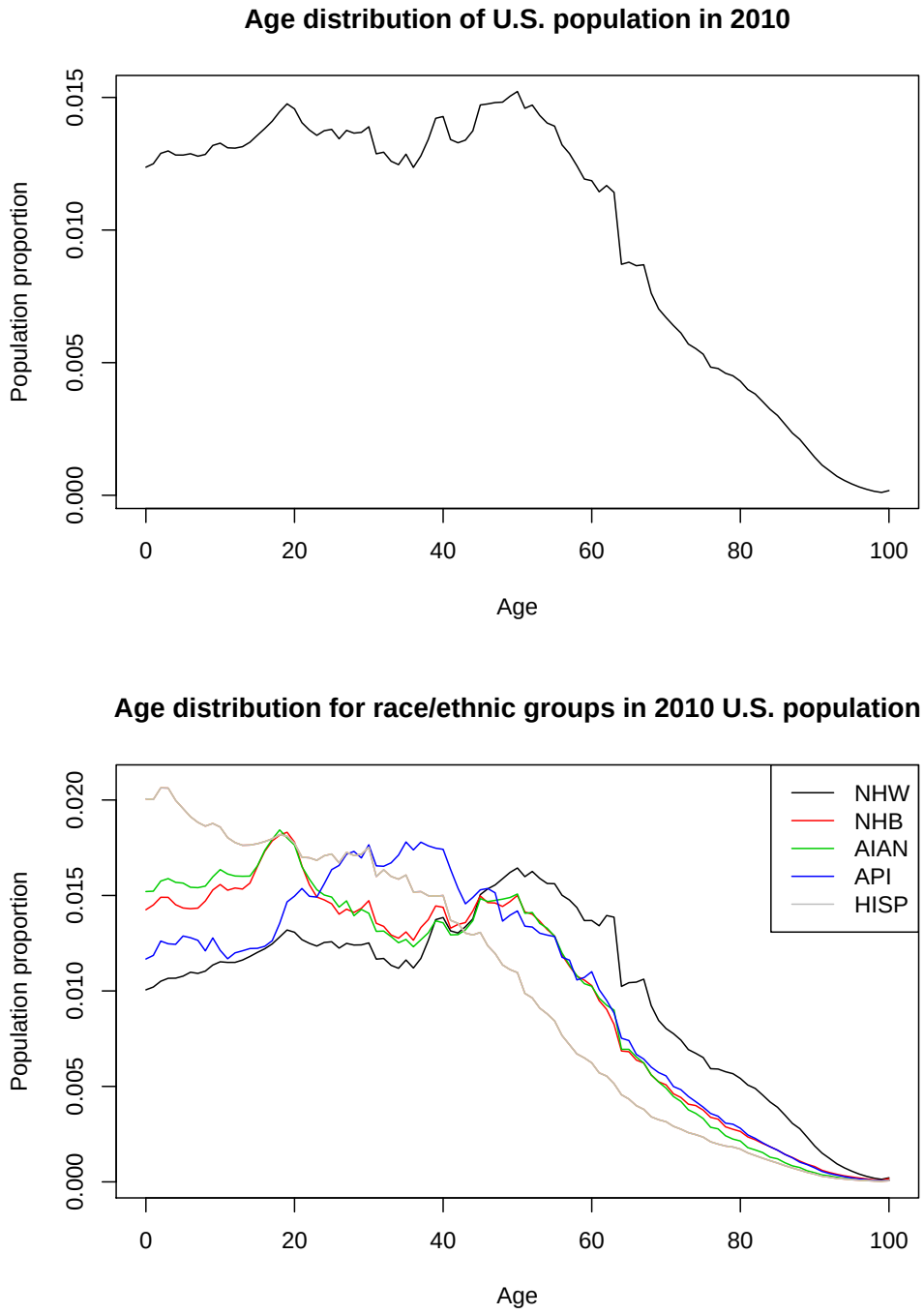


Figure 21: *Age distribution of 2010 U.S. population. Overall population is depicted on top, while the bottom displays for individual race/ethnic groups.*

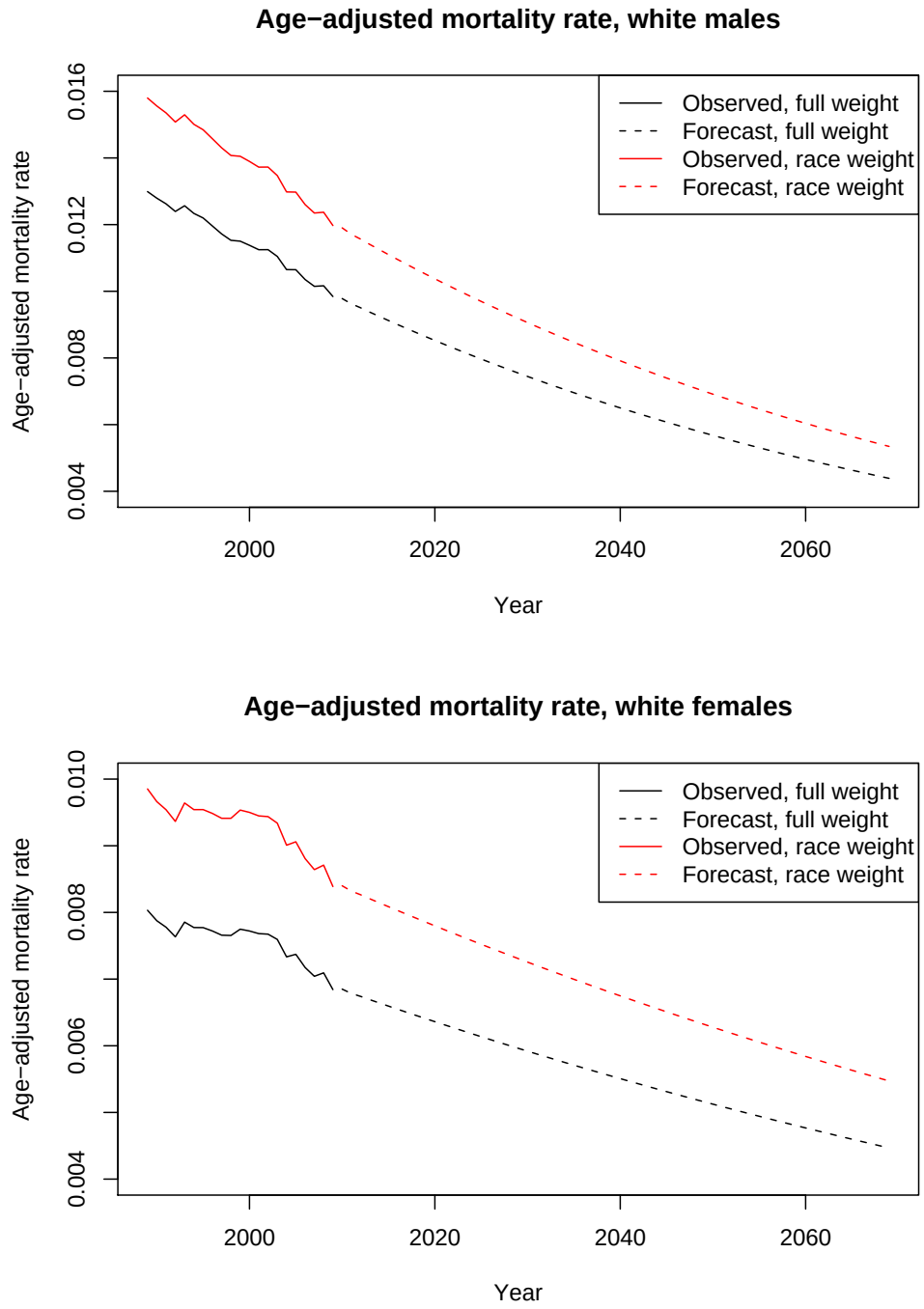


Figure 22: *Age-adjusted mortality rate for non-Hispanic whites (top - male; bottom - female) using weights based on U.S. population (black) and white population only (red).*

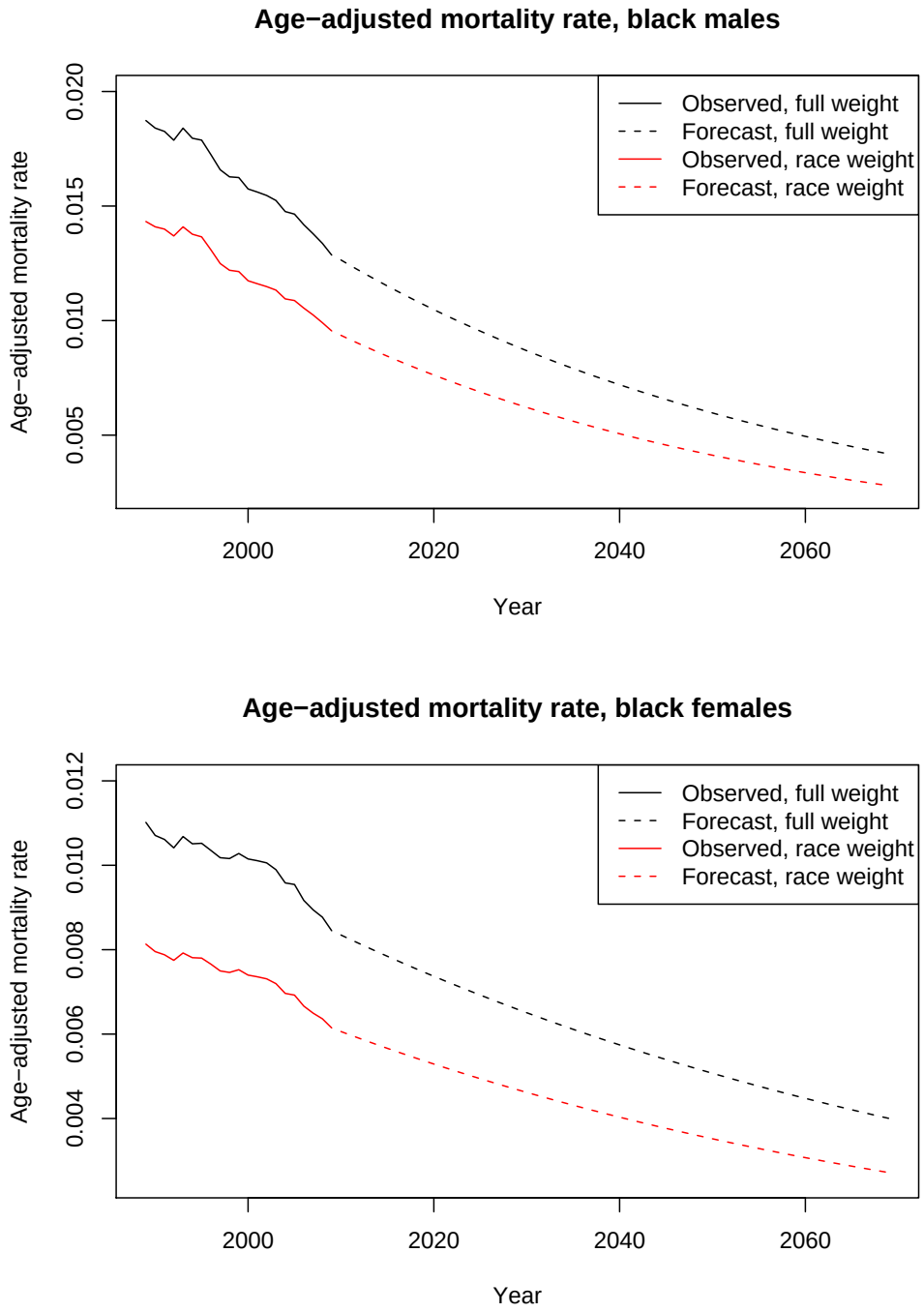


Figure 23: *Age-adjusted mortality rate for non-Hispanic blacks (top - male; bottom - female) using weights based on U.S. population (black) and black population only (red).*

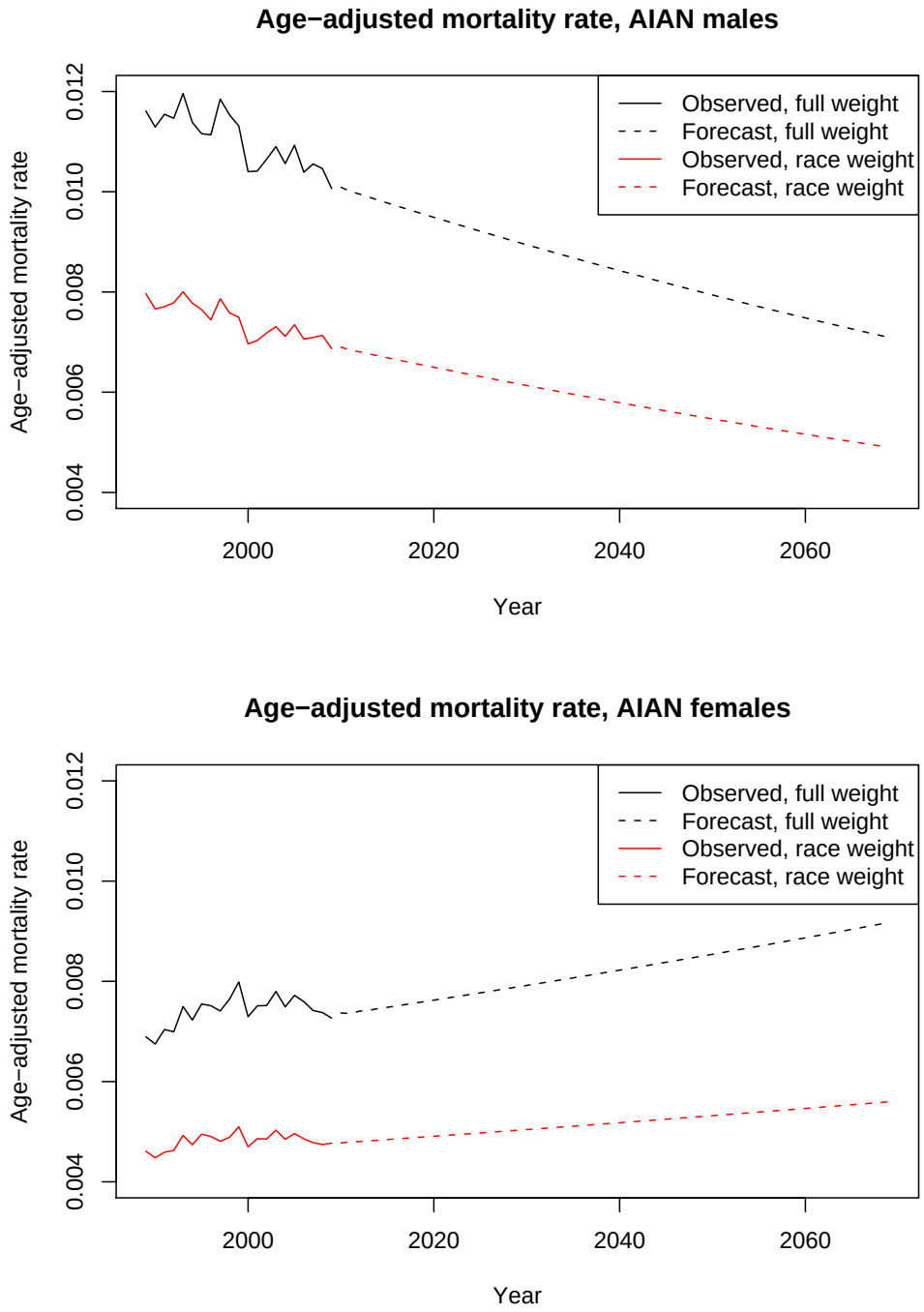


Figure 24: *Age-adjusted mortality rate for American Indian & Alaska natives (top - male; bottom - female) using weights based on U.S. population (black) and AIAN population only (red).*

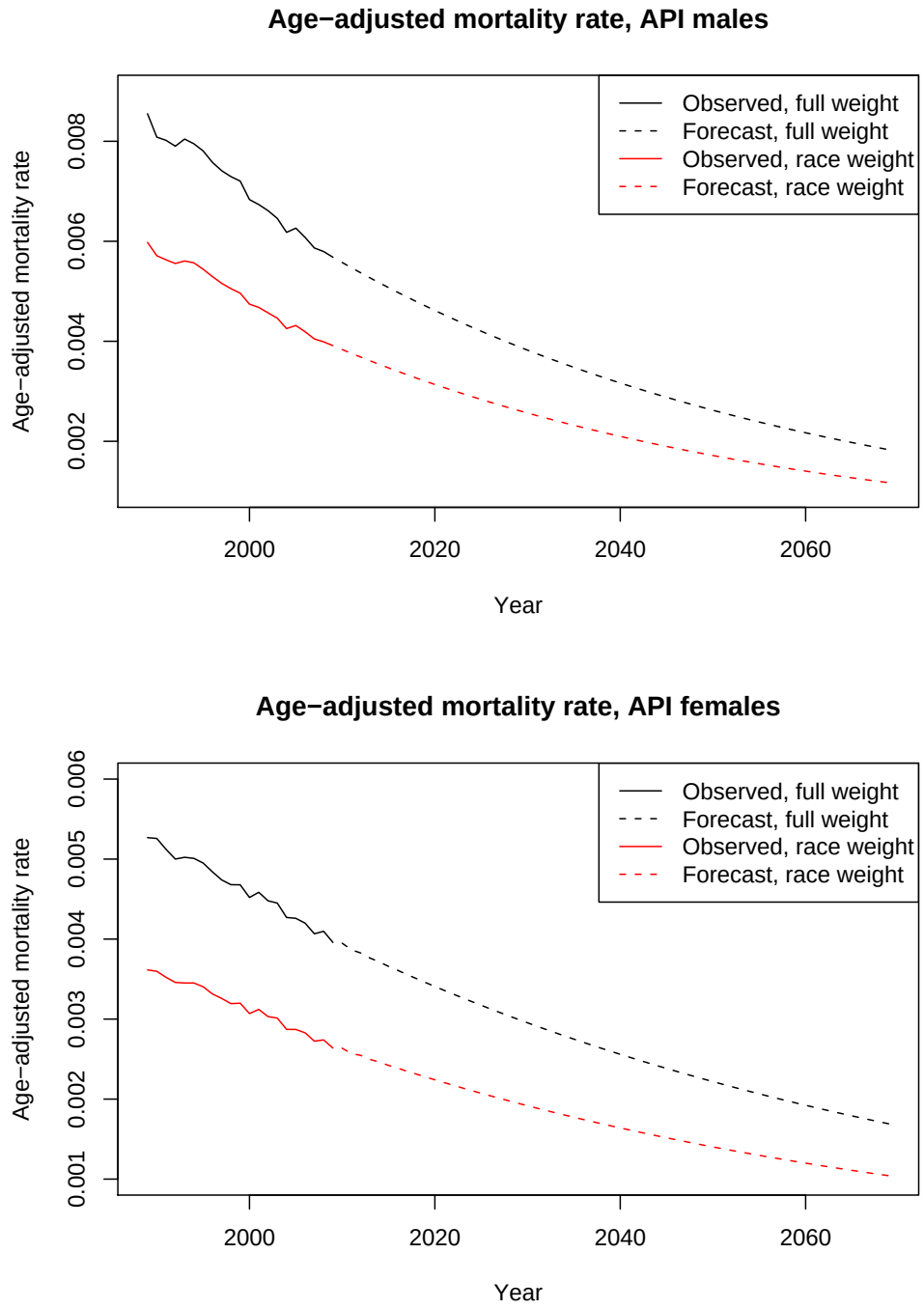


Figure 25: *Age-adjusted mortality rate for Asian & Pacific islanders (top - male; bottom - female) using weights based on U.S. population (black) and API population only (red).*

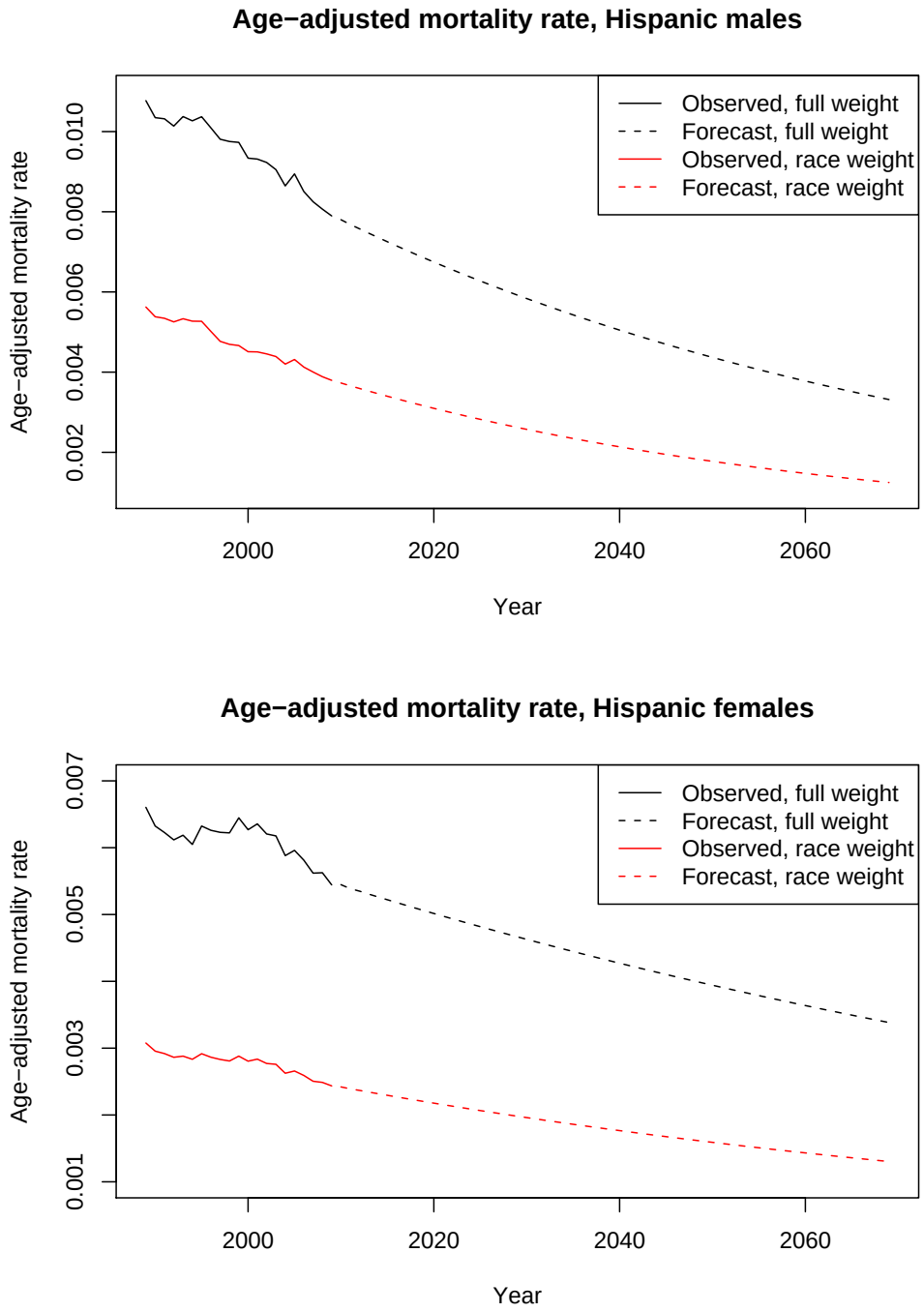
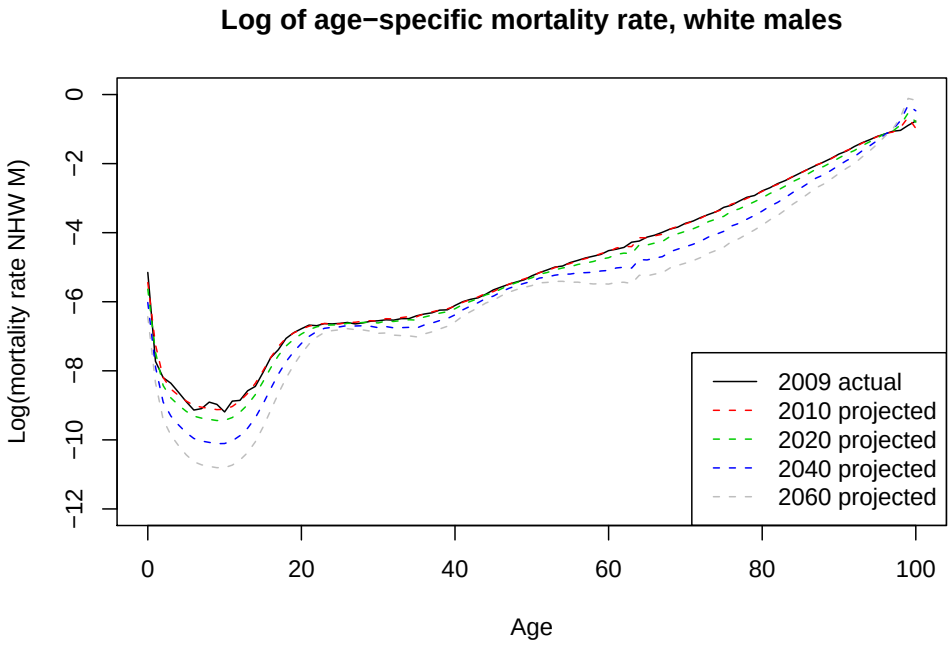


Figure 26: *Age-adjusted mortality rate for Hispanics (top - male; bottom - female) using weights based on U.S. population (black) and Hispanic population only (red).*



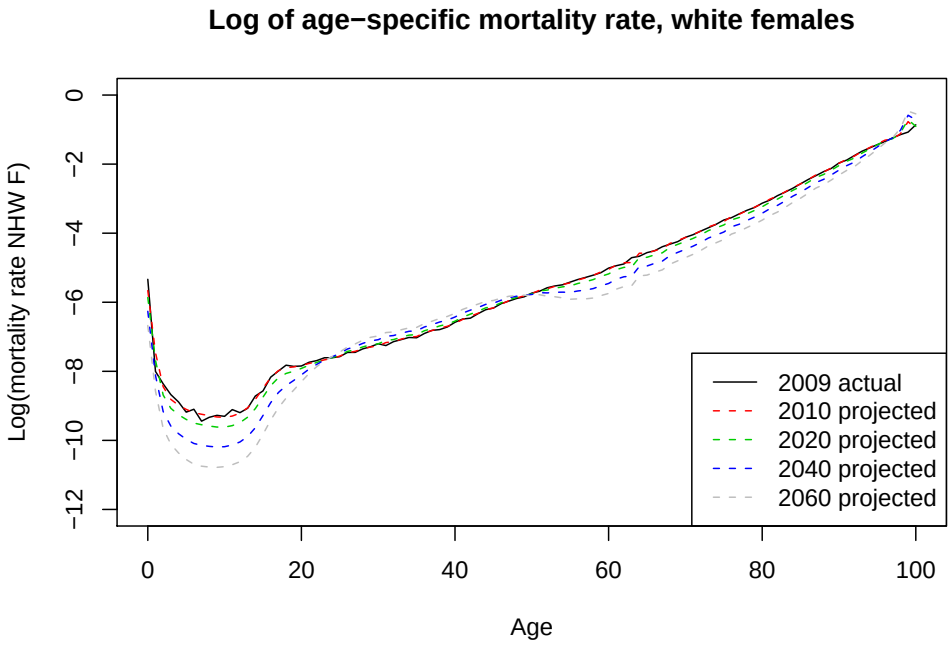


Figure 27: *Log of age-specific mortality rate produced using white population proportions for whites along with forecasts. Top shows males, while bottom shows females.*

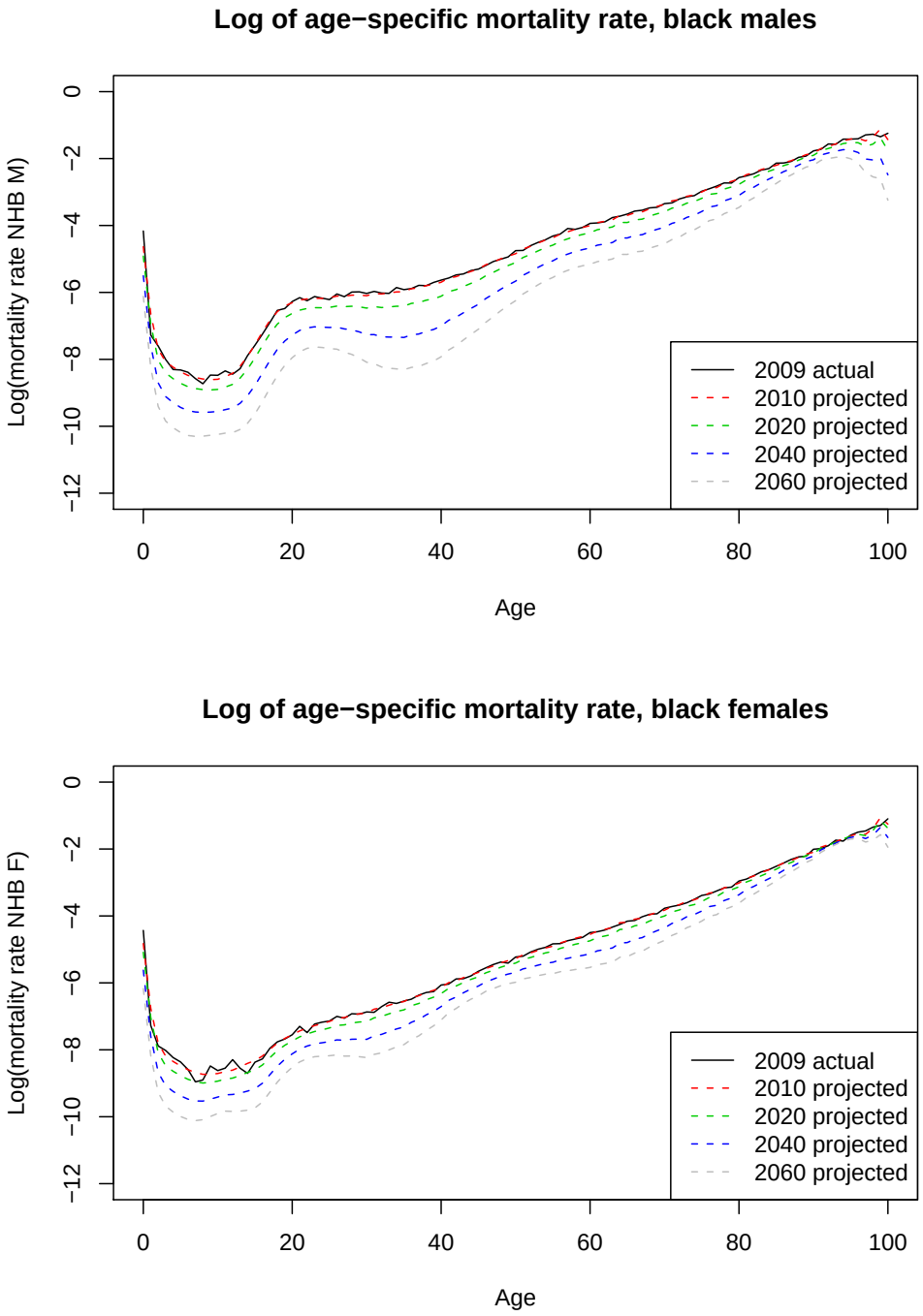


Figure 28: *Log of age-specific mortality rate produced using black population proportions for blacks along with forecasts. Top shows males, while bottom shows females.*

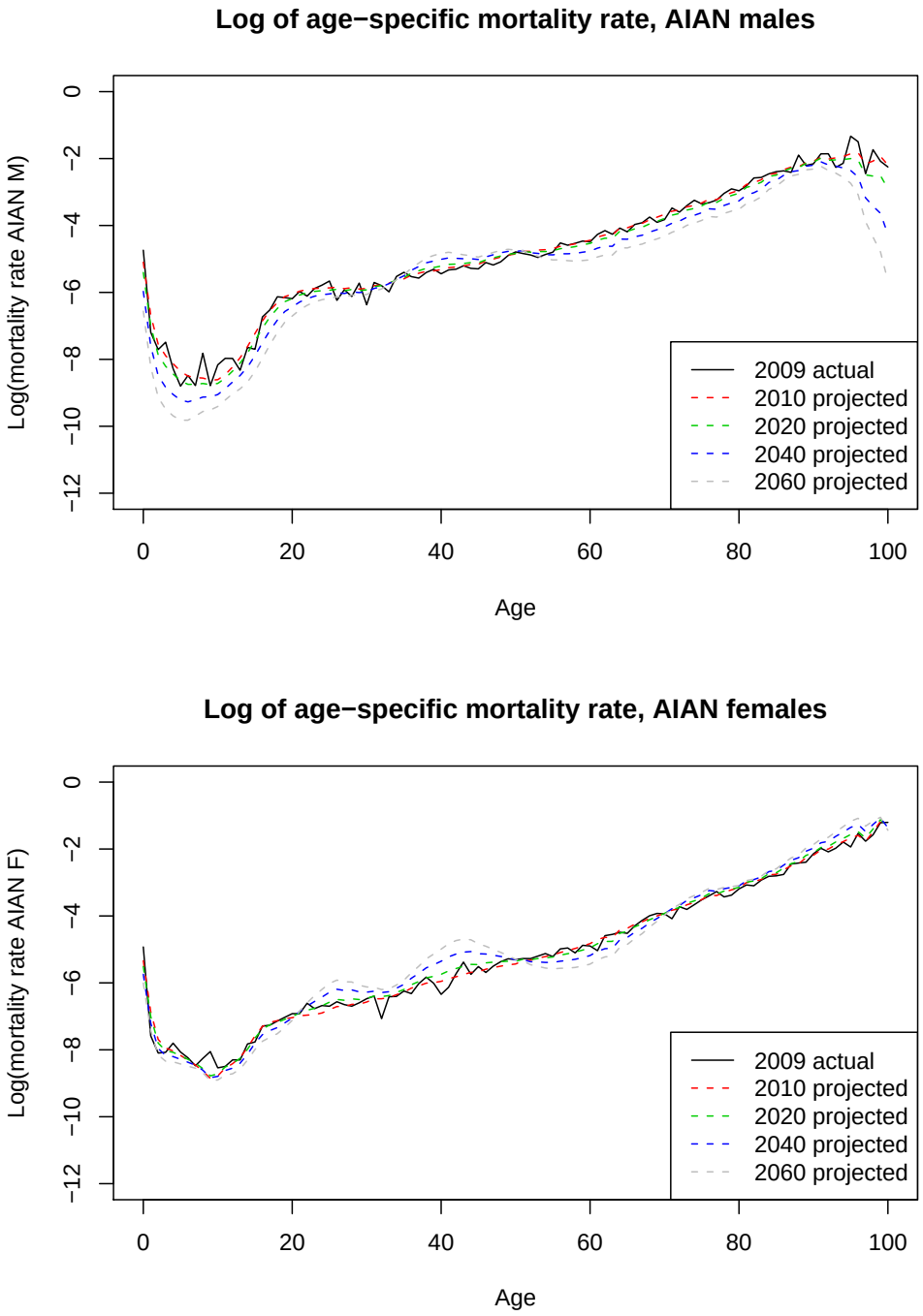


Figure 29: *Log of age-specific mortality rate produced using AIAN population proportions for American Indians & Alaska natives along with forecasts. Top shows males, while bottom shows females.*

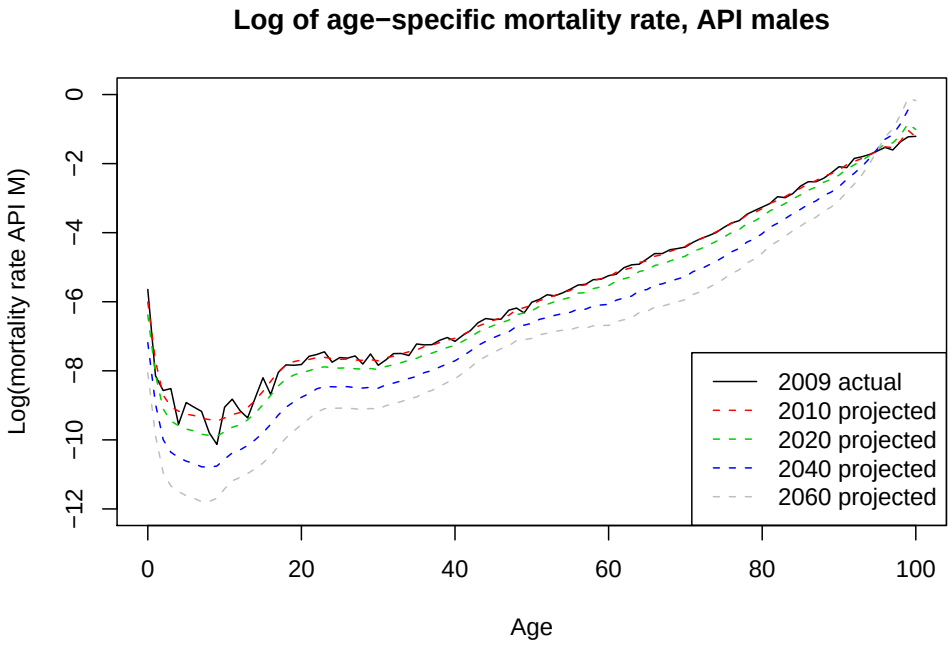
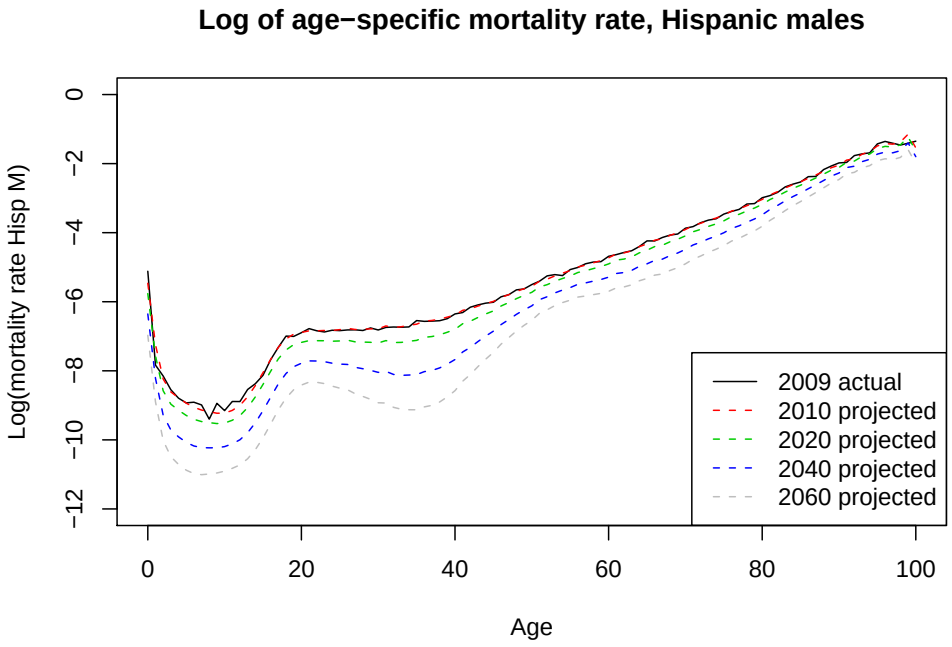


Figure 30: *Log of age-specific mortality rate produced using API population proportions for Asians & Pacific islanders along with forecasts. Top shows males, while bottom shows females.*



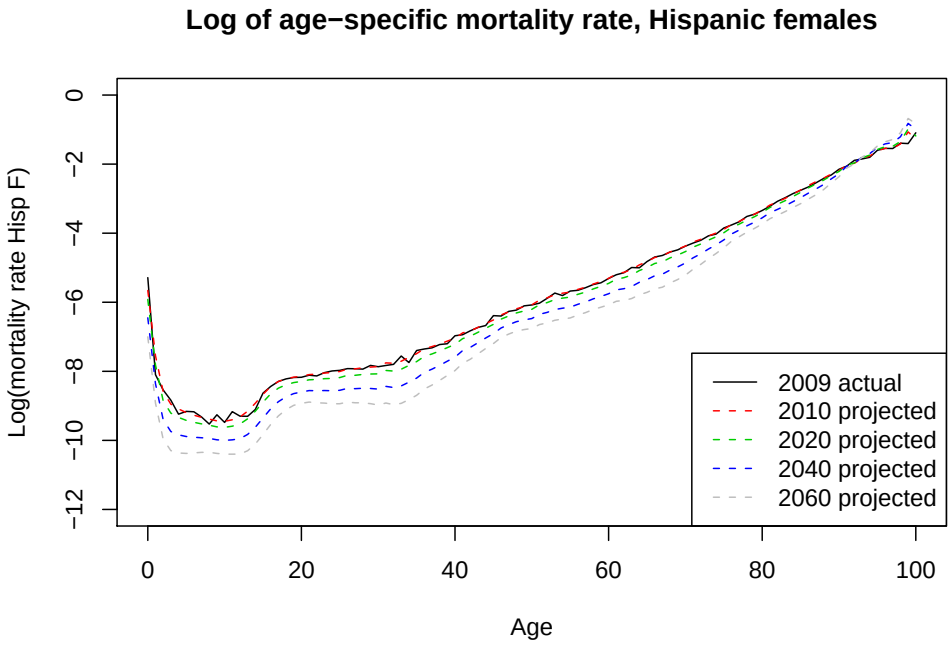


Figure 31: *Log of age-specific mortality rate produced using Hispanic population proportions for Hispanics along with forecasts. Top shows males, while bottom shows females.*