University of Michigan, Department of Statistics

Applied Qualifying Examination 2020

U.S. Mortality 2007-2018: Identifying Trends and Patterns in Mortality Trajectories

Simon Fontaine

University of Michigan, simfont@umich.edu

Examiners:
Kerby Shedden
Gongjun Xu
Ji Zhu

Contents

1 Introduction

Longitudinal analysis of mortality rates can provide significant insight to multiple levels of public officials. Indeed, observing an increase in mortality rate within some specific demographic group indicate that more attention has to be devoted to that group and its mortality patterns. In particular, when such an increase is attributed to a specific cause, this can warrant and guide targeted intervention.

For example, studies [Masters et al., 2017, Monnat, 2017, Woolf and Schoomaker, 2019] have found recent increases in mortality rates due to suicides, drug poisonings and alcohol-related illnesses, especially among middle-aged Americans. In turn, these mortality causes have been shown to be associated with unemployment [Institute for Work & Health, 2009, Henkel, 2011] whose rates reached historical highs during the Great Recession (2007-2009): Margerison-Zilk et al. [2017] find increased morbidity, psychological distress and suicide rate in the years following the event. The authors also claim that the U.S. was more severely impacted than European countries with stronger social safety nets. With the current COVID-19 epidemic, one can anticipate a second wave of increased mortality—this time not due the disease itself, but rather to its economic impacts. These considerations thus instruct increased government intervention in economic recovery as well as in mental illness prevention and treatment. Additionally, the ongoing opioid epidemic in the U.S.—possibly in conjunction with the economic situation—is a major contributing factor in the increase in mortality due to drug overdose and may call for improved detection, prevention and treatment as well as further substance control both within the U.S. and at its borders.

In order to act on these issues, they must first be detected. While time series of mortality rates can be particularly noisy, especially when considering smaller demographic groups or smaller mortality rates, it is important to be able to extract trends and patterns among them to conclude some variation is associated with a cause or a demographic. In this analysis, we analyze recent U.S. mortality data aggregated by sex and age to study what trends and patterns can be identified in mortality trajectories. Since all mortality causes are aggregated, we cannot directly confirm results such as those presented above: our analysis can be understood as a blueprint for de-noising mortality time series and the same methodology could be applied to finer data such as mortality rates per cause or by ethnicity and race. As a sanity check, we attempt to correlate our findings with those presented in literature.

1.1 Data Overview

We consider U.S. mortality rates from 2007 to 2018 extracted from the National Vital Statistics System [2020] and kindly aggregated by age group and sex by the examiners. The data set contains monthly death counts as well as yearly population estimates of each demographic group who are defined by sex (Female or Male) and age group (five years intervals for age 0 to 84 and 85 to 99). Hence, we have 72 time series (2 sexes $\times$ 18 age groups $\times$ 2 measurements), 36 of which are of length 144 (12 years $\times$ 12 months) for death counts and 36 of length 12 (12 years) for population estimates. We report no missing data. Figure 1 depicts the log mortality rate for those 36 demographic groups along the period of interest.

1.2 Research Questions

Our main research question is to identify what trends and patterns can be extracted from monthly mortality trajectories. We are particularly interested in building a model for deaths counts including

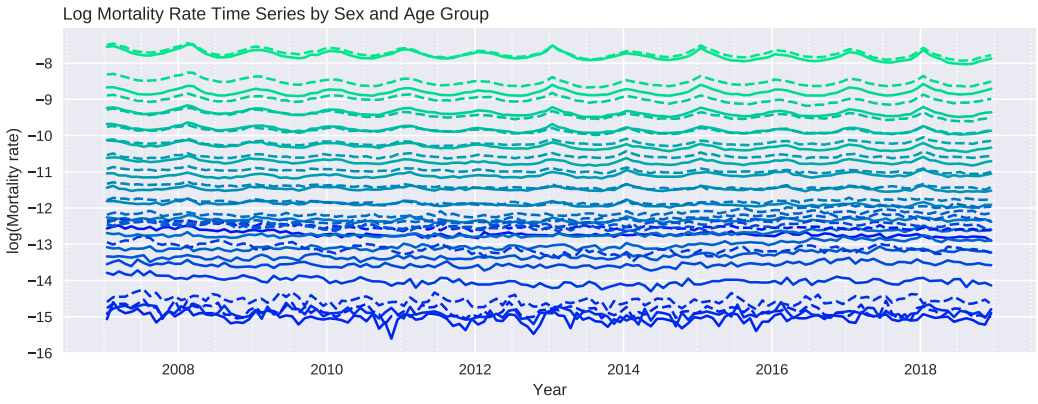

Figure 1: *Log mortality rate (using the exposure described in the text) by sex and age groups. Solid lines corresponds to females and dashed lines to males. The color gradient defines the sequence of age groups from 0-4 years old (blue) to 85-99 years old (green).*

estimated population, time effect, month effect and demographic information. The time effect will inform us on the average mortality rate of a demographic as well as extracting the trend over time of that rate; the month effect can tell us about seasonal patterns. The demographic information then allows us to model how these trends and patterns change with respect to sex and age group.

Emerging from the main question of interest are some secondary modeling consideration. First, death counts correspond to large positive integer data and many distribution assumptions could be used for our model: appropriate selection and diagnostic has to be performed. Second, the way in which each covariate (time, month, demographic group) influences the response and interact with each other has to be specified and selected.

Finally, we wish to learn a low-dimensional representation of mortality time series enabling us to easily describe and compare different time series: complex interaction between the different covariates may prevent us from extracting relevant insights from simply inspecting the marginal effect of each covariate.

2 Methodology

2.1 Data Pre-processing

Age Group Midpoints. Within a given age group, the distribution of ages may not be evenly distributed. In particular, older age groups will contain more people in the lower end of the interval: for example, we expect much more 85 years olds than 99 years olds in the 85-99 age group. Since some of the model we will consider use the age groups as a numerical value, we cannot simply use the midpoint of each interval as it would bias the actual average age within a group. To account for that fact, we will rather use a midpoint computed using linear interpolation. For each month, year, sex and age group, we identify a trend as the slope between the previous age group and the next age group (for fixed month, year and sex) and assume the current group's distribution follows the same slope. For the 0-4 age group, we set the previous age group's population as the same as group 0-4; for the 85-99 age group, we set the next age group's population to 0. The process is repeated for every month and year and by sex. In practice, the estimated midpoint for age groups up to 45 years old are

Age Groups Midpoint Estimation										
Age group	...	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85-99
Midpoint	...	47.00	52.00	57.00	62.00	67.00	72.00	77.00	82.00	92.00
Estimated	...	47.03	51.97	56.89	61.81	66.74	71.72	76.71	81.77	89.33

Table 1: *Average estimated midpoints per age group. Groups below 45 years are estimated almost exactly to the true midpoint of the corresponding interval.*

virtually the same as the true midpoint (the age pyramid is rather flat on that range); Table 1 contains the average (over all 144 months and sexes) estimated midpoint for each age group. We experimented with naive and estimated midpoints and found that the estimated ones perform slightly better in our models (not reported for brevity).

Estimated Monthly Population. The original data contains only the *yearly* estimated population in each demographic group. However, we have access to *monthly* death counts: computing the mortality rate using the yearly population may not be an accurate measurement. Instead, we will estimate a monthly population within each demographic group using linear interpolation. For a fixed sex and age group, we set the population of month 6 of every year to be the yearly population and interpolate linearly between those 12 points; the monthly population beyond those points is set to be constant. Figure 2 shows an example of the interpolation performed for females between 0 and 4 years old. The result is a continuously varying population and attenuates drastic changes in population: it is highly unlikely that the population of females between 0 and 4 years old dropped by over 500,000 on 2012's New Year Day. We experimented using the estimated yearly and monthly populations in our models and found that the monthly population exhibited marginally improved performance (not reported for brevity). From here on, *population* will refer to estimated monthly population without ambiguity.

Exposure and Daily Mortality Rate. In the first iterations of the current analysis, we identified a pattern of decrease in mortality rates in the month of February which was trivially explained by the shorter length of that month. Indeed, with less time, we expect less people to die. This justify to rather consider *daily* mortality rates instead of monthly rates. To achieved this, we compute the monthly *exposure* of a population as the product of its population and the length (in days) of that month. From here on, *mortality rate* refers to the death count divided by the exposure of that demographic group during a given month.

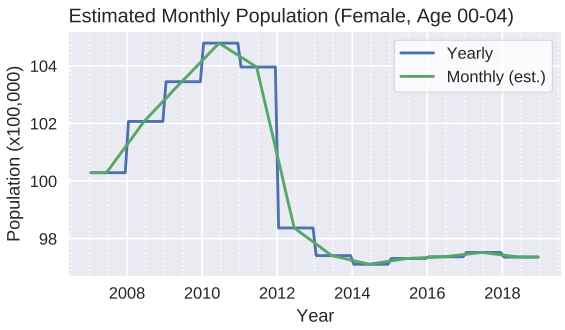

Figure 2: *Example of the linear interpolation estimation of the monthly population described in the text.*

2.2 Data Exploration

In order to summarize mortality trajectories for analysis and comparison, we consider a *Principal Component Analysis* (PCA) of the 36 log mortality rate time series. The study of the most important

components can unveil what properties of a trajectory are useful at describing it by capturing most of its variations. Then, comparing the scores of the time series by sex and age group can guide us toward a model description of the interactions between demographic and time. This will eventually help us defining the mean model requiring the specification of how covariates act on the response and how they interact with each other.

Furthermore, the same process can be repeated on the fitted values emanating from the selected model. Indeed, we can reconstruct the de-noised time series and perform PCA on them. This will tell us what our model understand from the collection of time series and the comparison between demographic groups will then allow us to extract insight on how mortality trajectories differ between groups when noise has been removed.

Note that we choose to act on the logarithm of the mortality rate for multiple reasons. First, the models we consider all model the relationship between the death count, exposure and covariates through a log link so performing PCA on the log rate is sensible for consistency. Second, it would be unwise to perform PCA on the raw counts as there is more variability in larger population groups and in groups with larger mortality rate. Third, since rates are proportions by nature, it would also make sense to use a logit transformation instead: however, all observed rates are close to 0 (the largest is 0.017) where the two functions are almost identical.

2.3 Mortality Rate Modeling

To model death counts using exposure, time, month, sex and age group, we consider *Generalized Linear Models* (GLM). We restrict our study to GLM families modeling non-negative data since death counts are non-negative integer data. Also, to account for the exposure, we consider only mean models using the log link which, effectively, implies that we model the mortality rate and not the counts directly.

Since our response is integer-valued, there are two obvious candidate families: Poisson and Negative Binomial (NB). Both distribution allow for non-negative integer response with unbounded domain. The main difference between the two families is that the Negative Binomial distribution is more suitable to over-dispersed counts since its mean-variance relationship is quadratic in the mean instead of linear (Poisson).

While our response is integer-values, right-continuous distributions are also relevant to our analysis. Indeed, the observed count are large (the smallest being 51) so using a continuous family does not suffer from the discretized values. Thus, we consider a Normal GLM with log link which has variance exponential in its mean and can then model severely over-dispersed data. Finally, we also consider a Tweedie model with power 1.5 which can be interpreted as halfway between a Poisson model and a NB model. Indeed, a Tweedie with power 1 is exactly Poisson and a Tweedie with power 2 is Gamma which closely resembles a continuous version of the NB distribution. A Tweedie (1.5) model has variance proportional to the 1.5th power of the mean.

Studying the fit of each model to the data allows us to study the over-dispersion of the death counts. In increasing order of over-dispersion, we have: Poisson, Tweedie(1.5), NB and Normal (log link). We inspect the Pearson residuals normalized by the estimated scale (Pearson's χ^2/degrees of freedom): better fit can be identified through residuals with constant variance along fitted means values.

3 Results

All `Python` code for data processing, analysis, producing tables and figures can be found at:

https://github.com/fontaine618/qr.

Data pre-processing and manipulation was performed using `pandas v1.0.3` [McKinney, 2010] and `numpy` v1.18.2 [van der Walt et al., 2011], PCA using `scikit-learn v0.22.2.post1` [Pedregosa et al., 2011], GLMs using `statsmodels v0.11.1` [Seabold and Perktold, 2010] and `patsy v0.5.1` [Smith, 2015] and plotting using `matplotlib v3.2.0` [Hunter, 2007].

3.1 Analysis of Observed Mortality Rates

Visual inspection of the log mortality rates (Figure 1) show some clear trends and patterns: increasing age is related to increased mortality rates and to larger seasonal effects; there appear to be some differences between males and females; young children (0-4 years olds) have higher mortality rates than older age groups (5+ years old); etc. Performing PCA on these 36 time series yields the results in Figure 3.

The inspection of the mean (Figure 3, 4th row) shows a clear average evolution over time—decreasing between 2007 and 2013 and increasing afterwards—as well as a clear seasonal effect—high for winter months and low for summer months. Both effects appear to be rather independent as the seasonal trend seems relatively constant through years. This indicates that our model should contain the main effect of time and of months.

The first principal component (Figure 3, 1st row), which explains 99.9% of the variance seem to capture most of the average rate over time as well as a corresponding adjustment of the seasonal effect: positive first components are associated with larger rates and stronger seasonal effect; negative first components are associated with smaller rates and seasonal effect in the opposite direction of that of the mean, that is, weaker seasonal effect when summing the two. Except for the 0-4 age group, we see a monotone increase in first component with age. These observations imply we should include the main effect of age, possibly interacted with sex, as well as the interaction between month and age (and possibly sex).

The second component (Figure 3, 2nd row), which explains 46% of the remaining variance (removing the effect of the first component), seem to capture the trend over time as well as some correction to the seasonal effect. Positive second components show a relatively linear increase over time and are observed for people between 20 and 40 years old. Negative second components show a somewhat linear decrease over time and are observed mostly for people under 20 years old. This indicates we should consider the interaction between time and age (and possibly sex and month, though the effects are less obvious).

The third component (Figure 3, 3rd row), which explains 14% of the remaining variance (removing the effect of the first component) seem to capture curvature over time and some seasonal effect. Positive third components, observed for males in the 15-19 age group, show a positive curvature while negative third components, observed for females of the 5-9 age group, show a small negative curvature over time. This also indicates we should include the interaction between time, age and sex.

Overall, we consider the following model, which contains all effects mentioned above,

$$\log \mathbb{E}\left\{\frac{\texttt{Deaths}}{\texttt{Exposure}}\right\} = \texttt{Age} * (\texttt{Date} + \texttt{Sex} * \texttt{Month}),$$

where $*$ denotes the inclusion of both main effects as well as the two-way interaction. In particular, we find no indication that the seasonal effect changes over time. In the next section, we use this model in order to de-noise the time series in order to get a clearer version of Figure 3.

3.2 Mortality Rate Modeling

Along with the four GLM families we consider, we also investigate different models for the marginal effects of the covariates. In the case of the age group, we either model it continuously using the midpoints with 9 cubic spline (denoted `sA`) or using the groups as categories (18 levels, denoted `cA`). For the time effect, we either model it using the date with 50 cubic splines (denoted `sD`) or using the year using 7 cubic splines (denoted `sY`). For the effect of months, we either consider 7 cubic splines (denoted `sM`) or as categories (12 levels, denoted `cM`). Finally, the sex covariate (denoted `S`) only admits two levels and any modeling will be equivalent. We therefore have 8 declinations for each family for a total of 32 models.

Within family, we select the best model using a combination of log-likelihood, AIC and BIC: we choose the formula where all three statistics are among the bests. The results for all families and formulas are in Table 3; the selected models are described in Table 2. The selected formulas are almost identical except that the Normal model with log link and the Poisson model uses months as a categorical variable instead of using splines; age is modeled using splines and time is modeled using splines on the date for all four families.

Log Mortality Rate: Best model by family

Family	Formula	Df Model	MSE	MSEL
Normal (log link)	`sA*(sD+S*cM)`	737	75256	0.00198
Poisson	`sA*(sD+S*cM)`	737	75740	0.00167
Negative Binomial	`sA*(sD+S*sM)`	657	77497	0.00171
Tweedie (1.5)	`sA*(sD+S*sM)`	657	76720	0.00172

Table 2: *Results for the selected model within each GLM family. See text for a description of the formulas. The MSEL is the mean squared error between the log mortality rates and the log predicted rates.*

All four families perform rather similarly in terms of prediction error. Indeed, both the mean squared error (MSE) and the mean squared error of the logs (MSEL) do not vary much across families. To select a model between those four candidates, we investigate how they model the over-dispersion of the data. Figure 4 depicts the normalized Pearson residuals along the fitted means. The Normal model with log link and the Poisson model perform badly for large fitted mean, which can be seen by large residuals at the right end of the domain; the NB model has larger residuals for smaller fitted means; the Tweedie (1.5) model has residuals with fairly constant variance throughout the domain. For this reason, we choose the Tweedie (1.5) model for our final model as it fits death counts uniformly well.

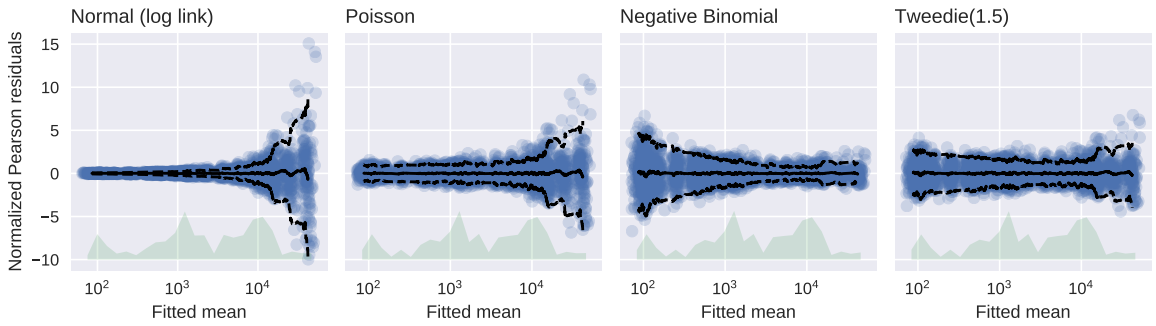

Figure 4: *Pearson residuals (blue) normalized by the estimated scale; running mean (solid, black) and* ± 1.96 *running standard deviations from the mean (dashed, black); distribution of the fitted means (green).*

3.3 Analysis of Predicted Mortality Rates

Figure 5 contains the same procedure that was applied to the observed log mortality rates of Section 3.1, but now performed on the fitted values resulting from our selected model. Overall, we observe the exact same trends and patterns, but the seasonal effects are now much clearer to see. Also, the associations with sex and age group are almost identical. For the mean, we still observe high mortality rates in the winter, but we now also see a small jump in June and July, relative to other summer months, that was not as obvious in the noisy data. In the case of component two, positive values, associated with increase over time, seem to remove the drop in mortality rates observed around September; negative values, associated with decrease over time seem to exacerbate that same drop. For component three, positive values seem to strengthen the time and seasonal effects observed in the mean while negative values tends to attenuate those same effects.

4 Discussion

4.1 Variations across Demographic Groups

Consolidating the observations regarding the principal components in Figure 5, we can finally compare the mortality trajectories between demographic groups defined by sex and age.

With respect to average mortality, we find that mortality rate increases with age, which is not surprising at all due the the nature of life itself. Also, we find that the 0-4 years old demographic exhibit large mortality rate as well, which could be explained by infant mortality and children diseases that do not occur for other age groups. We also point out that males between 15 and 39 show increased mortality rate compared with females of the same age: this effect is most likely due to increased mortality caused by of accidents, substance abuse and homicides within that demographic [Masters et al., 2017, Woolf and Schoomaker, 2019].

With respect to change over time, we find that the overall mortality rate decreased from 2007 to 2013 and increased between 2013 and 2017. During those years, however, we find that people under the age of 20 have shown a small decrease in mortality rate with respect to the mean trend and that people between the age of 25 and 39 have shown a larger increase in mortality compared to the mean. The decrease for children is partly due to a decrease in infant mortality [Callaghan et al., 2017] and to other

factors such as decrease in traumatic brain injuries [Cheng et al., 2020] while the increase for middle-aged people can possibly be attributed to the increased in suicides and substance abuse-related deaths discussed in the introduction. Another factor contributing to that increase is the relative increase of deaths related to obesity [Preston et al., 2018].

With respect to seasonal effects, the picture is slightly more complicated as the mean and all three components capture part of that effect. The overall effect is higher mortality rates during winter month (December-March) and lower mortality during the rest of the year with a small jump during June and July, but that effect seems only apparent for elderly people. There are many possible factors explaining this effect. First, winter months corresponds to flu season and flu-related death are majoritively observed within the older population [Center for Disease Control and Prevention, 2020]. Similarly, cardiovascular-related mortality was shown to peak during winter [Stewart et al., 2016]. Second, June and July corresponds to some of the warmer months of the year and the elderly show increased risk heat-related illnesses during heatwaves [Cheng et al., 2018].

4.2 On the Analysis

A major weakness of our modeling approach is the failure to account for time dependency. Indeed, mortality time series are inherently auto-correlated since a very large fraction of each demographic group is contained in consecutive months. The GLM approach considered here implicitly assume conditional independence of the count given the covariates. Therefore, we would need that, given the time (year and month) and given a demographic group and its exposure, the death count is independent from any other configuration. One way to check if this assumption is reasonable for our data is to look at the time series of the Pearson residuals. Under the assumption that our model is correct, the Pearson residuals should have constant variance and be uncorrelated. Hence, the time series of Pearson residuals should resemble a white noise series. Inspection of the auto-correlation function of all 36 series of residuals, shown in Figure 6, indicates that there remains some very minor auto-correlation. Some series show stronger correlation on small lags and others show some residual seasonal effect, but the overall trend does not show severe problems. This further indicates that we have successfully extracted most of the signal in these 36 time series.

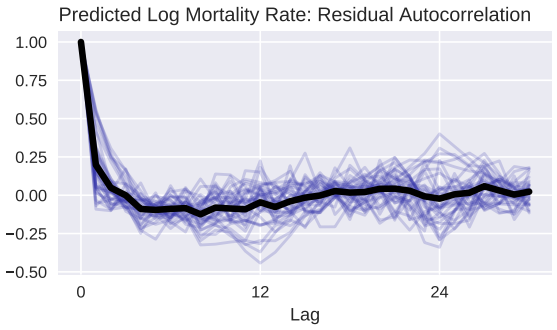

Figure 6: *Autocorrelation function of the Pearson residual series (blue) and their mean (black).*

On another note, the Negative Binomial and Tweedie (1.5) both seem to perform well when we inspect the Pearson residuals (Figure 4), but they seem to the fit the data differently well on opposing ends of the domain. Since NB is closely related to a Tweedie (2), it could interesting to further tune the Tweedie power to find an even better model (most likely with a power between 1.5 and 2) using profile likelihood [Dunn and Smyth, 2005] or the method described in Dunn and Smyth [2018]. Emerging from our analysis is that death counts are over-dispersed with respect to a Poisson model with an optimal mean-variance relationship that is polynomial with a power somewhere between 1.5 and 2.

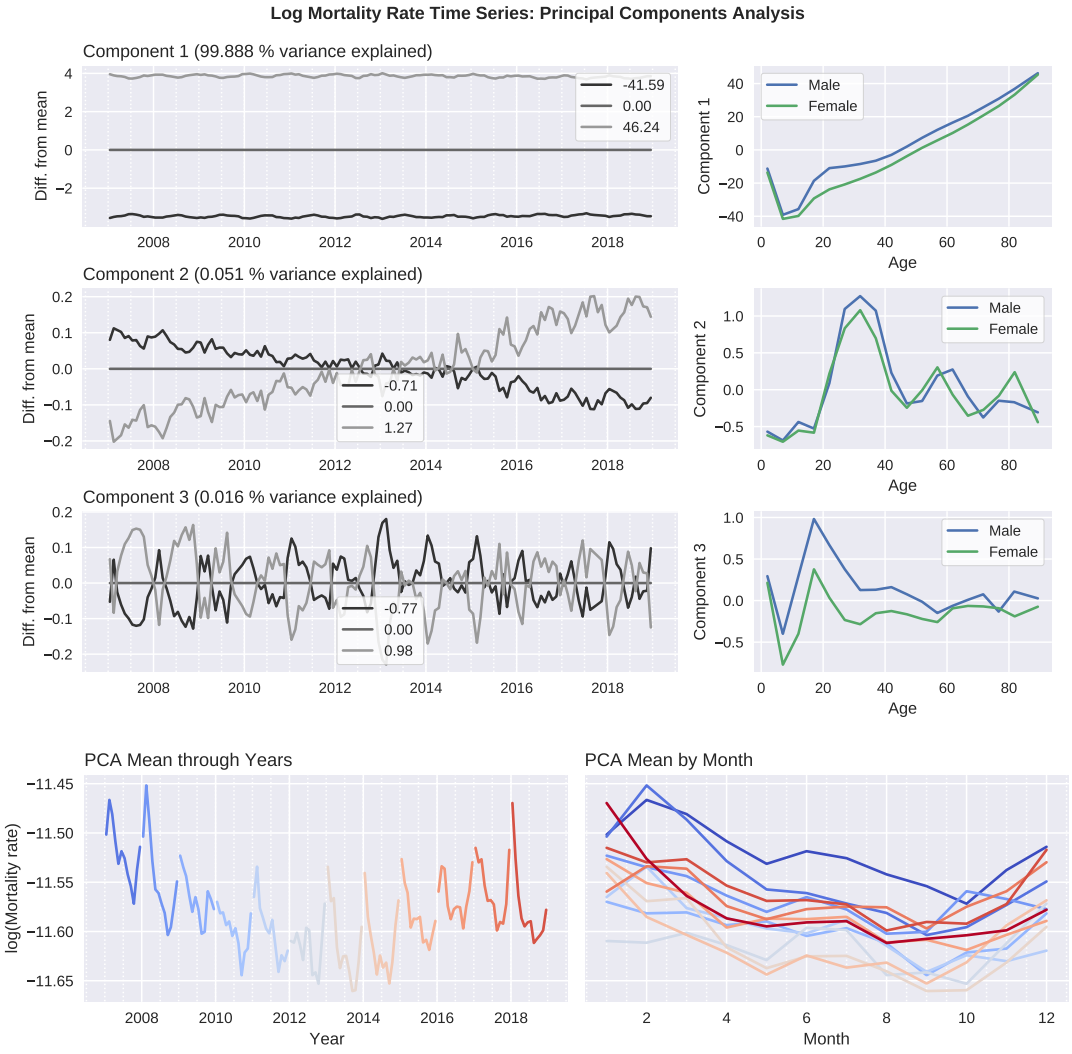

Figure 3: *Results from the PCA on the observed log mortality rate trajectories. The first three rows contain, for the first three principal components respectively, (left) the plot of the minimum and maximum effect of the component and (right) the value of the component by sex and age group. The fourth row contains the mean of the data (left) through years and (right) by months.*

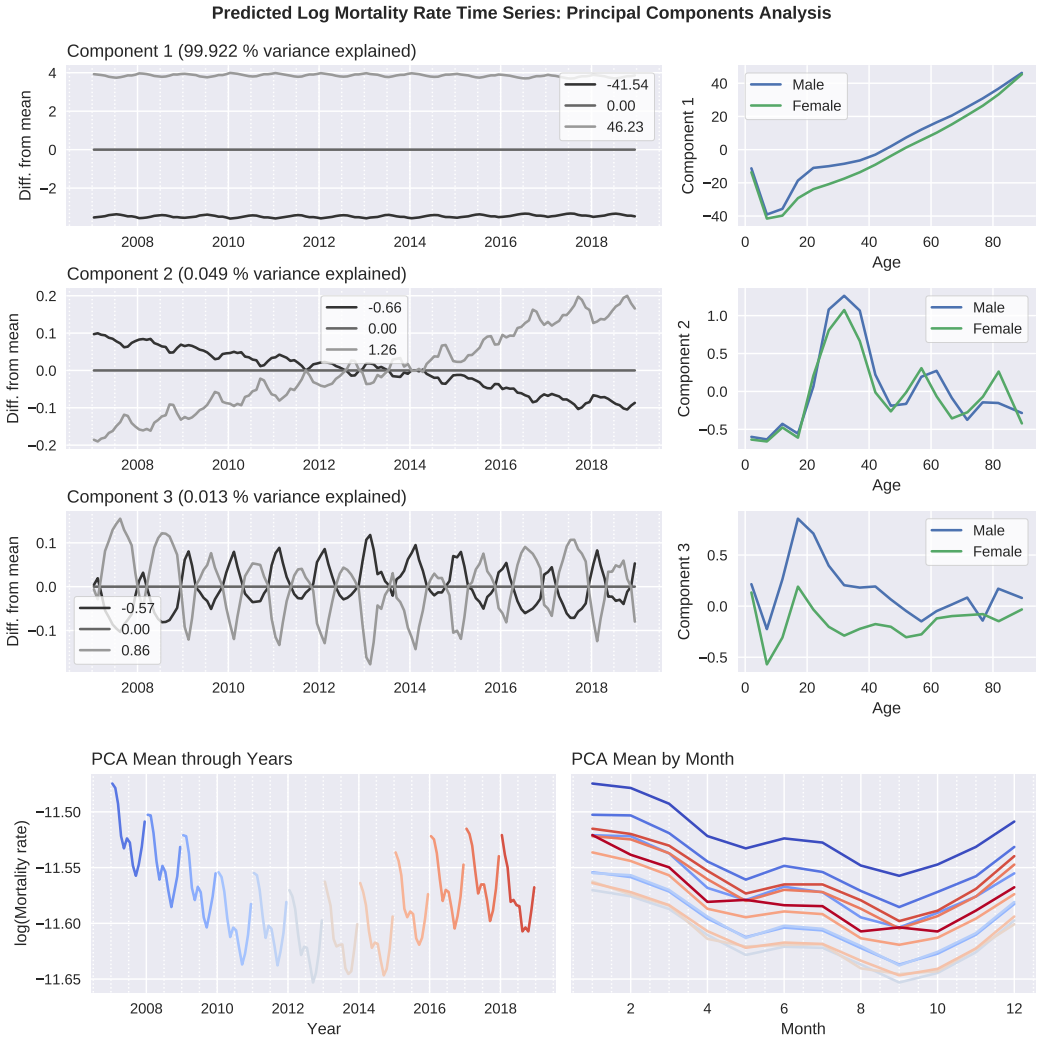

Figure 5: *Results from the PCA on the fitted log mortality rate trajectories using the selected model. The first three rows contain, for the first three principal components respectively, (left) the plot of the minimum and maximum effect of the component and (right) the value of the component by sex and age group. The fourth row contains the mean of the data (left) through years and (right) by months.*

References

W. M. Callaghan, M. F. MacDorman, C. K. Shapiro-Mendoza, and W. D. Barfield. Explaining the recent decrease in us infant mortality rate, 2007-2013. *American Journal of Obstetrics and Gynecology*, 216(1):73.e1 – 73.e8, 2017. ISSN 0002-9378. doi: https://doi.org/10.1016/j.ajog.2016.09.097. URL http://www.sciencedirect.com/science/article/pii/S0002937816308183.

Center for Disease Control and Prevention. Estimated influenza illnesses, medical visits, hospitalizations, and deaths in the united states – 2018-2019 influenza season. Online, 2020. URL https://www.cdc.gov/flu/about/burden/2018-2019.html.

J. Cheng, Z. Xu, H. Bambrick, H. Su, S. Tong, and W. Hu. Heatwave and elderly mortality: An evaluation of death burden and health costs considering short-term mortality displacement. *Environment International*, 115:334 – 342, 2018. ISSN 0160-4120. doi: https://doi.org/10.1016/j.envint.2018.03.041. URL http://www.sciencedirect.com/science/article/pii/S0160412017321980.

P. Cheng, R. Li, D. C. Schwebel, M. Zhu, and G. Hu. Traumatic brain injury mortality among u.s. children and adolescents ages 0-19 years, 1999-2017. *Journal of Safety Research*, 72:93 – 100, 2020. ISSN 0022-4375. doi: https://doi.org/10.1016/j.jsr.2019.12.013. URL http://www.sciencedirect.com/science/article/pii/S0022437519306711.

P. K. Dunn and G. K. Smyth. Series evaluation of tweedie exponential dispersion model densities. *Statistics and Computing*, 15(4):267–280, 2005.

P. K. Dunn and G. K. Smyth. Chapter 12: Tweedie glms. In *Generalized Linear Models With Examples in R*. Springer, New York, NY, 2018.

D. Henkel. Unemployment and substance use: A review of the literature (1990-2010). *Curr Drug Abuse Rev*, 4(1):4–27, Mar. 2011. doi: 10.2174/1874473711104010004.

J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3): 90–95, 2007. doi: 10.1109/MCSE.2007.55.

Institute for Work & Health. Unemployment and mental health. 2009. URL https://www.iwh.on.ca/summaries/issue-briefing/unemployment-and-mental-health.

C. Margerison-Zilk, S. Goldman-Mellor, A. Falconi, and J. Downing. Health impacts of the great recession: A critical review. *Curr Epidemiol Rep*, 3(1):81–91, Mar. 2017. doi: 10.1007/s40471-016-0068-6.

R. K. Masters, A. M. Tilstra, and D. H. Simon. Mortality from suicide, chronic liver disease, and drug poisonings among middle-aged u.s. white men and women, 1980-2013. *Biodemography and Social Biology*, 63(1):31–37, 2017. doi: 10.1080/19485565.2016.1248892. URL https://doi.org/10.1080/19485565.2016.1248892. PMID: 28287304.

W. McKinney. Data structures for statistical computing in python. In S. van der Walt and J. Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 51 – 56, 2010.

S. M. Monnat. Drugs, alcohol, and suicide represent growing share of us mortality. *Carsey School of Public Policy, University of New Hampshire*, 2017. URL https://carsey.unh.edu/publication/drugs-alcohol-suicide.

National Vital Statistics System. Mortality multiple cause-of-death. Online Dataset, Centers for Disease Control and Prevention, 2020. URL https://www.cdc.gov/nchs/nvss/mortality_public_use_data.htm.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

S. H. Preston, Y. C. Vierboom, and A. Stokes. The role of obesity in exceptionally slow us mortality improvement. *Proceedings of the National Academy of Sciences*, 115(5):957–961, 2018. ISSN 0027-8424. doi: 10.1073/pnas.1716802115. URL https://www.pnas.org/content/115/5/957.

S. Seabold and J. Perktold. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.

N. J. Smith. patsy: Describing statistical models in python. 2015.

S. Stewart, Y. Ahamed, J. F. Wiley, C. J. McDermott, J. Ball, A. K. Keates, M.-L. LÃžchen, and M. J. Carrington. Seasonal variations in cardiovascular-related mortality but not hospitalization are modulated by temperature and not climate type: a systematic review and meta-analysis of 4.5 million events in 26 countries. *Circulation*, 134(suppl1):A16759–A16759, 2016. doi: 10.1161/circ.134.suppl_1.16759. URL https://www.ahajournals.org/doi/abs/10.1161/circ.134.suppl_1.16759.

S. van der Walt, S. C. Colbert, and G. Varoquaux. The numpy array: A structure for efficient numerical computation. *Computing in Science Engineering*, 13(2):22–30, 2011.

S. H. Woolf and H. Schoomaker. Life Expectancy and Mortality Rates in the United States, 1959-2017. *JAMA*, 322(20):1996–2016, 11 2019. ISSN 0098-7484. doi: 10.1001/jama.2019.16932. URL https://doi.org/10.1001/jama.2019.16932.

A Supplemental Results

Log Mortality Rate: GLM Results						
Formula	Log likelihood	Df Model	AIC	BIC	MSE	MSEL
Negative Binomial						
`sA*(sD+S*sM)`	-45161	657	91638	-38704	77497	0.00171
`cA*(sD+S*sM)`	-45161	1115	92553	-34787	76677	0.00160
`sA*(sD+S*cM)`	-45161	737	91798	-38020	77085	0.00165
`cA*(sD+S*cM)`	-45161	1223	92769	-33863	76283	0.00153
`sA*(sY+S*sM)`	-45177	229	90813	-42333	127625	0.00805
`cA*(sY+S*sM)`	-45162	413	91152	-40789	113083	0.00206
`sA*(sY+S*cM)`	-45177	309	90973	-41649	127015	0.00799
`cA*(sY+S*cM)`	-45162	557	91439	-39558	112410	0.00198
Poisson						
`sA*(sD+S*sM)`	-33916	657	69148	-20377	76152	0.00173
`cA*(sD+S*sM)`	-33594	1115	69419	-17105	75507	0.00160
`sA*(sD+S*cM)`	-33808	737	69092	-19910	75740	0.00167
`cA*(sD+S*cM)`	-33482	1223	69413	-16404	75110	0.00154
`sA*(sY+S*sM)`	-45254	229	90968	-1362	119946	0.01116
`cA*(sY+S*sM)`	-38502	413	77832	-13293	110458	0.00208
`sA*(sY+S*cM)`	-45115	309	90851	-956	119314	0.01111
`cA*(sY+S*cM)`	-38329	557	77774	-12407	109777	0.00200
Tweedie (1.5)						
`sA*(sD+S*sM)`	-29492	657	60301	-38463	76720	0.00172
`cA*(sD+S*sM)`	-29374	1115	60980	-34561	76012	0.00160
`sA*(sD+S*cM)`	-29438	737	60352	-37785	76304	0.00166
`cA*(sD+S*cM)`	-29322	1223	61093	-33643	75618	0.00154
`sA*(sY+S*sM)`	-32817	229	66093	-41470	122236	0.00887
`cA*(sY+S*sM)`	-30289	413	61406	-40459	111520	0.00206
`sA*(sY+S*cM)`	-32799	309	66218	-40792	121610	0.00880
`cA*(sY+S*cM)`	-30234	557	61583	-39235	110842	0.00199
Normal (log link)						
`sA*(sD+S*sM)`	-36497	657	74311	392200895	75664	0.00205
`cA*(sD+S*sM)`	-36524	1115	75281	389137003	75072	0.00161
`sA*(sD+S*cM)`	-36489	737	74455	390086877	75256	0.00198
`cA*(sD+S*cM)`	-36526	1223	75501	387072493	74673	0.00154
`sA*(sY+S*sM)`	-37636	229	75732	613341153	118322	0.05572
`cA*(sY+S*sM)`	-37446	413	75719	568589935	109690	0.00211
`sA*(sY+S*cM)`	-37624	309	75868	610044032	117686	0.05568
`cA*(sY+S*cM)`	-37437	557	75990	565049572	109006	0.00203

Table 3: *Results from the 4 GLM families and 8 formulas fitted on the log mortality rates. See text for a description of the formulas. The MSEL is the mean squared error between the log mortality rates and the log predicted rates.*