Identification, characterization, and simulation of mortality shocks via **Hidden Markov Lee-Carter models**

Jonas Schöley
jschoeley.com

@jschoeley.com
0000-0002-3340-8518
schoeley@demogr.mpg.de

MAX PLANCK INSTITUTE
FOR DEMOGRAPHIC RESEARCH

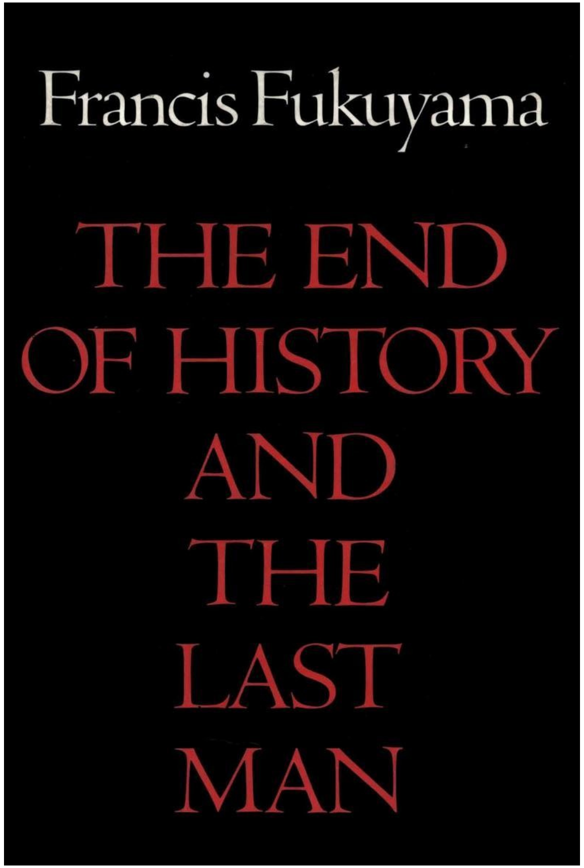

(c) Richter (November 1989). Berlin / GDR.

Forecasting at the End of History

(c) Richter (November 1989). Berlin / GDR.

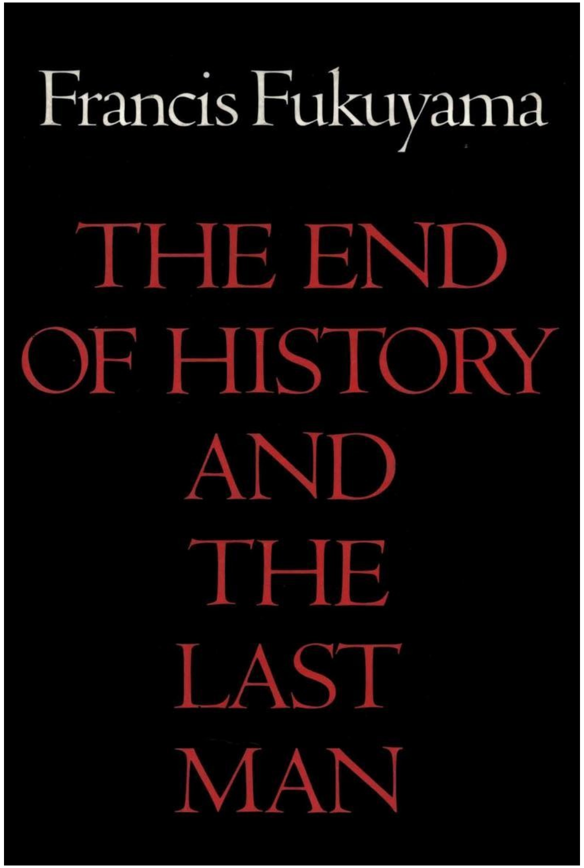
Forecasting at the end of history

Fukuyama (1992). Open Library



Forecasting at the end of history

Fukuyama (1992). Open Library



Lee & Carter (1992). 10.2307/2290201

Modeling and Forecasting U.S. Mortality

RONALD D. LEE and LAWRENCE R. CARTER*

Time series methods are used to make long-run forecasts, with confidence intervals, of age-specific mortality in the United States from 1990 to 2065. First, the logs of the age-specific death rates are modeled as a linear function of an unobserved period-specific intensity index, with parameters depending on age. This model is fit to the matrix of U.S. death rates, 1933 to 1987, using the singular value decomposition (SVD) method; it accounts for almost all the variance over time in age-specific death rates as a group. Whereas e_0 has risen at a decreasing rate over the century and has decreasing variability, $\mathbf{k}(t)$ declines at a roughly constant rate and has roughly constant variability, facilitating forecasting. $\mathbf{k}(t)$, which indexes the intensity of mortality, is next modeled as a time series (specifically, a random walk with drift) and forecast. The method performs very well on within-sample forecasts, and the forecasts are insensitive to reductions in the length of the base period from 90 to 30 years; some instability appears for base periods of 10 or 20 years, however. Forecasts of age-specific rates are derived from the forecasts of $\mathbf{k}$, and other life table variables are derived and presented. These imply an increase of 10.5 years in life expectancy to 86.05 in 2065 (sexes combined), with a confidence band of plus 3.9 or minus 5.6 years, including uncertainty concerning the estimated trend. Whereas 46% now survive to age 80, by 2065 46% will survive to age 90. Of the gains forecast for person-years lived over the life cycle from now until 2065, 74% will occur at age 65 and over. These life expectancy forecasts are substantially lower than direct time series forecasts of e_0, and have far narrower confidence bands; however, they are substantially higher than the forecasts of the Social Security Administration's Office of the Actuary.

KEY WORDS: Demography; Forecast; Life expectancy; Mortality; Population; Projection.

From 1900 to 1988, life expectancy in the United States rose from 47 to 75 years. If it were to continue to rise at this same linear rate, life expectancy would reach 100 years in 2065, about seventy five years from now. The increase would be welcomed by most of us, but it would come as a nasty surprise to the Social Security Administration, which plans on the more modest life expectancy of 80.5 years predicted by its Office of the Actuary. We scarcely need dwell on the importance of the future course of mortality in our aging society. In contrast to the past, now mortality decline is a powerful cause of population aging.

There are many ways to forecast mortality (Land 1986; Olshansky 1988). The new method we propose here is extrapolative and makes no effort to incorporate knowledge about medical, behavioral, or social influences on mortality change. Its virtues are that it combines a rich yet parsimonious demographic model with statistical time series methods, it is based firmly on persistent long-term historical patterns and trends dating back to 1900, and it provides probabilistic confidence regions for its forecasts. While many methods assume an upper limit to the human life span or rationalize in some other way the deceleration of gains in life expectancy, our method allows age-specific death rates to decline exponentially without limit; the deceleration of life expectancy follows without any special additional assumptions. We believe that our method has important advantages over other extrapolative procedures, albeit with the usual shortcomings of its genre.

In this article we first consider the available data and their limitations. We then develop our demographic model of mortality, which represents mortality level by a single index. Next we fit the demographic model to U.S. data and evaluate its historical performance. Using standard time series methods, we then forecast the index of mortality and generate associated life table values at five-year intervals. Because we intend our forecasts to be more than illustrative, we present them in some detail and provide information to enable the reader to calculate life table functions and their confidence intervals for each year of the forecast.

1. THE HISTORICAL DATA

Annual age-specific death rates for the entire U.S. population are available for the years 1933 to 1987. For the years 1900 to 1932, these data are available annually only for the death registration states, which form a varying subset of the total U.S. population, and have a cruder age specificity (see Grove and Hetzel 1968, table 51, p. 309). While data generally are available by race and sex, here we restrict our analysis to the age-specific mortality of the total population. (We plan to extend the analysis to population subgroups in the future, but are concerned about extrapolating differentials.) Death rates are available for infants and standard five-year age groups up to age 85, and for age 85 and over. There is reason to be skeptical about measures of mortality at the older ages. With 46% of the population already surviving to age 80, and with future gains in life to be concentrated at old ages, it is particularly important to deal carefully with the older age groups. We will use a new method proposed by Coale and Kisker (1990) and Coale and Guo (1989).

Figure 1 plots life expectancy at birth from the years 1900 to 1989. (It also plots forecasts, which should be ignored for now.) Not surprisingly, the 28-year increase in life expectancy between 1900 and 1987 was accompanied by more dramatic declines in death rates at some ages than at others. The mortality rate for infants fell to .067 of its initial value and that for age group 1–4 fell to .026 of its initial value, but that for

* Ronald D. Lee is Professor, Departments of Demography and Economics, University of California, Berkeley, CA 94720. Lawrence R. Carter is Professor, Department of Sociology, University of Oregon, Eugene, OR 97403. This article was prepared as part of a project on "Modeling and Forecasting Demographic Time Series," supported by NICHHD Grant RO1-HD24982. An earlier draft was presented at the 1990 annual meeting of the Population Association of America in Toronto. The authors thank Kenneth W. Wachter, John Wilmoth, George Alter, Nathan Keyfitz, Jay Olshansky, William Bell, Gregory Spencer, Leo Goodman and our referees and editor for helpful comments.

Journal of the American Statistical Association
September 1992, Vol. 87, No. 419, Applications & Case Studies

Forecasting at the end of history the Lee-Carter model

Modeling and Forecasting U.S. Mortality

RONALD D. LEE and LAWRENCE R. CARTER*

Time series methods are used to make long-run forecasts, with confidence intervals, of age-specific mortality in the United States from 1990 to 2065. First, the logs of the age-specific death rates are modeled as a linear function of an unobserved period-specific intensity index, with parameters depending on age. This model is fit to the matrix of U.S. death rates, 1933 to 1987, using the singular value decomposition (SVD) method; it accounts for almost all the variance over time in age-specific death rates as a group. Whereas e_0 has risen at a decreasing rate over the century and has decreasing variability, $\mathbf{k}(t)$ declines at a roughly constant rate and has roughly constant variability, facilitating forecasting. $\mathbf{k}(t)$, which indexes the intensity of mortality, is next modeled as a time series (specifically, a random walk with drift) and forecast. The method performs very well on within-sample forecasts, and the forecasts are insensitive to reductions in the length of the base period from 90 to 30 years; some instability appears for base periods of 10 or 20 years, however. Forecasts of age-specific rates are derived from the forecasts of $\mathbf{k}$, and other life table variables are derived and presented. These imply an increase of 10.5 years in life expectancy to 86.05 in 2065 (sexes combined), with a confidence band of plus 3.9 or minus 5.6 years, including uncertainty concerning the estimated trend. Whereas 46% now survive to age 80, by 2065 46% will survive to age 90. Of the gains forecast for person-years lived over the life cycle from now until 2065, 74% will occur at age 65 and over. These life expectancy forecasts are substantially lower than direct time series forecasts of e_0, and have far narrower confidence bands; however, they are substantially higher than the forecasts of the Social Security Administration's Office of the Actuary.

KEY WORDS: Demography; Forecast; Life expectancy; Mortality; Population; Projection.

From 1900 to 1988, life expectancy in the United States rose from 47 to 75 years. If it were to continue to rise at this same linear rate, life expectancy would reach 100 years in 2065, about seventy five years from now. The increase would be welcomed by most of us, but it would come as a nasty surprise to the Social Security Administration, which plans on the more modest life expectancy of 80.5 years predicted by its Office of the Actuary. We scarcely need dwell on the importance of the future course of mortality in our aging society. In contrast to the past, now mortality decline is a powerful cause of population aging.

There are many ways to forecast mortality (Land 1986; Olshansky 1988). The new method we propose here is extrapolative and makes no effort to incorporate knowledge about medical, behavioral, or social influences on mortality change. Its virtues are that it combines a rich yet parsimonious demographic model with statistical time series methods, it is based firmly on persistent long-term historical patterns and trends dating back to 1900, and it provides probabilistic confidence regions for its forecasts. While many methods assume an upper limit to the human life span or rationalize in some other way the deceleration of gains in life expectancy, our method allows age-specific death rates to decline exponentially without limit; the deceleration of life expectancy follows without any special additional assumptions. We believe that our method has important advantages over other extrapolative procedures, albeit with the usual shortcomings of its genre.

In this article we first consider the available data and their limitations. We then develop our demographic model of mortality, which represents mortality level by a single index. Next we fit the demographic model to U.S. data and evaluate its historical performance. Using standard time series methods, we then forecast the index of mortality and generate associated life table values at five-year intervals. Because we intend our forecasts to be more than illustrative, we present them in some detail and provide information to enable the reader to calculate life table functions and their confidence intervals for each year of the forecast.

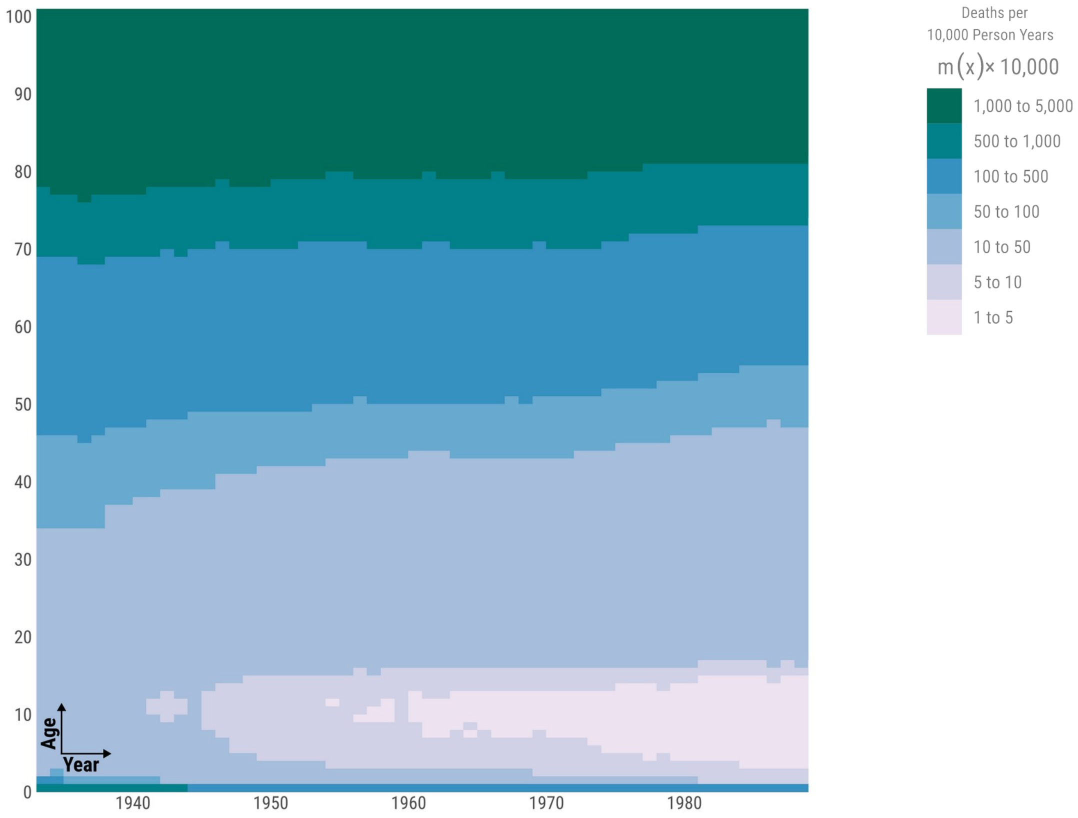
1. THE HISTORICAL DATA

Annual age-specific death rates for the entire U.S. population are available for the years 1933 to 1987. For the years 1900 to 1932, these data are available annually only for the death registration states, which form a varying subset of the total U.S. population, and have a cruder age specificity (see Grove and Hetzel 1968, table 51, p. 309). While data generally are available by race and sex, here we restrict our analysis to the age-specific mortality of the total population. (We plan to extend the analysis to population subgroups in the future, but are concerned about extrapolating differentials.) Death rates are available for infants and standard five-year age groups up to age 85, and for age 85 and over. There is reason to be skeptical about measures of mortality at the older ages. With 46% of the population already surviving to age 80, and with future gains in life to be concentrated at old ages, it is particularly important to deal carefully with the older age groups. We will use a new method proposed by Coale and Kisker (1990) and Coale and Guo (1989).

Figure 1 plots life expectancy at birth from the years 1900 to 1989. (It also plots forecasts, which should be ignored for now.) Not surprisingly, the 28-year increase in life expectancy between 1900 and 1987 was accompanied by more dramatic declines in death rates at some ages than at others. The mortality rate for infants fell to .067 of its initial value and that for age group 1–4 fell to .026 of its initial value, but that for

* Ronald D. Lee is Professor, Departments of Demography and Economics, University of California, Berkeley, CA 94720. Lawrence R. Carter is Professor, Department of Sociology, University of Oregon, Eugene, OR 97403. This article was prepared as part of a project on "Modeling and Forecasting Demographic Time Series," supported by NICHHD Grant RO1-HD24982. An earlier draft was presented at the 1990 annual meeting of the Population Association of America in Toronto. The authors thank Kenneth W. Wachter, John Wilmoth, George Alter, Nathan Keyfitz, Jay Olshansky, William Bell, Gregory Spencer, Leo Goodman and our referees and editor for helpful comments.

Journal of the American Statistical Association
September 1992, Vol. 87, No. 419, Applications & Case Studies



Human Mortality Database (2025). US death rates 1933-1989. mortality.org

Forecasting at the end of history the Lee-Carter model

Modeling and Forecasting U.S. Mortality

RONALD D. LEE and LAWRENCE R. CARTER*

Time series methods are used to make long-run forecasts, with confidence intervals, of age-specific mortality in the United States from 1990 to 2065. First, the logs of the age-specific death rates are modeled as a linear function of an unobserved period-specific intensity index, with parameters depending on age. This model is fit to the matrix of U.S. death rates, 1933 to 1987, using the singular value decomposition (SVD) method; it accounts for almost all the variance over time in age-specific death rates as a group. Whereas e_0 has risen at a decreasing rate over the century and has decreasing variability, $\mathbf{k}(t)$ declines at a roughly constant rate and has roughly constant variability, facilitating forecasting. $\mathbf{k}(t)$, which indexes the intensity of mortality, is next modeled as a time series (specifically, a random walk with drift) and forecast. The method performs very well on within-sample forecasts, and the forecasts are insensitive to reductions in the length of the base period from 90 to 30 years; some instability appears for base periods of 10 or 20 years, however. Forecasts of age-specific rates are derived from the forecasts of $\mathbf{k}$, and other life table variables are derived and presented. These imply an increase of 10.5 years in life expectancy to 86.05 in 2065 (sexes combined), with a confidence band of plus 3.9 or minus 5.6 years, including uncertainty concerning the estimated trend. Whereas 46% now survive to age 80, by 2065 46% will survive to age 90. Of the gains forecast for person-years lived over the life cycle from now until 2065, 74% will occur at age 65 and over. These life expectancy forecasts are substantially lower than direct time series forecasts of e_0, and have far narrower confidence bands; however, they are substantially higher than the forecasts of the Social Security Administration's Office of the Actuary.

KEY WORDS: Demography; Forecast; Life expectancy; Mortality; Population; Projection.

From 1900 to 1988, life expectancy in the United States rose from 47 to 75 years. If it were to continue to rise at this same linear rate, life expectancy would reach 100 years in 2065, about seventy five years from now. The increase would be welcomed by most of us, but it would come as a nasty surprise to the Social Security Administration, which plans on the more modest life expectancy of 80.5 years predicted by its Office of the Actuary. We scarcely need dwell on the importance of the future course of mortality in our aging society. In contrast to the past, now mortality decline is a powerful cause of population aging.

There are many ways to forecast mortality (Land 1986; Olshansky 1988). The new method we propose here is extrapolative and makes no effort to incorporate knowledge about medical, behavioral, or social influences on mortality change. Its virtues are that it combines a rich yet parsimonious demographic model with statistical time series methods, it is based firmly on persistent long-term historical patterns and trends dating back to 1900, and it provides probabilistic confidence regions for its forecasts. While many methods assume an upper limit to the human life span or rationalize in some other way the deceleration of gains in life expectancy, our method allows age-specific death rates to decline exponentially without limit; the deceleration of life expectancy follows without any special additional assumptions. We believe that our method has important advantages over other extrapolative procedures, albeit with the usual shortcomings of its genre.

In this article we first consider the available data and their limitations. We then develop our demographic model of mortality, which represents mortality level by a single index. Next we fit the demographic model to U.S. data and evaluate its historical performance. Using standard time series methods, we then forecast the index of mortality and generate associated life table values at five-year intervals. Because we intend our forecasts to be more than illustrative, we present them in some detail and provide information to enable the reader to calculate life table functions and their confidence intervals for each year of the forecast.

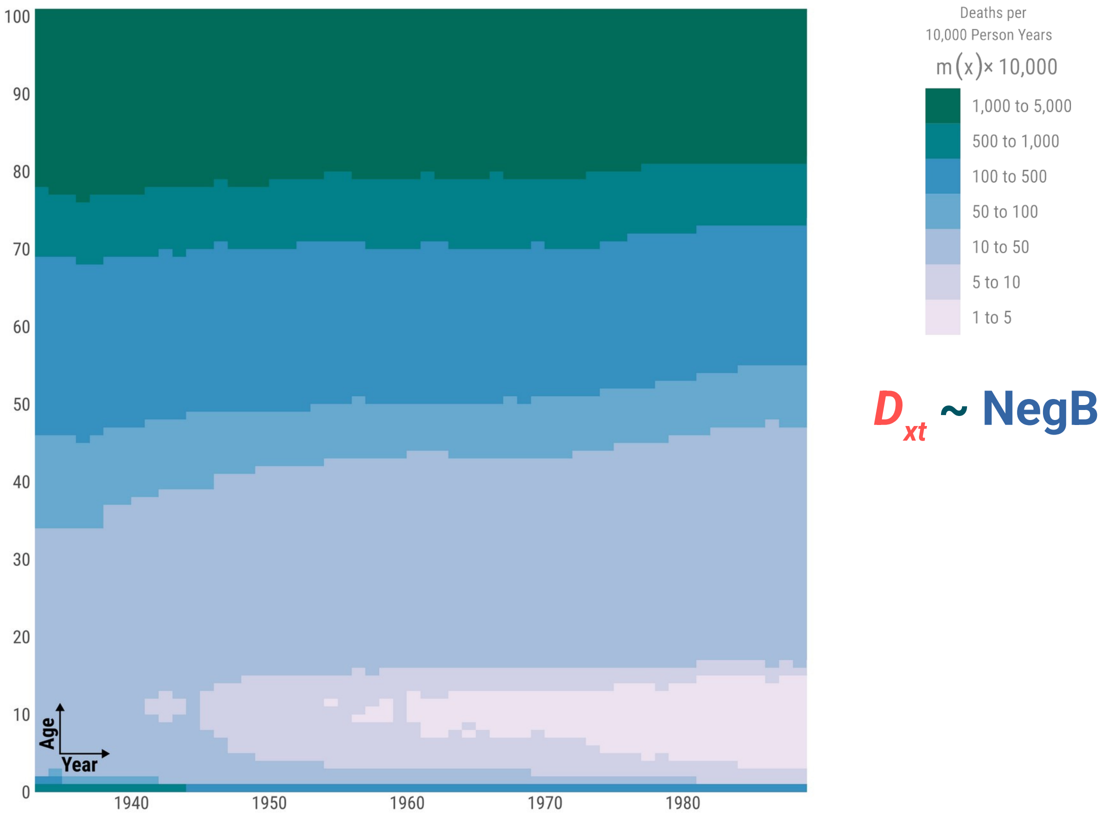
1. THE HISTORICAL DATA

Annual age-specific death rates for the entire U.S. population are available for the years 1933 to 1987. For the years 1900 to 1932, these data are available annually only for the death registration states, which form a varying subset of the total U.S. population, and have a cruder age specificity (see Grove and Hetzel 1968, table 51, p. 309). While data generally are available by race and sex, here we restrict our analysis to the age-specific mortality of the total population. (We plan to extend the analysis to population subgroups in the future, but are concerned about extrapolating differentials.) Death rates are available for infants and standard five-year age groups up to age 85, and for age 85 and over. There is reason to be skeptical about measures of mortality at the older ages. With 46% of the population already surviving to age 80, and with future gains in life to be concentrated at old ages, it is particularly important to deal carefully with the older age groups. We will use a new method proposed by Coale and Kisker (1990) and Coale and Guo (1989).

Figure 1 plots life expectancy at birth from the years 1900 to 1989. (It also plots forecasts, which should be ignored for now.) Not surprisingly, the 28-year increase in life expectancy between 1900 and 1987 was accompanied by more dramatic declines in death rates at some ages than at others. The mortality rate for infants fell to .067 of its initial value and that for age group 1–4 fell to .026 of its initial value, but that for

* Ronald D. Lee is Professor, Departments of Demography and Economics, University of California, Berkeley, CA 94720. Lawrence R. Carter is Professor, Department of Sociology, University of Oregon, Eugene, OR 97403. This article was prepared as part of a project on "Modeling and Forecasting Demographic Time Series," supported by NICHHD Grant RO1-HD24982. An earlier draft was presented at the 1990 annual meeting of the Population Association of America in Toronto. The authors thank Kenneth W. Wachter, John Wilmoth, George Alter, Nathan Keyfitz, Jay Olshansky, William Bell, Gregory Spencer, Leo Goodman and our referees and editor for helpful comments.

Journal of the American Statistical Association
September 1992, Vol. 87, No. 419, Applications & Case Studies



$$D_{xt} \sim \mathrm{NegBin}(\lambda_{xt} E_{xt}, \varphi)$$

Human Mortality Database (2025). US death rates 1933-1989. mortality.org

Forecasting at the end of history the Lee-Carter model

Modeling and Forecasting U.S. Mortality

RONALD D. LEE and LAWRENCE R. CARTER*

Time series methods are used to make long-run forecasts, with confidence intervals, of age-specific mortality in the United States from 1990 to 2065. First, the logs of the age-specific death rates are modeled as a linear function of an unobserved period-specific intensity index, with parameters depending on age. This model is fit to the matrix of U.S. death rates, 1933 to 1987, using the singular value decomposition (SVD) method; it accounts for almost all the variance over time in age-specific death rates as a group. Whereas e_0 has risen at a decreasing rate over the century and has decreasing variability, $\mathbf{k}(t)$ declines at a roughly constant rate and has roughly constant variability, facilitating forecasting. $\mathbf{k}(t)$, which indexes the intensity of mortality, is next modeled as a time series (specifically, a random walk with drift) and forecast. The method performs very well on within-sample forecasts, and the forecasts are insensitive to reductions in the length of the base period from 90 to 30 years; some instability appears for base periods of 10 or 20 years, however. Forecasts of age-specific rates are derived from the forecasts of $\mathbf{k}$, and other life table variables are derived and presented. These imply an increase of 10.5 years in life expectancy to 86.05 in 2065 (sexes combined), with a confidence band of plus 3.9 or minus 5.6 years, including uncertainty concerning the estimated trend. Whereas 46% now survive to age 80, by 2065 46% will survive to age 90. Of the gains forecast for person-years lived over the life cycle from now until 2065, 74% will occur at age 65 and over. These life expectancy forecasts are substantially lower than direct time series forecasts of e_0, and have far narrower confidence bands; however, they are substantially higher than the forecasts of the Social Security Administration's Office of the Actuary.

KEY WORDS: Demography; Forecast; Life expectancy; Mortality; Population; Projection.

From 1900 to 1988, life expectancy in the United States rose from 47 to 75 years. If it were to continue to rise at this same linear rate, life expectancy would reach 100 years in 2065, about seventy five years from now. The increase would be welcomed by most of us, but it would come as a nasty surprise to the Social Security Administration, which plans on the more modest life expectancy of 80.5 years predicted by its Office of the Actuary. We scarcely need dwell on the importance of the future course of mortality in our aging society. In contrast to the past, now mortality decline is a powerful cause of population aging.

There are many ways to forecast mortality (Land 1986; Olshansky 1988). The new method we propose here is extrapolative and makes no effort to incorporate knowledge about medical, behavioral, or social influences on mortality change. Its virtues are that it combines a rich yet parsimonious demographic model with statistical time series methods, it is based firmly on persistent long-term historical patterns and trends dating back to 1900, and it provides probabilistic confidence regions for its forecasts. While many methods assume an upper limit to the human life span or rationalize in some other way the deceleration of gains in life expectancy, our method allows age-specific death rates to decline exponentially without limit; the deceleration of life expectancy follows without any special additional assumptions. We believe that our method has important advantages over other extrapolative procedures, albeit with the usual shortcomings of its genre.

In this article we first consider the available data and their limitations. We then develop our demographic model of mortality, which represents mortality level by a single index. Next we fit the demographic model to U.S. data and evaluate its historical performance. Using standard time series methods, we then forecast the index of mortality and generate associated life table values at five-year intervals. Because we intend our forecasts to be more than illustrative, we present them in some detail and provide information to enable the reader to calculate life table functions and their confidence intervals for each year of the forecast.

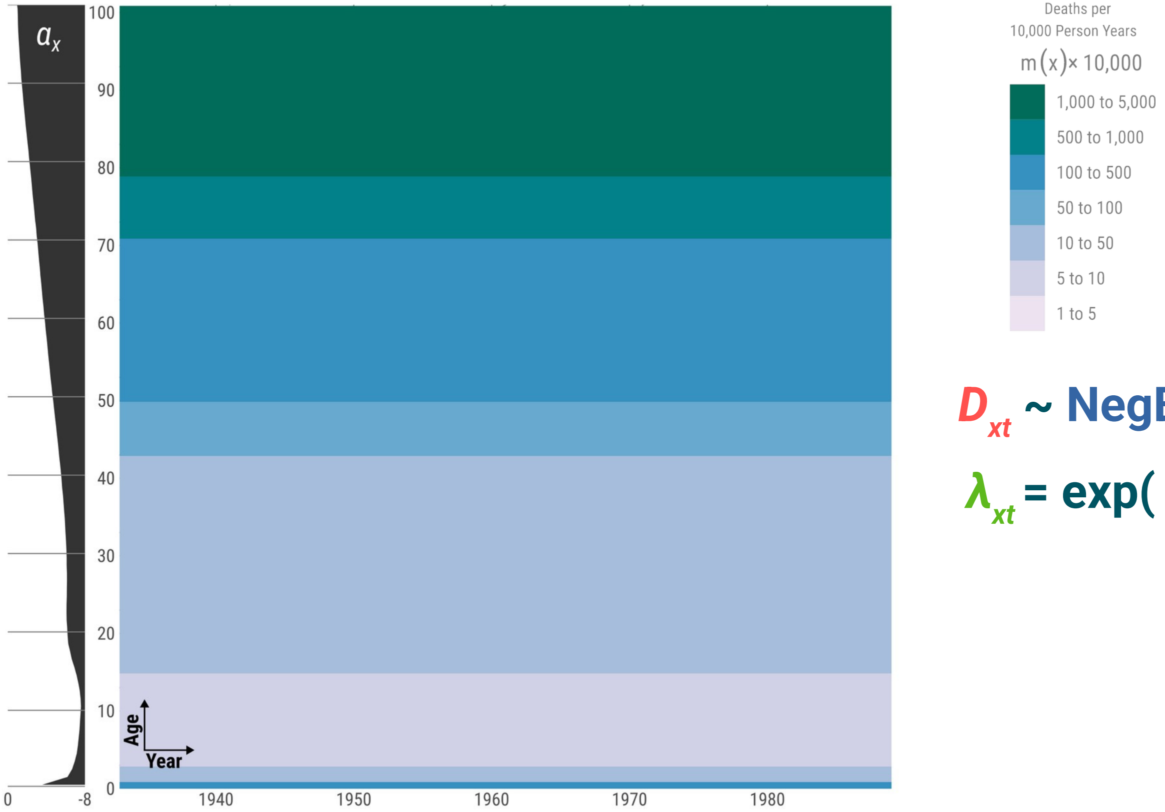
1. THE HISTORICAL DATA

Annual age-specific death rates for the entire U.S. population are available for the years 1933 to 1987. For the years 1900 to 1932, these data are available annually only for the death registration states, which form a varying subset of the total U.S. population, and have a cruder age specificity (see Grove and Hetzel 1968, table 51, p. 309). While data generally are available by race and sex, here we restrict our analysis to the age-specific mortality of the total population. (We plan to extend the analysis to population subgroups in the future, but are concerned about extrapolating differentials.) Death rates are available for infants and standard five-year age groups up to age 85, and for age 85 and over. There is reason to be skeptical about measures of mortality at the older ages. With 46% of the population already surviving to age 80, and with future gains in life to be concentrated at old ages, it is particularly important to deal carefully with the older age groups. We will use a new method proposed by Coale and Kisker (1990) and Coale and Guo (1989).

Figure 1 plots life expectancy at birth from the years 1900 to 1989. (It also plots forecasts, which should be ignored for now.) Not surprisingly, the 28-year increase in life expectancy between 1900 and 1987 was accompanied by more dramatic declines in death rates at some ages than at others. The mortality rate for infants fell to .067 of its initial value and that for age group 1–4 fell to .026 of its initial value, but that for

* Ronald D. Lee is Professor, Departments of Demography and Economics, University of California, Berkeley, CA 94720. Lawrence R. Carter is Professor, Department of Sociology, University of Oregon, Eugene, OR 97403. This article was prepared as part of a project on "Modeling and Forecasting Demographic Time Series," supported by NICHHD Grant RO1-HD24982. An earlier draft was presented at the 1990 annual meeting of the Population Association of America in Toronto. The authors thank Kenneth W. Wachter, John Wilmoth, George Alter, Nathan Keyfitz, Jay Olshansky, William Bell, Gregory Spencer, Leo Goodman and our referees and editor for helpful comments.

Journal of the American Statistical Association
September 1992, Vol. 87, No. 419, Applications & Case Studies



$$D_{xt} \sim \text{NegBin}(\lambda_{xt} E_{xt}, \varphi)$$

$$\lambda_{xt} = \exp(a_x)$$

Human Mortality Database (2025). US death rates 1933-1989. mortality.org

Forecasting at the end of history the Lee-Carter model

Modeling and Forecasting U.S. Mortality

RONALD D. LEE and LAWRENCE R. CARTER*

Time series methods are used to make long-run forecasts, with confidence intervals, of age-specific mortality in the United States from 1990 to 2065. First, the logs of the age-specific death rates are modeled as a linear function of an unobserved period-specific intensity index, with parameters depending on age. This model is fit to the matrix of U.S. death rates, 1933 to 1987, using the singular value decomposition (SVD) method; it accounts for almost all the variance over time in age-specific death rates as a group. Whereas e_0 has risen at a decreasing rate over the century and has decreasing variability, $\mathbf{k}(t)$ declines at a roughly constant rate and has roughly constant variability, facilitating forecasting. $\mathbf{k}(t)$, which indexes the intensity of mortality, is next modeled as a time series (specifically, a random walk with drift) and forecast. The method performs very well on within-sample forecasts, and the forecasts are insensitive to reductions in the length of the base period from 90 to 30 years; some instability appears for base periods of 10 or 20 years, however. Forecasts of age-specific rates are derived from the forecasts of $\mathbf{k}$, and other life table variables are derived and presented. These imply an increase of 10.5 years in life expectancy to 86.05 in 2065 (sexes combined), with a confidence band of plus 3.9 or minus 5.6 years, including uncertainty concerning the estimated trend. Whereas 46% now survive to age 80, by 2065 46% will survive to age 90. Of the gains forecast for person-years lived over the life cycle from now until 2065, 74% will occur at age 65 and over. These life expectancy forecasts are substantially lower than direct time series forecasts of e_0, and have far narrower confidence bands; however, they are substantially higher than the forecasts of the Social Security Administration's Office of the Actuary.

KEY WORDS: Demography; Forecast; Life expectancy; Mortality; Population; Projection.

From 1900 to 1988, life expectancy in the United States rose from 47 to 75 years. If it were to continue to rise at this same linear rate, life expectancy would reach 100 years in 2065, about seventy five years from now. The increase would be welcomed by most of us, but it would come as a nasty surprise to the Social Security Administration, which plans on the more modest life expectancy of 80.5 years predicted by its Office of the Actuary. We scarcely need dwell on the importance of the future course of mortality in our aging society. In contrast to the past, now mortality decline is a powerful cause of population aging.

There are many ways to forecast mortality (Land 1986; Olshansky 1988). The new method we propose here is extrapolative and makes no effort to incorporate knowledge about medical, behavioral, or social influences on mortality change. Its virtues are that it combines a rich yet parsimonious demographic model with statistical time series methods, it is based firmly on persistent long-term historical patterns and trends dating back to 1900, and it provides probabilistic confidence regions for its forecasts. While many methods assume an upper limit to the human life span or rationalize in some other way the deceleration of gains in life expectancy, our method allows age-specific death rates to decline exponentially without limit; the deceleration of life expectancy follows without any special additional assumptions. We believe that our method has important advantages over other extrapolative procedures, albeit with the usual shortcomings of its genre.

In this article we first consider the available data and their limitations. We then develop our demographic model of mortality, which represents mortality level by a single index. Next we fit the demographic model to U.S. data and evaluate its historical performance. Using standard time series methods, we then forecast the index of mortality and generate associated life table values at five-year intervals. Because we intend our forecasts to be more than illustrative, we present them in some detail and provide information to enable the reader to calculate life table functions and their confidence intervals for each year of the forecast.

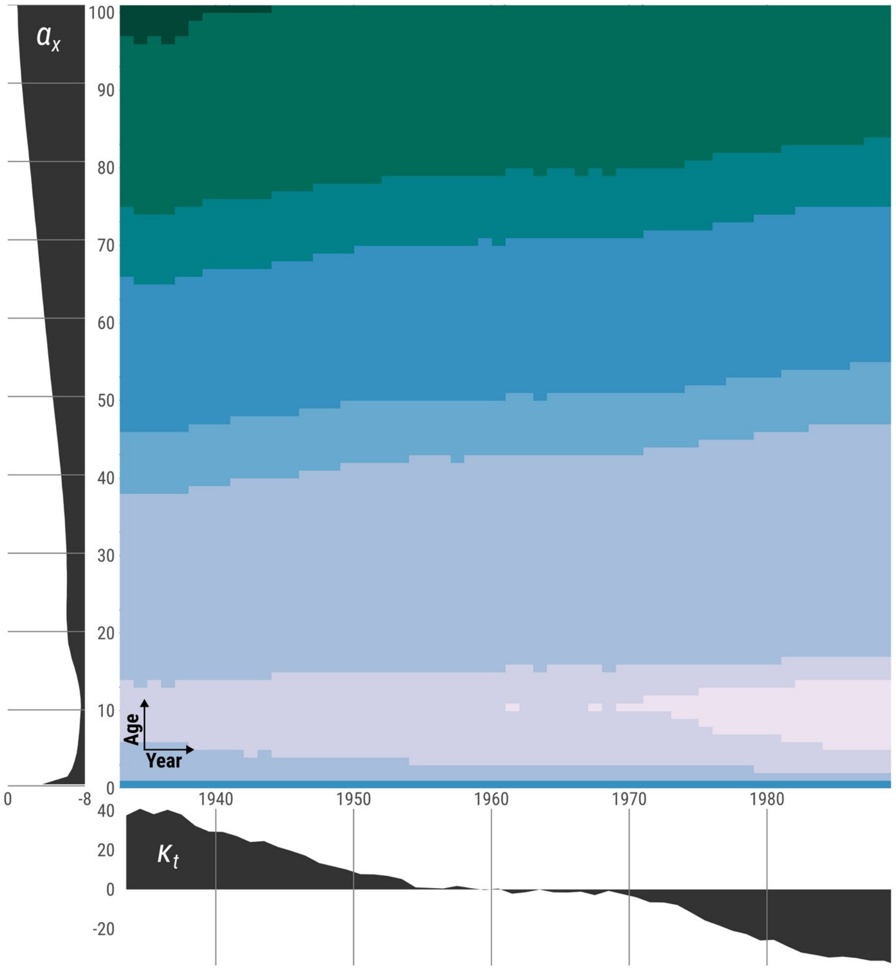
1. THE HISTORICAL DATA

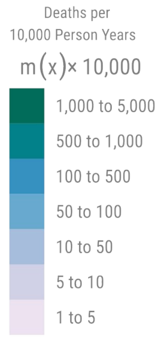
Annual age-specific death rates for the entire U.S. population are available for the years 1933 to 1987. For the years 1900 to 1932, these data are available annually only for the death registration states, which form a varying subset of the total U.S. population, and have a cruder age specificity (see Grove and Hetzel 1968, table 51, p. 309). While data generally are available by race and sex, here we restrict our analysis to the age-specific mortality of the total population. (We plan to extend the analysis to population subgroups in the future, but are concerned about extrapolating differentials.) Death rates are available for infants and standard five-year age groups up to age 85, and for age 85 and over. There is reason to be skeptical about measures of mortality at the older ages. With 46% of the population already surviving to age 80, and with future gains in life to be concentrated at old ages, it is particularly important to deal carefully with the older age groups. We will use a new method proposed by Coale and Kisker (1990) and Coale and Guo (1989).

Figure 1 plots life expectancy at birth from the years 1900 to 1989. (It also plots forecasts, which should be ignored for now.) Not surprisingly, the 28-year increase in life expectancy between 1900 and 1987 was accompanied by more dramatic declines in death rates at some ages than at others. The mortality rate for infants fell to .067 of its initial value and that for age group 1–4 fell to .026 of its initial value, but that for

* Ronald D. Lee is Professor, Departments of Demography and Economics, University of California, Berkeley, CA 94720. Lawrence R. Carter is Professor, Department of Sociology, University of Oregon, Eugene, OR 97403. This article was prepared as part of a project on "Modeling and Forecasting Demographic Time Series," supported by NICHHD Grant RO1-HD24982. An earlier draft was presented at the 1990 annual meeting of the Population Association of America in Toronto. The authors thank Kenneth W. Wachter, John Wilmoth, George Alter, Nathan Keyfitz, Jay Olshansky, William Bell, Gregory Spencer, Leo Goodman and our referees and editor for helpful comments.

Journal of the American Statistical Association
September 1992, Vol. 87, No. 419, Applications & Case Studies





$$D_{xt} \sim \text{NegBin}(\lambda_{xt} E_{xt}, \varphi)$$

$$\lambda_{xt} = \exp(a_x + \kappa_t \times 0.01)$$

Human Mortality Database (2025). US death rates 1933-1989. mortality.org

Forecasting at the end of history the Lee-Carter model

Modeling and Forecasting U.S. Mortality

RONALD D. LEE and LAWRENCE R. CARTER*

Time series methods are used to make long-run forecasts, with confidence intervals, of age-specific mortality in the United States from 1990 to 2065. First, the logs of the age-specific death rates are modeled as a linear function of an unobserved period-specific intensity index, with parameters depending on age. This model is fit to the matrix of U.S. death rates, 1933 to 1987, using the singular value decomposition (SVD) method; it accounts for almost all the variance over time in age-specific death rates as a group. Whereas e_0 has risen at a decreasing rate over the century and has decreasing variability, $\mathbf{k}(t)$ declines at a roughly constant rate and has roughly constant variability, facilitating forecasting. $\mathbf{k}(t)$, which indexes the intensity of mortality, is next modeled as a time series (specifically, a random walk with drift) and forecast. The method performs very well on within-sample forecasts, and the forecasts are insensitive to reductions in the length of the base period from 90 to 30 years; some instability appears for base periods of 10 or 20 years, however. Forecasts of age-specific rates are derived from the forecasts of $\mathbf{k}$, and other life table variables are derived and presented. These imply an increase of 10.5 years in life expectancy to 86.05 in 2065 (sexes combined), with a confidence band of plus 3.9 or minus 5.6 years, including uncertainty concerning the estimated trend. Whereas 46% now survive to age 80, by 2065 46% will survive to age 90. Of the gains forecast for person-years lived over the life cycle from now until 2065, 74% will occur at age 65 and over. These life expectancy forecasts are substantially lower than direct time series forecasts of e_0, and have far narrower confidence bands; however, they are substantially higher than the forecasts of the Social Security Administration's Office of the Actuary.

KEY WORDS: Demography; Forecast; Life expectancy; Mortality; Population; Projection.

From 1900 to 1988, life expectancy in the United States rose from 47 to 75 years. If it were to continue to rise at this same linear rate, life expectancy would reach 100 years in 2065, about seventy five years from now. The increase would be welcomed by most of us, but it would come as a nasty surprise to the Social Security Administration, which plans on the more modest life expectancy of 80.5 years predicted by its Office of the Actuary. We scarcely need dwell on the importance of the future course of mortality in our aging society. In contrast to the past, now mortality decline is a powerful cause of population aging.

There are many ways to forecast mortality (Land 1986; Olshansky 1988). The new method we propose here is extrapolative and makes no effort to incorporate knowledge about medical, behavioral, or social influences on mortality change. Its virtues are that it combines a rich yet parsimonious demographic model with statistical time series methods, it is based firmly on persistent long-term historical patterns and trends dating back to 1900, and it provides probabilistic confidence regions for its forecasts. While many methods assume an upper limit to the human life span or rationalize in some other way the deceleration of gains in life expectancy, our method allows age-specific death rates to decline exponentially without limit; the deceleration of life expectancy follows without any special additional assumptions. We believe that our method has important advantages over other extrapolative procedures, albeit with the usual shortcomings of its genre.

In this article we first consider the available data and their limitations. We then develop our demographic model of mortality, which represents mortality level by a single index. Next we fit the demographic model to U.S. data and evaluate its historical performance. Using standard time series methods, we then forecast the index of mortality and generate associated life table values at five-year intervals. Because we intend our forecasts to be more than illustrative, we present them in some detail and provide information to enable the reader to calculate life table functions and their confidence intervals for each year of the forecast.

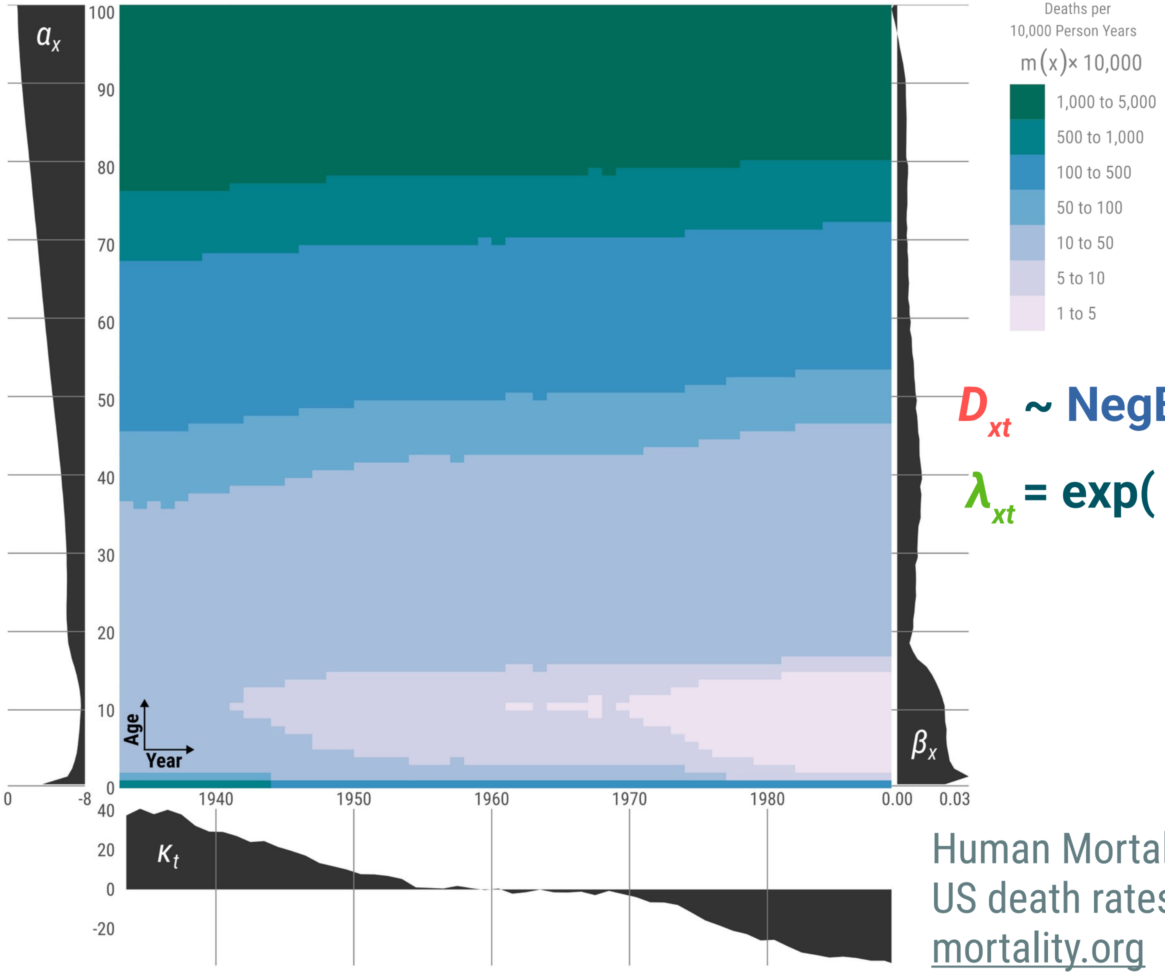
1. THE HISTORICAL DATA

Annual age-specific death rates for the entire U.S. population are available for the years 1933 to 1987. For the years 1900 to 1932, these data are available annually only for the death registration states, which form a varying subset of the total U.S. population, and have a cruder age specificity (see Grove and Hetzel 1968, table 51, p. 309). While data generally are available by race and sex, here we restrict our analysis to the age-specific mortality of the total population. (We plan to extend the analysis to population subgroups in the future, but are concerned about extrapolating differentials.) Death rates are available for infants and standard five-year age groups up to age 85, and for age 85 and over. There is reason to be skeptical about measures of mortality at the older ages. With 46% of the population already surviving to age 80, and with future gains in life to be concentrated at old ages, it is particularly important to deal carefully with the older age groups. We will use a new method proposed by Coale and Kisker (1990) and Coale and Guo (1989).

Figure 1 plots life expectancy at birth from the years 1900 to 1989. (It also plots forecasts, which should be ignored for now.) Not surprisingly, the 28-year increase in life expectancy between 1900 and 1987 was accompanied by more dramatic declines in death rates at some ages than at others. The mortality rate for infants fell to .067 of its initial value and that for age group 1–4 fell to .026 of its initial value, but that for

* Ronald D. Lee is Professor, Departments of Demography and Economics, University of California, Berkeley, CA 94720. Lawrence R. Carter is Professor, Department of Sociology, University of Oregon, Eugene, OR 97403. This article was prepared as part of a project on "Modeling and Forecasting Demographic Time Series," supported by NICHHD Grant RO1-HD24982. An earlier draft was presented at the 1990 annual meeting of the Population Association of America in Toronto. The authors thank Kenneth W. Wachter, John Wilmoth, George Alter, Nathan Keyfitz, Jay Olshansky, William Bell, Gregory Spencer, Leo Goodman and our referees and editor for helpful comments.

Journal of the American Statistical Association
September 1992, Vol. 87, No. 419, Applications & Case Studies

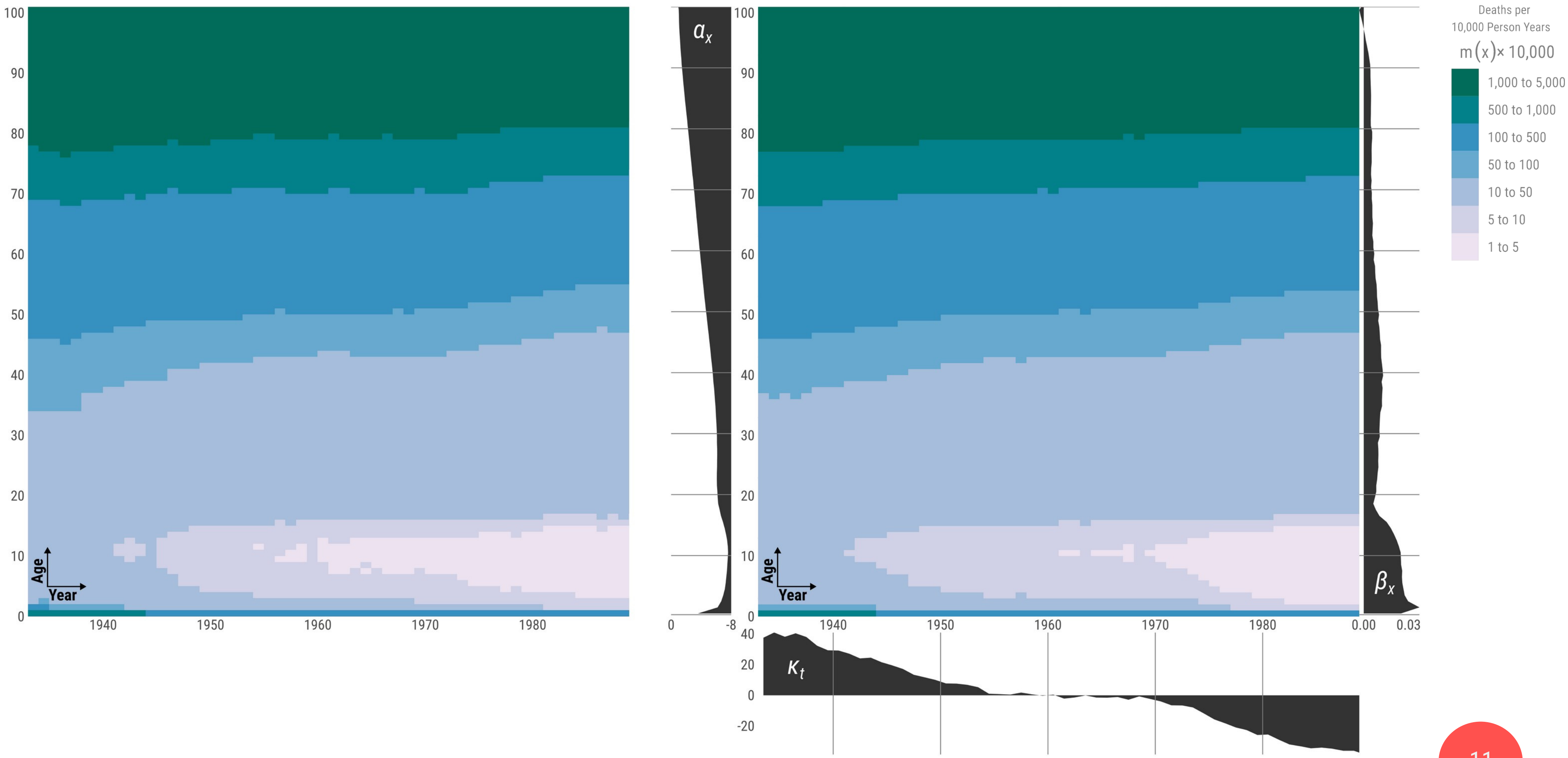


$$D_{xt} \sim \text{NegBin}(\lambda_{xt} E_{xt}, \varphi)$$

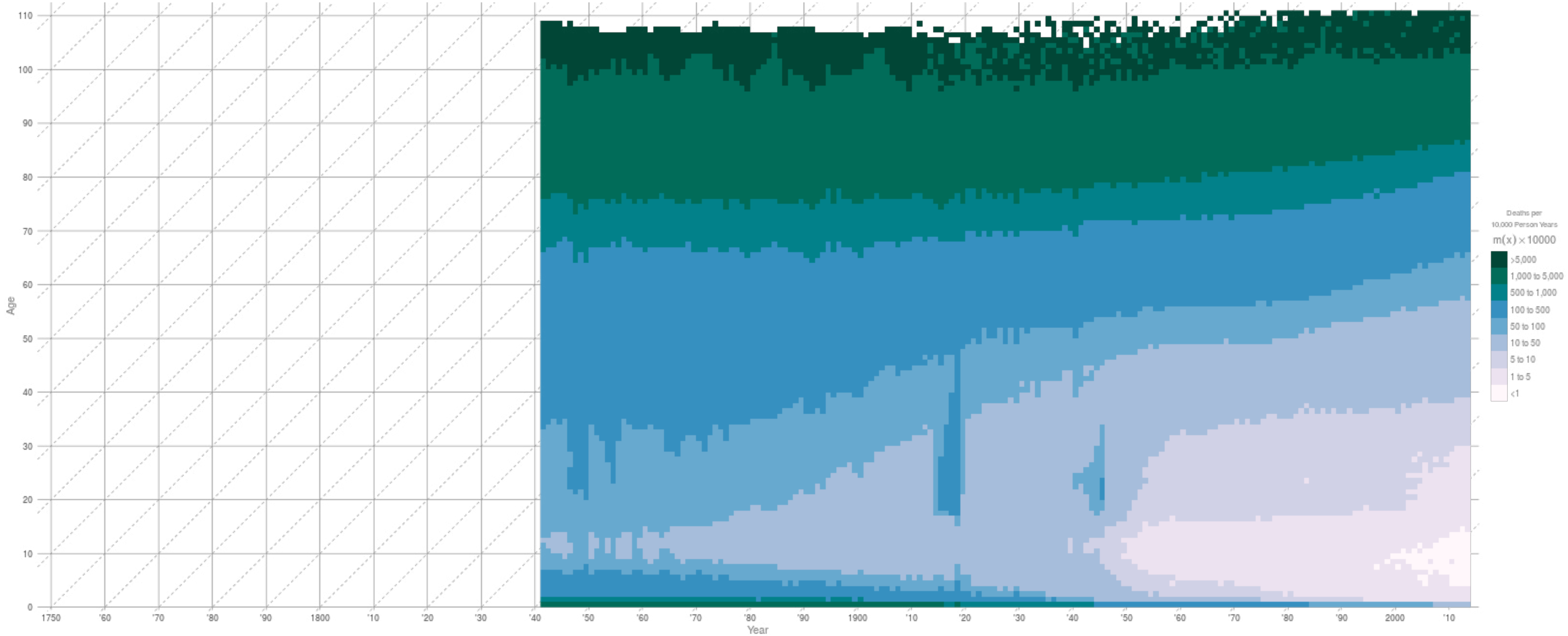
$$\lambda_{xt} = \exp(a_x + \beta_x \kappa_t)$$

Human Mortality Database (2025). US death rates 1933-1989. mortality.org

Forecasting at the end of history the Lee-Carter model



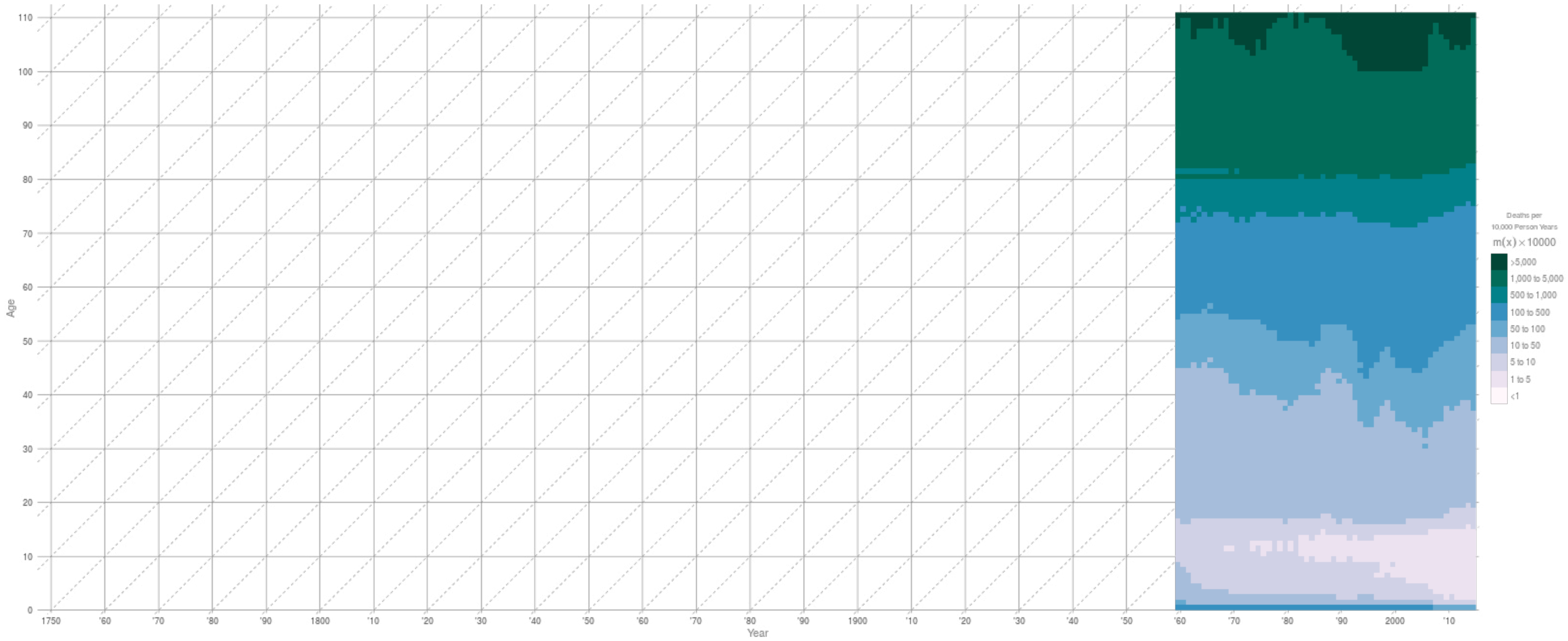
Forecasting at the end of history limits of the Lee-Carter model



Human Mortality Database (2016). England & Wales death rates. mortality.org

Schöley (2016). The Human Mortality Explorer. jschoeley.shinyapps.io/hmdexp/

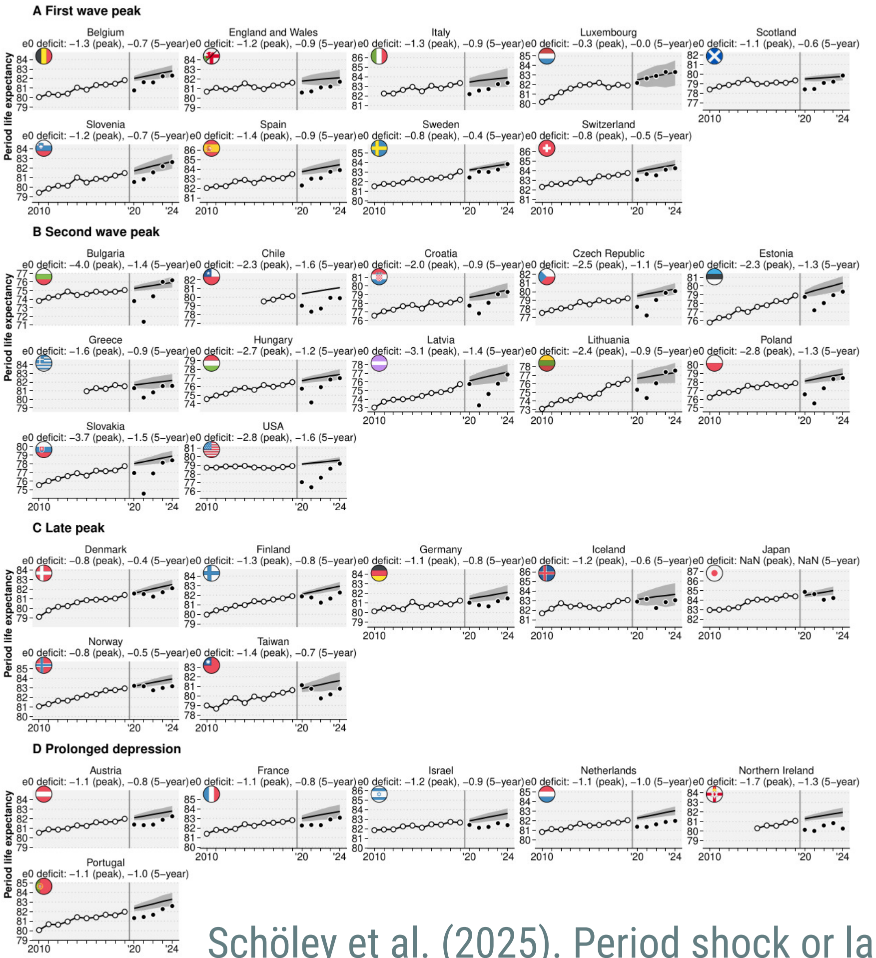
Forecasting at the end of history limits of the Lee-Carter model

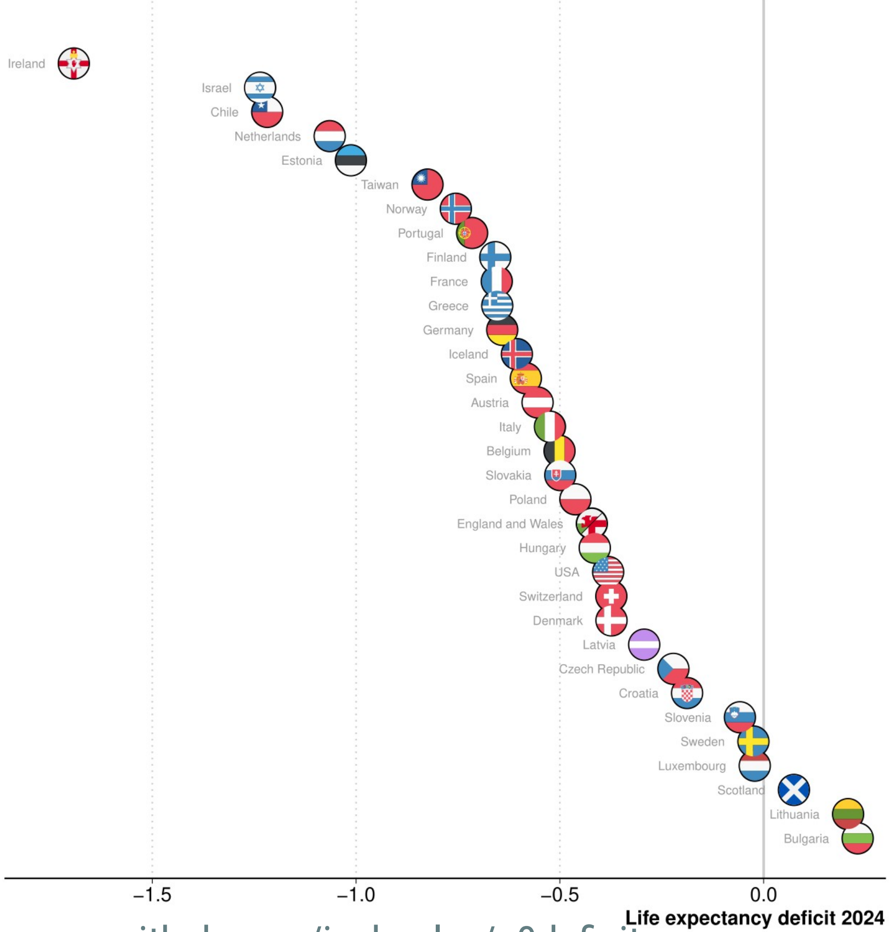


Human Mortality Database (2016). Russian death rates. mortality.org

Schöley (2016). The Human Mortality Explorer. jschoeley.shinyapps.io/hmdexp/

Forecasting at the end of history limits of the Lee-Carter model





Schöley et al. (2025). Period shock or lasting scar. github.com/jschoeley/e0deficit

Forecasting at the end of history limits of the Lee-Carter model

[3] E. Biffis. Affine processes for dynamic mortality and actuarial valuations. *Insurance: Mathematics and Economics*, 37(3):443–468, Dec. 2005. ISSN 01676687. doi:10.1016/j.insmatheco.2005.05.003.

[4] J. Goes, K. Barigou, and A. Leucht. Bayesian mortality modelling with pandemics: A vanishing jump approach. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 74(4):1150–1182, Aug. 2025. ISSN 0035-9254, 1467-9876. doi:10.1093/jrsssc/qlaf018.

[5] G. Gungah and J. Narsoo. A novel EVT-modified Lee-Carter model for mortality forecasting : An application to extreme mortality events. *Journal of Statistics and Management Systems*, 25 (1):211–243, Jan. 2022. ISSN 0972-0510, 2169-0014. doi:10.1080/09720510.2021.1878629.

[6] Y. Guo and J. S.-H. Li. Robust parameter estimation for the Lee-Carter family: A probabilistic principal component approach. *Insurance: Mathematics and Economics*, 125:103164, Nov. 2025. ISSN 01676687. doi:10.1016/j.insmatheco.2025.103164.

[7] D. Hainaut. Multidimensional Lee–Carter model with switching mortality processes. *Insurance: Mathematics and Economics*, 50(2):236–246, Mar. 2012. ISSN 01676687. doi:10.1016/j.insmatheco.2011.11.003.

[10] Y. Liu and J. S.-H. Li. The age pattern of transitory mortality jumps and its impact on the pricing of catastrophic mortality bonds. *Insurance: Mathematics and Economics*, 64: 135–150, Sept. 2015. ISSN 01676687. doi:10.1016/j.insmatheco.2015.05.005.

[11] C.-W. Wang, H.-C. Huang, and I.-C. Liu. A Quantitative Comparison of the Lee-Carter Model under Different Types of Non-Gaussian Innovations. *The Geneva Papers on Risk and Insurance - Issues and Practice*, 36(4):675–696, Oct. 2011. ISSN 1018-5895, 1468-0440. doi:10.1057/gpp.2011.20.

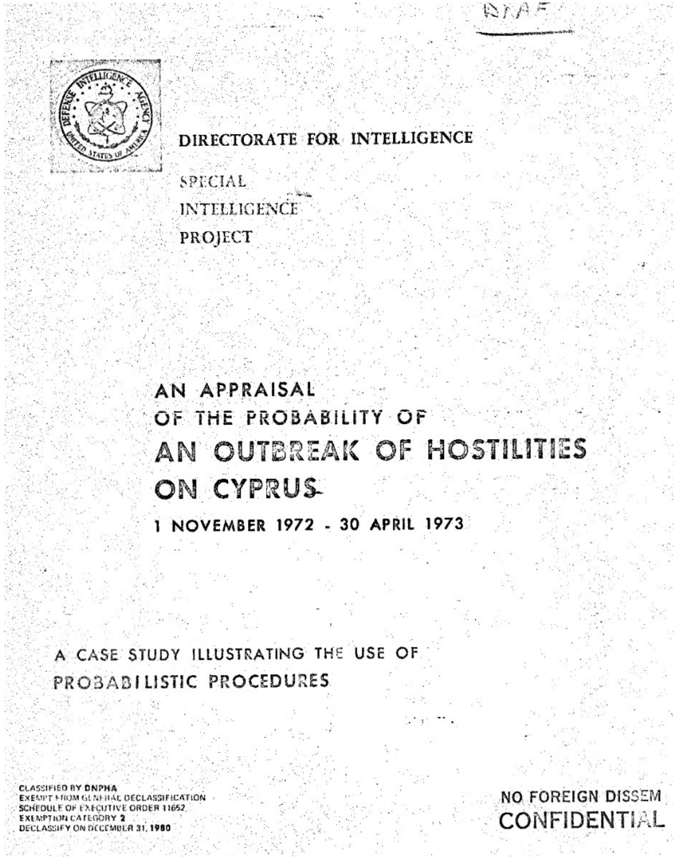
Forecasting at the end of history limits of the Lee-Carter model

Table 1: Comparison of approaches for dealing with mortality shocks in Lee-Carter models.

			Estimate shock					Forecast	
	Non-shock Lexis	Excess Lexis	Timing	Magnitude	Length	Probability	Age-profile	with shocks	without shocks
Shock removal	+	-	-	-	-	-	-	-	+
Shock indicator	+	+	-	+	-	+	-	+	+
Heavy tailed	-	-	-	-	-	+	-	+	-
Vanishing jumps	+	+	+	+	-	+	+	+	+
HIMALC	+	+	+	+	+	+	+	+	+

Kissinger's Markov Chain

Kissinger’s Markov chain



DIRECTORATE FOR INTELLIGENCE

SPECIAL
INTELLIGENCE
PROJECT

AN APPRAISAL
OF THE PROBABILITY OF
AN OUTBREAK OF HOSTILITIES
ON CYPRUS
1 NOVEMBER 1972 - 30 APRIL 1973

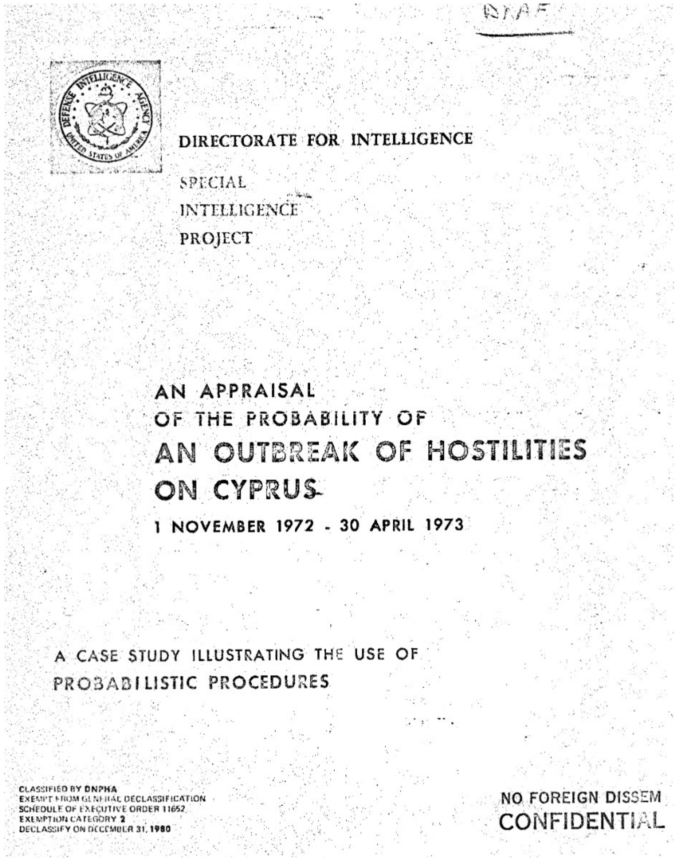
A CASE STUDY ILLUSTRATING THE USE OF
PROBABILISTIC PROCEDURES

CLASSIFIED BY DNPHA
EXEMPT FROM GENERAL DECLASSIFICATION
SCHEDULE OF EXECUTIVE ORDER 11652
EXEMPTION CATEGORY 2
DECLASSIFY ON DECEMBER 31, 1980

NO FOREIGN DISSEM
CONFIDENTIAL

Marshall. (1973). Experimental Intelligence Products to Improve the Communication of Uncertainty.
cia.gov/readingroom/docs/LOC-HAK-538-3-5-6.pdf

Kissinger's Markov chain



DIRECTORATE FOR INTELLIGENCE

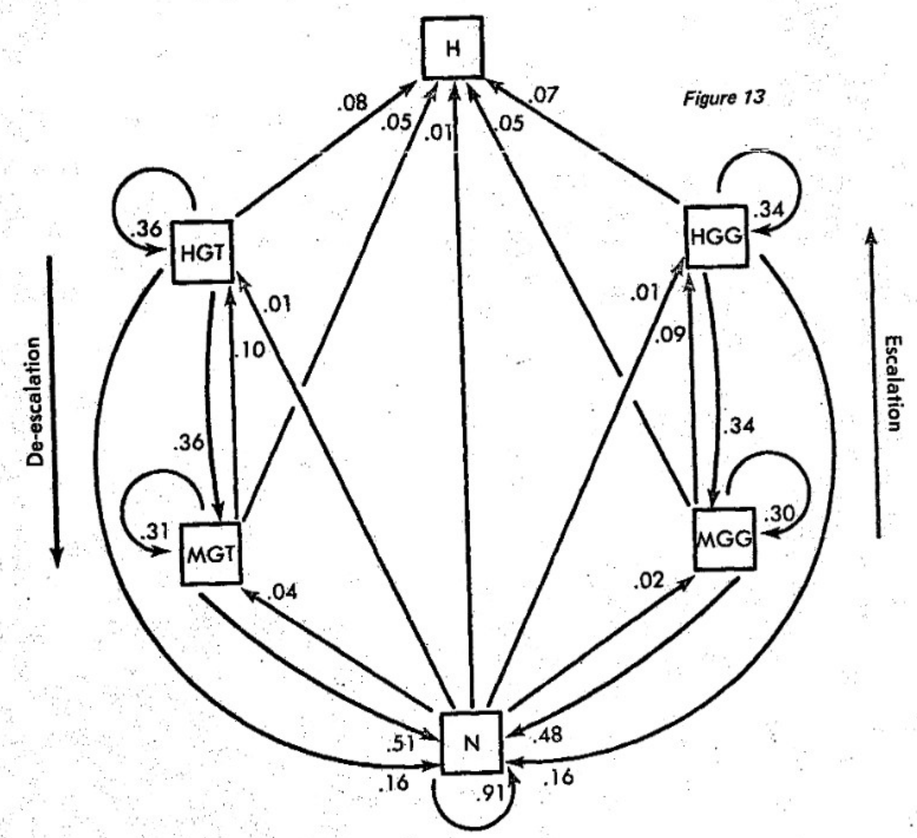
SPECIAL
INTELLIGENCE
PROJECT

AN APPRAISAL
OF THE PROBABILITY OF
AN OUTBREAK OF HOSTILITIES
ON CYPRUS
1 NOVEMBER 1972 - 30 APRIL 1973

A CASE STUDY ILLUSTRATING THE USE OF
PROBABILISTIC PROCEDURES

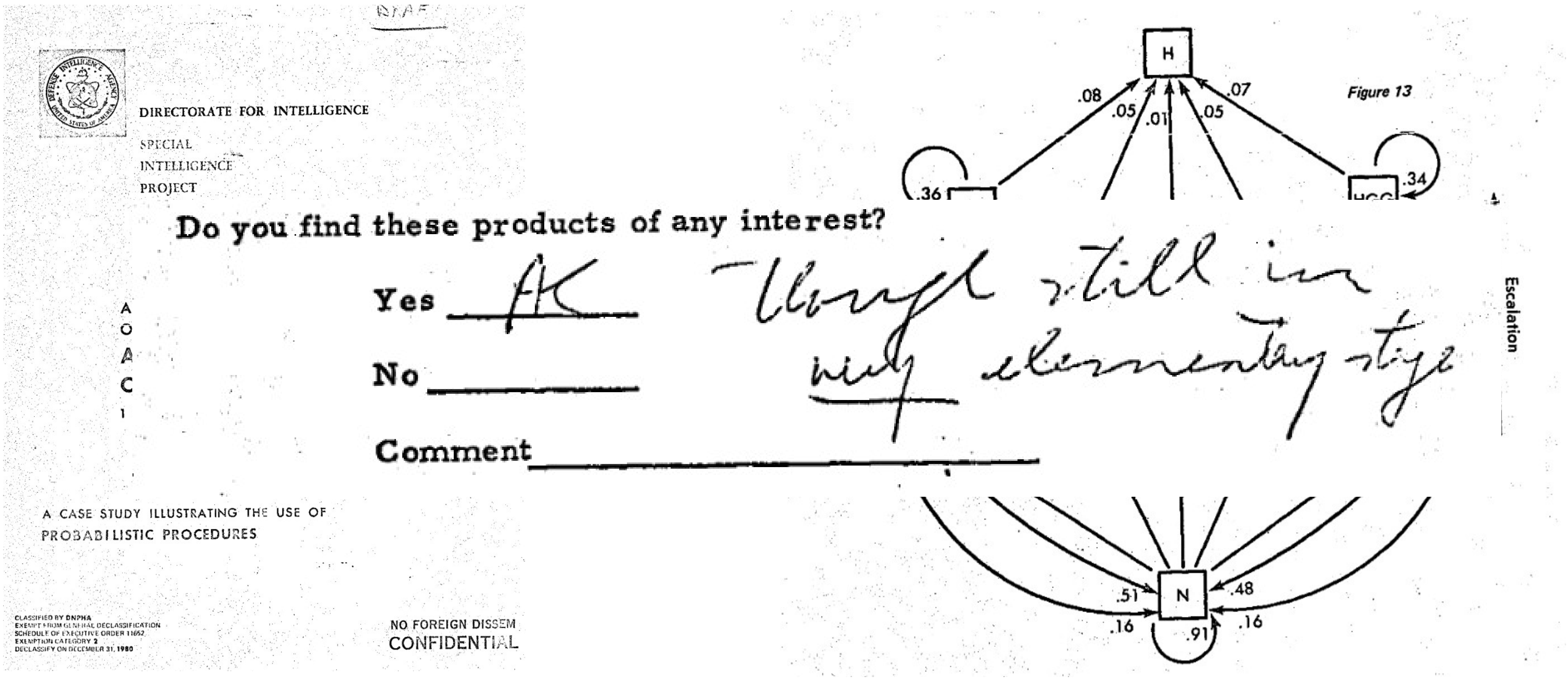
CLASSIFIED BY DNPHA
EXEMPT FROM GENERAL DECLASSIFICATION
SCHEDULE OF EXECUTIVE ORDER 11652
EXEMPTION CATEGORY 2
DECLASSIFY ON DECEMBER 31, 1980

NO FOREIGN DISSEM
CONFIDENTIAL



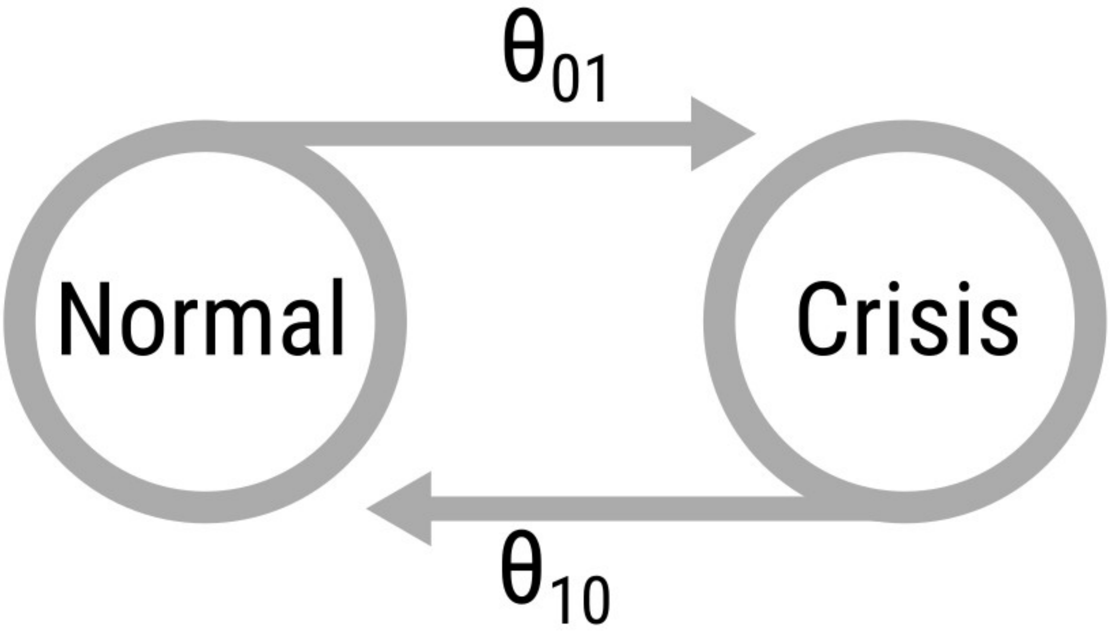
Marshall. (1973). Experimental Intelligence Products to Improve the Communication of Uncertainty.
cia.gov/readingroom/docs/LOC-HAK-538-3-5-6.pdf

Kissinger’s Markov chain



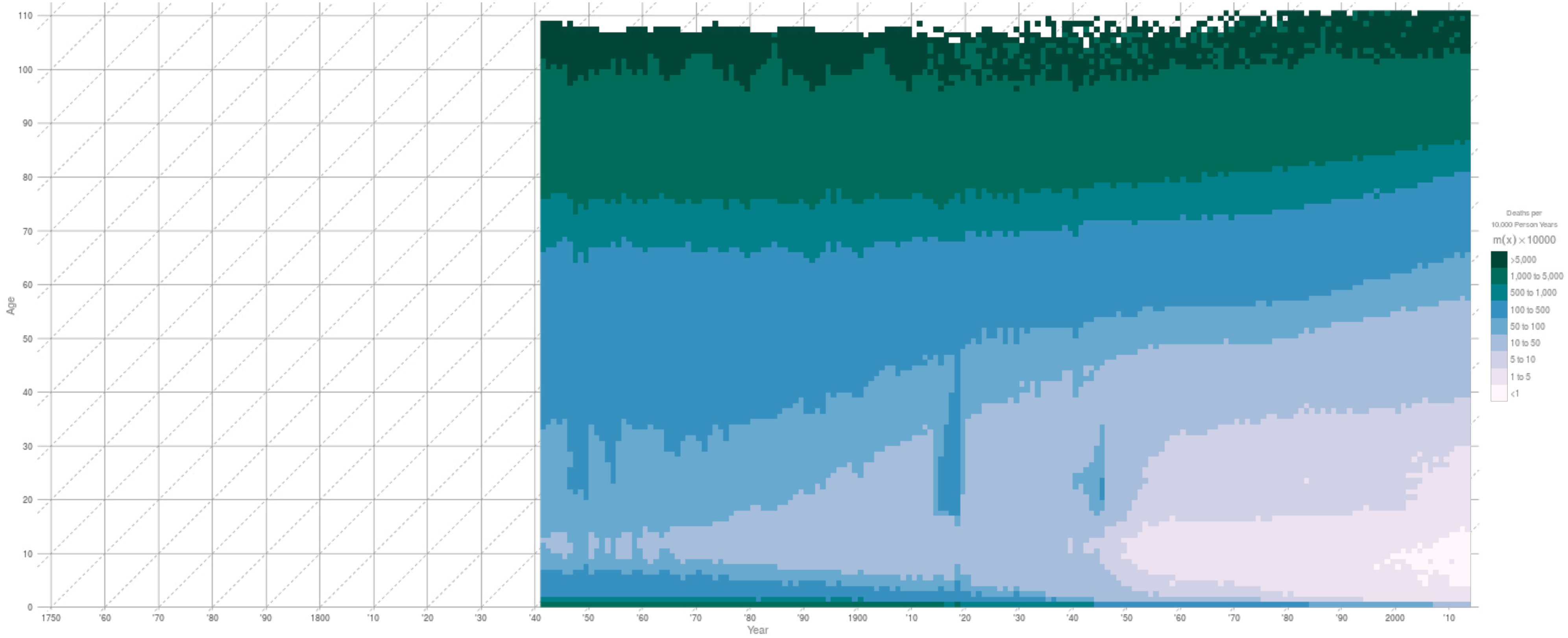
Marshall. (1973). Experimental Intelligence Products to Improve the Communication of Uncertainty.
cia.gov/readingroom/docs/LOC-HAK-538-3-5-6.pdf

Kissinger's Markov chain



Hidden Markov Lee-Carter

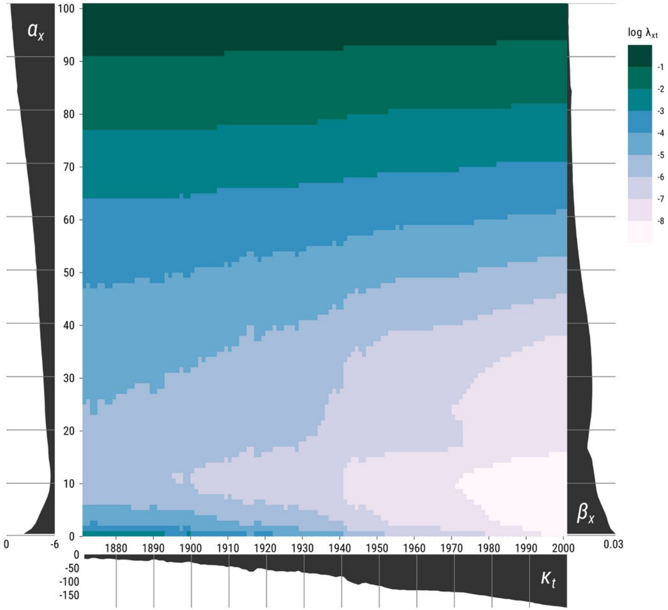
Hidden Markov Lee Carter



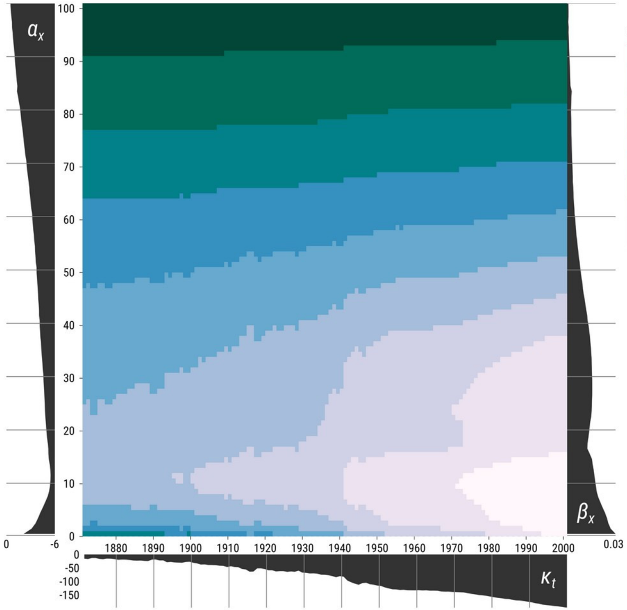
Human Mortality Database (2016). England & Wales death rates. mortality.org

Schöley (2016). The Human Mortality Explorer. jschoeley.shinyapps.io/hmdexp/

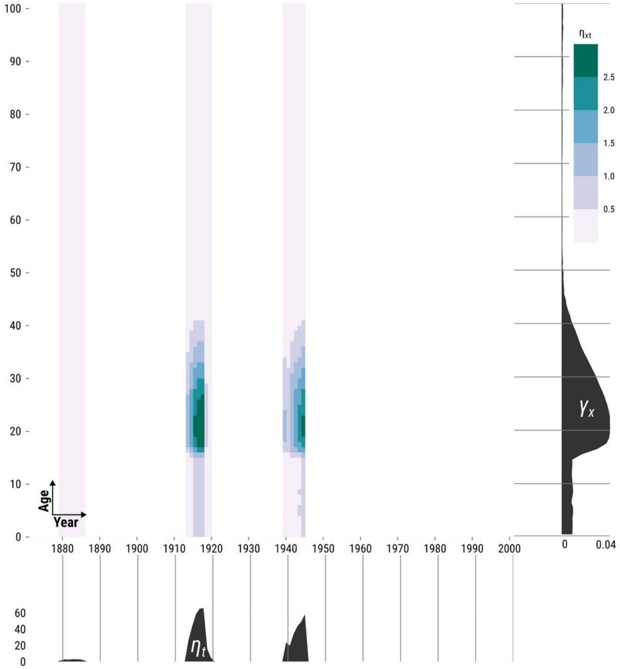
Hidden Markov Lee Carter



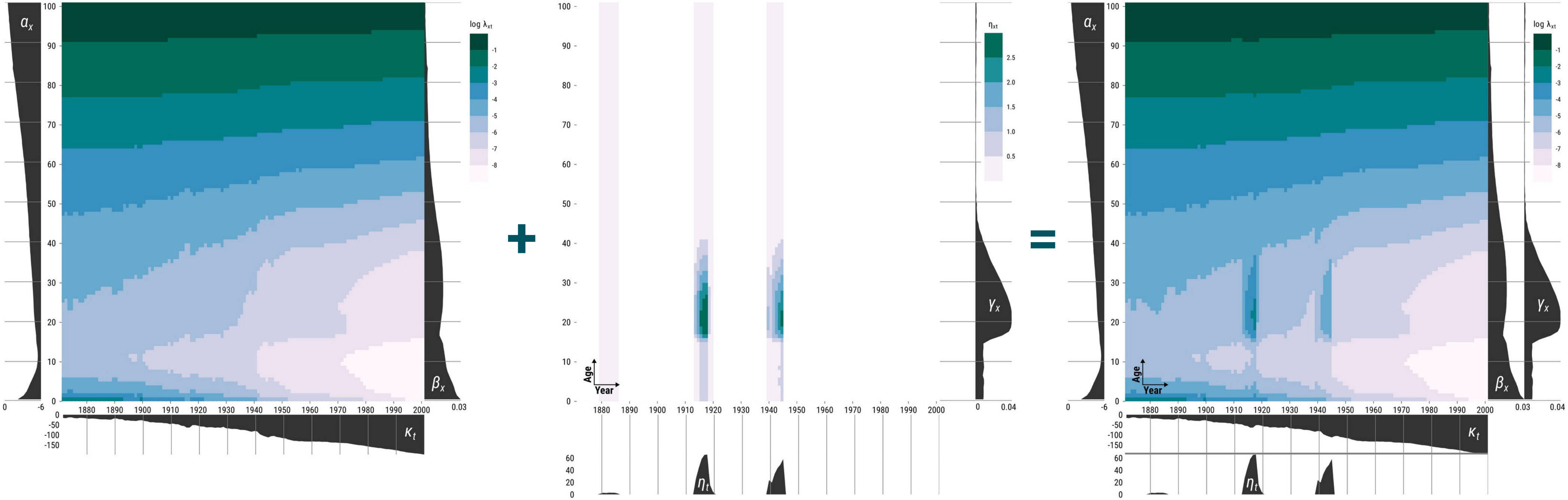
Hidden Markov Lee Carter



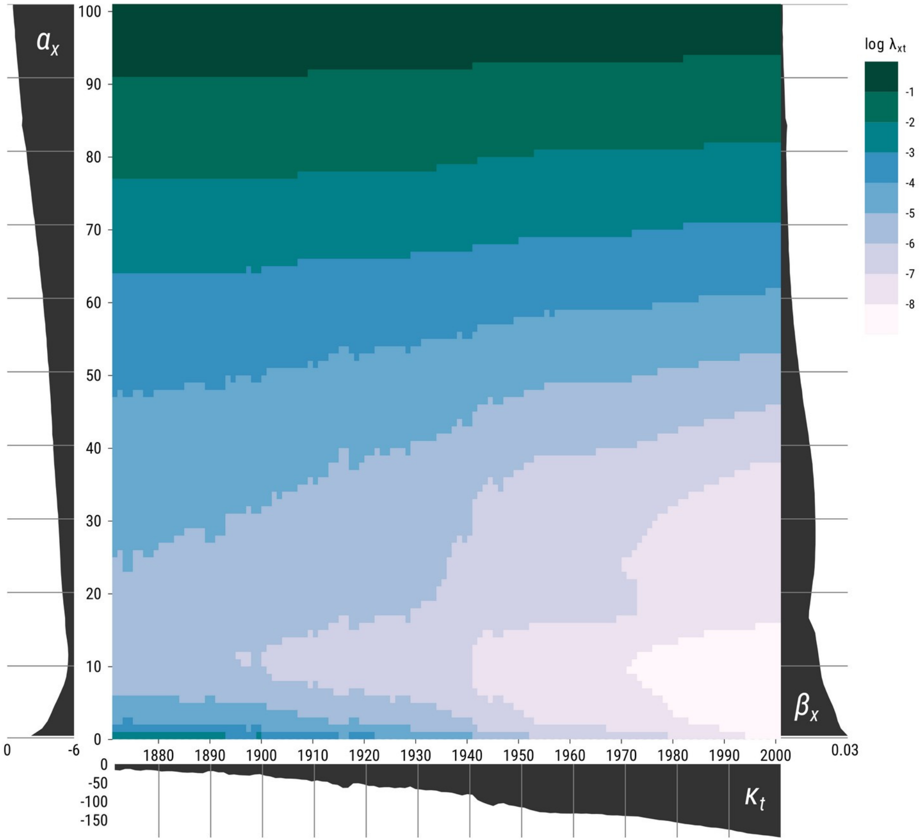
+



Hidden Markov Lee Carter



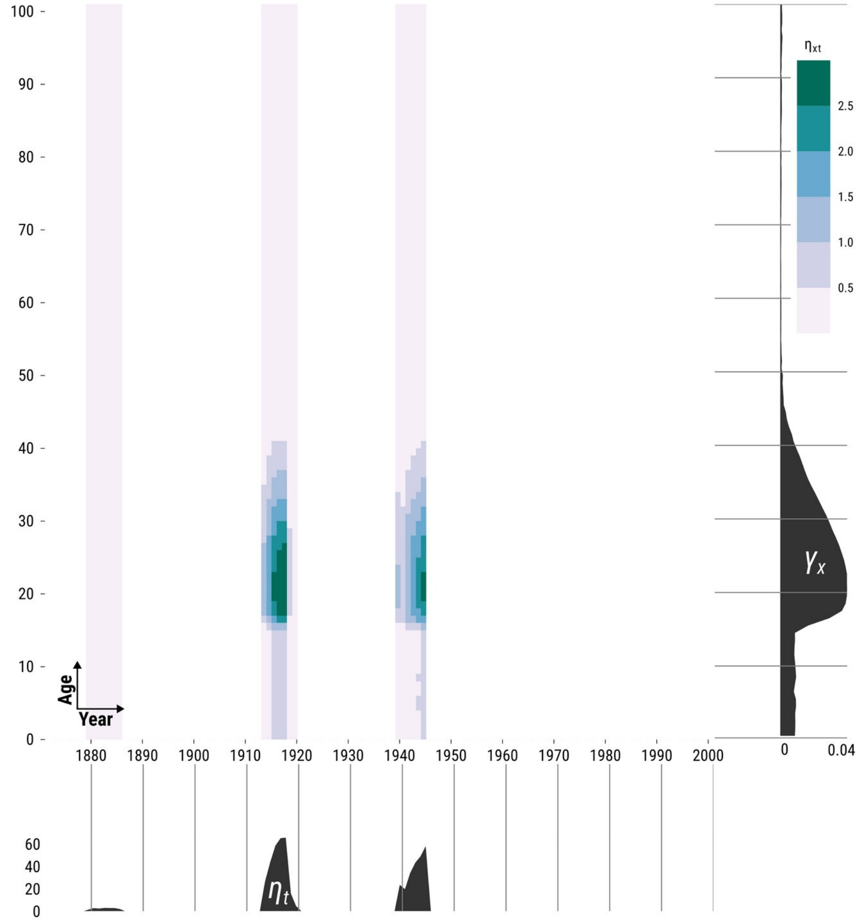
Hidden Markov Lee Carter



$$D_{xt} \sim \text{NegBin}(\lambda_{xt} E_{xt}, \varphi)$$

$$\lambda_{xt} = \exp(\alpha_x + \beta_x \kappa_t + \eta_t \gamma_x)$$

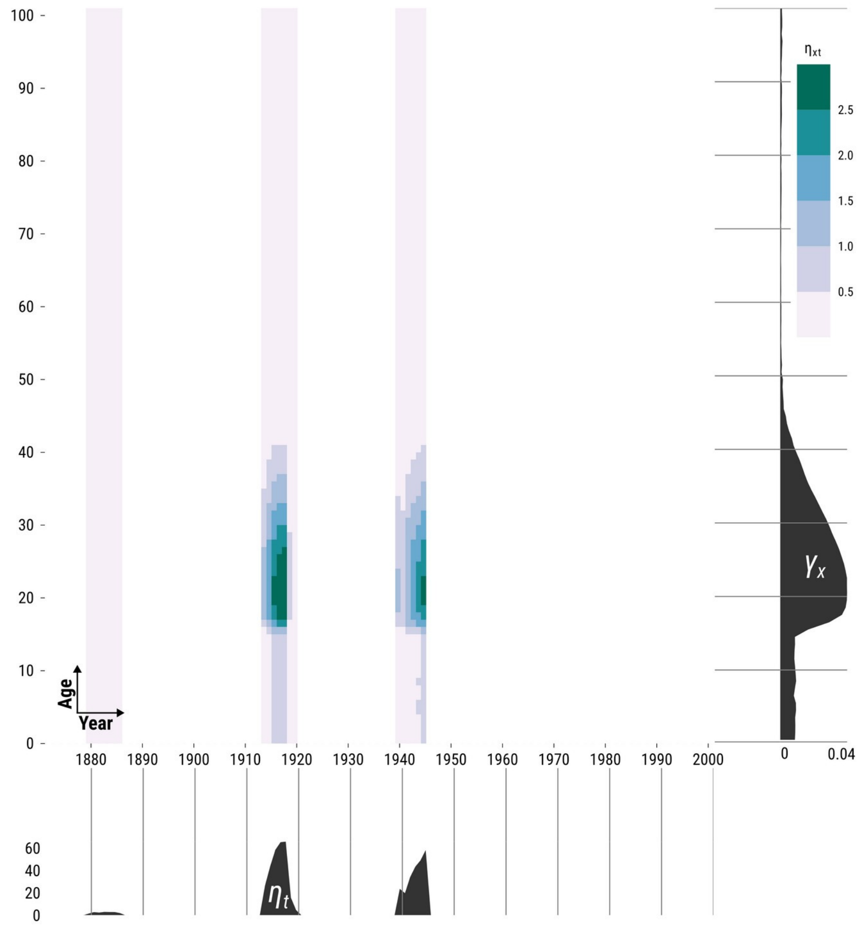
Hidden Markov Lee Carter



$$D_{xt} \sim \text{NegBin}(\,\lambda_{xt} E_{xt}\,,\, \varphi\,)$$

$$\lambda_{xt} = \exp(\,\alpha_x + \beta_x \kappa_t + \eta_t \gamma_x\,)$$

Hidden Markov Lee Carter

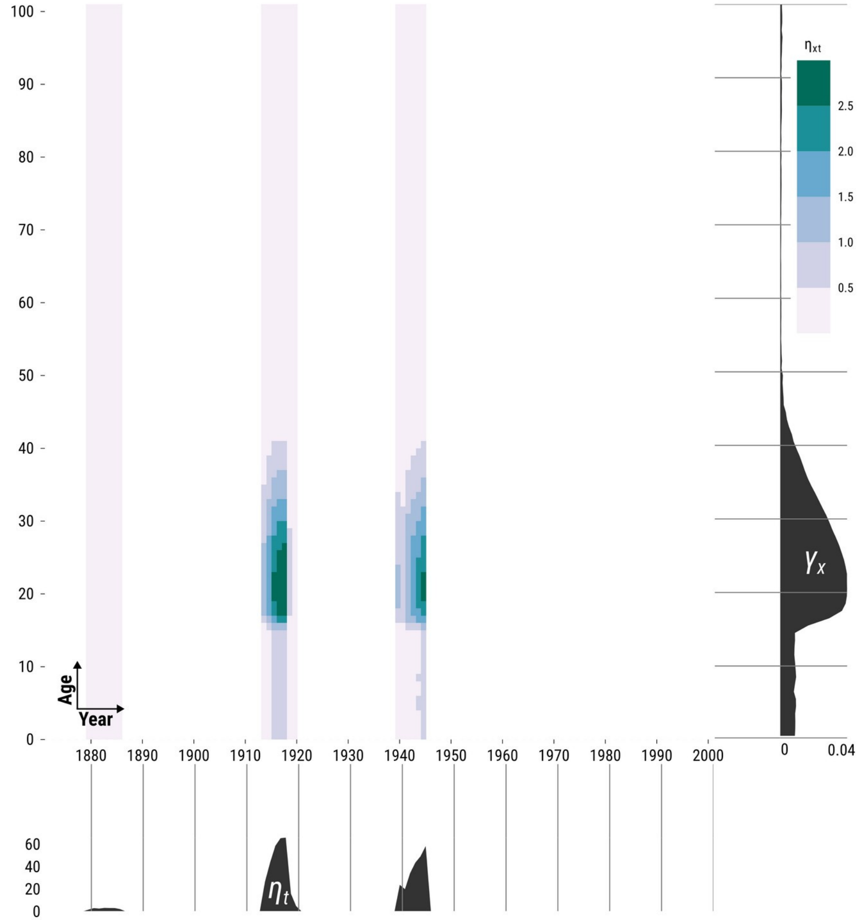


$$D_{xt} \sim \text{NegBin}(\lambda_{xt} E_{xt}, \varphi)$$

$$\lambda_{xt} = \exp(\alpha_x + \beta_x \kappa_t + \eta_t \gamma_x)$$

$$\eta_t = z_t u_t$$

Hidden Markov Lee Carter

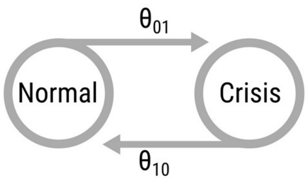


$$D_{xt} \sim \text{NegBin}(\lambda_{xt} E_{xt}, \varphi)$$

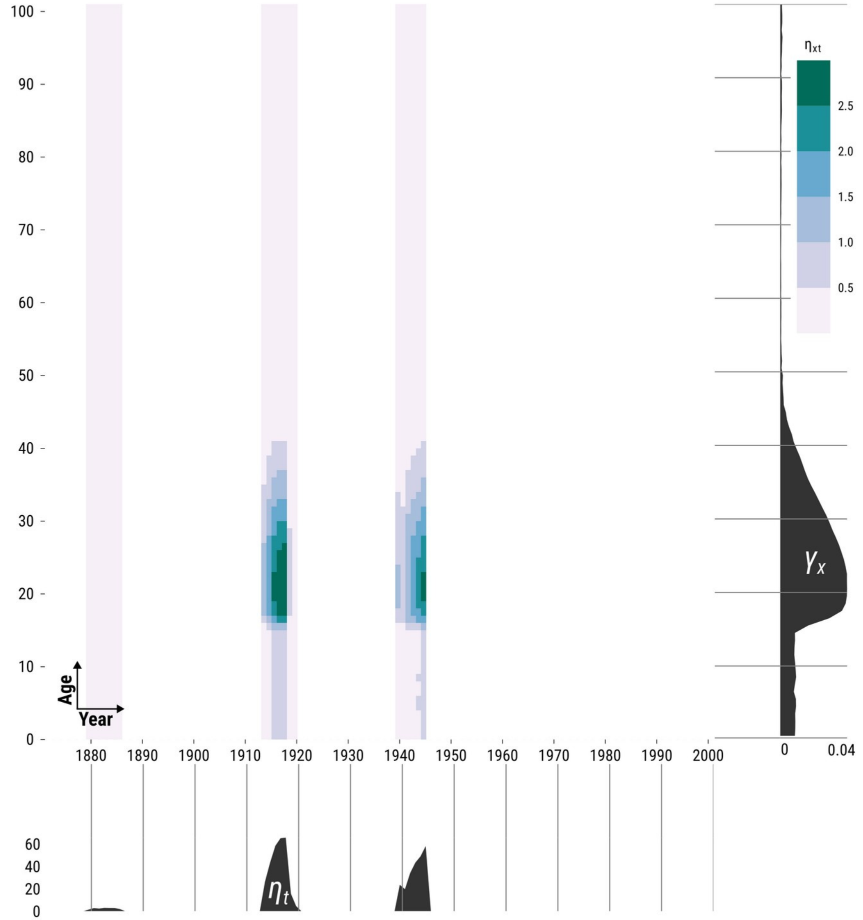
$$\lambda_{xt} = \exp(\alpha_x + \beta_x \kappa_t + \eta_t \gamma_x)$$

$$\eta_t = z_t u_t$$

$$z_t \sim \text{Bernoulli}(p_t)$$



Hidden Markov Lee Carter

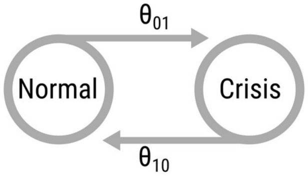


$$D_{xt} \sim \text{NegBin}(\lambda_{xt} E_{xt}, \varphi)$$

$$\lambda_{xt} = \exp(\alpha_x + \beta_x \kappa_t + \eta_t \gamma_x)$$

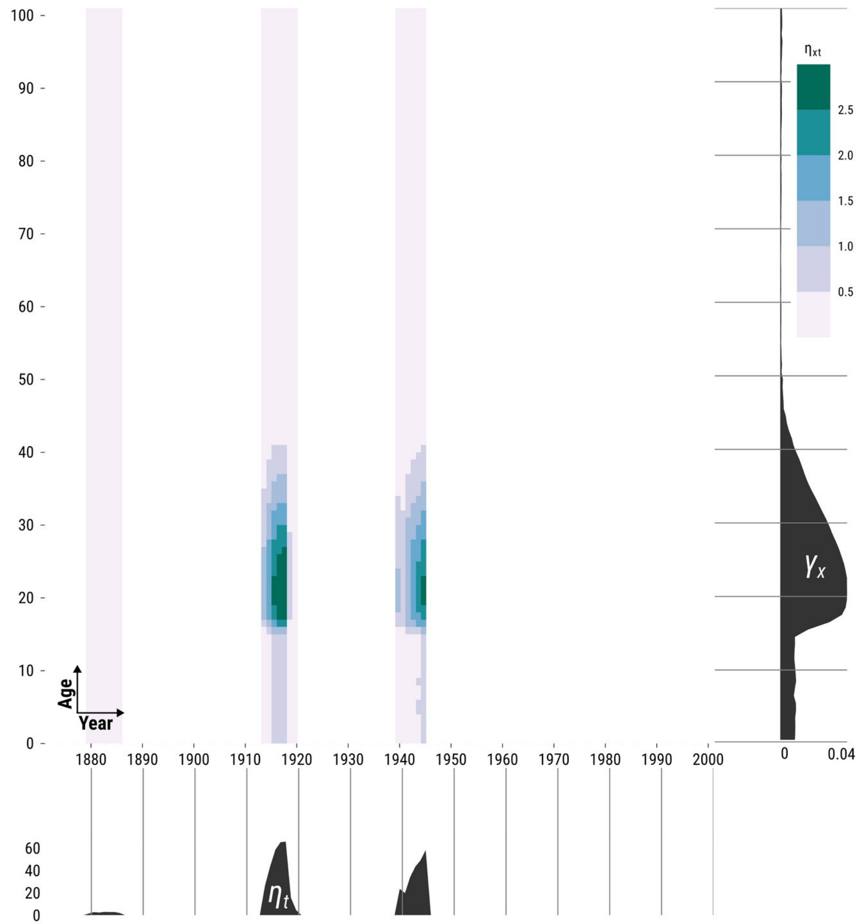
$$\eta_t = z_t u_t$$

$$z_t \sim \text{Bernoulli}(p_t)$$



$$p_t = z_{t-1}\theta_{1\to1} + (1-z_{t-1})\theta_{0\to1}$$

Hidden Markov Lee Carter

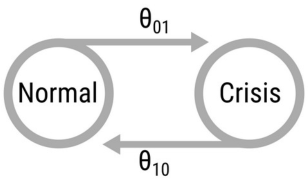


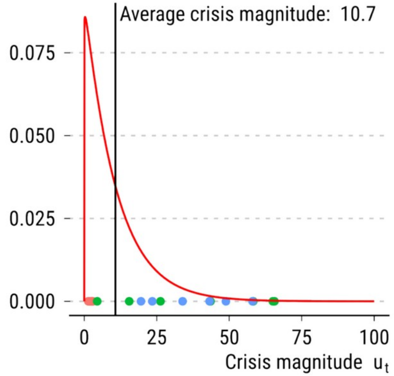
$$D_{xt} \sim \text{NegBin}(\lambda_{xt} E_{xt}, \varphi)$$

$$\lambda_{xt} = \exp(\alpha_x + \beta_x \kappa_t + \eta_t \gamma_x)$$

$$\eta_t = z_t u_t \qquad u_t \sim \text{Gamma}(a, b)$$

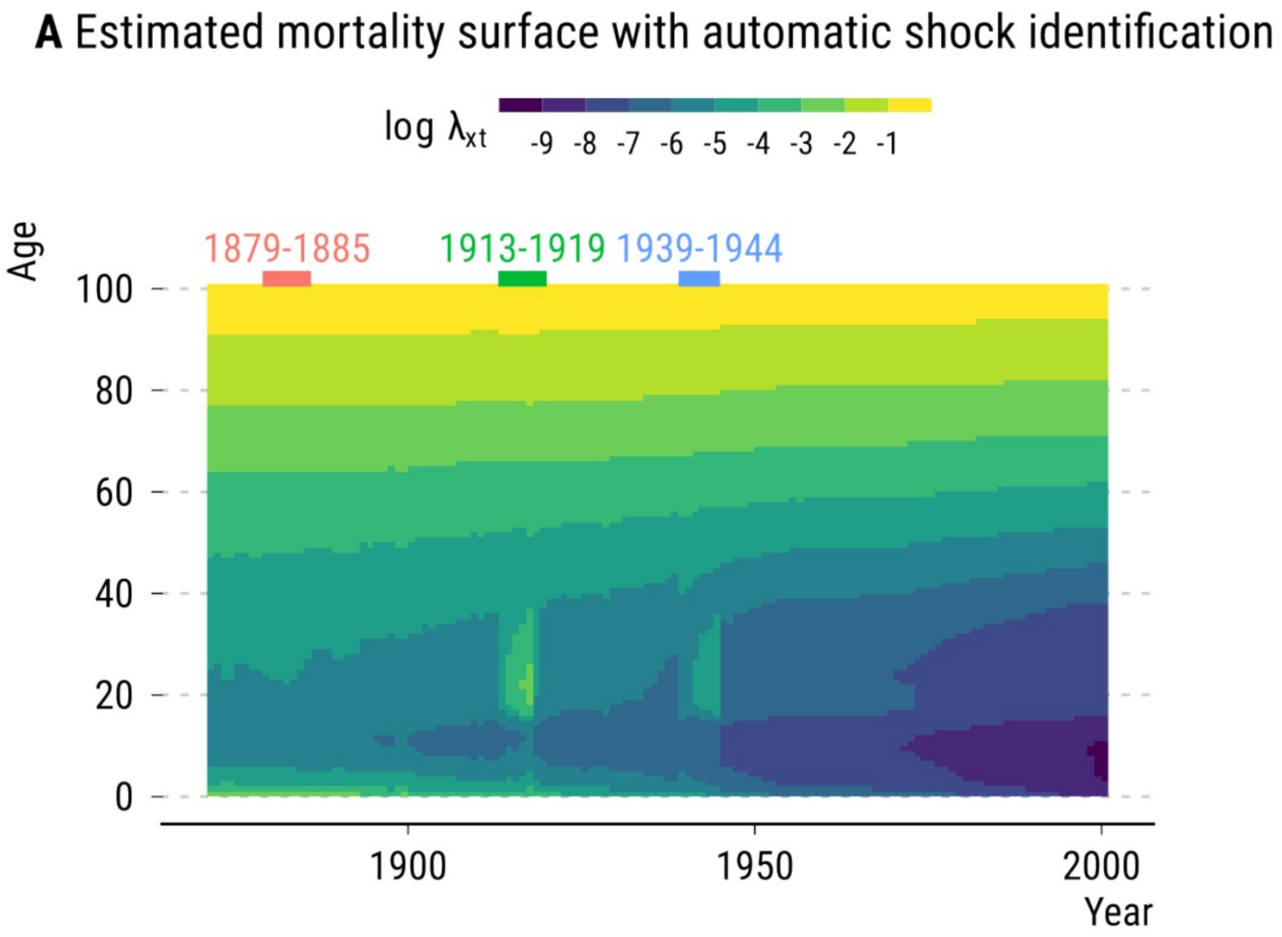
$$z_t \sim \text{Bernoulli}(p_t)$$



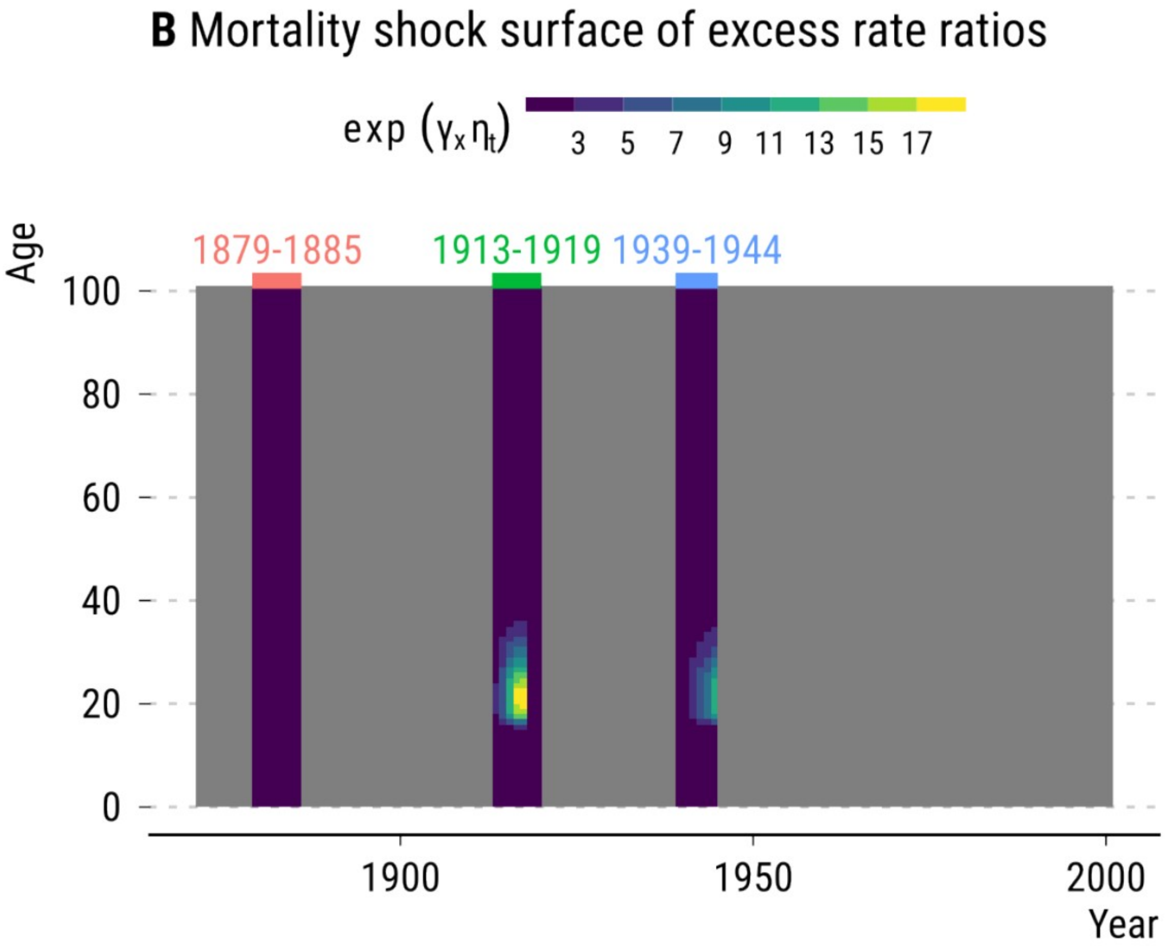


$$p_t = z_{t-1}\theta_{1\to1} + (1-z_{t-1})\theta_{0\to1}$$

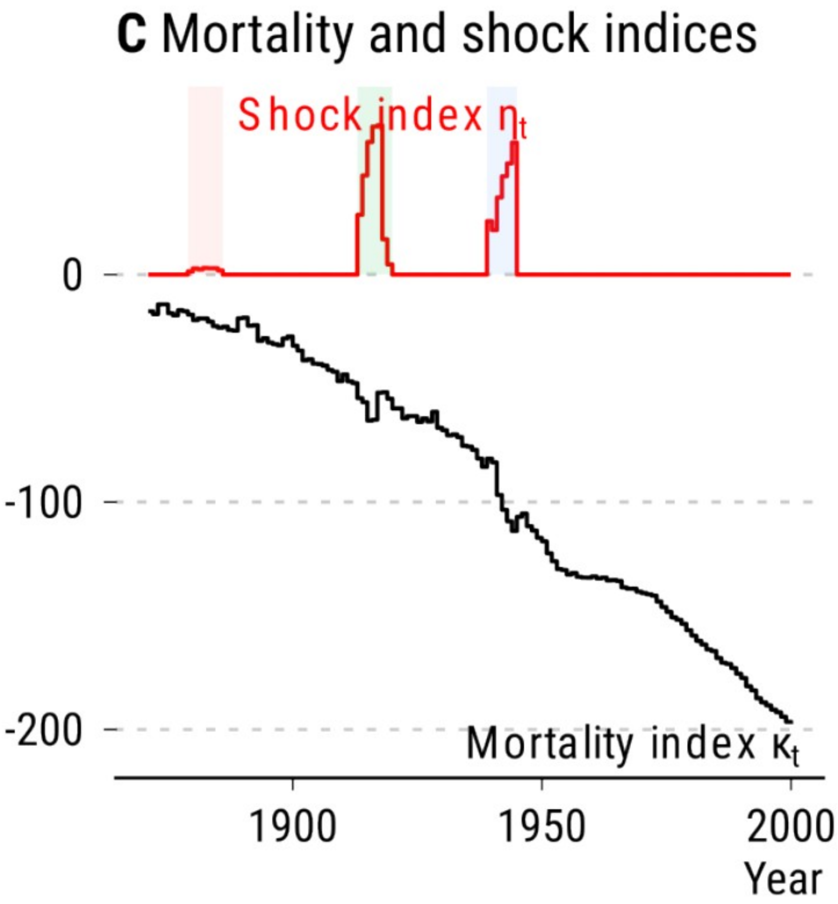
Identification



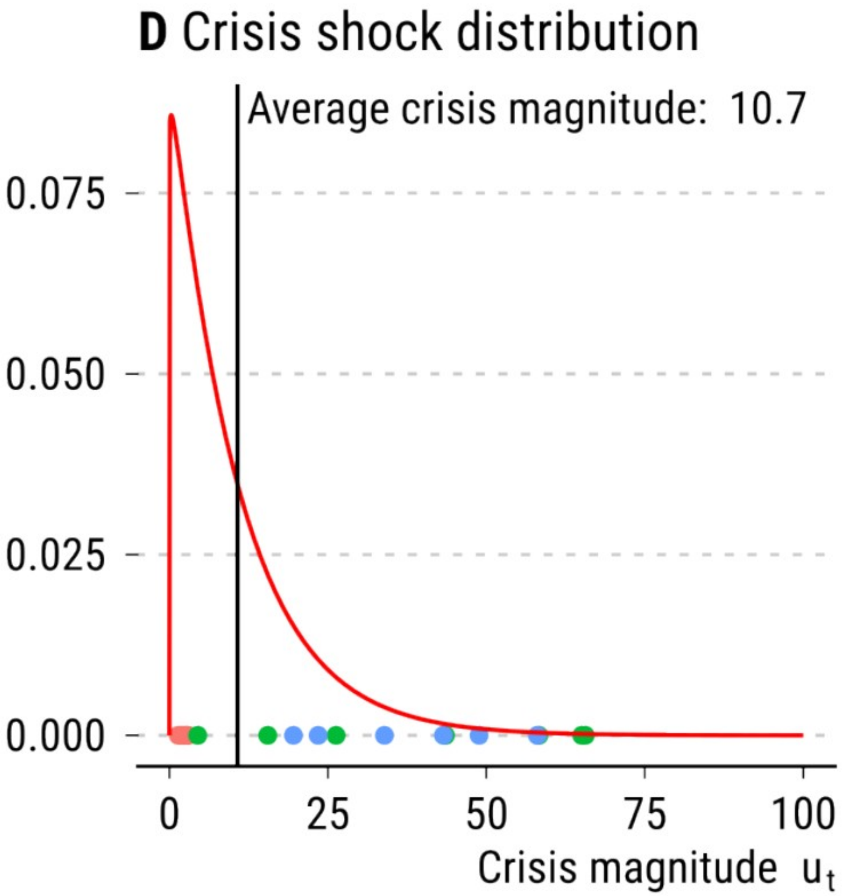
Characterization



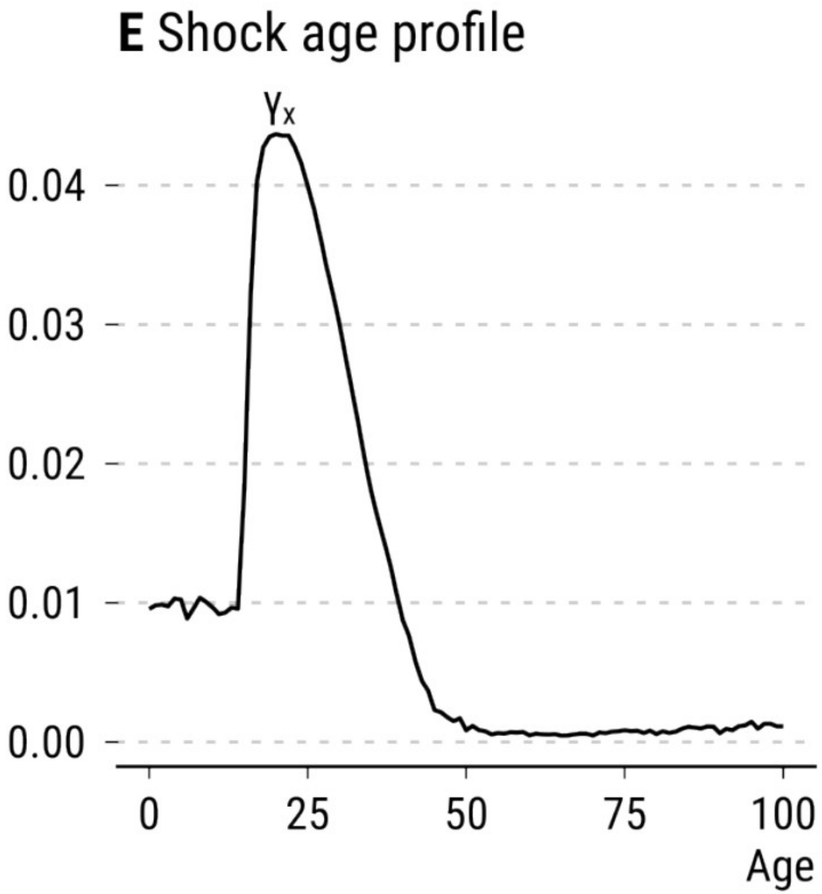
Characterization



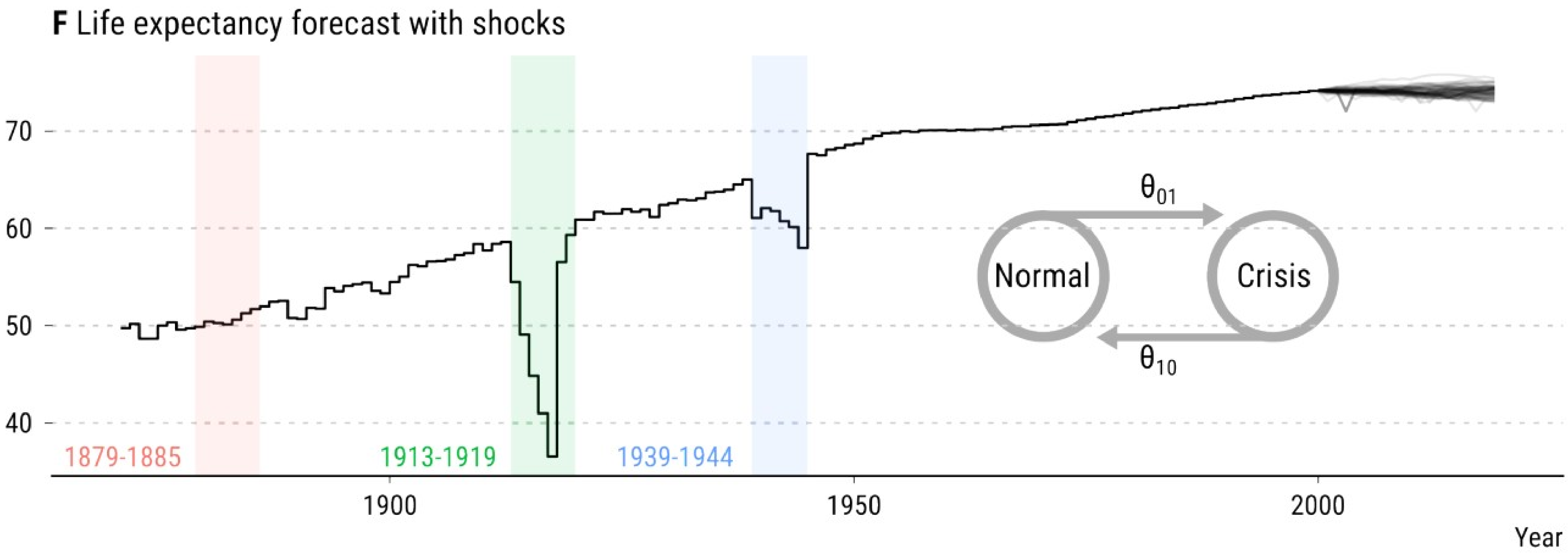
Characterization



Characterization



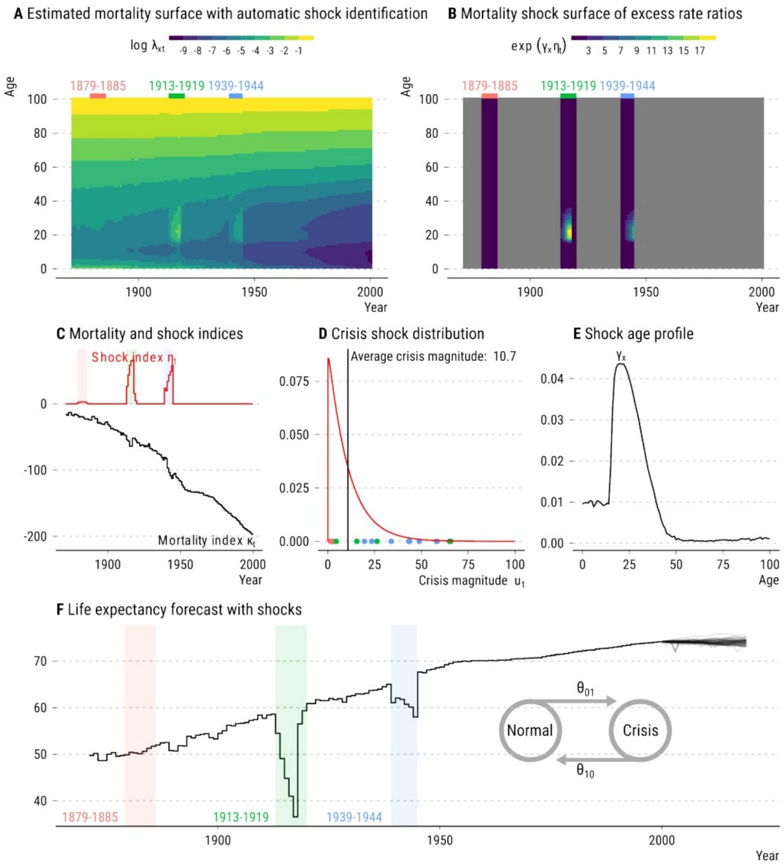
Simulation



F Life expectancy forecast with shocks
70
60
50
40
1879-1885
1913-1919
1939-1944
1900
1950
2000
Year
θ01
Normal
Crisis
θ10

Identification, Characterization, Simulation

Figure 1: Hidden Markov Lee-Carter Model estimates for male mortality in England and Wales (Data: HMD).



Jonas Schöley
jschoeley.com

@jschoeley.com
0000-0002-3340-8518
schoeley@demogr.mpg.de

MAX PLANCK INSTITUTE
FOR DEMOGRAPHIC RESEARCH