A Bayesian approach to modeling mortality rates

Cary Chi-Liang Tsai (with Dr. Tzuling Lin)

Department of Statistical and Actuarial Science
Simon Fraser University, British Columbia, Canada

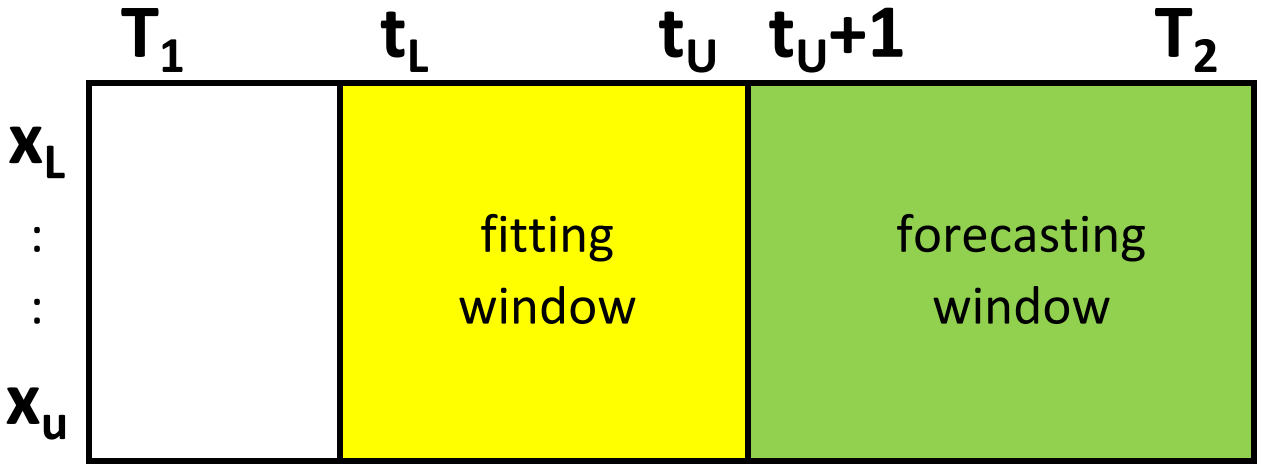
Snell Actuarial Science and Risk Management Seminar Series
Department of Finance
University of Nebraska-Lincoln, U.S.A.
September 6th, 2019

Outline

Motivations of developing effective mortality models

- Modeling the changes and dynamics of mortality rates is critical to the solvency of life insurers and social benefit programs.
- Mortality is also one of the key factors in pricing and reserving of life insurance and annuity products.
- Failing in capturing the downward trends in mortality rates would under-price/-reserve annuity products and then expose annuity providers and social security systems to the risk of financial insolvency.

Fitting window and forecasting window



- $q_{x,t}$ ($p_{x,t} = 1 - q_{x,t}$): one-year death (survival) probability, and $\mu_{x,t}$ ($m_{x,t}$): the associated force of mortality (central death rate).
- Assume $\mu_{x,t}$ is constant within each integer age x and year t (i.e., $\mu_{x+r,t+s} = \mu_{x,t}$ for $r, s \in [0, 1)$), $\Rightarrow \mu_{x,t} = m_{x,t} = -\ln(p_{x,t})$.
- Let $[x_L, x_L] \times [T_1, T_2]$ be the study age-year window where mortality rates are available.
- Stand at the end of year t_U and want to fit the models with mortality data in an age-year window $[x_L, x_U] \times [t_L, t_U]$, where $T_1 \leq t_L$.
- Project mortality rates for the window $[x_L, x_U] \times [t_U + 1, T_2]$, and compare them with the observed ones in $[x_L, x_U] \times [t_U + 1, T_2]$.

The Lee-Carter Model (I)

$$\ln(m_{x,t}) = \alpha_x + \beta_x \cdot k_t + \epsilon_{x,t},\ x = x_L, \ldots, x_U,\ t = t_L, \ldots, t_U$$

- α_x and β_x are the age-specific constants;
- k_t is the time-varying index;
- for each x, error terms $\epsilon_{x,t} \overset{i.i.d}{\sim} N(0, \sigma_x^2)$, $t = t_L, \ldots, t_U$;
- two constraints: $\sum_t k_t = 0$ and $\sum_x \beta_x = 1$.
- the fitting year span $[t_L,\ t_U]$ and the fitting age span $[x_L,\ x_U]$ $(x_U - x_L + 1 = m)$.

Estimation of parameters:

- $\sum_{t=t_L}^{t_U} k_t = 0 \Rightarrow \hat{\alpha}_x = \frac{1}{n} \sum_{t=t_L}^{t_U} \ln(m_{x,t})$,
- $\sum_{x=x_L}^{x_U} \beta_x = 1 \Rightarrow \hat{k}_t = \sum_{x=x_L}^{x_U} [\ln(m_{x,t}) - \hat{\alpha}_x]$,
- Regressing $[\ln(m_{x,t}) - \hat{\alpha}_x]$ on $\hat{k}_t$ without the intercept term $\Rightarrow \hat{\beta}_x = \frac{\sum_t [\ln(m_{x,t}) - \hat{\alpha}_x] \times \hat{k}_t}{\sum_t \hat{k}_t^2}$,

The Lee-Carter Model (II)

- k_t is assumed to follow a random walk with drift process, $\hat{k}_t = \hat{k}_{t-1} + \theta^{LC} + \varepsilon_t$, where $\varepsilon_t \sim N(0, \sigma^2)$, or $\hat{k}_t - \hat{k}_{t-1} \sim N(\theta^{LC}, \sigma^2)$.
- The drift θ^{LC} is estimated by

$$\hat{\theta}^{LC} = \frac{\sum_{t=t_L+1}^{t_U} (\hat{k}_t - \hat{k}_{t-1})}{t_U - t_L} = \frac{\hat{k}_{t_U} - \hat{k}_{t_L}}{t_U - t_L} = \frac{\sum_{x=x_L}^{x_U} [\ln(m_{x,\,t_U}) - \ln(m_{x,\,t_L})]}{t_U - t_L}.$$

- $\hat{k}_{t_U+\tau} = \hat{k}_{t_U} + \sum_{t=1}^{\tau} (\hat{k}_{t_U+t} - \hat{k}_{t_U+t-1}) = \hat{k}_{t_U} + \hat{\theta}^{LC} \times \tau$
- The projected $\ln(m_{x,\,t_U+\tau})$ is obtained by a linear function of τ,

$$\ln(\hat{m}^{LC}_{x,\,t_U+\tau}) = \hat{\alpha}_x + \hat{\beta}_x \times \hat{k}_{t_U+\tau} = \ln(\hat{m}^{LC}_{x,\,t_U}) + \hat{\mu}^{LC}_x \times \tau,$$

with intercept (starting point) $\ln(\hat{m}^{LC}_{x,\,t_U}) = \hat{\alpha}_x + \hat{\beta}_x \times \hat{k}_{t_U}$ and slope (annual mortality decrement) $\hat{\mu}^{LC}_x = \hat{\beta}_x \cdot \hat{\theta}^{LC}$.

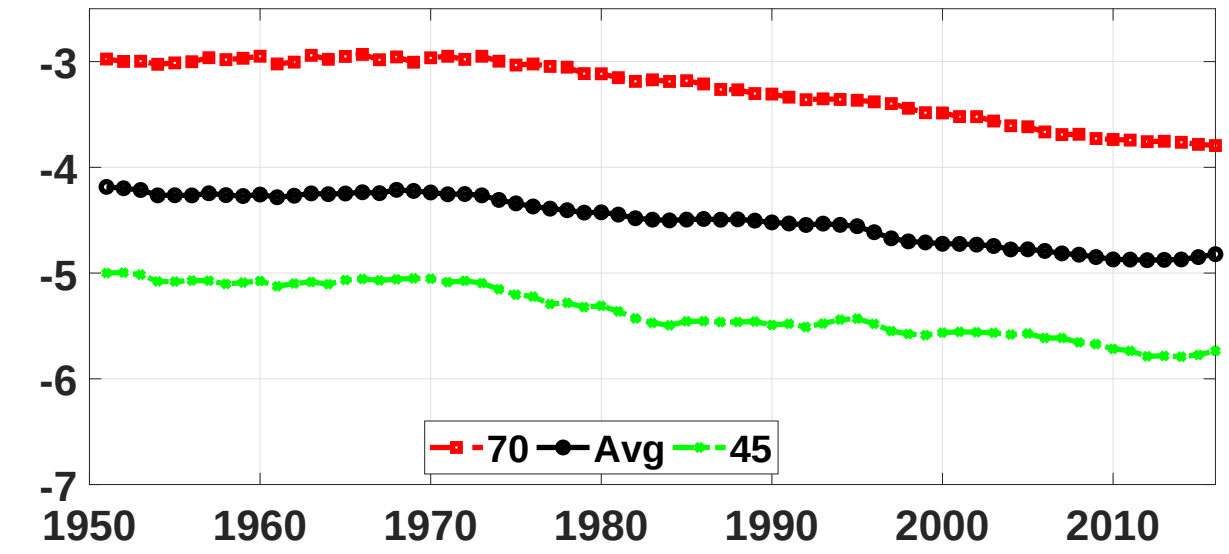
Random Walk with Drift Model

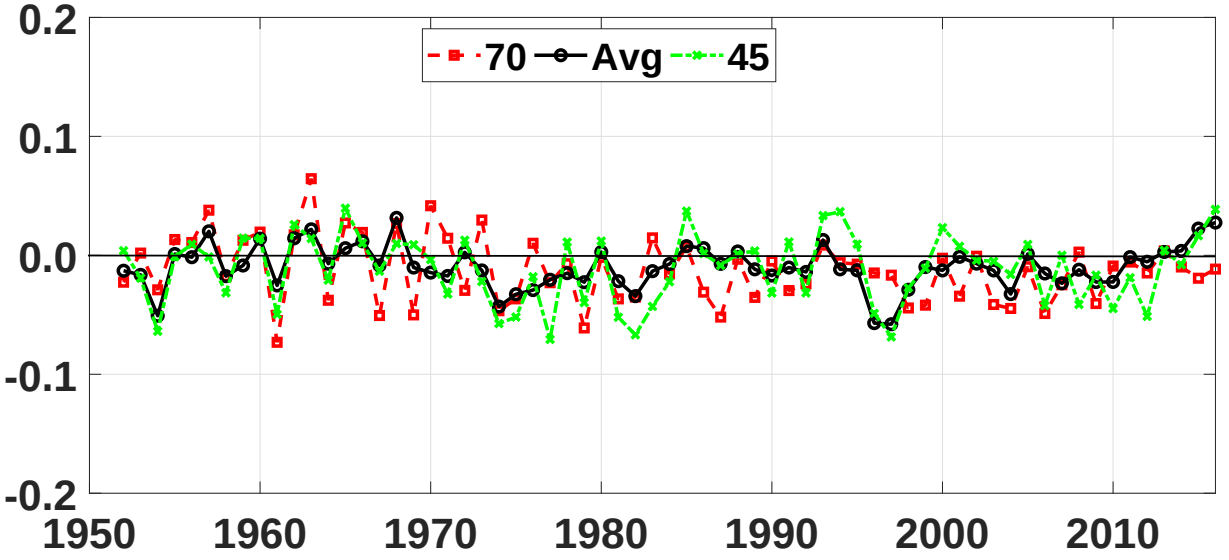
- $\ln(m_{x,t})$ is assumed to follow a random walk with drift process for each x, $\ln(m_{x,t+1}) = \ln(m_{x,t}) + \mu_x^{RW} + \varepsilon_{x,t}$, where $\varepsilon_{x,t} \overset{iid}{\sim} N(0, \sigma_x^2)$, or $Y_{x,t} \overset{\triangle}{=} \ln(m_{x,t_L+t}) - \ln(m_{x,t_L+t-1}) \overset{iid}{\sim} N(\mu_x^{RW}, \sigma_x^2)$, $t = 1, 2, \cdots$.
- The drift μ_x^{RW} is estimated by

$$\hat{\mu}_x^{RW} = \frac{\sum_{t=t_L}^{t_U-1} [\ln(m_{x,t+1}) - \ln(m_{x,t})]}{t_U - t_L} = \frac{\ln(m_{x,t_U}) - \ln(m_{x,t_L})}{t_U - t_L}.$$

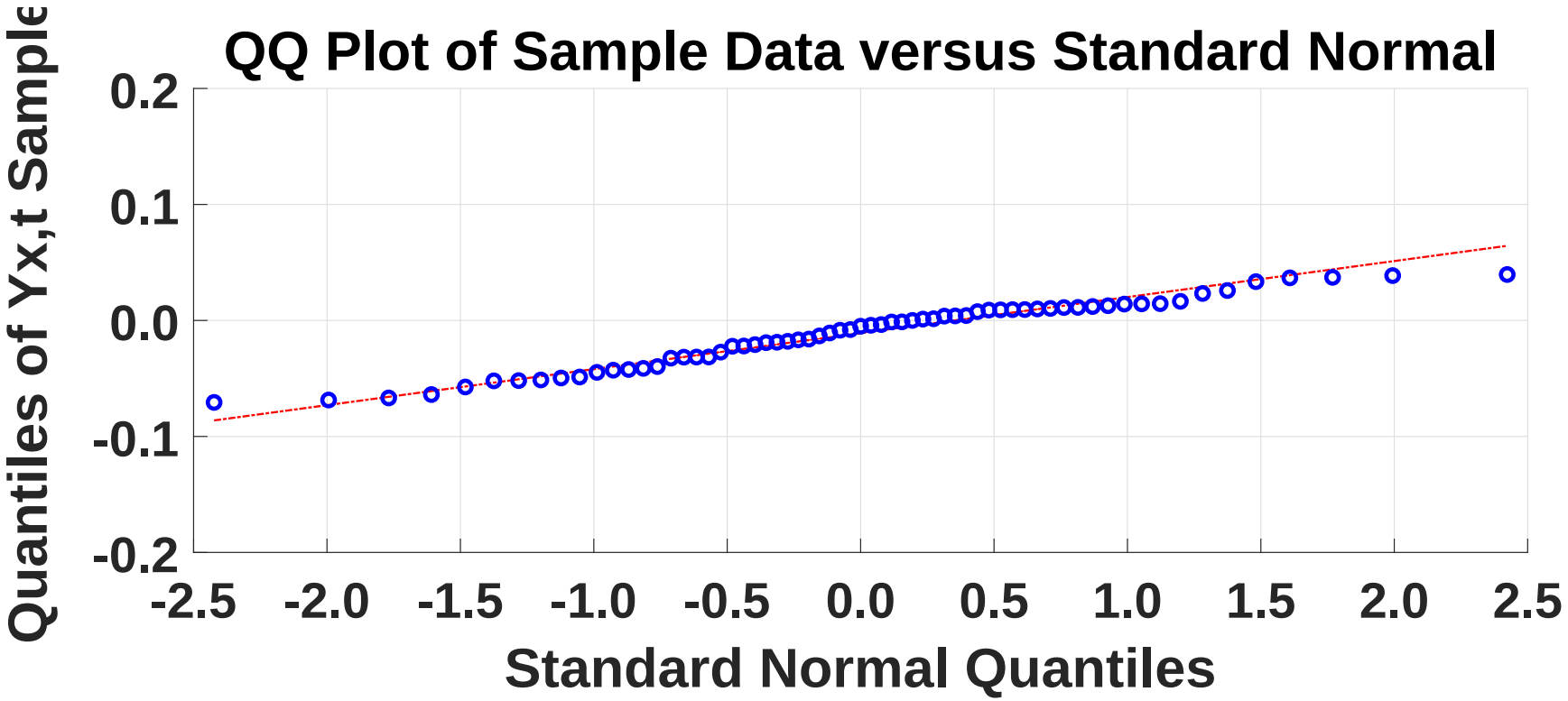
- The projected $\ln(m_{x,t_U+\tau})$ is obtained by a linear function of τ, $\ln(\hat{m}_{x,t_U+\tau}^{RW}) = \ln(m_{x,t_U}) + \sum_{t=1}^{\tau} [\ln(m_{x,t_U+t}) - \ln(m_{x,t_U+t-1})] = \ln(m_{x,t_U}) + \hat{\mu}_x^{RW} \times \tau$, with intercept $\ln(m_{x,t_U})$ and slope $\hat{\mu}_x^{RW}$.

$\ln(m_{x,t})$ and $Y_{x,t}$ for US males aged 45 and 70





Q-Q plot for US males aged 45



Bayesian Theory

- Assume we have observed $\tilde{Y} = \tilde{y}$, where $\tilde{Y} = (Y_1, \ldots, Y_n)^T$ and $\tilde{y} = (y_1, \ldots, y_n)^T$, and want to set a rate to cover Y_{n+1}.
- Let $Y_t|\Lambda, t = 1, \cdots, n+1$, be i.i.d random variables with probability density $f_{Y_t|\Lambda}(y|\lambda)$, where λ be a risk parameter from a r.v. Λ.
- the **prior** pdf of Λ: $\pi_\Lambda(\lambda)$.
- the **joint** pdf of $\tilde{Y}$ and Λ:
  $f_{\tilde{Y},\Lambda}(\tilde{y}, \lambda) = f_{\tilde{Y}|\Lambda}(y_1, \cdots, y_n|\lambda) \cdot \pi_\Lambda(\lambda) \overset{ind.}{=} [\prod_{j=1}^n f_{Y_j|\Lambda}(y_j|\lambda)] \cdot \pi_\Lambda(\lambda)$
- the **posterior** pdf of Λ: the conditional pdf of Λ given $\tilde{Y} = \tilde{y}$, $\pi_{\Lambda|\tilde{Y}}(\lambda|\tilde{y}) = f_{\tilde{Y},\Lambda}(\tilde{y}, \lambda)/f_{\tilde{Y}}(\tilde{y}) = [\prod_{j=1}^n f_{Y_j|\Lambda}(y_j|\lambda \cdot \pi_\Lambda(\lambda)]/f_{\tilde{Y}}(\tilde{y})$, where $f_{\tilde{Y}}(\tilde{y}) = \sum_\theta f_{\tilde{Y},\Lambda}(\tilde{y}, \lambda)$.
- the **predictive** pdf: the conditional pdf of Y_{n+1} given $\tilde{Y} = \tilde{y}$ can be derived as $f_{Y_{n+1}|\tilde{Y}}(y_{n+1}|\tilde{y}) = \sum_\lambda f_{Y_{n+1}|\Lambda}(y_{n+1}|\lambda) \cdot \pi_{\Lambda|\tilde{Y}}(\lambda|\tilde{y})$.
- The mean of the predictive distribution (the **Bayesian premium**) is
  $$\begin{aligned} E[Y_{n+1}|\tilde{Y} = \tilde{y}] &= \sum_y y \cdot f_{Y_{n+1}|\tilde{Y}}(y|\tilde{y}) \\ &= \sum_\lambda \mu_{n+1}(\lambda) \cdot \pi_{\Lambda|\tilde{Y}}(\lambda|\tilde{y}) = E[\mu_{n+1}(\Lambda)|\tilde{Y} = \tilde{y}], \end{aligned}$$
  where $\mu_{n+1}(\lambda) = E[Y_{n+1}|\Lambda = \lambda] = \sum_y y \cdot f_{Y_{n+1}|\Lambda}(y|\lambda)$.

Bayesian Premium (an example)

Λ	$f_{Y_j\|\Lambda}(y_j = 0\|\lambda_i)$	$f_{Y_j\|\Lambda}(y_j = 1\|\lambda_i)$	$E[Y_j\|\Lambda = \lambda_i]$	$\pi_\Lambda(\lambda_i)$
λ_1	$1/2$	$1/2$	$1/2$	$1/2$
λ_2	$1/4$	$3/4$	$3/4$	$1/2$

$\tilde{y}$	$f_{\tilde{Y},\Lambda}(\tilde{y}, \lambda_1)$	$f_{\tilde{Y},\Lambda}(\tilde{y}, \lambda_2)$	$f_{\tilde{Y}}(\tilde{y})$	$\pi_{\Lambda\|\tilde{Y}}(\lambda_1\|\tilde{y})$	$\pi_{\Lambda\|\tilde{Y}}(\lambda_2\|\tilde{y})$
$(0,0)$	$4/32$	$1/32$	$5/32$	$4/5$	$1/5$
$(0,1)$	$4/32$	$3/32$	$7/32$	$4/7$	$3/7$
$(1,1)$	$4/32$	$9/32$	$13/32$	$4/13$	$9/13$

$\tilde{y}$	$E[Y_3\|\tilde{Y} = \tilde{y}] = \sum_{i=1}^{2} E[Y_3\|\Lambda = \lambda_i] \cdot \pi_{\Lambda\|\tilde{Y}}(\lambda_i\|\tilde{y})$
$(0,0)$	$1/2 \times 4/5 + 3/4 \times 1/5 = 11/20 = 0.55$
$(0,1)$	$1/2 \times 4/7 + 3/4 \times 3/7 = 17/28 = 0.61$
$(1,1)$	$1/2 \times 4/13 + 3/4 \times 9/13 = 35/52 = 0.67$

Bayesian Mortality Model

- $Y_t|(\Lambda = \lambda_z) \overset{iid}{\sim} N(\theta_z, \sigma_z^2), t = 1, \cdots$ with probability density at y_t,

$$f_{Y_t|\Lambda}(y_t|\lambda_z) = \frac{1}{\sqrt{2\pi} \cdot \sigma_z} \exp\left[-\frac{1}{2}\left(\frac{y_t - \theta_z}{\sigma_z}\right)^2\right] = \phi(y_t;\, \theta_z,\, \sigma_z^2).$$

- Risk parameter λ_z for age z corresponds to $(\theta_z,\, \sigma_z^2)$ which generates $\tilde{y}_z = (y_{z,\, t_L+1}, \cdots, z_{x,\, t_U})$, where $y_{z,\, t} \overset{\triangle}{=} \ln(m_{z,\, t_L+t}) - \ln(m_{z,\, t_L+t-1})$.
- $\hat{\theta}_z = \frac{\sum_{t_L+1}^{t_U} y_{z,\, t}}{t_U - t_L} = \frac{\sum_{t_L+1}^{t_U} [\ln(m_{z,\, t}) - \ln(m_{z,\, t-1})]}{t_U - t_L} = \frac{\ln(m_{z,\, t_U}) - \ln(m_{z,\, t_L})}{t_U - t_L} = \hat{\theta}_z^{RW}$.
- $\hat{\sigma}_z^2 = \frac{1}{t_U - t_L - 1} \sum_{t_L+1}^{t_U} [y_{z,\, t} - \hat{\theta}_z]^2 = (\hat{\sigma}_z^{RW})^2$.
- $\hat{\mu}_x^{BS} \overset{\triangle}{=} \hat{E}[Y_{n+1}|\tilde{Y} = \tilde{y}_x] = \sum_{z=x_L}^{x_U} \hat{\theta}_z\, \pi_{\Lambda|\tilde{Y}}(\lambda_z|\tilde{y}_x) = \sum_{z=x_L}^{x_U} \hat{\theta}_z^{RW} \frac{f_{\tilde{Y},\Lambda}(\tilde{y}_x,\, \lambda_z)}{f_{\tilde{Y}}(\tilde{y}_x)}$

$$= \frac{\sum_{z=x_L}^{x_U} \hat{\theta}_z^{RW} \cdot \pi_\Lambda(\lambda_z) \cdot \prod_{t=t_L+1}^{t_U} \phi(y_{x,\, t};\, \hat{\theta}_z,\, \hat{\sigma}_z^2)}{\sum_{w=x_L}^{x_U} \pi_\Lambda(\lambda_w) \cdot \prod_{t=t_L+1}^{t_U} \phi(y_{x,\, t};\, \hat{\theta}_w,\, \hat{\sigma}_w^2)}$$

  Given observations $\tilde{y}_x = (y_{x,\, t_L+1}, \cdots, y_{x,\, t_U})$, $\hat{\mu}_x^{BS}$ for age x is a weighted average of $\hat{\mu}_z^{RW}$s with the posterior density $\pi_{\Lambda|\tilde{Y}}(\lambda_z|\tilde{y}_x)$ as the weight.
- Assume $\pi_\Lambda(\lambda_z) = 1/m$, $z = x_L, \cdots, x_U$, where $m = x_U - x_L + 1$.

$\pi_{\Lambda|\tilde{Y}}(\lambda_z|\tilde{y}_x)$ vs z with fitting window $[25, 84] \times [1951, 2006]$

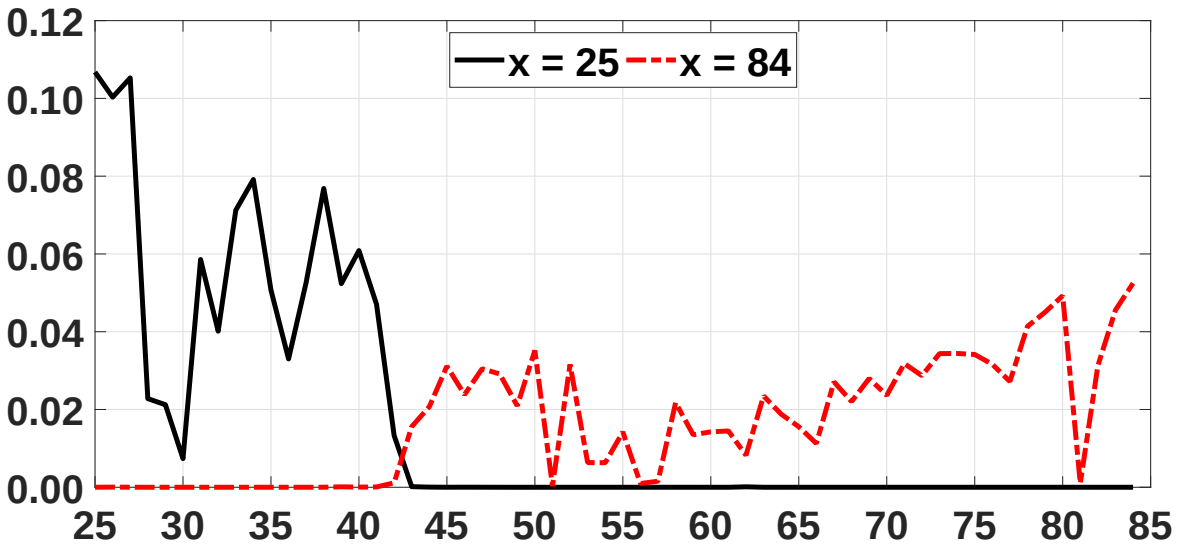
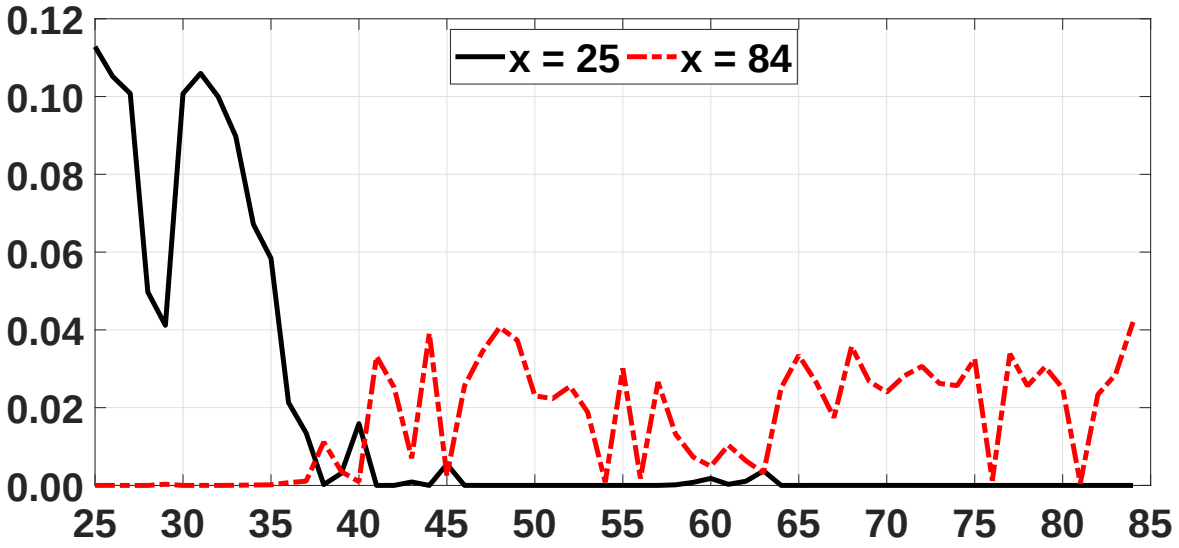


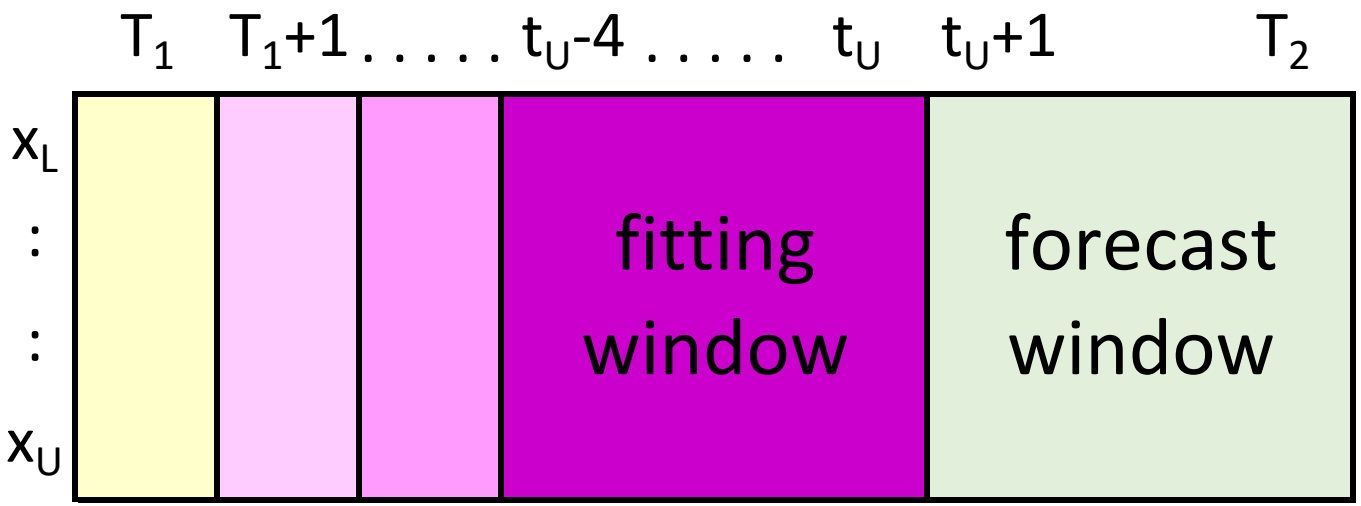
Figure: top: US males; bottom: US female



Models for comparisons

- LC-f model: $\ln(\hat{m}^{LC-f}_{x,\,t_U+\tau}) = \ln(\hat{m}^{LC}_{x,\,t_U}) + \hat{\mu}^{LC}_x \times \tau,\ \tau = 1, 2, \cdots,$ with intercept $\ln(\hat{m}^{LC}_{x,\,t_U})$ and slope $\hat{\mu}^{LC}_x = \hat{\beta}_x \cdot \hat{\theta}^{LC}$
- RW model: $\ln(\hat{m}^{RW}_{x,\,t_U+\tau}) = \ln(m_{x,\,t_U}) + \hat{\mu}^{RW}_x \times \tau,\ \tau = 1, 2, \cdots,$ with intercept $\ln(m_{x,\,t_U})$ and slope $\hat{\mu}^{RW}_x$.
- Bayesian model: $\ln(\hat{m}^{BS}_{x,\,t_U+\tau}) = \ln(m_{x,\,t_U}) + \hat{\mu}^{BS}_x \times \tau,\ \tau = 1, 2, \cdots,$ with intercept $\ln(m_{x,\,t_U})$ and slope $\hat{\mu}^{BS}_x = E[Y_{n+1}|\tilde{Y} = \tilde{y}_x]$.
- LC-r model: $\ln(\hat{m}^{LC-r}_{x,\,t_U+\tau}) = \ln(m_{x,\,t_U}) + \hat{\mu}^{LC}_x \times \tau,\ \tau = 1, 2, \cdots,$ with intercept $\ln(m_{x,\,t_U})$ and slope $\hat{\mu}^{LC}_x = \hat{\beta}_x \cdot \hat{\theta}^{LC}$
- Total intercepts over all ages for all models are preserved: $\sum_{x=x_L}^{x_U} \ln(\hat{m}^{LC}_{x,\,t_U}) = \sum_{x=x_L}^{x_U} \ln(m_{x,\,t_U})$.
- Total slopes over all ages for LC-f, LC-r and RW models are preserved: $\sum_{x=x_L}^{x_U} \hat{\mu}^{RW}_x = \sum_{x=x_L}^{x_U} \hat{\mu}^{LC} = \hat{\theta}^{LC}$.

Mortality Data



- study window: $[x_L,\ x_U] \times [T_1,\ T_2]$, where mortality data are available from the Human Mortality Database. We select $x_L = 25$, $x_U = 84$, $T_1 = 1951$, $T_2 = 2016$ for the study window $[25,\ 64] \times [1951,\ 2016]$.
- 3 forecasting windows: $[x_L,\ x_U] \times [t_U + 1,\ T_2]$, where $t_U = 2006, 1996, 1986$ for $[25,\ 84] \times [2007,\ 2016]$, $[25,\ 84] \times [1997,\ 2016]$ and $[25,\ 84] \times [1987,\ 2016]$.
- a series of fitting windows: $[x_L,\ x_U] \times [T_1,\ t_U]$, $[x_L,\ x_U] \times [T_1 + 1,\ t_U], \cdots, [x_L,\ x_U] \times [t_U - 4,\ t_U]$;
- study countries: both genders of the US, the UK, and Japan.

Forecasting Error Measures

- Data conversion using $q_{x,\,t} = 1 - e^{-m_{x,\,t}}$ under the assumption that $\mu_{x+r,\,t+s}$ is constant for $r, s \in [0,\,1)$.
- MAPE (Mean Absolute Percentage Error) over a forecasting window $[x_L,\,x_U] \times [t_U + 1,\,T_2]$ based on the fitting window $[x_L,\,x_U] \times [t_L,\,t_U]$:

$$MAPE^{[x_L,\,x_U]\times[t_L,\,t_U]}_{[x_L,\,x_U]\times[t_U+1,\,T_2]} = \frac{1}{T_2 - t_U} \cdot \frac{1}{m} \sum_{\tau=1}^{T_2-t_U} \sum_{x=x_L}^{x_U} \left| \frac{\hat{q}_{x,\,t_U+\tau} - q_{x,\,t_U+\tau}}{q_{x,\,t_U+\tau}} \right|.$$

- AMAPE (Average of $MAPE^{[x_L,\,x_U]\times[t_L,\,t_U]}_{[x_L,\,x_U]\times[t_U+1,\,T_2]}$ over all fitting windows $[x_L,\,x_U] \times [T_1,\,t_U], [x_L,\,x_U] \times [T_1 + 1,\,t_U], \cdots [x_L,\,x_U] \times [t_U - 4,\,t_U]$:

$$AMAPE_{[x_L,\,x_U]\times[t_U+1,\,T_2]} = \frac{1}{t_U - 4 - T_1 + 1} \sum_{t_L=T_1}^{t_U-4} MAPE^{[x_L,\,x_U]\times[t_L,\,t_U]}_{[x_L,\,x_U]\times[t_U+1,\,T_2]}.$$

- Comparison of the overall forecasting performances among mortality models is based on the $AMAPE_{[x_L,\,x_U]\times[t_U+1,\,T_2]}$ measure.

Average of $MAPE^{[25,84]\times[t_L,2006]}_{[25,84]\times[2007,2016]}$ over all 6 populations against $t_L = 1951, \cdots, 2002$

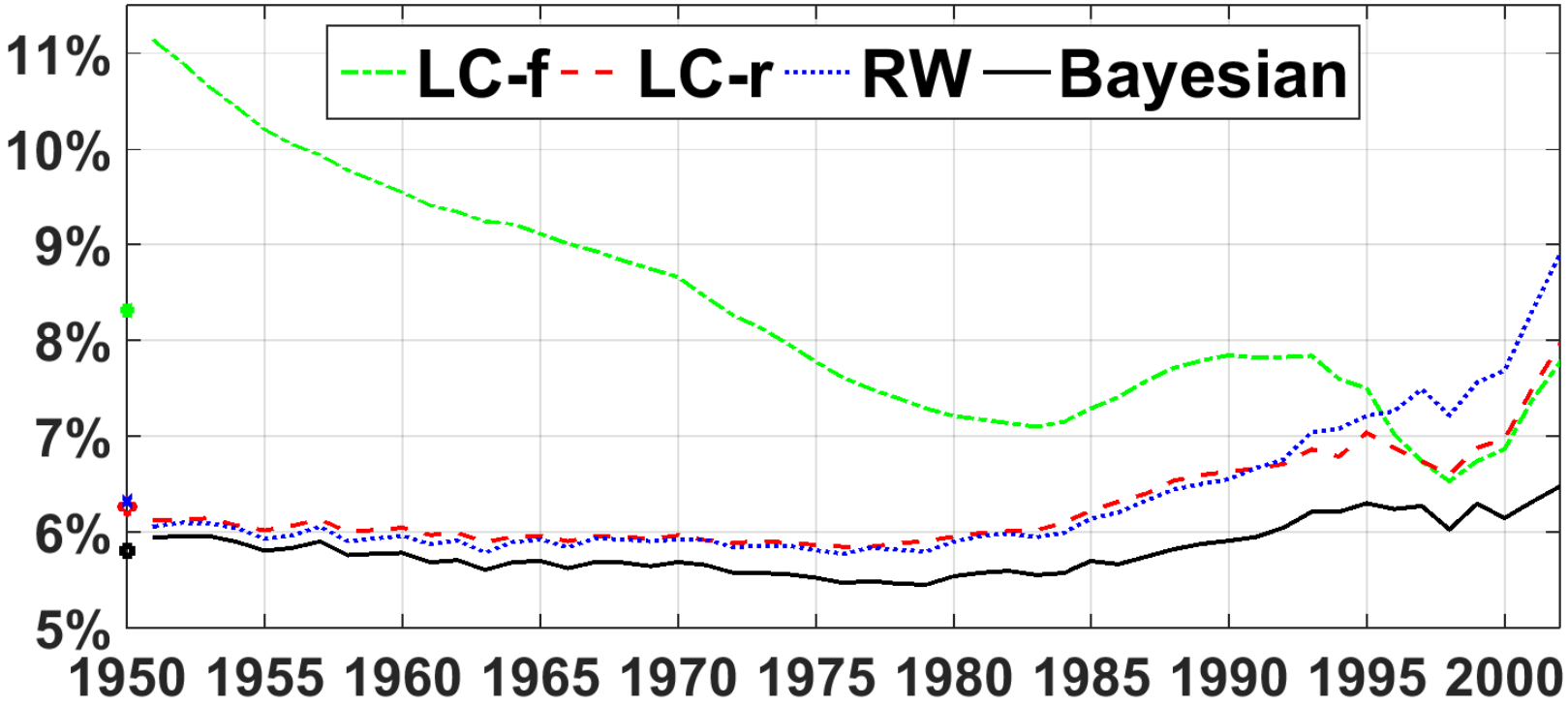


Figure: 10-year forecast for $[2007, 2016]$

Average of $MAPE^{[25,84]\times[t_L,1996]}_{[25,84]\times[1997,2016]}$ over all 6 populations against $t_L = 1951, \cdots, 1992$

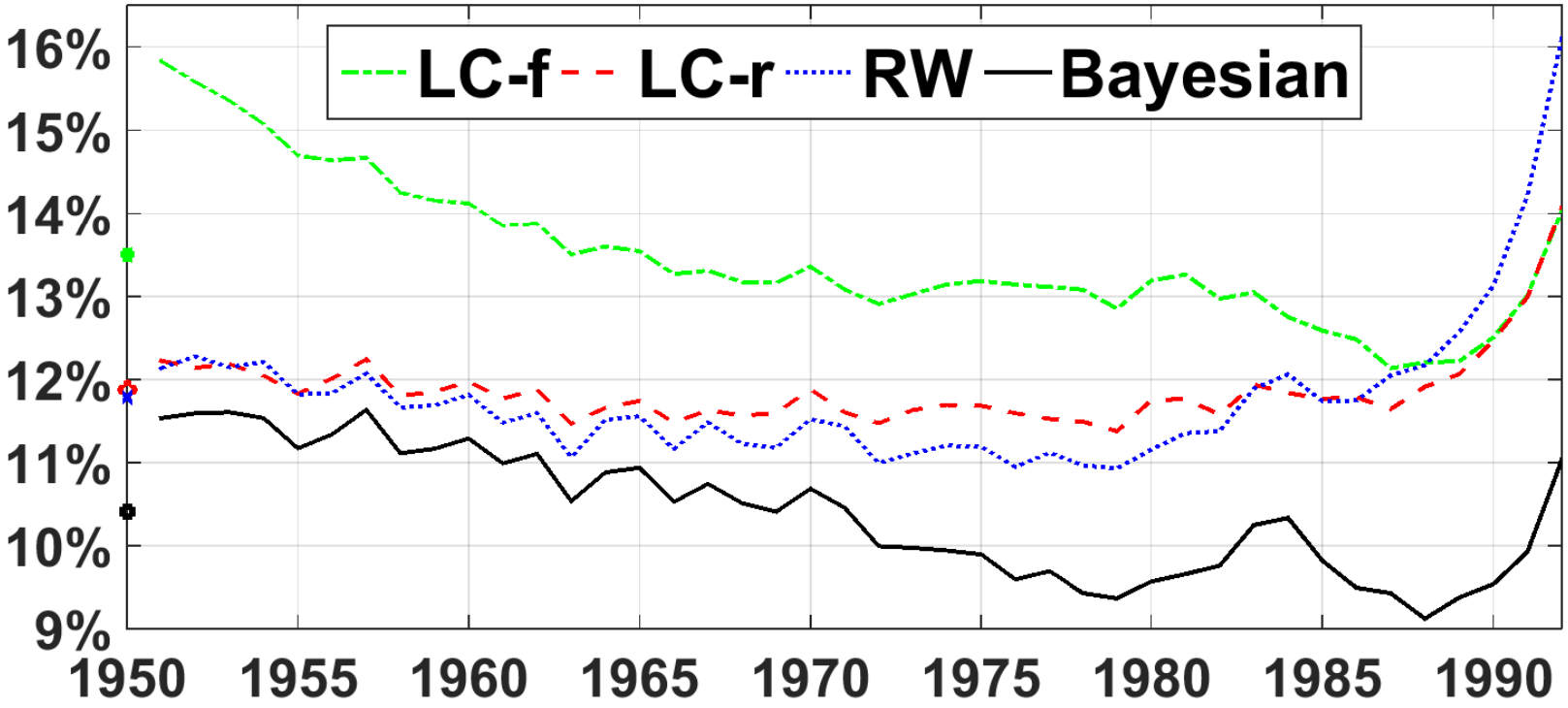


Figure: 20-year forecast for $[1997, 2016]$

Average of $MAPE_{[25,84]\times[1987,2016]}^{[25,84]\times[t_L,1986]}$ over all 6 populations against $t_L = 1951, \cdots, 1982$

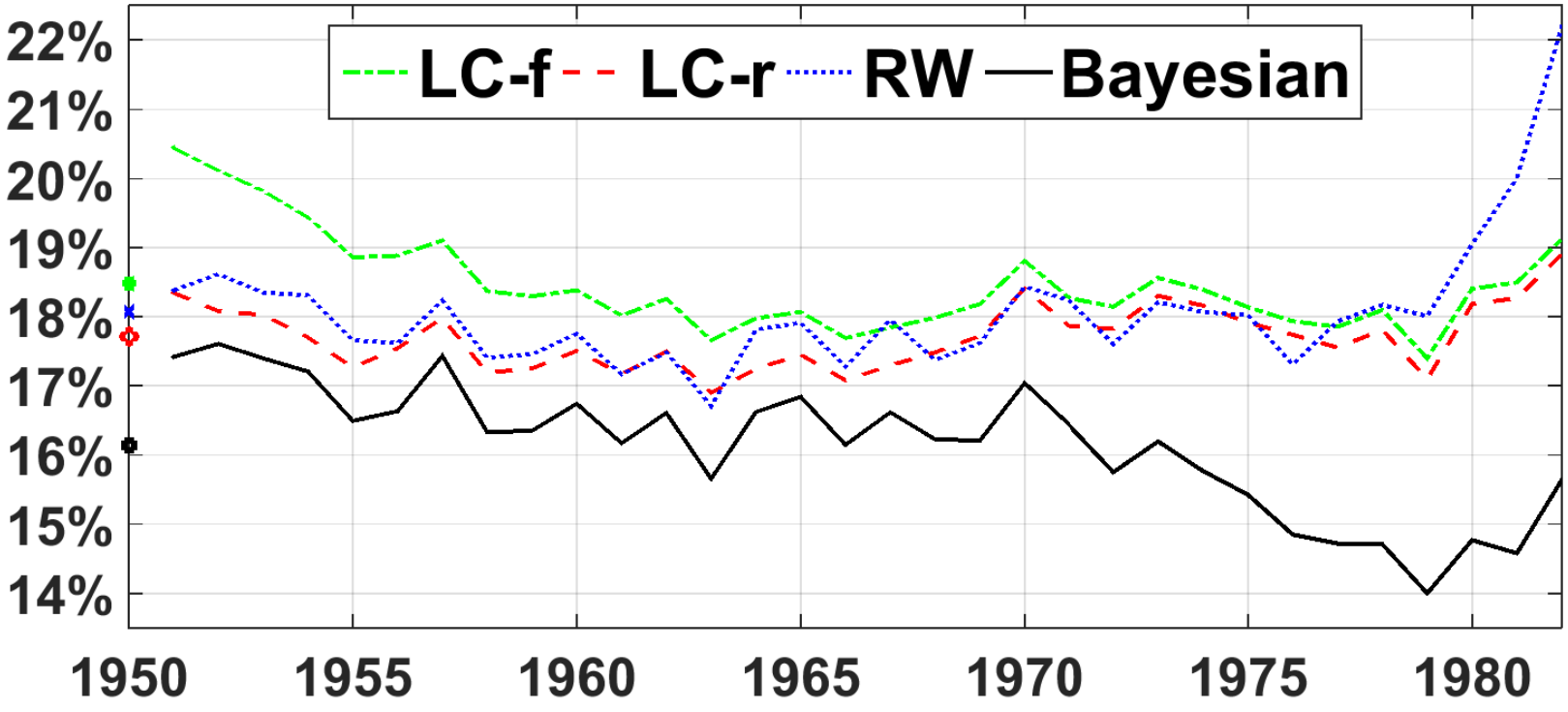


Figure: 30-year forecast for $[1987, 2016]$

SLOPE μ_x VERSUS x FOR US MALES

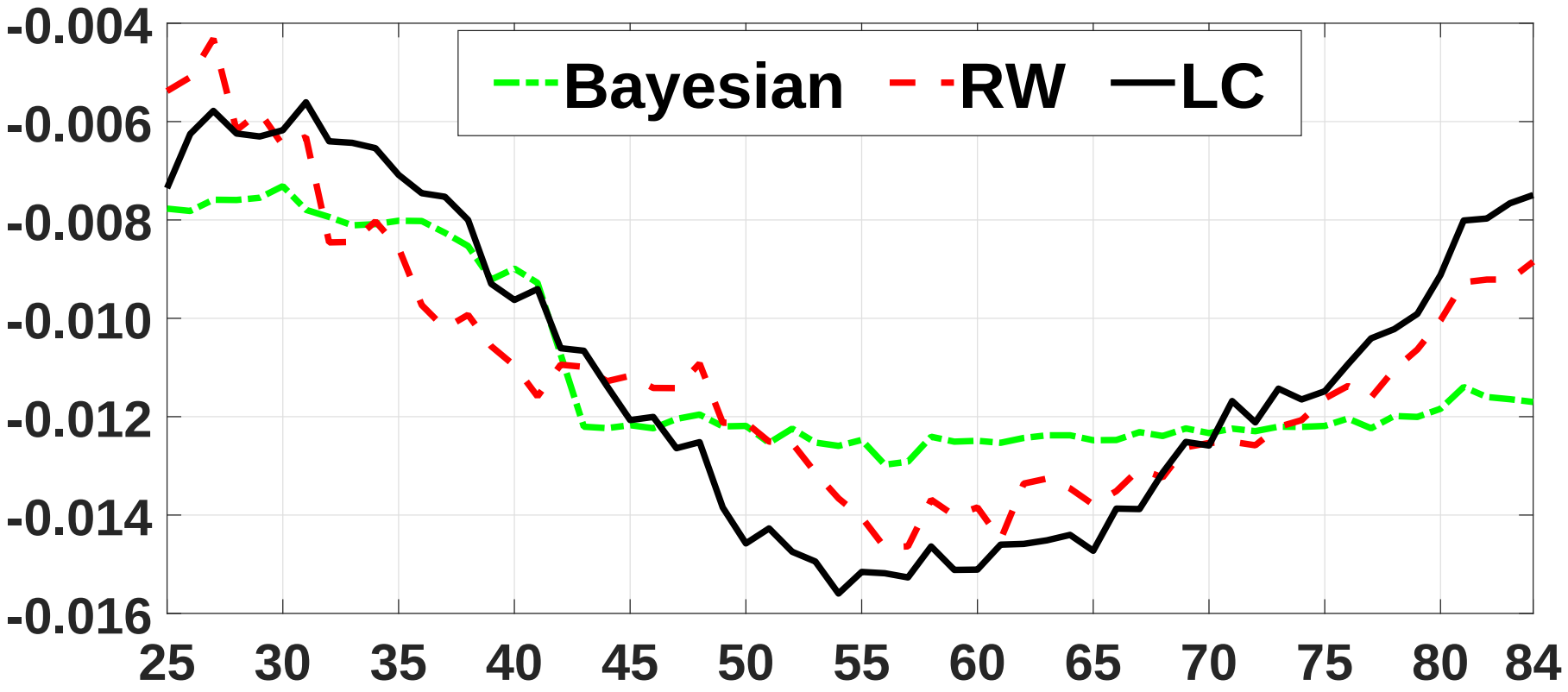


FIGURE: $MAPE^{[25,84]\times[1951,2006]}_{[25,84]\times[2007,2016]} = [4.50\%^{BS}, 4.81\%^{RW}, 8.60\%^{LC}]$

SLOPE μ_x VERSUS x FOR US MALES

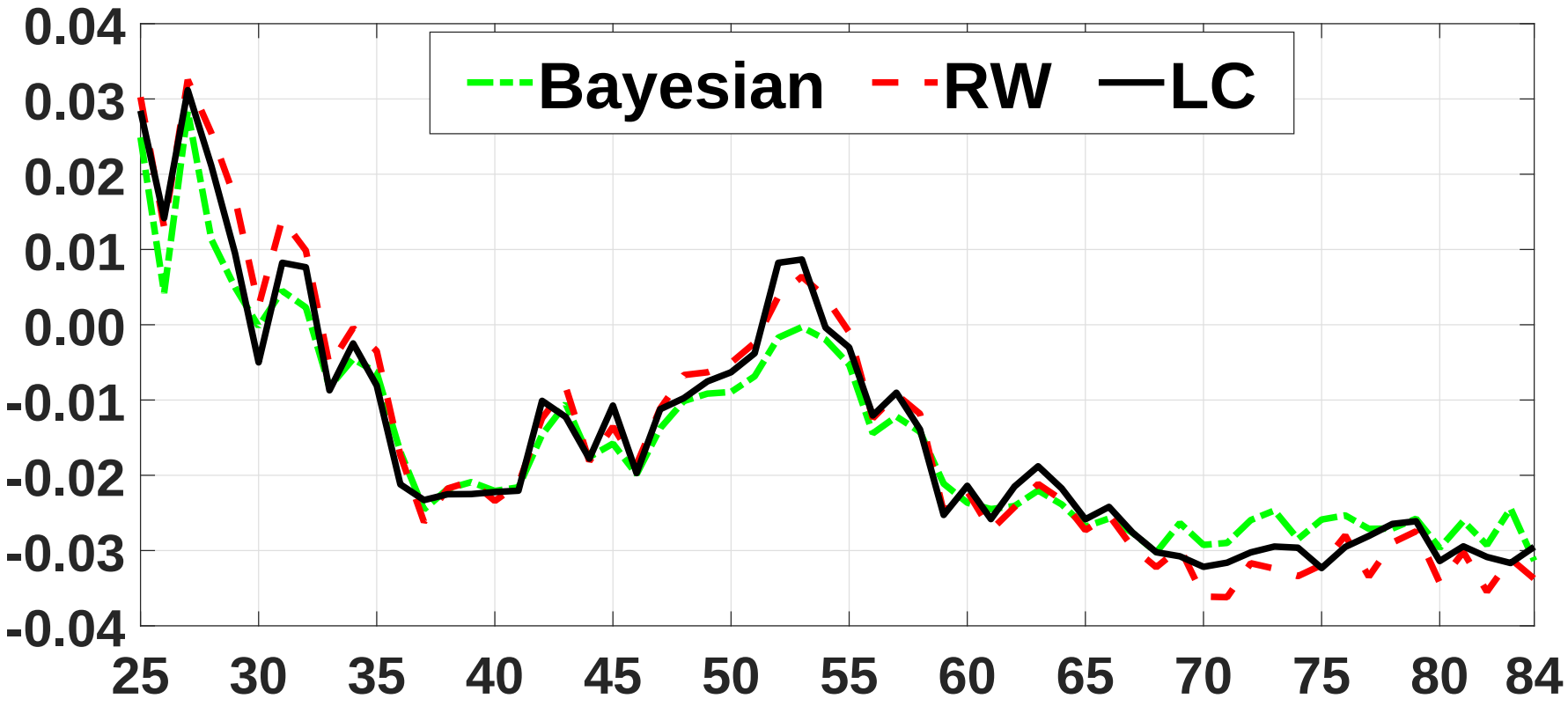


FIGURE: $MAPE^{[25,\,84]\times[2002,\,2006]}_{[25,\,84]\times[2007,\,2016]} = [6.96\%^{BS},\ 8.40\%^{RW},\ 7.60\%^{LC}]$

Forecasting Error Measures

$$AMAPE_{[25,\,84]\times[t_U+1,\,2016]}$$

Model	Avg-6	US-M	US-F	UK-M	UK-F	Japan-M	Japan-F
Panel A:AMAPE (%) for [25, 84]x[2007, 2016]							
LC-f	8.31%	8.55%	8.28%	9.63%	7.82%	7.27%	8.31%
LC-r	6.26%	5.79%	5.20%	7.05%	6.18%	5.91%	7.43%
RW	6.32%	5.69%	4.91%	7.23%	6.41%	6.00%	7.68%
BS	**5.81%**	**5.24%**	**4.58%**	**6.75%**	**5.82%**	**5.43%**	**7.03%**
Panel B: AMAPE (%) for [25, 84]x[1997, 2016]							
LC-f	13.50%	14.91%	9.18%	17.48%	12.69%	11.44%	15.28%
LC-r	11.87%	12.83%	8.34%	15.52%	10.61%	10.44%	13.49%
RW	11.78%	12.42%	8.33%	15.28%	10.78%	10.34%	13.54%
BS	**10.40%**	**10.92%**	**7.04%**	**13.57%**	**8.97%**	**9.07%**	**12.84%**
Panel C: AMAPE (%) for [25, 84]x[1987, 2016]							
LC-f	18.47%	14.66%	13.42%	23.71%	17.32%	18.58%	23.12%
LC-r	17.71%	13.95%	13.34%	22.44%	16.64%	17.94%	21.95%
RW	18.07%	15.71%	13.38%	22.50%	16.92%	17.86%	22.06%
BS	**16.14%**	**12.88%**	**11.64%**	**20.90%**	**13.56%**	**16.63%**	**21.24%**

Conclusions

- Change the starting location of forecast for the Lee-Carter model from the fitting value (LC-f) to the observed value (LC-r) can generally reduce the forecast error.
- The LC-r and RW models have similar forecasting performances based on the measure of Avg-6.
- The slope $\hat{\theta}_x^{RW}$ for age x under the RW model uses mortality data for a sole age x, and is not affected by $\hat{\theta}_z^{RW}$ for all $z \neq x$.
- Given observations $\tilde{Y} = \tilde{y}_x$, the slope $\hat{\theta}_x^{BS}$ for age x under the Bayesian model is a weighted average of $\hat{\mu}_z^{RW}$s (all slopes/drifts under the RW model) with the posterior density $\pi_{\Theta|\tilde{Y}}(\theta_z|\tilde{y}_x)$ as the weight:

  $\hat{\mu}_x^{BS} \triangleq E[Y_{n+1}|\tilde{Y} = \tilde{y}_x] = \sum_{z=x_L}^{x_U} \hat{\mu}_z^{RW} \cdot \pi_{\Theta|\tilde{Y}}(\theta_z|\tilde{y}_x)$
- Among the Bayesian, RW and LC-r models with the same starting location (intercept) for mortality forecast, the Bayesian model can better capture the downward mortality trend with more accurate slopes than the LC-r and RW models.

The End

Thank You