Homework 03

Spencer Pease

4/20/2020

Questions

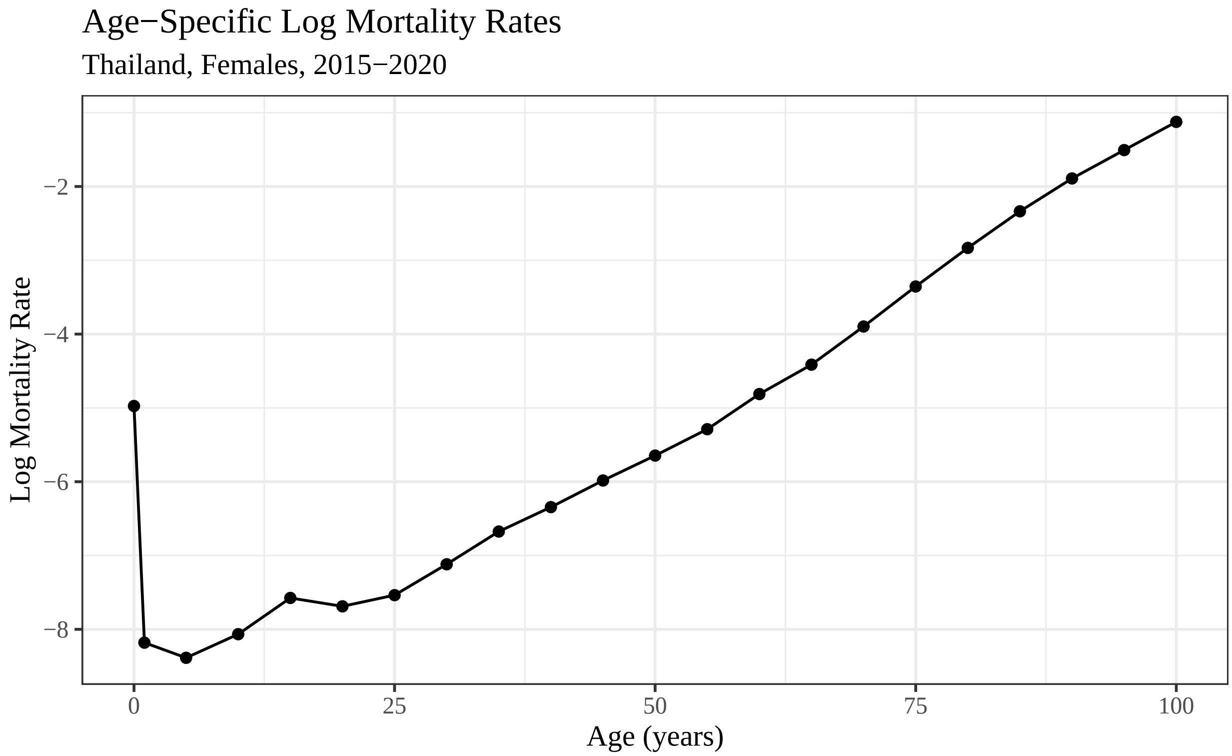
Q1

Q1.a

Using the *WPP2019* package, we can extract the age-specific mortality rates for females in Thailand in the period 2015-2020. This data is presented in a table and a log-transformed graph below.

Table 1: Thailand 2015-2020 female mortality rates

Age	n	${}_nM_x$
0	1	0.007
1	4	0.000
5	5	0.000
10	5	0.000
15	5	0.001
20	5	0.000
25	5	0.001
30	5	0.001
35	5	0.001
40	5	0.002
45	5	0.003
50	5	0.004
55	5	0.005
60	5	0.008
65	5	0.012
70	5	0.020
75	5	0.035
80	5	0.059
85	5	0.097
90	5	0.151
95	5	0.222
100	Inf	0.325



Q1.b

Two ways of modeling mortality rates are with the *Gompertz* model and *Gompertz-Makeham* model. These models estimate force of mortality $\mu(x)$ using a simple exponential, with the *Gompertz-Makeham* model including an additional constant term. A log-linear fit applied to each of the models, defined below, along with the estimated model parameters after fitting our observed data:

Gompertz

$(\alpha, \beta) = (0.003,\ 0.094)$

$$\mu(x) = \alpha e^{\beta x}$$
$$log[\mu(x)] = log(\alpha) + \beta x$$

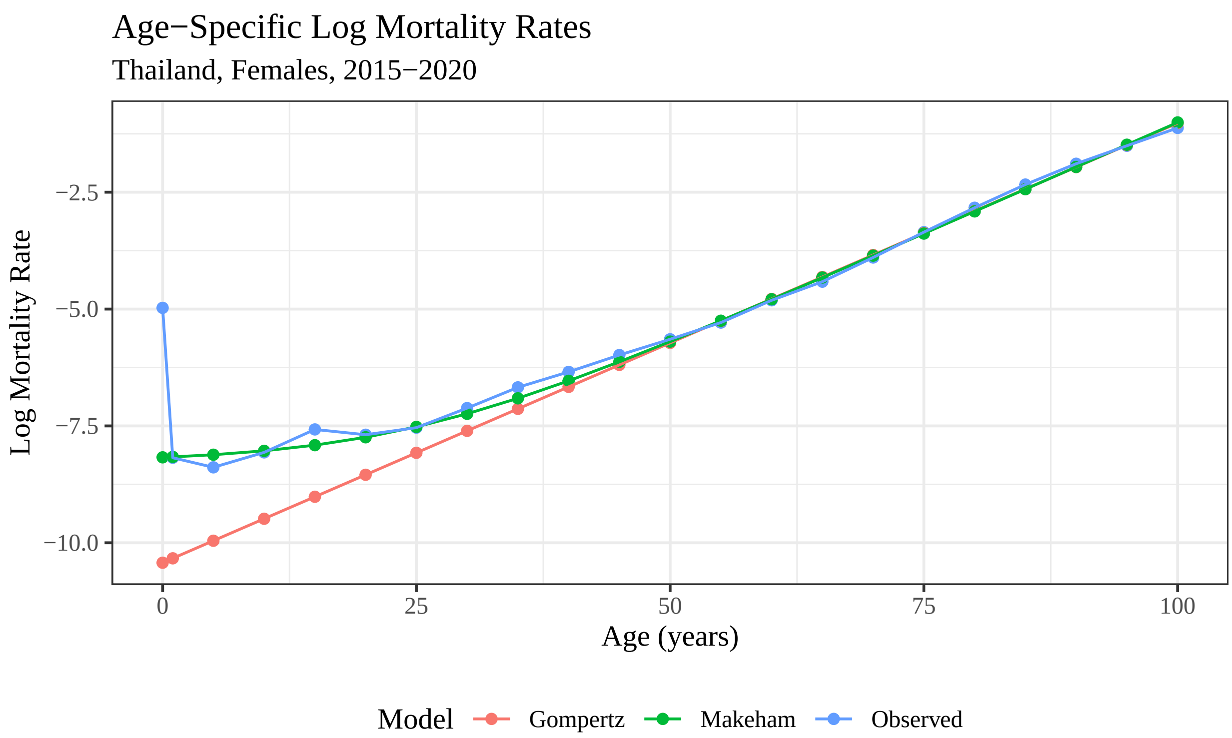
Gompertz-Makeham

$(\alpha, \beta, \gamma) = (0.0028,\ 0.0953,\ 3 \times 10^{-4})$

$$\mu(x) = \gamma + \alpha e^{\beta x}$$
$$log[\mu(x) - \gamma] = log(\alpha) + \beta x$$

Note: both of these models were fit using the MortalityLaws package, optimizing the function $log^2(\frac{est.}{obs.})$.

Q1.c



Plotting both fitted rates against the observed rates, we see that the additional constant in the *Gompertz-Makeham* enables the model to follow the observed trends in child and young adult mortality much closer. This presents a good case for using *Gompertz-Makeham* over just *Gompertz*.

Q1.d

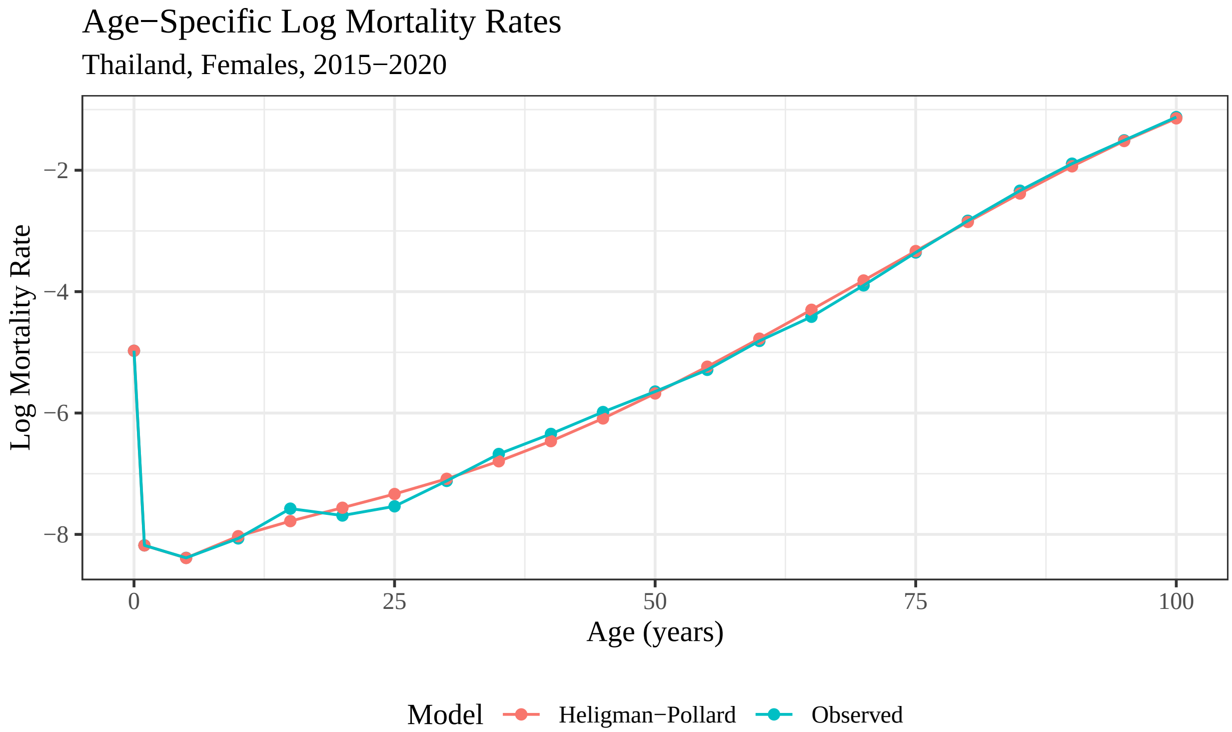
The *Heligman-Pollard* model uses three terms to capture child mortality, the adult accident bump, and old-age mortality to predict odds of death:

$$\frac{q_x}{1-q_x} = A^{(x+B)^2} + De^{-E(log(x)-log(F))^2} + GH^x$$

Fitting this model with our observed mortality rates, we get the model parameters:

(A, B, C, D, E, F, G, H) = 1, 8.46, 4.99, 2.53e+14, 0.00141, 2.48e+75, 1.72e-05, 1.11

Note: this model was fit using the MortalityLaws package, optimizing the function $log^2(\frac{est.}{obs.})$*, and using* M_x *as an input (internal* M_x *to* q_x *conversion).*



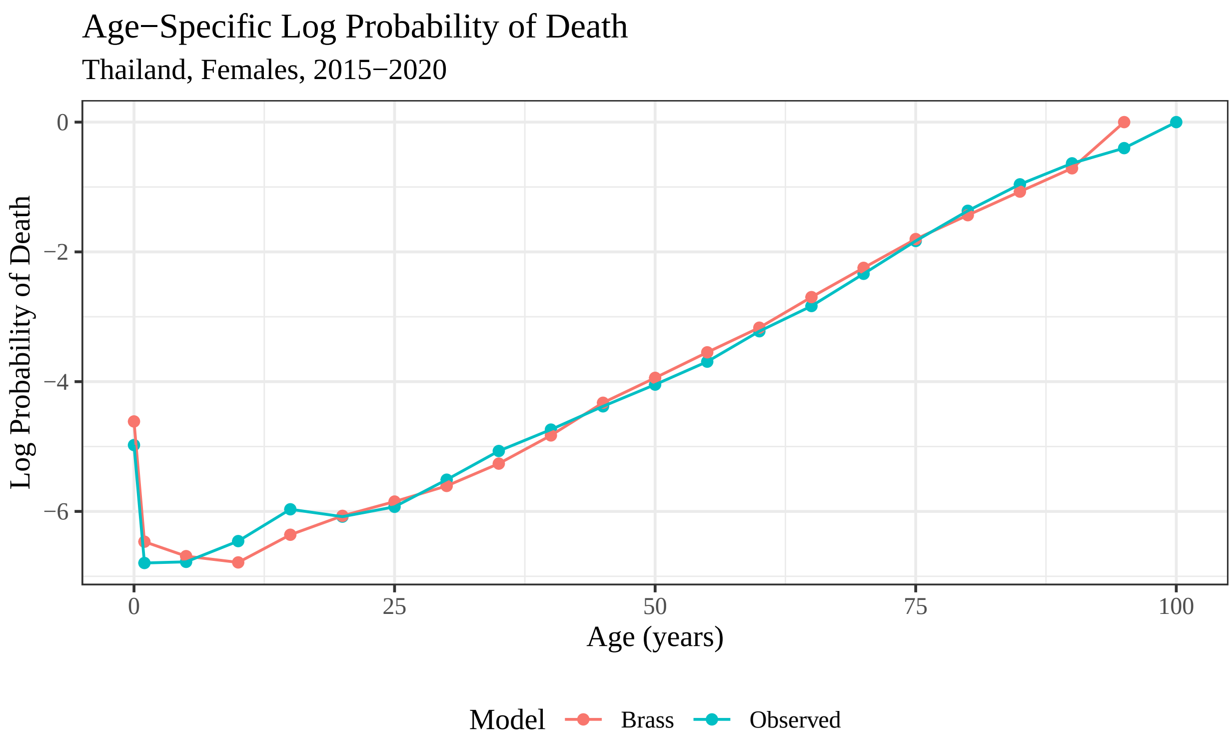
This model does a great job of capturing the high child mortality and gets closer to capturing the adult accident hump, following the observed data closely.

Q1.e

The *Brass relational model* estimates mortality given a "standard" reference mortality ($q^*(x)$) as an input. This standard mortality comes from the a *Coale-Demeny West* model life table, where the life table that was selected was the one with the lowest RMSE compared to our observed mortality (life table index #25). The model and estimated parameters are shown below:

$(\alpha, \beta) =$ -0.761, 0.71

$$\text{logit}(q_{\alpha,\beta}(x)) = \alpha + \beta \times \text{logit}(q^*(x))$$



Plotting the observed and estimated q_x values, we see that the *Brass* model also does a good job of capturing high child mortality, but underestimates the adult accident hump. In the old ages this model loses validity, since it was fix with a lower terminal age than is present in the observed data.

Q1.f

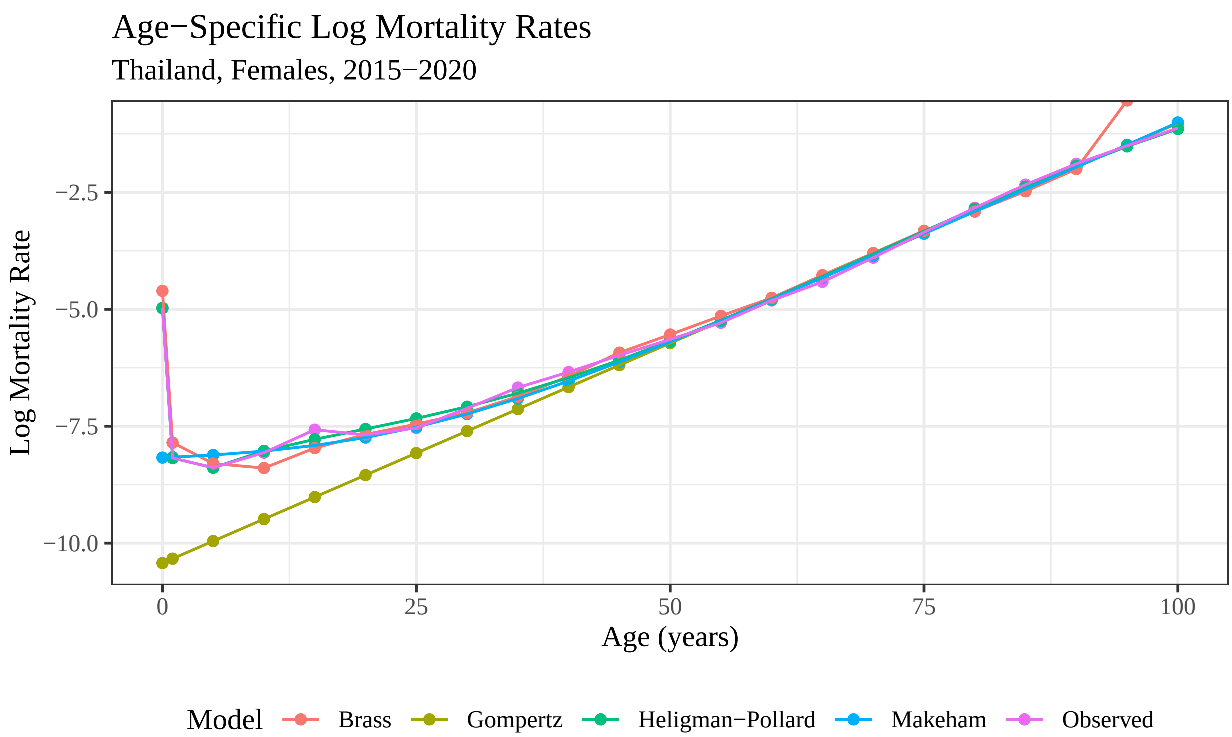


Table 2: $RMSE$ of estimated $log(M_x)$ against observed $log(M_x)$

Model	RMSE
Heligman-Pollard	5.989
Brass	5.992
Makeham	6.203
Gompertz	6.935

Comparing all four of these models to the observed data, both visually and via the root mean-squared-error of log-transformed M_x, shows that the *Heligman-Pollard* model most closely follows the observed data.

Q2

Q2.a

Using the *WPP2019 package*, we can extract the female age-specific mortality rates in Thailand in the for each five year period encompassing 1950-2020.

Table 3: Thailand female age-specific mortality rates, 1950-2020

age	1950	1955	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005	2010	2015
0	0.125	0.104	0.088	0.073	0.057	0.043	0.037	0.030	0.023	0.019	0.015	0.012	0.009	0.007
1	0.018	0.014	0.011	0.009	0.006	0.005	0.003	0.002	0.001	0.001	0.001	0.000	0.000	0.000
5	0.005	0.005	0.004	0.003	0.003	0.002	0.002	0.001	0.001	0.001	0.001	0.000	0.000	0.000

age	1950	1955	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005	2010	2015
10	0.003	0.003	0.002	0.002	0.002	0.001	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000
15	0.003	0.003	0.002	0.002	0.002	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001
20	0.005	0.004	0.004	0.003	0.003	0.002	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.000
25	0.006	0.005	0.004	0.004	0.003	0.002	0.002	0.001	0.001	0.002	0.002	0.001	0.001	0.001
30	0.007	0.006	0.005	0.005	0.004	0.003	0.002	0.001	0.001	0.002	0.002	0.002	0.001	0.001
35	0.008	0.008	0.007	0.006	0.005	0.004	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.001
40	0.009	0.009	0.008	0.007	0.006	0.005	0.004	0.003	0.003	0.003	0.003	0.002	0.002	0.002
45	0.010	0.009	0.009	0.009	0.008	0.007	0.006	0.004	0.004	0.004	0.004	0.003	0.003	0.003
50	0.012	0.012	0.011	0.011	0.011	0.010	0.009	0.007	0.006	0.006	0.006	0.005	0.004	0.004
55	0.015	0.014	0.015	0.014	0.014	0.013	0.012	0.009	0.009	0.008	0.008	0.008	0.006	0.005
60	0.022	0.022	0.020	0.021	0.019	0.019	0.017	0.014	0.013	0.012	0.011	0.010	0.009	0.008
65	0.029	0.030	0.029	0.029	0.029	0.026	0.025	0.021	0.021	0.019	0.018	0.017	0.014	0.012
70	0.045	0.056	0.048	0.048	0.046	0.044	0.037	0.034	0.033	0.030	0.028	0.028	0.023	0.020
75	0.064	0.070	0.067	0.069	0.068	0.062	0.056	0.050	0.052	0.048	0.048	0.044	0.039	0.035
80	0.101	0.113	0.121	0.118	0.116	0.109	0.100	0.094	0.101	0.086	0.080	0.079	0.066	0.059
85	0.133	0.147	0.155	0.155	0.156	0.143	0.149	0.128	0.140	0.129	0.127	0.121	0.107	0.097
90	0.181	0.200	0.216	0.214	0.218	0.198	0.227	0.184	0.204	0.193	0.195	0.188	0.166	0.151
95	0.242	0.266	0.292	0.287	0.295	0.267	0.329	0.257	0.287	0.278	0.287	0.279	0.244	0.222
100	0.330	0.356	0.392	0.385	0.398	0.362	0.459	0.359	0.399	0.394	0.412	0.404	0.355	0.325

Q2.b

The *Lee-Carter* model is defined as:

$$log(m_{x,t}) = a_x + k_t b_x + \epsilon_{x,t}$$

One way to fit this model is the apply a set of constraints to b_x and k_t. Under the constraints $\sum b_x = 1$ and $\sum k_t = 0$, we get:

$$\hat{a}_x = \frac{1}{T}\sum_{t=1}^{T} log(m_{x,t})\hat{k}_t = \sum_{x=0}^{A-1}\left[log(m_{x,t}) - a_x\right]$$

With these constraints, we can fit a least squares regression on $\hat{b}_x$ with an intercept of 0. Below are the estimated parameters and mortality rates

Note: This model was fit using the MortCast package.

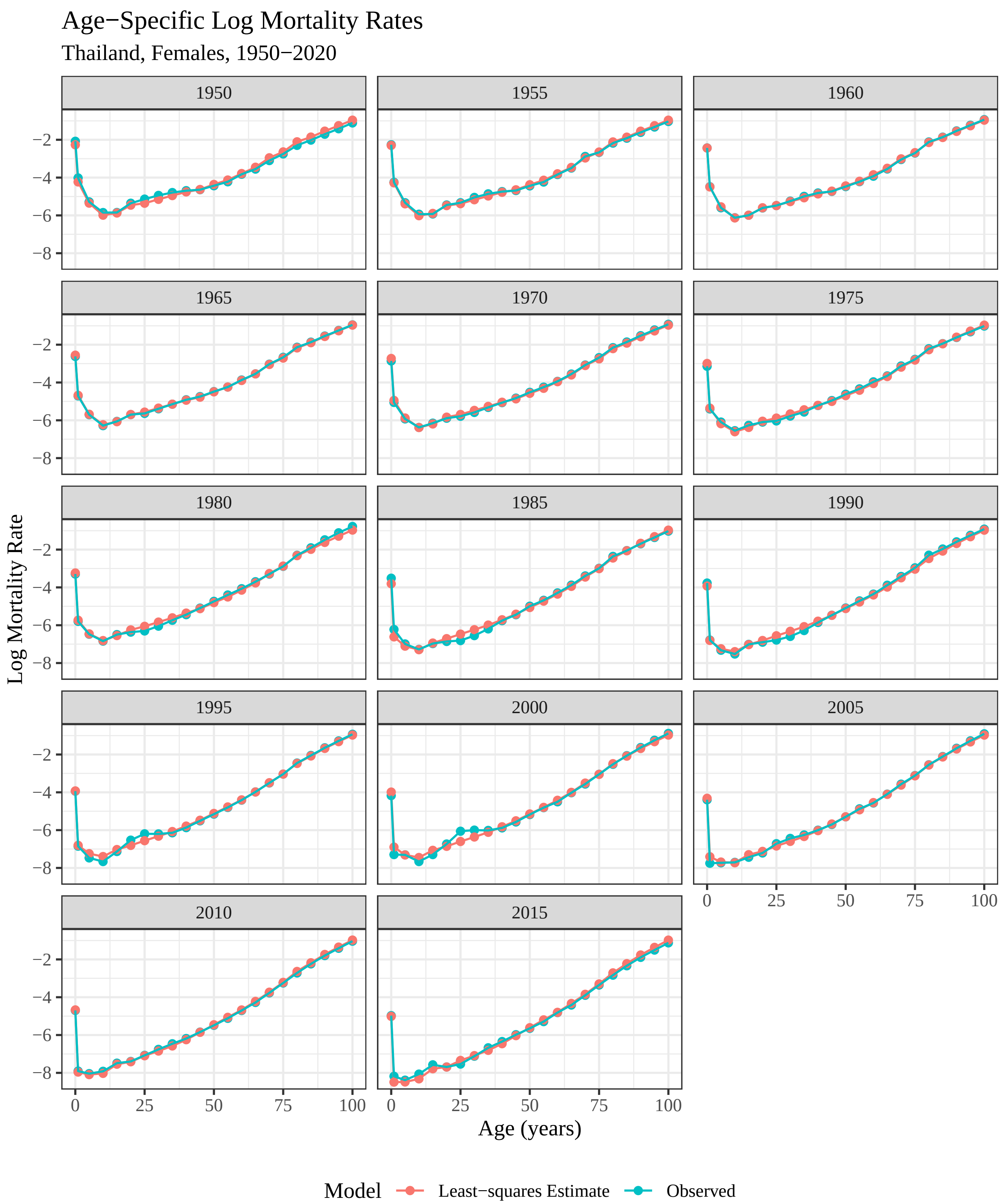
Table 4: Lee-Carter model parameter estimates (least squares method)

Age	$\hat{a}_x$	$\hat{b}_x$
0	-3.440	0.084
1	-6.050	0.131
5	-6.690	0.096
10	-6.984	0.071
15	-6.689	0.059
20	-6.412	0.069
25	-6.206	0.061
30	-5.978	0.059
35	-5.746	0.057

Age	$\hat{a}_x$	$\hat{b}_x$
40	-5.485	0.052
45	-5.229	0.043
50	-4.897	0.038
55	-4.585	0.033
60	-4.219	0.031
65	-3.828	0.027
70	-3.334	0.027
75	-2.921	0.020
80	-2.365	0.019
85	-2.013	0.011
90	-1.638	0.007
95	-1.300	0.003
100	-0.969	0.000

Table 5: Lee-Carter model parameter estimates (least squares method)

Year	$\hat{k}_t$
1950	13.915
1955	13.542
1960	11.896
1965	10.492
1970	8.435
1975	5.301
1980	2.414
1985	-4.338
1990	-5.774
1995	-5.770
2000	-6.479
2005	-10.383
2010	-14.595
2015	-18.656



Q2.c

A singular value decomposition can also be used to implement the *Lee-Carter* model, where the matrix $C = \left[log(m_{x,t}) - a_x\right]$ is used to get b_x and k_t with the decomposition:

$$\text{SVD}(C) = U\Lambda V^{\top}$$

Where:

$$b_x = U_{x,1} \div \sum_x U_{x,1} k_t = [V^{\top}]_{1,t} \times \sum U_{x,1} \times \Lambda_1$$

are the normalized estimates. The estimated parameters and fitted values are below.

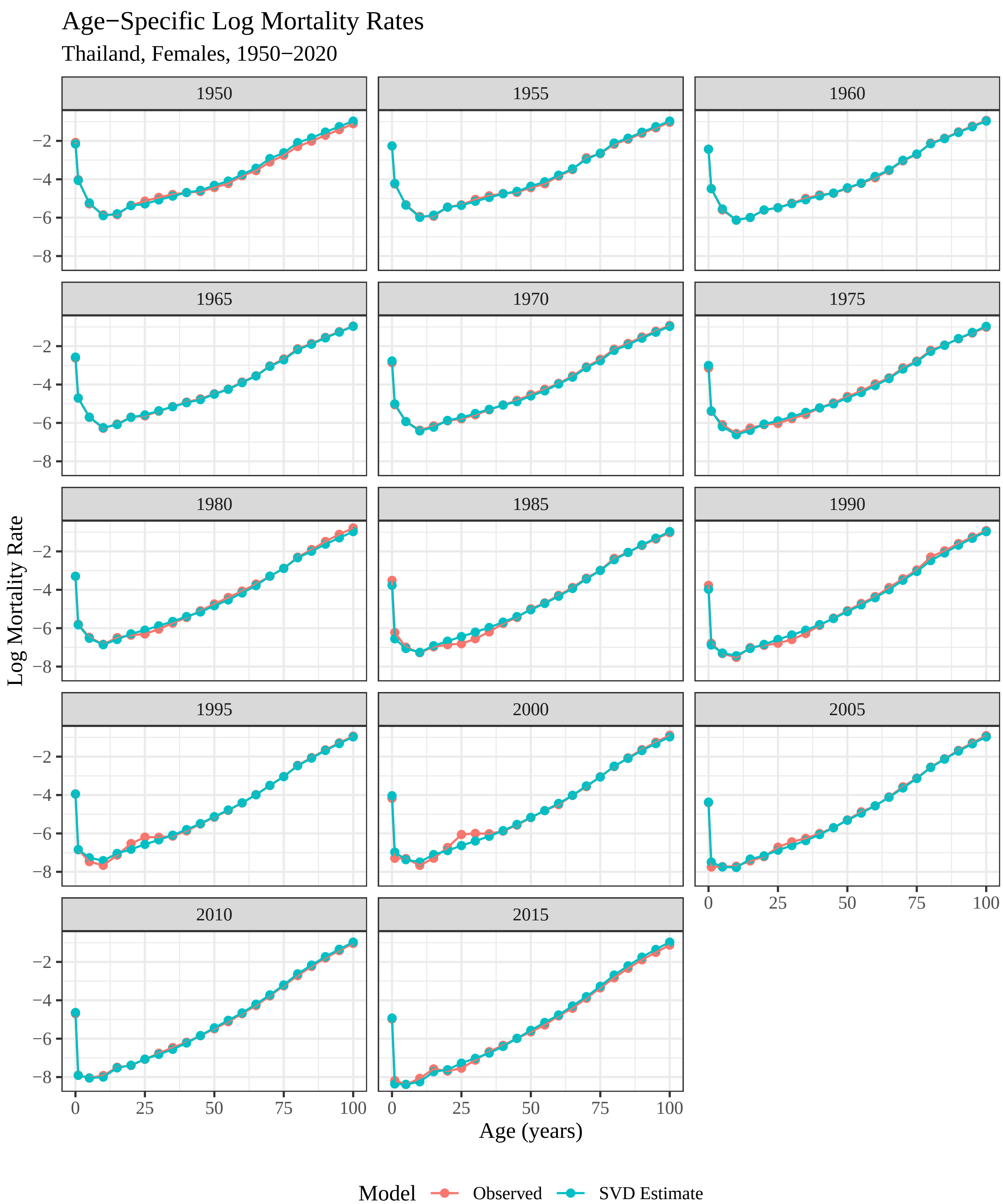
Table 6: Lee-Carter model parameter estimates (SVD method)

Age	b_x
0	0.085
1	0.132
5	0.096
10	0.072
15	0.059
20	0.069
25	0.061
30	0.060
35	0.057
40	0.052
45	0.043
50	0.038
55	0.033
60	0.031
65	0.027
70	0.027
75	0.020
80	0.018
85	0.011
90	0.006
95	0.003
100	0.000

Table 7: Lee-Carter model parameter estimates (SVD method)

Year	k_t
1950	15.106
1955	13.909
1960	11.840
1965	10.235
1970	7.883
1975	5.115
1980	1.693
1985	-3.835
1990	-6.281
1995	-6.002
2000	-7.028
2005	-10.977
2010	-14.120

Year	k_t
2015	-17.538



Comparing the estimated M_x from the normalized SVD to our observed data, we see that the estimates also track the observed data well. Note that if we were to use the unnormalized SVD, the estimated M_x would remain mostly constant over time, matching the middle years the closest (since those years are closest to the mean over time).

Q2.d

Table 8: Lee-Carter model comparison

Age	LS $\hat{b}_x$	SVD $\hat{b}_x$	diff
0	0.084	0.085	0.000
1	0.131	0.132	0.001
5	0.096	0.096	0.000
10	0.071	0.072	0.001
15	0.059	0.059	0.000
20	0.069	0.069	0.000
25	0.061	0.061	0.000
30	0.059	0.060	0.000
35	0.057	0.057	0.000
40	0.052	0.052	0.000
45	0.043	0.043	0.000
50	0.038	0.038	0.000
55	0.033	0.033	0.000
60	0.031	0.031	0.000
65	0.027	0.027	0.000
70	0.027	0.027	0.000
75	0.020	0.020	0.000
80	0.019	0.018	0.000
85	0.011	0.011	0.000
90	0.007	0.006	0.000
95	0.003	0.003	0.000
100	0.000	0.000	0.000

Table 9: Lee-Carter model comparison

Year	LS $\hat{k}_t$	SVD $\hat{k}_t$	diff
1950	13.915	15.106	1.191
1955	13.542	13.909	0.367
1960	11.896	11.840	0.055
1965	10.492	10.235	0.257
1970	8.435	7.883	0.551
1975	5.301	5.115	0.186
1980	2.414	1.693	0.721
1985	-4.338	-3.835	0.502
1990	-5.774	-6.281	0.508
1995	-5.770	-6.002	0.232
2000	-6.479	-7.028	0.548
2005	-10.383	-10.977	0.594
2010	-14.595	-14.120	0.475
2015	-18.656	-17.538	1.118

Table 10: $RMSE$ of estimated $log(M_x)$ against observed $log(M_x)$

Model	RMSE
Least-squares Estimate	4.7938
SVD Estimate	4.7940

Looking at the previous two plots, and comparing the root mean-squared-error of the log-transformed fitted M_x against the observed M_x, we see that, while close, the normalized SVD *Lee-Carter* method performs the best.

Q2.e

The *Lee-Carter* method can also be used to obtain a probabilistic forecast of mortality. Here we use the *demography R* package to forecast mortality for females in Mexico in the period 2020-2025. Below are the predicted mortality index k_t and mortality rate for females in the 75-80 age group. The confidence level is set to 95%.

Table 11: Forecast k_t

Est. k_t	95% Low	95% High
-1.996	-4.551	0.559

Table 12: Forecast mortality for age group 75-80

Est. M_x	95% Low	95% High
0.04	0.038	0.043

Appendix

```r
# Prep work ---------------------------------------------------------------

# Load libraries
library(dplyr)
library(tidyr)
library(tibble)
library(ggplot2)
library(MortalityLaws)
library(demogR)
library(MortCast)
library(demography)

# Helper functions
rmse <- function(x, y) sqrt(mean((x - y)^2))
mx_to_qx <- function(mx, n) 1 - exp(-1 * n * mx)
qx_to_mx <- function(qx, n) -1 * log(1 - qx) / n

# Load data
```

```r
data(mxF, package = "wpp2019")
tha_mort <- mxF %>%
  filter(name == "Thailand") %>%
  select(-country_code, -name) %>%
  pivot_longer(-age, names_to = "period", values_to = "Mx") %>%
  extract(period, "year", regex = "(^[:digit:]{4})", convert = TRUE) %>%
  filter(year < 2020) %>%
  select(year, age, everything()) %>%
  arrange(year, age)


# Question 1 --------------------------------------------------------------

tha_mort_2015 <- tha_mort %>%
  filter(year == 2015) %>%
  select(-year) %>%
  mutate(n = lead(age, default = Inf) - age) %>%
  mutate(qx = 1 - exp(-1 * n * Mx))


# Question 1a -------------------------------------------------------------

plot_log_mort <- function(data, ...) {

  ggplot(data, aes(...)) +
    geom_point() +
    geom_line() +
    theme_bw() +
    theme(
      text = element_text(family = "serif"),
      legend.position = "bottom"
    ) +
    labs(
      title = "Age-Specific Log Mortality Rates",
      subtitle = "Thailand, Females, 2015-2020",
      x = "Age (years)",
      y = "Log Mortality Rate",
      color = "Model"
    )

}

knitr::kable(
  select(tha_mort_2015, age, n, Mx),
  booktabs = TRUE,
  col.names = c("Age", "n", "${}_{n}M_x$"),
  caption = "Thailand 2015-2020 female mortality rates",
  eval = FALSE,
  digits = 3
)

plot_log_mort(tha_mort_2015, x = age, y = log(Mx))
```

```r
# Question 1b ------------------------------------------------------------

model_gompertz2 <- lm(log(Mx) ~ age, data = filter(tha_mort_2015, age >= 50))

model_gompertz <- with(
  filter(tha_mort_2015, age >= 50),
  MortalityLaw(x = age, mx = Mx, law = "gompertz")
)

model_makeham <- with(
  filter(tha_mort_2015, age >= 50),
  MortalityLaw(x = age, mx = Mx, law = "makeham")
)

coef_gompertz <- coef(model_gompertz)
coef_makeham <- coef(model_makeham)


# Question 1c ------------------------------------------------------------

tha_mort_2015_model <- tha_mort_2015 %>%
  rename(obs_qx = qx, obs_Mx = Mx) %>%
  mutate(
    Gompertz = predict(model_gompertz, x = age),
    Makeham = predict(model_makeham, x = age)
  ) %>%
  select(age, n, everything())

tha_mort_2015_model %>%
  select(-obs_qx, -n, age, Observed = obs_Mx, everything()) %>%
  pivot_longer(-age, names_to = "model", values_to = "Mx") %>%
  filter(model %in% c("Observed", "Gompertz", "Makeham")) %>%
  plot_log_mort(x = age, y = log(Mx), color = model)


# Question 1d ------------------------------------------------------------

model_HP_mx <- with(tha_mort_2015, MortalityLaw(x = age, mx = Mx, law = "HP"))
model_HP_qx <- with(tha_mort_2015, MortalityLaw(x = age, qx = qx, law = "HP"))

tha_mort_2015_model <- tha_mort_2015_model %>%
  mutate(
    `Heligman-Pollard` = predict(model_HP_mx, x = age),
    `Heligman-Pollard2` = -1 * log(1 - predict(model_HP_qx, x = age)) / n
  )

coef_HP <- coef(model_HP_mx)

tha_mort_2015_model %>%
  select(-obs_qx, -n, age, Observed = obs_Mx, everything()) %>%
  pivot_longer(-age, names_to = "model", values_to = "Mx") %>%
  filter(model %in% c("Observed", "Heligman-Pollard")) %>%
  plot_log_mort(x = age, y = log(Mx), color = model)
```

```r
# Question 1e -------------------------------------------------------------

coaleDemenyLTW <- demogR::cdmltw(sex = "F")
tha_Mx_0to95 <- tha_mort_2015 %>% filter(age < 100) %>% pull(Mx)

best_match_lt <-
  coaleDemenyLTW[["nmx"]] %>%
  apply(1, function(x) rmse(x, log(tha_Mx_0to95))) %>%
  which.min()

tha_mort_2015_model <- tha_mort_2015_model %>%
  mutate(
    obs_qx_95 = obs_qx,
    standard_qx = c(coaleDemenyLTW[["nqx"]][best_match_lt, ], NA)
  )

tha_mort_2015_model[["obs_qx_95"]][21:22] <- c(1, NA)

model_brass <- lm(
  qlogis(obs_qx_95) ~ qlogis(standard_qx),
  data = filter(tha_mort_2015_model, age < 95)
)

tha_mort_2015_model <- tha_mort_2015_model %>%
  mutate(
    brass_pred_fit_qx = c(plogis(model_brass$fitted.values), 1, NA),
    brass_pred_obs_qx = plogis(predict(model_brass, data.frame(standard_qx = obs_qx))),
    brass_pred_fit_mx = -1 * log(1 - brass_pred_fit_qx) / n
  )

coef_brass <- coef(model_brass)

tha_mort_2015_model %>%
  select(
    -n, age, Observed = obs_qx,
    Brass = brass_pred_fit_qx, Brass2 = brass_pred_obs_qx
  ) %>%
  pivot_longer(-age, names_to = "model", values_to = "qx") %>%
  filter(model %in% c("Observed", "Brass")) %>%
  plot_log_mort(x = age, y = log(qx), color = model) +
  labs(
    title = "Age-Specific Log Probability of Death",
    y = "Log Probability of Death"
  )


# Question 1f -------------------------------------------------------------

tha_mort_2015_compare <- tha_mort_2015_model %>%
  select(
    age, n,
    Observed = obs_Mx,
```

```r
    Gompertz,
    Makeham,
    `Heligman-Pollard`,
    Brass = brass_pred_fit_mx
  )

# knitr::kable(
#   tha_mort_2015_compare,
#   booktabs = TRUE,
#   digits = 3,
#   caption = "Comparison of models against observed Mx in Thailand females, 2015-2020"
# )

tha_mort_2015_compare %>%
  pivot_longer(!c(age, n), names_to = "model", values_to = "Mx") %>%
  filter(model %in% c("Observed", "Gompertz", "Makeham", "Heligman-Pollard", "Brass")) %>%
  plot_log_mort(x = age, y = log(Mx), color = model)


tha_mort_2015_compare_fit <- tha_mort_2015_compare %>%
  filter(age < 95) %>%
  select(-age, -n) %>%
  mutate_all(log) %>%
  summarise_all(~rmse(., Observed)) %>%
  select(-Observed) %>%
  pivot_longer(everything(), names_to = "Model", values_to = "RMSE") %>%
  arrange(RMSE)

knitr::kable(
  tha_mort_2015_compare_fit,
  booktabs = TRUE,
  digits = 3,
  caption = "$RMSE$ of estimated $log(M_x)$ against observed $log(M_x)$"
)

# Question 2 ------------------------------------------------------------

tha_mort_wide <- tha_mort %>% pivot_wider(names_from = year, values_from = Mx)
tha_mort_mat <- tha_mort_wide %>% column_to_rownames("age") %>% as.matrix()


# Question 2a -----------------------------------------------------------

knitr::kable(
  tha_mort_wide,
  booktabs = TRUE,
  digits = 3,
  caption = "Thailand female age-specific mortality rates, 1950-2020"
)


# Question 2b -----------------------------------------------------------
```

```r
model_LC <- leecarter.estimate(tha_mort_mat)

LC_lsq_mx <- with(model_LC, apply(
  matrix(bx) %*% matrix(kt, nrow = 1), 2, function(x) ax + x
))
dimnames(LC_lsq_mx) <- dimnames(tha_mort_mat)

tha_mort_lsq <- LC_lsq_mx %>%
  as_tibble(rownames = "age") %>%
  pivot_longer(-age, names_to = "year", values_to = "log_Mx") %>%
  mutate(
    year = as.integer(year),
    age = as.integer(age),
    `Least-squares Estimate` = exp(log_Mx),
  ) %>%
  select(-log_Mx)

tha_LC_compare <- tha_mort %>%
  left_join(tha_mort_lsq, by = c("age", "year")) %>%
  rename(Observed = Mx)

model_LC_ls_x_tbl <- tibble(
  age = names(model_LC$ax),
  ax = model_LC$ax,
  least_squares_bx = model_LC$bx
)

model_LC_ls_t_tbl <- tibble(
  year = names(model_LC$kt),
  least_squares_kt = model_LC$kt
)

knitr::kable(
  model_LC_ls_x_tbl,
  booktabs = TRUE,
  digits = 3,
  col.names = c("Age", "$\\hat{a}_x$", "$\\hat{b}_x$"),
  eval = FALSE,
  caption = "Lee-Carter model parameter estimates (least squares method)"
)

knitr::kable(
  model_LC_ls_t_tbl,
  booktabs = TRUE,
  digits = 3,
  col.names = c("Year", "$\\hat{k}_t$"),
  eval = FALSE,
  caption = "Lee-Carter model parameter estimates (least squares method)"
)

tha_LC_compare %>%
  pivot_longer(!c(year, age), names_to = "model", values_to = "Mx") %>%
  filter(model %in% c("Observed", "Least-squares Estimate")) %>%
```

```r
  plot_log_mort(x = age, y = log(Mx), color = model) +
  facet_wrap(vars(year), ncol = 3) +
  labs(subtitle = "Thailand, Females, 1950-2020")


# Question 2c -----------------------------------------------------------

ax_hat <- rowMeans(log(tha_mort_mat))
model_LC_svd <- svd(apply(log(tha_mort_mat), 2, function(x) x - ax_hat), 1, 1)

# Get normalized bx and kt
LC_svd_bx <- model_LC_svd$u / sum(model_LC_svd$u)
LC_svd_kt <- t(model_LC_svd$v) * sum(model_LC_svd$u) * model_LC_svd$d[1]

LC_svd_mx <- apply(LC_svd_bx %*% LC_svd_kt, 2, function(x) ax_hat + x)
dimnames(LC_svd_mx) <- dimnames(tha_mort_mat)

tha_mort_svd <- LC_svd_mx %>%
  as_tibble(rownames = "age") %>%
  pivot_longer(-age, names_to = "year", values_to = "log_Mx") %>%
  mutate(
    year = as.integer(year),
    age = as.integer(age),
    `SVD Estimate` = exp(log_Mx),
  ) %>%
  select(-log_Mx)

tha_LC_compare <- tha_LC_compare %>%
  left_join(tha_mort_svd, by = c("age", "year"))

model_LC_svd_x_tbl <- tibble(
  age = names(model_LC$bx),
  svd_bx = as.vector(LC_svd_bx)
)

model_LC_svd_t_tbl <- tibble(
  year = names(model_LC$kt),
  svd_kt = as.vector(LC_svd_kt),
)

knitr::kable(
  model_LC_svd_x_tbl,
  booktabs = TRUE,
  digits = 3,
  col.names = c("Age", "$b_x$"),
  eval = FALSE,
  caption = "Lee-Carter model parameter estimates (SVD method)"
)

knitr::kable(
  model_LC_svd_t_tbl,
  booktabs = TRUE,
  digits = 3,
```

```r
  col.names = c("Year", "$k_t$"),
  eval = FALSE,
  caption = "Lee-Carter model parameter estimates (SVD method)"
)

tha_LC_compare %>%
  pivot_longer(!c(year, age), names_to = "model", values_to = "Mx") %>%
  filter(model %in% c("Observed", "SVD Estimate")) %>%
  plot_log_mort(x = age, y = log(Mx), color = model) +
  facet_wrap(vars(year), ncol = 3) +
  labs(subtitle = "Thailand, Females, 1950-2020")


# Question 2d -------------------------------------------------------------

LC_compare_bx_tbl <- model_LC_ls_x_tbl %>%
  left_join(model_LC_svd_x_tbl, by = "age") %>%
  select(-ax) %>%
  mutate(abs_diff = abs(least_squares_bx - svd_bx))

LC_compare_kt_tbl <- model_LC_ls_t_tbl %>%
  left_join(model_LC_svd_t_tbl, by = "year") %>%
  mutate(abs_diff = abs(least_squares_kt - svd_kt))

tha_LC_compare_fit <- tha_LC_compare %>%
  group_by(year) %>%
  select(-age) %>%
  mutate_at(vars(-group_cols()), log) %>%
  summarise_all(~rmse(., Observed)) %>%
  select(-Observed) %>%
  ungroup() %>%
  select(-year) %>%
  summarize_all(mean) %>%
  pivot_longer(everything(), names_to = "Model", values_to = "RMSE") %>%
  arrange(RMSE)

knitr::kable(
  LC_compare_bx_tbl,
  booktabs = TRUE,
  digits = 3,
  col.names = c("Age", "LS $\\hat{b}_x$", "SVD $\\hat{b}_x$", "diff"),
  eval = FALSE,
  caption = "Lee-Carter model comparison"
)

knitr::kable(
  LC_compare_kt_tbl,
  booktabs = TRUE,
  digits = 3,
  col.names = c("Year", "LS $\\hat{k}_t$", "SVD $\\hat{k}_t$", "diff"),
  eval = FALSE,
  caption = "Lee-Carter model comparison"
)
```

```
knitr::kable(
tha_LC_compare_fit,
booktabs = TRUE,
digits = 4,
caption = "$RMSE$ of estimated $log(M_x)$ against observed $log(M_x)$"
)


# Question 2e -------------------------------------------------------------

data(popF, package = "wpp2019")

mex_pop <- popF %>%
  filter(name == "Mexico") %>%
  select(-country_code, -name, -`2020`) %>%
  extract(age, "age", convert = TRUE) %>%
  column_to_rownames("age") %>%
  as.matrix() %>%
  `*`(1000)

mex_mort_tbl <- mxF %>%
  filter(name == "Mexico") %>%
  select(-country_code, -name) %>%
  pivot_longer(-age, names_to = "period", values_to = "Mx") %>%
  extract(period, "year", regex = "(^[:digit:]{4})", convert = TRUE) %>%
  filter(year < 2020)

# Collapse 0 and 1-4 age group Mx
mex_mort_tbl_u5 <- mex_mort_tbl %>%
  filter(age < 5) %>%
  mutate(
    n = age * 3 + 1,
    sx = 1 - mx_to_qx(Mx, n)
  ) %>%
  group_by(year) %>%
  summarise(sx = prod(sx)) %>%
  mutate(
    age = 0,
    Mx = qx_to_mx(1 - sx, 5)
  ) %>%
  select(age, year, Mx)

mex_mort <- mex_mort_tbl %>%
  filter(age >= 5) %>%
  bind_rows(mex_mort_tbl_u5) %>%
  pivot_wider(names_from = year, values_from = Mx) %>%
  arrange(age) %>%
  column_to_rownames("age") %>%
  as.matrix()

demog_data <- demogdata(
  data = mex_mort,
  pop = mex_pop,
```

```r
  ages = seq(0, 100, 5),
  years = seq(1950, 2015, 5),
  type = "mortality",
  label = "Mexico",
  name = "Female"
)

model_mex_LC <- demography::lca(demog_data)

forecast_mex <- forecast(model_mex_LC, h = 1, level = 95, se = "innovonly")

forecast_mex_kt <- as_tibble(forecast_mex[["kt.f"]])

forecast_mex_mx <- forecast_mex$rate %>%
  as_tibble() %>%
  mutate_all(as.vector) %>%
  mutate(age = rownames(forecast_mex$rate$Female)) %>%
  select(age, Mx = Female, lower_95 = lower, upper_95 = upper)

forecast_mex_mx_75 <- forecast_mex_mx %>% filter(age == 75)

knitr::kable(
  forecast_mex_kt,
  booktabs = TRUE,
  digits = 3,
  eval = FALSE,
  col.names = c("Est. $k_t$", "95% Low", "95% High"),
  caption = "Forecast $k_t$"
)

knitr::kable(
  select(forecast_mex_mx_75, -age),
  booktabs = TRUE,
  digits = 3,
  eval = FALSE,
  col.names = c("Est. $M_x$", "95% Low", "95% High"),
  caption = "Forecast mortality for age group 75-80"
)
```